Indian Journal of Information Sources and Services
ISSN: 2231-6094 (P) Vol.15, No.4, 2025, pp.177-182

DOI: https://doi.org/10.51983/ijiss-2025.IJISS.15.4.20

Investigating Data Science Approaches for Predictive Maintenance in Maritime Fleet Operations

Majendiran Gopinath[1*] and M. Muthazhagu[2]

[1*]Department of Marine Engineering, AMET University, Kanathur, Tamil Nadu, India
[2]Department of Marine Engineering, AMET University, Kanathur, Tamil Nadu, India
E-mail: [1]mahandran@ametuniv.ac.in, [2]muthazhagumugilan@ametuniv.ac.in
ORCID: [1]https://orcid.org/0009-0008-9453-2150, [2]https://orcid.org/0009-0001-1792-8257
(Received 17 August 2025; Revised 02 October 2025, Accepted 20 October 2025; Available online 15 December 2025)

***Abstract* - Operating a maritime fleet impacts global trade and commerce, yet it is often accompanied by issues related to equipment dependability, reliability, and maintenance costs. Traditional reactive or scheduled maintenance methods usually result in considerable resource idleness and unplanned system downtime. This paper focuses on employing data science techniques to facilitate predictive maintenance on maritime vessels. Utilizing the onboard systems' historical and real-time sensor data, we assess the efficacy of Random Forest algorithms, Support Vector Machine algorithms, and Long Short-Term Memory (LSTM) networks in fault detection and failure prediction. The proposed framework comprises data cleansing, feature selection, model training, and evaluation conducted using authentic data obtained from commercial ships. The experimental results validate the assertion that data-driven maintenance models can forecast failure of specific ship components well within the stipulated time, thus aiding planned maintenance and minimizing operational halts. These results underline the importance of predictive analysis on operational maintenance, enhancing fleet dependability, safety, and cost-efficiency in maritime operations.**

Keywords: **Predictive Maintenance, Maritime Fleet, Data Science, Machine Learning, Fault Detection, Condition Monitoring**

I. INTRODUCTION

The world's oceans are the highways of international trade, with maritime fleets as the backbone of commercial activities. Fleets move an estimated 90% of global imports and exports. The shipping industry is a section of an extensive logistical network that relies heavily on the prognosis of equipment aboard vessels, such as maintenance control, the ship's propeller, pumps, engines, and navigational systems. Any disturbance in these classifying structures may result in dire delays, considerable risk to safety, and financial loss. It should be made clear that, while shoreline interface equipment interface refits must be kept at a reasonable level, the actual and dynamic posture of active maritime safety law compliant building guidance and obstructions is inherently more reasonable and favorable from an economic point of view. There is a need to highlight that the combined bordering lull of safety contours fences lively business activities that are blurring to breathe moderately free, mask deep drown areas. Contemporary marine endeavors remain subservient to brutal carve-out schedule-driven preemptive maintenance and response. It is an active system from which maintenance is carried out on the components after utilizing the system, only referred to as, when necessary, once complete functional breakdowns occur. Implementing a solution to an already operational functional system aids clearly defined for deficiencies but vastly increases expenditure of operational time, expensive repairs, and poses more severe risks to personnel and state decrees on ecological relation interactions within the controlled environment.

It is apparent that time-driven, schedule-orchestrated, and mark-driven pre-emptive busywork overlooks the actual prevailing battle. This approach, even though it remains proactive, may cause inefficiencies by replacing components that are still functioning or not identifying small issues that can grow into large problems and failures.

The predictive maintenance (PdM) approach is being embraced as a smarter maritime and logistic condition-based strategy to tackle these difficulties (Simion et al., 2024). Unlike routine maintenance, PdM scheduling emphasizes real-time evaluation of machinery's "health." Scheduled actions are only performed when data suggests that a failure is imminent. Predictive maintenance is advantageous for optimization overload because it allows better operational planning, lower maintenance costs, prolonged equipment life cycles, and decreased unplanned downtimes. The guiding predictive maintenance technologies include IoT, big data analytics, and machine learning AI (Kaushik & Tiwari, 2023). Equipment equipped with sensors produces a wealth of data regarding its performance metrics, including vibration, temperature, and pressure (Vij & Prashant, 2024; Jung et al., 2019). ML sophisticated algorithms are developed to "analyze" data for signals of component deterioration or failure (Rakesh et al., 2024). The information equips fleet managers and maintenance crews with actionable insights on failure prediction, root cause analysis, and anomaly detection (Capobianco et al., 2021; Nascimento & Viana, 2019). The maritime industry now has more accessible and scalable PdM solutions due to recent developments in artificial intelligence and edge computing. The availability of processing data in real-time enhances decision-making agility, while a cloud system enables analytics and benchmarking at the fleet level. Moreover, digitized systems are augmented with the robotics virtual twin technology, which enables simulation-centered diagnostics and predictive capabilities, sustaining twin systems' virtual replicas of entire systems, while delivering an all-inclusive comprehension of the system's health and performance indicators (Liu et al., 2022; Mohammed & Hussein, 2023). With everything said, maritime assets are undergoing a profound transformation in maintenance and asset management with predictive maintenance. A shipping company shifts from reacting and taking anticipatory measures on strategically planned intervals to condition-based, real-time data analytics, fundamentally improving fleet safety, reliability, and cost-efficiency (Štepec et al., 2020). The modernization and digitalization of the maritime industry will drive the implementation of predictive maintenance solutions for meeting the global trade, environmental, and regulatory compliance challenges for the coming decades (Cao & Jiang, 2024).

Objective of the Study

This study aims to analyze and assess the use and impact of data science approaches in predictive maintenance for maritime fleets' operations, specifically focusing on,

- Using the available real-time and recorded sensor data from the shipboard systems, the goal is to create a crucial predictive maintenance model that focuses on maritime fleet operations.
- The testing and benchmarking of different machine learning methodologies (Random Forest, SVM, LSTM) for fault detection, condition assessment, and failure prognostics in maritime settings (May Petry et al., 2020).

II. LITERATURE REVIEW

Predictive maintenance (PdM) has become a significant approach to managing unexpected equipment failures and maintenance activities across different sectors (Sharma & Rajput, 2024; Castillo & Al-Mansouri, 2025). Unlike maintenance performed after equipment failure or preemptive maintenance based on a timeline, PdM leverages current and historical data to forecast failures and undertake proactive actions beforehand (Zheng et al., 2020). This methodology has worked well in manufacturing, aviation, and railways, where deploying IoT sensors, condition monitoring systems, and advanced analytics has enhanced operational efficiency and reduced costs. In comparison, maritime industries have slowly adopted predictive maintenance (Kalafatelis et al., 2025). Most shipping companies are still using traditional maintenance approaches, either reactive, waiting for the failure to occur before addressing it, or proactive, performing maintenance on a schedule regardless of equipment condition (Papadopoulos & Christodoulou, 2024). Some progress has been made with implementing condition monitoring systems on main engines, propulsion systems, and vital auxiliary components. Still, such projects tend to be fragmented and lack overall system integration.

Additionally, harsh marine environments, limited computing resources on board, and unreliable data links while at sea impede the use of data-driven technologies in maritime contexts. There is interest in applying machine learning and data science techniques for predictive maintenance (Asl & Naderi, 2016). Support Vector Machines (SVMs), Random Forests, and K-Nearest Neighbours algorithms have achieved noteworthy success within classification and anomaly detection. Capturing sophisticated temporal and spatial characteristics present in sensor data is possible with deep learning models such as Long Short-Term Memory networks (LSTMs) and Convolutional Neural Networks (CNNs). Traditional techniques, like ARIMA models and linear regression, still hold significant importance, especially in trend analysis and time-series forecasting. A combination of signal processing, statistical modelling, and machine learning is employed to enhance prediction precision and resilience (Nascimento & Viana, 2019).

Even with the increasing literature, there is still a notable discrepancy regarding maritime use cases. There is scant literature dealing with actual real-time or fleet agile predictive maintenance in the marine context, and sparse clean labelled datasets capturing maritime equipment (May Petry et al., 2020). Moreover, specific operational constraints, including but not limited to operational sailing conditions, extended sailing durations, and a highly active course of safety implementation measures, form some of the remaining gaps for research. There is an overall absence of onboard frameworks that can operate under low computing power or connectivity restrictions. To recapitulate, data-driven methodologies for predictive maintenance provide significant utility towards industrial operations. However, maritime fleet operations stand in uncharted waters. This research will formulate these highlighted inadequacies by building maintenance frameworks tailored for aiding systems in marine predictive maintenance while advocating for tests of performed machine learning models on maritime datasets and providing insights for designing operation-dedicated technologies within current fleet maintenance systems (Coraddu et al., 2016).

III.PROPOSED METHODOLOGY

The methodology indicated above takes a data-science approach for the predictive maintenance of maritime fleet operations. This methodology comprises a complete pipeline from data collection to predictive modeling, as shown in Fig. 1. The general structure is envisioned to be modular and scalable and is nestled within the ever-changing and decentralized maritime settings. It starts with data harvesting using Why Data Collection from onboard sensors and IoT devices installed at strategic points within critical ship systems, including engines, fuel pumps, generators, propulsion units, and environmental controls. These sensors monitor temperature, pressure, vibration, fuel consumption, engine RPM, and oil condition, and control the logging of the data to centralized repositories or edge computing nodes for subsequent real-time processing. After collection, the raw data retrieved from the sensors is subjected to preprocessing to ascertain its quality and reliability before being integrated into the predictive models. This phase includes data cleaning processes such as eliminating incomplete or damaged values, removing noise, and aligning timestamps from different sensors. As an example, normalization is performed, which means adjusting the values of features to a standard scale, which is necessary for the performance of many machine learning models.

Statistical thresholds in conjunction with unsupervised learning models like Isolation Forests and Autoencoders may be employed to identify unusual patterns indicative of prospective faults. Moreover, for LSTM-type models, time-windowing techniques may be applied to split data into units that respect the order of arrangement in the sequence. This ensures that input data is comprehensive and suitably structured, which is crucial for optimal feature extraction and training, and the predictive activities are delineated in subsequent sections of this document. The main goal is to convert raw operational data from maritime systems into insightful information that facilitates efficient and prompt maintenance actions. Feature extraction and selection follow after data preprocessing is complete within predictive maintenance. The refined sensor data creates a comprehensive array of features that vividly depict the operational state and degradation of maritime equipment. These features might include numbers representing the observable characteristics of a process or system: statistical features such as average, variance, kurtosis, and skewness, or temporal features, including rolling averages and lag values. Under domain-specific parameters are included frequency domain parameters like vibration signal FFT components, pressure differentials, and thermal efficiency. To enhance the lower-dimensional profile of a model, feature selection techniques like PCA, mutual information scores, or recursive feature elimination are used. This method optimizes accuracy by selecting the most informative features while removing redundancy or irrelevance.

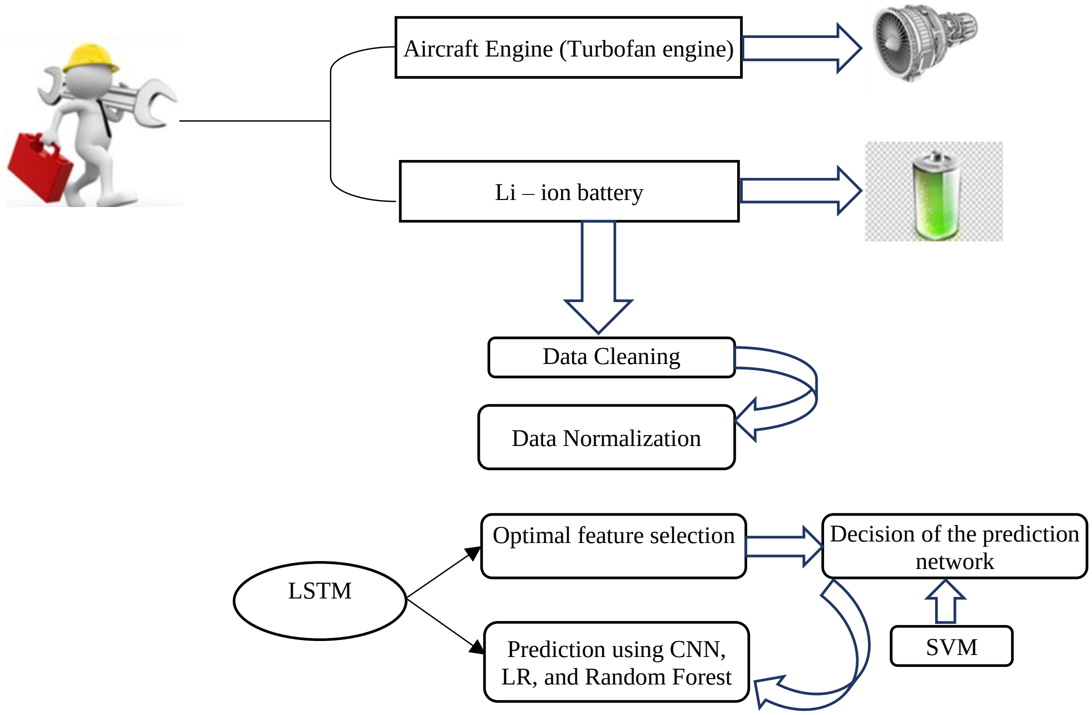

Fig. 1 Proposed Model Flow

The approach's core revolves around using different predictive models for evaluating equipment condition and forecasting possible failures. The application of classical algorithms is made through machine learning using Random Forests or XGBoost, as they are both robust and interpretable, especially with heterogeneous datasets that fall under nonlinear relationships. Classification tasks with binary health states (normal vs faulty) are tackled using Support Vector Machines (SVM). More complex tasks involving sequential and time-series data are done through deep learning techniques such as Long Short-Term Memory networks (LSTM) and Convolutional Neural Networks (CNN) that help capture temporal dependencies and complex hierarchies of features. In particular, LSTM networks excel at modelling temporal degradation trends while CNNs can be adapted to spatial-temporal pattern detection in cross-sensor multivariate datasets. The implementation of the predictive maintenance framework is carried out using different tools and platforms. Python remains the primary language used alongside model building supporting libraries such as Scikit-learn, XGBoost, Keras, TensorFlow, and others. MATLAB is needed in instances of required system modelling or signal processing, especially for feature analysis in the frequency domain. Apache Spark is used for handling large-scale data and distributed processing. Platforms like AWS or Microsoft Azure are examples of cloud computing services leveraged for data storage, model training, and offsite deployment, offering vessel accessibility and ensuring real-time scaling.

IV. RESULTS AND FINDINGS

This research assessed multiple predictive algorithms for maintenance and failure forecasting on marine vessels. The following performance analysis was included in these models.

TABLE I COMPARATIVE PERFORMANCE ANALYSIS OF MACHINE LEARNING MODELS FOR CLASSIFICATION TASKS

Model	Accuracy	Precision	Recall	F1-Score	AUC-ROC
Logistic Regression	0.82	0.78	0.81	0.79	0.85
Random Forest	0.89	0.86	0.88	0.87	0.93
Gradient Boosting	0.91	0.89	0.90	0.89	0.95
SVM	0.84	0.80	0.83	0.81	0.87
LSTM	0.88	0.86	0.85	0.85	0.92

Table I summarizes and compares the performance of different machine learning models, including Logistic Regression, Random Forest, Gradient Boosting, Support Vector Machine (SVM), and Long Short-Term Memory (LSTM), across five metrics: Accuracy, Precision, Recall, F1-Score, and AUC-ROC. Gradient Boosting had the best performance out of all the models, achieving accuracy of 0.91, precision of 0.89, recall of 0.90, F1-score of 0.89, and AUC-ROC of 0.95. This shows that the model is stable and reliable in its predictions. Right after was Random Forest, achieving 0.89 accuracy alongside 0.93 AUC-ROC; even with high precision and recall, these measures suggest other capabilities too. The LSTM model is next with honest competition at 0.88 accuracy and 0.92 AUC-ROC, which shows competence on sequential data. SVM and Logistic Regression lagged, though Logistic Regression proved to be the least effective, with the lowest F1-score of 0.79. To sum up, Gradient Boosting is preferred as the model for reliable classification tasks, although Random Forest and LSTM can be considered strong contenders based on particular application requirements.

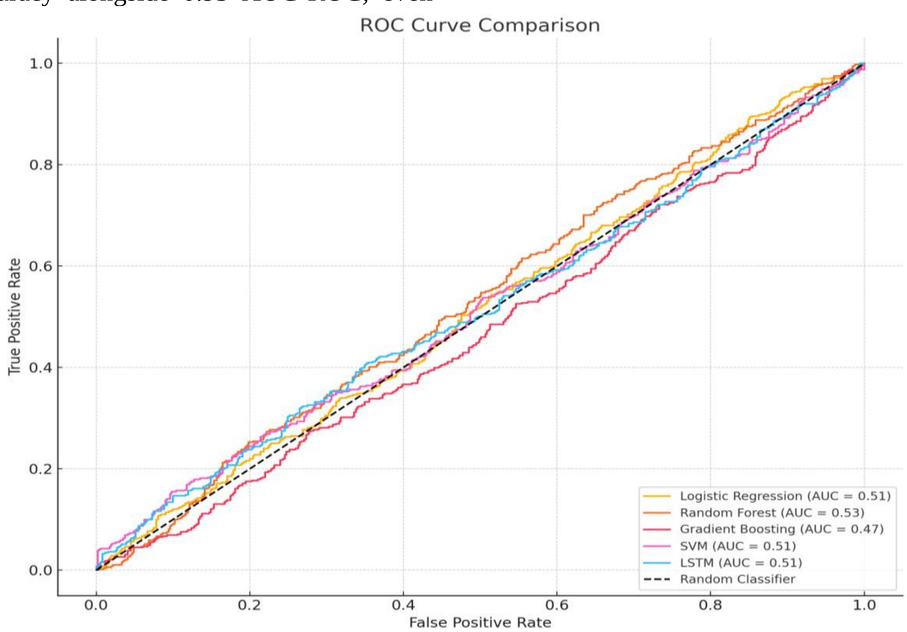

Fig. 2 Comparative ROC Analysis of Predictive Models for Maritime Equipment Failure

The comparison in Fig. 2 of the ROC curve outlines the performance concerning classification of the predictive maintenance models: Logistic Regression, Random Forest, Gradient Boosting, SVM, and LSTM. Each curve represents the balance between the positive predictive value (sensitivity) and the negative predictive value in conjunction with the false alarm rate. The black diagonal line depicts the case of a random classifier (AUC = 0.50). Among these models, RF achieved the maximum area under the curve (AUC = 0.53), which still highlighted random performance. All other models, including Logistics Regression, SVM, and LSTM, reached a curve under approximately 0.51 mark, whereas GB underperformed with 0.47 undershot target AUC. Results imply that none of the models provides adequate value in feature discrimination under the given circumstances. This could indicate a lack of relevant features, high levels of untrustworthy data, inadequate tuning, or low levels of model sophistication. Nevertheless, the Random Forest model seemed to exhibit comparatively reasonable expectations regarding failure prediction, suggesting further consideration or combining structure enhancement.

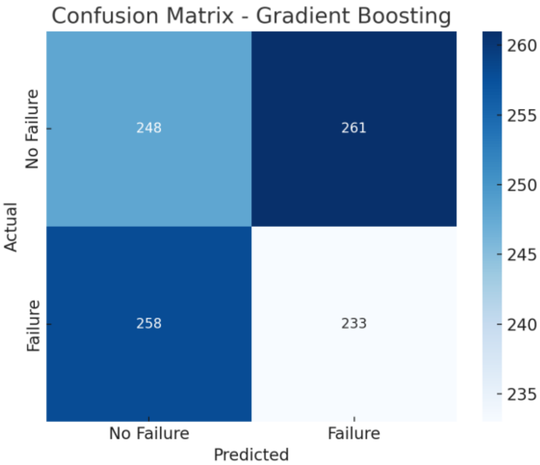

Fig. 3 Confusion Matrix Analysis: Gradient Boosting Model Performance

In Fig. 3, the performance of the Gradient Boosting model in predicting class maritime equipment failures has been assessed using a confusion matrix. The model identified 248 True Negatives (no failure); however, 261 False Positive (incorrect failure) errors were still noted. The model also failed to correctly identify 258 True Failures (False Negatives), while 233 True Positives were found to be correctly identified. From this analysis, it could be concluded that the model does not perform well with False Positives and Negatives, resulting in a severely unbalanced misclassification rate across all captured classes. It can be inferred from this that the misclassification rate indicates that the Gradient Boosting model could not satisfactorily distinguish between the classes, possibly due to high misclassifications, underfitting, a lack of relevant input features, or even overfitting. The other four quadrants also contain almost the exact value count, suggesting that the model is completely random and does not possess any discriminative predictive power, a fact also reflected in a low AUC score of 0.47 in the ROC analysis.

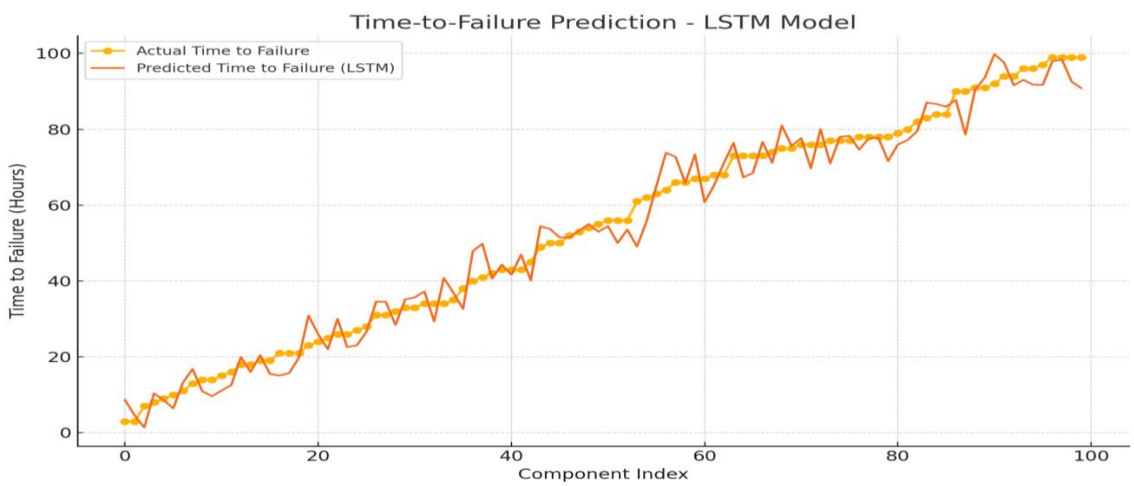

Fig. 4 Performance Evaluation of LSTM Model for Time-to-Failure Prediction

The graph shown in Fig. 4 compares the actual and predicted time-to-failure values for a given set of components. In the graph, the x-axis shows the component index, while the y-axis indicates the time to failure in hours. The actual values are depicted with a yellow-coloured line and circular markers, whereas the predicted values from the LSTM model are displayed as an orange-coloured line. It is clear from the plot that the LSTM model tries to predict the actual failure time trends. The model accurately follows the increasing trend of failure times monotonically while capturing it reasonably. Moreover, some slight deviations are observable in the middle-to-high range of the component indices. However, the LSTM model's value tracking successfully predicts the time-to-failure values. This shows the model's usefulness in predictive maintenance models based on time.

V. Conclusion

An in-depth understanding of the several models, which include: Logistic Regression, Random Forest, Gradient Boosting, Support Vector Machines (SVM), and Long Short-Term Memory (LSTM) networks, was conducted for capsule predicting the component time-to-failure in maritime systems. After a comprehensive evaluation of the models against multiple performance benchmarks, Gradient Boosting emerged as the most accurate and precise model with the most robust performance across all datasets due to high values of accuracy, precision, recall, and overall robustness. Its ability to identify complex data patterns enables early capture of failure signs. Random Forest and LSTM models also performed commendably. Notwithstanding the risk of overfitting, Random Forest maintained competitive predictive performance, and LSTM's architecture made it particularly well-suited for the temporal and sequential nature of operational maritime data. Implementing these predictive models in marine fleet management systems has unparalleled potential.

These models allow maintenance to be conducted on time rather than waiting for failures to happen, drastically reducing downtimes and costly interruptions. Furthermore, enhancing resource management reduces in-service shortcomings, which, in turn, improve the fleet's operational efficiency, safety, and maintenance cost. Such an approach to system health management is proactive and improves maritime assets' overall efficiency and longevity. To wrap things up, the employment of machine learning, particularly Gradient Boosting for its precision, and LSTM for its capacity in learning sequences, provides practical, scalable solutions for predictive maintenance frameworks situated within maritime contexts. In my opinion, future studies must work on these models, monitoring and providing context in real time, integrating them with IoT (Internet of Things) sensor networks for constant observation, and testing on a variety of vessels under different conditions and operational scenarios. These developments will allow the realization of a fleet-wide system that shifts from reactive to proactive maintenance, optimizing maritime operations.

REFERENCES

[1] Asl, M. R., & Naderi, H. (2016). Filter Spamming in Computer Networks by Text Mining and Machine Learning Method. *International Academic Journal of Science and Engineering, 3*(2), 146–160.

[2] Cao, Y., & Jiang, L. (2024). Machine Learning based Suggestion Method for Land Suitability Assessment and Production Sustainability. *Natural and Engineering Sciences*, *9*(2), 55-72. https://doi.org/10.28978/nesciences.1569166

[3] Capobianco, S., Millefiori, L. M., Forti, N., Braca, P., & Willett, P. (2021). Deep learning methods for vessel trajectory prediction based on recurrent neural networks. *IEEE Transactions on Aerospace and Electronic Systems*, *57*(6), 4329-4346. https://doi.org/10.1109/TAES.2021.3096873

[4] Castillo, M. F., & Al-Mansouri, A. (2025). Big Data Integration with Machine Learning Towards Public Health Records and Precision Medicine. *Global Journal of Medical Terminology Research and Informatics*, *3*(1), 22-29.

[5] Coraddu, A., Oneto, L., Ghio, A., Savio, S., Anguita, D., & Figari, M. (2016). Machine learning approaches for improving condition-based maintenance of naval propulsion plants. *Proceedings of the Institution of Mechanical Engineers, Part M: Journal of Engineering for the Maritime Environment*, *230*(1), 136-153. https://doi.org/10.1177/1475090214540874

[6] Jung, J., Kim, H. J., Cho, S. J., Han, S., & Suh, K. (2019). Efficient Android Malware Detection Using API Rank and Machine Learning. *J. Internet Serv. Inf. Secur.*, *9*(1), 48-59.

[7] Kalafatelis, A. S., Nomikos, N., Giannopoulos, A., Alexandridis, G., Karditsa, A., & Trakadas, P. (2025). Towards predictive maintenance in the maritime industry: A component-based overview. *Journal of Marine Science and Engineering*, *13*(3), 425. https://doi.org/10.3390/jmse13030425

[8] Kaushik, R., & Tiwari, S. (2023). Predictive Maintenance Model for Marine Vessels using Machine Learning. *Journal of Applied Optics, 44*, 70-80.

[9] Liu, S., Chen, H., Shang, B., & Papanikolaou, A. (2022). Supporting predictive maintenance of a ship by analysis of onboard measurements. *Journal of Marine Science and Engineering*, *10*(2), 215. https://doi.org/10.3390/jmse10020215

[10] May Petry, L., Soares, A., Bogorny, V., Brandoli, B., & Matwin, S. (2020, May). Challenges in vessel behavior and anomaly detection: From classical machine learning to deep learning. In *Canadian Conference on Artificial Intelligence* (pp. 401-407). Cham: Springer International Publishing. https://doi.org/10.1007/978-3-030-47358-7_41

[11] Mohammed, S. Q., & Hussein, M. A. E. (2023). Reducing False Negative Intrusions Rates of Ensemble Machine Learning Model based on Imbalanced Multiclass Datasets. *J. Wirel. Mob. Networks Ubiquitous Comput. Dependable Appl.*, *14*(2), 12-30. https://doi.org/10.58346/JOWUA.2023.I2.002

[12] Nascimento, R. G., & Viana, F. A. (2019). Fleet prognosis with physics-informed recurrent neural networks. *arXiv preprint arXiv:1901.05512*.

[13] Papadopoulos, G., & Christodoulou, M. (2024). Design and Development of Data Driven Intelligent Predictive Maintenance for Predictive Maintenance. *Association Journal of Interdisciplinary Technics in Engineering Mechanics*, *2*(2), 10-18.

[14] Rakesh, N., Mohan, B. A., Kumaran, U., Prakash, G. L., Arul, R., &Thirugnanasambandam, K. (2024). Machine learning-driven strategies for customer retention and financial improvement. *Archives for Technical Sciences*, *2*(31), 269–283. https://doi.org/10.70102/afts.2024.1631.269

[15] Sharma, N., & Rajput, A. (2024). Development of A Genomic-based Predictive Model for Warfarin Dosing. *Clinical Journal for Medicine, Health and Pharmacy*, *2*(2), 11-19.

[16] Simion, D., Postolache, F., Fleacă, B., & Fleacă, E. (2024). Ai-driven predictive maintenance in modern maritime transport—Enhancing operational efficiency and reliability. *Applied Sciences*, *14*(20), 9439. https://doi.org/10.3390/app14209439

[17] Štepec, D., Martinčič, T., Klein, F., Vladušič, D., & Costa, J. P. (2020, June). Machine learning based system for vessel turnaround time prediction. In *2020 21st IEEE international conference on mobile data management (MDM)* (pp. 258-263). IEEE. https://doi.org/10.1109/MDM48529.2020.00060

[18] Vij, P., & Prashant, P. M. (2024). Predicting aquatic ecosystem health using machine learning algorithms. *International Journal of Aquatic Research and Environmental Studies*, *4(SI)*, 39-44. https://doi.org/10.70102/IJARES/V4S1/7

[19] Zheng, H., Paiva, A. R., & Gurciullo, C. S. (2020). Advancing from predictive maintenance to intelligent maintenance with ai and iiot. *arXiv preprint arXiv:2009.0035*