

## Article

# The Development of a Machine Learning-Based Carbon Emission Prediction Method for a Multi-Fuel-Propelled Smart Ship by Using Onboard Measurement Data

Juhyang Lee <sup>1</sup>, Jeongon Eom <sup>1</sup> , Jumi Park <sup>1</sup>, Jisung Jo <sup>2</sup> and Sewon Kim <sup>1,\*</sup> 

<sup>1</sup> Department of Intelligent Mechatronics Engineering, Sejong University, Seoul 05006, Republic of Korea; wngid0822@naver.com (J.L.); djawjddhs@naver.com (J.E.); jumi.park@hanwha.com (J.P.)

<sup>2</sup> Logistics and Maritime Industry Research Department, Korea Maritime Institute, Busan 49111, Republic of Korea; jisungjo@kmi.re.kr

\* Correspondence: sewonkim@sejong.ac.kr; Tel.: +82-10-3123-7176

**Abstract:** Zero-carbon shipping is the prime goal of the seaborne trade industry at this moment. The utilization of ammonia and liquid hydrogen propulsion in a carbon-free propulsion system is a promising option to achieve net-zero emission in the maritime supply chain. Meanwhile, optimal ship voyage planning is a candidate to reduce carbon emissions immediately without new buildings and retrofits of the alternative fuel-based propulsion system. Due to the voyage options, the precise prediction of fuel consumption and carbon emission via voyage operation profile optimization is a prerequisite for carbon emission reduction. This paper proposes a novel fuel consumption and carbon emission quantity prediction method which is based on the onboard measurement data of a smart ship. The prediction performance of the proposed method was investigated and compared to machine learning and LSTM-model-based fuel consumption and gas emission prediction methods. The results had an accuracy of 81.5% in diesel mode and 91.2% in gas mode. The SHAP (Shapley additive explanations) model, an XAI (Explainable Artificial Intelligence), and a CO<sub>2</sub> consumption model were employed to identify the major factors used in the predictions. The accuracy of the fuel consumption calculated using flow meter data, as opposed to power load data, improved by approximately 21.0%. The operational and flow meter data collected by smart ships significantly contribute to predicting the fuel consumption and carbon emissions of vessels.

**Keywords:** dual-fuel engines; fuel consumption prediction; smart ship; carbon emission calculation; machine learning; LSTM; SHAP



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

Academic Editor: Tingsong Wang

Received: 2 February 2024

Revised: 28 February 2024

Accepted: 8 March 2024

Published: 13 March 2024



**Copyright:** © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

## 1. Introduction

### 1.1. Background

Maritime transportation is the predominant mode of transportation, responsible for 90% of world trade [1]. The process of operating ships for maritime transportation consumes fuel and emits carbon dioxide. This makes the shipping industry an important contributor to global carbon emissions. According to the IMO Greenhouse Gas Study, ship emissions account for about 3% of global anthropogenic emissions [2]. Therefore, decarbonization is important for the sustainable development of the shipping industry. Kim et al. [3] proposed an algorithm to reduce fuel oil consumption and Greenhouse Gas (GHG) emissions by controlling the speed in a ship's route planning procedure. Perera and Mo [4] studied voyage planning to reduce emissions and performed an analysis of the ship's draft and trim. Laasma et al. [5] analyzed and evaluated alternative fuels for small ships. The study concluded that the most suitable methods are fully electric or diesel–electric hybrid solutions. Zhou and Zhang [6] considered the economic and social performance based on the fuel type of ships to study emission control technologies in the port supply chain within a cap-and-trade scheme. To reduce carbon emissions from ships,

various efforts are being made to manage ship operations, analyze the performance of ship components, and consider alternative fuels. Among these efforts, the IMO has set decarbonization targets and wants to tighten carbon regulations further. The targets set by the IMO require the shipping industry to reduce carbon emissions by at least 20% by 2030, at least 40% by 2024, and strive to achieve net-zero by 2050. With the goal of reducing carbon, the shipping industry is increasingly interested in researching sustainable operating models, environmentally friendly technologies, renewable energy, and reward schemes.

Studies have analyzed the atmospheric impact of fuel and carbon emissions from ships. He, Shaoli et al. [7] found that emissions from ships have long-term effects on air quality around waterways. Corbett, James J. et al. [8] calculated emissions by considering ship characteristic information and traffic density. The study showed an increase in premature deaths from cardiopulmonary diseases due to vessel emissions. Jonson, Jan Eiof et al. [9] investigated air quality with weather data and emissions from ships and found that the impact of emissions is greater near the shore than on the high seas, and Saxe, H. and T. Larsen [10] observed that it can cause health problems for people near ports. Ballini, Fabio, and Riccardo Bozzo [11] evaluated air pollution caused by ships in terms of external health cost. Chatzinikolaou, Stefanos D. et al. [12] also analyzed the impact of ship emissions on human health for a port in Greece but found that the analysis needed to be supplemented with information on actual ship operations. These studies show that managing emissions is an important issue as ships' fuel and carbon emissions can affect air quality and worsen people's health. Therefore, we expect to be able to estimate fuel consumption from the ship's onboard data to improve carbon emissions management.

## 1.2. Literature Review

Many studies have estimated fuel consumption and the CO<sub>2</sub> emission of ships. We categorized the literature on ship fuel consumption estimation models into physical models and data-driven artificial intelligence models. We aimed to investigate the relevant literature and review models that predict the fuel consumption of ships based on the characteristics of the estimation model and the utilized data.

Models that estimate fuel consumption using ship specification information or organization standards mainly have physical model characteristics. Miola, A and Ciuffo, B [13] calculated the energy consumed by the main and auxiliary engines to estimate the fuel consumption of a ship. They used the specific fuel oil consumption (SFOC) and emission factors discussed by the IMO and utilized the calculation procedure agreed upon by the IMO Marine Environmental Protection Committee (MEPC). Lin, Y. H. et al. [14] utilized the engine information of a specific container ship and the specific fuel consumption rate of a container ship defined by the International Organization for Standardization (ISO) to estimate fuel consumption. Vettor, R. and Soares, C. G [15]; Veneti, Aphrodite et al. [16]; and Kim, S.W., and Eom, J.O. [17] calculated fuel consumption using the ship's engine model information. Chen, Dongsheng et al. [18] defined an emission estimation model with power, load factor, operating time, and emission substance coefficients according to engine type. Moreno-Gutiérrez, Juan et al. [19] compared the energy consumption calculation methods of the EPA, IMO, Jalkanen, and MAN and proposed an advantageous method that can calculate energy consumption in real-time by calculating the power coefficient based on ship types and estimating fuel consumption for the main engine, auxiliary engine, and operation mode. Wijaya, A. T. A. et al. [20] used the Holtrop method and the STAWAVE method to monitor ship fuel and derive real-time emissions using empirical formulas that consider basic resistance, wind resistance, waves, draft, and water characteristics. Kim, K. S. and Roh, M. I. [21] improved the fuel consumption estimation procedure based on ISO15016:2015 [22]. The authors accounted for the required power in fuel consumption by considering the marine environment and additional resistance. Guo, Bingjie et al. [23] calculated the power of the main engine from the estimated propeller efficiency, considering the additional drag caused by waves and wind.

Physical models can predict fuel consumption based on the characteristics of the vessel but are affected by details such as vessel-specific engine information and operating conditions. On the other hand, artificial intelligence models analyze relationships between elements and can more easily process large datasets, making it possible to estimate fuel consumption using real-time voyage and weather data. Therefore, we review studies that apply artificial intelligence models when estimating fuel consumption using a ship's AIS data, weather data, or onboard data.

Wang, Shengzheng et al. [24] analyzed the correlation between a ship's cargo weight, draft, and engine load and predicted fuel consumption with a LASSO regression algorithm considering operational and weather data. Okumuş, F. et al. [25] developed and compared methodologies for four regression algorithms to estimate engine power based on ship length, gross tonnage, and age data collected from seven different ship types. Among the simple linear regression, polynomial regression, K-nearest neighbors (KNN) regression, and gradient boosting machine (GBM) regression algorithms, the GBM regression algorithm provided the most accurate estimation and was expected to be applicable to emission calculations. Chen, Z. S. et al. [26] collected hourly operational data and meteorological data of a tugboat to predict fuel consumption. The authors collected ship speed, main engine power, ambient temperature, ambient relative humidity, and fuel consumption from the engine and analyzed the correlation and found that the ship speed had the largest influence on the engine power. The random forest model exhibited significantly smaller error than the ridge regression and support vector regression models. Le, Luan Thanh et al. [27] utilized operational data from 100 to 243 container ships and found that cargo weight had the highest correlation with fuel consumption and that the ANN model had better prediction than multiple regression analysis. Panapakidis, I. et al. [28] used speed, wind power, the number of passengers, main engine hours, and distance information for passenger ships to develop a deep learning-based fuel consumption estimation model. Ahlgren, F. et al. [29] conducted a study on predicting dynamic fuel consumption with an automated machine learning algorithm to complement the installation of additional mass flow meters. However, if the ship's operational data are obtained through the AIS or weather systems, the data have a long cycle time and may differ from the ship's actual operating speed. Research that utilizes operational data obtained through sensors or systems on a ship to estimate fuel consumption complements previous studies.

Uyanık, T. et al. [30,31] proposed a fuel consumption estimation model using an artificial intelligence model by collecting information from measurement data on the ship's main engine, noon report, and engine logbook. Tarelko, W. and Rudzki, K. [32] developed a model for predicting a ship's fuel consumption using an artificial neural network algorithm. They employed input variables such as the rotation speed of the combustion engine, a controllable pitch propeller, marine conditions, and time since the last docking from the onboard sensors and decision models of the Barquentine. Kaklis, Dimitrios et al. [33,34] estimated fuel consumption using Automatic Identification System (AIS) data and onboard sensor data from a 3000 TEU container ship. Tran, Tien Anh [35] developed a fuel consumption estimation model with an ANN algorithm considering the factors of cargo load, diesel engine load, and engine operating status obtained from a Data Acquisition System (DAS) and validated it on a bulk carrier. The state-of-the-art studies used measured information from the ship but combined various sources such as the noon report and engine logbook to build the dataset. As a result, although the utilization of past voyage data is possible, obtaining real-time data is not feasible, and the voyage data cycle is extended due to the data management approach. In addition, it is difficult to apply to various ship types by utilizing variables that affect ship type such as cargo volume.

In this study, we aim to utilize onboard sensor data from smart ships to build the machine learning-based fuel consumption and carbon emission estimation model. By utilizing data directly measured on the vessel, the operator can obtain information closely related to actual navigation, and by analyzing the data in minutes, the accuracy of the prediction model can be improved by the real-time measurement. It is expected that it can

be used to estimate the actual fuel consumption and carbon emissions of ships based on the precise learning model based on the actual vessel voyage operation profile. The proposed model provides an accurate and tailor-made fuel consumption prediction model that fits the specific vessel and accumulates the operation data. As the smart ship has developed more, this approach has become more powerful than the conventional resistance-based fuel consumption estimation method.

### 1.3. Problem Definition

In this study, we propose a method to predict carbon emissions based on the fuel oil consumption of ships. It establishes a framework to analyze fuel consumption by utilizing operational data measured onboard a smart ship. There are three main contributions to this study.

- Actual operational data of dual-fuel propulsion vessels with explainable AI (XAI) were analyzed, and a suitable machine learning prediction model was designed.
- The fuel consumption predicted based on the vessel's operational data was compared with the fuel consumption predicted based on power.
- The prediction model was verified using mass flow meter data measured onboard a smart ship.

Firstly, we preprocessed the minute-by-minute navigation data collected from smart ships and analyzed it with XAI to derive feature importance. We compared the accuracy of machine learning methods according to the mode of operation by fuel and designed a fuel consumption prediction model with an LSTM algorithm, which can easily handle long sequences. Secondly, the fuel consumption was predicted based on the mass flow meter of the fuel and the power obtained from the actual operation data of a dual-fuel propulsion ship, and the two values were compared. In this study, we evaluated the value of the proposed fuel consumption prediction model by comparing the predicted values based on the mass flow meter of the fuel with the predicted values based on the power and analyzed their accuracy. Thirdly, mass flow meter data measured by onboard sensors were utilized to verify the fuel consumption of the actual voyage. By calculating fuel consumption and carbon dioxide emissions from actual operational data obtained from the smart vessel, we obtained data that closely approximate the actual consumption of the vessel. By forecasting consumption for current and future operations that closely approximates actual consumption, ships can plan routes that meet their carbon emissions targets. We expect to be able to respond to the IMO's strengthened carbon regulations by establishing a sailing plan in line with target carbon emissions.

The main contents of the study are as follows. Section 2 defines the methodology. Section 3 describes the experiments, and Section 4 discusses the study. Section 2 defines the dataset and compares machine learning methods to define a model to predict fuel consumption. We set up a detailed case study to validate the model and experiments utilizing onboard data from a smart ship, as described in Section 3. Section 4 discusses the fuel consumption prediction model and results.

## 2. Materials and Methods

### 2.1. Data

Fuel consumption is the outcome of the ship voyage operation data such as the ship speed, weather, and route. The novel method that will be proposed in this study is a data-driven fuel consumption and carbon emissions estimation method based on the onboard measurement data of a smart ship. Therefore, the measured onboard data collected from November 2022 to August 2023 is employed to build the proposed model. A smart ship can measure operational data such as the vessel's position, wind direction, wind speed, and water depth. Using these operational data, the authors tried to predict the fuel consumption data measured by the flow meter, which are regarded as the true target values. There are various propulsion modes, but for the purpose of this study, diesel and gas modes were selected as the target modes. The subject smart ship weighs 2700 tons and is 89 m long,

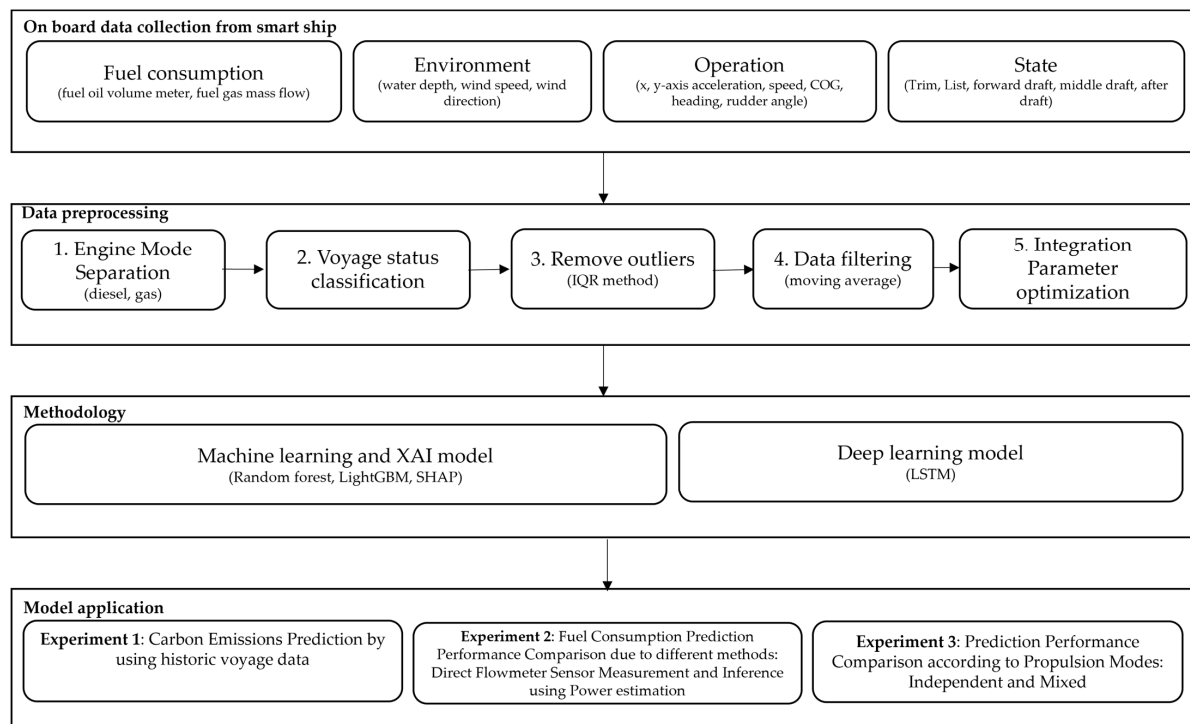
13 m wide, and 5.4 m high. It can operate at 14 knots and up to 16 knots to tour a near-shore area. Table 1 presents a detailed representation of the feature set collected for estimating fuel consumption in vessels, along with their respective abbreviations.

**Table 1.** The features of our dataset.

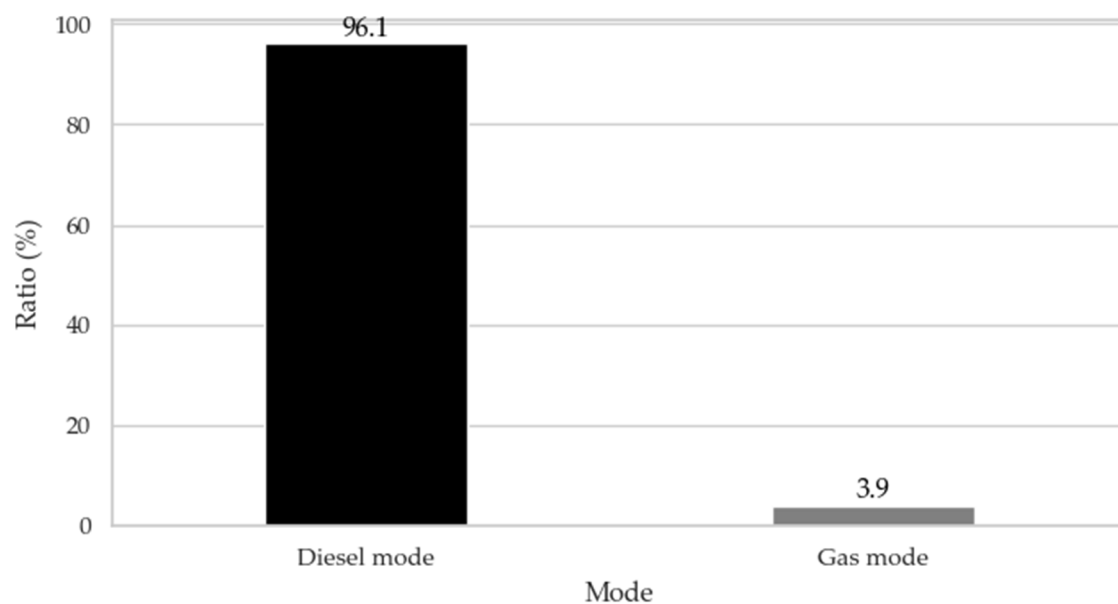
| Id | Feature                         | Abbreviation   | Minimum Value | Maximum Value | Measurement Unit   |
|----|---------------------------------|----------------|---------------|---------------|--------------------|
| 1  | Distance between waypoints      | Distance_point | 0.0           | 0.283         | NM (Nautical mile) |
| 2  | X-axis acceleration             | ACCELER_X      | −93.0         | 117.0         | m/s <sup>2</sup>   |
| 3  | Y-axis acceleration             | ACCELER_Y      | −19.0         | 3.0           | m/s <sup>2</sup>   |
| 4  | Forward draft                   | ForwardDraft   | 1.6           | 4.1           | m                  |
| 5  | Course of ground                | COG            | 0.5           | 360.0         | ° (degree)         |
| 6  | After draft                     | AfterDraft     | 2.2           | 3.8           | m                  |
| 7  | Heading                         | Heading        | 0             | 360           | ° (degree)         |
| 8  | List tilt                       | List           | −5.5          | 6.5           | ° (degree)         |
| 9  | Log of speed through water      | LogSpeed       | 1.0           | 17.0          | Kn (knot)          |
| 10 | Mid draft on the port side      | MidPDraft      | 2.4           | 3.7           | m                  |
| 11 | Mid draft on the starboard side | MidSDraft      | 2.4           | 3.6           | m                  |
| 12 | Rudder Angle                    | RudderAngle    | −35.0         | 35.0          | ° (degree)         |
| 13 | Trim tilt                       | Trim_Tilt1     | −0.1          | 0.8           | ° (degree)         |
| 14 | Water depth                     | WaterDepth     | 0.0           | 132.5         | m                  |
| 15 | Wind direction                  | WindDirection  | 0.0           | 360.0         | ° (degree)         |
| 16 | Wind speed                      | WindSpeed      | 0.0           | 27.4          | m/s                |
| 17 | Fuel oil consumption            | FOC            | 0.0           | 735.8         | kg/h               |
| 18 | Fuel gas consumption            | FGC            | 0.0           | 219           | kg/h               |

### 2.1.1. Engine Mode Separation: Diesel and Gas

Figure 1 includes the overall preprocessing process. The time series data were then divided into diesel mode and gas mode. Subsequently, the analysis matched the FOVolumeMeter and FGMassFlow data, representing the fuel flow for each mode. Figure 2 shows, diesel mode was 96.1% and gas mode was 3.9% of the whole voyage period. The target vessel used MDO (Marine Diesel Oil) in diesel mode and LNG (liquefied natural gas) in gas mode as fuel for propulsion.



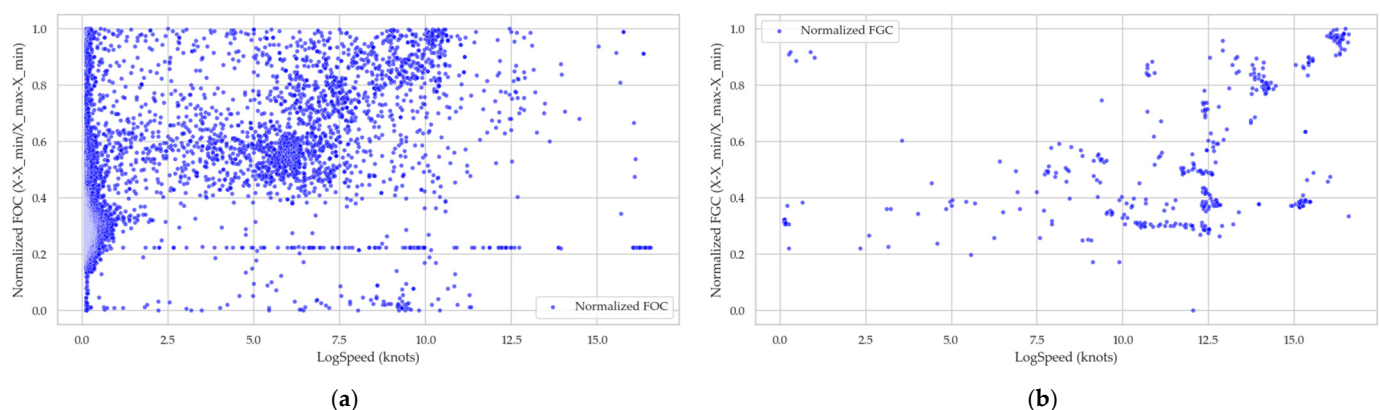
**Figure 1.** Schematic diagram of research.



**Figure 2.** Ratio of modes.

### 2.1.2. Voyage Status Classification: Berthing and Voyage

Data from ships at berthing are not suitable for predicting fuel consumption and thus must be removed. This study utilized speed data to classify the data into berthing and voyage modes. Figure 3 shows the correlation between speed and fuel consumption in diesel and gas modes. Operations at speeds of 10 knots or below used diesel mode, while those exceeding 10 knots used gas mode. Therefore, it is necessary to apply different filtering speeds for diesel and gas modes. To preserve more low-speed operational data, the filtering speed for diesel mode was set to 1 knot. For gas mode, to focus on relatively high-speed operational data, the filtering speed was set to 5 knots.

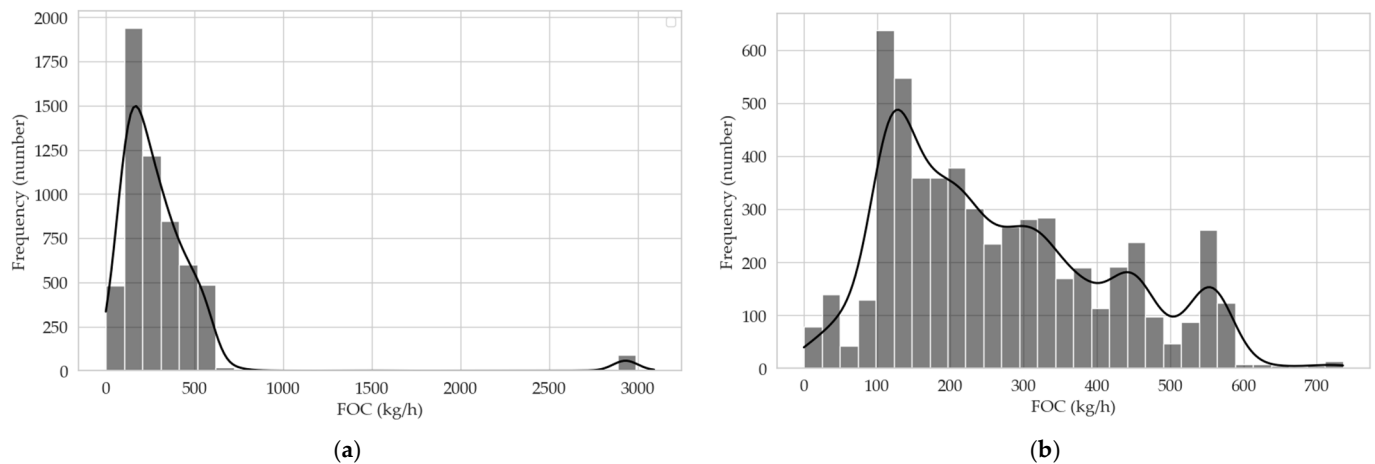


**Figure 3.** Distribution plots. (a) FOC (fuel oil consumption) based on speed; (b) FGC (fuel gas consumption) based on speed.

### 2.1.3. Noise Removal: Interquartile Range (IQR)

The raw data collected from the ship's sensors, presented in a time series (minute-by-minute) format, tend to be noisy and contain errors. As observed in (a) of Figure 4, the data for this vessel contain a small number of outliers. While the values of other features are similar, there are instances where only the ship's fuel consumption values differ significantly. The IQR method employed by Chen [36] was adopted to filter out such data. This method is useful when data do not follow a normal distribution or have a skewed distribution. It is less sensitive to outliers than the mean and standard deviation and is particularly

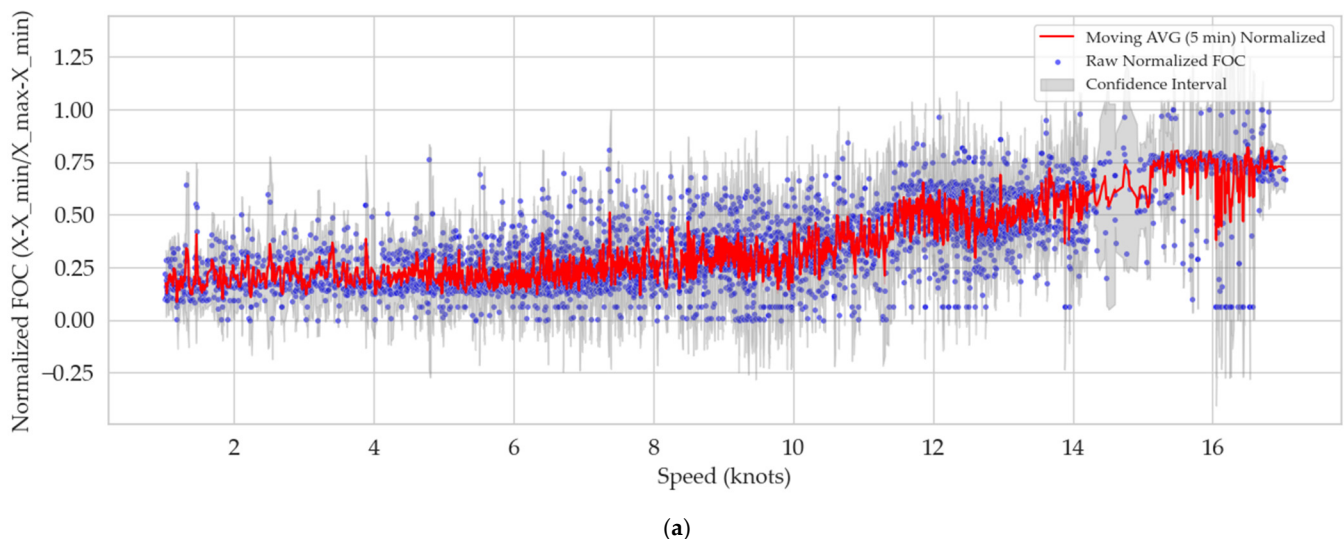
effective at detecting outliers in skewed data distributions. The data used in this study show such a distribution, so this method is appropriate. In Figure 4, (a) is before applying the IQR method and (b) is after. The histogram illustrates the frequency distribution of fuel consumption values, while the overlaid black line represents a Kernel Density Estimate (KDE) providing a smooth approximation of the data's probability distribution.



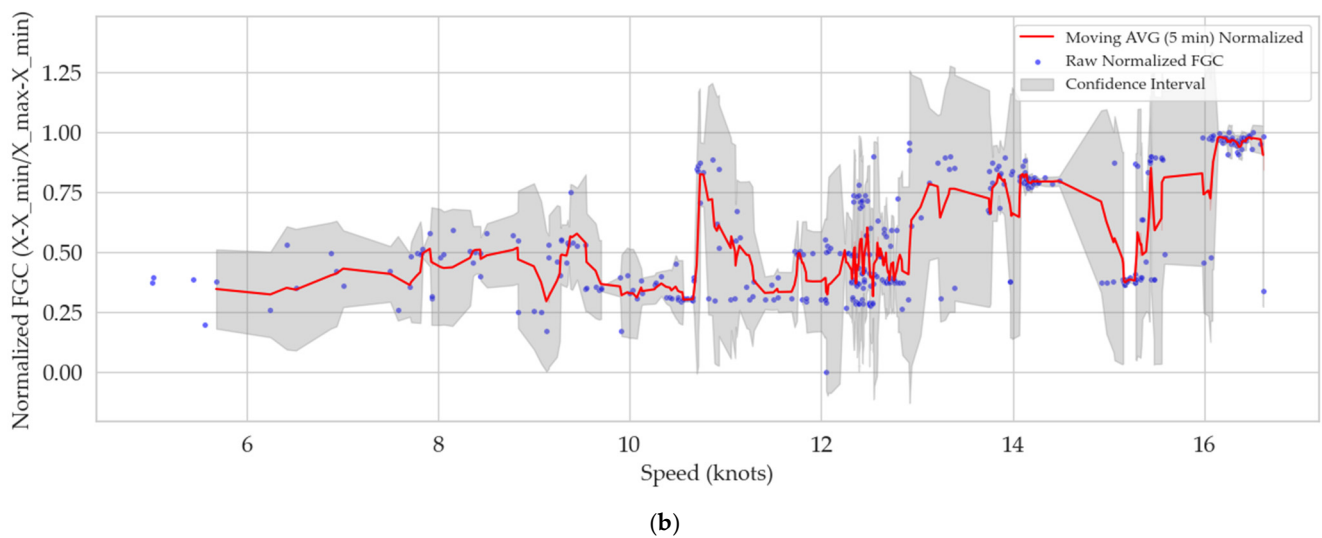
**Figure 4.** Distribution plots. (a) Before FOC noise removal; (b) After FOC noise removal.

#### 2.1.4. Moving Average for Data Filtering

The Simple Moving Average (SMA) is a technique used to smooth out trends in time series data by calculating the average value over a specific period. Although outliers in fuel consumption data were removed using the IQR method, the data still exhibit spikes reminiscent of sparks. By applying a moving average, spikes and anomalies caused by the ship's flow meter sensor can be more smoothly addressed [33]. The data were transformed into a 5 min rolling window average, allowing fuel consumption predictions to be based on average consumption rather than instantaneous use. Figure 5 presents the raw fuel consumption data collected from the sensor alongside the 5 min rolling window average.



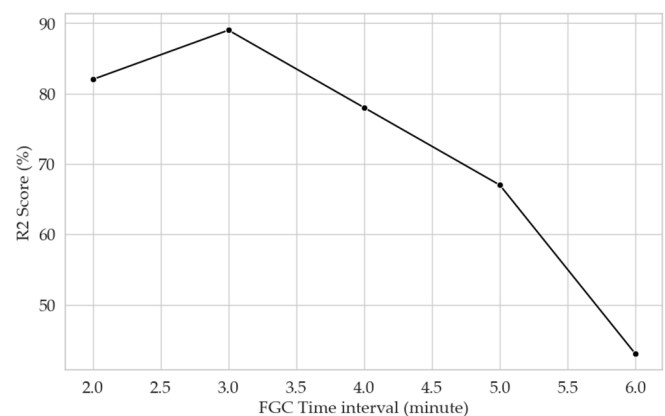
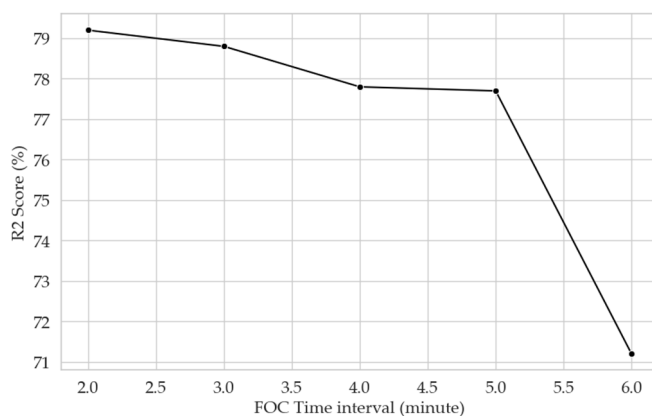
**Figure 5.** Cont.



**Figure 5.** Graph showing the moving average and normalized fuel consumption according to speed. (a) FOC; (b) FGC.

#### 2.1.5. Integration Parameter (Time Interval) Convergence Test

The prediction of a ship's fuel consumption utilizes obtained flow meter data. The form of the fuel flow meter is mass flow, so the integration of mass flow is essential for fuel consumption prediction. Therefore, an optimal time interval for accurate prediction and performing numerical integration was identified using the convergence test. Consequently, the focus was on determining the time interval that yields the highest R2 score in predictions made using a random forest regressor, thereby enhancing the precision of fuel consumption forecasting using an incremental test from 2 min to 6 min. Figure 6 shows fuel oil consumption (FOC) achieved its best performance at a 2 min interval, and fuel gas consumption (FGC) reached its peak performance at a 3 min interval.



**Figure 6.** R2 score plots over time intervals. (a) FOC graph; (b) FGC graph.

#### 2.2. Methodology

This section explains the methodology of the proposed fuel and carbon prediction method under voyage conditions using onboard measurement data. First, the background theory achieved high performance due to the selection of the random forest method for fuel consumption models. Ref. [23] utilized the random forest method to estimate fuel consumption using data. The random forest approach has demonstrated its strength as a powerful ensemble model across various studies. Therefore, this study also employed

the random forest method. LightGBM was chosen from among the ensemble models for comparison with the random forest method. Additionally, LSTM, previously used in prior research for fuel consumption prediction models, was selected for performance comparison.

This study divided the data into 70% for training and 30% for testing. The diesel mode data used for learning include 5600 rows and 18 columns, while the gas mode comprises 395 rows and 18 columns. The test data were randomly selected. Hyperparameters were optimized through grid search and experimentation.

### 2.2.1. Random Forest Regressor (RFR)

The random forest regressor employs an ensemble of decision trees. The algorithm integrates two main ideas: the bagging method and the random subspace methodology. The main idea is to combine predictions from multiple models to produce more accurate results than a single model. The random subspace methodology reduces correlation between trees and avoids overfitting. Basic algorithms are trained using various subsets, which are randomly allocated.

$$\hat{y} = \frac{1}{N} \sum_{i=1}^N t_i(x) \quad (1)$$

where  $\hat{y}$  is the final predicted output,  $N$  is the number of decision trees, and  $t_i(x)$  is the predicted output for  $x$  in the  $i$  – th decision tree. Table 2 shows the main hyper parameters of the RFR optimized by grid search.

**Table 2.** Main RFR hyperparameters.

| Hyperparameters   | Value                         |
|-------------------|-------------------------------|
| N_estimators      | 100 (Diesel), (Gas)           |
| Criterion         | Squared error (Diesel), (Gas) |
| Max_depth         | 40 (Diesel), (Gas)            |
| min_samples_split | 2 (Diesel)                    |
| min_samples_leaf  | 4 (Diesel)                    |
| max_features      | None                          |

### 2.2.2. LightGBM Regressor

LightGBM stands out as an efficient and distributed decision tree-based gradient boosting algorithm that utilizes a histogram approach, which contributes to reduced memory usage and computational demands. The key difference from other methods is that it splits the tree based on leaves, that is, it can detect key points and stop calculations (other lifting algorithms are depth-based or level-based). This leaf-focused strategy generally leads to a reduction in loss and increased predictive precision when a leaf node expands. LightGBM adds a maximum depth limit leaf-wise, ensuring high efficiency while preventing overfitting.

For each data point  $i$ , the gradient (first derivative)  $g_i$  and the Hessian (second derivative)  $h_i$  of the loss function  $L$  are computed.

$$g_i = \frac{\partial L(y_i, \hat{y}_i)}{\partial \hat{y}_i} \quad (2)$$

$$h_i = \frac{\partial^2 L(y_i, \hat{y}_i)}{\partial \hat{y}_i^2} \quad (3)$$

where  $y_i$  is the actual value, and  $\hat{y}_i$  is the predicted value to date by the model. At each iteration, update the model by building a new tree  $f_t$ . The tree uses gradient and Hessian information to find the optimal split.

$$\hat{y}_i^{(t)} = \hat{y}_i^{(t-1)} + \eta f_t(x_i) \quad (4)$$

where  $f_t(x_i)$  is the predictive contribution by the new tree, and  $\eta$  is the learning rate. The new tree  $f_t$  is added to the updated model, which means that each tree contributes to minimizing the loss function. Table 3 shows the main hyperparameters of the LightGBM optimized by grid search.

**Table 3.** LightGBM main hyperparameters.

| Hyperparameters   | Value                |
|-------------------|----------------------|
| N_estimators      | 100 (Diesel), (Gas)  |
| Learning_rate     | 0.05 (Diesel), (Gas) |
| Max_depth         | 10 (Diesel), (Gas)   |
| Min_child_samples | 30 (Diesel)          |
| Num_leaves        | 31 (Diesel)          |

### 2.2.3. Feature Selection

In the process of predicting a ship's fuel consumption, certain features have the potential to interfere with learning. By removing these features, the model is able to more easily learn the essential patterns and relationships present in the data. The goal was to identify the combination of 2 to 18 columns that results in the highest R2 score, utilizing the RFE (recursive feature elimination) library from Scikit-learn. The code employed for this analysis is presented in Algorithm 1.

**Algorithm 1** Feature selection using RFE with Model M

**Require:** Training data  $X_{train}$ , Test data  $X_{test}$ , Target labels  $Y_{train}$ , Target labels  $Y_{test}$

**Ensure:** Best feature subset for model M

```

1: Initialize  $max\_value$  to  $-\infty$ 
2: for  $i$  in range(2, 18) do
3:   Create a new instance of Model M
4:    $selector \leftarrow RFE(M, n\_features\_to\_select = i, step = 1)$ 
5:   Fit  $selector$  on  $X_{train}$ ,  $Y_{train}$ 
6:    $X_{train\_sel} \leftarrow X_{train}$  with columns selected by  $selector$ 
7:    $X_{test\_sel} \leftarrow X_{test}$  with columns selected by  $selector$ 
8:   Re-train Model M on  $X_{train\_sel}$ ,  $Y_{train}$ 
9:    $Y_{pred} \leftarrow$  Predict using M on  $X_{test\_sel}$ 
10:   $value \leftarrow R2\_score(Y_{test}, Y_{pred})$ 
11:  if  $value > max\_value$  then
12:     $max\_value \leftarrow value$ 
13:     $X_{train\_final} \leftarrow X_{train\_sel}$ 
14:     $X_{test\_final} \leftarrow X_{test\_sel}$ 
15:     $mem \leftarrow i$ 
16:     $best\_M \leftarrow M$ 
17:  end if
18: end for
19: print "Maximum R2 Score: ",  $max\_value$ , "with features count: ",  $mem$ 

```

Table 4 shows that, in diesel mode, the application of feature selection enhanced the performance of the random forest method by 0.4% and the LightGBM method by 0.3%. In gas mode, the improvements were slightly greater, with random forest showing a 2.6% increase in performance and LightGBM exhibiting a 1.9% improvement.

**Table 4.** Comparison of R2 scores before and after using feature selection of ML model.

| Models           | R2 Score (%)<br>before Feature Selection | R2 Score (%)<br>after Feature Selection |
|------------------|------------------------------------------|-----------------------------------------|
| RFR for FOC      | 92.4                                     | 92.6                                    |
| LightGBM for FOC | 92.1                                     | 92.4                                    |
| RFR for FGC      | 90.9                                     | 93.5                                    |
| LightGBM for FGC | 90.1                                     | 92.0                                    |

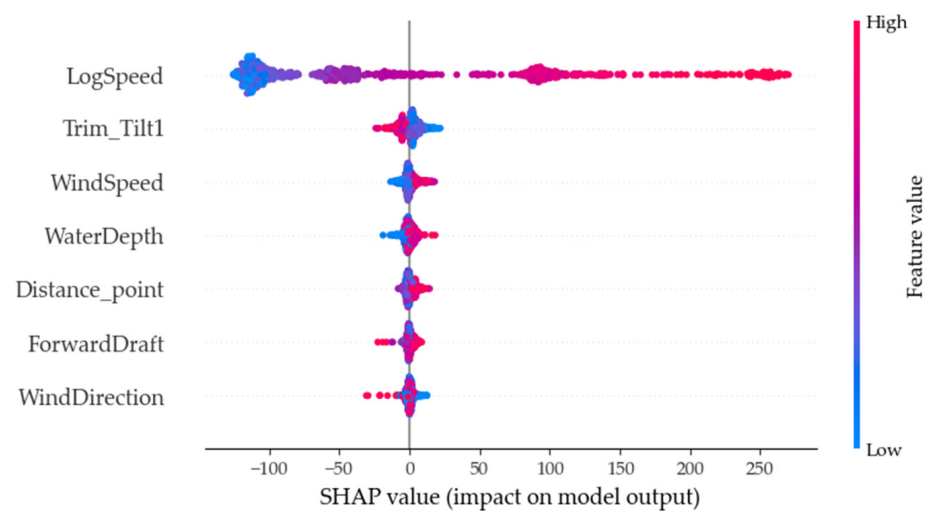
#### 2.2.4. Feature Importance Using SHAP (Shapley Additive Explanations)

SHAP serves as a method to describe a model's predictions, relying on Shapley values from game theory. This approach quantitatively evaluates the impact of each input characteristic on a machine learning model's prediction. Originating from cooperative game theory, Shapley values offer a method to distribute 'payoffs'—the game's rewards—fairly among multiple collaborating players. In the context of machine learning, these 'players' represent the attributes or features, and the 'payoffs' correspond to the 'predicted outcomes'.

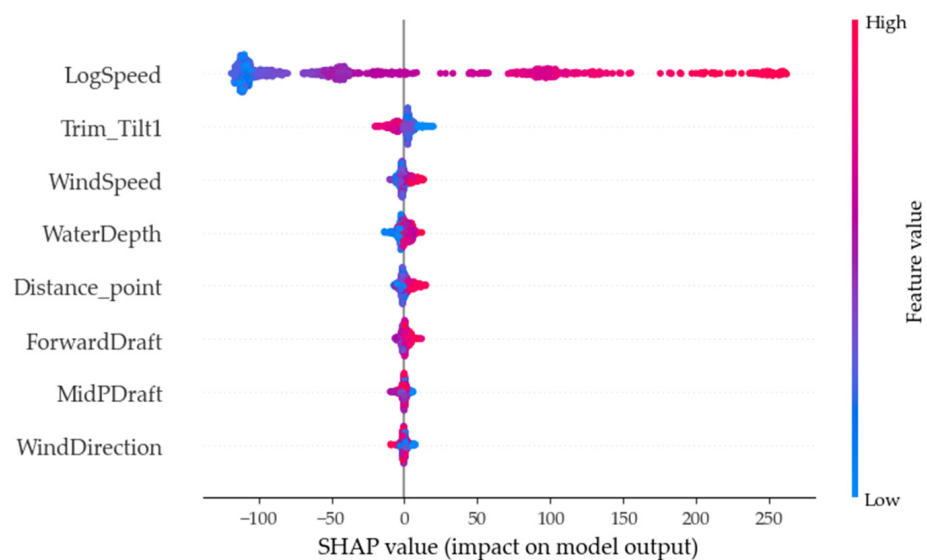
$$\phi_i = \sum_{S \subseteq N \setminus \{i\}} \frac{|S|!(N-|S|-1)!}{N!} [f(S \cup \{i\}) - f(S)] \quad (5)$$

where  $\phi_i$  is the SHAP value for attribute  $i$ ,  $N$  is the set of all input attributes, and  $S$  represents the subset of  $N$  except attribute  $i$ .  $f(S)$  and  $f(S \cup \{i\})$  are the model's predictions.

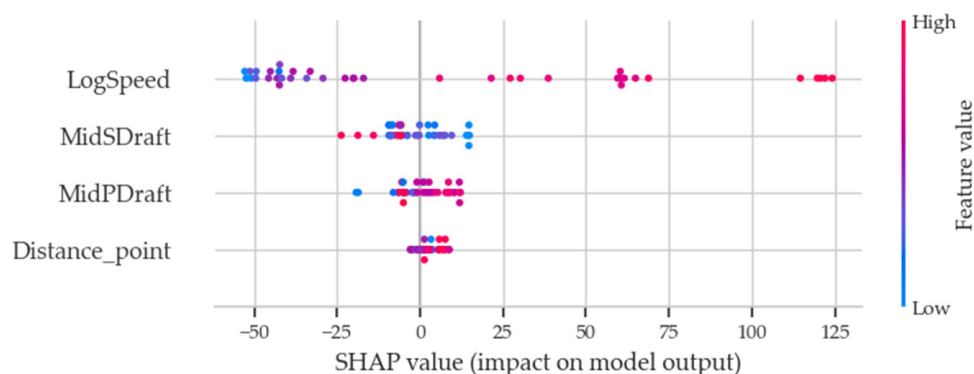
According to Figures 7–10, the analysis focuses on the correlation between the vessel's fuel consumption and various operational data points. The LogSpeed value emerged prominently as the most significantly correlated factor, primarily due to its direct role as a measure of fuel consumption.



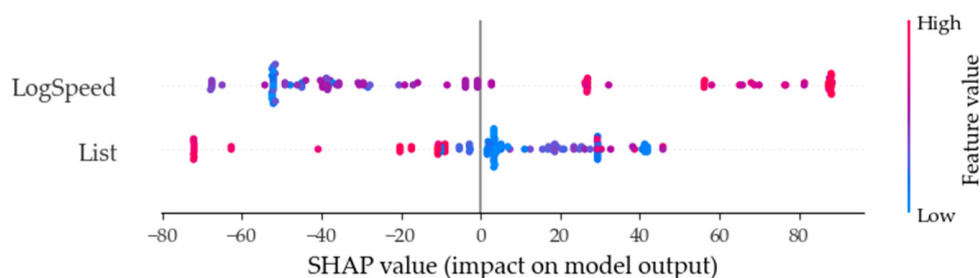
**Figure 7.** Feature importance of RFR (diesel).



**Figure 8.** Feature importance of LightGBM (diesel).



**Figure 9.** Feature importance of RFR (gas).



**Figure 10.** Feature importance of LightGBM (gas).

Further into the analysis, in diesel mode, the most significant predictor of fuel consumption following speed is the trim of the vessel. The trim refers to the horizontal state of a ship, denoting the difference in height between the bow and the stern. The trim significantly influences the vessel's stability, fuel efficiency, speed, and its ability to maintain course during navigation. Proper trim adjustment can reduce the resistance the ship faces when moving through water, thereby increasing its speed. Additionally, the method of cargo loading significantly impacts the trim, making efficient cargo stowage crucial. An appropriate trim setting, deeply intertwined with fuel consumption, enhances the vessel's stability and fuel efficiency.

Environmental influences such as wind speed, wind direction, and water depth are also important features. This refers to the actual value as data measured directly from the ship. The data are optimized for the ship's situation. For ships primarily navigating coastal regions and short distances, water depth is the most variable environmental factor. It is imperative to navigate towards areas where sufficient water depth is assured to avoid groundings and collisions. This necessitates a careful consideration of water depth in operational planning, especially for vessels operating in such variable coastal environments.

In gas mode, the draft difference between the starboard and port sides is a critical feature, indicative of the vessel's list, which refers to the state of the ship being tilted laterally. Like the trim, the list can increase resistance, so it is best to load the cargo in a balanced manner and adjust the draft appropriately to avoid creating extreme list.

#### 2.2.5. Long Short-Term Memory (LSTM)

A recurrent neural network (RNN) is a type of neural network architecture for processing time series data. It analyzes the relationship between the preceding and present data and estimates the succeeding data. However, the RNN does not work well in long-term dependent problems and past information is lost. LSTM consists of cell states, forgetting gates, input gates, and output gates and uses cell states to improve on the limitations of RNNs.

## 1. Forget gate

$$f_t = \sigma(W_f \cdot [h_{t-1}, x_t] + b_f) \quad (6)$$

where  $f_t$  is the output of the forgetting gate that decides what information to discard at the current time step, and  $W_f$  and  $b_f$  are the weights and biases of the forgetting gate.  $\sigma$  is a sigmoidal function.

## 2. Input gate

$$i_t = \sigma(W_i \cdot [h_{t-1}, x_t] + b_i) \quad (7)$$

$$\tilde{C}_t = \tanh(W_c \cdot [h_{t-1}, x_t] + b_c) \quad (8)$$

where  $i_t$  determines how much new information to add to the cell state, and  $\tilde{C}_t$  is the new candidate value.

## 3. Cell state update

$$C_t = f_t \times C_{t-1} + i_t \times \tilde{C}_t \quad (9)$$

The previous cell state  $C_{t-1}$  loses some information through the forgetting gate, and new information  $\tilde{C}_t$  is added through the input gate.

## 4. Output gate and hidden state

$$o_t = \sigma(W_o \cdot [h_{t-1}, x_t] + b_o) \quad (10)$$

$$h_t = o_t \times \tanh(C_t) \quad (11)$$

The output gate determines the output for this time step, from which the hidden state  $h_t$  is calculated. In an LSTM regression model, this hidden state  $h_t$  is passed to the final output layer to predict successive target values. When using a linear output layer, the prediction is calculated as follows.

$$\hat{y}_t = W_y h_t + b_y \quad (12)$$

where  $\hat{y}_t$  is the predicted value at time step  $t$ , and  $W_y$  and  $b_y$  are the weights and biases of the output layer. LSTM regression models use this structure to learn the long-term dependencies and complex patterns in time series data and use them to predict future values or sequences. Figure 11 shows the structure of LSTM.

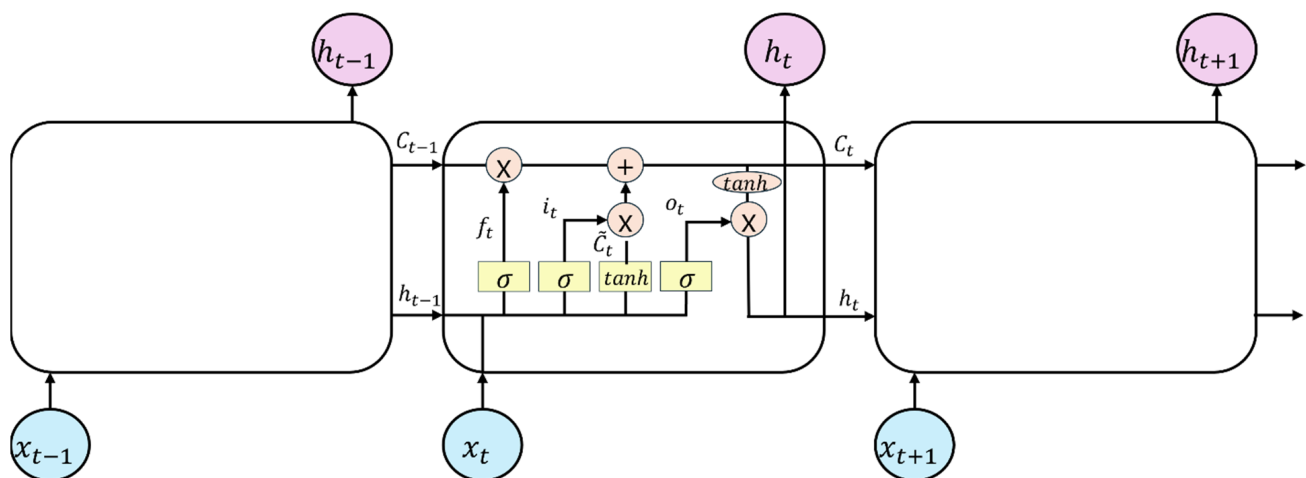


Figure 11. Structural form diagram of the LSTM.

Various factors, such as weather conditions, ocean currents, and ship speed, influence a vessel's fuel consumption, creating temporal patterns. LSTM effectively captures these temporal characteristics in time series data, enabling learning from past data to predict future fuel consumption. Table 5 presents the hyperparameters that optimize the LSTM model's prediction of vessel fuel consumption in each mode. Among them, sequence length stands out as the most crucial hyperparameter. It defines the number of consecutive data points the model processes as input, thereby determining the extent of past information the model can utilize. While longer sequence lengths can encompass older information, they also increase computational complexity and the risk of overfitting. Diesel mode, with its larger data availability and longer sequence length, showed superior performance compared to gas mode. Specifically, in diesel mode, the sequence length was 12, which means that 24 min of historical data went into the input, considering that each data point was 2 min apart. In gas mode, the sequence length was 1, which means that 3 min of historical data was input, considering that each data point was 3 min apart. On average, every 8 min, a change in the ship's operational data affected fuel consumption.

**Table 5.** LSTM model main hyperparameters.

| Hyperparameters | Value                 |
|-----------------|-----------------------|
| Sequence length | 12 (Diesel), 1 (Gas)  |
| Batch size      | 64 (Diesel), 32 (Gas) |
| Hidden layer    | 10 (Diesel), 1 (Gas)  |
| Optimizer       | Adam                  |
| Learning rate   | 0.1 (Diesel, Gas)     |

The length of the sequences was deliberately optimized to be relatively short, taking into consideration the operational patterns of the vessels, which predominantly navigate within the coastal vicinity of Ulsan Port. As smart ships are equipped with the capability to measure data in real-time, it is anticipated that the precision of the training model will be enhanced through periodic updates, considering the determined sequence length. The observed superior performance of the diesel mode compared to the gas mode is likely attributable to the greater volume of available data. This abundance of data renders the deep learning model, particularly the LSTM architecture, more effective due to its ability to learn from larger datasets.

Table 6 shows the performance of RFR, LighGBM, and LSTM in each mode. In diesel mode, RFR and LightGBM perform better than LSTM. In gas mode, RFR performed significantly better, but LSTM had the lowest RMSE. Overall, RFR outperformed in both modes. In the context of operational data utilized for fuel consumption prediction, it was found that the performance of Random Forest Regression (RFR), adept at capturing complex non-linear relationships, was superior to that of LSTM, which is significant for time-sequential data, and LightGBM, which is suitable for processing large volumes of data. Unearthing the intricate relationships and patterns within the data is of paramount importance in this domain.

**Table 6.** Comparison of model performances.

| Models           | R2 Score (%) | MAPE (%) | RMSE (Value) |
|------------------|--------------|----------|--------------|
| LSTM for FOC     | 84.8         | 13.8     | 44.1         |
| RFR for FOC      | 92.7         | 11.5     | 32.9         |
| LightGBM for FOC | 92.4         | 11.5     | 33.5         |
| LSTM for FGC     | 75.1         | 81.9     | 14.7         |
| RFR for FGC      | 93.5         | 5.6      | 19.0         |
| LightGBM for FGC | 90.0         | 7.1      | 23.6         |

### 3. Results

#### 3.1. Experiment 1: Calculating Carbon Emissions (CO<sub>2</sub>) from Specific Routes

This experiment assessed the carbon emission prediction performance of the proposed method and compared the real measurement data under voyage conditions near the Ulsan area. This experiment was designed to measure the accuracy of the proposed method.

In diesel mode, data for each case were excluded from the training process. However, in gas mode, due to the limited amount of data, information pertaining to each case was also included in the training. When the operator sets the operating conditions such as ship speed and terrain information, a route that minimizes carbon emissions or provides an estimate of carbon emissions can be recommended. The International Council on Clean Transportation provides the carbon emission factor for different fuels [37]. MGO powers the ship in diesel mode, while LNG serves as the fuel in gas mode. Conversion to the appropriate units for the emission factors was carried out.

$$CO_2 = FOC \times 3.206(g\ CO_2/g\ fuel) \quad (13)$$

$$CO_2 = FGC \times 2.750(g\ CO_2/g\ fuel) \quad (14)$$

This study employed normalization to prevent the disclosure of the actual ship data values like carbon emissions due to security considerations.

$$X_{normalized} = \frac{X - X_{min}}{X_{max} - X_{min}} \quad (15)$$

where  $X$  is the original data value,  $X_{min}$  is the minimum value, and  $X_{max}$  is the maximum value.  $X_{normalized}$  is the normalized value, which lies between 0 and 1.

##### 3.1.1. Case 1: Operating Case with Mixed Diesel and Gas Modes

Figure 12 shows the first case, the route is from Ulsan to Jeju. The beginning and end are operated in diesel mode, and the middle is operated in gas mode. In Figure 12, yellow circles are diesel mode and red triangles are gas mode.

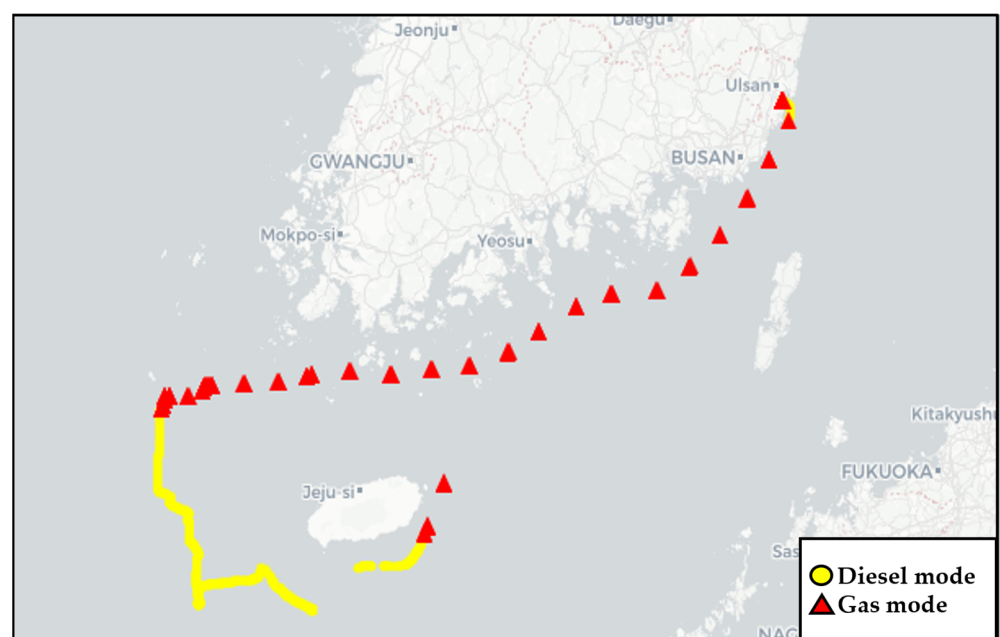
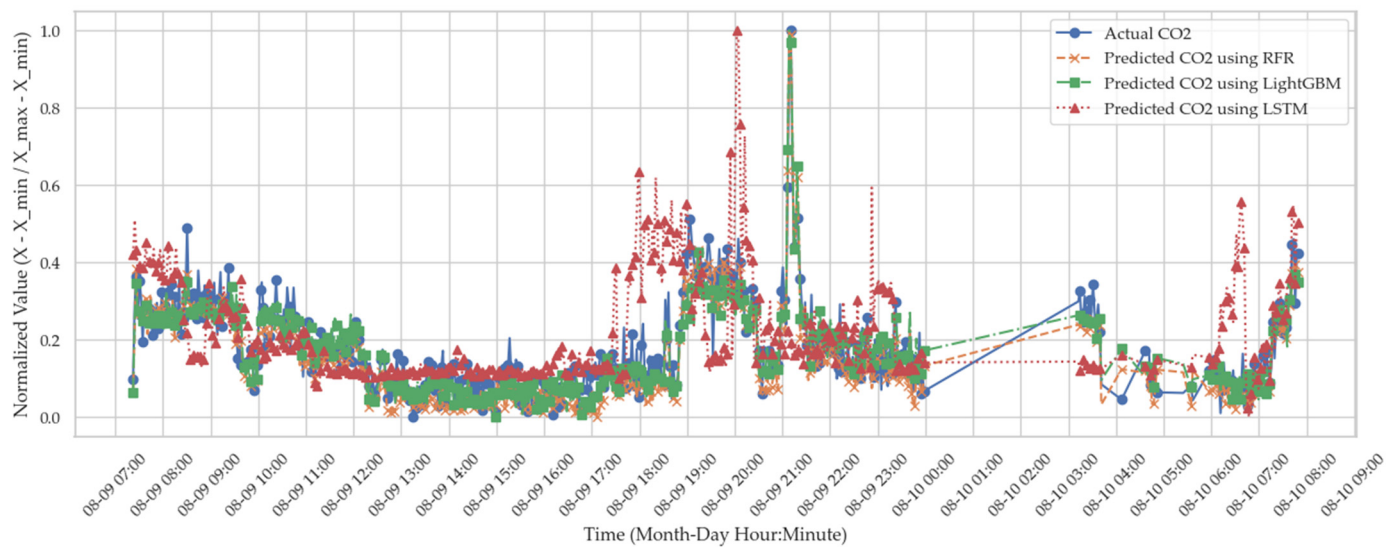


Figure 12. Case 1 route.

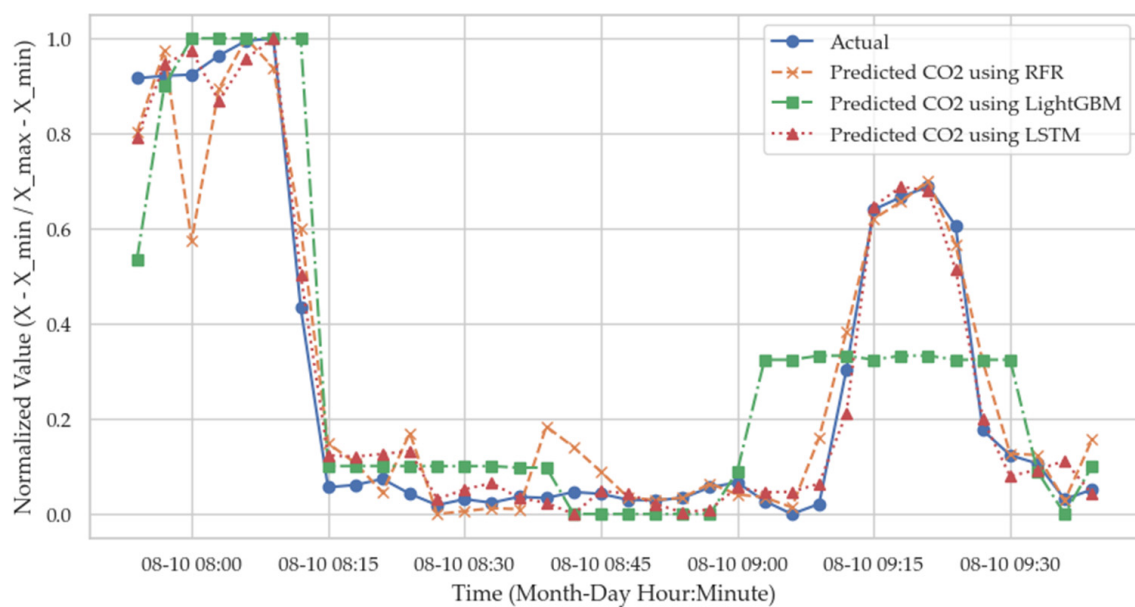
Figure 13 shows the CO<sub>2</sub> prediction results in diesel mode. The model that excels across all three evaluation metrics is the RFR. There are about four instances of sharp

transitions, where the ability to accurately predict these trends is a critical factor. Both RFR and LightGBM generally predicted well; however, LSTM did not perform as effectively. LSTM was notably less proficient during sections of rapid trend changes.



**Figure 13.** Comparison of actual and predicted CO<sub>2</sub> emissions in diesel mode for Case 1.

Figure 14 shows, in gas mode, based on the R2 score and RMSE, RFR exhibited good performance. When considering MAPE, LSTM showed superior performance. There are approximately two instances of rapid changes in carbon emissions. LightGBM struggled to predict these sections accurately.



**Figure 14.** Comparison of actual and predicted CO<sub>2</sub> emissions in gas mode for Case 1.

According to Table 7, the RFR demonstrated outstanding performance across all modes. It proved to be a robust model, even on routes with significant fluctuations.

**Table 7.** Performance metrics results for Case 1.

| Models           | R2 Score (%) | MAPE (%) | RMSE (Value) |
|------------------|--------------|----------|--------------|
| RFR for FOC      | 86.4         | 9.6      | 22.0         |
| LightGBM for FOC | 80.5         | 11.7     | 26.4         |
| LSTM for FOC     | 66.2         | 10.8     | 29.0         |
| RFR for FGC      | 84.3         | 12.6     | 21.0         |
| LightGBM for FGC | 64.3         | 21.7     | 31.8         |
| LSTM for FGC     | 75.3         | 8.7      | 35.7         |

### 3.1.2. Case 2: Operating Case with Gas Mode

In Figure 15, the route is from Ulsan to Busan, primarily operating in gas mode.

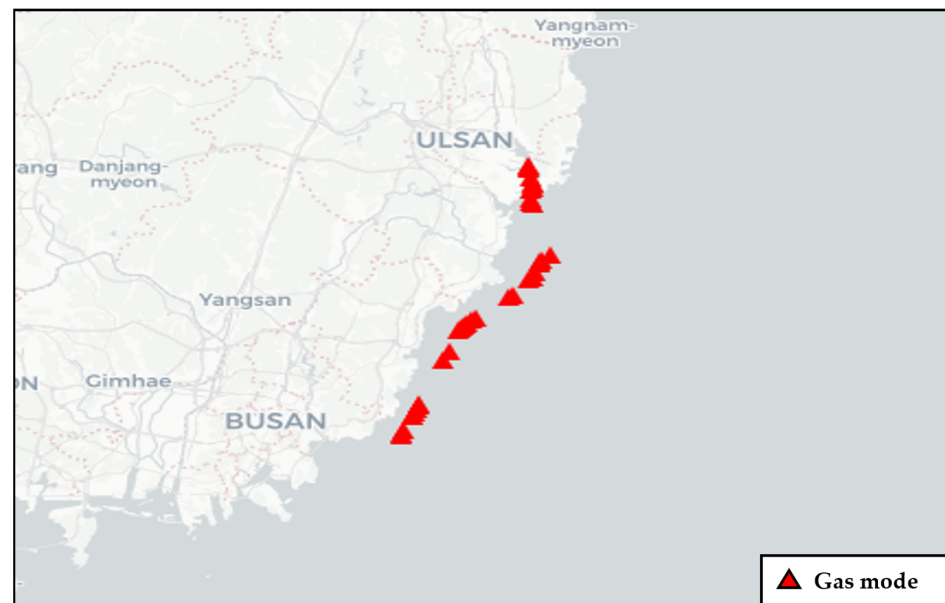
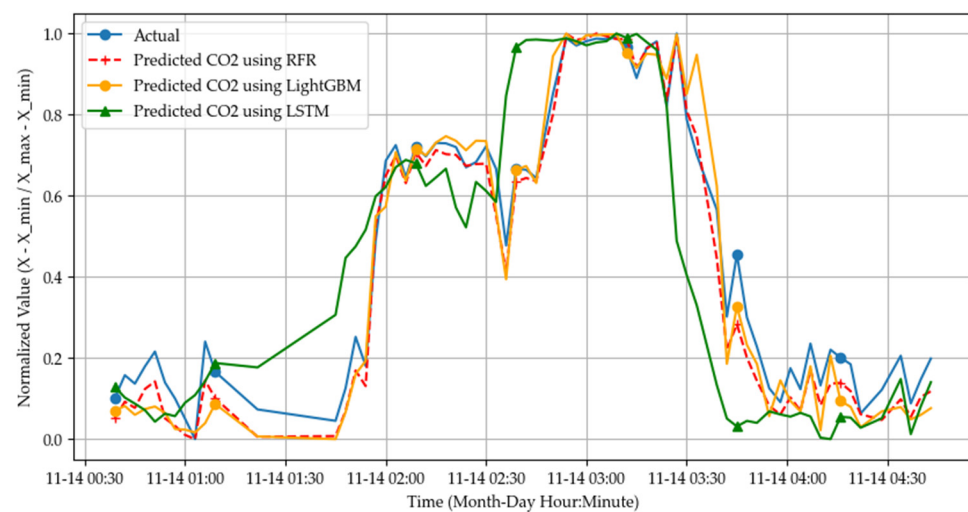
**Figure 15.** Case 2 route.

Figure 16 shows, there were no sharp changes, but rather gradual increases and decreases. According to Table 8, RFR performed best in all evaluation indicators.

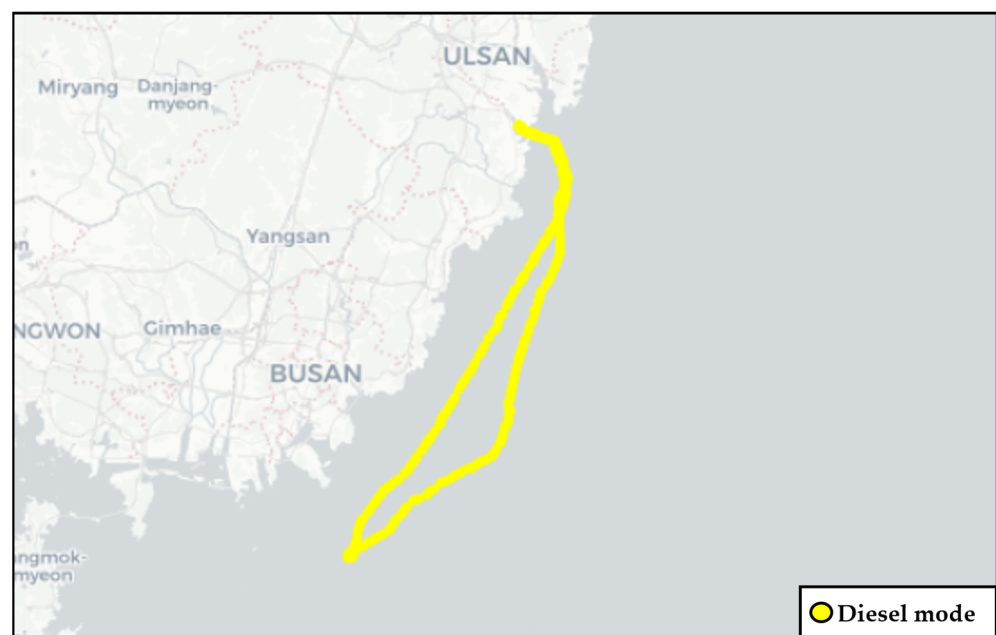
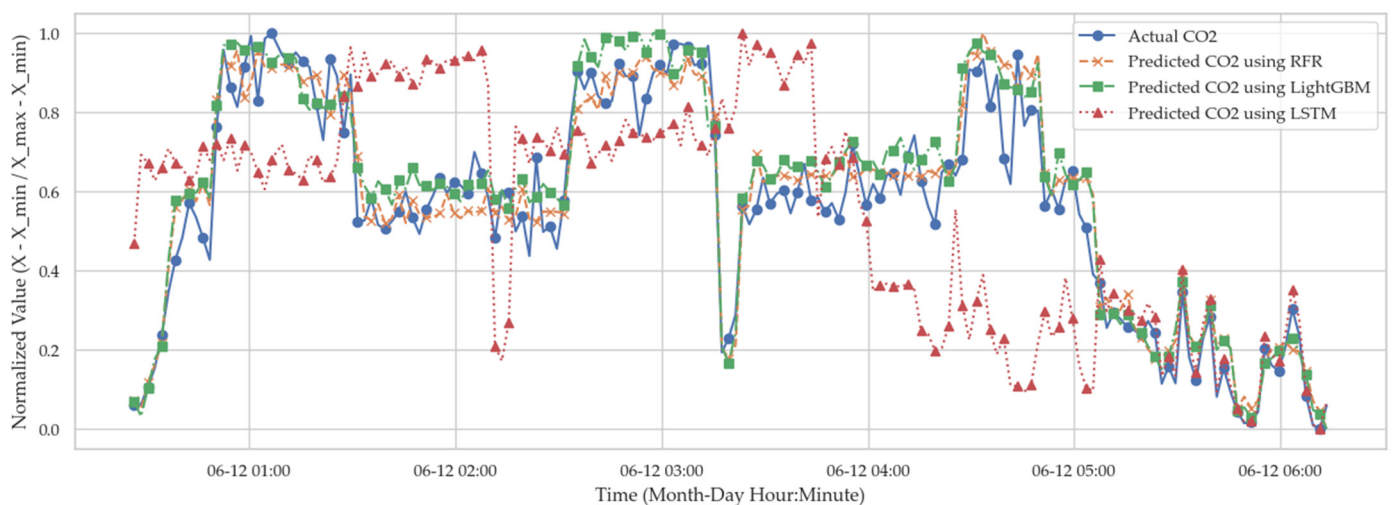
**Figure 16.** Comparison of actual and predicted CO<sub>2</sub> emissions in gas mode for Case 2.

**Table 8.** Performance metrics results for Case 2.

| Models   | R2 Score (%) | MAPE (%) | RMSE (Value) |
|----------|--------------|----------|--------------|
| RFR      | 99.0         | 1.9      | 7.6          |
| LightGBM | 98.4         | 2.7      | 9.8          |
| LSTM     | 87.6         | 6.8      | 26.8         |

### 3.1.3. Case 3: Operating Case with Diesel Mode

In Figure 17, the third case involves a voyage from Ulsan to Busan, like the second, but entails a longer round trip and operates solely in diesel mode. In Figure 18, sections exhibiting rapid changes were periodically observed three times. According to Table 9, LightGBM outperformed other methodologies in three evaluation metrics. RFR followed with the next best performance. Compared to the previous two methodologies, LSTM was less effective in predicting sections with changes.

**Figure 17.** Case 3 route.**Figure 18.** Comparison of actual and predicted CO<sub>2</sub> emissions in diesel mode for Case 3.

**Table 9.** Performance metrics results for Case 3.

| Models   | R2 Score (%) | MAPE (%) | RMSE (Value) |
|----------|--------------|----------|--------------|
| RFR      | 90.0         | 7.8      | 30.8         |
| LightGBM | 91.8         | 6.8      | 34.0         |
| LSTM     | 83.0         | 11.0     | 50.2         |

### 3.2. Experiment 2: Estimation of Fuel Consumption Using Flow Meter Data and Power Data

The aim of this experiment was to compare the CO<sub>2</sub> and fuel consumption prediction when the direct fuel flow meter data are recorded, or the fuel quantity is obtained from the power load data. The calculation of FOC and FGC, utilizing power load data, employed the SFOC (specific fuel oil consumption) curve from a ship with similar specifications to the vessel under study. Yoon's [38] proposed model identifies ships with similar specifications. According to Tables 10 and 11, in terms of prediction accuracy for the vessel's fuel consumption, flow meter data outperformed other methods. As a direct measurement obtained using a flow meter, flow meter data provide a more precise account of a vessel's fuel consumption.

$$\text{fuel oil consumption} = \text{Power} \times \text{Specific Fuel Oil Consumption} \quad (16)$$

$$\text{fuel gas consumption} = \text{Power} \times \text{Specific Fuel Gas Consumption} \quad (17)$$

**Table 10.** R2 score for prediction FOC and FGC using flow meter data.

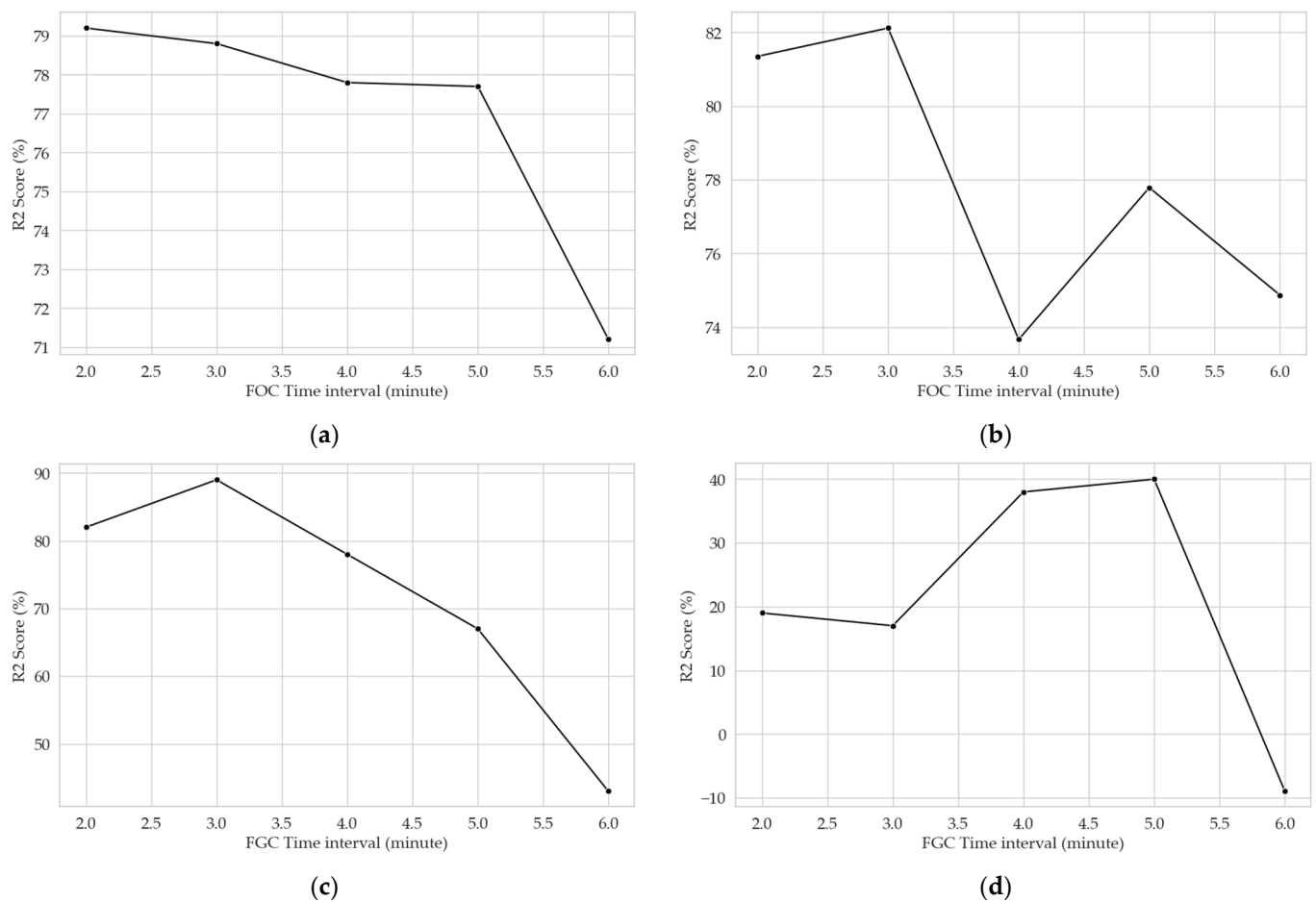
| Models           | R2 Score (%) |
|------------------|--------------|
| RFR for FOC      | 79.82        |
| LightGBM for FOC | 81.16        |
| RFR for FGC      | 91.2         |
| LightGBM for FGC | 88.45        |

**Table 11.** R2 score for prediction FOC and FGC using power data.

| Models           | R2 Score (%) |
|------------------|--------------|
| RFR for FOC      | 50.66        |
| LightGBM for FOC | 50.0         |
| RFR for FGC      | 78.56        |
| LightGBM for FGC | 77.06        |

### 3.3. Experiment 3: Predicting Fuel Consumption for Ships Considering Mixed Mode

Separating the modes to create individual models for each, this study acknowledged the existence of mixed mode timelines. These timelines, however, were not removed, limiting performance improvement. The aim of the research was to assess the model's performance when considering mixed mode intervals. To isolate these intervals, data with zero FGC were eliminated when predicting FOC, and the reverse was applied for FGC predictions. Consequently, this led to approximately 22% of the data being discarded in diesel mode. According to Figure 19, identifying the optimal numerical integration bins, ranging from 2 to 6 min, and eliminating the mixed mode bin enhanced performance. In gas mode, discarding around 12% of the data did not result in any performance improvements. The deletion of the mixed mode portion significantly impacted performance due to the data reduction having a larger effect on performance.



**Figure 19.** R2 score plots over time intervals. (a) FOC graph; (b) FOC graph considering mixed mode; (c) FGC graph; (d) FGC graph considering mixed mode.

#### 4. Discussion

Research on dual-fuel propulsion ships is a crucial step towards eco-friendly maritime transportation. In this study, we propose a model for predicting fuel consumption and carbon emissions using operational data from dual-fuel propulsion ships. Among random forest, LightGBM, and LSTM, the random forest model exhibited the highest performance with an accuracy of 92.7% in diesel mode and 93.5% in gas mode. We also employed the SHAP model to analyze which features played a significant role in predicting fuel consumption and carbon emissions.

The experiments detailed in Section 3 present the model's feasibility. CO<sub>2</sub> emissions for voyages from Ulsan to Busan or Jeju were calculated using data from ships operating in two different fuel modes. Operators can leverage the developed models to predict CO<sub>2</sub> emissions based on operational data before commencing the voyage. This information enables the provision of services that offer eco-friendly navigation routes. The data we employed, measured from flow meters, provide more accurate fuel consumption figures than those calculated using power data.

In our final experiment, we explored the potential for further research. Although operating modes are distinct, fuel is often mixed during the conversion process and navigation. Considering these aspects, creating a separate mixed-fuel mode label and training machine learning or deep learning models could lead to more accurate predictions. However, this would require the collection of sufficient data. Particularly, conducting research with the collection of various data on long-distance operations from smart ships will yield better results. The proposed method can utilize the vessel route planning considering the mixed engine mode under severe carbon emission conditions. The contribution of the wind and

waves to the fuel consumption and the CO<sub>2</sub> emission prediction is worth further exploration. A limitation of our research is the inclusion of only the measured wind data during the voyage mode because the anemometer's wind speed and direction data are obtained from the target ship. Meanwhile, the waves and the current contribution are significant to an ocean-going vessel's CO<sub>2</sub> emission. In a hybrid method, the voyage data and the wind and current data provided by the meteorological organization can be explored further to improve the performance of the forecast model. The proposed model in this paper is designed to apply to the vessel, which is equipped to accumulate voyage-related and fuel oil consumption-related data. The model can be applied to the measured voyage data of the vessel, which has a single- or dual-fuel engine mode. Another limitation of the developed method is that the prediction performance of the proposed model was validated for near-shore cruising ferries, which have dual-fuel engines only. Therefore, the prediction performance could show different tendencies when the application boundary is expanded to other sizes, types, and operation regions, such as canal and ocean-going vessels. The proposed method could be expanded with the onboard voyage data accumulation of a smart ship, which generally has onboard voyage data measurement equipment. The proposed method's applicability to other types of vessels is valuable and should be investigated.

**Author Contributions:** Conceptualization, J.L. and S.K.; methodology, J.L.; software, J.L.; validation, J.L. and S.K.; formal analysis, J.J. and J.E.; investigation, J.P.; data curation, J.L.; writing—original draft preparation, J.L., J.E. and S.K.; visualization, J.L.; supervision, S.K. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research was funded by the Ministry of Science and ICT (MSIT), Republic of Korea, through the ICT Challenge and Advanced Network of HRD (ICAN) program (IITP-2024-RS-2022-00156345), under the supervision of the Institute for Information and Communications Technology, Planning and Evaluation (IITP).

**Institutional Review Board Statement:** Not applicable.

**Informed Consent Statement:** Not applicable.

**Data Availability Statement:** Data are contained within the article.

**Conflicts of Interest:** The authors declare no conflicts of interest.

## References

1. Ocean Shipping and Shipbuilding. Available online: <https://www.oecd.org/ocean/topics/ocean-shipping/> (accessed on 31 January 2024).
2. Fourth Greenhouse Gas Study 2020. Available online: <https://www.imo.org/en/OurWork/Environment/Pages/Fourth-IMO-Greenhouse-Gas-Study-2020.aspx> (accessed on 31 January 2024).
3. Kim, S.W.; Yun, S.W.; You, Y.J. Eco-Friendly speed control algorithm development for autonomous vessel route planning. *J. Mar. Sci. Eng.* **2021**, *9*, 583. [\[CrossRef\]](#)
4. Perera, L.P.; Mo, B. Emission control based energy efficiency measures in ship operations. *Appl. Ocean Res.* **2016**, *60*, 29–46. [\[CrossRef\]](#)
5. Laasma, A.; Otsason, R.; Tapaninen, U.; Hilmola, O.P. Evaluation of Alternative Fuels for Coastal Ferries. *Sustainability* **2022**, *14*, 16841. [\[CrossRef\]](#)
6. Zhou, H.; Zhang, W. Choice of Emission Control Technology in Port Areas with Customers' Low-Carbon Preference. *Sustainability* **2022**, *14*, 13816. [\[CrossRef\]](#)
7. He, S.; Wu, X.; Wang, J. A calculation algorithm for ship pollutant gas emissions and diffusions based on real-time meteorological conditions and its application. *Ocean Eng.* **2023**, *287*, 115825. [\[CrossRef\]](#)
8. Corbett, J.J.; Winebrake, J.J.; Green, E.H.; Kasibhatla, P.; Eyring, V.; Lauer, A. Mortality from ship emissions: A global assessment. *Environ. Sci. Technol.* **2007**, *41*, 8512–8518. [\[CrossRef\]](#)
9. Jonson, J.E.; Gauss, M.; Schulz, M.; Jalkanen, J.P.; Fagerli, H. Effects of global ship emissions on European air pollution levels. *Atmos. Chem. Phys.* **2020**, *20*, 11399–11422. [\[CrossRef\]](#)
10. Saxe, H.; Larsen, T. Air pollution from ships in three Danish ports. *Atmos. Environ.* **2004**, *38*, 4057–4067. [\[CrossRef\]](#)
11. Ballini, F.; Bozzo, R. Air pollution from ships in ports: The socio-economic benefit of cold-ironing technology. *Res. Transp. Bus. Manag.* **2015**, *17*, 92–98. [\[CrossRef\]](#)
12. Chatzinikolaou, S.D.; Oikonomou, S.D.; Ventikos, N.P. Health externalities of ship air pollution at port—Piraeus port case study. *Transp. Res. Part D Transp. Environ.* **2015**, *40*, 155–165. [\[CrossRef\]](#)

13. Miola, A.; Ciuffo, B. Estimating air emissions from ships: Meta-analysis of modelling approaches and available data sources. *Atmos. Environ.* **2011**, *45*, 2242–2251. [\[CrossRef\]](#)
14. Lin, Y.-H.; Fang, M.-C.; Yeung, R.W. The optimization of ship weather-routing algorithm based on the composite influence of multi-dynamic elements. *Appl. Ocean Res.* **2013**, *43*, 184–194. [\[CrossRef\]](#)
15. Vettor, R.; Soares, C.G. Development of a ship weather routing system. *Ocean Eng.* **2016**, *123*, 1–14. [\[CrossRef\]](#)
16. Veneti, A.; Makrygiorgos, A.; Konstantopoulos, C.; Pantziou, G.; Vetsikas, I.A. Minimizing the fuel consumption and the risk in maritime transportation: A bi-objective weather routing approach. *Comput. Oper. Res.* **2017**, *88*, 220–236. [\[CrossRef\]](#)
17. Kim, S.W.; Eom, J.O. Ship Carbon Intensity Indicator Assessment via Just-in-Time Arrival Algorithm Based on Real-Time Data: Case Study of Pusan New International Port. *Sustainability* **2023**, *15*, 13875. [\[CrossRef\]](#)
18. Chen, D.; Zhao, Y.; Nelson, P.; Li, Y.; Wang, X.; Zhou, Y.; Lang, J.; Guo, X. Estimating ship emissions based on AIS data for port of Tianjin, China. *Atmos. Environ.* **2016**, *145*, 10–18. [\[CrossRef\]](#)
19. Moreno-Gutiérrez, J.; Pájaro-Velázquez, E.; Amado-Sánchez, Y.; Rodríguez-Moreno, R.; Calderay-Cayetano, F.; Durán-Grados, V. Comparative analysis between different methods for calculating on-board ship's emissions and energy consumption based on operational data. *Sci. Total Environ.* **2019**, *650*, 575–584. [\[CrossRef\]](#)
20. Wijaya, A.T.A.; Ariana, I.M.; Handani, D.W.; Abdillah, H.N. Fuel Oil Consumption Monitoring and Predicting Gas Emission Based on Ship Performance using Automatic Identification System (AISITS) Data. In Proceedings of the 2nd Maritime Safety International Conference (MASTIC), Surabaya, Indonesia, 18 July 2020; IOP Conference Series: Earth and Environmental Science. IOP Publishing: Bristol, UK, 2020; Volume 557.
21. Kim, K.S.; Roh, M.I. ISO 15016: 2015-based method for estimating the fuel oil consumption of a ship. *J. Mar. Sci. Eng.* **2020**, *8*, 791. [\[CrossRef\]](#)
22. ISO. ISO 15016:2015-Ship and Marine Technology-Guidelines for the Assessment of Speed and Power Performance by Analysis of Speed Trial Data; ISO: Geneva, Switzerland, 2015.
23. Guo, B.; Liang, Q.; Tvette, H.A.; Brinks, H.; Vanem, E. Combined machine learning and physics-based models for estimating fuel consumption of cargo ships. *Ocean Eng.* **2022**, *255*, 111435. [\[CrossRef\]](#)
24. Wang, S.; Ji, B.; Zhao, J.; Liu, W.; Xu, T. Predicting ship fuel consumption based on LASSO regression. *Transp. Res. Part D Transp. Environ.* **2018**, *65*, 817–824. [\[CrossRef\]](#)
25. Okumuş, F.; Ekmekçioglu, A.; Kara, S.S. Modelling ships main and auxiliary engine powers with regression-based machine learning algorithms. *Pol. Marit. Res.* **2021**, *1*, 83–96. [\[CrossRef\]](#)
26. Chen, Z.S.; Lam, J.S.L.; Xiao, Z. Prediction of harbour vessel fuel consumption based on machine learning approach. *Ocean Eng.* **2023**, *278*, 114483. [\[CrossRef\]](#)
27. Le, L.T.; Lee, G.; Park, K.S.; Kim, H. Neural network-based fuel consumption estimation for container ships in Korea. *Marit. Policy Manag.* **2020**, *47*, 615–632. [\[CrossRef\]](#)
28. Panapakidis, I.; Sourtzi, V.M.; Dagoumas, A. Forecasting the fuel consumption of passenger ships with a combination of shallow and deep learning. *Electronics* **2020**, *9*, 776. [\[CrossRef\]](#)
29. Ahlgren, F.; Mondejar, M.E.; Thern, M. Predicting dynamic fuel oil consumption on ships with automated machine learning. *Energy Procedia* **2019**, *158*, 6126–6131. [\[CrossRef\]](#)
30. Uyanık, T.; Karatug, Ç.; Arslanoğlu, Y. Machine learning approach to ship fuel consumption: A case of container vessel. *Transp. Res. Part D Transp. Environ.* **2020**, *84*, 102389. [\[CrossRef\]](#)
31. Uyanık, T.; Yalman, Y.; Kalenderli, Ö.; Arslanoğlu, Y.; Terriche, Y.; Su, C.L.; Guerrero, J.M. Data-Driven Approach for Estimating Power and Fuel Consumption of Ship: A Case of Container Vessel. *Mathematics* **2022**, *10*, 4167. [\[CrossRef\]](#)
32. Tarelko, W.; Rudzki, K. Applying artificial neural networks for modelling ship speed and fuel consumption. *Neural Comput. Appl.* **2020**, *32*, 17379–17395. [\[CrossRef\]](#)
33. Kaklis, D.; Eiriniakis, P.; Giannakopoulos, G.; Spyropoulos, C.; Varelas, T.J.; Varlamis, I. A big data approach for Fuel Oil Consumption estimation in the maritime industry. In Proceedings of the 2022 IEEE Eighth International Conference on Big Data Computing Service and Applications (BigDataService), Newark, CA, USA, 15–18 August 2022.
34. Kaklis, D.; Varlamis, I.; Giannakopoulos, G.; Spyropoulos, C.; Varelas, T.J. Online training for fuel oil consumption estimation: A data driven approach. In Proceedings of the 2022 23rd IEEE International Conference on Mobile Data Management (MDM), Paphos, Cyprus, 6–9 June 2022.
35. Tran, T.A. Comparative analysis on the fuel consumption prediction model for bulk carriers from ship launching to current states based on sea trial data and machine learning technique. *J. Ocean Eng. Sci.* **2021**, *6*, 317–339. [\[CrossRef\]](#)
36. Chen, C.; Verwilligen, J.; Mansuy, M.; Eloit, K.; Lataire, E.; Delefortrie, G. Tracking controller for ship manoeuvring in a shallow or confined fairway: Design, comparison and application. *Appl. Ocean Res.* **2021**, *115*, 102823. [\[CrossRef\]](#)
37. International Council on Clean Transportation. Available online: <https://theicct.org/> (accessed on 31 January 2024).
38. Yoon, J.H.; Kim, S.W.; Eom, J.O.; Oh, J.; Kim, H.J. Coastal Air Quality Assessment through AIS-Based Vessel Emissions: A Daesan Port Case Study. *J. Mar. Sci. Eng.* **2023**, *11*, 2291. [\[CrossRef\]](#)

**Disclaimer/Publisher's Note:** The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.