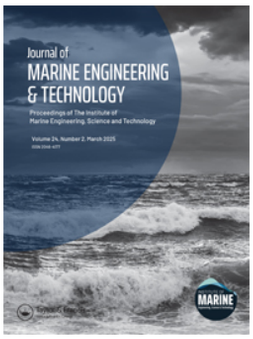

Journal of Marine Engineering & Technology

ISSN: (Print) (Online) Journal homepage: www.tandfonline.com/journals/tmar20

Parametric machine learning integrated approach for assessing environmental and engine variables on fuel consumption and carbon intensity

Onur Yuksel, Murat Bayraktar & Olgun Konur

To cite this article: Onur Yuksel, Murat Bayraktar & Olgun Konur (30 Apr 2025): Parametric machine learning integrated approach for assessing environmental and engine variables on fuel consumption and carbon intensity, Journal of Marine Engineering & Technology, DOI: 10.1080/20464177.2025.2499346

To link to this article: https://doi.org/10.1080/20464177.2025.2499346

Published online: 30 Apr 2025.

JOURNAL OF MARINE ENGINEERING & TECHNOLOGY
https://doi.org/10.1080/20464177.2025.2499346

Taylor & Francis
Taylor & Francis Group

Parametric machine learning integrated approach for assessing environmental and engine variables on fuel consumption and carbon intensity

Onur Yuksel [a,b], Murat Bayraktar [b] and Olgun Konur [c]

[a]Liverpool Logistics Offshore and Marine Research Institute (LOOM), Faculty of Engineering, Liverpool John Moores University, Liverpool, UK; [b]Marine Engineering, Maritime Faculty, Zonguldak Bülent Ecevit University, Ereğli/Zonguldak, Türkiye; [c]Marine Engineering, Maritime Faculty, Dokuz Eylül University, Buca/İzmir, Türkiye

ABSTRACT

The study aims to identify the most optimal machine learning (ML) algorithm for predicting fuel consumption (FC) based on noon report (NR) data and to explore the impact of environmental and operational engine variables on the FC through a parametric study. The M5 Rules, Artificial Neural Networks, and Random Forests algorithms have been compared in this context. This study's innovative aspect lies in parametric analysis within the best-performing algorithm to explore how variations in selected control parameters influence FC and the Carbon Intensity Indicator rating. The NR data has been gathered from a tanker ship's noon reports over a year. After feature selection for the parametric study, the adjusted data comprising the identified variables have been used to run the chosen model. The results showed that the M5 Rules algorithm is the most appropriate for the specific data, and the Beaufort scale/slip and scavenge pressure have the highest effects on the FC. The Beaufort scale/slip varies the FC annually between a 1341.27 t (−23.48%) reduction and a 2088.05 t (36.55%) increase. Similarly, the changes in scavenge pressure impact the FC from a decrease of 859.99 t (−15.05%) or increment up to 733.72 t (12.84%).

ARTICLE HISTORY
Received 8 January 2025
Accepted 24 April 2025

KEYWORDS
M5 Model Rules; Artificial Neural Networks; Random Forest; Carbon Intensity Indicator (CII); Fuel Consumption Prediction

1. Introduction

Considering that over 80% of all trade is conducted by sea, air pollution and global warming are serious issues for maritime transportation (Lindstad et al. 2020; Baştürk and Erol 2023). The marine industry plans to transition to a carbon-neutral shipping sector by 2050 in the scope of goals set by the International Maritime Organization (IMO). The short-term reduction aims to cover a 40% decrease in CO_2 by 2030 compared to 2018 (Seyfi et al. 2023). In this regard, beginning on 1 January 2023, the IMO mandated the computation and reporting of the Energy Efficiency Existing Ship Index (EEXI) and the Carbon Intensity Indicator (CII) for ships larger than 400 gross tonnage (IMO 2022; Bayraktar and Yuksel 2023).

Reliable fuel consumption (FC) estimations for marine systems are important from an environmental and financial viewpoint. Ship management effectiveness is increased when FC implications are recognised and predictions can be used to optimise systems (Prpić-Oršić and Faltinsen 2012). Estimates for the coming years are also useful for evaluating how well a vessel or system complies with IMO standards (Eide et al. 2011). FC prediction can be ensured by utilising the Machine Learning (ML) approach, which scrutinises the historical data of a system, evaluates its current state, and forecasts categorises, or groups target values relevant to its intended application. ML methodologies serve as potent tools for discerning the impact of weather patterns and operational variables on FC (El Naqa and Murphy 2015). The construction of such models facilitates precise FC predictions depending upon environmental and operational parameters. These predictions, in turn, aid in the selection of optimal routes, velocities, trim adjustments, or draft variations (Akyuz et al. 2019).

One of the most important methods for managing the variables that affect the fuel usage is to evaluate the ship's noon report (NR) data. Various information is included in the NR, including the amount of fuel used each day, the speed, the direction of sail, and external factors, including wind, waves, and currents (Safaei et al. 2018). Despite the potential for human error, NRs remain the most commonly used and cost-effective method of data collection onboard ships (Zwart et al. 2023). Consequently, they are widely adopted by most shipping companies. In this context, an appropriate ML algorithm selection is crucial. The chosen algorithm must effectively process NR data, even with limited data points, and should be readily implementable across various ship types, while also being easily understandable by operators and managers (Panda 2023).

1.1. Literature review

Several researchers have conducted studies in recent years about the prediction of FC, required power or energy, and their effects on marine vessel conditions by utilising ML and deep learning (DL) approaches. Research has evaluated the effectiveness of various algorithms in forecasting the FC of marine vessels under diverse environmental and engine operational conditions.

Le et al. (2020) compared the Multiple Linear Regression (MLR) and Artificial Neural Networks (ANN) for forecasting the FC of container vessels in Korea. The results demonstrated that the ANN

CONTACT Onur Yuksel O.Yuksel@ljmu.ac.uk Liverpool Logistics Offshore and Marine Research Institute (LOOM), Faculty of Engineering, Liverpool John Moores University, Byrom Street, Liverpool L3 3AF, UK; Marine Engineering, Maritime Faculty, Zonguldak Bülent Ecevit University, Kepez District, Hacı Eyüp Street, No:1 67300, Kdz, Ereğli/Zonguldak, Türkiye

This article has been corrected with minor changes. These changes do not impact the academic content of the article.

outperformed the MLR regarding predictive accuracy. Furthermore, the model effectively validated the benefits of slow-steaming practices for energy efficiency.

Uyanık et al. (2020) conducted a benchmarking study of various predictive models, including MLR, Ridge Regression, Least Absolute Shrinkage and Selection Operator (LASSO), Support Vector Regression (SVR), Tree-Based Algorithms, and Boosting Algorithms. he results demonstrated that both Multiple Linear Regression (MLR) and Ridge Regression exhibited optimal performance, achieving a Root Mean Square Error (RMSE) of 0.0001, a Mean Absolute Error (MAE) of 0.002, and a coefficient of determination (R^2) of 0.999.

Panapakidis et al. (2020) forecasted the FC of a passenger ship through various types of ANN, including Long Short-Term Memory (LSTM), Feed-Forward Neural Networks (FNN), and Elman Neural Networks (ENN). The study achieved the highest R^2 of 0.829. Notably, the most concise estimations were delivered by the ENN and FNN.

Hu et al. (2021) investigated the performance of single and hybrid models, including Extreme Random Trees (ERT), Random Forest (RF), and Extreme Gradient Boosting (XGB), alongside ANN and MLR, trained with sensory data. They introduced a data preprocessing methodology that significantly enhanced the quality of ship fuel consumption data, increasing the correlation coefficient (R-value) from 0.777 to 0.9179. The findings indicated that the developed model achieved the highest predictive accuracy, followed by ERT, XGB, RF, ANN, Support Vector Machines (SVM), and MLR.

Kim et al. (2021) compared MLR and ANN to predict the FC of a container vessel. ANN models gave more accurate results with the R^2 values ranging from 0.971 to 0.994. The study further conducted a sensitivity analysis on the ship's draught, identifying an optimal draught of 14.79 m, closely aligned with the vessel's design specifications, as the condition yielding the best fuel efficiency.

Alexiou et al. (2021) proposed a hybrid methodology that compared several algorithms, including ANN, Decision Tree Regression, RF, k-Nearest Neighbours (KNN), MLR, and AdaBoost, to predict power output from the main engine (ME) based on Automatic Data Logging Monitoring (ADLM) data. The proposed strategy yielded a considerable reduction in both the absolute error of forecasts and the computational power required for calculations, significantly enhancing the overall efficiency of the prediction process.

Gupta et al. (2022) constructed Principal Component Regression (PCR), Partial Least Squares Regression (PLSR), and probabilistic ANN using two sisters' ships data for the performance monitoring. The probabilistic ANN model exhibited superior performance. The results from PCR and PLSR were closely aligned, indicating that simpler methods can effectively address similar challenges when integrated with domain knowledge.

Parkes et al. (2022) trained ANNs using data collected from various vessels to estimate power requirements. The model demonstrated validated ship powering predictions with a mean error of less than 2% and achieved a 4% error for vessels with no prior data available.

dos Santos Ferreira et al. (2022) benchmarked MLR, Polynomial Regression, Regression Tree (RT), Gradient Boosting Tree (GBT), RF, and ANN using the big data collected from two sister container vessels. RF provided an R^2 of 0.997, which is the highest among the compared methods.

Yüksel and Köseoğlu (2022) evaluated MLR, SVR, and ANN models to predict the FC of marine diesel generators aboard a crude oil tanker. The study assessed both prediction accuracy and computational efficiency, identifying MLR as the optimal algorithm for this specific application.

Guo et al. (2022) demonstrated the integration of RF and physics-based approaches to predict the FC of cargo ships across various case studies. With an error rate of 2.7%, the proposed approach performed satisfactorily in predicting the target values.

Yuksel et al. (2023) conducted a benchmarking analysis comparing MLR, SVR, M5 model rules, and RT-based NRs data to forecast the ship FC and prioritise the influencing environmental conditions. The M5 model rules demonstrated the optimum accuracy, achieving an R^2 of 0.966.

Nguyen et al. (2023) compared XGB and ANN and physical models to forecast the FC using the NRs. For certain ship types (general cargo and containerships), physical models outperformed ML. ANN still provided reasonably accurate FC predictions for oil tankers, bulk carriers, and Roll-On/Roll-Off ships.

Zhang et al. (2023) performed a comparative benchmarking of ANN, SVR, LASSO, XGB, and RF for a CII estimation approach. The ANN model achieved the highest accuracy, with an average absolute error of 0.0336, while the LASSO model displayed the greatest inaccuracy, with an error of 0.2817.

Bilgili (2023) constructed An ANN architecture that calculates ship resistance with an R-value of 0.745 based on environmental conditions. Classification and Regression Trees (CART) were used to assess the significance of each ANN feature. Ship resistance is mostly influenced by wave orientations and swell, according to the results.

Agand et al. (2023) presented a comparative study of ML algorithms to estimate the FC of a passenger ship. MLR, ANN, decision tree, and XGB were among the tried algorithms. The best prediction accuracy was obtained from the XGB methodology with an R^2 of 0.988.

Bayraktar and Sokukcu (2024) built MLR and ANN models with NRs for energy efficiency determination by predicting the slip. R^2 at 0.927 was obtained, and the regression formula of the slip was provided regarding the environmental variables.

Handayani et al. (2023) developed an XGB model using NRs of two vessels to forecast FC. Results indicated that the evaluated XGB model achieved a satisfactory prediction performance with an R^2 of 0.95.

Karatuğ et al. (2023) presented a decision support system for condition-based maintenance of ship machinery systems using the adaptive neuro-fuzzy inference system (ANFIS) approach. A case study of a container ship predicted diesel engine power using exhaust gas outlet temperatures and ME shaft revolution. Two strategies were employed: a cylinder-based model and an overall system model. Comparative analyses identified the optimal ANFIS structure, with the finalised model showing superior accuracy ($R^2 = 0.9806$, RMSE $= 1.6588$ MW).

Taghavifar and Perera (2023a) examined the application of ANNs, incorporating a modified homogeneity factor, to estimate exergetic parameters resulting from combustion and/or mixing dynamics in a marine engine. Results showed that ANNs effectively forecast examined terms related to spray and injection parameters, enhancing energy efficiency and reducing emissions in diesel engines.

Fan et al. (2024) provided a comprehensive study on FC prediction through ML approaches based on sensory data from a ship. RF and XGB were detected as the most appropriate candidates, with the R^2 value reaching up to 0.986.

Hajli et al. (2024) performed the MLR for bulk carriers depending on propeller speed and weather factors. The developed model successfully predicts FC for over 80% of the voyages in the dataset, with an MAE and RMSE below 0.01 t per nautical mile.

Le et al. (2024) developed a ship FC prediction framework utilising two distinct algorithms: Huber Regression (HR) and Light Gradient Boosting Machines (LGBM). In terms of predictive accuracy, the HR model outperformed LGBM, achieving an R^2 of 0.979 compared to LGBM's 0.917.

Zhang et al. (2024) presented a Bidirectional Long Short-Term Memory (BiLSTM) model to estimate the FC of a Kamsarmax bulk

carrier, leveraging sensory data collected from the vessel. The model demonstrated robust performance, achieving an average R^2 of 0.85 across eight different voyages.

Senol and Seyhan (2024) built an ANN model to estimate the ship emissions during the manoeuvring. Over 300,000 rows of ME speed data, collected during berthing manoeuvres performed by 92 professional maritime pilots, were analysed in the study. The results of berthing manoeuvring emissions exhibited variability of up to 1.85 times depending on the pilot. The ANN model provided an accuracy ratio of 73%.

Vorkapić et al. (2024) employed the 'Cross Industry Standard Process for Data Mining' framework to analyse the two-stroke propulsion engine of a very large gas carrier. Linear regression and decision tree models were developed. The linear regression model achieved an RMSE of 23.16 and a mean relative absolute error (MRAE) of 14.7%, while the decision tree models demonstrated accuracies ranging from 96.4% to 97.69%.

Karatuğ et al. (2024) ensured a decision support system to improve ship energy efficiency by integrating an optimisation model with an ANN. The model was based on real-time data of ship and engine performance, validated through Ricardo Wave software. FC predictions were made using various ANN configurations, with the most effective model having one hidden layer and five neurons, achieving high accuracy ($R^2 = 0.99697$, RMSE = 0.00035).

Zhou et al. (2025) developed a grey-box deep learning approach for predicting FC in tuna purse seiners, addressing the financial concerns of fuel costs. The proposed model combines a Multi-head BiLSTM network with domain knowledge to identify operating modes for improved accuracy. Utilising data from onboard sensors and the Copernicus service, the model captured the distinct characteristics of FC across various operational modes. It demonstrated high accuracy, achieving at least 97.66% across ten unseen cruising trips and 90.93% during eight fishing events.

Su et al. (2025) built an ML-based model to forecast FC n costs for roll-on/roll-off carriers, addressing the rising fuel expenses and environmental concerns. Analysing a dataset of 16,189 observations from a shipping company, the research utilised the categorical boosting algorithm, achieving an impressive R^2 value of 0.976. Key variables influencing fuel costs included distance, sea days, speed, duration, and port call days.

Nguyen et al. (2025) applied an Internet of Things-driven approach integrated with explainable ML models to estimate ship FC based on sensory data. Among five ML techniques tested, XGBoost outperformed the others, achieving the highest R^2 values (0.997 for training and 0.95 for testing) and demonstrating the lowest error rates. The analysis revealed that ME shaft power was the most significant parameter for FC prediction, emphasising the critical role of engine performance metrics in enhancing model accuracy and interpretability.

Piao et al. (2025) performed the ship FC prediction by integrating engine room data and using various ML algorithms for feature selection. Five methods, including MLR and RF, identified key predictive features, with the full-navigational dataset yielding the most accurate predictions due to engine temperature and pressure inclusion. The findings indicated that adding overlooked features not directly associated with propulsive force can significantly improve FC prediction accuracy, enhancing operational efficiency.

The literature review demonstrated that ANNs emerged as the most frequently utilised algorithm in the investigated studies. SVM, MLR, RF, and XGB were also prominently employed, underscoring their efficacy in various contexts.

While benchmarking solely on a single metric is suboptimal, R^2 serves as a valuable comparative measure for assessing algorithmic performance across diverse datasets. As a rate-based metric, R^2 provides a concise overview of the overall effectiveness of the applied methodologies. In the reviewed studies, reported R^2 values ranged between 0.73 and 0.99, reflecting variability attributable to differences in data sources and analytical scope.

The primary data sources included ADLM and NRs for operational parameters, along with the Automatic Identification System for vessel location. The studies mostly examined the FC estimation regarding the changes in environmental factors and their level of impact on the target column. ME power output forecasting and emission prediction regarding different operation modes were also investigated. Some studies used ML or DL techniques for route or speed optimisation by predicting the vessel speed alongside the FC. While predicting these output columns, the benchmarking of various algorithms was performed.

1.2. Research gap, novelty, and motivation

The literature review revealed that the M5 Rules algorithm, known for its high understandability and performance, has not been utilised in FC prediction for marine vessels. The only instance where this algorithm has been applied is in our previous research paper on a related subject. The impact of environmental and ME variables of FC has been evaluated by determining their importance level on the algorithm in the previous studies. A parametric investigation, presented in the next section, has not been encountered in the review.

The innovative aspect of this study lies in the application of parametric analysis, which utilises various levels of selected control parameters within the best-performing algorithm to assess how these variations impact FC and the CII rating. Additionally, this study highlights the use of the M5 Rules algorithm as a grey-box model, which combines the interpretability of traditional models with the predictive power of complex algorithms. Unlike black-box models that often obscure the relationships between input and output, M5 Rules provides clear, understandable rules that facilitate insights into the underlying mechanisms affecting FC. This advantage is particularly significant given the algorithm's demonstrated performance with NR data and a limited number of data points in our previous study, further reinforcing the novelty and applicability of our approach (Yuksel et al. 2023).

This ML-integrated parametric approach can be beneficial for shipping companies that monitor vessel performance analytics through NR data. By enabling numerical simulations of operational and environmental impacts, this method provides high clarity for vessel operators and superintendents. Specifically, the term 'parametric approach' refers to a methodology that relies on selected parameters to characterise a system or model. Within this framework, a simulation-based sensitivity analysis is conducted on designated variables, facilitating the examination of how variations in these parameters influence the model's outcomes. This comprehensive analysis enhances our understanding of the relationships and effects among different variables within the system under investigation, thereby supporting informed decision-making in vessel performance management.

The motivation of this study is to identify the most suitable ML algorithm for FC prediction based on navigational data, with the aim of providing shipping companies with valuable carbon intensity projections. These projections are essential for planning future operations and investments in sustainable practices. To achieve this, the M5 Rules algorithm, characterised as a grey box model, was benchmarked against ANN and RF, which are commonly employed 'black box' models in the literature. In our previous research (Yuksel et al. 2023), we compared the M5 Rules algorithm with MLR and SVR, where it outperformed both.

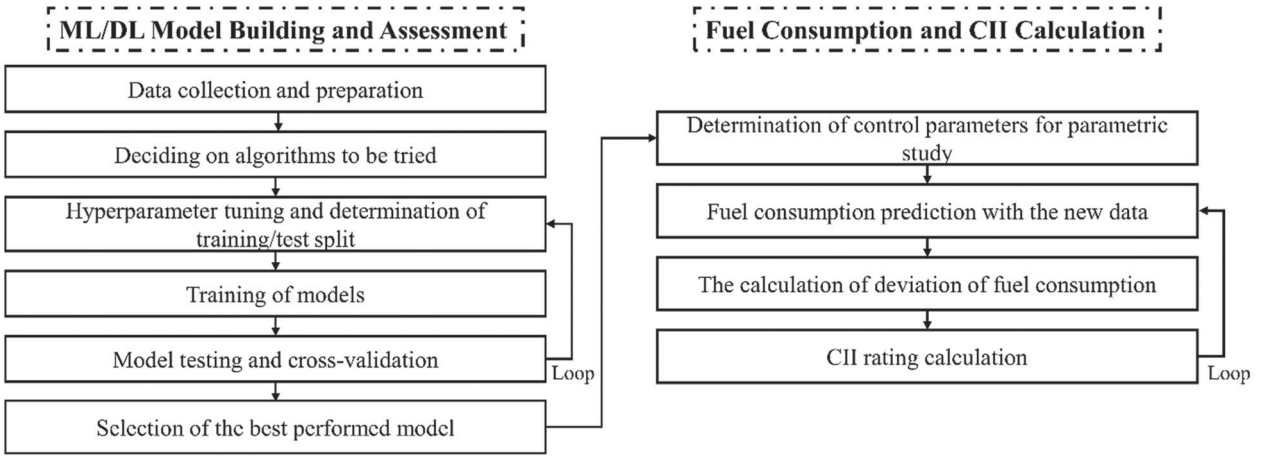

Figure 1. Methodology flowchart of the paper.

One year of navigational data was collected from an ocean-going tanker vessel. A parametric study was then conducted using the model derived from the most successful algorithm. The potential effects of varying selected features on other aspects of ship performance were modelled based on the identified relationships in the dataset. Changes in FC, CO_2 emissions, attained CII and CII ratings were analysed and discussed.

2. Research method structure

This section presents the methodology employed in the paper and highlights the properties of the investigated system. Figure 1 illustrates the methodology flowchart of the study, emphasising the steps involved in the analysis.

The data were gathered for one-year operational recordings kept on board from an oceangoing tanker vessel. The preparation of the data involved cleaning the missing rows and separating the navigation data, including the running of the ME. Three different ML techniques were tried to ensure the importance of features and the FC prediction, which are M5 Model Rules, ANN, and RF. These algorithms were determined as well-suited techniques for such an application, and a low number of rows (Vilaça et al. 2015; Baumann and Klingauf 2020; Yüksel and Köseoğlu 2020; Yuksel et al. 2023). Before the training of models, the hyperparameter tuning and the ideal training-test split percentages of data were decided by trials. Then, the model training was ensured, and performance measurements with test data and 10-fold cross-validation (CV) were conducted. The best-performing model was used to estimate FC and CII Ratings of the parametric study. By considering the model outcomes, the control parameters that can impact the FC were decided for the parametric investigation. The variation of FC ws detected by employing the model fed by new data, and the possible alteration of CII ratings has been calculated and discussed. The structure of this section continues as the feature and system description, which explains the properties of the data and the investigated ship, and the mathematical background that clarifies the model details and calculation of CII ratings.

2.1. Feature and system description

The reference ship was an oil/chemical tanker having 131,433 t of displacement when fully loaded and 50,697.1 t of deadweight (DWT). The ship particulars are demonstrated in Table 1.

The ME specified in Table 1 is the MAN B&W 6 S50 MC-C model, featuring a two-stroke, direct injection design with six cylinders. It delivers a power output of 9480 kW at Maximum Continuous Rating (MCR) and operates at 127 revolutions per minute (rpm). With a bore of 500 mm and a stroke of 2000 mm, it has a compression ratio of 17.2. The specific fuel consumption (SFC) is 187 g/kWh at 85% of MCR.

Table 1. Ship and ME particulars.

Parameter	Value	Unit
Length Overall	183	m
Beam	32	m
DWT	50,697.1	t
Gross Tonnage	30,056	–
Reference Speed (V_{ref})	11.68	Knots
Auxiliary Engine (AE) Model	NISHISHIBA NTAKL-VE	–
Number of AEs	3	–
AE power	900	kW
ME Data		
Parameter	Value	Unit
Model	MAN B&W 6 S50 MC-C	–
Process type	Two-stroke	–
Number of cylinders	6	–
Power at 85% MCR	9480 KW	kW
Revolutions at 85% MCR	127	rpm
Bore	500	mm
Stroke	2000	mm
Compression Ratio	17.2	–
SFC	187 @ 85% MCR	g/kWh

The data source utilised was NRs, from which faulty entries and missing cells were eliminated during the data preprocessing phase. The dataset comprised one year of operational recordings. After data preparation, the final dataset used in the analysis consists of 221 rows and 19 columns. The target column for the analysis is the FC, while the remaining 18 columns were utilised as the input columns. Table 2 defines the features used in the model and sample dataset.

A small dataset for ML, comprising only 221 records, poses several limitations, particularly an increased susceptibility to overfitting. Small datasets often lack representative diversity, leading models to capture noise rather than true patterns, which can adversely affect performance on unseen data (Rather et al. 2024). Strategies involving CV, effective hyperparameter tuning, and a train-test split are employed to mitigate the overfitting risks. CV partitions the dataset into multiple subsets, allowing the model to be trained and validated on different portions, providing a more comprehensive performance evaluation and reducing overfitting (Seraj et al. 2023). The train-test split further ensures that a distinct subset of the data is reserved for

Table 2. The description of the parameters used in the ML model and sample dataset.

Parameter	Description
Displacement	The displacement of the reference vessel during the voyage.
Distance	Travelled distance during the navigation.
BFScale	The Beaufort scale for wind power is used when the journal record has been entered.
Speed	Ship speeds in knots.
Slip	The variance in velocity between the propulsion system's operational speed, denoted as the engine speed, and the empirically determined velocity of the vessel, referred to as the observed speed.
SeaState	The statistical portrayal of characteristics of wind-induced oceanic wave dynamics comprises parameters such as wave heights, periods, and orientations.
MinExhT	Minimum exhaust temperature of the ME in °C.
MaxExhT	Maximum exhaust temperature of the ME in °C.
Draft	The vertical distance from the waterline to the lowest point of a vessel's hull
SWT	Sea water temperature in °C.
ER_T	Engine room temperature in °C.
ScavT	Scavenge temperature in °C.
ScavP	Scavenge pressure in bars.
AvgHeading	The average heading angle of the ship.
Event	The operational state of the ship. It can be 'at sea' or 'end of a sea passage (EOSP)'.
WindDirection	Description of the wind's direction as it affects the ship.
WaveLength	The natural length of a wave is the same speed as the ship.
WaveHeight	The variation between the crest and trough of the wave is measured in metres.
FC	ME fuel consumption in t.

Sample input data

Displacement	Event	Distance	BFScale	Speed	Slip
131433	At Sea	91	9	6	54
131433	At Sea	179	8	7.2	45
131433	At Sea	284	5	11.8	14
131433	At Sea	291	4	12.1	13
131433	EOSP	62	4	11.7	16
105767	At Sea	185	5	12.2	10
WindDirection	SeaState	WaveLength	WaveHeight	AvgHeading	Draft
W	8	Long	High	WNW	14.6
NW	7	Long	High	WNW	14.6
NE	4	Avg	Mod	WNW	14.6
SW	3	Avg	Mod	WNW	14.6
NW	3	Avg	Mod	NW	14.6
SE	4	Avg	Mod	WNW	12
MinExhT	MaxExhT	SWT	ER_T	ScavT	ScavP
251	281	27	30	47	1.8
254	288	25	27	42	1.09
248	284	21	26	42	1.3
245	280	21	26	42	1.32
245	280	21	26	42	1.32
252	287	18	22	39	1.2

Sample output data

1	**2**	**3**	**4**	**5**	6
24.2	39.2	33.2	32.9	7.8	20.1

testing, allowing for an unbiased assessment of the model's performance (Rácz et al. 2021). Meanwhile, hyperparameter tuning optimises model complexity by systematically adjusting key parameters, balancing flexibility and rigidity (Aftab et al. 2025). Together, these strategies enhance the model's robustness and predictive power, enabling it to generalise effectively despite the limitations of a small dataset. Table 3 presents summary statistics for the numerical data, while Figure 2 depicts the distribution of the categorical features.

During the reporting phase, anemometers were used to collect and save data more effectively regarding the Beaufort scale. Anemometers, often known as 'Marine Vanes', are devices that monitor wind direction and speed simultaneously aboard ships (Tamaya 2020). Marine Vanes are used to measure direction and relative speed. The course and speed of the vessel are therefore crucial during the recording period. These circumstances should be considered by the authorised crew when completing the registration procedure (Yuksel et al. 2023). The daily wavelength height, sea state, average direction, and wind force are demonstrated in the NRs, even though they are documented in the logbook. Displacement and draft were recorded at the beginning of the voyage, while the speed was taken as the average during the day. The authorised ship crew used the Beaufort scale to record the sea state entered in the logbook.

The Beaufort scale is an empirical tool that correlates wind speed with observable conditions both on land and at sea, has been employed in the assessment of fundamental performance criteria for the ship's crew (MetMatters 2025). Table 4 delineates details of the Beaufort scale.

The wind force and wave height represent the sea state. Wave height is recorded using both manual methods and automated instruments, such as buoys and radar systems. Personnel take regular measurements, noting the maximum height of waves observed within specific intervals. To calculate the average wave height, individual measurements are summed and divided by the total number of observations (Hwang et al. 2023). In this case study, ship personnel systematically monitored wave height, quantified in metres, alongside the prevailing sea state. This data was recorded by manually comparing the vessel's height with the observed wave heights. The wind direction and average heading are demonstrated according to the '32-Point Compass Rose' (Wheeler 2005). The distance was computed using the ships' coordinates.

The ship's propeller is driven by the ME and transmission mechanisms, facilitating movement contingent upon the rpm, operational duration, and the pitch value. All these values are utilised for theoretical distance range calculations. However, the actual distance travelled may vary due to factors such as wind conditions, currents, wave dynamics, ship draft, and the degree of fouling on the hull. The disparity between the theoretical distance and the actual distance is quantified as 'slip' (Bayraktar and Sokukcu 2024). The factors like fouling or additional drag conditions, were included in ML under the slip feature.

The parameters, such as seawater temperature, engine room temperature, exhaust temperatures, scavenge temperature, and pressure, were collected by authorised engine crew or duty engineers and recorded in the engine logbook. Seawater temperature is a significant factor influencing the FC of a diesel engine, as the freshwater cooling system relies on seawater for effective heat dissipation within marine systems. Elevated seawater temperatures can diminish cooling efficiency, resulting in increased operating temperatures within the engine (Pariotis et al. 2019). Similarly, engine room temperature also affects cooling efficiency and the intake air temperature, which can lead to performance-related issues affecting the FC (Yang et al. 2022). Exhaust temperature is crucial for marine diesel engines, as it indicates combustion efficiency and performance; excessively high exhaust temperatures can signify incomplete combustion or potential engine malfunctions, while optimal temperatures ensure effective operation and reduced emissions (Drazdauskas and Lebedevas 2024)

Scavenging is a vital process that expels exhaust gases from the combustion cylinder and replenishes it with fresh air. The scavenging temperature, representing the temperature of the incoming air mixture, significantly affects the density and mass of the air–fuel mixture, thereby influencing combustion efficiency and engine performance. Likewise, scavenging pressure is the pressure of the incoming air, essential for displacing residual exhaust gases and ensuring a sufficient supply of fresh air for effective combustion (Yu et al. 2024).

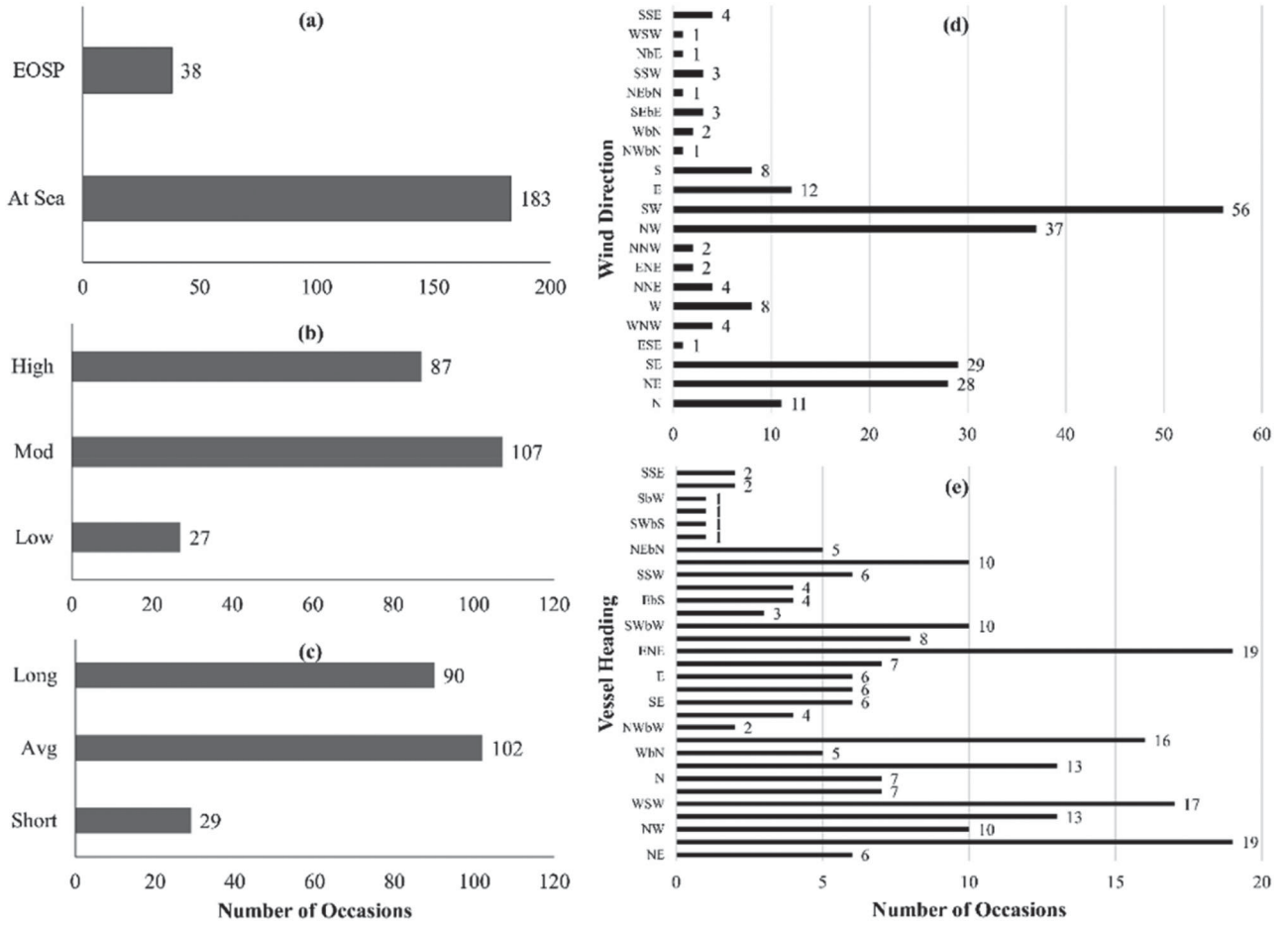

Figure 2. The categorical data briefing of the: (a) Event, (b) WaveHeight, (c) WaveLength, (d) WindDirection, (e) AvgHeading.

Table 3. The summary statistics of numerical features.

Metric	Minimum	Maximum	Mean	Standard deviation	75% percentile	50% percentile	25% percentile
Displacement	58, 723	131, 433	98, 256.61	27, 145.05	119, 846	110, 493	58, 877
Distance	1.00	336	232.01	92.35	291	272	207
BFScale	3.00	9	5.14	1.13	6	5	4
Speed	2.70	14.6	11.68	1.37	12.5	11.8	11.2
Slip	−44.00	54	9.62	9.36	14	10	5
SeaState	2.00	8	4.17	1.11	5	4	3
Draft	8.00	14.6	11.5	2.43	13.45	12.5	8
MinExhT	190	295	255.24	14.14	265	256	248
MaxExhT	227	332	281.52	16.46	290	286	275
SWT	4	32	22.94	7.66	30	25	18
ER_T	22	40	32.09	5.38	36	33	28
ScavT	1.2	48	36.90	6.42	40	39	36
ScavP	0.12	1.8	1.01	0.31	1.2	1.05	0.91
FC	0.1	39.3	24.88	11.12	32.9	29.8	17.2

The correlation between features is demonstrated by using Pearson's correlation matrix, as shown in Figure 3.

The highest correlation with FC was observed with distance, as illustrated in Figure 3. It can be anticipated that the algorithms will be developed using distance as a threshold value. The displacement, draft, BFScale, exhaust temperatures, scavenge pressure, sea state, and slip exhibited moderate collinearity with FC. However, this does not imply that the models will necessarily include all these variables or exclude those with low correlation to the target feature.

2.2. Mathematical and logical background

This subsection demonstrates the benchmarked algorithms, the calculations of their performance measurement metrics, and the computation of CII ratings. The algorithms were executed using the dataset in the WEKA software. The WEKA project aims to provide a comprehensive suite of machine learning techniques and data processing tools for both practitioners and researchers. It enables users to efficiently test and compare various ML algorithms by integrating new datasets. A variety of algorithms can be employed to

Table 4. The Beaufort scale (MetMatters 2025).

Force	Wind force	Wave Height (m)	Visual clues
0	0	0	Calm – The sea is flat
1	1–3	< 0.1524	Breeze – Very small ripples
2	4–6	mean 0.15, max 0.30	Light wind – Small distinct ripples
3	7–10	mean 0.61, max 0.9	Sweet wind – Ripple crests begin to break
4	11–16	mean 0.91, max 1.52	Medium wind – small waves get bigger
5	17–21	mean 1.83, max 2.44	Strong wind – White crests, and foaming waves
6	22–27	mean 2.74, max 3.66	Strong wind – large waves form
7	28–33	mean 3.96, max 5.79	Light storm – The sea begins to rise
8	34–40	mean 5.49, max 7.62	Storm – High waves occur around
9	41–47	mean 7.01, max 9.75	Strong storm – High wave crests begin to roll
10	48–55	mean 8.84, max 12.5	Full storm – the sea often appears white.
11	56–63	mean 11.28, max 15.85	Very severe storm
12	64+	13.716+	Tornado – The sky is covered with foam, and visibility is very low

Table 5. OHE example for the categorical variable 'WaveHeight'.

WaveHeight (Original Data)	WaveHeight_Low	WaveHeight_Mod	WaveHeight_High
Mod	0	1	0
Low	1	0	0
Mod	0	1	0
Mod	0	1	0
Mod	0	1	0
Low	1	0	0
Low	1	0	0
Mod	0	1	0
High	0	0	1
High	0	0	1

develop complex data mining operations, facilitated by the user-friendly interface (Frank et al. 2016). The formation and application logic of investigated ML algorithms are illustrated in the flowchart presented in Figure 4.

The model's training-test split ratio and hyperparameter tuning were determined based on the optimal results from several trials. A 10-fold CV and performance metrics were applied to the test set. Predictions of FC models were evaluated to assess the final error against the original data.

Categorical data, which refers to variables representing distinct classifications or groups, are often a challenge in ML algorithms due to its non-numeric nature. To effectively utilise categorical data in ML models, it is essential to convert it into a numerical format that can be interpreted by algorithms. One widely adopted technique to accomplish this transformation is the one-hot encoding (OHE) approach. The OHE method generates a separate binary column for each unique category within the categorical variable. This creates multiple new columns in the dataset, with each column corresponding to a distinct categorical value. Each original observation is then represented by a binary vector – an array composed exclusively of 0 s and 1 s. In this binary vector representation, a value of '1' signifies the presence of the corresponding categorical value at that specific index, while all other indices are assigned a value of '0' (Fawcett 2021). Table 5 illustrates an OHE methodology example applied to the 'WaveHeight' categorical column, utilising a selected subset of the dataset.

In Table 5, the categorical variable 'WaveHeight' is delineated, comprising three distinct values: 'Low', 'Mod', and 'High'. Upon the application of the OHE technique, three new binary columns are generated: 'WaveHeight_Low', 'WaveHeight_Mod', and 'WaveHeight_High'. For an observation classified under the 'Mod' category, the resulting encoded representation manifests as a binary vector of [0, 1, 0].

2.2.1. M5 model rules

One of the approaches for predicting the FC with varying environmental conditions and engine variables was the M5 model rules algorithm. The performance of this algorithm was compared to algorithms like SVR and MLR in our previous study (Yuksel et al. 2023). The results indicated that the algorithm excels in terms of understandability, applicability to small datasets, and suitability for FC prediction problems.

Quinlan (1992) introduced the concept of an M5 model tree, which combines binary decision tree structures with MLR algorithms at the leaf elements. This model is specifically designed for forecasting continuous numerical features. The first step is to use a division criterion to create a decision tree. The dividing criterion is the anticipated decrease in the error that results from evaluating every feature at that node. The error at a node is measured using the standard deviation of the class values that arrive at that node. The term is called standard deviation reduction (SDR) and is calculated by employing Equation 1 (Jothiprakash and Kote 2011).

$$SDR = sd(T) - \sum \frac{|T_i|}{T} sd(T_i) \quad (1)$$

The parameter T represents a set of instances reaching a specific node, with Ti denoting a subset of examples corresponding to the ith possible outcome, while sd signifies the standard deviation. During the splitting process, the data in child nodes typically shows greater purity and lower standard deviation than the parent node.

The M5 model systematically evaluates potential splits, selecting the one that maximises the reduction in anticipated error. However, this iterative division can result in a complex tree-like structure prone to overfitting. To mitigate this, pruning is necessary, which involves replacing a subtree with a leaf node. Consequently, the secondary phase of the model tree's design focuses on trimming the overgrown structure and substituting subtrees with MLR functions (Quinlan 1992; Pal and Deswal 2009). The model structure and its steps are illustrated in Figure 5 (Fang et al. 2020).

The process begins with inputting 20 instances into the M5 Rules Learner, which generates a pruned decision tree. From this tree, rules are derived based on the best leaf nodes. The algorithm iterates until all instances are encompassed within the established rules. Finally, the rules are constructed, producing the desired output (Fang et al. 2020).

Python was utilised to access the WEKA Application Programming Interface (API) in a grid search approach to optimise the model parameters. The maximum depth was examined from 0 to 10 in increments of 1, while the batch size was tested at intervals of 10–200. Following hyperparameter tuning of the M5 Model Rules, the final model was constructed using a batch size of 100 and a minimum number of instances set to 4 in the WEKA software.

2.2.2. Artificial neural networks

ANNs were first proposed as a component of computer systems intended to mimic the brain's information-processing powers. Nowadays, ANNs are a common computational framework for many applications, such as pattern recognition, prediction, and optimisation (Biçen and Celik 2024). Because they can learn from data and

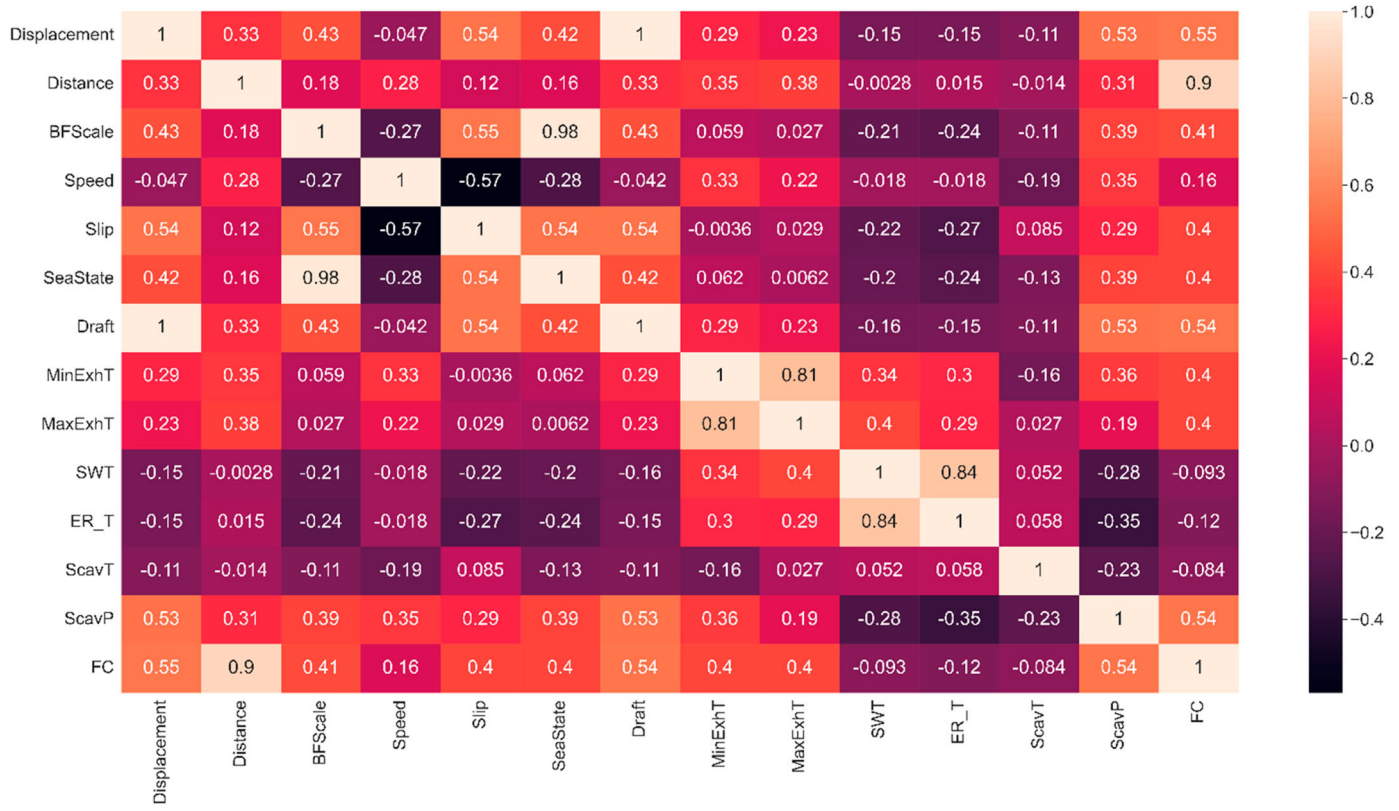

Figure 3. Pearson correlation matrix.

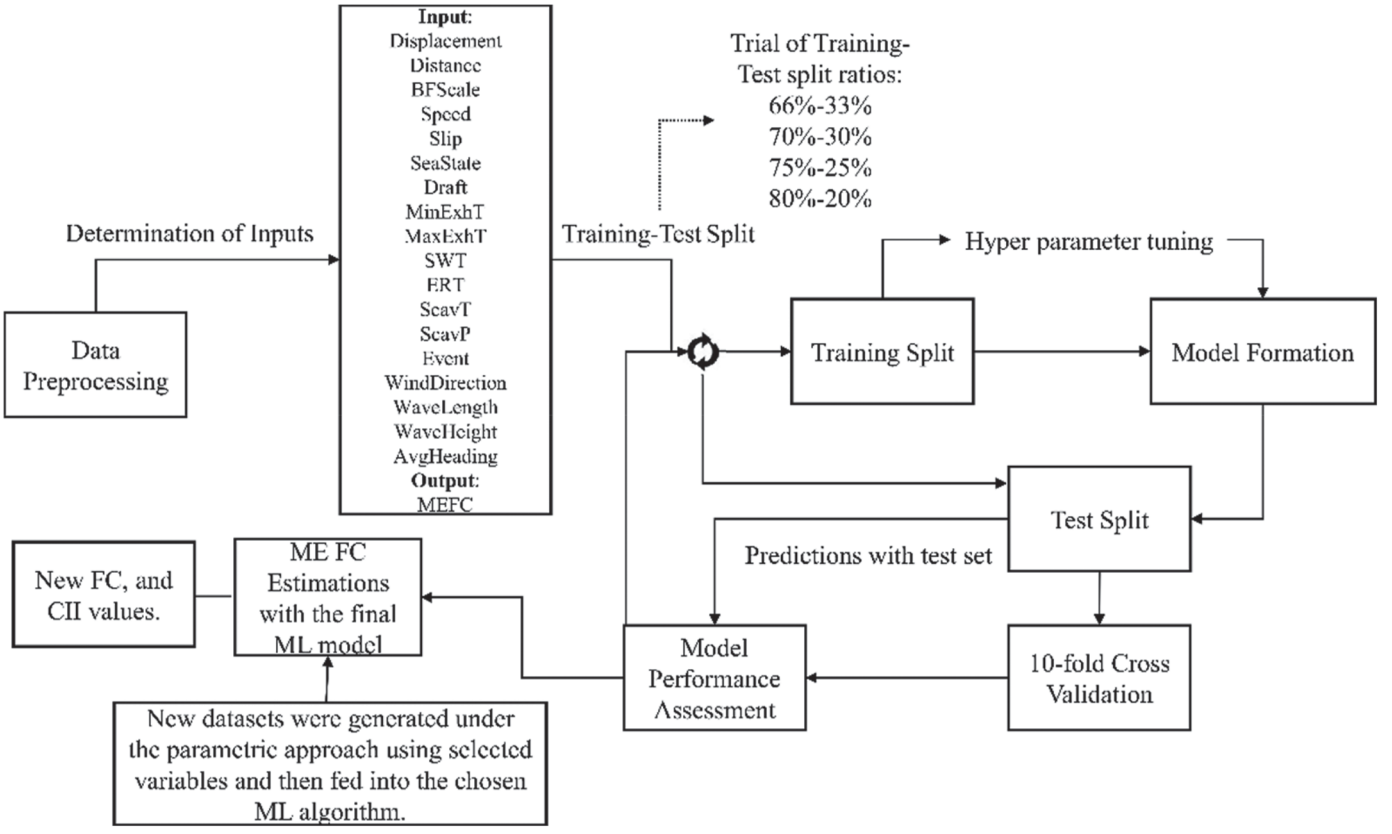

Figure 4. The flowchart of the ML methodology usage.

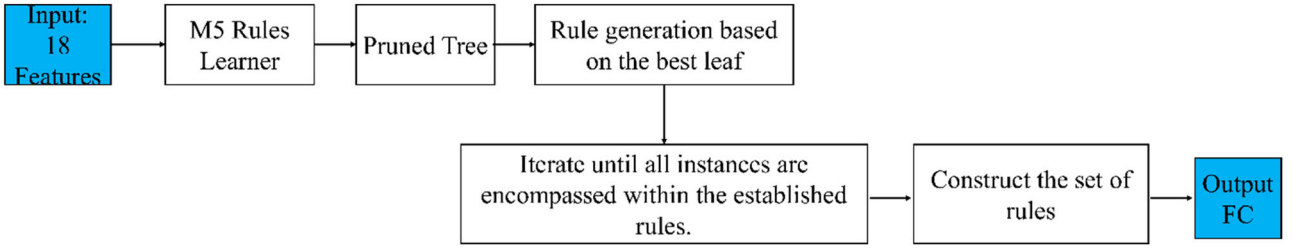

Figure 5. M5 rules algorithm stages and structure

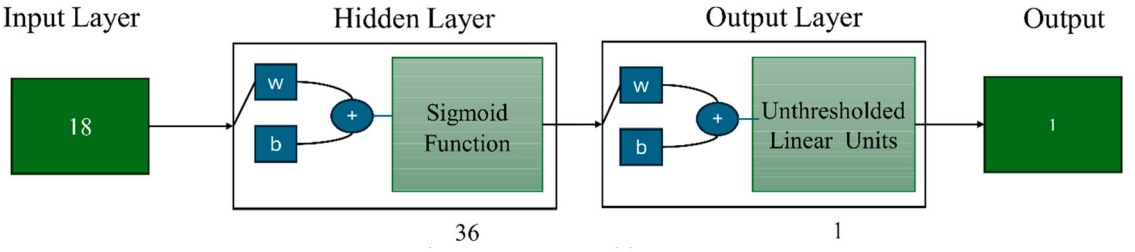

Figure 6. ANN architecture.

apply that learning to new situations, ANNs are widely used. As a result, ANNs are becoming a crucial tool for resolving challenging issues across a range of industries (Ertogan and Wilson 2024; Kurucan et al. 2024).

Three-layer types make up an ANN in general: an input layer, hidden layers, and an output layer (Bal Beşikçi et al. 2016). The ANN working principle is described in several key stages. Initially, the input layer gets the information, which serves as the entry point for data. This information is subsequently transmitted to the first hidden layer. Each neuron calculates a linear combination of the input parameter weights in this layer, followed by an activation function application. This activated output is then propagated to the next hidden layer, and this iterative process continues through multiple hidden layers. Ultimately, the data reaches the output layer, where the final prediction or output of the ANN is generated. This structured flow of information enables the ANN to learn and model complex relationships within the data (Yüksel and Köseoğlu 2020).

The multi-layer perceptron (MLP) algorithm was selected to implement the ANN in the software. An MLP is a modern FNN comprising fully connected neurons with nonlinear activation functions. It acts as a classifier that employs backpropagation for instance classification. The network can be built manually or via a simple heuristic, allowing for monitoring and modification of parameters during training. All nodes typically use sigmoid activation functions, except when the output class is numeric, leading to unrestricted linear units for output nodes (Cybenko 1989). Through an extensive hyperparameter tuning process utilising a grid search methodology within the WEKA API, optimal parameter values were identified to enhance model performance to its fullest potential. Figure 6 indicates the ANN architecture employed in the study after the hyper parameter tuning.

The learning rate and momentum parameters were systematically tested, ranging from 0.01 to 1 in increments of 0.01. The batch size was examined within the range of 10–200, with intervals of 10, while the number of hidden layers has varied from 1 to 3. The model was configured with a single hidden layer in accordance with the grid search findings. In this configuration, the batch size was set to 100, the learning rate was adjusted to 0.01, and the momentum parameter was established at 0.05. The hidden layer was configured to contain 36 neurons, with 'Sigmoid' as the hidden layer activation function and 'Unthresholded Linear Units' as the output layer transfer function. To optimise training, the model underwent 500 epochs, utilising these parameters to execute the algorithm for the analysis. Equation 2 demonstrates the sigmoid function (S(x)) (Ngah et al. 2016).

$$S(x) = \frac{1}{1 + e^{-Z}} \tag{2}$$

where e is the Euler's number, and x is the input vector in Equation 2. The parameter Z represents the input information transmitted to the hidden layer and can be computed utilising Equation 3 (Najafi et al. 2018).

$$Z = \sum_{i=1}^{40} w_{ij} \times p_i + b_j \tag{3}$$

The parameter w_{ij} denotes the weights connecting the neuron 'i' in the preceding layer to the neuron 'j' while p_i represents the outcome of neuron 'i'. Additionally, the bias is represented by b_j (Najafi et al. 2018). The transfer function of the output layer (g) is characterised as a linear unit, as delineated in Equation 4 (Rajoub 2020).

$$g = w_L \times Z + b_0 \tag{4}$$

The interconnection weights between the final hidden layer and the output layer are represented by w_L while b_0 denotes the bias associated with the output layer (Taghva et al. 2018).

2.2.3. Random forest

RF generates multiple decision trees and merges their outputs to produce more accurate and reliable estimations. Due to the tree-based structure of the model, the results are presented in a discrete format. The RF regression model is developed using the bagging technique, which involves repeatedly extracting samples from the dataset to create new trees. The RF is constructed by aggregating these trees, thereby enhancing the model's robustness and accuracy (Breiman 2001). This ensemble method leverages the diversity of individual trees to enhance predictive performance and reduce overfitting, making it a powerful tool for regression tasks (Uyanık et al. 2020).

In other words, given an input vector (x) that contains the values of various evidentiary aspects examined for a specific training

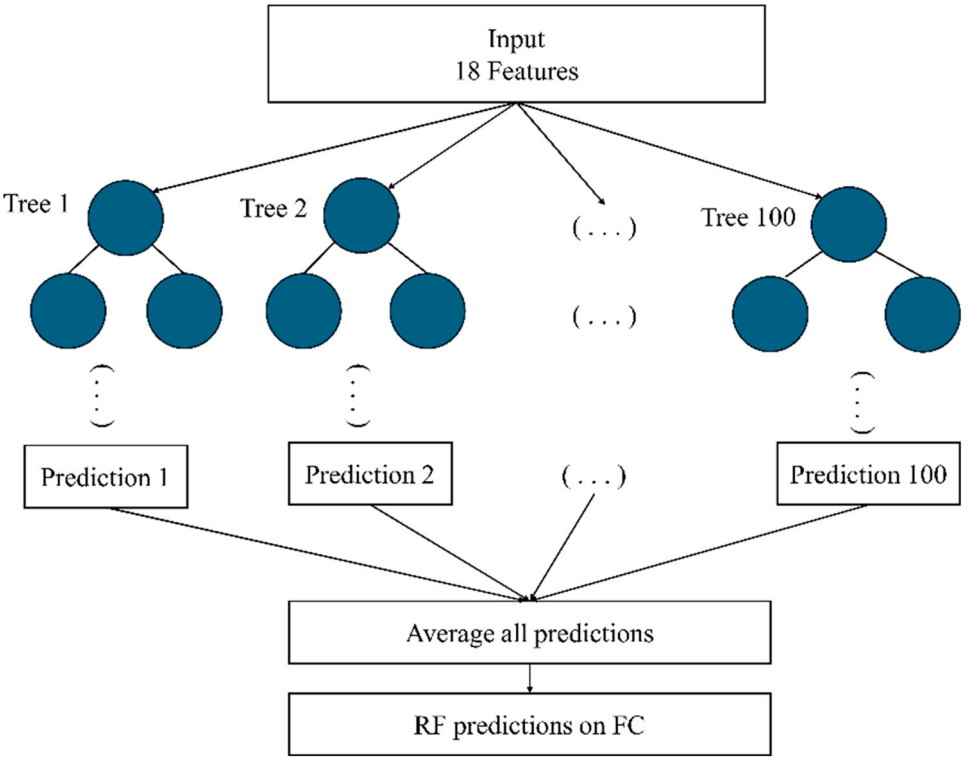

Figure 7. RF architecture.

region, the RF algorithm builds K regression trees. Each tree independently predicts the output based on the input vector. The final prediction of the RF is obtained by averaging the outcomes of these K regression trees. The $T(x)_1^K$ growth of K trees yields a regression predictor indicated in Equation 5 (Rodriguez-Galiano et al. 2015). Figure 7 illustrates the structure and simplified working principles of the RF model.

$$f_{RF} = \frac{1}{K} x \sum_{k=1}^{K} T(x) \tag{5}$$

The RF process involves an input of 18 features, which are fed into a random forest model comprising 100 decision trees. Each tree generates a prediction based on the input features. The final output is obtained by averaging the predictions from all 100 trees, resulting in the random forest predictions for the given feature set (Breiman 2001).

The identical grid search methodology employed in the ANN and M5 Rules was also utilised to optimise the RF parameters within the WEKA API. The batch size was evaluated in increments of ten, ranging from ten to two hundred, while the maximum depth was systematically tested from 0 to 10 in increments of one. Ultimately, the maximum depth was established at four, while the bagging size percentage and batch size were configured at 100, following the hyperparameter tuning of the RF model.

2.2.4. Model evaluation metrics

The R^2, MAE, relative absolute error (RAE), RMSE, and root relative square error (RRSE) assess ML models' performance. Equations 6–10 demonstrate the calculation process of the performance evaluation metrics demonstrated in Table 6.

In the formulas shown in Table 6, n represents the data number, y_p depicts the estimated or predicted variable, y_t is the true or original data, and $\bar{y}$ denotes the average of the target column (Chai and Draxler 2014). The well-known R^2 statistic, which indicates how closely the regression analyses estimated and actual outcomes relate to one another, is used to evaluate the success of models. A closer to one R^2 indicates a higher correlation and performance (Kasuya 2019). The MAE measures the variance between real and projected parameters. Consequently, a smaller MAE indicates greater accuracy in predictions (Willmott and Matsuura 2005). Another often-used statistical metric to assess the predictive power of models is the root mean square error (RMSE). Because the RMSE penalises large errors, but the MAE assigns equal weight to all errors, when both measures are established, the MAE usually has a lower average than the RMSE (Chai and Draxler 2014).

Table 6. Metrics used for the model performance evaluation (Chai and Draxler 2014; Kasuya 2019; Padhma 2021).

Metric	Equation	Equation No
R^2	$1 - \frac{\sum_{i=1}^{n} (y_p - y_t)^2}{\sum_{i=1}^{n} (y_t - \bar{y})^2}$	(6)
MAE	$\frac{\sum_{i=1}^{n} \lvert y_p - y_t \rvert}{n}$	(7)
RMSE	$\sqrt{\frac{\sum_{i=1}^{n} (y_p - y_t)^2}{n}}$	(8)
RAE	$\frac{\sum_{i=1}^{n} \lvert y_p - y_t \rvert}{n \sum_{i=1}^{n} \lvert \bar{y} - y_t \rvert}$	(9)
RRSE	$\sqrt{\frac{\sum_{i=1}^{n} (y_p - y_t)^2}{\sum_{i=1}^{n} (\bar{y} - y_t)^2}}$	(10)

A 10-fold CV separating the data into ten different parts to avoid the impact of randomness was conducted for each ML model. Their performance for each fold has been evaluated and illustrated. The average of the metrics in each fold has been taken as the final CV metrics. In addition to the CV, various data training and test split ratios (66%−33%, 70%−30%, 75%−25%, and 80%−20%) were tried for each algorithm. The highest performer was selected to perform the variances assessed in the parametric study.

2.2.5. The parametric study

In the initial phase of the parametric study, the model was provided with the minimum, maximum, and average values of selected input variables to investigate the extreme effects and delineate the boundaries of variation. Concurrently, the other input variables were held constant to isolate the specific impact of the parameter under examination. These control parameters were established based on the results obtained from the highest-performing algorithm, as detailed in the results section.

Another criterion for selecting the control variables was to ensure their independence from other variables to the greatest extent possible. For instance, while changes in speed or distance can influence all parameters, there was insufficient data to effectively model such interdependencies. Consequently, the selected control variables either do not affect others or their effects can be adequately modelled. For example, the BFScale can impact the slip parameter, which is modelled using Equation 11. The remaining parameters were adjusted independently to maintain this isolation in the analysis.

$$Slip = 4.5562 * BFScale - 13.7795 \quad (11)$$

Equation 11 was derived from the data using the M5 model rules with a 75-25 train-test split ratio. In the second phase of the parametric study, the two most influential parameters were examined using a normal distribution (ND) approach. This methodology was adopted because, in real-time operations, conditions are unlikely to remain consistently extreme. Two NDs for each selected input variable were generated using the truncnorm function from the rv_continuous class in the Python SciPy library. The first dataset utilises the minimum value as the lower bound and the mean of the 221 observations for the corresponding parameter as the upper bound. The second dataset is established with the mean as the starting point, extending to the maximum value.

2.2.6. Carbon intensity calculations

The FC is predicted using the new datasets, and the potential impacts on CII ratings are calculated. The resulting CII computation is ensured by employing Equation 12 (IMO 2021; Bayraktar and Yuksel 2023).

$$Attained\ CII = \frac{Annual\ CO_2(t)}{Distance\ (NM)x\ Capacity\ (t)} \quad (12)$$

The capacity is DWT in tons for tanker vessels, and operational CO_2 is found by multiplying the FC of A/Es and ME with 3.206 (MDO) and 3.114 (HFO), respectively (Yuksel 2023). SFC is taken at 187 g/kWh at 85% load. The required CII (CII_{req}) is computed by applying Equation 13 (IMO 2021; Yuksel 2023).

$$CII_{req} = \left(1 - \frac{Z}{100}\right) x5247xDWT^{-0.61} \quad (13)$$

The reduction factor (Z) is determined at 5% for 2023 compared to reference CII calculated from a regression line in 2019 shown in Equation 10 (IMO 2021; Yuksel 2023). The ratings of CII are decided considering the dd vectors given for tanker vessels. Figure 8 indicates the dd vectors used in the analysis and illustrates the scale of CII ratings (ClassNK 2021; Bayraktar et al. 2023).

The rating limit points can be established by multiplying the coefficient depicted in Figure 5 by the CIIreq. By evaluating these bounds in conjunction with the attained CII, it is possible to determine the ship's rating concerning the Z for a specific year (IMO 2021).

3. Findings

The section was divided into four subsections, which are the 'Algorithm Benchmarking', 'Selected Algorithm Performance Results', 'Parametric Study of Engine Variables and Environmental Impacts on FC and CII', and 'Discussion'.

Table 7. Performance metrics of algorithms by train-test split and CV.

	M5 rules				
Training-test split	66–33	70–30	75–25	80–20	CV
R^2	0.9848	0.9757	0.9652	***0.9888***	0.949
MAE	1.4023	1.5986	1.7973	***1.2828***	1.6053
RMSE	1.934	2.4418	2.9103	***1.7047***	2.4091
RAE	15.59%	17.97%	20.27%	***14.30%***	17.18%
RRSE	17.29%	21.85%	26.31%	***15.00%***	21.57%
RF					
Training-test split	66–33	70–30	75–25	80–20	CV
R^2	***0.9486***	0.9401	0.9434	0.9198	0.943
MAE	***3.2646***	3.3591	3.2145	3.851	1.903
RMSE	***4.4826***	4.5568	4.3876	5.1256	4.965
RAE	***36.30%***	37.76%	36.25%	42.94%	37.92%
RRSE	***40.08%***	40.78%	39.67%	45.11%	44.46%
ANN					
Training-test split	66-33	70-30	75-25	80-20	CV
R^2	0.9604	0.9626	0.9619	***0.9823***	0.95
MAE	2.136	2.2776	2.3007	***1.6443***	1.617
RMSE	3.0989	3.1159	3.1995	***2.1143***	2.6261
RAE	23.75%	25.60%	25.94%	***18.34%***	20.37%
RRSE	27.71%	27.88%	28.92%	***18.61%***	23.52%
Test data size	75	66	55	44	221

3.1. Algorithm benchmarking

The M5 Rules, RF, and ANN models was run with the data set using different training test splits, and their performance was measured by metrics performed utilising the test set. The 10-fold CV provided validation and insurance against overfitting. Table 7 illustrates the performance metrics regarding the different split ratios and CV for each algorithm.

Best performances in identifying the training-test split ratios are highlighted in bold and italics in Table 7. The M5 Rules algorithm achieved the smallest error with an 80-20 training-test split. The ANN and RF algorithms ranked second and third in terms of performance, respectively. Notably, the RF demonstrated its optimal outcomes using a 66-33 split, setting it apart from the other algorithms.

The performance rankings of the algorithms remained consistent when evaluated using the CV metrics. M5 Rules and the ANN performed similarly in the CV; however, the RF metrics lagged. While the R^2 values for each model were comparable, the ANN achieved the highest R^2, whereas the M5 Rules algorithm was recorded the lowest MAE. Figure 9 illustrates the CV performance for each algorithm across the respective folds.

The average performance metrics for each fold provided the overall CV performance of the algorithms, as shown in Table 7. In terms of stability, each algorithm has performed adequately; however, RF demonstrated the least variance in the R^2 metric. The M5 Rules algorithm has achieved the lowest average MAE, whereas the ANN has produced the highest mean R^2.

The highest and the lowest MAE for the M5 Rules algorithm were recorded in the 8th and 1st folds, respectively, as illustrated in Figure 7(a). For the ANN, the metrics varied in the 9th and 3rd folds, as shown in Figure 7(b), while the RF metrics fluctuated in the 4th and 8th folds. Although the CV outcomes significantly differentiated between each algorithm, the M5 Rules algorithm yielded more accurate predictions when considering the different training and test split ratios. Table 8 presents a comparative analysis of the algorithms, focusing on training time, interpretability, maximum achieved accuracy, and the ease of hyperparameter tuning.

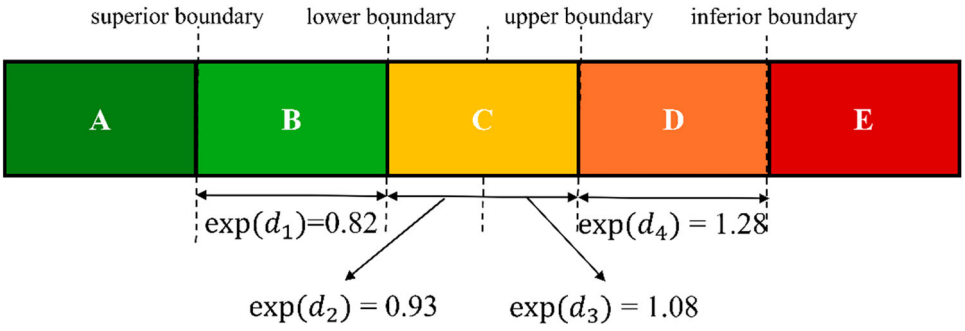

Figure 8. CII rating scale.

Table 8. Algorithm comparison.

Model	Training time	Interpretability	Maximum obtained R^2	Hyperparameter tuning
ANN	2.21 s	Low	0.9823	Hard
M5 rules	0.38 s	High	0.9888	Easy
RF	0.03 s	Low	0.9486	Easy

Training times were measured on a laptop equipped with a 13th Generation Intel® Core™ i7-13850HX processor, 32 GB of random-access memory, and an NVIDIA RTX 2000 Ada Generation graphics card featuring 16 GB of dedicated graphics memory. ANN exhibited the slowest training times, with M5 ranking second and RF being the fastest to build in WEKA. The authors assessed the 'Interpretability' in terms of the understandability and usability of the output formula, as well as the overall comprehensibility of the methods for an audience with limited ML knowledge. In this context, both RF and ANN output formulas proved challenging to interpret, whereas M5 Rules delivered the most comprehensible results for the maritime audience targeted in the analysis. Regarding performance, as discussed in previous sections, M5 achieved the highest accuracy, closely followed by ANN, while RF slightly lagged. The ease of hyperparameter tuning was evaluated based on the simplicity of model construction and the number of variables involved. RF and M5 demonstrated a more straightforward tuning process, whereas the ANN model required extensive adjustments and time to optimise for the best outcomes. Consequently, a parametric analysis was conducted to investigate changes in carbon intensity based on variations in environmental parameters, utilising the M5 Rules algorithm.

3.2. Selected algorithm performance results

Considering the findings and the potential additional benefits mentioned above, the FC prediction model provided by the M5 Rules algorithm with an 80%-20% training-test split was utilised in the parametric investigation. Figure 10 compares the real test data and the predicted values from the selected algorithm, along with the error distribution in instances.

The model predictions compared to the true values in the test data was considerably successful and consistent, as observed in Figure 10(a). The maximum error was noted in the 19th instance at 5.12 t of deviation, followed by the 32nd data point with a decline of 4.23 t, as shown in Figure 10(b).

The MAE was calculated at 1.283, yielding a total FC of 5713.25 t when the algorithm runs with the whole dataset. The vessel FC for a year has been 5499.02 t in the original data, which means the algorithm miscalculated 214.23 t in total, resulting in a 3.75% deviation. This error rate was found satisfactory, considering the low number of rows in the data and the possibility of faulty entries onboard.

The following benchmarks for CO_2 emissions and CII ratings were carried out using the results of the M5 model rules, since the model calculated the FC in the parametric study. Figure 11 benchmarks actual-predicted data and errors in each CV fold.

The distribution of R^2 and MAE metrics across each fold was well-balanced. The highest MAE occured in the fourth fold with the third lowest R^2, while the lowest R^2 is found in the first fold. These values were not significantly underperforming, and in the other folds, the algorithm performs satisfactorily.

The high MAE in the fourth fold was linked to negative values predicted by the model. Since the actual values were close to zero, this suggests that these data points represent EOSP or drifting operations, leading to slight miscalculations by the model. The rules established by the M5 algorithm are illustrated in Figure 12.

The algorithm identified the distance column, which exhibited the highest collinearity with the target column, as a threshold parameter, shaping the rules accordingly. The first two rules incorporated categorical variables, utilising separate if-else constructs for adaptation. For example, if AvgHeading is 'SbE', the coefficient is 0.0922, while it is −2.3106 for 'SSW'. The OHE of categorical values was detailed in the methodology section. The third rule introduces a regression line based on highly correlated dependent variables.

The selected parameters by the M5 Rules algorithm were found to have the most substantial influence on FC due to their significant collinearity with the target variable. By prioritising the parameters with the highest collinearity to fuel consumption, M5 Rules effectively constructed a final rule that maximises predictive accuracy while minimising complexity. According to the coefficients/rules established by the M5 Rules algorithm, the engine variables and environmental conditions that had most significantly impacted FC are SWT, BFScale/Slip, MaxExhT, ScavT, and ScavP. Thus, the parametric study focused on investigating the variations of these variables.

The CTGANSynthesizer from the Python Synthetic Data Vault (SDV) library was employed to simulate a realistic scenario, utilising a randomised set of inputs. The parameters for the CTGANSynthesizer were configured as follows: learning rate of 0.0001, 1000 epochs, an embedding dimension of 256, and generator and discriminator dimensions of (512, 216). Approximately one and half months of navigation data, comprising 43 rows, were generated. Table 9 provides a preview of the synthetic data sample.

This approach enables the generation of synthetic data that closely resembles the characteristics of real-world datasets. Subsequently,

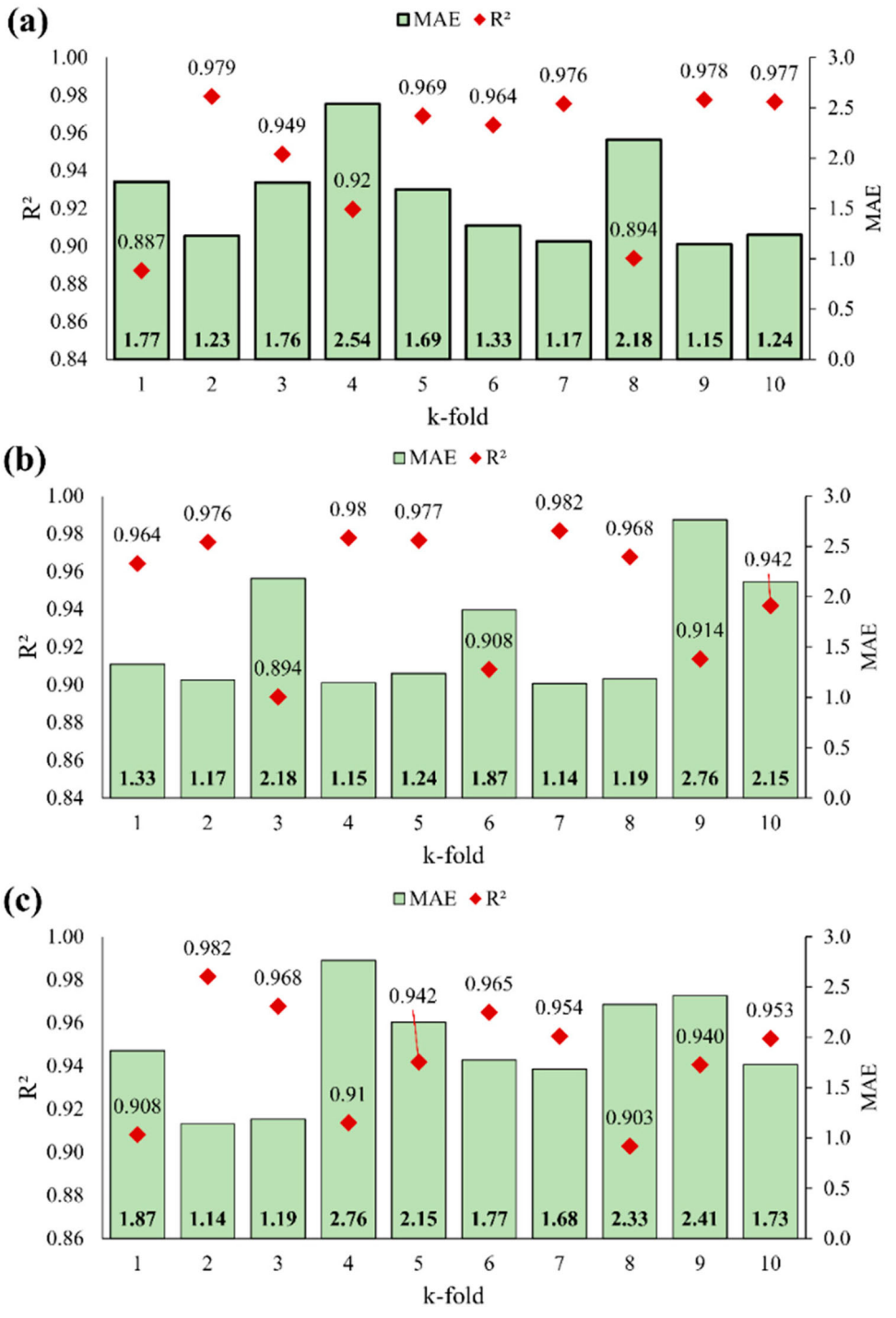

Figure 9. The CV performances of (a) M5 Rules, (b) ANN, and (c) RF.

the performance of the M5 Rules algorithm was evaluated using this synthetic data, specifically testing its efficacy with unseen inputs that mimic actual operational conditions. This methodology not only validates the algorithm's robustness but also enhances its applicability in practical settings. Figure 13 illustrates the performance of the M5 Rules algorithm when applied to the unseen data presented in Table 9.

The MAE of the M5 Rules algorithm's FC predictions was 1.52 t, and the R^2 was 0.8938. It should be noted that the synthetic data contains some noise and uncertainties introduced by the synthesiser. Since the M5 Rules algorithm was built on real-world data, these R^2 and MAE values are quite satisfactory. An ideal approach, given the ease of building the M5 Rules algorithm, is to create a new set of rules by training the model with new data.

3.3. Parametric study

Figure 14 illustrates the FC estimations and variations in comparison to predictions using the original data set. These results were obtained when the model is provided with data encompassing the minimum, mean, and maximum values of the control parameters.

In Figure 14(a), the FC variation of the selected control parameters was adjusted to minimum, mean, and maximum values and

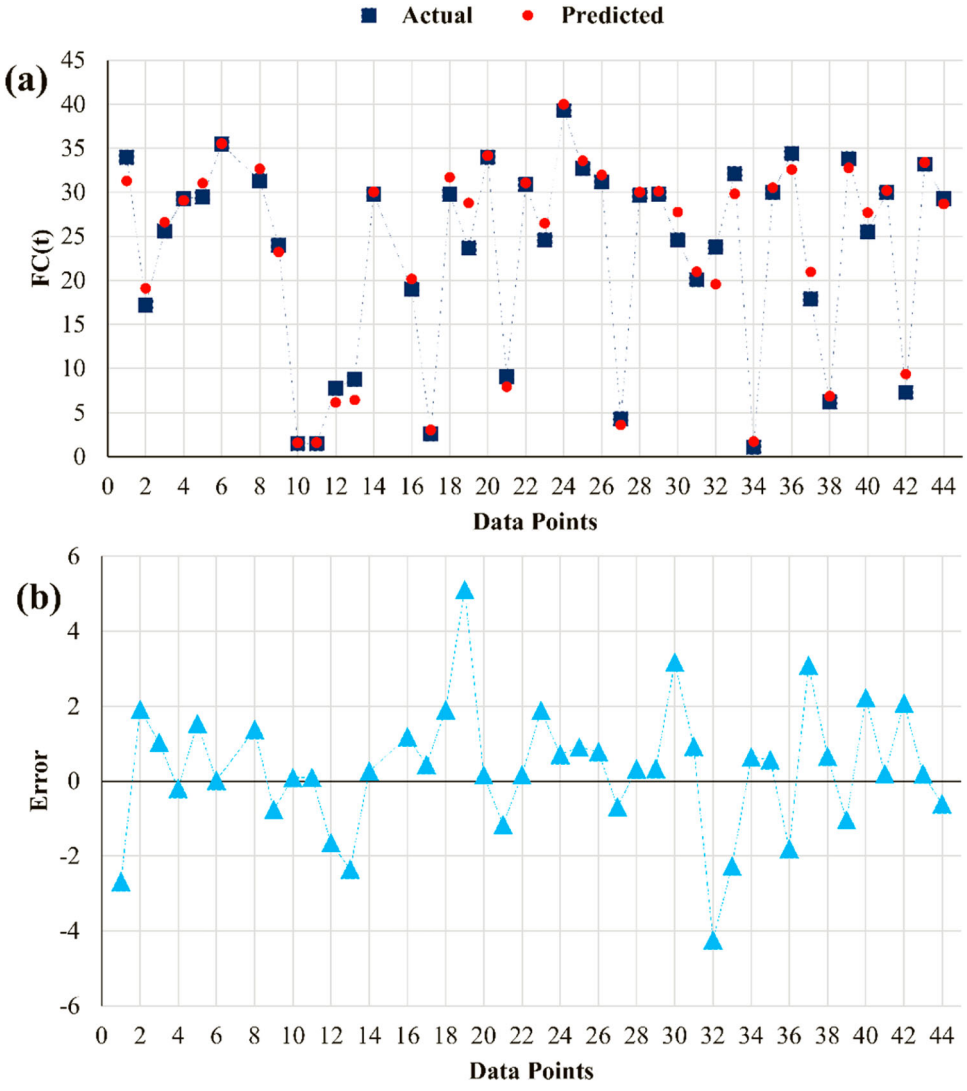

Figure 10. The comparison of actual and predicted values (a) and error distribution (b).

Table 9. Synthetic test data mimicking a real-world navigation scenario.

Displacement	Event	Distance	BFScale	Speed	Slip
111585	At Sea	294	6	4	15
58723	At Sea	224	3	12.9	8
123579	At Sea	285	6	12	2
117694	At Sea	215	4	11	16
118004	At Sea	250	6	12	25
WindDirection	**SeaState**	**WaveLength**	**WaveHeight**	**AvgHeading**	**Draft**
NW	8	Avg	High	NNW	12.78
NW	4	Avg	Mod	NE	9.06
WSW	4	Long	Low	SWbW	12.03
NEbN	3	Long	High	E	14.6
SW	3	Long	High	N	14.57
MinExhT	**MaxExhT**	**SWT**	**ER_T**	**ScavT**	**ScavP**
255	274	29	29	25.5	1.05
242	281	27	33	30.6	1.27
243	288	29	33	36.5	1.21
267	260	29	39	34.9	0.74
265	268	11	30	35.7	0.99

fed into the M5 Rules algorithm, as depicted in Figure 12. The consumption of A/E was 651.37 t, resulting in emissions of 2088.3 t of CO_2, which is used as a constant in calculations. The most impactful control parameters on FC were identified as scavenge pressure and BF Scale/Slip, while the maximum exhaust temperature had minimal effect.

Assuming the ship operates with a BF Scale of 9 for a year, the FC increased by 36.55%, equating to 2088.05 t annually. Conversely, adjusting the BF Scale to its minimum of 3 achieved a FC reduction of 23.48% (1341.27 t). Similarly, a reduction in scavenge pressure decreased the FC by 859.99 t (−15.05%), while increasing it results in an FC rise of 733.72 t (12.84%). Navigating in cold waters could yield fuel savings of up to 227.38 t for this vessel. However, a reduced scavenge temperature negatively impacted FC, increasing it by 395.99 t.

Notably, parameters with a significant impact can trigger increased CO_2 emissions, resulting in a higher attained CII for the vessel. The base scenario CII was calculated at 7.65 using the FC predicted by the M5 Rules, while the original data yielded a CII of 7.39. To simulate a more realistic change scenario, the analysis focused on the variables with the highest impact on FC: ScavP and BF Scale/Slip. Figure 15 displays the distributions of BF Scale/Slip and ScavP used in the parametric study and inputted into the ML model.

Since the BFScale and Slip parameters are interrelated, the NDs for BFScale were created, with Slip values subsequently calculated using Equation 11. The 'Mean to Maximum' NDs were adjusted to reflect a distribution that begins at the mean of 221 data points and

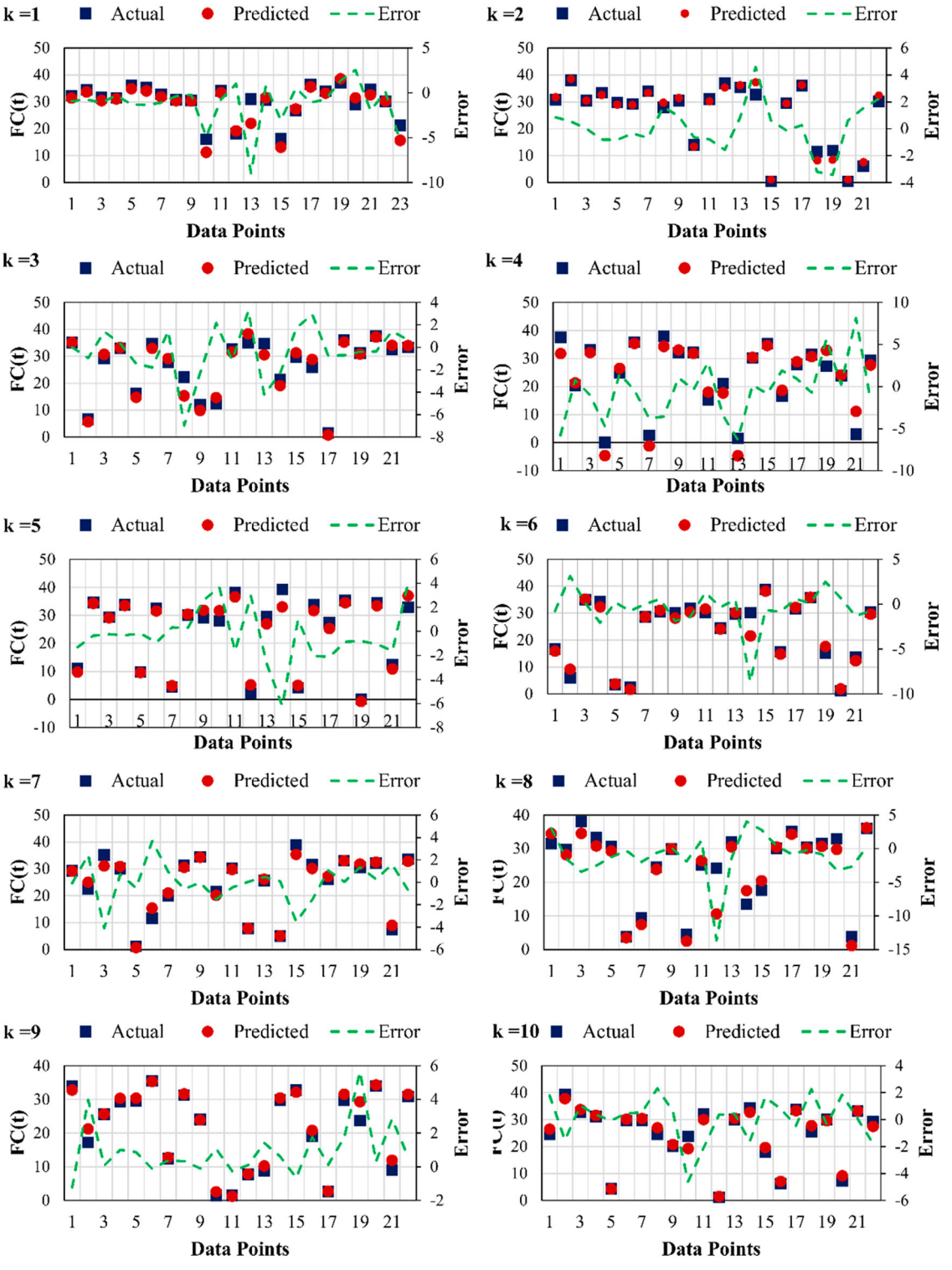

Figure 11. The benchmark of the actual-predicted data and errors of each fold.

extends to the maximum value. Similarly, the 'Min to Mean' NDs, illustrated in Figure 15, were constructed starting from the minimum value and concluding at the average.

Running the M5 Rules model, as presented in the previous subsection, with the data shown in Figure 15 yielded more moderate variation rates applicable to the daily operations of the vessel regarding FC. Figure 16 illustrates the predicted FC for both parameters and data sets, along with the change ratio compared to the predicted FC for the base case.

The BFScale/Slip, identified as the most influential control parameter on FC, resulted in a 7.63% increase in FC when the data was distributed normally from the mean to the maximum, using the standard deviation of the original data. Conversely, there was a 10.46% reduction in FC when the distribution ranged between the minimum and the mean. Similarly, scavenge pressure led to a 3.48% increase in FC and a 4.33% decrease. These intervals were more pronounced when real operational states were considered. Figure 14(d) illustrates the variation in attained CIIs

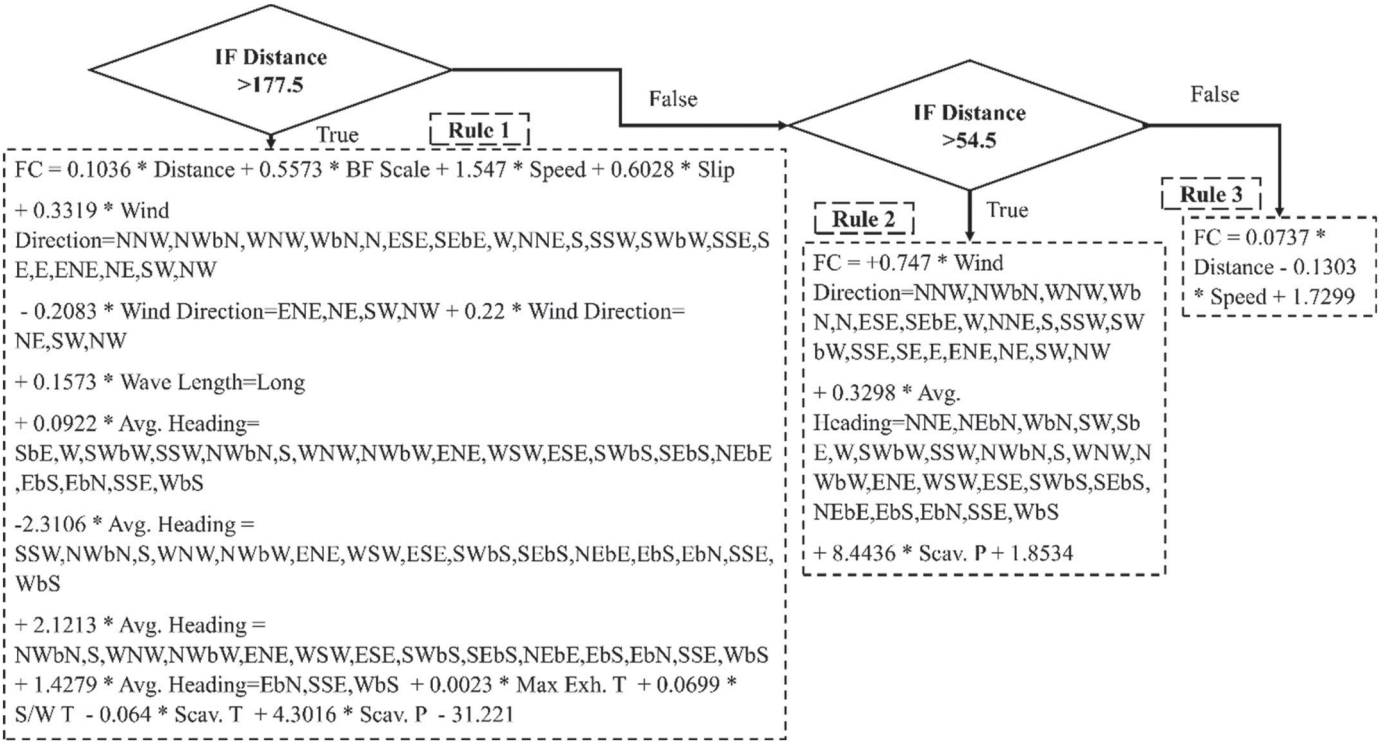

Figure 12. The determined rules of the M5 model rules algorithm.

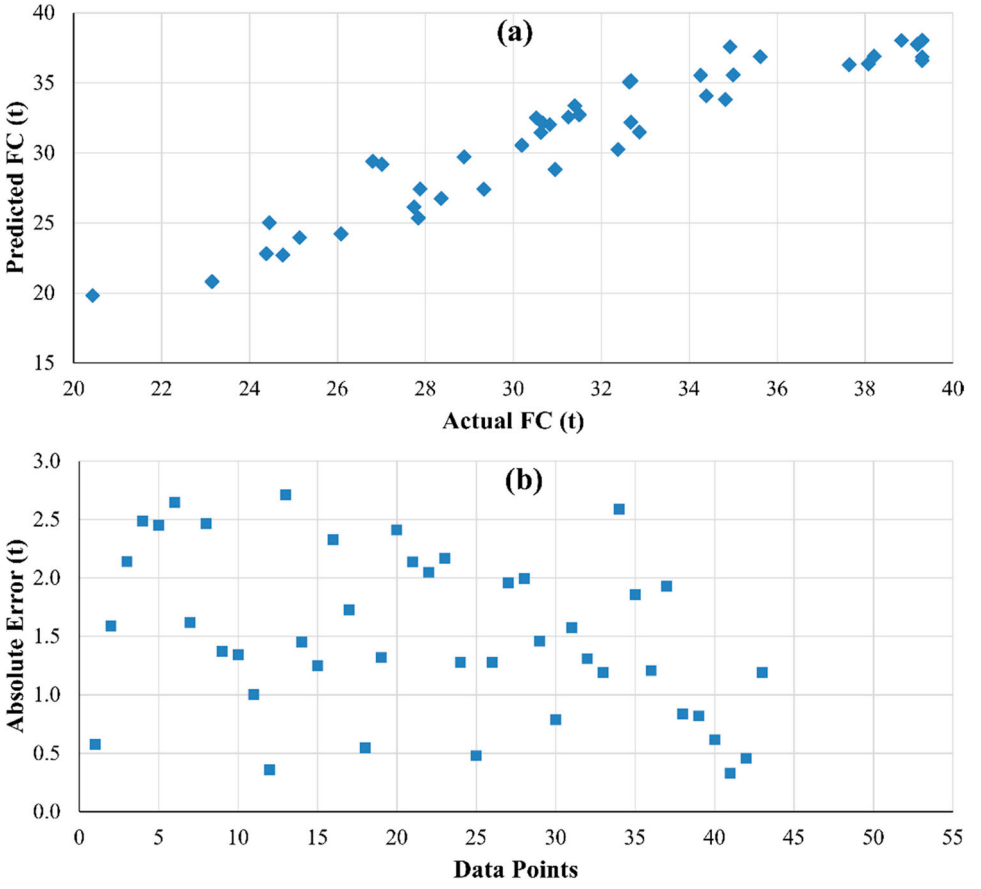

Figure 13. Predicted vs actual FC (a) and absolute error (c) with the synthetic data.

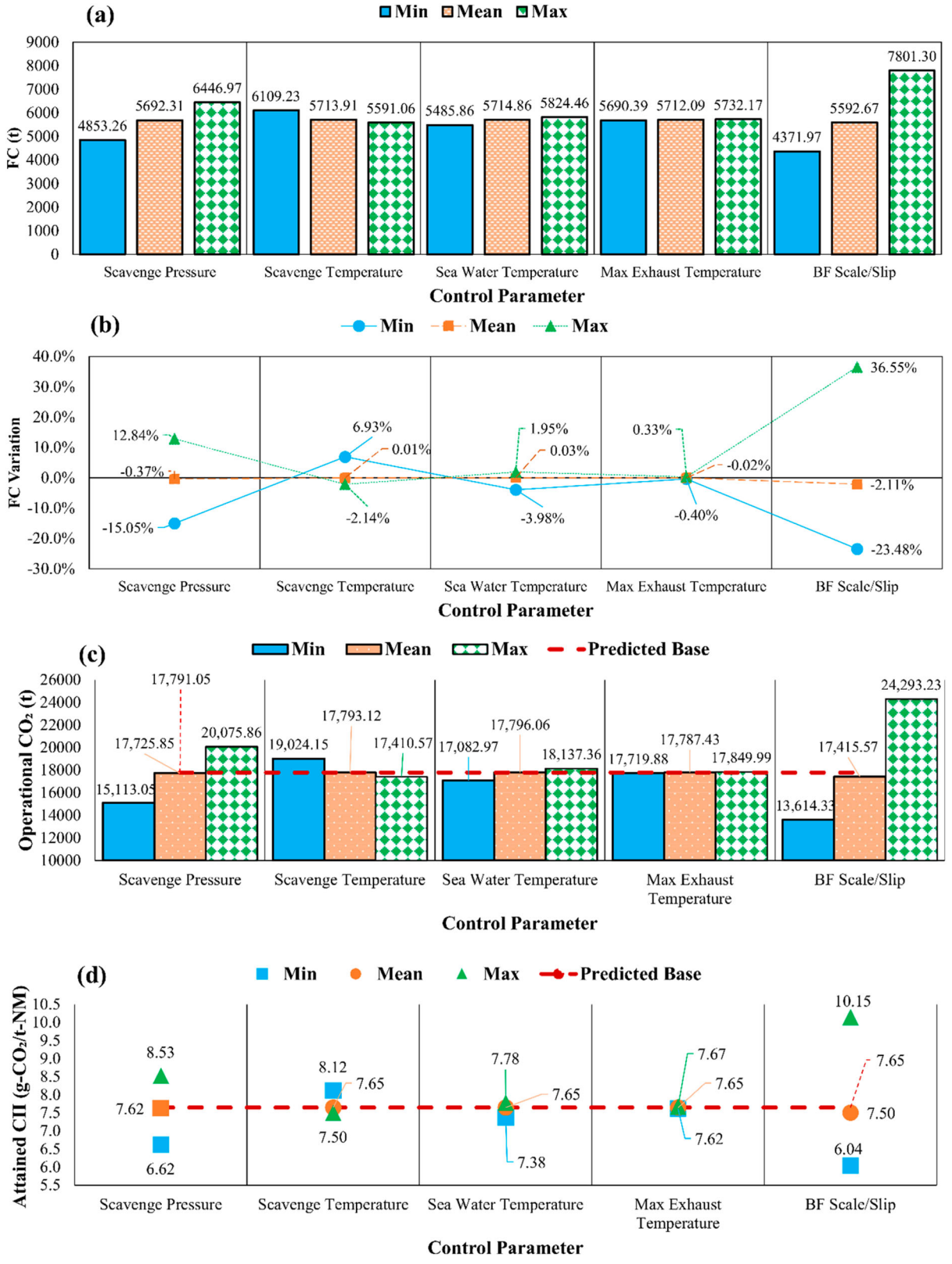

Figure 14. ME FC estimations (a), their variations compared to the predictions with the original data set (b), CO_2 estimations (c), and CII calculations (d) of the parametric study.

under extreme conditions, while Figure 17 depicts the change in CII ratings, reflecting the impact of parameters under such conditions.

The base scenario rating was D, requiring corrective action in subsequent years. Minimising the scavenge pressure improved the rating to a C level. Similarly, operating with a minimum BFScale/Slip

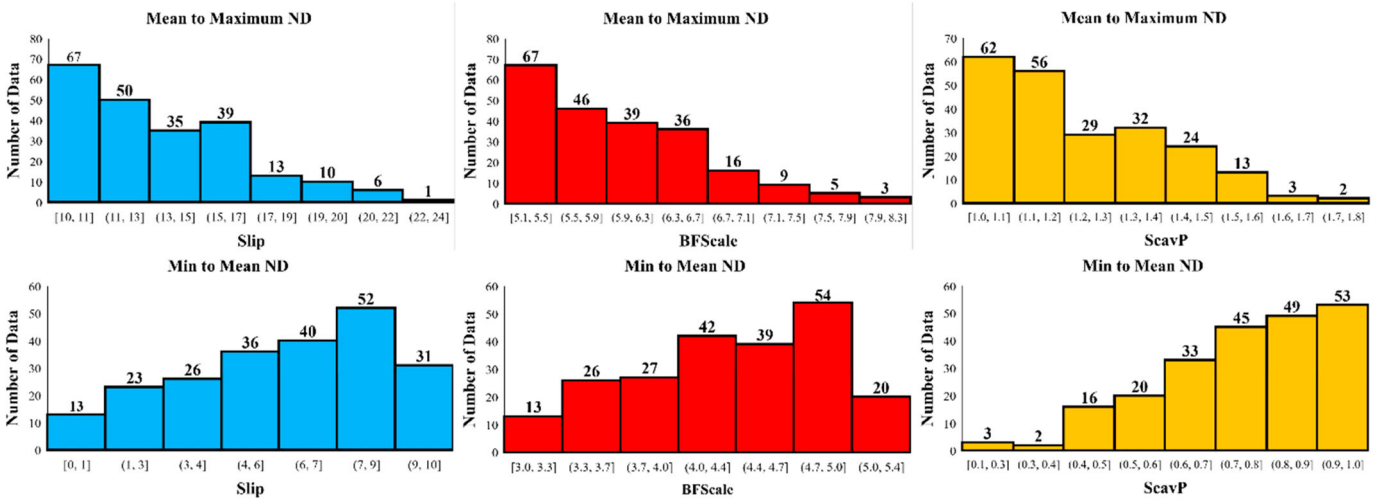

Figure 15. NDs for BF Scale/Slip and ScavP.

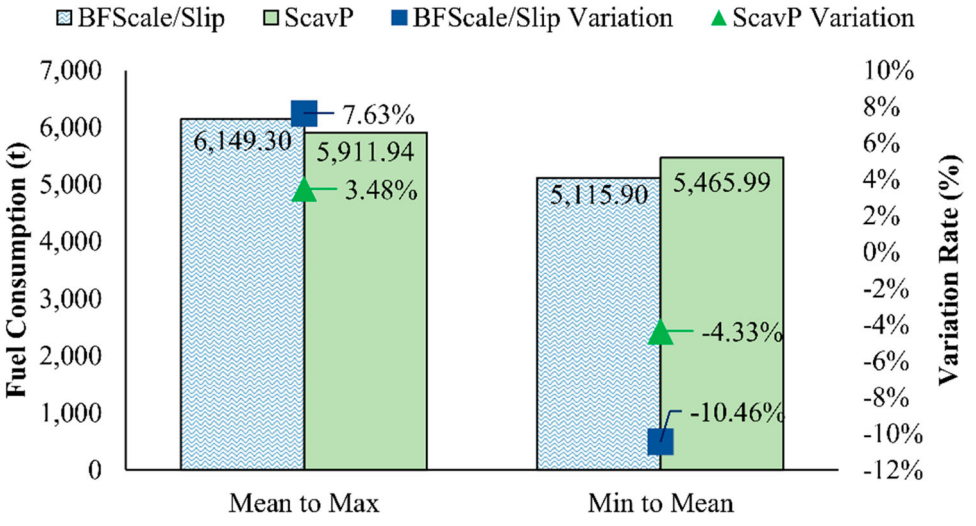

Figure 16. FC and change rates compared to the predicted FC for the base scenario.

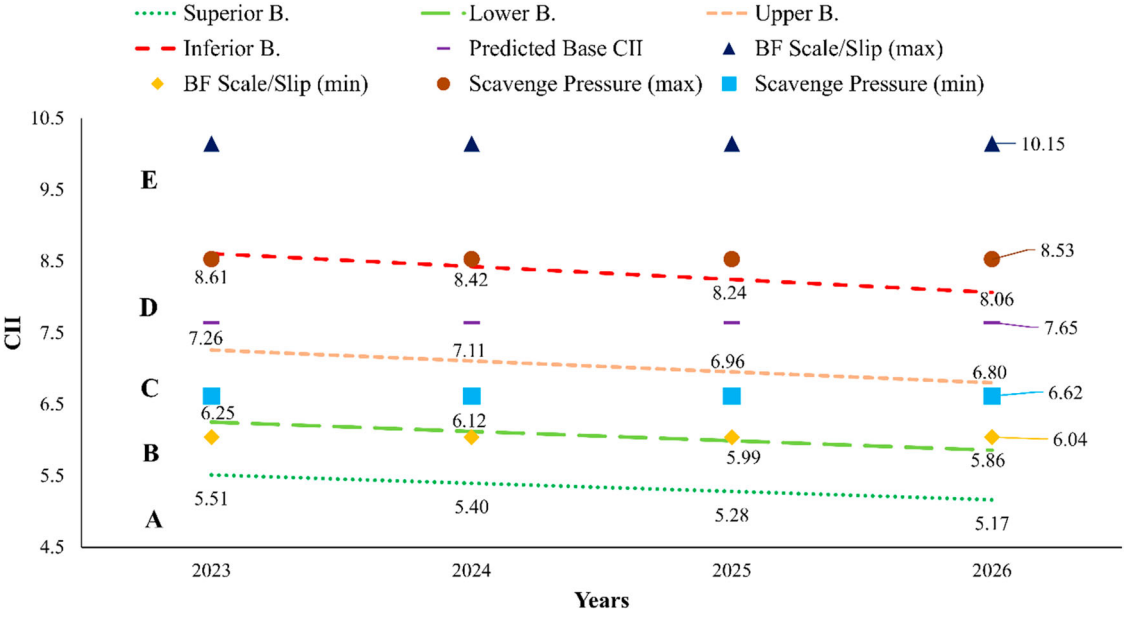

Figure 17. CII Ratings of some scenarios that have a notable change in FC.

enhanced the rating to C at the beginning of 2025. In contrast, the vessel's CII rating dropped to E when the worst-case scenarios for these parameters are applied.

The fluctuations in CII ratings highlight the sensitivity of vessel performance to operational parameters. This phenomenon underscores the importance of optimising control variables to maintain or improve efficiency and compliance with energy efficiency standards. The significant drop in ratings under worst-case scenarios also emphasises the potential risks associated with suboptimal operational practices.

3.4. Discussion

The most influential operational variables affecting FC were identified as BFScale/Slip and ScavP, following the distance travelled. Weather conditions have been shown to significantly impact ship performance (Bialystocki and Konovessis 2016), while optimised engine operation also plays a critical role (Park et al. 2023). In this context, the findings of this study align with the existing literature, reinforcing the importance of considering both environmental and operational factors in enhancing vessel efficiency.

The M5 Rules algorithm achieved the predictions with R^2 reaching 0.9888, outperforming the ANN and RF for the examined dataset. Additionally, when compared to algorithms discussed in the literature, the M5 Rules algorithm demonstrated competitive performance. Despite the advantages of ADLM-based data collection, which provides more reliable and frequent data, the maritime industry still heavily relies on NRs for measuring ship performance. This reliance is primarily attributed to the easy implementation and lower cost of using NRs (Safaei et al. 2018; Zwart et al. 2023).

Due to the human error factor, data losses in the NR records are significant, resulting in a remaining dataset that is even smaller in size compared to data collected using an ADLM system (Gkerekos et al. 2019). In addition, the maritime education system and the current structure of the sector's workforce do not typically include extensive knowledge of ML or artificial intelligence (Oladapo Adeboye et al. 2024). Considering these factors, beyond its high accuracy, the algorithm's shorter run time, greater comprehensibility, and ability to work with NR-based limited data relative to the other algorithms investigated, make it an excellent choice for this type of application (Baumann and Klingauf 2020).

The integration of the parametric approach with the M5 Rules algorithm can be easily applied across an entire company's fleet, offering improved weather routing and optimisation of engine parameters for the vessels. Particularly, shifts in the CII rating can become critical when carbon prices rise and when these changes are assessed at the fleet level (Bayraktar et al. 2023). This FC estimation approach can be integrated with life cycle cost analysis to evaluate the GHG emissions associated with alternative fuels throughout all phases of the fuel life cycle, including extraction/refinery, transportation, and storage (Taghavifar and Perera 2023c). Operational efficiencies, electricity mixes, and transportation dynamics significantly affect GHG emissions. Furthermore, assessing carbon credit rates under different carbon allowance frameworks for alternative fuel-powered vessels will enhance the life-cycle carbon emission cost assessment (Taghavifar and Perera 2023b).

The proposed approach can play a crucial role in helping shipping companies prepare for the EEXI/CII regulations and decarbonisation targets by simulating various case scenarios. Employing ML-based techniques integrated with the proposed parametric approach, even with NRs, can significantly reduce costs while enhancing environmental impact (Le et al. 2024). Further improvements in these methods can be achieved by reducing human error in NR recording through company policies, more frequent training, and the implementation of automated systems (Gkerekos et al. 2019).

The proposed M5 Rules model based on NRs should be updated daily or weekly in alignment with the submission of these reports. However, the frequency of updates may need to increase based on operational changes, such as variations in routing or weather conditions, as well as any modifications to the ship's equipment. Continuous monitoring of the model's predictive performance against actual FC will help identify when recalibration is necessary. Additionally, the incorporation of supplementary data sources, such as weather forecasts and sea conditions, may also warrant more frequent updates to maintain optimal accuracy in predictions.

Incorporating interactions in a more detailed way between parameters, such as sea state and speed, can further enhance the accuracy of FC predictions. Building models addressing these interactions often reveals complex, non-linear relationships that individual parameters may overlook, allowing models to include the nuances of real-world conditions. By including interaction terms in the feature set, models can improve their robustness and generalisation to unseen data. However, this approach requires sufficient training data to avoid overfitting. Overall, modelling these interactions can lead to more reliable and efficient FC predictions in varying operational scenarios(Karaçay et al. 2024).

4. Conclusion

The paper compared ANN, RF, and M5 Rules algorithms for an FC prediction target and determined the highest performer. Then, parametric research was employed to observe the impact of some environmental and operational variables. The selected parameters were SWT, BFScale/Slip, MaxExhT, ScavT, and ScavP since most of these parameters were independent or their variation could be modelled in the dataset. The effect on the FC and CII Ratings of parameter changes was investigated. The key findings derived from this analysis were as follows:

- The M5 Rules algorithm performed best among the tried methods for this data set and problem.
- BFScale/Slip and ScavP were the most impactful parameters on the variation of FC, CO_2 and attained CII.
- The extreme increases on the BFScale/Slip could increase the FC up to 36.55%.
- The variation of the determined parameters could reduce the CII rating near the border values.
- MaxExhT had the least effect on the FC, CO_2 and attained CII.
- The route optimisation and condition monitoring-based operation of ME were suggested to reduce the impact of possible fluctuations.

The limitations of the study were as follows:

- The annual navigational data had a limited number of rows. The performance of algorithms may vary if they are fed with more data.
- The NR data were recorded by the ship personnel. Even though it is pre-processed, there is still a possibility of human error.
- The parametric study investigation limits were kept in the minimum and maximum values encountered in the data set.
- The modelling of interactive columns, when one is changed, was ensured by using the data solely.

The study contributed to the literature by ensuring an FC prediction model selection process and the investigation of several environmental and engine operational variables' impact on FC and CII

rating This research paper can be beneficial for academicians working on a similar subject, industry partners interested in fuel economy or planning for ships, and other maritime authorities related to the regulatory side.

Future research should consider expanding the analysis to encompass multi-variable variations, investigating the interactions among multiple parameters simultaneously to achieve a more nuanced understanding of the complex dynamics within the system. This comprehensive analysis may uncover interdependencies overlooked by single-variable approaches, yielding deeper insights into vessel performance and operational efficiencies. Moreover, future studies could involve evaluations based on sensor data alongside the hybrid application of the proposed methodology, further enhancing informed decision-making and optimising outcomes for shipping companies.

Disclosure statement

No potential conflict of interest was reported by the author(s).

ORCID

Onur Yuksel http://orcid.org/0000-0002-5728-5866
Murat Bayraktar http://orcid.org/0000-0001-7252-4776
Olgun Konur http://orcid.org/0000-0002-2677-6048

References

Aftab M, Tanvir A, Shahid A, Haider BS, Irfan M. 2025. Hyper-parameter tuning through innovative designing to avoid over-fitting in machine learning modelling: a case study of small data sets. J Stat Comput Simul. 95(7):1595–1609. doi:10.1080/00949655.2025.2465787.

Agand P, Kennedy A, Harris T, Bae C, Chen M, Park EJ. 2023. Fuel consumption prediction for a passenger ferry using machine learning and in-service data: a comparative study. Ocean Eng. 284:115271–115271. doi:10.1016/j.oceaneng.2023.115271.

Akyuz E, Cicek K, Celik M. 2019. A comparative research of machine learning impact to future of maritime transportation. Procedia Comput Sci. 158:275–280. doi:10.1016/j.procs.2019.09.052.

Alexiou K, Pariotis EG, Zannis TC, Leligou HC. 2021. Prediction of a ship's operational parameters using artificial intelligence techniques. J Mar Sci Eng. 9(6):681. doi:10.3390/jmse9060681.

Bal Beşikçi E, Arslan O, Turan O, Ölçer AI. 2016. An artificial neural network based decision support system for energy efficient ship operations. Comput Oper Res. 66:393–401. doi:10.1016/j.cor.2015.04.004.

Baştürk S, Erol S. 2023. Optimizing ship speed depending on cargo and wind-sea conditions for sustainable blue growth and climate change mitigation. J Mar Sci Technol. 28(3):659–674. doi:10.1007/s00773-023-00947-4.

Baumann S, Klingauf U. 2020. Modeling of aircraft fuel consumption using machine learning algorithms. CEAS Aeronaut J. 11(1):277–287. doi:10.1007/s13272-019-00422-0.

Bayraktar M, Sokukcu M. 2024. Marine vessel energy efficiency performance prediction based on daily reported noon reports. Ships Offsh Struct. 19(6):831–840. doi:10.1080/17445302.2023.2214490.

Bayraktar M, Yuksel O. 2023. A scenario-based assessment of the energy efficiency existing ship index (EEXI) and carbon intensity indicator (CII) regulations. Ocean Eng. 278:114295. doi:10.1016/j.oceaneng.2023.114295.

Bayraktar M, Yuksel O, Pamik M. 2023. An evaluation of methanol engine utilization regarding economic and upcoming regulatory requirements for a container ship. Sustainable Prod Consumption. 39:345–356. doi:10.1016/j.spc.2023.05.029.

Bialystocki N, Konovessis D. 2016. On the estimation of ship's fuel consumption and speed curve: A statistical approach. J Ocean Eng Sci. 1(2):157–166. doi:10.1016/j.joes.2016.02.001.

Biçen S, Celik M. 2024. A deep learning approach to analyse ship inspection reports via natural language processing integrated with artificial neural network. J Marine Eng Technol. 23(4):291–304. doi:10.1080/20464177.2024.2353407.

Bilgili L. 2023. Determination of the weights of external conditions for ship resistance. Ocean Eng. 276(February):114141–114141. doi:10.1016/j.oceaneng.2023.114141.

Breiman L. 2001. Random forests. Mach Learn. 45(1):5–32. doi:10.1023/A:1010933404324.

Chai T, Draxler RR. 2014. Root mean square error (RMSE) or mean absolute error (MAE)? -arguments against avoiding RMSE in the literature. Geosci Model Dev. 7(3):1247–1250. doi:10.5194/gmd-7-1247-2014.

ClassNK. 2021. Outlines of EEXI regulation; [accessed 2024 May 24]. https://www.classnk.or.jp/hp/pdf/activities/statutory/eexi/eexi_rev3e.pdf.

Cybenko G. 1989. Approximation by superpositions of a sigmoidal function. Math Control Signals Syst. 2(4):303–314. doi:10.1007/BF02551274.

dos Santos Ferreira R, Padilha de Lima JV, Caprace J-D. 2022. Comparative analysis of machine learning prediction models of container ships propulsion power. Ocean Eng. 255:111439. doi:10.1016/j.oceaneng.2022.111439.

Drazdauskas M, Lebedevas S. 2024. Optimization of combustion cycle energy efficiency and exhaust Gas emissions of marine dual-fuel engine by intensifying ammonia injection. J Mar Sci Eng. 12(2):309. doi:10.3390/jmse12020309.

Eide MS, Longva T, Hoffmann P, Endresen Ø, Dalsøren SB. 2011. Future cost scenarios for reduction of ship CO2 emissions. Marit Policy Manage. 38(1):11–37. doi:10.1080/03088839.2010.533711.

El Naqa I, Murphy MJ. 2015. What Is Machine Learning?. In: Machine Learning in Radiation Oncology. Cham: Springer; p. 3–11. doi:10.1007/978-3-319-18305-3_1.

Ertogan M, Wilson PA. 2024. Modelling using neural networks and dynamic position control for unmanned underwater vehicles. J ETA Marit Sci. 12(1):64–73. doi:10.4274/jems.2024.46514.

Fan A, Wang Y, Yang L, Tu X, Yang J, Shu Y. 2024. Comprehensive evaluation of machine learning models for predicting ship energy consumption based on onboard sensor data. Ocean Coast Manag. 248:106946–106946. doi:10.1016/j.ocecoaman.2023.106946.

Fang Q, Nguyen H, Bui X-N, Nguyen-Thoi T. 2020. Prediction of blast-induced ground vibration in open-pit mines using a new technique based on imperialist competitive algorithm and M5 rules. Nat Resour Res. 29(2):791–806. doi:10.1007/s11053-019-09577-3.

Fawcett A. 2021. Data science in 5 minutes: What is one hot encoding?; [accessed 2024 Aug 21]. https://www.educative.io/blog/one-hot-encoding.

Frank E, Hall MA, Witten IH. 2016. Data mining: practical machine learning tools and techniques, 4th ed. Burlington: Morgan Kaufmann.

Gkerekos C, Lazakis I, Theotokatos G. 2019. Machine learning models for predicting ship main engine fuel Oil consumption: a comparative study. Ocean Eng. 188:106282. doi:10.1016/j.oceaneng.2019.106282.

Guo B, Liang Q, Tvete HA, Brinks H, Vanem E. 2022. Combined machine learning and physics-based models for estimating fuel consumption of cargo ships. Ocean Eng. 255(March):111435–111435. doi:10.1016/j.oceaneng.2022.111435.

Gupta P, Rasheed A, Steen S. 2022. Ship performance monitoring using machine-learning. Ocean Eng. 254:111094. doi:10.1016/j.oceaneng.2022.111094.

Hajli K, Rönnqvist M, Dadouchi C, Audy J-F, Cordeau J-F, Warya G, Ngo T. 2024. A fuel consumption prediction model for ships based on historical voyages and meteorological data. J Marine Eng Technol. 23(6):439–450.

Handayani MP, Kim H, Lee S, Lee J. 2023. Navigating energy efficiency: a multifaceted interpretability of fuel oil consumption prediction in cargo container vessel considering the operational and environmental factors. J Mar Sci Eng. 11(11):2165. doi:10.3390/jmse11112165.

Hu Z, Zhou T, Osman MT, Li X, Jin Y, Zhen R. 2021. A novel hybrid fuel consumption prediction model for ocean-going container ships based on sensor data. J Mar Sci Eng. 9(4):449–449. doi:10.3390/jmse9040449.

Hwang I-K, Lee M, Han J, Choi J. 2023. Wave height measurement scheme using wave detector based on convolutional neural network and PPM calculator with ocean wave images. Int J Nav Archit Ocean Eng. 15:100542. doi:10.1016/j.ijnaoe.2023.100542.

IMO. 2021. 2021 guidelines on the operational carbon intesnity rating of ships (CII Rating Guidelines, G4); [accessed 2024 May 25]. https://wwwcdn.imo.org/localresources/en/OurWork/Environment/Documents/Air%20pollution/MEPC.339(76).pdf.

IMO. 2022. Improving the energy efficiency of ships; [accessed 2024 Aug 5]. https://www.imo.org/en/OurWork/Environment/Pages/Improving%20the%20energy%20efficiency%20of%20ships.aspx.

Jothiprakash V, Kote AS. 2011. Effect of pruning and smoothing while using M5 model tree technique for reservoir inflow prediction. J Hydrol Eng. 16(7):563–574. doi:10.1061/(ASCE)HE.1943-5584.0000342.

Karaçay ÖE, Karatuğ Ç, Uyanık T, Arslanoğlu Y, Lashab A. 2024. Prediction of ship main particulars for harbor tugboats using a Bayesian network model and non-linear regression. Appl Sci. 14(7):2891. doi:10.3390/app14072891.

Karatuğ Ç, Arslanoğlu Y, Guedes Soares C. 2023. Design of a decision support system to achieve condition-based maintenance in ship machinery systems. Ocean Eng. 281:114611. doi:10.1016/j.oceaneng.2023.114611.

Karatuğ Ç, Tadros M, Ventura M, Soares CG. 2024. Decision support system for ship energy efficiency management based on an optimization model. Energy. 292:130318. doi:10.1016/j.energy.2024.130318.

Kasuya E. 2019. On the use of r and r squared in correlation and regression. Ecol Res. 34(1):235–236. doi:10.1111/1440-1703.1011.
Kim YR, Jung M, Park JB. 2021. Development of a fuel consumption prediction model based on machine learning using ship in-service data. J Mar Sci Eng. 9(2):1–25.
Kurucan M, Özbaltan M, Yetgin Z, Alkaya A. 2024. Applications of artificial neural network based battery management systems: a literature review. Renewable Sustainable Energy Rev. 192:114262–114262. doi:10.1016/j.rser.2023.114262.
Le LT, Lee G, Park KS, Kim H. 2020. Neural network-based fuel consumption estimation for container ships in Korea. Marit Policy Manage. 47(5):615–632. doi:10.1080/03088839.2020.1729437.
Le TT, Sharma P, Pham NDK, Le DTN, Le VV, Osman SM, Rowinski L, Tran VD. 2024. Development of comprehensive models for precise prognostics of ship fuel consumption. J Marine Eng Technol. 23(6):451–465.
Lindstad E, Eskeland GS, Rialland A, Valland A. 2020. Decarbonizing maritime transport: the importance of engine technology and regulations for LNG to serve as a transition fuel. Sustainability. 12(21):8793–8793. doi:10.3390/su12218793.
MetMatters. 2025. The beaufort wind scale; [accessed 2025 Apr 3]. https://www.rmets.org/metmatters/beaufort-wind-scale.
Najafi A, Nowruzi H, Ghassemi H. 2018. Performance prediction of hydrofoil-supported catamarans using experiment and ANNs. Appl Ocean Res. 75:66–84. doi:10.1016/j.apor.2018.02.017.
Ngah S, Bakar RA, Embong A, Razali S. 2016. Two-steps implementation of sigmoid function for artificial neural network in field programmable gate array. ARPN J Eng Appl Sci. 7:4882–4888.
Nguyen S, Fu X, Ogawa D, Zheng Q. 2023. An application-oriented testing regime and multi-ship predictive modeling for vessel fuel consumption prediction. Transp Res Part E Logist Transp Rev. 177:103261. doi:10.1016/j.tre.2023.103261.
Nguyen VN, Chung N, Balaji GN, Rudzki K, Hoang AT. 2025. Internet of things-driven approach integrated with explainable machine learning models for ship fuel consumption prediction. Alexandria Eng J. 118:664–680. doi:10.1016/j.aej.2025.01.067.
Oladapo Adeboye P, Michael Oladipo A, Godwin N, Excel GC, Chukwuekem David O. 2024. The impact of automation on maritime workforce management: A conceptual framework. Int J Manage Entrepreneurship Res. 6(5):1467–1488. doi:10.51594/ijmer.v6i5.1095.
Padhma M. 2021. End-to-end introduction to evaluating regression models; [accessed 2024 Aug 22]. https://www.analyticsvidhya.com/blog/2021/10/evaluation-metric-for-regression-models/.
Pal M, Deswal S. 2009. M5 model tree based modelling of reference evapotranspiration. Hydrol. Process. 23(10):1437–1443.
Panapakidis IP, Sourtzi VM, Dagoumas AS. 2020. Forecasting the fuel consumption of passenger ships with a combination of shallow and deep learning. Electronics (Switzerland). 9(5):776–801.
Panda JP. 2023. Machine learning for naval architecture, ocean and marine engineering. J Mar Sci Technol. 28(1):1–26. doi:10.1007/s00773-022-00914-5.
Pariotis EG, Zannis TC, Katsanis JS. 2019. An integrated approach for the assessment of central cooling retrofit using variable speed drive pump in marine applications. J Mar Sci Eng. 7(8):253. doi:10.3390/jmse7080253.
Park M-H, Hur J-J, Lee W-J. 2023. Prediction of diesel generator performance and emissions using minimal sensor data and analysis of advanced machine learning techniques. J Ocean Eng Sci. 10(1):150–168.
Parkes AI, Savasta TD, Sobey AJ, Hudson DA. 2022. Power prediction for a vessel without recorded data using data fusion from a fleet of vessels. Expert Syst Appl. 187:115971. doi:10.1016/j.eswa.2021.115971.
Piao S, Park M-H, Yeo S, Chun KW, Jee J-H, Lee W-J. 2025. Expanding the range of ship fuel consumption prediction: a multi-algorithm feature selection approach. Ocean Eng. 316:119944. doi:10.1016/j.oceaneng.2024.119944.
Prpić-Oršić J, Faltinsen OM. 2012. Estimation of ship speed loss and associated CO2 emissions in a seaway. Ocean Eng. 44:1–10. doi:10.1016/j.oceaneng.2012.01.028.
Quinlan JR. 1992. Learning with continuous classes. 5th Australian joint conference on artificial intelligence; World Scientific.
Rácz A, Bajusz D, Héberger K. 2021. Effect of dataset size and train/test split ratios in QSAR/QSPR multiclass classification. Molecules. 26(4):1111. doi:10.3390/molecules26041111.
Rajoub B. 2020. Chapter 3 – Supervised and unsupervised learning. In: Zgallai W, editor. Biomedical signal processing and artificial intelligence in healthcare. London: Academic Press; p. 51–89.
Rather IH, Kumar S, Gandomi AH. 2024. Breaking the data barrier: a review of deep learning techniques for democratizing AI with small datasets. Artif Intell Rev. 57(9):226. doi:10.1007/s10462-024-10859-3.
Rodriguez-Galiano V, Sanchez-Castillo M, Chica-Olmo M, Chica-Rivas M. 2015. Machine learning predictive models for mineral prospectivity: an evaluation of neural networks, random forest, regression trees and support vector machines. Ore Geol Rev. 71:804–818. doi:10.1016/j.oregeorev.2015.01.001.
Safaei AA, Ghassemi H, Ghiasi M. 2018. Correcting and enriching vessel's noon report data using statistical and data mining methods. Eur Transport. 67(1).
Senol YE, Seyhan A. 2024. A novel machine-learning based prediction model for ship manoeuvring emissions by using bridge simulator. Ocean Eng. 291:116411. doi:10.1016/j.oceaneng.2023.116411.
Seraj A, Mohammadi-Khanaposhtani M, Daneshfar R, Naseri M, Esmaeili M, Baghban A, Habibzadeh S, Eslamian S. 2023. Chapter 5 – Cross-validation. In: Eslamian S, Eslamian F, editors. Handbook of hydroinformatics. Amsterdam: Elsevier; p. 89–105.
Seyfi H, Hitchmough D, Armin M, Blanco-Davis E. 2023. A study on the viability of fuel cells as an alternative to diesel fuel generators on ships. J Mar Sci Eng. 11(8):1489. doi:10.3390/jmse11081489.
Su M, Jasmine LH, Xueqin W, Bae S-H. 2025. Fuel consumption cost prediction model for ro-ro carriers: a machine learning-based application. Maritime Policy & Management. 52(2):229–249. doi:10.1080/03088839.2024.2303120.
Taghavifar H, Perera LP. 2023a. Data-driven modeling of energy-exergy in marine engines by supervised ANNs based on fuel type and injection angle classification. Process Saf Environ Prot. 172:546–561. doi:10.1016/j.psep.2023.02.034.
Taghavifar H, Perera LP. 2023b. The effect of LNG and diesel fuel emissions of marine engines on GHG-reduction revenue policies under life-cycle costing analysis in shipping.
Taghavifar H, Perera LP. 2023c. Life cycle emission and cost assessment for LNG-retrofitted vessels: the risk and sensitivity analyses under fuel property and load variations. Ocean Eng. 282:114940. doi:10.1016/j.oceaneng.2023.114940.
Taghva HR, Ghassemi H, Nowruzi H. 2018. Seakeeping performance estimation of the container ship under irregular wave condition using artificial neural network. Am J Civil Eng Archit. 6(4):147–153.
Tamaya. 2020. Marine vanes; [accessed 2024 Aug 22]. https://tamaya-technics.com/en/wp-content/uploads/2020/10/pdf_20201005.pdf.
Uyanık T, Karatuğ Ç, Arslanoğlu Y. 2020. Machine learning approach to ship fuel consumption: a case of container vessel. Transp Res Part D Transp Enviro. 84:102389. doi:10.1016/j.trd.2020.102389.
Vilaça A, Aguiar A, Soares C. 2015. Estimating fuel consumption from GPS data. Lect Notes Comput Sci (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics). 9117:672–682.
Vorkapić A, Martinčić-Ipšić S, Piltaver R. 2024. Interpretable machine learning: a case study on predicting fuel consumption in VLGC ship propulsion. J Mar Sci Eng. 12(10):1849. doi:10.3390/jmse12101849.
Wheeler D. 2005. An examination of the accuracy and consistency of ships' logbook weather observations and records. Clim Change. 73(1–2):97–116. doi:10.1007/s10584-005-6950-8.
Willmott CJ, Matsuura K. 2005. Advantages of the mean absolute error (MAE) over the root mean square error (RMSE) in assessing average model performance. Clim Res. 30(1):79–82. doi:10.3354/cr030079.
Yang Z, Radchenko R, Radchenko M, Radchenko A, Kornienko V. 2022. Cooling potential of ship engine intake air cooling and its realization on the route line. Sustainability. 14(22):15058. doi:10.3390/su142215058.
Yu H, Gao J, Zhang P, Han FJ, Yang Q, Cui B. 2024. The impact of scavenging air state on the combustion and emission performance of marine two-stroke dual-fuel engine. Sci Rep. 14(1):15776. doi:10.1038/s41598-024-66826-z.
Yuksel O. 2023. A comprehensive feasibility analysis of dual-fuel engines and solid oxide fuel cells on a tanker ship considering environmental, economic, and regulation aspects. Sustainable Prod Consumption. 42:106–124. doi:10.1016/j.spc.2023.09.012.
Yuksel O, Bayraktar M, Sokukcu M. 2023. Comparative study of machine learning techniques to predict fuel consumption of a marine diesel engine. Ocean Eng. 286:115505. doi:10.1016/j.oceaneng.2023.115505.
Yüksel O, Köseoğlu B. 2020. Modelling and performance prediction of a centrifugal cargo pump on a chemical tanker. J Marine Eng Technol. 19(4):278–290. doi:10.1080/20464177.2019.1665330.
Yüksel O, Köseoğlu B. 2022. Regression modelling estimation of marine diesel generator fuel consumption and emissions. Trans Marit Sci. 11(1):79–94. doi:10.7225/toms.v11.n01.w08.
Zhang C, Lu T, Wang Z, Zeng X. 2023. Research on carbon intensity prediction method for ships based on sensors and meteorological data. J Mar Sci Eng. 11(12):2249. doi:10.3390/jmse11122249.
Zhang M, Tsoulakos N, Kujala P, Hirdaris S. 2024. A deep learning method for the prediction of ship fuel consumption in real operational conditions. Eng Appl Artif Intell. 130:107425–107425. doi:10.1016/j.engappai.2023.107425.
Zhou Y, Pazouki K, Norman R. 2025. A grey-box deep learning modelling strategy for fuel oil consumption prediction: a case study of tuna purse seiner. Ocean Eng. 324:120733. doi:10.1016/j.oceaneng.2025.120733.
Zwart RH, Bogaard J, Kana AA. 2023. A grey-box model approach using noon report data for trim optimization. Int Shipbuilding Prog. 70:41–63. doi:10.3233/ISP-220009.