

Interpretable Machine Learning for NO_x Emission Anomaly Detection in Marine Diesel Engines Using Real Operational Data

Imam Nur Rokhim^{1*}, Teguh Junian Kuswanto², Muhammad Iqbal Firdaus³

(Received: 13 July 2026 / Revised: 16 July 2026 / Accepted: 8 August 2026 / Available Online: 30 September 2026)

Abstract—Nitrogen oxide (NO_x) emissions from marine diesel engines remain a major environmental concern in the maritime industry, requiring effective monitoring strategies under actual operating conditions. This study proposes a Random Forest-based approach for detecting anomalous operating conditions associated with NO_x emissions using operational data collected from an actual merchant vessel during an 11-day voyage. Following data preprocessing and threshold-based operational screening, a total of 347,587 valid observations were retained for model development. The Random Forest classifier was evaluated using Accuracy, Precision, Recall, F1-score, and ROC-AUC. The proposed model achieved an Accuracy of 96.87%, Precision of 97.35%, Recall of 93.18%, F1-score of 95.22%, and a ROC-AUC of 99.39%, demonstrating reliable classification performance for distinguishing normal and anomalous operating conditions. Feature importance analysis revealed that operational parameters contributed differently to the classification process, with ship speed and fuel system parameters exhibiting the greatest influence. An engineering interpretation of these findings indicates that the identified parameters are consistent with established combustion mechanisms associated with thermal NO_x formation in marine diesel engines. The proposed approach extends the application of machine learning beyond anomaly detection by providing an interpretable framework that links data-driven classification results with engineering knowledge, offering potential support for onboard condition monitoring and operational decision-making in NO_x emission management.

Keywords—Random Forest; anomaly detection; feature importance; NO_x emission; marine diesel engine; ship operational data.

*Corresponding Author: imamrokhim@ppns.ac.id

I. INTRODUCTION

Maritime transportation plays a fundamental role in the global economy, carrying more than 80% of the world's international trade by volume. Despite being recognized as one of the most energy-efficient transportation modes, the shipping industry remains a significant contributor to atmospheric pollution through exhaust emissions from marine diesel engines. Among these pollutants, nitrogen oxides (NO_x) have attracted considerable attention due to their adverse impacts on air quality, human health, and the environment. NO_x emissions contribute to the formation of acid rain, photochemical smog, and secondary particulate matter, making their reduction a major objective in the pursuit of sustainable maritime transportation [1]. Consequently, international regulations governing marine engine emissions have become increasingly stringent, encouraging ship operators and engine manufacturers to develop more effective emission monitoring and control strategies [2].

The amount of NO_x emitted by a marine diesel engine is influenced not only by engine design but also

by the operational conditions experienced throughout a voyage. Variations in ship speed, engine load, fuel supply rate, fuel temperature, and combustion characteristics continuously alter the combustion process, resulting in dynamic fluctuations in NO_x emissions [3]. These operational parameters interact in a highly nonlinear manner, making conventional analytical approaches insufficient for accurately characterizing emission behavior. Meanwhile, modern onboard monitoring systems continuously record large volumes of operational data with high temporal resolution, providing valuable information regarding engine performance and environmental compliance. However, manually identifying abnormal operating conditions from such large and multidimensional datasets has become increasingly challenging [4].

The rapid advancement of machine learning has provided new opportunities for extracting meaningful information from large-scale operational datasets collected from marine systems. Numerous studies have successfully employed machine learning techniques for fuel consumption prediction, engine performance estimation, fault diagnosis, predictive maintenance, machinery condition monitoring, and anomaly detection in marine propulsion systems [5]. Algorithms such as Artificial Neural Networks (ANN), Support Vector Machines (SVM), eXtreme Gradient Boosting (XGBoost), Random Forests, and various deep learning models have demonstrated excellent capability in capturing nonlinear relationships among operational parameters [6]. In particular, data-driven approaches have become increasingly attractive because they are capable of identifying complex operational patterns that are often difficult to describe using conventional

Imam Nur Rokhim is with Departement of Shipbuilding Engineering, Politeknik Perkapalan Negeri Surabaya, Surabaya, 61111, Indonesia. E-mail: imamrokhim@ppns.ac.id

Teguh Junian Kuswanto is with Departement of Shipbuilding Engineering, Politeknik Perkapalan Negeri Surabaya, Surabaya, 61111, Indonesia. E-mail: tjunian@ppns.ac.id

Muhammad Iqbal Firdaus is with Departement of Shipbuilding Engineering, Politeknik Perkapalan Negeri Surabaya, Surabaya, 61111, Indonesia. E-mail: m.firdaus@ppns.ac.id

physics-based or rule-based methods [7].

The Random Forest algorithm was selected because of its capability to model complex nonlinear relationships among operational parameters while maintaining high robustness against overfitting [8]. Unlike single decision trees, Random Forest constructs multiple decision trees using bootstrap sampling and random feature selection, thereby improving classification stability and generalization performance

[9][10]. Another important advantage of Random Forest is its ability to estimate feature importance directly during the training process, allowing not only accurate anomaly classification but also interpretation of the relative contribution of each operational parameter [11]. These characteristics make Random Forest particularly suitable for analyzing multidimensional operational data collected from marine propulsion systems.

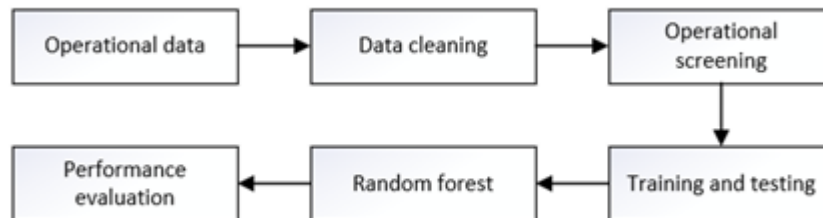


Figure. 1. Overall Research Methodology

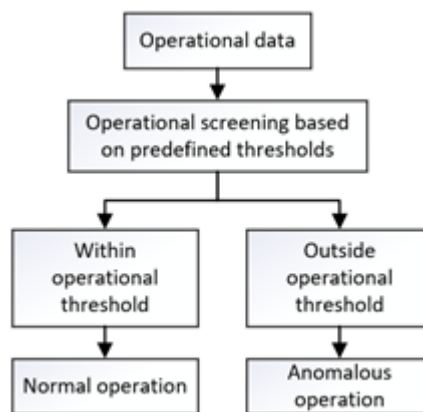


Figure. 2. Operational Screening Procedure Based on Predefined Thresholds

Previous studies on anomaly detection have primarily focused on identifying machinery faults, abnormal fuel consumption, sensor failures, or equipment degradation using operational data collected from marine engines [12]. Various machine learning algorithms, including Isolation Forest, Local Outlier Factor (LOF), K-Nearest Neighbors (KNN), ensemble learning methods, and Random Forest, have demonstrated promising performance in distinguishing normal and abnormal operating conditions [13]. Most of these studies, however, have concentrated on improving classification accuracy through evaluation metrics such as accuracy, precision, recall, F1-score, and the area under the receiver operating characteristic curve (ROC-AUC) [5]. Although these performance indicators demonstrate the effectiveness of the developed models, they provide limited insight into the engineering mechanisms responsible for the detected anomalies.

Similarly, existing studies concerning NOx emissions have mainly focused on emission prediction, combustion optimization, or compliance assessment using laboratory experiments, numerical simulations, or machine learning models [14][15]. While these approaches have achieved high predictive performance, machine learning is frequently employed as a black-box prediction tool, with

limited effort devoted to explaining how operational parameters contribute to the occurrence of NOx emission anomalies [16]. Consequently, the engineering significance of the developed models often remains unclear, despite their high predictive capability[17]. Understanding which operational parameters most strongly influence NOx anomalies is essential not only for improving model interpretability but also for supporting operational decision-making and developing intelligent emission monitoring systems for marine diesel engines [18][19].

Based on the existing literature, three major research gaps can be identified. First, most machine learning studies emphasize predictive performance rather than providing engineering interpretations of the detected anomalies [20]. Second, investigations utilizing real operational data collected from commercial ships during actual voyages remain relatively limited because of restricted access to industrial datasets. Third, the application of feature importance analysis to explain the influence of operational parameters on the formation of NOx emission anomalies has received little attention [21]. As a result, previous studies generally conclude with model performance evaluation without establishing a meaningful relationship between machine learning

findings and the physical mechanisms governing NO_x formation during marine engine operation.

TABLE 1.
OPERATIONAL PARAMETERS USED IN THIS STUDY

Parameter	Unit	Description
Speed Through Water	105	Vessel speed relative to the surrounding water
ME Supply Fuel Flow	L/h	Fuel flow supplied to the main engine
ME Return Fuel Flow	L/h	Fuel flow returning from the fuel circulation system
ME Supply Fuel Temperature	°C	Fuel temperature measured at the engine supply line
ME Return Fuel Temperature	°C	Fuel temperature measured at the engine return line
NO _x Emission	g/kWh	Nitrogen oxide emission generated by the main engine

II. METHOD

Therefore, this study proposes a Random Forest-based classification model for detecting NO_x emission anomalies using real operational data collected from a commercial vessel during an actual voyage. The operational dataset consists of ship speed, fuel flow rate, fuel temperature, shaft torque, and NO_x emission measurements continuously recorded through the vessel's onboard monitoring system. Unlike previous studies, this work extends beyond conventional model performance evaluation by employing feature importance analysis to identify the operational parameters that contribute most significantly to NO_x emission anomalies. Furthermore, the obtained findings are interpreted based on the combustion characteristics of marine diesel engines to establish the relationship between operational conditions and NO_x formation mechanisms. The proposed approach not only demonstrates the effectiveness of machine learning for anomaly detection but also provides engineering insights that can support the development of intelligent real-time emission monitoring and operational decision-support systems for marine applications.

This study employed operational data collected from an international merchant vessel, hereafter referred to as the Case Vessel, during an 11-day voyage. The dataset was continuously recorded using an onboard operational monitoring system that captures key propulsion and engine performance parameters throughout normal ship operations [22]. To preserve the confidentiality of the vessel and the data acquisition system, specific information regarding the ship identity and monitoring equipment is intentionally omitted. The overall research methodology adopted in this study is illustrated in Figure 1.

The operational dataset consisted of seven parameters representing the propulsion and fuel system conditions, namely Speed Through Water (STW), Main Engine (ME) Supply Fuel Flow, ME Return Fuel Flow, ME Supply Fuel Temperature, ME Return Fuel Temperature, Shaft Torque, and NO_x emission. These parameters were selected because they directly reflect engine operating conditions that influence the combustion process and consequently affect NO_x formation. A detailed

TABLE 2.
OPERATIONAL PARAMETERS USED IN THIS STUDY

Parameter	Lower Threshold	Upper Threshold	Threshold Basis
Speed Through Water	0.58	13.36	Engineering operational criterion
ME Supply Fuel Flow	2480	3922	Operational data distribution
ME Return Fuel Flow	2059	3211	Operational data distribution
ME Supply Fuel Temperature	123.145	133.401	Operational data distribution
ME Return Fuel Temperature	119.265	129.078	Operational data distribution
NO _x Emission	0.000	11.966	Engineering operational criterion

As presented in Table 2, the operational thresholds were established according to the characteristics of each monitored parameter. The thresholds for Speed Through Water and NO_x emission were determined based on engineering operational criteria because both parameters are governed by operational requirements and regulatory

considerations during ship operation. The lower speed threshold (0.58 knots) was selected to exclude maneuvering and near-stationary conditions, whereas the upper threshold (13.36 knots) represents the maximum operating speed observed under normal voyage conditions. Likewise, the upper NO_x threshold (11.966

g/kWh) was defined according to the acceptable emission level adopted in this study to distinguish normal operation from abnormal emission conditions.

In contrast, the thresholds for ME Supply Fuel Flow, ME Return Fuel Flow, ME Supply Fuel Temperature, and ME Return Fuel Temperature were derived from the distribution of the cleaned operational dataset. Since no universally accepted operational limits are available for these parameters under actual ship operating conditions, the observed operational ranges were adopted to represent normal engine behavior during the monitoring period. Consequently, observations falling outside these ranges were considered indicative of abnormal operational conditions and were assigned as anomalous during the adaptive labeling process.

Each observation was subsequently assigned an operational label according to the screening results. Observations with all parameters remaining within their respective operational thresholds were classified as Normal Operation, whereas observations containing one or more parameters outside the predefined thresholds were categorized as Anomalous Operation. The adaptive labeling procedure is illustrated in Figure 2, and the resulting labels were subsequently used as the target classes for Random Forest model development.

III. RESULTS AND DISCUSSION

A. Dataset Characteristics

The operational dataset used in this study was collected continuously over an 11-day voyage, providing a comprehensive representation of the vessel's operating conditions under actual service. Prior to model development, the raw dataset underwent a series of preprocessing procedures, including the removal of duplicate records, missing values, and observations

containing zero values. Zero-value observations were excluded because they generally correspond to engine shutdown conditions or invalid sensor readings that do not represent actual engine operation. Following the preprocessing stage, a total of 347,587 valid observations were retained for subsequent analysis. The large number of observations provides sufficient operational variability to capture different engine operating conditions, thereby improving the robustness of the developed machine learning model.

Subsequently, each observation was automatically labeled using the threshold-based operational screening procedure described in Method. Observations with all monitored parameters remaining within their respective operational thresholds were classified as Normal Operation, whereas observations containing one or more parameters outside the predefined thresholds were classified as Anomalous Operation. This labeling process established the ground truth required for supervised learning. To preserve the temporal characteristics of the operational data and to better represent practical implementation, the labeled dataset was divided chronologically into 278,069 observations (79.998%) for model training and 69,518 observations (20.002%) for model testing.

The distribution of the training and testing datasets is presented in Figure 3, while a summary of the processed dataset is provided in Table 3. The adopted data split provides sufficient observations for model learning while maintaining an independent testing dataset for evaluating the model's generalization capability. Moreover, the chronological splitting strategy minimizes the possibility of data leakage between the training and testing phases, resulting in a more realistic assessment of model performance under actual ship operating conditions.

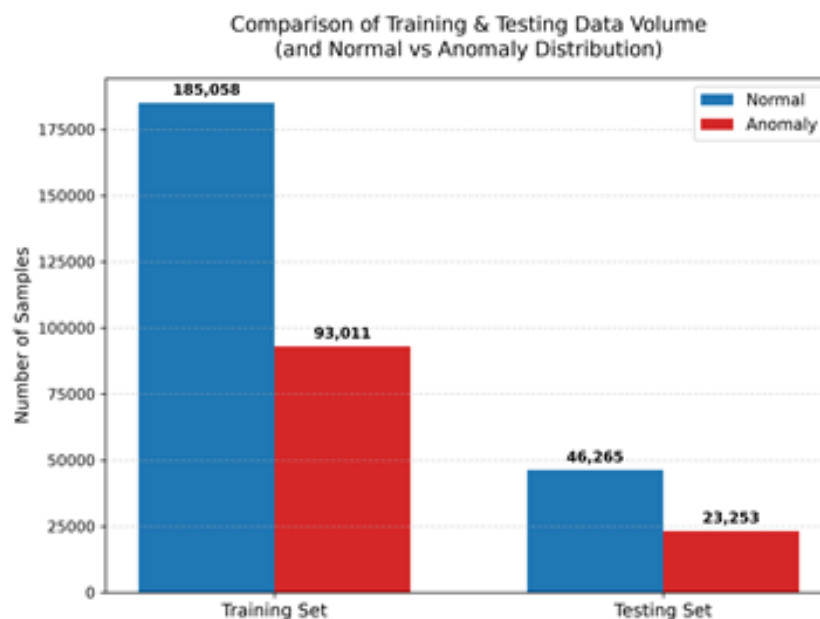


Figure 3. Training–Testing Distribution

TABLE 3.
OPERATIONAL PARAMETERS USED IN THIS STUDY

Description	Value
Observation period	11 days
Total valid observations	347587
Training dataset	278069
Testing dataset	69518
Training proportion	79.998%
Testing proportion	20.002%

B. Performance of Random Forest Classification

The performance of the developed Random Forest classifier was evaluated using five widely adopted classification metrics, namely Accuracy, Precision, Recall, F1-score, and the Area Under the Receiver Operating Characteristic Curve (ROC-AUC). The evaluation results are summarized in Figure 5, while the

corresponding confusion matrix and ROC curve are presented in Figures 4 and 6, respectively. Overall, the proposed model achieved high classification performance, indicating its capability to effectively distinguish between Normal Operation and Anomalous Operation based on the selected operational parameters.

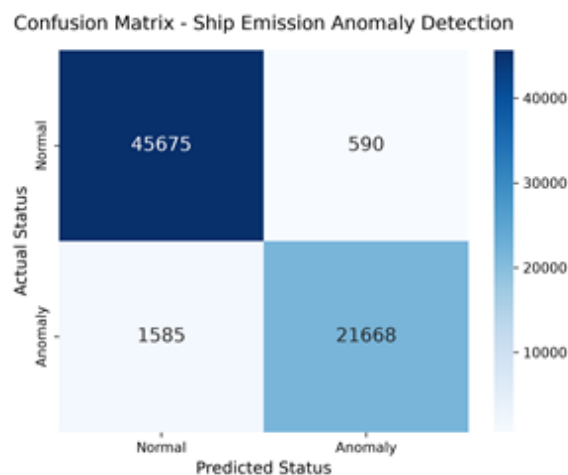


Figure 4. Confusion matrix of the Random Forest classifier.

The confusion matrix presented in Figure 4 provides a detailed visualization of the classification results for both operational classes. The large number of correctly classified observations (True Positives and True Negatives) demonstrates that the model successfully learned the operational patterns contained in the training dataset. Meanwhile, the relatively small number of False Positives and False Negatives indicates that only a limited number of observations were misclassified. From

an operational perspective, minimizing false-negative predictions is particularly important because undetected anomalous operating conditions may delay corrective actions and potentially increase engine emissions or reduce operational efficiency. Conversely, a small number of false-positive predictions is generally acceptable since they primarily trigger additional inspections rather than overlooking actual anomalies.

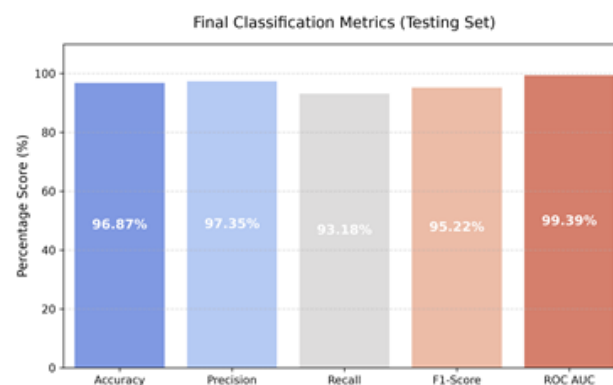


Figure 5. Classification performance metrics of the Random Forest model.

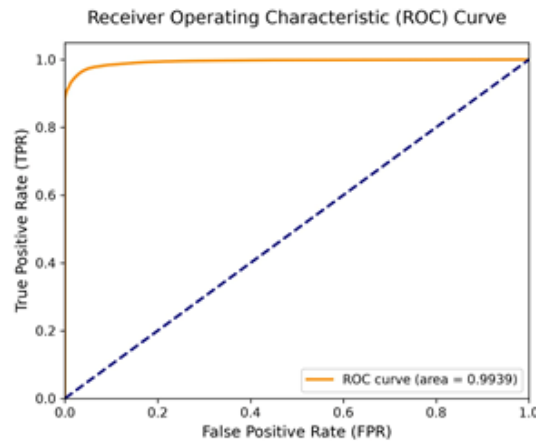


Figure. 6. Receiver Operating Characteristic (ROC) curve of the proposed model.

The ROC curve shown in Figure 6 further confirms the discrimination capability of the developed Random Forest model. The ROC curve is located close to the upper-left corner of the graph, indicating a favorable balance between the True Positive Rate and False Positive Rate over various decision thresholds. Consistent with this observation, the high ROC-AUC value demonstrates that the proposed model possesses strong generalization capability and can reliably distinguish between normal and anomalous operating conditions. These results suggest that the developed classifier is sufficiently robust for anomaly detection using real operational data collected during actual ship operations.

In addition to its classification capability, the Random Forest algorithm provides feature importance scores that quantify the relative contribution of each operational parameter to the classification process. The feature importance values are calculated based on the cumulative reduction of Gini impurity contributed by each feature across all decision trees within the forest. Consequently, parameters that consistently provide greater improvements in node purity during tree construction receive higher importance scores. Although feature importance does not indicate a direct causal relationship, it provides valuable information regarding which operational variables contribute most significantly to distinguishing between Normal Operation and Anomalous Operation.

C. Feature Importance Analysis

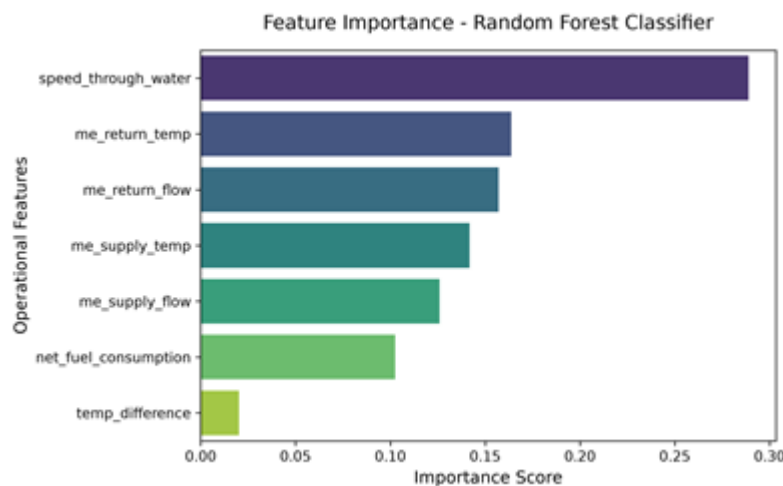


Figure. 7. Feature importance of the operational parameters obtained from the Random Forest model.

The feature importance results are presented in Figure 7. As shown, each operational parameter contributes differently to the developed classification model. Parameters with higher feature importance values were more influential in determining the classification outcome, whereas parameters with lower values contributed less to the decision-making process.

These findings indicate that the operational parameters do not equally influence the occurrence of anomalous operating conditions. Therefore, the

parameters with the highest feature importance values were further analyzed from an engineering perspective to explain their relationship with the physical mechanisms governing NO_x formation in marine diesel engines.

D. Engineering Interpretation of NO_x Anomaly Formation

Based on the feature importance ranking presented in Figure 7, the operational parameters are discussed in descending order according to their contribution to

anomaly classification. The results indicate that the occurrence of NO_x emission anomalies is not governed by a single operational parameter but rather by the combined effects of ship operating conditions, fuel system characteristics, and combustion behavior. Therefore, the engineering interpretation presented in this section focuses on the dominant parameters identified by the Random Forest model and their relationship with the physical mechanisms responsible for NO_x formation.

Among the evaluated parameters, Speed Through Water (STW) exhibited the highest contribution to anomaly classification. From an engineering perspective,

variations in ship speed directly affect the propulsion power demand and consequently alter the engine load [23]. As engine load increases, a greater quantity of fuel is injected into the combustion chamber, resulting in higher in-cylinder pressure and combustion temperature. Elevated combustion temperatures promote the formation of thermal NO_x through the extended Zeldovich mechanism, which is recognized as the dominant NO_x formation pathway in marine diesel engines operating under high-temperature conditions. The relationship between ship speed, engine load, combustion temperature, and NO_x formation is illustrated in Figure 8.



Figure 8. Relationship between ship speed, engine load, combustion temperature, and NO_x formation.

The next influential parameters are associated with the fuel system, particularly fuel temperature and fuel flow. Fuel temperature strongly influences fuel viscosity, which subsequently affects spray atomization, fuel-air mixing, and combustion quality [24]. Maintaining the fuel temperature within its recommended operating range

improves atomization and promotes more stable combustion. Conversely, deviations from the optimum fuel temperature may deteriorate combustion characteristics, leading to localized high-temperature regions that increase NO_x formation. This combustion mechanism is conceptually illustrated in Figure 9.



Figure 9. Relationship between fuel temperature, fuel atomization, combustion quality, and NO_x formation.

Similarly, fuel flow reflects the energy demand of the engine under different operating conditions. An increase in fuel flow generally corresponds to higher engine load and greater heat release during combustion. Under efficient combustion conditions, the resulting increase in peak flame temperature favors thermal NO_x formation

[24]. On the other hand, abnormal variations in fuel flow may indicate irregularities in the fuel supply system, potentially causing unstable combustion and abnormal emission behavior. The relationship between fuel flow, heat release, combustion temperature, and NO_x formation is presented in Figure 10.



Figure 10. Relationship between fuel flow, heat release, combustion temperature, and NO_x formation.

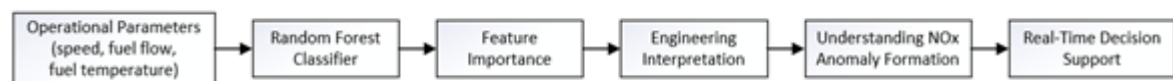


Figure 11. Proposed framework integrating feature importance analysis with engineering interpretation to support NO_x anomaly monitoring and operational decision-making.

Overall, the engineering interpretation demonstrates that the operational parameters identified by the Random Forest model are closely associated with the fundamental mechanisms governing NO_x formation in marine diesel engines. Rather than serving solely as statistical indicators, the feature importance results provide meaningful engineering insights into the operational factors that contribute to emission anomalies. Consequently, the proposed approach not only improves the interpretability of the machine learning model but also establishes a direct connection between data-driven anomaly detection and the physical combustion

processes occurring during ship operation.

Unlike previous studies that primarily employed machine learning to improve predictive accuracy or estimate NO_x emissions, the present study extends the role of machine learning by providing an engineering interpretation of anomaly formation through feature importance analysis [25]. Rather than treating feature importance solely as a statistical ranking, the identified operational parameters were systematically related to established combustion mechanisms governing thermal NO_x formation. This approach enables the classification results to be interpreted in terms of

physical engine behavior, thereby bridging data-driven anomaly detection with engineering knowledge [6]. Consequently, the proposed framework not only improves model interpretability but also provides practical information that can support operational decision-making for onboard emission monitoring.

E. Practical Implications

This study demonstrates that feature importance analysis derived from the Random Forest model can be used as an approach to identify the operational parameters that contribute most significantly to the classification of normal and anomalous operating conditions. The resulting information provides a basis for engineering interpretation of the potential relationship between operational parameters and the occurrence of NOx emission anomalies. The proposed implementation framework is illustrated in Figure 11.

IV. CONCLUSION

This study proposed a Random Forest-based approach for detecting anomalous operating conditions associated with NOx emissions using operational data collected from an actual merchant vessel. A total of 347,587 valid operational observations were obtained after data preprocessing and adaptive threshold-based labeling. The developed Random Forest classifier demonstrated excellent classification performance, achieving an Accuracy of 96.87%, Precision of 97.35%, Recall of 93.18%, F1-score of 95.22%, and ROC-AUC of 99.39%, indicating its capability to reliably distinguish between normal and anomalous operating conditions using real ship operational data.

Beyond the classification performance, the feature importance analysis revealed that the operational parameters contributed differently to anomaly detection, with ship speed and fuel system parameters exhibiting the greatest influence on the classification results. The subsequent engineering interpretation suggested that these parameters are closely associated with engine load, fuel injection characteristics, combustion temperature, and the mechanisms governing thermal NOx formation. Therefore, the proposed approach extends the application of machine learning beyond anomaly detection by providing an interpretable framework that links data-driven classification results with engineering knowledge. This framework has the potential to support onboard condition monitoring and operational decision-making for improved NOx emission management in marine diesel engines. Future studies are recommended to validate the proposed approach using operational datasets collected from different vessel types and engine configurations to further improve its generalization capability.

REFERENCES

- [1] P. Ni, X. Wang, and H. Li, "A review on regulations, current status, effects and reduction strategies of emissions for marine diesel engines," *Fuel*, vol. 279, p. 118477, Nov. 2020, doi: 10.1016/j.fuel.2020.118477.
- [2] A. Fan *et al.*, "Analysing ship emissions under complex operating conditions: Insights from onboard measurement data," *Mar. Pollut. Bull.*, vol. 209, p. 117280, 2024, doi: <https://doi.org/10.1016/j.marpolbul.2024.117280>.
- [3] A. Y. H. Noura, A. E. L. Amrani, and E. El Adel, "A Survey on Data-Driven Fault Diagnostic Techniques for Marine Diesel Engines."
- [4] M. Tadros and E. Boulougouris, "Calibration Procedures for NOx Emissions Model of a High-Speed Marine Diesel Engine Using Optimization Procedures," *Journal of Marine Science and Engineering*, vol. 13, no. 8, p. 1585, 2025, doi: 10.3390/jmse13081585.
- [5] 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," *J. Ocean Eng. Sci.*, vol. 11, no. 1, pp. 278–296, 2026, doi: <https://doi.org/10.1016/j.joes.2025.11.011>.
- [6] A. T. Hoang, T. Anh, E. Bui, X. P. Nguyen, V. H. Bui, and Q. C. Nguyen, "Explainable machine learning-based prediction of fuel consumption in ship main engines using operational data," vol. 76, no. 4, pp. 1–24, 2025.
- [7] D. Upadrashta and T. Wijaya, "AI/ML Based Anomaly Detection and Fault Diagnosis of Turbocharged Marine Diesel Engines: Experimental Study on Engine of an Operational Vessel," *Information*, vol. 17, no. 1, p. 16, 2026, doi: 10.3390/info17010016.
- [8] E. Cahya, W. Pribadi, A. Ghofur, R. Subagyo, and A. Syarif, "A Machine Learning-Based Approach for Designing SEEMP on Ships: Case Study of CO₂ Emissions at a Container Port," vol. 10, no. 4, pp. 1218–1227, 2025.
- [9] V. Şahin, "Explainable machine learning-based prediction of CO₂ emissions from passenger vessels," *Ocean Eng.*, vol. 341, p. 122752, 2025, doi: <https://doi.org/10.1016/j.oceaneng.2025.122752>.
- [10] N. Doanh, N. Dang, and K. Pham, "Analysis of weather and ship-type effects on fuel efficiency and emissions for green maritime operations," vol. 3, no. 2, pp. 1–10, 2025.
- [11] S. Adiyono and T. Hidayat, "Predictive Health Monitoring of a Marine Propulsion Plant Using Multi-Output Random," vol. 11, no. 1, pp. 248–258, 2026.
- [12] M. Michel, A. Mehanna, S. N. Saleh, and A. S. Shehata, "Proactive fault prediction in marine diesel engines using multivariate machine learning," pp. 1–30, 2026.
- [13] H. Je-Gal *et al.*, "Time-Series Explanatory Fault Prediction Framework for Marine Main Engine Using Explainable Artificial Intelligence," *Journal of Marine Science and Engineering*, vol. 12, no. 8, p. 1296, 2024, doi: 10.3390/jmse12081296.
- [14] Q. Shen, G. Wang, Y. Wang, B. Zeng, X. Yu, and S. He, "Prediction Model for Transient NOx Emission of Diesel Engine Based on CNN-LSTM Network," *Energies*, vol. 16, no. 14, p. 5347, 2023, doi: 10.3390/en16145347.
- [15] Fouzi Harrou, Abdelhakim Dorbane, and Ying Sun, "Explainable machine learning with SMOTE for predictive maintenance of ship engines," *Proc. Inst. Mech. Eng. Part M J. Eng. Marit. Environ.*, p. 14750902251400372, Dec. 2025, doi: 10.1177/14750902251400373.
- [16] M. B. Sanjeevannavar *et al.*, "Machine Learning Prediction and Optimization of Performance and Emissions Characteristics of IC Engine," *Sustainability*, vol. 15, no. 18, p. 13825, 2023, doi: 10.3390/su151813825.
- [17] T. Žvirblis, K. Čiziūnienė, and J. Matijošius, "Application of Machine Learning for Fuel Consumption and Emission Prediction in a Marine Diesel Engine Using Diesel and Waste Cooking Oil," *Journal of Marine Science and Engineering*, vol. 13, no. 7, p. 1328, 2025, doi: 10.3390/jmse13071328.
- [18] H. P. Selvam, B. Jayaprakash, Y. Li, and S. Shekhar, "Physics-based machine learning framework for predicting NOx emissions from compression ignition engines using on-board diagnostics data," pp. 1–36.
- [19] O. Yuksel, M. Bayraktar, and M. Sokuc, "Comparative study of machine learning techniques to predict fuel consumption of a marine diesel engine," *Ocean Eng.*, vol. 286, p. 115505, 2023, doi: <https://doi.org/10.1016/j.oceaneng.2023.115505>.
- [20] H. N. K. A. kai Z. K. N. Modabberian, "Machine Learning Methods Methods for for Emissions Emissions Machine Machine Learning Methods for Emissions Machine Learning Methods for Prediction in Combustion with Machine Learning Methods Engines for Emissions Emissions Prediction in Combusti," vol. 2, pp. 3072–3078, 2023, doi:

- 10.1016/j.ifacol.2023.10.1437.
- [21] Samet Memis; Ramazan Şener, "Classification of NO_x Emission in Marine Engines Utilizing kNN-Based Machine Learning Algorithms," no. 02, pp. 74–79, 2024.
 - [22] S. Lim and J. Oh, "Real-time emission prediction for ship engine using stacked time- series learning : a transformer- XGBoost hybrid framework," pp. 1–16, 2025.
 - [23] J. Deng, X. Wang, Z. Wei, L. Wang, C. Wang, and Z. Chen, "A review of NO_x and SO_x emission reduction technologies for marine diesel engines and the potential evaluation of liquefied natural gas fuelled vessels," *Sci. Total Environ.*, vol. 766, p. 144319, 2021, doi: <https://doi.org/10.1016/j.scitotenv.2020.144319>.
 - [24] K.-S. Jo, K.-J. Kong, and S.-H. Han, "Comparative Analysis of Combustion Characteristics and Emission Formation in Marine Diesel Engines Using Biofuels: Chemical Mechanism Analysis and Computational Fluid Dynamics Simulation," *Journal of Marine Science and Engineering*, vol. 13, no. 6. p. 1098, 2025, doi: 10.3390/jmse13061098.
 - [25] D. Kim, M. P. Handayani, S. Lee, and J. Lee, "Feature Attribution Analysis to Quantify the Impact of Oceanographic and Maneuverability Factors on Vessel Shaft Power Using Explainable Tree-Based Model.," *Sensors (Basel)*, vol. 23, no. 3, Jan. 2023, doi: 10.3390/s23031072.