journal homepage: www.brodogradnja.fsb.hr

Brodogradnja

An International Journal of Naval Architecture and Ocean Engineering for Research and Development

Explainable machine learning-based prediction of fuel consumption in ship main engines using operational data

Anh Tuan Hoang[1,2], *Thi Anh Em Bui*[3], *Xuan Phuong Nguyen*[4], *Van Hung Bui*[5], *Quang Chien Nguyen*[6,7], *Thanh Hai Truong*[4], *Nghia Chung*[8,*]

[1] *Faculty of Engineering, Dong Nai Technology University, Bien Hoa City, Vietnam*
[2] *Graduate School of Energy and Environment, Korea University, 145 Anam-ro, Seongbuk-gu, Seoul, 02841, South Korea*
[3] *Institute of Engineering, HUTECH University, Ho Chi Minh City, Vietnam*
[4] *PATET Research Group, Ho Chi Minh City University of Transport, Ho Chi Minh City, Vietnam*
[5] *University of Technology and Education, The University of Danang, Danang, Vietnam*
[6] *Institute of Research and Development, Duy Tan University, Da Nang, Vietnam*
[7] *School of Engineering & Technology, Duy Tan University, Da Nang, Vietnam*
[8] *Institute of Maritime, Ho Chi Minh City University of Transport, Ho Chi Minh City, Vietnam*

ARTICLE INFO

Keywords:
Ship fuel consumption
Prediction model
Local interpretable model-agnostic explanations
Shapley additive explanations
Explainable artificial intelligence

ABSTRACT

A significant percentage of fuel consumption and emissions from transportation activities is related to maritime transportation. Hence, accurate prediction models for fuel consumption are quite important. Machine learning offers a data-driven approach to improving fuel consumption prediction, thereby promoting environmental sustainability, lowering operational costs, and enhancing financial viability. This work explores several machine learning approaches by employing statistical measures, including mean squared error (MSE), coefficient of determination (R^2), and Kling-Gupta efficiency (KGE), to develop main engine fuel consumption (MEFC) prediction models. Hyperparameter optimization via grid search was conducted to improve the generalizability and robustness of the models. With the lowest test MSE (0.69), a robust testing R^2 (0.9867), and a high KGE (0.9681), the Random Forests proved to be the most appropriate model for MEFC modeling among all others. Extreme Gradient Boosting followed closely with competitive accuracy, with MSE values of 0.75 and a robust testing R^2 (0.9856). Using Shapley additive explanations and Local interpretable model-agnostic explanations, this study improves model interpretability even more and indicates that main engine speed and wind speed were revealed to be the most important factors controlling MEFC. Explainable artificial intelligence techniques offer transparency in decision-making, thereby helping marine operators maximize fuel economy. Employing reliable and interpretable predictive modeling, this study offers insightful information for sustainable shipping, hence lowering operating costs and emissions.

1. Introduction

As the backbone of international trade, maritime transport accounts for over 80% of global trade volume

* Corresponding author.
E-mail address: chungnghia@ut.edu.vn

https://doi.org/10.21278/brod76405
Received 15 March 2025; Received in revised form 6 June 2025; Accepted 1 July 2025
Available online 1 August 2025
ISSN 0007-215X; eISSN 1845-5859

and plays a crucial part in global commerce and economic growth [1, 2]. However, with the expansion of shipping and maritime activities, annual greenhouse gas emissions (GHG) from vessels have surpassed one billion tons [3, 4]. Therefore, the International Maritime Organization (IMO) has implemented various regulatory measures to mitigate emissions, including the Energy Efficiency Design Index (EEDI), the Energy Efficiency Operational Indicator (EEOI), and the Carbon Intensity Indicator (CII), all aimed at improving ship energy efficiency and decreasing GHGs [5-7]. Since 1997, the IMO has enacted progressive policies to address maritime emissions, culminating in the adoption of the Initial GHG Strategy in 2018, and subsequently, the 2023 Revised GHG Strategy [8, 9]. The updated strategy issued in 2023 outlined ambitious decarbonization targets, seeking to reduce GHG emissions by 30% by 2030 and 80% by 2040, relative to 2008 levels [9, 10].

Efforts to decarbonize maritime activities require a comprehensive strategy that combines technological innovations, operational enhancements, and the use of alternative fuels [11-14]. The IMO has underscored the significance of low-carbon and zero-carbon fuels, including ammonia, hydrogen, and biofuels, in attaining the ambitious objectives outlined in the 2023 GHG Strategy [9, 15]. The transition to alternative fuels necessitates substantial infrastructural advancements, including the retrofitting of current fleets and the establishment of fuel supply lines, presenting economic and logistical obstacles [16, 17]. In addition, innovations in energy efficiency technology such as waste heat recovery [18, 19] and hull optimization [20, 21] have been progressively investigated to improve fuel economy and reduce emissions. Digitalization and artificial intelligence (AI) can enhance these initiatives by facilitating real-time monitoring and predictive maintenance, therefore optimizing voyage planning and reducing fuel usage [22, 23]. Notwithstanding these achievements, considerable obstacles persist in the shift towards a low-carbon maritime industry. The substantial capital expenditures required for converting vessels and implementing modern propulsion systems pose financial obstacles, especially for small and medium-sized maritime enterprises [24, 25]. Additionally, uncertainty in international regulatory frameworks and the absence of a globally unified carbon pricing mechanism deter investment and innovation in sustainable technologies. Addressing these issues requires coordinated action among stakeholders like governments, shipowners, operators, and research institutions to develop supportive legislation and incentive frameworks that encourage low-emission technology adoption [26, 27]. More importantly, regional regulatory regimes, including the European Union's Fit for 55 packages and the incorporation of maritime emissions into the Europe Emissions Trading System, impose supplementary compliance obligations on ship operators [28, 29]. Although these methods expedite the implementation of sustainable practices, they also pose a risk of generating market discrepancies among various shipping corridors. The implementation of carbon intensity regulations has compelled industry stakeholders to investigate fuel-blending techniques and hybrid propulsion systems to achieve compliance standards [30]. Indeed, attaining the IMO's decarbonization objectives necessitates a comprehensive and flexible strategy that harmonizes technological advancement with economic viability. Ongoing legislative enhancements and industry-driven initiatives will be crucial in advancing a resilient and environmentally sustainable maritime sector [31, 32].

During a vessel's operational cycle, fuel consumption, EEOI, EEDI, and CII are directly correlated with its emission levels [33]. To adhere to these requirements and sustain a competitive advantage, shipping companies are proactively pursuing measures to reduce fuel use [34, 35]. Consequently, predictive models for fuel use have become indispensable in the maritime industry. Accurate forecasting of fuel use under diverse ship and climatic conditions is essential for assessing emissions and energy efficiency. Furthermore, accurate forecasts enable informed decision-making concerning route optimization, speed modifications, and other operational methods designed to enhance overall ship efficiency [36, 37]. Despite various modeling efforts, traditional statistical methods often fail to capture the nonlinear relationships and complex interactions between operational parameters and environmental factors influencing main engine fuel consumption (MEFC). Conventional regression methods limit their capacity to consider dynamic marine circumstances since they assume linear dependencies. Moreover, empirical models are less flexible to fit different ship profiles, paths, and weather conditions since they may demand great domain expertise and parameter tuning [38, 39]. These constraints draw attention to a notable literature gap and underline the need for sophisticated data-driven methods able to manage complicated connections and offer strong MEFC projections. Harnessing the capabilities of nonlinear modeling, feature interaction capture, and adaptive learning, machine learning

(ML) algorithms such as Extreme Gradient Boosting (XGBoost), Decision Trees (DT), Extra Trees, and Random Forests (RF) have shown significant potential for improving MEFC prediction models [40-42]. XGBoost, an improved Gradient Boosting method, employs iterative learning and regularization to improve the prediction accuracy. XGBoost is suitable for capturing complex fuel consumption patterns under many operating circumstances [41, 42]. Conversely, DT models provide interpretable decision-making frameworks that effectively segregate the data depending on influential characteristics, therefore enabling organized and rule-based MEFC estimates. Extra tree regression models are observed to be resilient against overfitting. Extra tree regression is an ensemble learning method that helps to lower variance by utilizing randomized feature selection and multiple decision trees, thus helping to improve model generalization [43, 44]. These ML models are particularly effective in processing high-dimensional input features such as vessel speed, shaft power, wind speed, wave height, and ocean currents. Unlike conventional models, they do not rely on predefined equations to capture relationships but learn directly from data patterns [45]. Their ability to incorporate real-time operational inputs allows for adaptive forecasting, supporting dynamic and informed fuel management decisions. Such capabilities position ML-based approaches as superior alternatives to traditional statistical methods in MEFC modeling. Notwithstanding these developments, present models can suffer from interpretability and fail to accurately measure individual feature impacts, so restricting their practical value for end users. For evidence-based decision-making, these gaps in the literature impede the operational transparency ship operators demand. Through Shapley additive explanations (SHAP) and Local interpretable model-agnostic explanations (LIME), the extraction of transparent and actionable insights from intricate ML models helps to overcome these constraints, strengthening trust and enabling more informed fuel management policies.

The primary objective of this study is to develop a robust and interpretable ML framework for MEFC prediction by integrating multiple ML algorithms to improve predictive accuracy and generalizability, thereby addressing key limitations and gaps identified in the current body of literature. The development of an interpretable white-box modeling framework will be used to explore feature importance and underlying decision-making logic, thereby ensuring transparency, reliability, and practical applicability in real-world maritime fuel management scenarios. The study intends to use two different interpretable approaches, SHAP and LIME to attain interpretability and model transparency. Together, SHAP and LIME bridge the critical gap between ML accuracy and model transparency, often cited as a limitation in black-box ML approaches. Their inclusion ensures that the developed predictive models are not only highly accurate but also interpretable and trustworthy, which is essential for practical deployment in maritime operations. The integration of explainable AI tools within the MEFC modeling framework thus empowers maritime decision-makers with transparent, data-driven insights, supporting more effective and accountable fuel management.

2. Materials and methods

2.1 Data collection and background

The data used in this analysis comes from a published research project by Uyanik et al. [46] analyzing a ship's MEFC using noon reports gathered through a 35-day trip. The adoption of this dataset is justified by its empirical basis since it is obtained from actual operational data kept on a commercial vessel. Traditionally utilized for ship performance monitoring, noon reports offer insightful analysis of important operational factors, including fuel usage, speed, weather, and engine performance. This work ensured that the fuel consumption calculations were based on real-world voyage conditions by using this dataset, thereby improving the dependability and resilience of the developed ML models. Furthermore, using actual voyage data helps to reduce the restrictions related to theoretical models that might not adequately explain the complexity of actual marine operations.

2.2 Machine learning algorithms

2.2.1 Random Forest

Random Forest (RF) is a powerful ensemble learning technique that enhances predictive performance by combining multiple DTs. The fundamental concept is to create several separate DTs and combine their results to produce a final aggregated prediction [47]. Using a method known as bootstrapping aggregating, each tree in the forest is trained on a randomly chosen portion of the data, thus improving generalization and lowering overfitting [48, 49]. Furthermore, a random subset of features is selected for the split at every node of a tree, therefore guaranteeing variety among the trees and thereby reducing overfitting. RF uses majority voting - that is, each tree votes for a class, where the class with the most votes ends as the final prediction in classification problems [50, 51]. In regression, a reduction in variance results from averaging the outputs from all trees. This averaging approach guarantees consistency in forecasts and builds the model to be strong against noisy data. The capacity of RF to efficiently manage high-dimensional data is one of its main benefits. Operating on several subsets of features, it remains robust against irrelevant or redundant features. RF has a few shortcomings, notwithstanding its merits. First, especially in cases of a lot with trees, the model can lead to higher computational costs [52, 53]. As the number of trees increases, training and inference times likewise change dramatically. Furthermore, even if it works well on structured data, it might not be as successful on very complicated patterns where XGBoost or Gradient Boosting could produce better outcomes. RF is a dependable tool for many ML applications, including medical diagnosis, fraud detection, and recommendation systems, since it balances bias with variance generally. In predictive modeling, this popular and extensively used method can generalize effectively across several datasets and combined resistance to overfitting.

2.2.2 Gradient Boosting

Gradient Boosting (GB) is a sequential ensemble learning technique that builds models iteratively, with each new tree correcting the errors of the previous ones. GB creates trees consecutively, maximizing a preset loss function at every step, unlike RF, in which trees are built randomly [54]. This method produces a model with decreasing bias and enhanced predictive accuracy that is ever more sophisticated. GB’s basic idea is to fit trees to the residuals (errors) of the prior model, thereby reducing the loss function. Each tree is trained to minimize the error in anticipating the variation between the actual values and the model's current predictions rather than on the raw labels. Through the capture of complex, data-driven relationships, this iterative improvement technique ensures that every new tree enhances the performance of the model [55, 56]. The learning rate of GB is a fundamental factor since it determines the degree of contribution each tree makes to the resultant prediction. Whereas a high learning rate may cause overfitting, a lower learning rate produces slower but more steady learning. While subsampling generates randomization by choosing a part of the data for every tree, hence lowering variance, shrinkage scales down the contribution of every tree [57, 58]. GB has certain disadvantages, even if it boasts great accuracy. Particularly for big datasets, the sequential character makes it computationally costly. Furthermore, being sensitive to hyperparameters, it calls for careful optimization of the learning rate, tree depth, and tree count to reach the best performance. The model can overfit the training data without appropriate tweaking, thereby reducing generalization to unprocessed input. Predictive analytics is a great tool because of its ability to replicate intricate patterns and interactions among features.

2.2.3 XGBoost

XGBoost is an optimized implementation of GB that improves both computational efficiency and predictive performance. Popular for large-scale machine-learning projects and competitions, it combines many improvements over conventional GB [59]. XGBoost improves over GB by including second-order derivatives, therefore producing more accurate updates, while conventional GB reduces the loss function using first-order gradients [60, 61]. Furthermore, included in the model are regularizing terms in the goal function, hence guiding tree complexity and avoiding overfitting. XGBoost uses shrinkage to improve generalization by lowering the impact of every tree's input and therefore facilitating more gradual learning. It also presents column subsampling, in which at each split a random subset of features is chosen to foster diversity and lower

variation. XGBoost's capacity to effectively manage missing values using a sparsity-aware algorithm, which gives ideal default directions for missing data during tree building, adds a significant advantage [62].

XGBoost's parallelized execution, which distributes computations over several CPUs, therefore accelerates training and is a computational advantage. XGBoost uses backward pruning, eliminating pointless splits to increase efficiency, unlike conventional boosting techniques that develop trees in a depth-first way [63, 64]. The weighted quantile sketching method also improves performance on imbalanced data, therefore guaranteeing better treatment of rare classes. XGBoost needs careful hyperparameter adjustment to avoid overfitting, even if it offers benefits. Maximizing performance depends on choosing the regularization parameters, tree depth, and an ideal learning rate. XGBoost is increasingly applied in fields such as fraud detection, consumer analytics, and automated trading systems because of its great efficiency and high accuracy. Among the most potent gradient-boosting systems available in contemporary ML, its mix of speed, adaptability, and strength is quite effective.

2.2.4 Linear Regression

Linear Regression (LR) is a statistical method that models the relationship between a dependent variable and one or more independent variables using a linear function. The model assumes a linear relationship between the independent variables and the corresponding variations in the dependent variable. By using the ordinary least squares function, the regression coefficients are approximated, hence reducing the sum of squared residuals [65, 66]. The error term explains the inexplicable variance. The model cannot be effective without certain important assumptions being satisfied: linearity, homoscedasticity, absence of multicollinearity, and lack of normality of residuals. Although LR suffers from capturing nonlinear relationships, it is basic and interpretable. In the case of high-dimensional datasets, overfitting becomes a problem and calls for Lasso Regression using an L1 penalty to enhance generalization, and ridge regression using an L2 penalty. Due in great part to its computing efficiency and simplicity of interpretation, LR is extensively applied in many disciplines, including economics, social sciences, and engineering, despite its restrictions [67, 68]. Non-linear models such as DT, RF, or neural networks often exceed LR in predicting accuracy, but in situations where interactions between variables are complicated.

2.2.5 Decision Tree

A Decision Tree (DT) is a non-parametric, tree-based ML model used for both classification and regression tasks. It works by recursively partitioning the input space into homogeneous subsets based on feature values, forming a tree-like structure. Each internal node represents a decision rule on a feature, while branches show several results, and leaf nodes have final forecasts. Built using techniques such as classification and Regression Trees, which choose the optimum feature splits depending on parameters such as Gini impurity for classification or mean square error. The tree grows until it reaches a minimum number of samples per leaf or a maximum depth, therefore satisfying a stopping criterion. One of the primary advantages of decision trees is their interpretability. They provide transparent decision paths and can handle both categorical and numerical variables without assuming any underlying data distribution. Moreover, they require minimal preprocessing, making them suitable for a wide range of practical applications [69, 70]. DT has a significant disadvantage in that it overfits, particularly in deep trees that pick up noise in the training data. Pruning removes branches that do not significantly affect prediction accuracy, helping to reduce this overfitting problem. Control overfitting also comes from methods including minimum impurity decrease and maximum depth restriction. Notwithstanding their restrictions, DT constitutes the basis for strong ensemble techniques such as RF and GB, which use several trees to raise predictive performance. Applications including credit scoring, client segmentation, and medical diagnosis all benefit from DT’s efficiency and interpretability [71, 72].

2.2.6 Support Vector Regression

Support Vector Regression (SVR) is an extension of Support Vector Machine for regression tasks. SVR seeks to discover a function that deviates from the actual target values by no more than a defined margin (ε), while also being as smooth as possible, unlike conventional regression models that minimize the error directly. Maintaining a tolerance level for deviations, SVR creates a hyperplane in a high-dimensional feature space

that best fits the data. The aim is to minimize the epsilon-insensitive loss function, which solely penalizes deviations beyond this threshold and ignores errors smaller than ε [73, 74]. SVR is robust enough to work against noise in data and minor fluctuations. Using kernel functions, e.g., linear, polynomial, radial basis function, could allow SVR to model complicated, nonlinear relationships by converting data into a higher-dimensional space where an LR function may be used. The selection of the kernel greatly affects model performance and should be done depending on the distribution of the data. SVR comprises two important hyperparameters: The regularization parameter manages the trade-off between preserving model simplicity and reducing error [75, 76]. Although a greater regularization parameter produces a better fit, overfitting may follow from this. Defines the error tolerance threshold, and epsilon margin. Greater flexibility and lessening of extreme sensitivity to noise are made possible by a higher ε. SVR offers more flexibility and is quite successful for small-to-medium-sized datasets with complicated relationships than conventional LR. But as it solves quadratic optimization issues, its computational cost may be high, especially for big datasets. Because SVR can easily manage both linear and nonlinear regression problems with great accuracy, it is extensively used in traffic prediction, financial forecasting, and biological modeling.

2.2.7 Extra Tree Regressor

The Extra Tree Regressor (ETR) is an ensemble learning method similar to RF, but with a key difference: it introduces additional randomness when constructing DTs. ETR chooses feature splits at random, hence reducing variance and accelerating training durations; RF chooses the optimum feature split by optimizing impurity reduction. Each tree in the ETR ensemble is constructed from the entire training set without bootstrapping, and at each node, a random feature is chosen together with a random threshold value for splitting [77, 78]. While guaranteeing strong generalization, this enhanced randomness reduces the model's overfitting susceptibility. ETR are more efficient than RF since they do not need to look for the ideal split at every node, hence, accelerating training speed is one of their various benefits. Furthermore, they offer improved regularization since the extra randomization lowers variance and improves model stability, especially in noisy data. Moreover, ETR show great resistance against overfitting and perform well on high-dimensional datasets without much hyperparameter adjustment needed [79, 80]. The ETR has a bigger bias than RF, though, since the random splits could produce fewer ideal trees, therefore lowering the general accuracy. It also lacks the interpretability of more basic models such as DTs, which makes it less appropriate for uses where explainability is important [81]. ETR are beneficial for uses including bioinformatics, climate modeling, and real estate price prediction, even with these restrictions, since they show great efficiency in managing huge datasets with complex feature relationships.

2.2.8 Kernel Ridge Regression

Kernel Ridge Regression (KRR) combines Ridge Regression with Kernel methods to model complex, nonlinear relationships. Using kernel techniques could extend LR and enable LR to capture complex patterns without clearly changing input features. With an L2 regularization term, ridge regression reduces the sum of squared residuals, hence punishing high coefficients and avoiding overfitting [82, 83]. By substituting a higher-dimensional feature space created by kernel functions such as polyn, Gaussian, and sigmoid kernels for the input space, KRR improves this even more. This change allows the LR function to work in the transformed domain, hence capturing nonlinear dependencies in the data. Strong generalization is one of the various benefits KRR presents since regularizing helps to reduce overfitting while preserving model flexibility. For nonlinear modeling, it is quite successful since kernel methods enable it to capture complicated interactions without explicit feature engineering. Furthermore, KRR provides an analytical solution, so it is computationally effective for small datasets, unlike iterative gradient-based techniques. KRR has certain restrictions, too, including high computational costs resulting from kernel-based transformations, which can be costly for big datasets and poorly scaled with sample count [84, 85]. It is also memory-intensive since storing kernel matrices calls for a lot of memory, therefore restricting its use to really big datasets. Notwithstanding these difficulties, KRR has been extensively applied in fields including signal processing, geographic modeling, and biological data analysis, where nonlinear interactions abound. Particularly for structured or very complicated data, KRR offers a strong substitute for conventional regression methods by

aggregating the strengths of ridge regression and kernel approaches [86, 87]. A schematic showing a flowchart of the ML implementation is depicted in Figure 1.

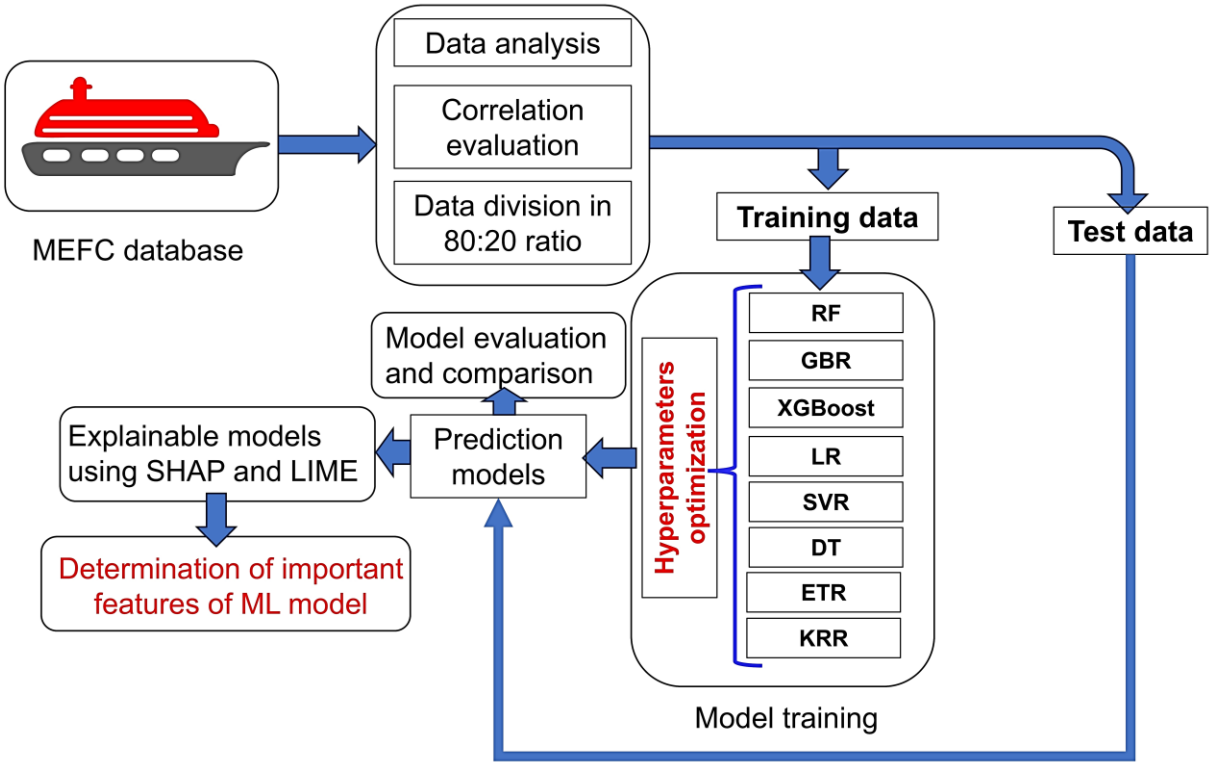


Fig. 1 Schematic flow chart of ML implementation for MEFC

2.3 Explainable machine learning approach

2.3.1 SHapley additive explanations

Robust interpretability method SHapley additive explanations (SHAP) quantifies the contribution of every input feature in ML-based prediction models. It is based on cooperative game theory, which considers all conceivable feature combinations to guarantee a fair distribution of feature importance. SHAP aids in the identification of important factors that greatly affect fuel consumption in MEFC prediction, including vessel speed, weather conditions, and wave height [88]. The system improves ML models by including local and global explanations, therefore helping ship operators maximize fuel economy. Whereas dependence and force charts provide instance-specific information, the SHAP summary plot presents a whole picture of feature contributions [89]. In the maritime industry, in which fuel expenses and pollutants are the main issues, this interpretability is important. Using SHAP allows academics and business experts to create more accurate models and make data-driven decisions to raise operational efficiency and sustainability in shipping [90]. The flowchart for SHAP analysis is given in Figure 2.

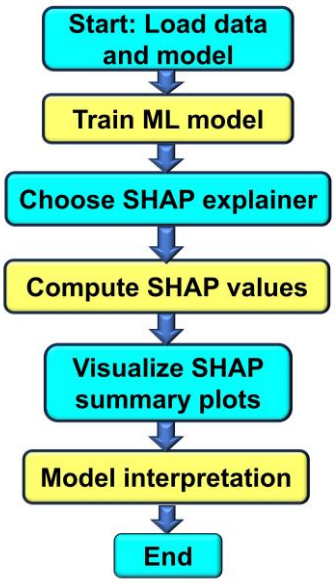


Fig. 2 Flowchart for SHAP analysis

2.3.2 Local interpretable model-agnostic explanations

Local interpretable model-agnostic explanations (LIME) is an interpretability method that explains individual predictions of ML models. Unlike global techniques, LIME highlights the most important elements in every instance by producing local surrogate models around particular predictions. LIME can help to clarify in MEFC forecast how factors like speed, sea state, and wind speed support a specific fuel consumption estimate [91]. LIME generates interpretable linear approximations by varying input values and examining their effects, so guiding marine operators in their understanding of decision-making. Real-time monitoring, anomaly identification, and fuel economy optimization all benefit particularly from this approach. LIME's visual explanations make ML models' outputs more transparent and actionable, thus enhancing trust in them. By allowing shipowners and engineers to adjust operational methods, this improved interpretability helps to save fuel costs and emissions while yet guaranteeing environmental compliance [92, 93]. The schematic flowchart of LIME is depicted in Figure 3.

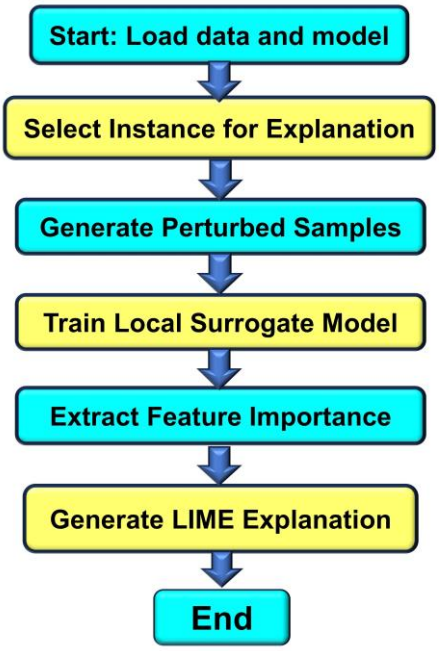


Fig. 3 Schematic flowchart for LIME

2.4 Statistical metrics for model evaluation

In the present work, three statistical measures namely mean squared error (MSE), coefficient of determination (R^2), and Kling-Gupta efficiency (KGE) were employed to assess the predictive performance of the ML models. Each one of these measures provides a different angle of view on the performance traits of the models. The mean of the squared variations between predicted and observed values is measured by MSE. A lower MSE value closer to 0 denotes higher prediction accuracy. In addition, R^2 represents the proportion of variance in the observed data that the model can explain. R^2 values fall between 0 and 1, while higher values indicate a better fit of the model for the observed variance. Moreover, KGE is the composite measure of three fundamental components correlation, bias, and variability. The following expressions were used for the measurement of MSE [94], R^2 [95], and KGE [96]:

$$R^2 = 1 - [\Sigma(y_i - \hat{y}_i)^2 / \Sigma(y_i - \bar{y})^2] \quad (1)$$

$$MSE = (1/n)\Sigma(y_i - \hat{y}_i)^2 \quad (2)$$

$$KGE = 1 - \sqrt{(r-1)^2 + (\alpha - 1)^2 + (\beta - 1)^2} \quad (3)$$

Herein, y_i denotes the actual value, $\hat{y}_i$ represents the predicted value, $\bar{y}$ denotes the mean of actual values, n is the number of observations, r is the Pearson correlation coefficient, $\alpha = \sigma_{\hat{y}}/\sigma_y$ (ratio of standard deviations), $\beta = \mu_{\hat{y}}/\mu_y$ (ratio of means).

3. Results and discussion

3.1 Dendrogram analysis with correlation heatmap

The graphic shown in Figure 4 displays a correlation heatmap integrated with a hierarchical clustering dendrogram to examine the links among several metrics associated with a ship's performance and environmental circumstances. The heatmap illustrates the Pearson correlation coefficients among main engine speed (MES, rpm), shaft power (SP, kW), wind speed (WS, kts), wave height (WH, m), current speed (CS, kts), and main engine fuel consumption (MEFC, tons) since these parameters were found to affect the MEFC the most [97]. Notable findings from the correlation values reveal a robust positive correlation of 0.88 between MES and SP, which is rational, as an increase in engine speed typically leads to elevated shaft power. Likewise, MES and MEFC exhibit a correlation of 0.89, signifying that increased engine speeds result in elevated fuel consumption. SP and MEFC exhibit a robust correlation of 0.79 since heightened power production necessitates greater fuel consumption. Indeed, WS exhibits a moderate correlation with MES (0.76) and MEFC (0.68), indicating that increased wind intensity affects ship resistance and fuel consumption. Conversely, WH and CS demonstrate weaker correlations with other variables, suggesting that these environmental elements may not significantly influence engine power and fuel consumption compared to speed-related measures.

The dendrogram on the left categorizes analogous parameters according to their correlation values. MES, MEFC, and SP constitute a closely associated cluster, indicating their direct correlation with engine performance and fuel efficiency. WH and CS clusters operate independently, as their impact on engine performance is diminished. WS is situated between these clusters, influencing ship resistance, however, it is not as intimately associated with engine operations as MES or SP. Practically, the clustering suggests that enhancing engine speed and power can augment fuel efficiency, whilst environmental variables such as WH and CS exert a minimal effect on direct fuel use, although they may still affect overall energy efficiency. Operational parameters, including velocity, energy consumption, and fuel efficiency, must be optimized while accounting for external environmental factors. The findings are corroborated by the pairwise correlation matrix in Figure 5.

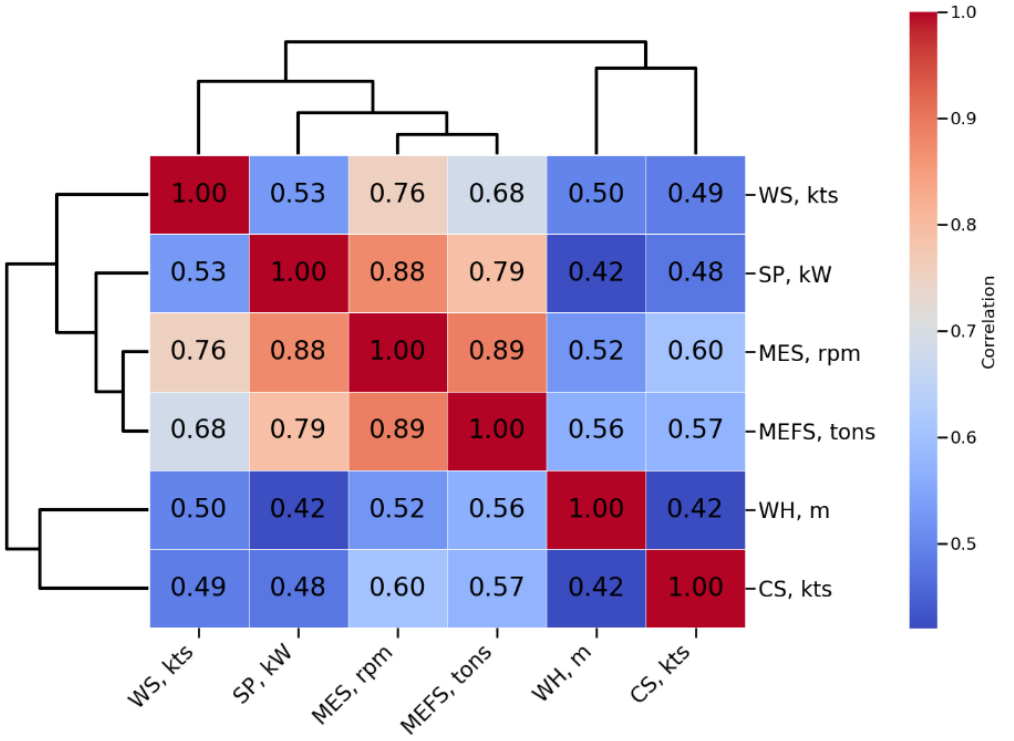


Fig. 4 Dendrogram with correlation heatmap

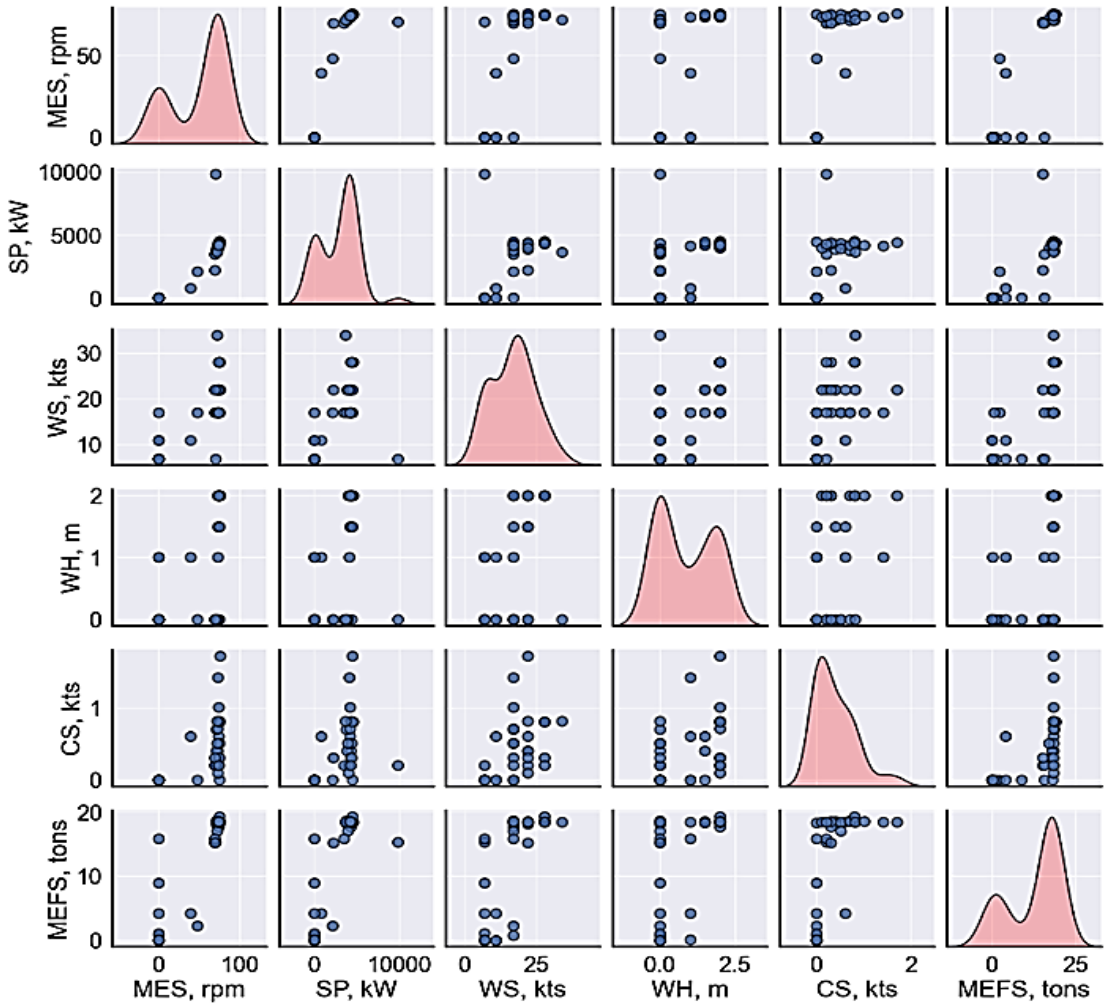


Fig. 5 MEFC data's pairwise correlation matrix

3.2 Descriptive statistics

The dataset comprises 35 observations across six variables about ship operations and environmental conditions. The descriptive statistics are listed in Table 1. The primary engine speed averages 50.59 rpm, with a standard deviation of 33.24, signifying considerable fluctuation. The speed has a median of 71.83 rpm, and the maximum rpm value is 75.32 rpm, signifying that the engine frequently operates close to its upper threshold. The mean shaft power is 2919.42 kW, accompanied by a substantial standard deviation of 2248.65 kW, indicating considerable variability. The power varies from 0 to 9824.72 kW, with a median of 3926.73 kW, indicating that power usage frequently exceeds the average. The average WS is 16.99 knots, with a standard deviation of 7.44, and it ranges from 6.9 to 33.9 knots. The median of 16.9 knots indicates that the majority of data cluster around the mean, implying a generally normal distribution. The average wave height is 0.87 meters, with a standard deviation of 0.91, ranging from 0 meters in calm waters to a maximum of 2 meters. A median of 1 meter signifies that the majority of waves are rather modest. The CS has a mean of 0.39 knots, a standard deviation of 0.43 knots, and a maximum of 1.71 knots, suggesting predominantly weak currents. The fuel consumption of the main engine exhibits considerable variability, averaging 12.89 tons with a standard deviation of 7.71. The lowest of 0 tons implies non-operational intervals, whereas the median of 18.1 tons signifies elevated consumption during active phases.

Table 1 Descriptive statistics

Column	Count	Mean	std.	Min	25%	50%	75%	Max
MES, rpm	35	50.59	33.24	0	0	71.83	74.045	75.32
SP, kW	35	2919.42	2248.65	0	0	3926.73	4296.17	9824.72
WS, kts	35	16.99	7.44	6.9	10.9	16.9	21.9	33.9
WH, m	35	0.87	0.91	0	0	1	2	2
CS, kts	35	0.39	0.43	0	0	0.30	0.65	1.71
MEFC, tons	35	12.89	7.71	0	4.22	18.1	18.46	19.25

3.3 Prediction model development

Python-based code was written to employ the predictive modeling ML framework with multiple approaches, including ensemble methods, tree-based methods, and neural approaches. It utilized fundamental Python modules, including pandas for data manipulation, numpy for numerical analysis, and scikit-learn for ML functionalities. XGBoost performed gradient-boosted regression, while matplotlib facilitated visualization. To ensure a uniform model assessment, the dataset was imported via pandas, randomized, and subsequently divided into training and testing subsets in a ratio of 8:2. Several regression models were trained and tested, including XGBoost, RF, GB Regression, DTs, SVR, and KRR. The evaluation of model performance was conducted utilizing MSE, R^2, and KGE. The accuracy of predictions was enhanced through model training and hyperparameter optimization employing a grid search approach. The statistical evaluation of the models is listed in Table 2, while the range of hyperparameters used and their optimized values are shown in Table 3.

Table 2 Statistical evaluation of developed models for MEFC

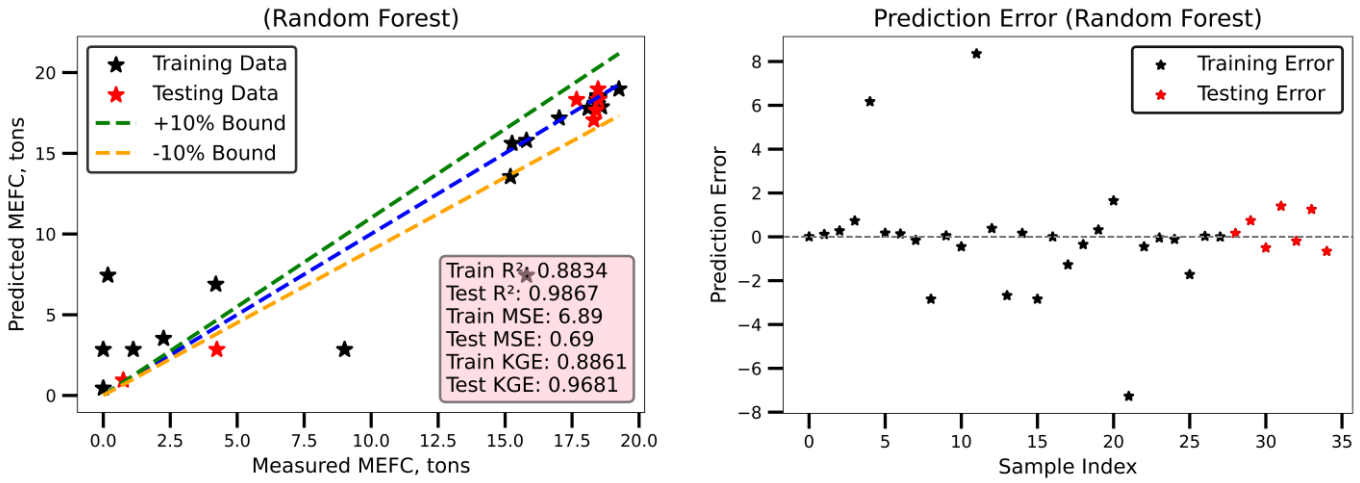
Metrics	**MSE**		R^2		**KGE**	
Model	Train	Test	Train	Test	Train	Test
RF	6.89	0.69	0.8834	0.9867	0.8861	0.9681
GB	6.39	0.76	0.8918	0.9854	0.9213	0.9218
XGBoost	6.39	0.75	0.8918	0.9856	0.9211	0.9101
LR	13.33	2.63	0.7741	0.9491	0.8303	0.9628
DT	6.39	1.05	0.8918	0.9796	0.9213	0.9569
SVR	16.77	3.64	0.7162	0.9295	0.8456	0.8695
ETR	6.39	0.75	0.8918	0.9854	0.9213	0.9409
KRR	59.17	238.46	-0.0015	-3.6126	0.2509	-1.006

Table 3 Range of training hyperparameters and their optimized values

Model	**Hyperparameters range**	**Selected values**
RF	Number of estimators: range (50 to 200) and Maximum depth: range (5 to 20)	100 and 10
GBR	Number of estimators: range (50 to 200), learning rate (0.01 to 0.2), and Maximum depth: range (5 to 20)	100, 0.1 and 5
LR	Fit of intercept: (True or False)	True
DT	Maximum depth: range (3 to 10)	10
SVR	C: (0.1, 1, and 10); epsilon: (0.01, 0.1, to 0.2); and kernel: linear or rbf	1, 0.1, and linear
ETR	Number of estimators: range (50 to 200) and Maximum depth: range (5 to 20)	100 and 10
KRR	alpha: (0.1, 1, 10) and kernel: linear or rbf	1 and rbf
XGBoost	Number of estimators: range (50 to 200), learning rate (0.01 to 0.2), and Maximum depth: range (3 to 10)	100, 0.1, and 5

3.3.1 Random Forest regressor

The RF model demonstrated robust predictive efficacy as shown in Figure 6a, with a training MSE of 6.89 and a lower test MSE of 0.69. This implied that the model generalized effectively without overfitting. Statistical measures of the RF model are listed in Table 2. The R^2 values were on the higher side for both training (0.8834) and testing (0.9867), indicating that the model efficiently elucidates the variance in MEFC. The KGE values of 0.8861 for training and 0.9681 for testing corroborated the model's reliability in identifying patterns within the dataset. Good results can be obtained since the RF has ensemble characteristics; the RF leverages several DTs by collaborating to reduce variation and enhance prediction stability. The low error distribution, as shown in Figure 6b, further corroborates these findings. As an ensemble approach, it necessitated considerable computational resources and hyperparameter manipulation, including the number of estimators and tree depth, to enhance performance. This model proved especially effective in managing non-linearity, rendering it an appropriate option for forecasting MEFC under diverse scenarios.

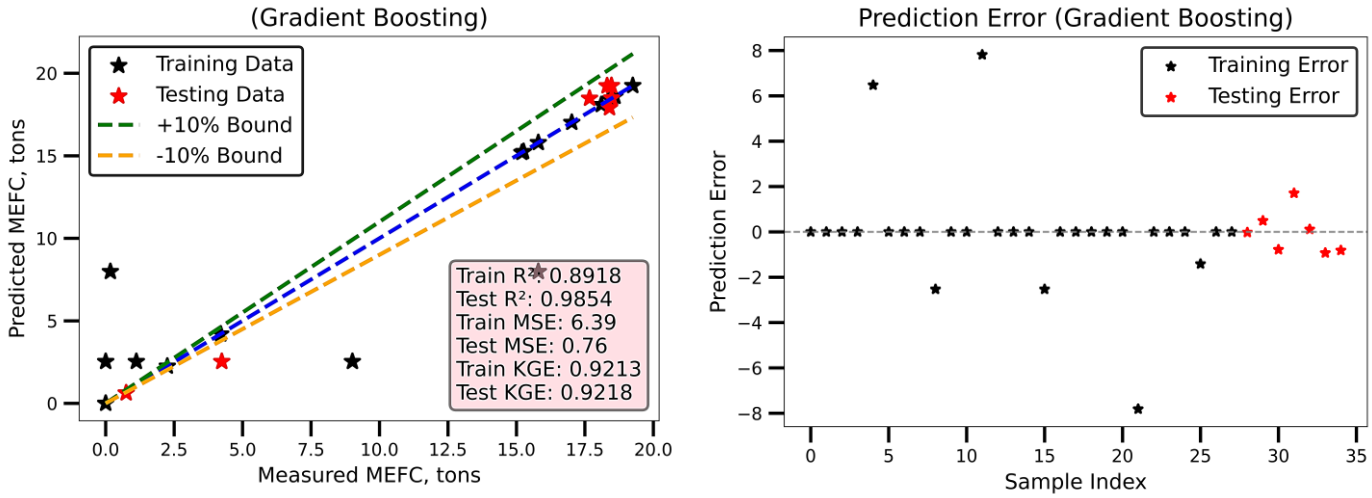


(a) (b)

Fig. 6 RF-based MEFC model: (a) Measured vs. predicted (b) Error distribution

3.3.2 Gradient Boosting regressor

Gradient Boosting (GB) regressor, as illustrated in Figure 7a, demonstrated robust performance with a training MSE of 6.39 and a notably low test MSE of 0.76. Statistical measures of the GB regression model are listed in Table 2. The substantial decrease in error throughout testing demonstrated that the model effectively adapted to novel data, likely benefiting from its iterative learning process that rectified errors from prior trees. The R^2 values of 0.8918 for training and 0.9854 for testing underscored its capacity to elucidate the majority of the variance in fuel usage, indicating exceptional generalization. The KGE values of 0.9213 for training and 0.9218 for testing further corroborated its predictive reliability.



(a) (b)

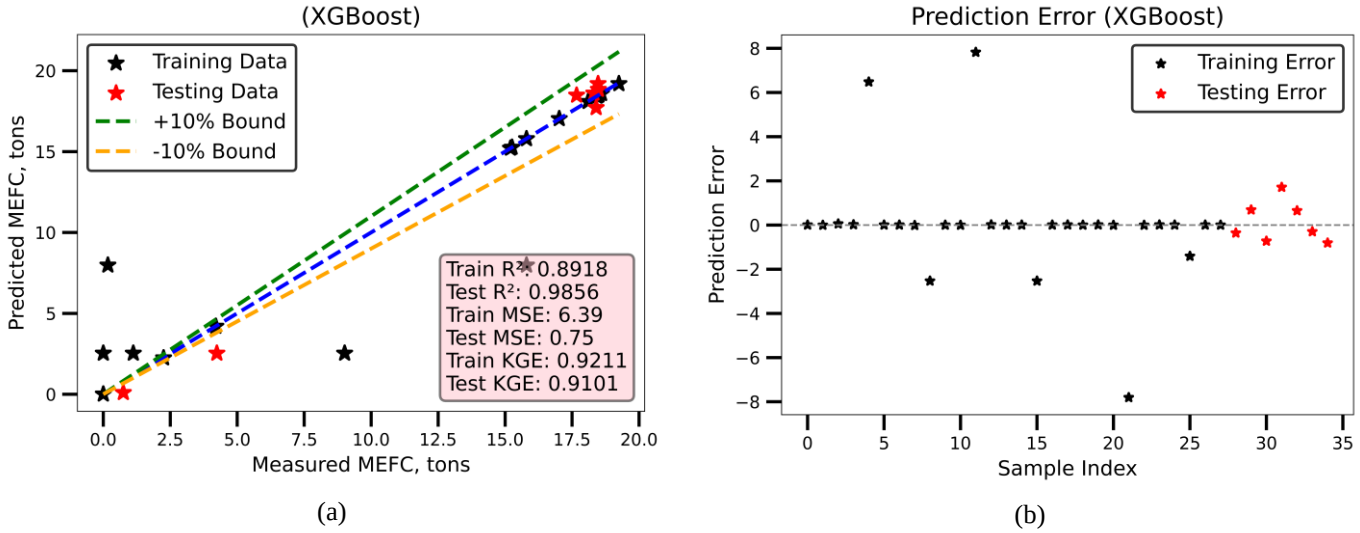
Fig. 7 GB-based MEFC model; (a) Measured vs. predicted (b) Error distribution

Error distribution plot depicted in Figure 7b shows that errors were low. Nonetheless, despite its precision, GB proved computationally demanding, necessitating meticulous hyperparameter optimization to avert overfitting. The model's dependence on sequential learning resulted in slower training compared to parallelizable models such as RF. Nevertheless, its robust test results indicated that it may serve as an effective model for forecasting fuel use, especially in the context of intricate, nonlinear interactions.

3.3.3 Extreme Gradient Boosting

The Extreme Gradient Boosting (XGBoost) Regressor exhibited a minimal test MSE of 0.75, markedly

surpassing the performance of alternative models in terms of error minimization, as shown in Figure 8a, while the error distribution is depicted in Figure 8b. Statistical measures of the XGBoost model are listed in Table 2. The training MSE was 6.39, signifying a well-fitted model devoid of significant overfitting. The R^2 values were notably high, measuring 0.8918 for training and 0.9856 for testing, so affirming its efficacy in capturing variance. The KGE scores of 0.9211 for training and 0.9101 for testing indicated a robust concordance between anticipated and actual values. The model's outstanding performance resulted from its optimized boosting structure, which effectively integrated weak learners to create a highly accurate predictive model. Nonetheless, XGBoost exhibited greater computational complexity compared to more straightforward tree-based models, requiring meticulous adjustment of parameters such as learning rate and tree depth. Moreover, its susceptibility to outliers and skewed datasets necessitated preprocessing to ensure stability. Notwithstanding these issues, its exceptional test performance indicated that it was one of the most dependable choices for predicting MEFC.

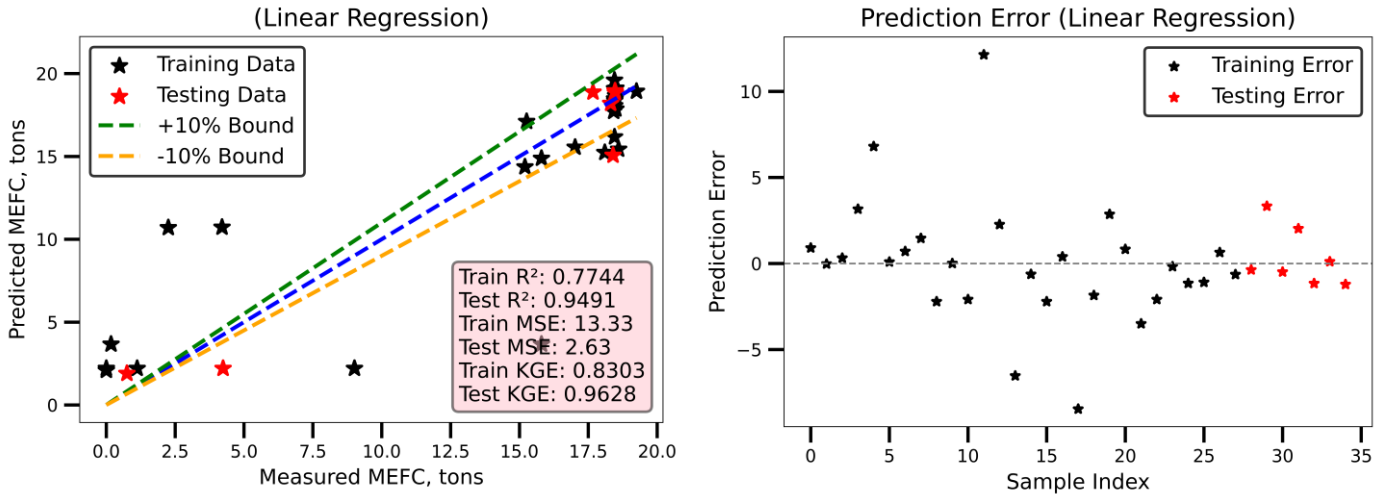


(a) (b)

Fig. 8 XGBoost-based MEFC model; (a) Measured vs. predicted (b) Error distribution

3.3.4 Linear Regression

The LR model plotted in Figure 9a had moderate predictive power, with a training MSE of 13.33 and a higher test MSE of 2.63. The prediction errors over the entire data range are depicted in Figure 9b. Statistical measures of the LR model are listed in Table 2. The model's R^2 value decreased markedly from training (0.7741) to testing (0.9491), suggesting it had difficulty generalizing effectively. The KGE values exhibited a comparable trend, featuring a substantially elevated training score of 0.8303, contrasted with a diminished test score of 0.9628. This indicated that the model did not overfit to the training data. As a parametric model, LR presumes a linear correlation between input features and fuel usage, which may not apply to actual maritime data. Although it is straightforward and interpretable, this restriction rendered it less appropriate for the extremely varied fuel consumption patterns observed in ships.

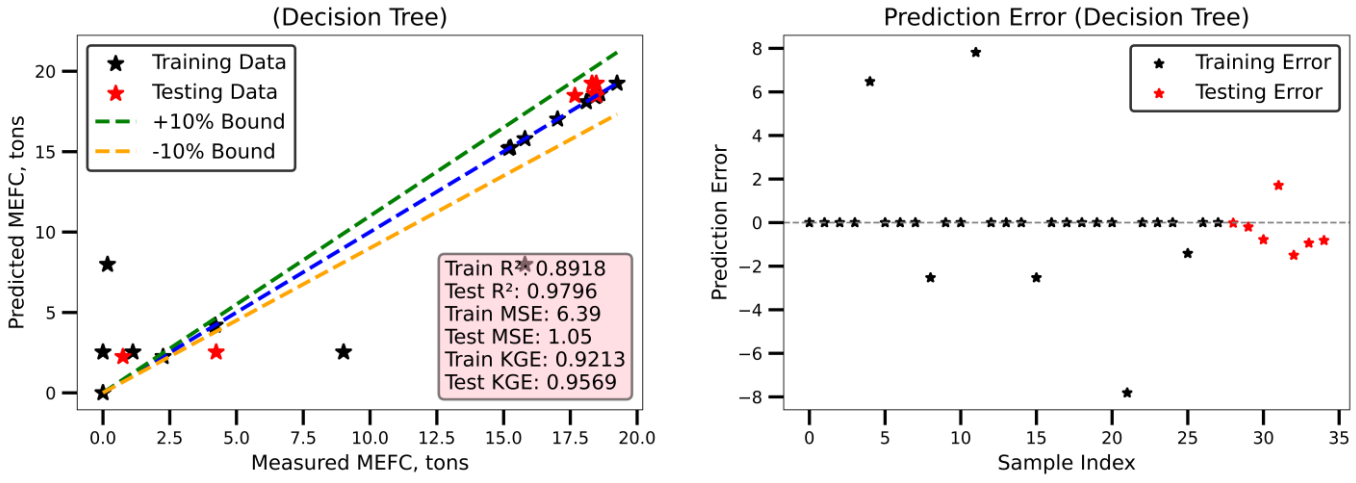


(a) (b)

Fig. 9 LR-based MEFC model; (a) Measured vs. predicted (b) Error distribution

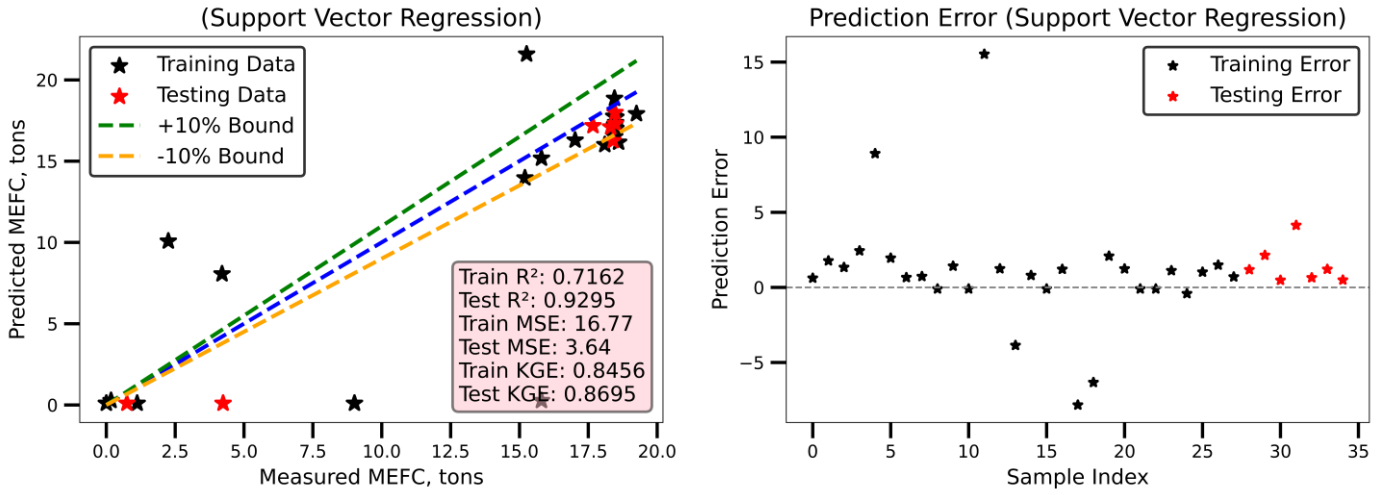
3.3.5 Decision Tree Regressor

As illustrated in Figure 10a, the Decision Tree (DT) model attained a training MSE of 6.39, which is significantly lower than its test MSE of 1.05. Statistical measures of the DT model are listed in Table 2. The R^2 values of 0.8918 for training and 0.9796 for testing demonstrated that the model proficiently captured variation and exhibited strong performance on unseen data. KGE scores exhibited a comparable pattern, recording 0.9213 for training and 0.9569 for testing, thus affirming the model's predictive reliability. The low error distribution as shown in Figure 10b, further corroborates these findings. DTs inherently offer interpretability, facilitating a straightforward display of feature significance and decision paths. Nevertheless, in the absence of meticulous pruning and parameter optimization, they frequently experienced overfitting, which was not evident in this instance. The robust test results indicated that the model effectively reconciled complexity and generalization. Although it exhibited comparable performance to ensemble approaches, its independent architecture may have constrained its efficacy in managing high-dimensional data.



(a) (b)

Fig. 10 DT-based MEFC model's (a) Measured vs. predicted (b) Error distribution



(a) (b)

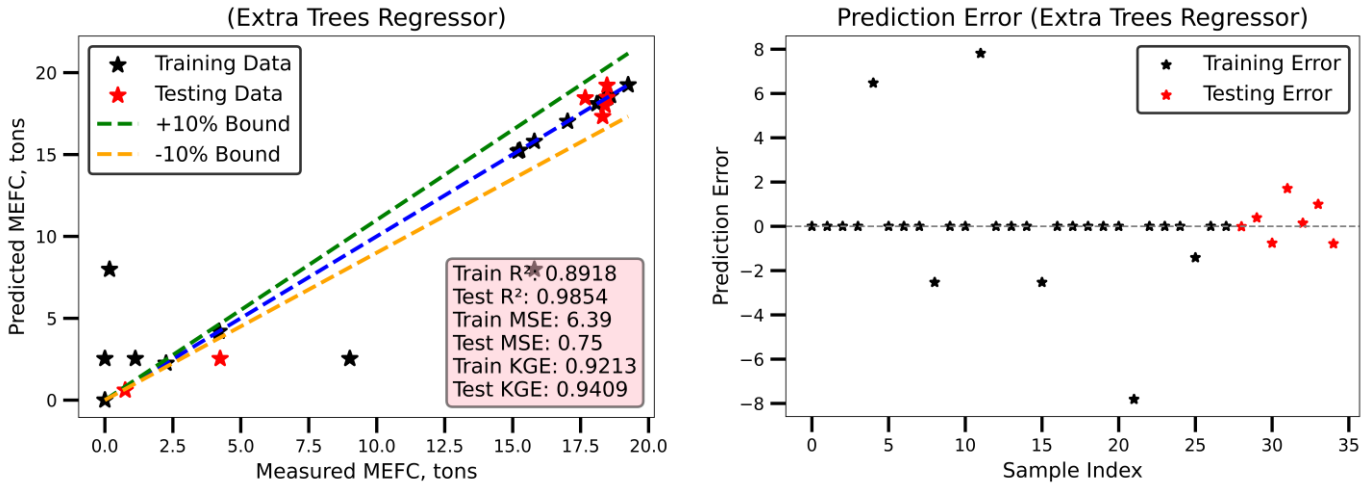
Fig. 11 SVR-based MEFC model: (a) Measured vs. predicted (b) Error distribution

3.3.6 Support Vector Regression

SVR demonstrated inferior performance relative to the tree-based models, evidenced by a training MSE of 16.77 and a test MSE of 3.64, as plotted in Figure 11a. Statistical measures of the SVR model are listed in Table 2. The R^2 values of 0.7162 for training and 0.9295 for testing indicated a good performance in explaining variance, especially for unseen data. KGE values exhibited a comparable trend, recording 0.8456 for training and an improved 0.8695 for testing, signifying reduced consistency in predictions. In addition, the error distribution as depicted in Figure 11b also reveals lower errors in this model. SVR, although theoretically sound for high-dimensional issues, frequently necessitated meticulous adjustment of kernel values and regularization terms to avert suboptimal performance. The substantial rise in error for test data indicated that the model may not have generalized effectively, possibly due to suboptimal kernel selection or inadequate feature scaling. While SVR may be effective in certain contexts, its suboptimal performance in this instance rendered it unsuitable for predicting MEFC.

3.3.7 Extra Trees Regressor

The Extra Trees Regressor (ETR) model exhibited commendable performance, with a training MSE of 6.39 and a testing MSE of 0.75, as depicted in Figure 12a, while the error distribution over the entire data range is depicted in Figure 12b. The statistical measures of the ETR model are listed in Table 2. It could be seen that the R^2 values were 0.8918 for training and 0.9854 for testing, indicating robust predictive accuracy. KGE values of 0.9213 for training and 0.9409 for testing demonstrated strong concordance between predicted and actual MEFC values. Moreover, ETR, as an ensemble method, utilized randomly chosen splits at each node, enhancing variety and diminishing variation. In comparison to RF, it incorporated greater randomness, which improves generalization. The marginal improvement in KGE during testing indicated good generalization of the model on the unseen data supplied. Its computational efficiency was analogous to that of RF; nevertheless, hyperparameter adjustment was essential to equilibrate bias and variance. Due to its stability and robust testing performance, ETR emerged as a viable choice for predicting fuel usage.

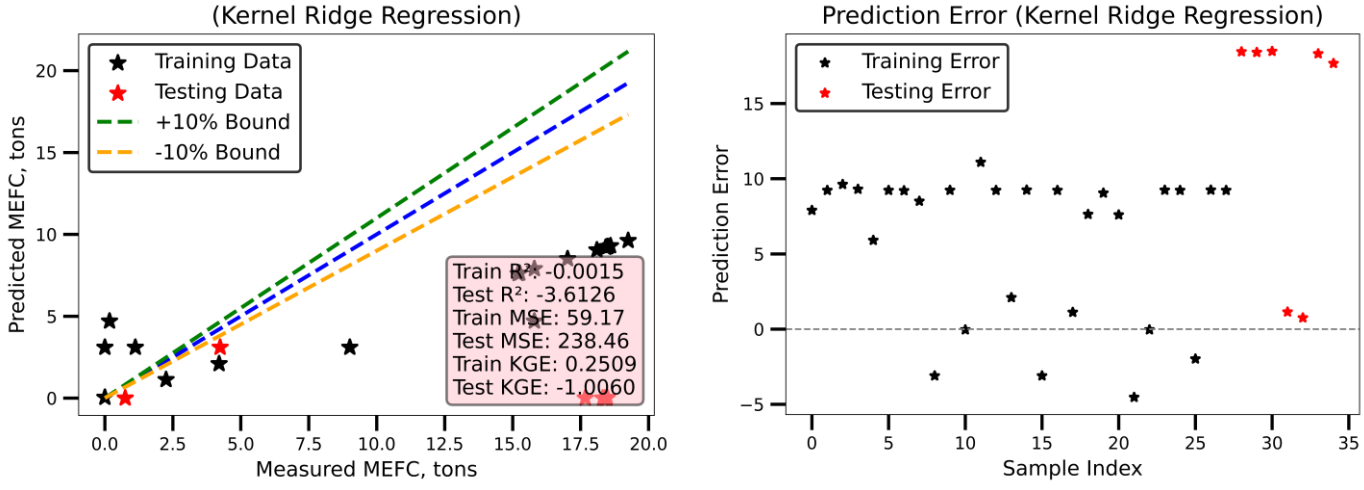


(a) (b)

Fig. 12 ETR-based MEFC model; (a) Measured vs. predicted (b) Error distribution

3.3.8 Kernel Ridge Regression

Kernel Ridge Regression (KRR) in Figure 13a exhibited the poorest performance among all models, demonstrating an exceedingly high training MSE of 59.17 and an even greater test MSE of 238.46. Statistical measures of the KRR model are listed in Table 2. The R^2 values were markedly low at -0.0015 for training and -3.6126 for testing, signifying that the model inadequately accounted for the variance in MEFC. KGE values exhibited a comparable pattern, registering 0.2509 for training and -1.006 for testing, thus affirming its poor generalization. The error distribution was on the higher side as shown in Figure 13b further corroborates that it was a poor model. The negative R^2 and KGE indicated that the model's predictions were substantially inaccurate, potentially due to an unsuitable kernel function or excessive regularization. Due to its inadequate predictive capability, it was improbable that it could serve as a feasible choice for forecasting MEFC without substantial alterations.



(a) (b)

Fig. 13 KRR-based MEFC model; (a) Measured vs. predicted (b) Error distribution

3.4 Model comparison

The examination of several ML models based on statistical measures such as MSE, R^2, and KGE reveals information on their prediction accuracy and dependability for MEFC modeling. RF achieved the lowest test MSE of 0.69 and the highest test R^2 of 0.9867 and test KGE of 0.9681, indicating strong accuracy and generalization. GB and XGBoost showed nearly identical training MSE (6.39) and R^2 (0.8918), with test MSEs

of 0.76 and 0.75, respectively. GB and XGBoost showed almost exactly similar training MSE (6.39) and R^2 (0.8918). Though their test R^2 values were close at 0.9854 and 0.9856, GB's test KGE of 0.9218 somewhat exceeded XGBoost's 0.9101. Although it had a higher test MSE of 1.05 and somewhat lower test R^2 (0.9796) and KGE (0.9569), DT also shared the same training metrics as GB and XGBoost, making it less favorable. Strong performance was indicated by ETR matching GB and XGBoost in training metrics, recording a test MSE of 0.75, test R^2 of 0.9854, and test KGE of 0.9409. Although their test R^2 values, 0.9491 and 0.9295, remain reasonable, LR and SVR had notably higher test MSEs (2.63 and 3.64). Their lower KGE scores (0.9628 and 0.8695), however, indicated less dependability. With negative R^2 and KGE values on the test set and an exceptionally high test MSE of 238.46, KRR performed the lowest across all measures. RF, GB, and XGBoost are the top three performing models based on a balanced evaluation of MSE, R^2, and KGE; they show both high accuracy and strong generalizing capability.

3.5 Development of explainable machine learning

The first plot shown in Figure 14 is a SHAP summary plot, which provides insights into the impact of different features on the MEFC model. Every dot is a data point; colors show feature values, blue for low values and red for high values. With SHAP values ranging from almost -10 to +6, the MES characteristic has the most important impact since it indicates that changes in MES significantly affect the forecasts of the model. WH is also quite important as its SHAP values range from roughly -4 to +4. Though their effects are less apparent, other elements such as WS, CS, and SP have significance.

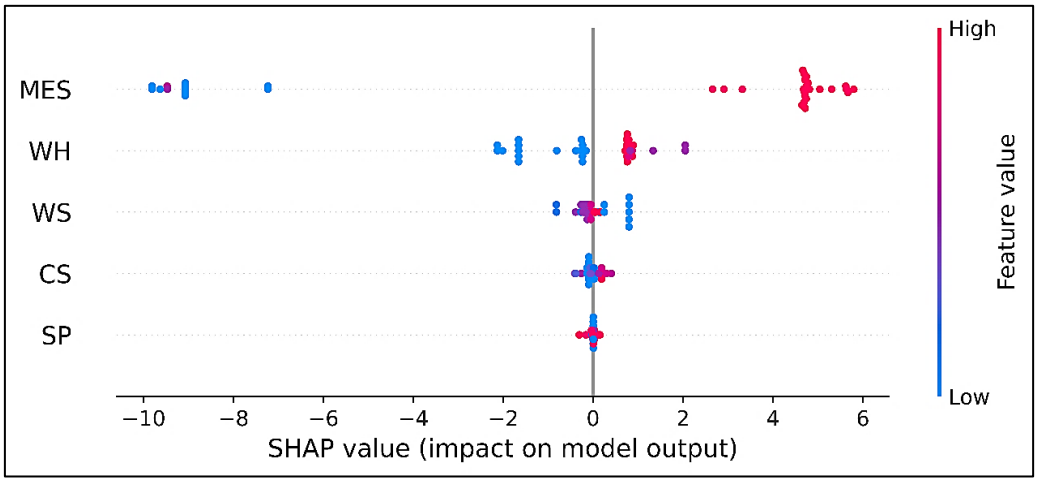


Fig.14 MEFC model's SHAP summary plot

The second plot depicted in Figure 15 represents a LIME-based local explanation for a specific prediction, illustrating which features contributed the most. With a contribution of almost 6.5 units, the MES characteristic (71.83 MES $\leq$ 74.05) has the strongest positive impact. WH (1.00 < WH $\leq$ 2.00) also has a significant influence, adding roughly 2 units. By contrast, WS, CS, and SP barely contribute; their effects are shown by either tiny or almost zero values. Red bars show either minimal or negative effects; green bars show favorable contributions. Employing an explainable interpretation of the model's decision-making process, this LIME visualization aids in understanding how particular feature values form a given forecast.

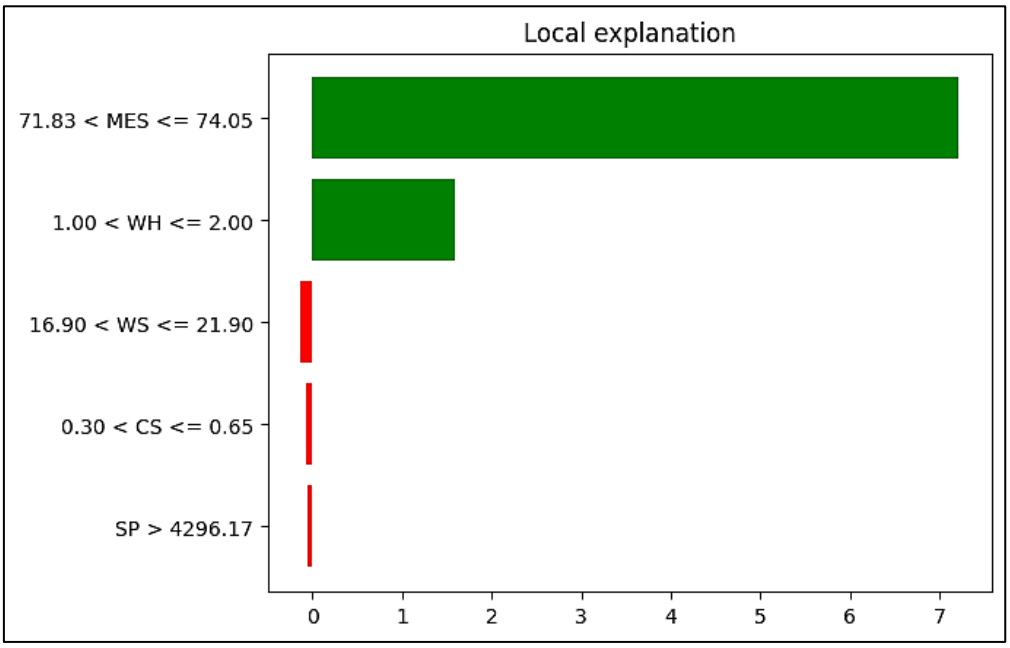


Fig. 15 MEFC model's LIME plots

4. Conclusion

Ship's main engine fuel consumption (MEFC) is a critical factor in both economic and environmental sustainability, as the shipping industry seeks to reduce fuel costs and minimize emissions. Maximizing operational strategies, following legal requirements, and advancing green shipping projects all depend on accurate predictive modeling of fuel consumption. The present study used multiple ML approaches for the development of prediction models, combined with the use of explainable approaches of SHAP and LIME. Indeed, the RF attained the minimal test MSE of 0.69, the maximal test R^2 of 0.9867, and the highest test KGE of 0.9681, signifying robust accuracy and generalization. GB and XGBoost exhibited nearly equivalent training MSEs (6.39) and R^2 (0.8918), with test MSEs of 0.76 and 0.75, respectively. In addition, GB's test KGE of 0.9218 marginally exceeded XGBoost's 0.9101, whereas their test R^2 values were comparable at 0.9854 and 0.9856. Moreover, the DT exhibited identical training metrics to GB and XGBoost; however, it recorded a higher test MSE of 1.05 and marginally lower test R^2 (0.9796) and KGE (0.9569), rendering it less advantageous. ETR equalled GB and XGBoost in training metrics, achieving a test MSE of 0.75, a test R^2 of 0.9854, and a test KGE of 0.9409, signifying robust performance. LR and SVR exhibited markedly elevated test MSEs (2.63 and 3.64), while their test R^2 values (0.9491 and 0.9295) remained acceptable. Nonetheless, their diminished KGE scores (0.9628 and 0.8695) indicated decreased reliability. KRR exhibited the poorest performance across all metrics, demonstrating negative R^2 and KGE values on the test set, along with an exceptionally high test MSE of 238.46.

In summary, findings from this current work provide a robust foundation for selecting appropriate machine learning models for MEFC optimization. SHAP and LIME were used to provide explainability, thereby improving the practical usability of the best model. The SHAP summary figure highlights the worldwide impact of important features, with MES and WH having the biggest influence on fuel consumption projections. MES exhibits SHAP values ranging from approximately -10 to +6, while WH has SHAP values between -4 and +4. Meanwhile, LIME offers localized explanations for individual forecasts, demonstrating that MES contributes roughly 6.5 units favorably and WH adds around 2 units, although WS, CS, and SP have minimal impact. This explainability ensures that the models are accurate and interpretable, allowing maritime operators to make informed decisions. In general, this work supports sustainable marine operations by using explainable machine learning. Future research could investigate hybrid models combining physics-based techniques with machine learning for even more accuracy and interpretability.

ACKNOWLEDGMENTS

This work was supported by Ho Chi Minh City University of Transport, Vietnam, via project KHTĐ2441. The authors also wish to express their appreciation to the Ministry of Education and Training Vietnam for supporting this research under project B2024.DNA.12 entitled "Smart controller for engine fueled with flexible gaseous fuels in hybrid renewable energy system."

REFERENCES

[1] UNCTAD, 2022. Review of Maritime Transport 2022. *UNCTAD (United Nations Conference on Trade and Development)-Review of Maritime Transport*.

[2] Prados, J. M. M., 2024. The decarbonisation of the maritime sector: Horizon 2050. *Brodogradnja*, 75(2), 1-26. https://doi.org/10.21278/brod75202

[3] Trivyza, N. L., Rentizelas, A., Theotokatos, G., Boulougouris, E., 2022. Decision support methods for sustainable ship energy systems: A state-of-the-art review. *Energy*, 239, 122288. https://doi.org/10.1016/j.energy.2021.122288

[4] Uyanık, T., Karatuğ, Ç., Arslanoğlu, Y., 2020. Machine learning approach to ship fuel consumption: A case of container vessel. *Transportation Research Part D: Transportation Environment*, 84, 102389. https://doi.org/10.1016/j.trd.2020.102389

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

[6] Tzu, F.-M., Su, D.-T., 2024. Evaluation of carbon dioxide emission based on energy efficiency existing ship index during oceanographic navigation. *Journal of Operational Oceanography*, 17(3), 151-164. https://doi.org/10.1080/1755876X.2023.2254133

[7] Ammar, N.R., Almas, M., Nahas, Q., 2023. Economic Analysis and the EEXI Reduction Potential of Parallel Hybrid Dual-Fuel Engine–Fuel Cell Propulsion Systems for LNG Carriers. *Polish Maritime Research*, 30(3), 59-70. https://doi.org/doi:10.2478/pomr-2023-0039

[8] Ampah, J. D., Yusuf, A. A., Afrane, S., Jin, C., Liu, H., 2021. Reviewing two decades of cleaner alternative marine fuels: Towards IMO's decarbonization of the maritime transport sector. *Journal of Cleaner Production*, 320, 128871. https://doi.org/10.1016/j.jclepro.2021.128871

[9] IMO, 2023. 2023 IMO strategy on reduction of ghg emissions from ships, Annex 1. https://www.imo.org/en/OurWork/Environment/Pages/2023-IMO-Strategy-on-Reduction-of-GHG-Emissions-from-Ships.aspx. accessed 14th March 2025.

[10] Bilgili, L., Ölçer, A. I., 2024. IMO 2023 strategy-Where are we and what's next?. *Marine Policy*, 160, 105953. https://doi.org/10.1016/j.marpol.2023.105953

[11] Olszewski, W., Dzida, M., Nguyen, V. G., Cao, D. N., 2023. Reduction of CO2 Emissions from Offshore Combined Cycle Diesel Engine-Steam Turbine Power Plant Powered by Alternative Fuels. *Polish Maritime Research,* 30, 71-80. https://doi.org/10.2478/pomr-2023-0040

[12] Guzelbulut, C., Badalotti, T., Suzuki, K., 2025. Impact of Control Strategies for Wind-Assisted Ships on Energy Consumption. *Brodogradnja,* 76, 1–14. https://doi.org/10.21278/brod76104

[13] Kołodziej, R., Hoffmann, P., 2024. Determination of Propeller-Rudder-Hull Interaction Coefficients in Ship Manoeuvring Prediction. *Polish Maritime Research,* 31, 15–24. https://doi.org/10.2478/pomr-2024-0032

[14] Wei, C., Jiang, G., Wu, G., Zhou, Y., Liu, Y., 2024. Effects on of Blended Biodiesel and Heavy Oil on Engine Combustion and Black Carbon Emissions of a Low-Speed Two-Stroke Engine. *Polish Maritime Research,* 31, 94-101. https://doi.org/10.2478/pomr-2024-0010

[15] IMO, 2021. Fourth IMO GHG Study 2020. *International Maritime Organisation*, 6(11), 524.

[16] Wang, Y., Wright, L. A., 2021. A Comparative Review of Alternative Fuels for the Maritime Sector: Economic, Technology, and Policy Challenges for Clean Energy Implementation. *World*, 2, 456–81. https://doi.org/10.3390/world2040029

[17] Nguyen, V. N., Rudzki, K., Marek, D., Pham, N. D. K., Pham, M. T., Nguyen, P. Q. P., Nguyen, X. P., 2023. Understanding fuel saving and clean fuel strategies towards green maritime. *Polish Maritime Research,* 30, 146-64. https://doi.org/10.2478/pomr-2023-0030

[18] Suárez, F. S., Cao, T., Pujol, A. G., Romagnoli, A. 2022. Waste heat recovery on ships. *Sustainable Energy Systems on Ships*, 1, 123–95. https://doi.org/10.1016/B978-0-12-824471-5.00011-6

[19] Naveiro, M., 2023. Energy, exergy, economic and environmental analysis of a regasification system integrating simple ORC and LNG open power cycle in floating storage regasification units. *Brodogradnja,* 74, 39-75. https://doi.org/10.21278/brod74203

[20] Trimulyono, A., Hakim, M. L., Ardhan, C., Ahmad, S. T. P., Tuswan, T., Santosa, A. W. B., 2023. Analysis of the double steps position effect on planing hull performances. *Brodogradnja*, 74, 41–72. https://doi.org/10.21278/brod74403

[21] Karczewski, A., Kozak, J. A., 2023. Generative Approach to Hull Design for a Small Watercraft. *Polish Maritime Research*, 30, 4–12. https://doi.org/doi:10.2478/pomr-2023-0001

[22] Tay, Z. Y., Hadi, J., Chow, F., Loh, D. J., Konovessis, D., 2021. Big Data Analytics and Machine Learning of Harbour Craft Vessels to Achieve Fuel Efficiency: A Review. *Journal of Marine Science and Engineering*, 9(12), 1351. https://doi.org/10.3390/jmse9121351

[23] Li, Y., Cui, J., Zhang, X., Yang X., 2023. A Ship Route Planning Method under the Sailing Time Constraint. *Journal of Marine Science and Engineering,* 11, 1242. https://doi.org/10.3390/jmse11061242

[24] Gray, N., McDonagh, S., O'Shea, R., Smyth, B., Murphy, J. D., 2021. Decarbonising ships, planes and trucks: An analysis of suitable low-carbon fuels for the maritime, aviation and haulage sectors. *Advances in Applied Energy*, 1, 100008. https://doi.org/10.1016/j.adapen.2021.100008

[25] Council of the EU, 2023. REGULATION (EU) 2023/1805 OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL of 13 September 2023 on the use of renewable and low-carbon fuels in maritime transport, and amending Directive 2009/16/EC (Text with EEA relevance). *Office Journal of the European Union*.

[26] UNCTAD, 2020. Review of Maritime Transport 2020. *UNCTAD (United Nations Conference on Trade and Development)-Review of Maritime Transport*.

[27] Singh, D. V., Pedersen, E., 2016. A review of waste heat recovery technologies for maritime applications. *Energy Conversion and Management*, 111, 315-328. https://doi.org/10.1016/j.enconman.2015.12.073

[28] Council of the EU, 2025. Fit for 55. https://www.consilium.europa.eu/en/policies/fit-for-55. accessed 14th March 2025.

[29] European Commission (EC), 2013. EU Emissions Trading System. https://climate.ec.europa.eu/eu-action/eu-emissions-trading-system-eu-ets_en. accessed 14th March 2025.

[30] Verde, S. F., Borghesi, S., 2022. The International Dimension of the EU Emissions Trading System: Bringing the Pieces Together. *Environmental and Resource Economics*, 83(1), 23-46. https://doi.org/10.1007/s10640-022-00705-x

[31] Inal, O. B., 2024. Decarbonization of shipping: Hydrogen and fuel cells legislation in the maritime industry. *Brodogradnja,* 75, 1-13. https://doi.org/10.21278/brod75205

[32] Pham, N. D. K., Dinh, G. H., Pham, H. T., Kozak, J., Nguyen, H. P., 2023. Role of Green Logistics in the Construction of Sustainable Supply Chains. *Polish Maritime Research*, 30, 191-211. https://doi.org/10.2478/pomr-2023-0052

[33] Kondratenko, A.A., Zhang, M., Tavakoli, S., Altarriba, E., Hirdaris, S., 2025. Existing technologies and scientific advancements to decarbonize shipping by retrofitting. *Renewable and Sustainable Energy Reviews*, 212, 115430. https://doi.org/10.1016/j.rser.2025.115430

[34] Wang, H., Yan, R., Wang, S., Zhen, L., 2023. Innovative approaches to addressing the tradeoff between interpretability and accuracy in ship fuel consumption prediction. *Transportation Research Part C: Emerging Technologies*, 157, 104361. https://doi.org/10.1016/j.trc.2023.104361

[35] Nguyen, P. Q. P., Nguyen, D. T., D., Ha, T. Y. N., Le, Q., Nguyen, N. T., Pham, N. D. K., 2025. Machine Learning-Driven Insights for Optimizing Ship Fuel Consumption : Predictive Modeling and Operational Efficiency. *International Journal on Advanced Science, Engineering and Information Technology*, 15, 27-35. https://doi.org/10.18517/ijaseit.15.1.12374

[36] Nguyen, V. G., Rajamohan, S., Rudzki, K., Kozak, J., Sharma, P., Pham, N. D. K., Nguyen, P. Q. P., Xuan, P. N., 2023. Using Artificial Neural Networks for Predicting Ship Fuel Consumption. *Polish Maritime Research*, 30(2), 39-60. https://doi.org/10.2478/pomr-2023-0020

[37] Zhou, Y., Pazouki, K., Murphy, A. J., Uriondo, Z., Granado, I., Quincoces, I., Fernandes-Salvador, J. A., 2023. Predicting ship fuel consumption using a combination of metocean and on-board data. *Ocean Engineering*, 285, 115509. https://doi.org/10.1016/j.oceaneng.2023.115509

[38] Fan, A., Yang, J., Yang, L., Wu, D., Vladimir, N., 2022. A review of ship fuel consumption models. *Ocean Engineering*, 264, 112405. https://doi.org/10.1016/j.oceaneng.2022.112405

[39] Yan, R., Wang, S., Du, Y., 2020. Development of a two-stage ship fuel consumption prediction and reduction model for a dry bulk ship. *Transportation Research Part E: Logistics and Transportation Review*, 138, 101930. https://doi.org/10.1016/j.tre.2020.101930

[40] Zhang, M., Tsoulakos, N., Kujala, P., Hirdaris, S., 2024. A deep learning method for the prediction of ship fuel consumption in real operational conditions. *Engineering Applications of Artificial Intelligence*, 130, 107425. https://doi.org/10.1016/j.engappai.2023.107425

[41] Xie, X., Sun, B., Li, X., Olsson, T., Maleki, N., Ahlgren, F., 2023. Fuel Consumption Prediction Models Based on Machine Learning and Mathematical Methods. *Journal of Marine Science and Engineering*, 11(4), 738. https://doi.org/10.3390/jmse11040738

[42] Wang, K., Wang, J., Huang, L., Yuan, Y., Wu, G., Xing, H., Wang, Z., Wang, Z., Jiang, X., 2022. A comprehensive review on the prediction of ship energy consumption and pollution gas emissions. *Ocean Engineering*, 266, 112826. https://doi.org/10.1016/j.oceaneng.2022.112826

[43] Yang, H., Sun, Z., Han, P., Ma, M., 2024. Data-driven prediction of ship fuel oil consumption based on machine learning

models considering meteorological factors. *Proceedings of the Institution of Mechanical Engineers, Part M: Journal of Engineering for the Maritime Environment*, 238(3), 483-502. https://doi.org/10.1177/14750902231210047

[44] Gkerekos, C., Lazakis, I., Theotokatos, G., 2019. Machine learning models for predicting ship main engine Fuel Oil Consumption: A comparative study. *Ocean Engineering*, 188, 106282. https://doi.org/10.1016/j.oceaneng.2019.106282

[45] Zhang, M., Taimuri, G., Zhang, J., Zhang, D., Yan, X., Kujala, P., Spyros, H., 2025. Systems driven intelligent decision support methods for ship collision and grounding prevention: Present status, possible solutions, and challenges. *Reliability Engineering & System Safety*, 253, 110489. https://doi.org/10.1016/j.ress.2024.110489

[46] Uyanik, T., Kalenderli, O., Arslanoglu, Y., 2019. Ship Fuel Consumption Prediction with Machine Learning. *4th International Mediterranean Science and Engineering Congress*, 25-27 April, Alanya, Antalya, Turkey, 757-759.

[47] Ramadhan, A., Susetyo, B., -Indahwati., 2021. Classification Modelling of Random Forest to Identify the Important Factors in Improving the Quality of Education. *International Journal on Advanced Science, Engineering and Information Technology*, 11, 501-507. https://doi.org/10.18517/ijaseit.11.2.8878

[48] Chen, J., Wang, X., Lei, F., 2024. Data-driven multinomial random forest: a new random forest variant with strong consistency. *Journal of Big Data*, 11(1), 34. https://doi.org/10.1186/s40537-023-00874-6

[49] Alruqi, M., Sharma, P., Algburi, S., Khan, M. A., Alsubih, M., Islam, S., 2024. Biomass energy transformation: Harnessing the power of explainable ai to unlock the potential of ultimate analysis data. *Environmental Technology & Innovation*, 35, 103652. https://doi.org/10.1016/j.eti.2024.103652

[50] Talekar, B., 2020. A Detailed Review on Decision Tree and Random Forest. *Bioscience Biotechnology Research Communications*, 13(14), 245-248. http://doi.org/10.21786/bbrc/13.14/57

[51] Villegas-Mier, C., Rodriguez-Resendiz, J., Álvarez-Alvarado, J., Jiménez-Hernández, H., Odry, Á., 2022. Optimized Random Forest for Solar Radiation Prediction Using Sunshine Hours. *Micromachines*, 13(9), 1406. https://doi.org/10.3390/mi13091406

[52] Tyralis, H., Papacharalampous, G., Langousis, A., 2019. A Brief Review of Random Forests for Water Scientists and Practitioners and Their Recent History in Water Resources. *Water*, 11(5), 910. https://doi.org/10.3390/w11050910

[53] Belgiu, M., Drăguţ, L., 2016. Random forest in remote sensing: A review of applications and future directions. *ISPRS Journal of Photogrammetry and Remote Sensing*, 114, 24-31. https://doi.org/10.1016/j.isprsjprs.2016.01.011

[54] Deif, A. M., Hammam, E. R., Solyman, A. A., 2021. Gradient Boosting Machine Based on PSO for prediction of Leukemia after a Breast Cancer Diagnosis. *International Journal on Advanced Science, Engineering and Information Technology*, 11(2), 508-515. https://doi.org/10.18517/ijaseit.11.2.12955.

[55] Xu, N., Wang, Z., Dai, Y., Li, Q., Zhu, W., Wang, R., Finkelman, R. B., 2023. Prediction of higher heating value of coal based on gradient boosting regression tree model. *International Journal of Coal Geology,* 274, 104293. https://doi.org/10.1016/j.coal.2023.104293

[56] Yang, F., Wang, D., Xu, F., Huang, Z., Tsui, K.-L., 2020. Lifespan prediction of lithium-ion batteries based on various extracted features and gradient boosting regression tree model. *Journal of Power Sources,* 476, 228654. https://doi.org/10.1016/j.jpowsour.2020.228654

[57] Wen, H.-T., Lu, J.-H., Phuc, M.-X., 2021. Applying Artificial Intelligence to Predict the Composition of Syngas Using Rice Husks: A Comparison of Artificial Neural Networks and Gradient Boosting Regression. *Energies*, 14(10), 2932. https://doi.org/10.3390/en14102932

[58] Kumar K, P., Alruqi, M., Hanafi, H. A., Sharma, P., Wanatasanappan, V. V., 2024. Effect of particle size on second law of thermodynamics analysis of Al2O3 nanofluid: Application of XGBoost and gradient boosting regression for prognostic analysis. *International Journal of Thermal Sciences*, 197, 108825. https://doi.org/10.1016/j.ijthermalsci.2023.108825

[59] Rozam, N. F., Riasetiawan, M., 2023. XGBoost Classifier for DDOS Attack Detection in Software Defined Network Using sFlow Protocol. *International Journal on Advanced Science Engineering and Information Technology*, 13, 718-725. https://doi.org/10.18517/ijaseit.13.2.17810

[60] Handayani, M. P., 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. *Journal of Marine Science and Engineering*, 11(11), 2165. https://doi.org/10.3390/jmse11112165

[61] Du, Y., Chen, Y., Li, X., Schönborn, A., Sun, Z., 2022. Data fusion and machine learning for ship fuel efficiency modeling: Part III – Sensor data and meteorological data. *Communications in Transportation Research*, 2, 100072. https://doi.org/10.1016/j.commtr.2022.100072

[62] Li, X., Du, Y., Chen, Y., Nguyen, S., Zhang, W., Schönborn, A., Sun, Z., 2022. Data fusion and machine learning for ship fuel efficiency modeling: Part I – Voyage report data and meteorological data. *Communications in Transportation Research*, 2, 100074. https://doi.org/10.1016/j.commtr.2022.100074

[63] Hu, Z., Zhou, T., Osman, M. T., Li, X., Jin, Y., Zhen, R., 2021. A Novel Hybrid Fuel Consumption Prediction Model for Ocean-Going Container Ships Based on Sensor Data. *Journal of Marine Science and Engineering*, 9(4), 449. https://doi.org/10.3390/jmse9040449

[64] Chen, T., Guestrin, C., 2016. XGBoost: A scalable tree boosting system. *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 785-794. https://doi.org/10.1145/2939672.2939785

[65] Shanmugasundar, G., Vanitha, M., Čep, R., Kumar, V., Kalita, K., Ramachandran, M., 2021. A Comparative Study of Linear, Random Forest and AdaBoost Regressions for Modeling Non-Traditional Machining. *Processes*, 9(11), 2015. https://doi.org/10.3390/pr9112015

[66] Pandit, P., Dey, P., Krishnamurthy, K. N., 2021. Comparative Assessment of Multiple Linear Regression and Fuzzy Linear Regression Models. *SN Computer Science*, 2(2), 76. https://doi.org/10.1007/s42979-021-00473-3

[67] Maulud, D., Abdulazeez, A. M., 2020. A Review on Linear Regression Comprehensive in Machine Learning. *Journal of Applied Science and Technology Trends*, 1(2), 140-147. https://doi.org/10.38094/jastt1457

[68] al-Swaidani, A. M., Khwies, W. T., Al-Baly, M., Lala, T., 2022. Development of multiple linear regression, artificial neural networks and fuzzy logic models to predict the efficiency factor and durability indicator of nano natural pozzolana as cement additive. *Journal of Building Engineering*, 52, 104475. https://doi.org/10.1016/j.jobe.2022.104475

[69] Mienye, I. D., Sun, Y., Wang, Z., 2019. Prediction performance of improved decision tree-based algorithms: a review. *Procedia Manufacturing*, 35, 698-703. https://doi.org/10.1016/j.promfg.2019.06.011

[70] Nanfack, G., Temple, P., Frénay, B., 2022. Constraint Enforcement on Decision Trees: A Survey. *ACM Computing Surveys*, 54(10s), 1-36. https://doi.org/10.1145/3506734

[71] Canete-Sifuentes, L., Monroy, R., Medina-Perez, M. A., 2021. A Review and Experimental Comparison of Multivariate Decision Trees. *IEEE Access*, 9, 110451-110479. https://doi.org/10.1109/ACCESS.2021.3102239

[72] Sagi, O., Rokach, L., 2021. Approximating XGBoost with an interpretable decision tree. *Information Sciencesm*, 572, 522-542. https://doi.org/10.1016/j.ins.2021.05.055

[73] Sabzekar, M., Hasheminejad, S. M. H., 2021. Robust regression using support vector regressions. *Chaos, Solitons & Fractals*, 144, 110738. https://doi.org/10.1016/j.chaos.2021.110738

[74] Dong, H., Jin, X., Lou, Y., Wang, C., 2014. Lithium-ion battery state of health monitoring and remaining useful life prediction based on support vector regression-particle filter. *Journal of Power Sources*, 271, 114-123. https://doi.org/10.1016/j.jpowsour.2014.07.176

[75] Sharifzadeh, M., Sikinioti-Lock, A., Shah, N., 2019. Machine-learning methods for integrated renewable power generation: A comparative study of artificial neural networks, support vector regression, and Gaussian Process Regression. *Renewable and Sustainable Energy Reviews*, 108, 513-538. https://doi.org/10.1016/j.rser.2019.03.040

[76] Ban, B., Yang, J., Chen, P., Xiong, J., Wang, Q., 2017. Ship Track Regression Based on Support Vector Machine. *IEEE Access*, 5, 18836 - 18846. https://doi.org/10.1109/ACCESS.2017.2749260

[77] Malakouti, S. M., Menhaj, M. B., Suratgar, A. A., 2023. The usage of 10-fold cross-validation and grid search to enhance ML methods performance in solar farm power generation prediction. *Cleaner Engineering and Technology*, 15, 100664. https://doi.org/10.1016/j.clet.2023.100664

[78] Hazare, S. R., Vala, S. V., Patil, C. S., Joshi, A. J., Joshi, J. B., Vitankar, V. S., Patwardhan, A. W., 2023. Correlating Interfacial Area and Volumetric Mass Transfer Coefficient in Bubble Column with the Help of Machine Learning Methods. *Industrial & Engineering Chemistry Research*, 62(5), 2104-2123. https://doi.org/10.1021/acs.iecr.2c02820

[79] Chen, C., Wang, N., Chen, M., 2021. Prediction Model of End-point Phosphorus Content in Consteel Electric Furnace Based on PCA-Extra Tree Model. *ISIJ International*, 61(6), 1908-1914. https://doi.org/10.2355/isijinternational.ISIJINT-2020-615

[80] Ahmad, M. W., Reynolds, J., Rezgui, Y., 2018. Predictive modelling for solar thermal energy systems: A comparison of support vector regression, random forest, extra trees and regression trees. *Journal of Cleaner Production*, 203, 810-821. https://doi.org/10.1016/j.jclepro.2018.08.207

[81] Alsariera, Y. A., Adeyemo, V. E., Balogun, A. O., Alazzawi, A. K., 2020. AI Meta-Learners and Extra-Trees Algorithm for the Detection of Phishing Websites. *IEEE Access*, 8, 142532-142542. https://doi.org/10.1109/ACCESS.2020.3013699

[82] Elton, D. C., Boukouvalas, Z., Butrico, M. S., Fuge, M. D., Chung, P. W., 2018. Applying machine learning techniques to predict the properties of energetic materials. *Scientific Reports*, 8(1), 9059. https://doi.org/10.1038/s41598-018-27344-x

[83] Mengesha, B. N., Shaeri, M. R., Sarabi, S., 2022. Application of Machine Learning to Predict Thermal Performances of Heat Sinks. *Proceedings of the 8th World Congress on Mechanical, Chemical, and Material Engineering*, 31 July-2 August, Prague, Czech Republic. https://doi.org/10.11159/htff22.138

[84] Wu, X. H., Lu, Y. Y., Zhao, P. W., 2022. Multi-task learning on nuclear masses and separation energies with the kernel ridge regression. *Physics Letters B*, 834, 137394. https://doi.org/10.1016/j.physletb.2022.137394

[85] Zhang, R., Li, Y., Gui, Y., 2023. Prediction of rock blasting induced air overpressure using a self-adaptive weighted kernel ridge regression. *Applied Soft Computing*, 148, 110851. https://doi.org/10.1016/j.asoc.2023.110851

[86] Stuke, A., Rinke, P., Todorović, M., 2021. Efficient hyperparameter tuning for kernel ridge regression with Bayesian optimization. *Machine Learning: Science and Technology*, 2(3), 035022. https://doi.org/10.1088/2632-2153/abee59

[87] Ahmed, A. A. M., Sharma, E., Jui, S. J. J., Deo, R. C., Nguyen, H. T., Ali, M., 2022. Kernel Ridge Regression Hybrid Method

for Wheat Yield Prediction with Satellite-Derived Predictors. *Remote Sensing*, 14(5), 1136. https://doi.org/10.3390/rs14051136

[88] Ma, Y., Zhao, Y., Yu, J., Zhou, J., Kuang, H., 2023. An Interpretable Gray Box Model for Ship Fuel Consumption Prediction Based on the SHAP Framework. *Journal of Marine Science and Engineering*, 11(5), 1059. https://doi.org/10.3390/jmse11051059

[89] Lee, J., Eom, J., Park, J., Jo, J., Kim, S., 2024. The Development of a Machine Learning-Based Carbon Emission Prediction Method for a Multi-Fuel-Propelled Smart Ship by Using Onboard Measurement Data. *Sustainability*, 16(6), 2381. https://doi.org/10.3390/su16062381

[90] Kuo, C., 2019. Explain your model with the SHAP values. Towards Data Science. https//towardsdatascience.com/explain-your-model-with-the-shap-values-bc36aac4de3d. accessed 14th March 2025.

[91] Pham, N. D. K., Dinh, G. H., Nguyen, C. L., Dang, H. Q., Pham, H. T., Nguyen, Q. T., Tran, M. C., 2025. Forecasting and Feature Analysis of Ship Fuel Consumption by Explainable Machine Learning Approaches. *Polish Maritime Research*, 32, 81-94. https://doi.org/10.2478/pomr-2025-0008

[92] Nguyen, V. N., Chung, N., Balaji, G. N., Rudzki, K., Hoang, A. T., 2025. Internet of things-driven approach integrated with explainable machine learning models for ship fuel consumption prediction. *Alexandria Engineering Journal*, 118, 664-680. https://doi.org/10.1016/j.aej.2025.01.067

[93] Barhrhouj, A., Ananou, B., Ouladsine, M., 2025. Exploring Explainable Machine Learning for Enhanced Ship Performance Monitoring. *Machine Learning, Optimization, and Data Science*, 15509, 1–13. https://doi.org/10.1007/978-3-031-82484-5_1

[94] Chicco, D., Warrens, M.J., Jurman, G., 2021. The coefficient of determination R-squared is more informative than SMAPE, MAE, MAPE, MSE and RMSE in regression analysis evaluation. *Peerj Computer Science*, 7, e623. https://doi.org/10.7717/peerj-cs.623

[95] Zhang, M., Wang, H., Conti, F., Manderbacka, T., Remes, H., Hirdaris, S., 2025. A hybrid deep learning method for the real-time prediction of collision damage consequences in operational conditions. *Engineering Applications of Artificial Intelligence*. 145, 110158. https://doi.org/10.1016/j.engappai.2025.110158

[96] Vrugt, J.A., de Oliveira, D.Y., 2022. Confidence intervals of the Kling-Gupta efficiency. *Journal of Hydrology*, 612, 127968. https://doi.org/10.1016/j.jhydrol.2022.127968

[97] Zhang, M., Tsoulakos, N., Kujala, P., Hirdaris, S., 2024. A deep learning method for the prediction of ship fuel consumption in real operational conditions. *Engineering Applications of Artificial Intelligence*, 130, 107425. https://doi.org/10.1016/j.engappai.2023.107425