

---

This is an electronic reprint of the original article.  
This reprint may differ from the original in pagination and typographic detail.

Zhang, Mingyang; Tsoulakos, Nikolaos; Kujala, Pentti; Hirdaris, Spyros

## A deep learning method for the prediction of ship fuel consumption in real operational conditions

*Published in:*  
Engineering Applications of Artificial Intelligence

*DOI:*  
[10.1016/j.engappai.2023.107425](https://doi.org/10.1016/j.engappai.2023.107425)

Published: 01/04/2024

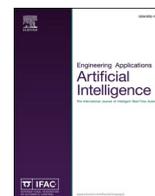
*Document Version*  
Publisher's PDF, also known as Version of record

*Published under the following license:*  
CC BY

*Please cite the original version:*  
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, Article 107425. <https://doi.org/10.1016/j.engappai.2023.107425>

---

This material is protected by copyright and other intellectual property rights, and duplication or sale of all or part of any of the repository collections is not permitted, except that material may be duplicated by you for your research use or educational purposes in electronic or print form. You must obtain permission for any other use. Electronic or print copies may not be offered, whether for sale or otherwise to anyone who is not an authorised user.



# A deep learning method for the prediction of ship fuel consumption in real operational conditions

Mingyang Zhang<sup>a,\*</sup>, Nikolaos Tsoulakos<sup>b</sup>, Pentti Kujala<sup>a</sup>, Spyros Hirdaris<sup>a,\*\*</sup>

<sup>a</sup> Department of Mechanical Engineering, Marine Technology Group, Aalto University, Espoo, Finland

<sup>b</sup> Laskaridis Shipping Co. Ltd., Athens, Greece

## ARTICLE INFO

### Keywords:

Ship fuel consumption  
Decarbonisation  
Big data science  
Deep learning  
Bulk carrier

## ABSTRACT

In recent years, the European Commission and the International Maritime Organization (IMO) implemented various operational measures and policies to reduce ship fuel consumption and related emissions. The effectiveness of these measures relies upon developing accurate predictive models encompassing the influence of real operational conditions. This paper presents a deep learning method for the prediction of ship fuel consumption. The method utilizes big data analytics from sensors, voyage reporting and hydrometeorological data, comprising of 266 variables made available following sea trials of a Kamsarmax bulk carrier of Laskaridis Shipping Co. Ltd. A variable importance estimation model using a Decision Tree (DT) is used to understand the underlying relationships in the available dataset. Consequently, a deep learning model is developed to understand the influence of sailing speed, heading, displacement/draft, trim, weather, sea conditions, etc. on ship fuel consumption (SFC). This is achieved by incorporating attention mechanism into Bi-directional Long Short-Term Memory (Bi-LSTM) network. The potential of the new method is demonstrated by training data streams corresponding to real ship fuel consumption rates as well as internal and external operational conditions. A comprehensive comparison with existing methods indicates that the Bi-LSTM with attention mechanism presents the best fit when using high frequency data. It is concluded that subject to further testing and validation the method could be used for the development of decision support systems for monitoring environmentally sustainable ship operations.

## 1. Introduction

Maritime transport, as highlighted by the United Nations Conference on Trade and Development (UNCTAD), assumes a paramount role in the global trade and transportation supply chain. Today shipping is responsible for approximately over 80% of the transportation of global cargo volume while it ensures the uninterrupted movement of goods fostering global economic growth (UNCTAD, 2022). Whereas shipping may be considered relatively safe and clean in relation to other modes of transportation (Kevin and Kodak, 2023; Probha and Hoque, 2018), within the context of the targets set by the Paris Agreement decarbonisation remains a top priority (UNFCCC, 2022). According to the International Maritime Organisation (IMO) in 2018 shipping accounted for 2.89% of anthropogenic emissions worldwide (IMO, 2020a; IMO, 2020b). Hence to reduce Greenhouse Gas (GHG) emissions by 2050 by 50% in relation to the 2008 baseline, it is imperative to decarbonise via

retrofitting ships in service and reduce the emissions footprint of new-build specifications.

To respond to the pressing needs of the decarbonisation agenda there is an urgent need to develop a precise model for the accurate prediction of Ship Fuel Consumption (SFC) rates in real operational conditions. Review papers by Yan et al. (2021) and Fan et al. (2022) highlight that such a model can be further harnessed to optimize ship voyages in real operational conditions, with the primary objectives of achieving improved fuel efficiency and emission reductions.

Common research on SFC prediction/estimation can be broadly categorized into two main areas namely (a) physics-based models and (b) machine learning models.

Physics-based models offer a reliable means of approximating SFC based on specific assumptions (Lang and Mao, 2020). Common methods used are based on ship-propeller-engine performance models, and ship resistance estimation models, see Fan et al. (2022). The former are used

\* Corresponding author.

\*\* Corresponding author.

E-mail addresses: [mingyang.0.zhang@aalto.fi](mailto:mingyang.0.zhang@aalto.fi) (M. Zhang), [spyros.hirdaris@aalto.fi](mailto:spyros.hirdaris@aalto.fi) (S. Hirdaris).

to assess the overall performance of ship energy systems, e.g., engine, propeller, and associated components (Dai et al., 2022; Saydam et al., 2022). And they are useful in terms of estimating the energy needed to overcome the total resistance encountered by a ship during operations (Wang, 2020; Wang et al., 2021). Tillig and Ringsberg (2019) developed a ship fuel consumption prediction model based on the generic energy systems theory. Their approach accounts for ship motions in four degrees of freedom, external forces and moments associated with waves. More recently, Kim et al. (2023) proposed a comprehensive approach for ship powering prediction. Their method includes data pre-processing for ship resistance estimation, and propulsion efficiency. These studies collectively suggest that the accuracy of physics-based models for ship fuel consumption are closely tied to ship resistance.

To comprehensively evaluate the total resistance experienced by a ship, both calm water and added resistance components should be considered, e.g., see Lang and Mao. (2020) and Wang et al. (2020). Calm water resistance refers to the hydrodynamic effects a ship encounters when she is moving through calm and undisturbed waters. Sub-components of this resistance include the influences from frictional and viscous effects (Molland et al., 2017). Added resistance refers to additional hydrodynamic influences a ship hull may encounter from the waves, wind, other ships, or obstacles (Vinayak et al., 2021; Liu et al., 2020). In most cases, ship resistance is evaluated in still waters (Liu and Papanikolaou, 2020). The Holtrop and Mennen (1982) model and its derivatives, e.g., see Julianto et al. (2021) and Elkafas et al. (2019), have been commonly employed to estimate ship resistance at the preliminary ship design stage. Over the years, the International Towing Tank Conference (ITTC) proposed empirical methods (ITTC, 2002). For example, Choi et al. (2010) employed the model-based ship performance analysis method based on the revised ITTC'78 method to predict speed performance using resistance and propulsion characteristics. On the same basis, Min and Kang (2010) introduced an improved extrapolation procedure based on the ITTC'78 method to predict full-scale ship resistance performance.

In recent years, the use of Computational Fluid Dynamics (CFD) became increasingly useful within the context of ship design (Yaakob et al., 2015; Haase et al., 2016). A recent contribution by Grlj et al. (2023) introduced a Reynolds Averaged Navier-Stokes (RANS)- CFD model to investigate the influences of various container configurations, trim settings, and ship motion on wind and air resistance. Farkas et al. (2020) presented a novel method for estimating the effects of fouling on ship resistance and propulsion characteristics using the same method. Campbell et al. (2022) used CFD to explore the influence of trim and draught variations on the resistance of a ship in confined waters. The ISO (2015) guidance, the empirical methods proposed by Lewis (1988), and the CFD models of Kim et al. (2017); Luo et al. (2016) can be employed to estimate the added resistance due to wind. The empirical methods put forth by Hasselmann et al. (1973) and numerical modelling procedures introduced by Sadat-Hosseini et al. (2013), Luo et al. (2016), and Cho et al. (2023) can be utilized to estimate wave resistance. To account for the combined effects of wave resistance on ship motions, semi-empirical models have been proposed. These models are based on experimental data and incorporate empirical equations. Examples are given by Fujii and Takahashi (1975), Tsujimoto et al. (2008), and Liu et al. (2016, 2020). Jinkine and Ferdinande. (1974), Liu et al. (2016), ITTC (2014, 2017), and Valanto and Hong (2015) also explored the overall influence of added resistance effects on ship motions.

Physics-based models are often used to estimate the total resistance a ship may encounter during operations (e.g., Lang, 2023; Carlton, 2018). Nevertheless, it is worth noting that their accuracy may be limited when it comes to changing or extreme operational conditions (Vinayak et al., 2021) that pose unique challenges not accounted for in existing models (Fan et al., 2020, 2022; Yan et al., 2021; Haranen et al., 2016).

Artificial Intelligence (AI) technologies and big data theory could offer a potentially valid solution to address the challenges associated with capturing influencing factors and interactions from various

**Table 1**

The existing machine learning methods for ship fuel consumption.

Methods		Number of ships	Ship type	Literature
Machine learning methods	Adaptive boosting (AB)	1	Container ship	Uyanik et al. (2020)
	Decision trees (DT)	1	Dry bulk ship	Yan et al. (2020)
	Gaussian process regression (GPR)	1	Container ship	Hu et al. (2019)
	Extreme gradient boosting (XGBoost)	1	VLCC	Papandreou and Ziakopoulos (2022)
	Least absolute shrinkage and selection operator (LASSO)	1	Container ship	Wang et al. (2018)
	Multiple linear regression (MLR)	2 (Sister ships)	Cruise ship	Bocchetti et al. (2013)
	Random forest (RF)	2 (Sister ships)	Tuna seiner	Zhou et al. (2023)
	Support vector regression (SVR)	1	Container ship	Uyanik et al. (2020)
Deep learning methods	Long short-term memory networks (LSTM)	1	Inland cargo ship	Yuan et al. (2020)
	Gated recurrent units (GRU)	1	Inland cargo ship	Yuan et al. (2020)
	Artificial neural networks (ANN)	1	Container ship	Kim et al. (2021)

The term "sister ships" typically refers to a ship that has an identical or nearly identical sister ship. These sister ships are often built using the same design and specifications, making them very similar or nearly identical in terms of size, shape, and functionality.

resistance components encompassed in physics-based models (Chen et al., 2023; Lang, 2023; Shang et al., 2023). This is because AI could be employed to elucidate the intricate relationship between measured SFC and a multitude of influential parameters, encompassing realistic navigational patterns, ship operational status, engineering systems specifics, weather conditions, etc. To date, several review papers have extensively discussed the potential application of machine and deep learning methods to predict SFC (Yan et al., 2021; Fan et al., 2022; Huang et al., 2022). These papers identified three main clusters of algorithms commonly utilized, namely: (i) supervised Machine Learning Methods (MLMs), (ii) unsupervised MLMs, and (iii) Deep Learning Methods (DLMs), see Table 1. Supervised MLMs encompass a range of techniques including multiple linear regression (Bocchetti et al., 2013), random forest (Coraddu et al., 2017), least absolute shrinkage and selection operator regression (Wang et al., 2018), support vector regression (Uyanik et al., 2020), extreme gradient boosting (Papandreou and Ziakopoulos, 2022), adaptive boosting (Uyanik et al., 2020) and decision trees (Yan et al., 2020). Unsupervised MLMs involve Gaussian process regression (Hu et al., 2019) and Gaussian mixture model (Jeon et al., 2018). DLMs consist of artificial neural networks (Kim et al., 2021), long short-term memory network (Yan et al., 2020a), and gated recurrent units (Yan et al., 2020b).

All methods have been employed to predict the SFC of different ships by utilizing diverse data sources such as voyage reports, Automatic Identification System (AIS) data, as well as hydrometeorological and other sensor data (Dai et al., 2022a; Li et al., 2022a; Li and Yang, 2023; Li et al., 2023b). Understandably, these diverse data sources provide valuable information that enables the accurate predictions and enhances the understanding of the factors that may impact the evaluation of SFC. Nevertheless, statistical analysis reveals that the existing MLMs may be sensitive to the varying numbers of influencing factors (inputs), apparently ranging from 7 to 75 (Chen et al., 2023).

MLM and DLM algorithms can be used to predict SFC on the basis of

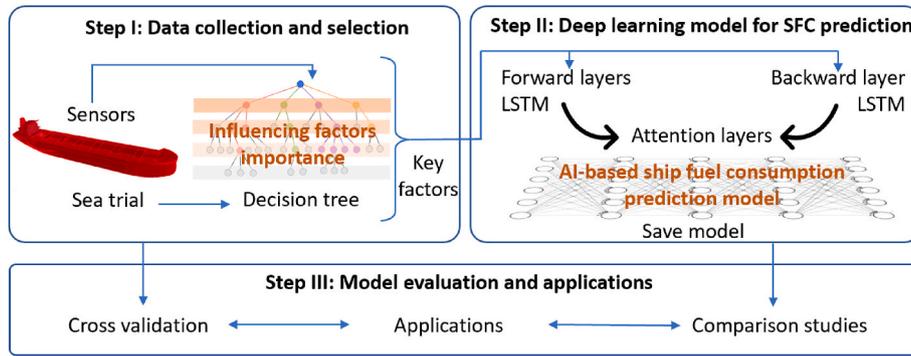


Fig. 1. The attention-based Bi-LSTM framework for the prediction of ship fuel consumption.

well-defined scenarios for a specific ship (e.g., [Chen et al., 2023](#); [Yan et al., 2020](#); [Du et al., 2022a,b](#)). However, they often fail to accurately predict SFC in real time or for an entire ship voyage or a fleet in complex or extreme conditions. This could be attributed to the following reasons.

- (i) Existing models tend to omit key operational factors under complex or extreme operational conditions ([Chen et al., 2023](#)). A key objective of the existing SFC prediction models should be to accurately account for the effects of intricate ship energy systems during operations. Subsequently, a trained model should function as a digital twin. Results may then be used in aiding routing and OPEX (operational expenditure) optimization.
- (ii) Existing models do not consider the taxonomy of influencing factors and their combined or retrospective effects on SFC prediction. For example, certain big data analytics may introduce noise to data patterns and result in overfitting issues that compromise the performance of a trained model ([Zhou et al., 2022](#); [Fiskin et al., 2021](#)).
- (iii) MLM and DLM models struggle to effectively deal with the complexity of big data patterns and their combinations. The selective use of information on the basis of attention mechanisms entails limitations. Thus, more sophisticated methods should be

developed to improve model performance and big data interpretability ([Zhang et al., 2023](#)).

To address the above limitations, this paper proposes an AI informed method for the prediction of SFC in real operational conditions. The method accounts for complex big data patterns, combinatorial influencing factors, and DLM using a Bi-LSTM method with attention mechanisms. The practical application of the proposed approach is exemplified through the utilization of big data records collected over sea trials of a bulk carrier operated by Laskaridis Shipping Co. Ltd. The new approach is validated by comparison against current methods. The paper concludes on the perspective potential of the ideas developed to contribute toward environmentally sustainable shipping operations.

## 2. Methods

In this paper, a decision tree (DT) is used to evaluate the importance of factors encompassed on extensive big data analytics records and consequently select key influencing factors with impact on SFC. A Bi-LSTM with attention mechanism method is employed. The methodology presented comprises of the following three steps (see [Fig. 1](#)).

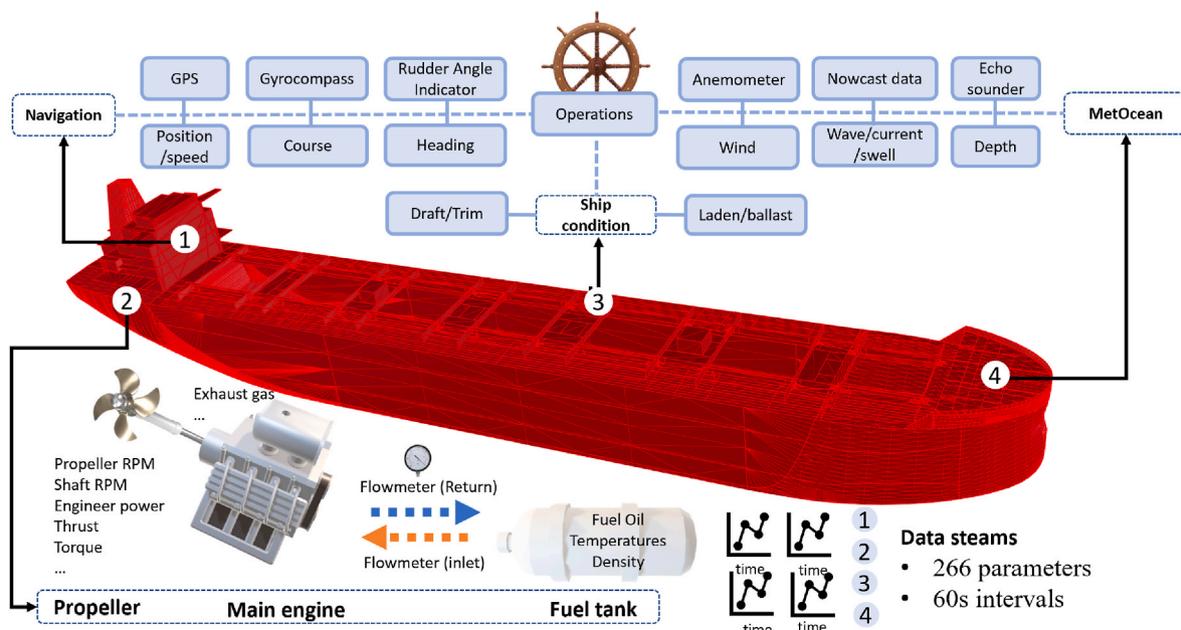


Fig. 2. Data collection and multisource-information fusion.

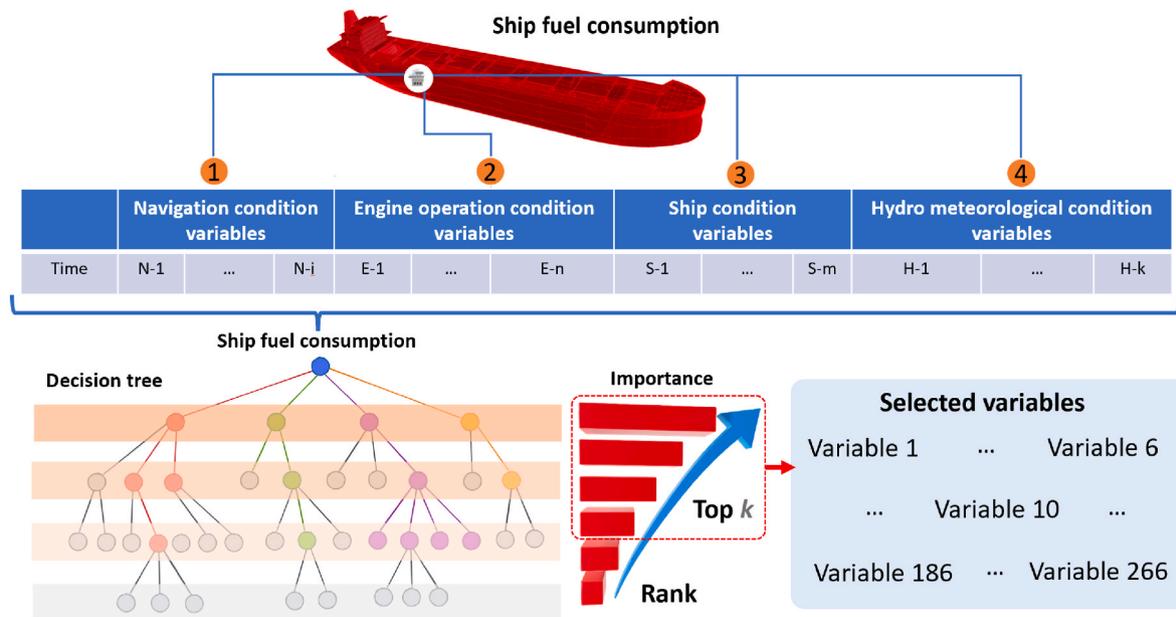


Fig. 3. Variable importance calculation for inputs selection of deep learning method.

### • Step I: Data collection and the importance evaluation of influencing factors.

To collect streams of big data analytics from sea trials, numerous sensors are deployed. These sensors encompass a range of functionalities, including GPS, gyrocompass, anemometer, torque meter, flow meter, thermometer, echosounder, density meter, etc. In this study, the sensors used collectively measured 266 parameters including navigation data (i.e., speed, heading, course, etc.), ship operational status (i.e., draft, trim, etc.), the functionality of engineering systems (i.e., fuel oil volumetric flow, fuel oil density, fuel oil temperature, propeller revolutions, engine power), Metocean data (i.e., air pressure and temperature, hydrometeorological data, etc.). Subsequently, a DT method is employed to assess the significance of big data parameters on SFC in real operational conditions.

### • Step II: A DLM for the prediction of ship fuel consumption.

The DLM incorporates attention mechanisms into a Bi-LSTM network (Lin et al., 2019; Li et al., 2023). The network focuses on specific parts of the data streams and assigns varying degrees of importance to different parts of the training data. During training, the network can capture information from both past and future data sequences. Consequently, the model developed is capable to capture long term dependencies from ship energy systems in real operational conditions.

### • Step III: Cross validation, comparison, and applications.

The trained model is validated using a  $k$ -fold cross-validation procedure. This ensures accuracy and reliability (Ma et al., 2022; Zhang et al., 2023). Comparative studies are then conducted to assess the performance of the method against MLMs used in existing studies. Finally, the generalisation of the model is tested by direct application to new long-period ship voyages (see Section 3.3).

#### 2.1. Data collection and feature importance evaluation

To gain a comprehensive view of the performance of ship energy system in real operational conditions, sensors were installed on a Kamsarmax bulk carrier and big data streams were collected during

extensive sea trials. Sensors included a GPS for position tracking, a gyrocompass for heading measurements, an anemometer and a torque meter for wind speed and torque monitoring, a flow meter for fuel flow assessment, thermometers for temperature measurements, an echosounder for water depth monitoring, a density meter for assessing the fuel oil properties, etc., see Fig. 2. The combined deployment of this hardware enabled the simultaneous measurement of 266 parameters (see Fig. A1 in appendix A) related to ship navigation, engine, ship condition and operational conditions. The data collection interval was 60 s and big data were classified as follows.

- *Navigation data* including parameters such as ship speed, heading, course, and position. These data streams are crucial for route planning and speed optimization.
- *Ship operational status data*, such as draft, trim, ballast, cargo load, and hull fouling. These parameters may impact ship hydrodynamic performance and fuel efficiency.
- *Engineering systems data*, e.g., fuel oil volumetric flow, fuel oil density, fuel oil temperature, propeller RPM, and engine power. This information can provide improved insights into engine performance and fuel consumption.
- *Metocean data* including information related to atmospheric and oceanic conditions such as air pressure and temperature, sea temperature, wave height, and current speed. Such data streams are crucial for understanding the influence of environmental conditions on ship safety and performance (Zhang et al., 2023a,b).

Whereas the data collection system captures a comprehensive an extensive set of time domain parameters, not all of those are equally important to determine SFC. They also do not contribute equally to the accuracy of results derived from models. On this basis, it was considered useful to derive a variable selection model to effectively reduce data dimensions.

A DT method has been used to examine the significance of variables with respect to operational targets (Karabadjji et al., 2014; Kazemitabar et al., 2017; Perner and Apte, 2004). By definition, a DT model has a hierarchical structure comprising of nodes, branches, and leaves, each serving a specific purpose, see Fig. 3 and (Abreu et al., 2023; Zhou et al., 2021). The fundamental principle of decision tree involves the recursive division of datasets based on selected variables. Such process may lead

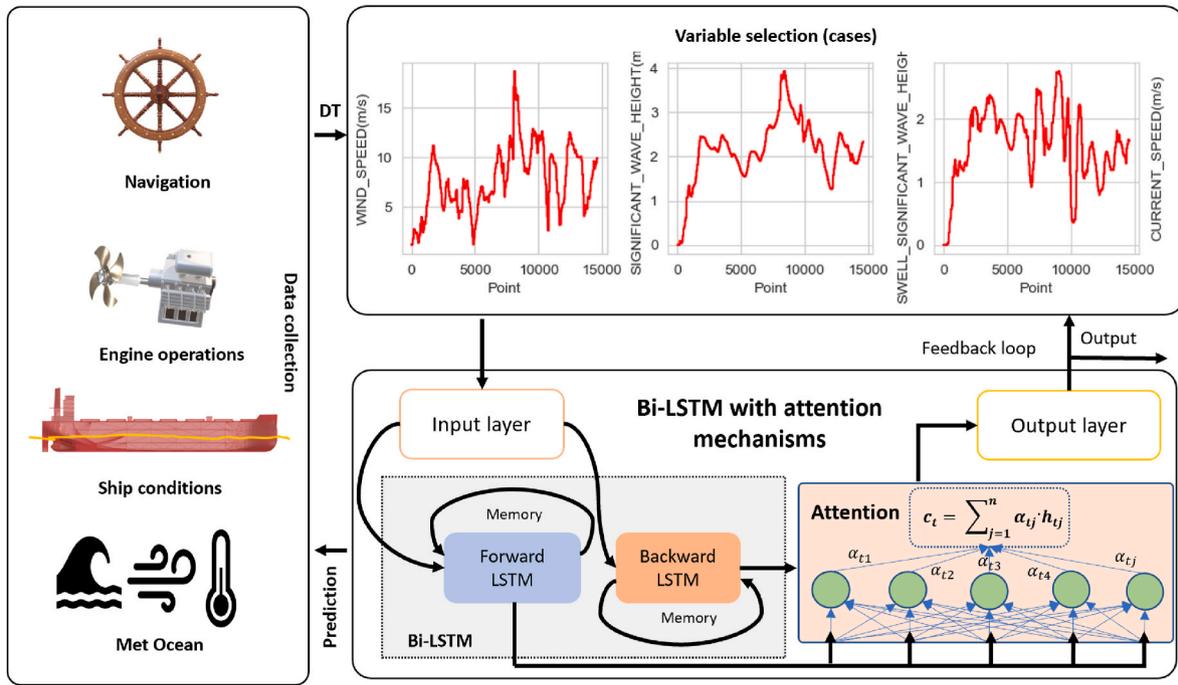


Fig. 4. The schematic of the deep learning method architecture for ship fuel consumption prediction.

to the generation of big data subsets. The process may ultimately lead to the formation of decision rules (Cai et al., 2018). A notable advantage of the method is that it captures intricate relationships between the input and the output nodes (Kazemitabar et al., 2017; Abreu et al., 2023). Unlike other modelling techniques, such as logistic regression and correlation methods (Hall, 1999), it may also handle complex interactions and nonlinear relationships between variables (Wu et al., 2023). This is because it considers the importance of different variables by examining their splits and hierarchy. Last but not least it produces interpretable results (Sagi and Rokach, 2021).

The process of constructing the DT regression model was based on the idea of evaluating the importance of variables that may involve attribute selection measures, tree growth, pruning, etc. (Kotsiantis, 2013; Zhou et al., 2021). Generally, attribute selection measures are utilized to identify the most informative features and for partitioning the

data streams. In this paper, the Mean Absolute Error (MAE) is used for attribute selection, see Eq. (1).

$$i^2 = \frac{1}{n} \sum_{i=1}^n [y_i - \hat{y}_i]^2 \quad (1)$$

where,  $n$  is the number of predictions,  $y_i$  is the actual value,  $\hat{y}_i$  denotes the predicted value.

A dataset can be recursively partitioned using the selected variables. This leads to the creation of DT branches until termination conditions are met, e.g., maximum depth, minimum number of instances, random states, etc. (Rajagopal et al., 2020). To prevent overfitting (i.e., implementation of unnecessary branches or nodes), cost complexity pruning is applied (Raghavan, 2010; Karabadjji et al., 2014) as denoted in Eq. (2).

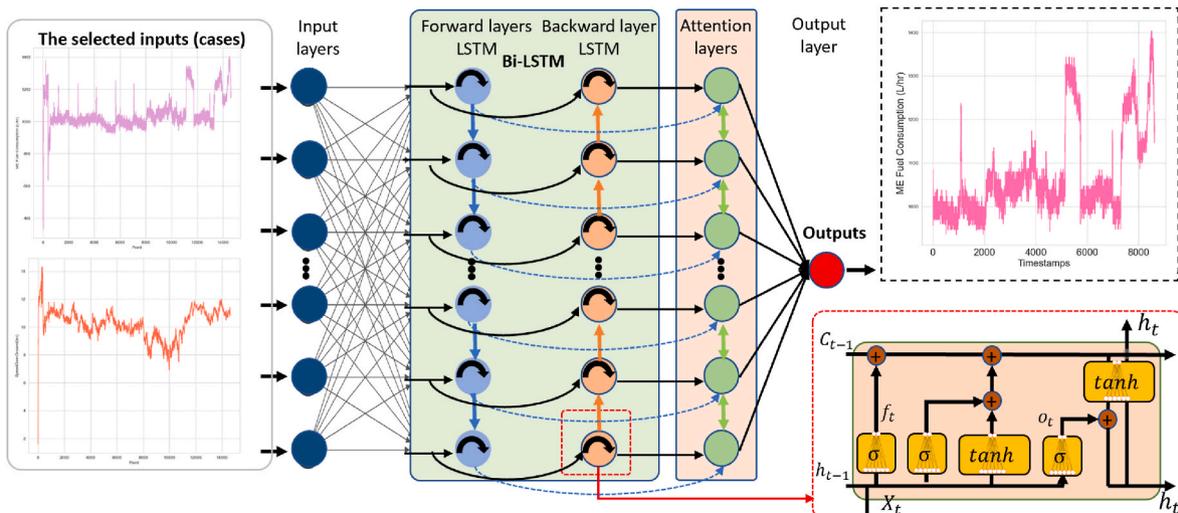


Fig. 5. The architecture of Bi-LSTM with attention mechanism for ship fuel consumption prediction.

**Table 2**

Algorithm 1- Bi-LSTM with attention mechanism for ship fuel consumption prediction.

Algorithm: Bi-LSTM with attention mechanisms for ship fuel consumption (SFC) prediction	
1	<b>Input:</b> Data collection SFC(n), see more in Section 2.1
2	<b>Output:</b> AI-based ship fuel consumption surrogate model
3	Select key variables X(n) and ship fuel consumption SFC (n) in the time domain
4	Split the training data set $data_{train}$ and $data_{test}$ from X(n) using k folds cross-validation method.
5	For batch $data_{batchsize}$ in $data_{train}$
6	For L-length data $data_i$ in $data_{batchsize}$
7	For $i = 1$ to L
8	Using forward LSTM to an encoder $\overline{h}_t$
9	Using backward LSTM to an encoder $\overleftarrow{h}_t$
10	End For
11	Compute Attention score $\alpha_{tj}$ and $c_t$
12	Compute ship fuel consumption $SFC_t$ from $c_t$
13	End For
14	Training the model to identify ship energy system in real operational conditions
15	End For
16	Save the prediction model: AI-based ship fuel consumption prediction model

$$J(\theta) = L(\theta) + \alpha \cdot R(\theta) \quad (2)$$

In Eq. (2)  $J(\theta)$  represents the objective function,  $L(\theta)$  is the loss function that measures the model performance on the training data,  $R(\theta)$  denotes the regularization term that penalizes the complexity of the model,  $\alpha$  is a hyperparameter that controls the trade off between the loss and regularization term.

The critical attribute of a node with children is denoted as  $U(N_j)$  in Eq. (3). Thus, when the MAE of the children node is small, see Eq. (1), the importance of the node, particularly its feature for splitting, becomes significant.

$$U(N_j) = w_j i_j^2 - W(N_k^l) i^2(N_k^l) - W(N_k^r) i^2(N_k^r) \quad (3)$$

In Eq. (3)  $j$  denotes the number of nodes of the developed tree,  $N_j, j \in \{1, 2, \dots, J\}$  presents all the nodes of the decision tree. Each  $N_j$  has a left child  $N_k^l$  and a right child  $N_k^r$ ,  $W(N_j) = w_j \in (0, 1]$  and denote the weight of node  $N_j$ .  $I^2(N_j)$  represents the mean absolute error of  $N_j$ .

The DT model can capture the complex relationships between navigation conditions, engine, and ship condition data, hydrometeorological conditions, and their impact on SFC. By evaluating the significance of features within the DT such as the frequency of feature usage for splitting and the resulting variance reduction, the importance of these parameters can be determined. Variables with high importance values indicate a strong influence on SFC, while those with low values may have minimal impact. The selection of the top  $k$  variables is based on their demonstrated ability to yield the highest prediction accuracy when used as input for the deep learning model (see Section 2.2). It is worth noting that the analysis of big data records demonstrated that modifying (increasing or decreasing), the number of variables may have a negative effect on the accuracy of the model. Thus, the top  $k$  variables were set as inputs in DLM, see Fig. 3.

## 2.2. A DLM for the prediction of ship fuel consumption

To predict SFC the paper introduces a DLM based on Bi-LSTM with attention mechanism, see Fig. 4. Since the fuel consumption prediction model can capture both past and future dependencies, and attention mechanisms are considered, the model introduced has by definition the potential to demonstrate good performance in terms of effectively capturing and utilizing historical information.

The architecture of the SFC prediction model as illustrated in Fig. 5,

consists of four main components namely: (i) input layer, (ii) Bi-LSTM layer, (iii) attention mechanism layer, and (iv) output layer. Table 2 summarizes the algorithm, i.e., it outlines the step-by-step procedure followed, including data pre-processing, model architecture setup, training process, and prediction generation. Details about the four main components can be summarised as follows.

### (i) Input layer

The model input includes key variables obtained from navigation data, engineering systems, ship status, and hydrometeorological conditions using the DT regression model, see Fig. 5. These selected variables are then pre-processed and fed into the subsequent layers, see Section 2.1.

### (ii) Bi-LSTM layer

Each Bi-LSTM layer consists of two LSTM sub-layers. In turn, each LSTM unit comprises of 4 interconnections that encompass the input and control signals for the input, forgotten, and output gates (Staudemeyer and Morris, 2019; Gao et al., 2020). These components work together to regulate memory storage, retention, and output. Fig. 5 depicts the internal structure of the LSTM unit (see red box). The input  $x_t$  at a time increment  $t$  and the output  $h_{t-1}$  of the hidden layer neuron at a time increment  $t-1$  represent the joint inputs to the hidden layer. These inputs are then multiplied by distinct weight vectors, and upon application of the activation function, the control signals  $f_t$ ,  $i_t$ ,  $o_t$ , of the forgotten gate, input gate, and output gate are generated. The mathematical basis is denoted by Eqs. (4)–(6), where the weight vector is represented by  $w$  and the activation amount by  $b$ . The biases for different connection weights are represented by  $b_f$ ,  $b_i$ , and  $b_o$ , while  $\sigma(\cdot)$  is the sigmoid activation function.

$$f_t = \sigma(w_f[h_{t-1}, x_t] + b_f) \quad (4)$$

$$i_t = \sigma(w_i[h_{t-1}, x_t] + b_i) \quad (5)$$

$$o_t = \sigma(w_o[h_{t-1}, x_t] + b_o) \quad (6)$$

The state of a cell  $\tilde{C}_t$  is presented as per Eq. (7). The value of the memory unit  $C_t$  is updated according to Eq. (8).

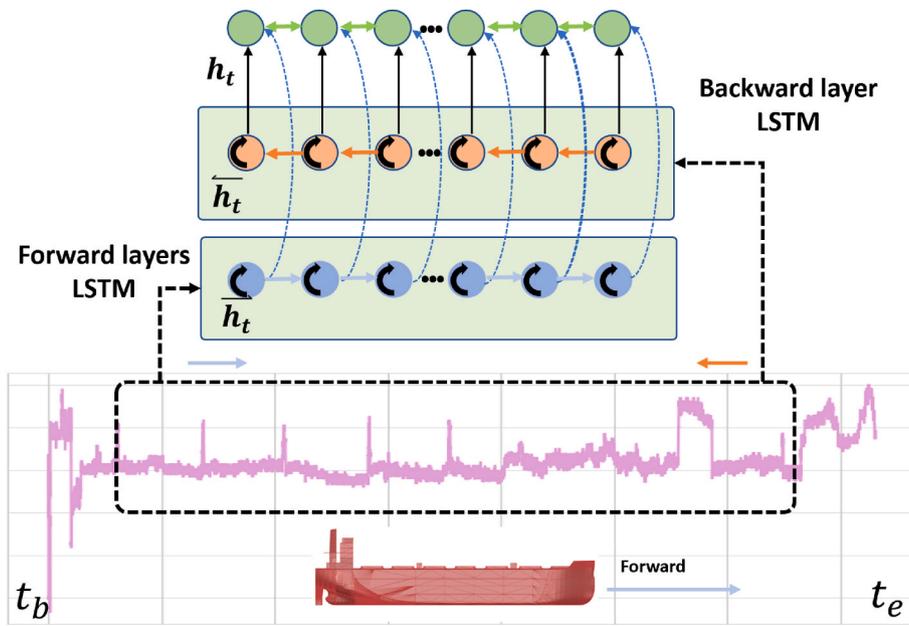


Fig. 6. Capturing and utilizing information from both forward and backward directions.

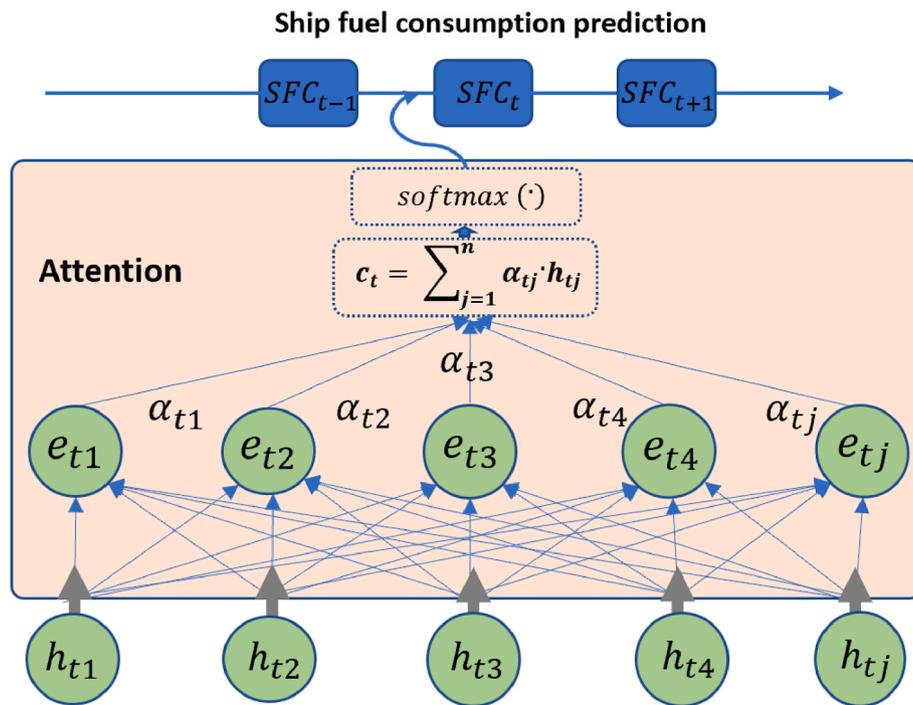


Fig. 7. Attention mechanism for ship fuel consumption prediction.

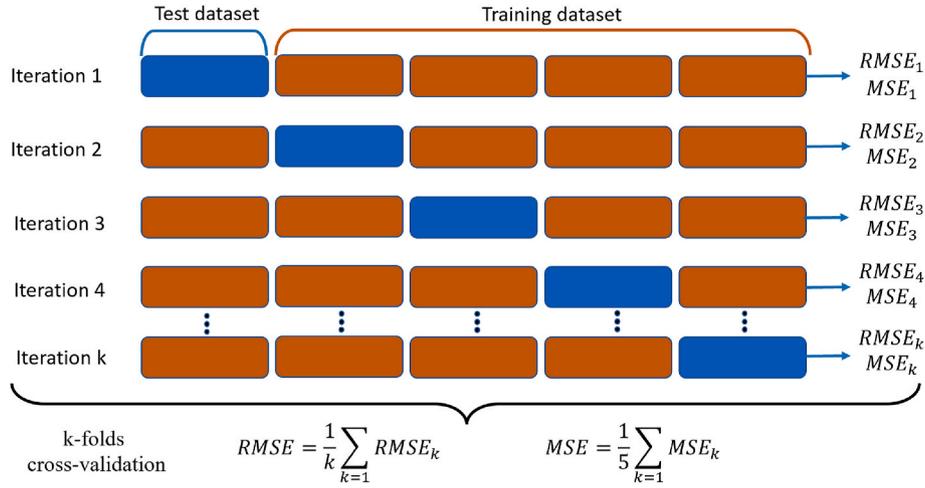


Fig. 8.  $k$  folds cross-validation for model evaluation.

$$\tilde{C}_i = \tanh(w_c[h_{i-1}, x_i] + b_c) \quad (7)$$

$$C_i = f_i * C_{i-1} + i_i * \tilde{C}_i \quad (8)$$

The output of hidden layer neurons  $h_t$  is defined as:

$$h_t = o_t * \tanh(C_t) \quad (9)$$

where  $\tanh(\cdot)$  represents a hyperbolic tangent activation function.

Ships are systems of systems that navigate through complex operational conditions (Zhang et al., 2021). To effectively capture the complexity of ship systems using real operational data for SFC prediction, it is essential for the DL model to learn from both past and future information in the data streams. While a standard LSTM model only considers information from past time frames in the data stream, the Bi-LSTM layer overcomes this limitation by addressing the problem of disregarding relevant past information (Ma et al., 2020; Xu et al., 2022; Gao et al., 2023). This is possible as it comprises of both forward and backward LSTM sublayers, see Fig. 5.

Forward sublayers process the input stream in a forward fashion, i.e., from the beginning  $t_b$  to the end  $t_e$ . The sublayers operate conversely to this (i.e., in a backward direction: from the end  $t_e$  to the beginning  $t_b$ ), see Fig. 6.

Given an input data stream  $X = [x_1, x_2, \dots, x_n]$ , a Bi-LSTM generates hidden states in both forward and backward directions as per Eqs. 10 and 11. Then, the hidden states from both directions are concatenated to obtain a comprehensive representation at each time step see Eq. (12).

$$\vec{h}_t = \text{LSTM}(x_t, \vec{h}_{t-1}) \quad (10)$$

$$\overleftarrow{h}_t = \text{LSTM}(x_t, \overleftarrow{h}_{t-1}) \quad (11)$$

$$h_t = [\vec{h}_t ; \overleftarrow{h}_t] \quad (12)$$

where  $t$  represents the time step, and  $\vec{h}_t$  is the hidden state in the forward direction.  $\overleftarrow{h}_t$  is the hidden state in the backward direction,  $[\cdot]$  denotes the concatenation.

### (iii) Attention mechanism

The attention mechanism allows the model to dynamically focus on different parts of the input data streams as well as assign varying levels of importance to different time steps (Zhang et al., 2023). This enhances its ability to emphasize critical information on the basis of the data streams available. This feature may be useful in terms of capturing extreme scenarios. The attention layers calculate attention scores and weights for each time step based on the hidden representations from the Bi-LSTM layers, see Fig. 7.

Given a specific time step  $t$ , the attention weight  $\alpha_{ij}$  of other hidden layers for the current input of  $x_t$  is calculated as given in Eq. (13).

$$\alpha_{ij} = \frac{\exp(e_{ij})}{\sum_{j=1}^n \exp(e_{ij})} \quad (13)$$

The attention score  $e_{ij}$  computed using an additive attention model is defined according to Eq. (14), where  $w_\mu$  and  $w_w$  represent the weights of the fully connected layers. The bias for different connection weights is represented by  $b_w$ . Consequently, the information is used to compute the attention weights for the attention mechanism (Chorowski et al., 2015).

$$e_{ij} = w_\mu^T * \tanh(w_w h_t + b_w) \quad (14)$$

The context vector  $c_t$  is the weighted sum of the hidden states. It reflects the attended information at a time step, see Eq. (15).

$$c_t = \sum_{j=1}^n \alpha_{ij} h_{ij} \quad (15)$$

The final output  $SFC_t$  at a time step  $t$  is generated by passing the context vector  $c_t$  through linear transformation and an activation function, see Eq. (16).

$$SFC_t = \text{softmax}(w_c c_t + b_c) \quad (16)$$

In the above expression,  $w_c$ , and  $b_c$  are parameters (weight matrices or

Table 3

Ship specification of the KAMSARMAX class bulk carrier.

Information			Real
IMO	9843405		
Vessel Type	Bulk Carrier		
DWT	81,600.0 t		Digital Twin
Length x Breadth	229.0 x 32.0 m		
Year Built	2020		

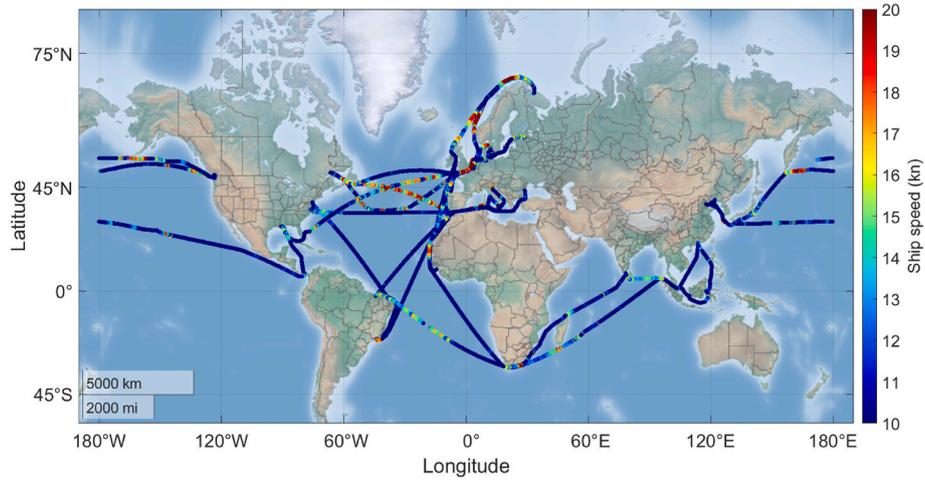


Fig. 9. Ship trajectories of sea trial data of a bulk carrier from 01.2021 to 02.2023.

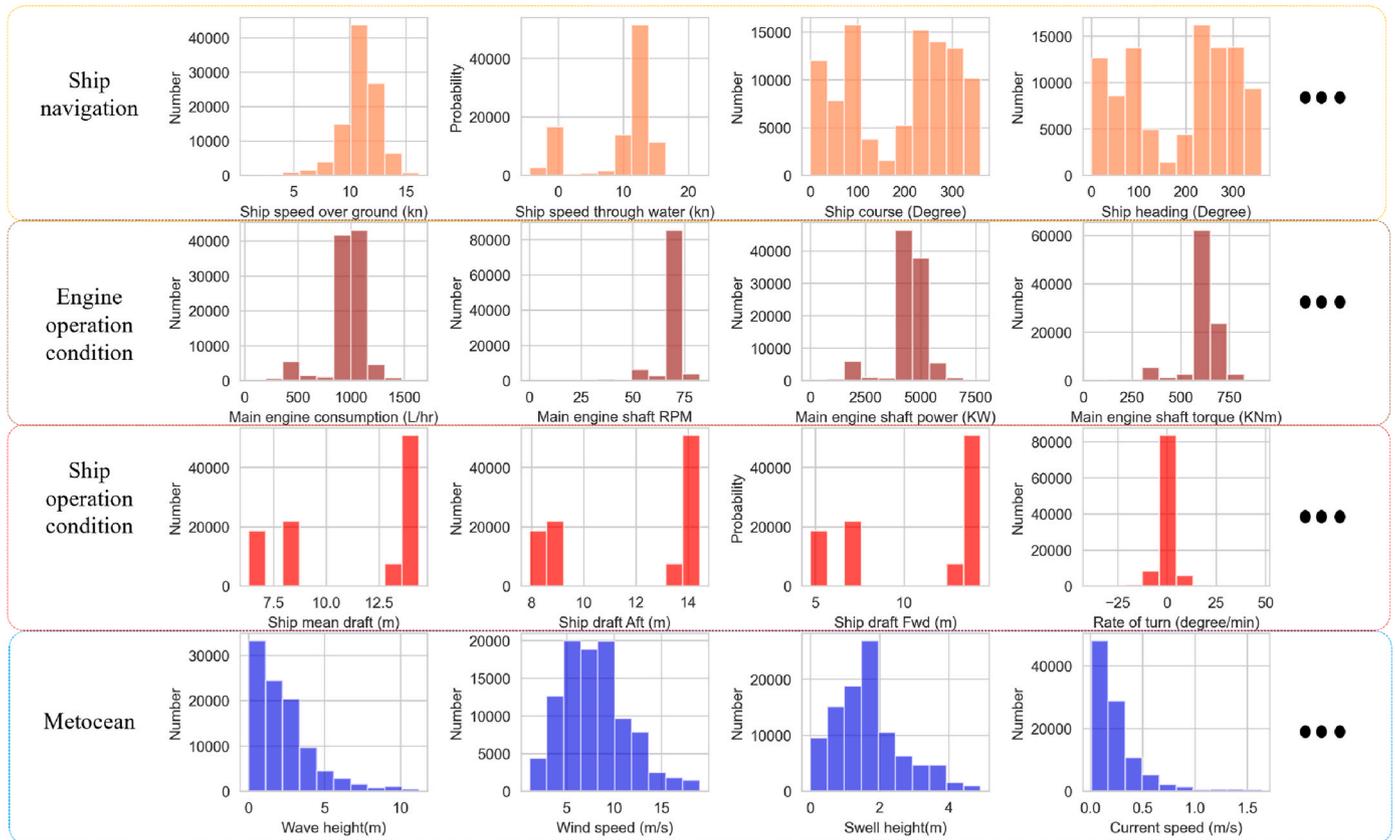
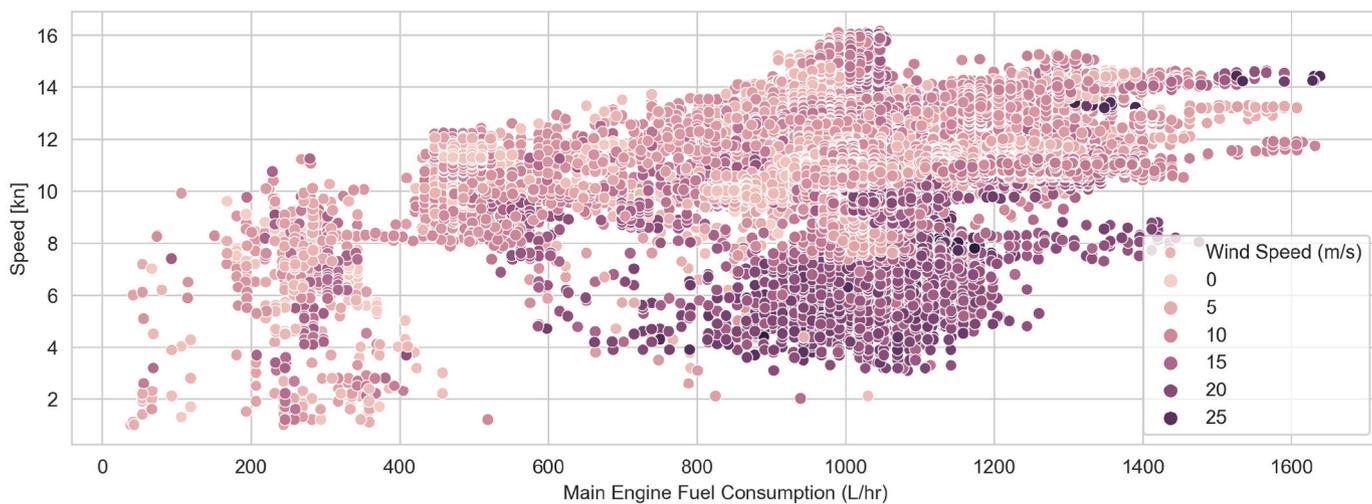
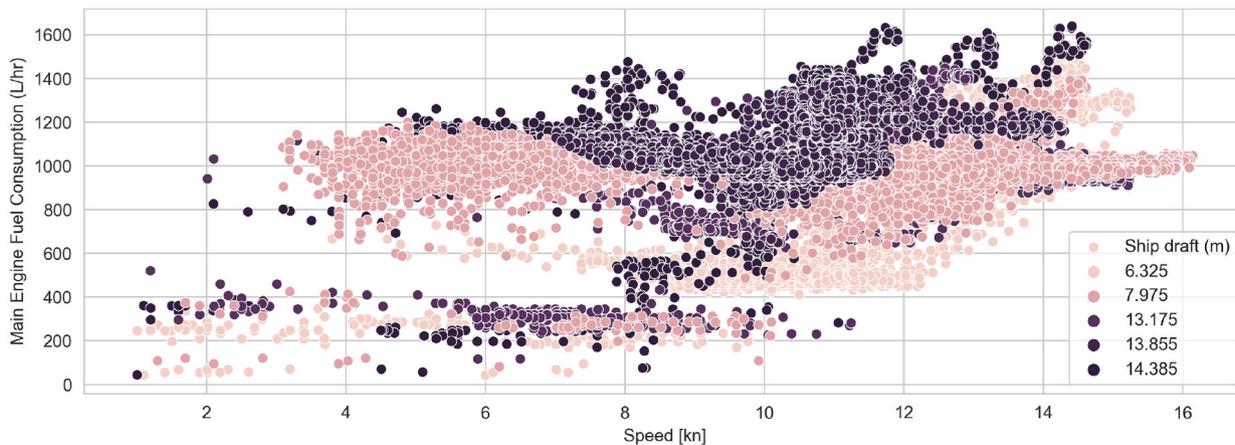


Fig. 10. Visual representation of the collected data samples.

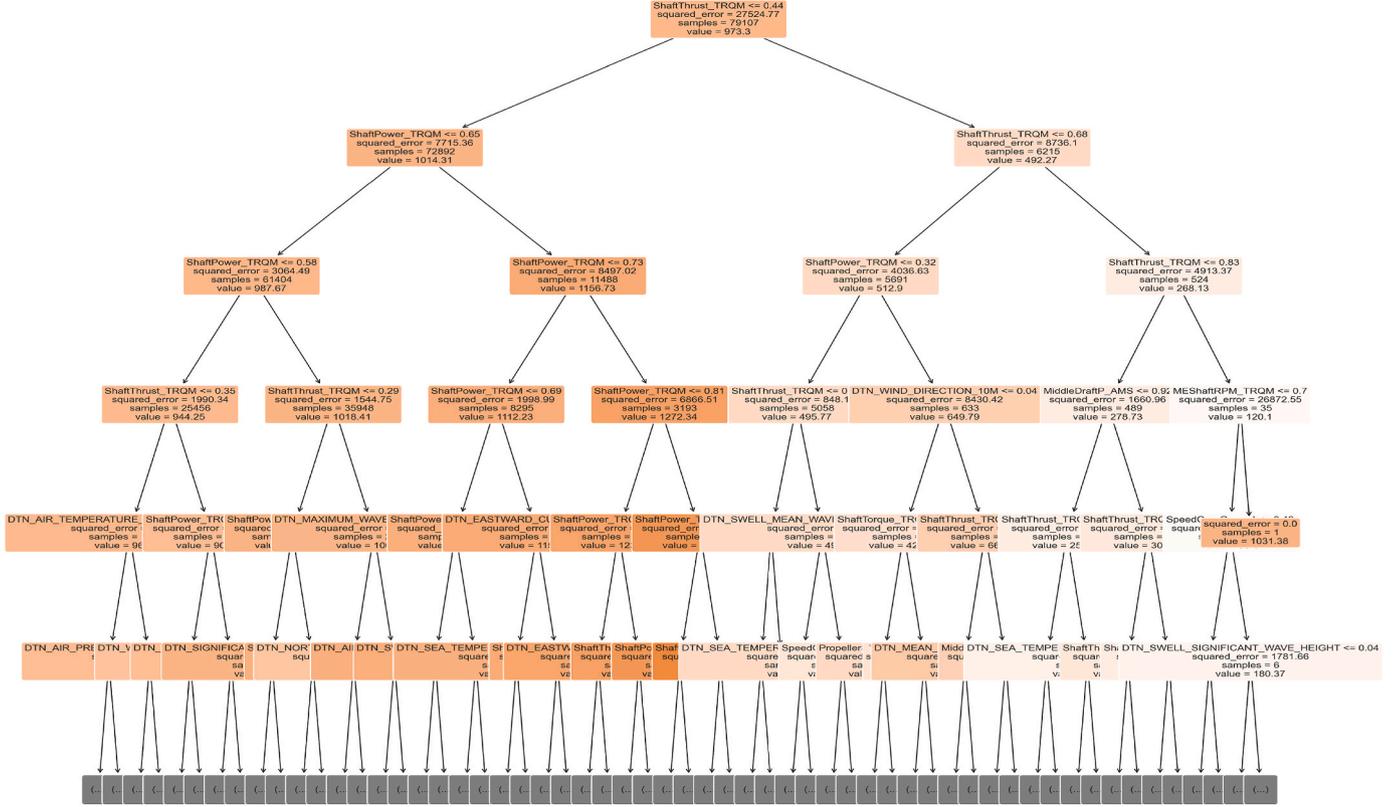


(a) ship fuel consumption distribution under different ship speed and wind speed conditions.



(b) ship fuel consumption distribution under different ship speed and draft conditions.

Fig. 11. The relationship between ship fuel consumption and external conditions (speed, draft, wind).



**Fig. 12.** Visualization of the decision tree model with the best-performing hyperparameters obtained from the grid search (the tree includes 90 levels (depths), and only the first five levels of the tree are displayed here).

vectors) that are typically trained and determined during the process of model training.

#### (iv) Output layer

The output layer receives the context vector, which is the weighted combination of the hidden representations obtained from the attention mechanism process. In turn, the context vector captures the most relevant information from the input data stream. The output layer processes the context vectors and predicts SFC for a given time step, see Fig. 7.

### 2.3. Model evaluation and applications

To evaluate the performance and quantify the errors between the real and the predicted SFC, Root Mean Square Error (RMSE), Mean Square Error (MSE), and error rates  $e_n$  were evaluated, see Eqs. 17–19. In addition, the  $R^2$  value (coefficient of determination) was used to measure the generalization potential of the method, see Eq. (20).

$$RMSE = \sqrt{\frac{1}{N} \sum_{n=1}^N (y_n - \hat{y}_n)^2} \quad (17)$$

$$MSE = \frac{1}{N} \sum_{n=1}^N (y_n - \hat{y}_n)^2 \quad (18)$$

$$e_n = (y_n - \hat{y}_n) / (y_n) \quad (19)$$

$$R^2 = 1 - \frac{\sum_{n=1}^N (y_n - \hat{y}_n)^2}{\sum_{n=1}^N (y_n - \bar{y})^2} \quad (20)$$

In the above expressions,  $y_n$  is the actual value,  $\hat{y}_n$  denotes the predicted value,  $\bar{y}_n$  is the mean value.

The so-called  $k$ -fold cross-validation is a widely employed machine learning method that can be used for the evaluation of model performance (Zhang et al., 2023). It involves partitioning the dataset into  $k$  equally sized subsets or folds, see Fig. 8. The evaluation process is then conducted  $k$  times, with each fold serving as the validation set once. The remaining  $k-1$  folds as the training set which allows for a comprehensive assessment of model performance by iteratively rotating through all the folds. After each iteration, performance metrics such as accuracy or error are calculated. Eventually, results are averaged to obtain a robust estimation of a model's generalization potential.

To compare the adopted model with existing machine learning models, a series of comparisons were conducted using the same data streams. To assess the practical applicability and use the trained model, new input data streams encompassing records from the entire voyages were collected and used for the prediction of SFC. Key findings are summarised in Section 3.

### 3. Case studies

To validate the adopted deep learning method, big data records from extensive sea trials of a Kamsarmax bulk carrier of Laskaridis Shipping Co. Ltd. were used, see Table 3. Fig. 9 illustrates the voyages undertaken over a period of two years (February 2021 to January 2023). The big dataset used comprised of more than 1 million data records, each consisting of 266 parameters as indicated in the parameter word cloud provided in Appendix A. The data collection interval was 60 s.

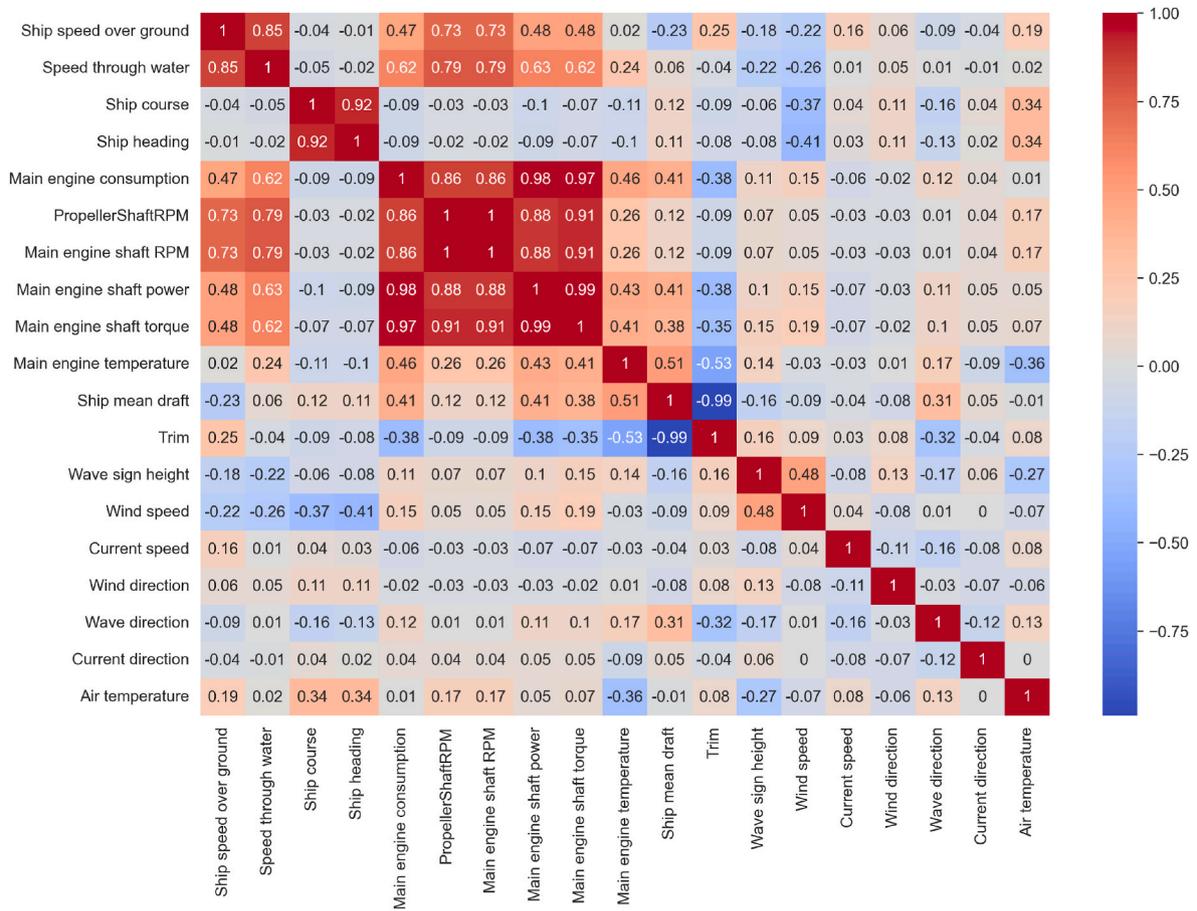


Fig. 13. The correlation relationships between the selected key influencing factors on ship fuel consumption using the collected data streams.

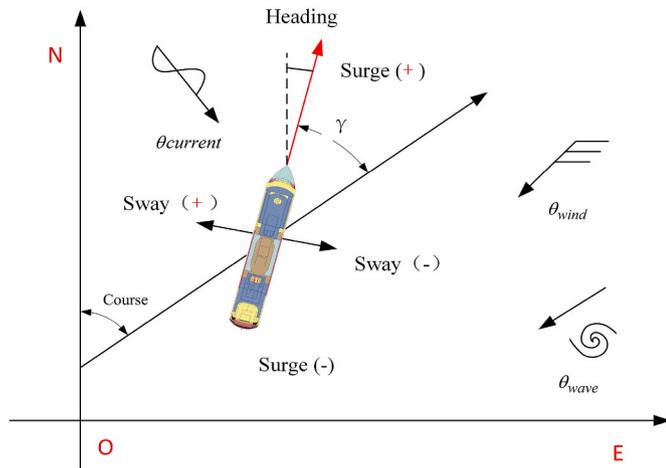


Fig. 14. The relationship between hydrometeorological factors and ship motions in real operational conditions (Zhang et al., 2023b).

### 3.1. Decision tree model for key variables selection

Big data analytics involved acquiring information on ship navigation, engine and ship operational conditions as well as meteocean data, see Fig. 2. At first instance, anomalous data points were carefully identified, and data streams were included if the recorded speed exceeded 1 knot, the SFC exceeded 0 L/h but remained below 2000 L/h, or the main engine shaft were greater than 10 rpm. Approximately 7.5%

of the dataset available for the analysis was deemed aberrant and was subsequently removed. After filtering, the data presented the ship performance during actual operations at sea, provided a realistic representation of the ship performance rather than being limited to periods when the ship was at anchor or in port. A visual representation of the collected data instances is given in Fig. 10.

This comprehensive dataset facilitated for a detailed analysis of SFC encompassing the influences of navigational patterns, engineering systems performance, operational conditions, and the prevailing hydro-meteorological conditions. Fig. 11 illustrates the relationship between SFC and external factors, such as speed, draft, and wind conditions.

To tackle the challenge of identifying the key influencing factors while disregarding irrelevant data, the study employed the DT regression method outlined in Section 2.1. The method accounted for 266 sets of variables that specifically comprised of 265 parameters and SFC. The main aim has been to use time domain signals and extract meaningful insights on predictors that may impact SFC. To achieve this, 265 variables were designated as the X database, while SFC served as the Y database. The X database and Y database were split into training and testing sets, with the training set accounting for 80% of the available data streams and the testing set encompassing the remaining 20%. Eventually big data sets were normalized to facilitate for the training of the DT model.

To determine the most effective combination of hyperparameters for the DT model and assess its performance, the DT regression model was optimized and evaluated using a grid search with cross-validation (Krstajic et al., 2014). Specifically, a hyperparameter grid was constructed with different values for max depth (d), min number of instances per leaf (m), and random state (r) of the DT model. Subsequently, a DT regressor was initialized to serve as the base

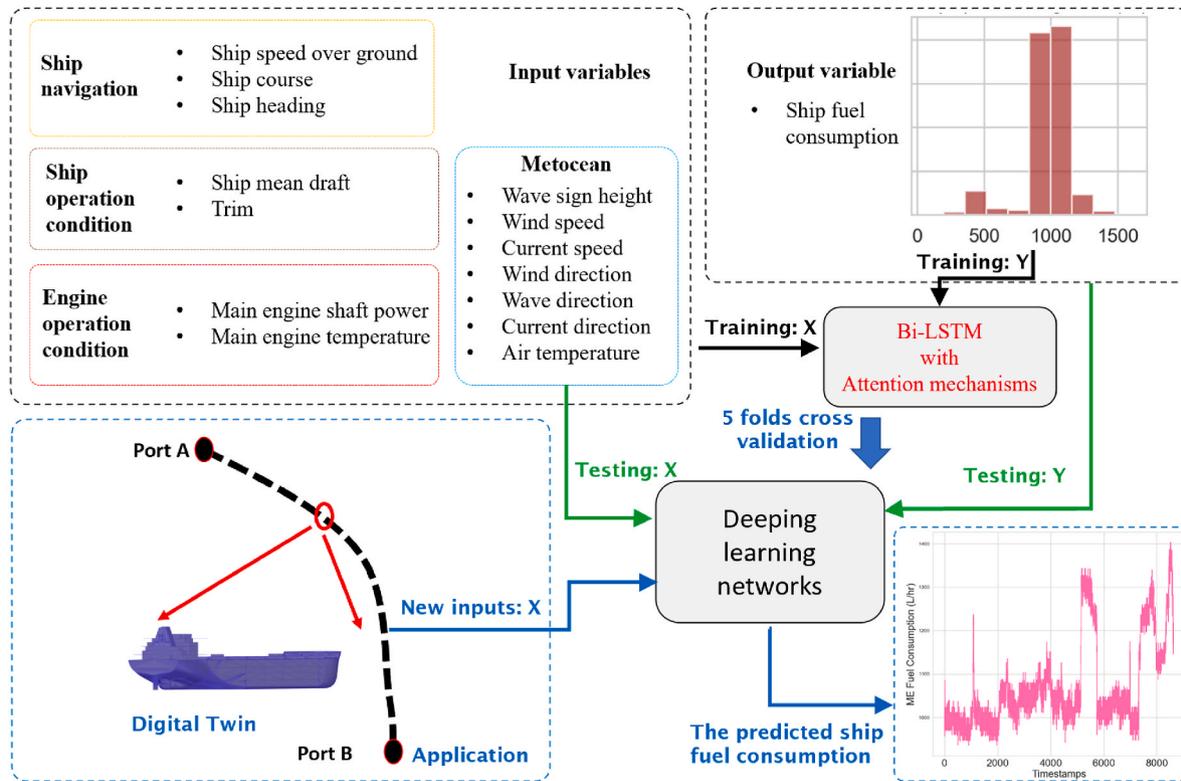


Fig. 15. The deep learning processing of ship fuel consumption prediction for model training, testing and application.

Table 4  
The model characteristics and the optimal hyperparameters.

Model	Input variables	Output variable	Layers
Bi-LSTM with attention mechanisms	14	1	7
Hidden units per layer	Optimizer	Batch Size	Early stopping
128	Adam	48	Patience = 10
Dropout rate	Leaning rate	Epochs	Regularization param
0.2	5e-05	178	0.1

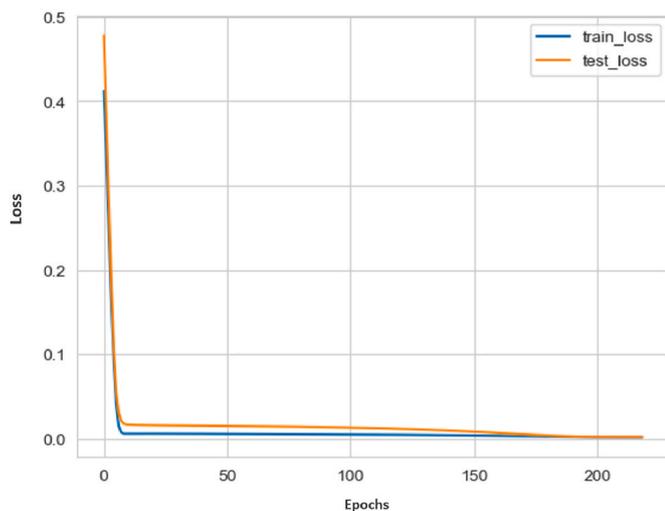


Fig. 16. The model performance evaluation.

estimator for the grid search.

The grid search was performed using the grid search cross validation function (Ranjan et al., 2019), that uses a cross-validation scheme with five folds for the hyperparameters of maximum depth (d), minimum number of instances per leaf (m), and random state (r). During the grid search, the negative MAE was used as key evaluation metric, see Eq. (1). The DT model was trained and evaluated using various hyperparameter combinations. The main aim of the process has been to select the model that achieved the lowest negative MAE.

The best DT model was obtained for  $d = 90$ ,  $m = 4$ , and  $r = 1$ , see Fig. 12. Notably, the subsequent evaluation using MAE allowed for an assessment of the model’s performance on both the training and test sets. The results indicated a MAE of 0.61 on the training set and 1.21 on the testing set. This suggests that the developed model did not experience overfitting.

The optimized DT model and collected dataset helped to assess the importance of variables, see Section 2.1. For example, Fig. 13 illustrates the correlation among key influencing factors and SFC. These findings reveal a robust positive correlation between propeller shaft RPM, main engine shaft RPM, and main engine shaft torque with SFC. A moderate correlation is evident between speed, engine temperature, trim, and SFC. It is noteworthy that ship course, heading, and hydrometeorological factors demonstrate no relevance or even negative correlation against SFC (Yan et al., 2023). This could be attributed to the disparity between the coordinate systems associated to the hydrometeorological factors and ship motions, see Fig. 14. The combination of hydrometeorological factor values (e.g., wave height, wind speed, current speed, etc.) and their absolute directional components (i.e., absolute wave, wind, and current directions) is of key importance for the determination of weather routing and not necessarily forecasting. Notwithstanding this, it is important to note that Fig. 13 employs a correlation method exclusively designed to assess linear relationships between the selected key influencing factors and SFC. To account for potential nonlinear relationships the use of a DT model is critical.

The hydrometeorological conditions data used is provided by a

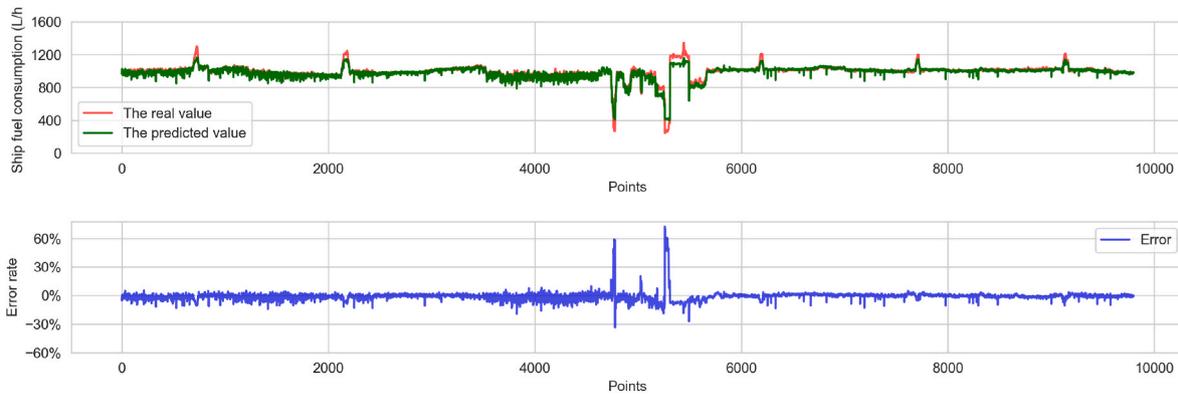


Fig. 17. The results of ship fuel consumption prediction (In the upper figure, the red line represents the real values of ship fuel consumption, while the green line represents the predicted results. In the bottom figure, the blue line represents the error rate in the time domain.)

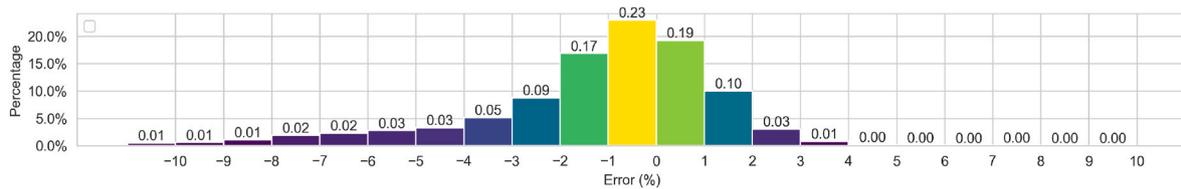


Fig. 18. The analysis of prediction errors using the adopted model.

Table 5

The accuracy evaluation for ship fuel consumption prediction using various tools.

Methods		MSE: Mean Square Error					
		Fold 1	Fold 2	Fold 3	Fold 4	Fold 5	Mean
Machine learning methods	Ada boosting	0.0340	0.0217	0.0326	0.0264	0.0275	0.0283
	DT	0.0275	0.0263	0.0287	0.0279	0.0268	0.0272
	GPR	0.0302	0.0288	0.0295	0.0281	0.0299	0.0291
	XG Boost	0.0305	0.0281	0.0292	0.0287	0.0286	0.0288
	LASSO	0.0315	0.0310	0.0302	0.0308	0.0305	0.0307
	MLR	0.0226	0.0240	0.0238	0.0233	0.0229	0.0338
	RF	0.0230	0.0239	0.0237	0.0233	0.0235	0.0234
	SVR	0.0272	0.0259	0.0276	0.0261	0.0269	0.0264
Deep learning methods	LSTM	0.0245	0.0238	0.0241	0.0237	0.0243	0.0240
	GRU	0.0227	0.0219	0.0222	0.0220	0.0223	0.0221
	ANN	0.0273	0.0276	0.0272	0.0273	0.0275	0.0274
	Bi-LSTM	0.0258	0.0263	0.0260	0.0262	0.0260	0.0261
	LSTM with attention	0.0222	0.0223	0.0219	0.0221	0.0220	0.0220
	Bi-LSTM with attention (The adopted method)	0.0198	0.0206	0.0202	0.0203	0.0205	0.0204

weather provider and represents absolute directions. This is because the primary goal of the SFC prediction model is to optimize routing and absolute directions are more useful for weather routing based on weather forecasting. Furthermore, a DLM can capture the relationship between ship heading and absolute directions of weather conditions as well as the combined effects on SFC. Therefore, it is not necessary to adjust them by subtracting the ship heading to make them relative in DLM.

### 3.2. Ship fuel consumption prediction using the deep learning method

Based on the collected data streams (see Fig. 2), a DT model was utilized to rank the variables and identify the key influencing factors on ship fuel consumption (Fig. 12). In this section, an attention mechanism-based Bi-LSTM architecture was designed to train a DLM.

During maritime operations, ship navigation information (speed, course, heading), ship operation condition information (draft, trim), engine operation information (propeller and main engine shaft RPM and

torque), and external operational conditions serve as the primary control parameters or influencing factors for managing the ship energy and navigation systems. The propeller and main engine shaft RPM, and torque are closely related and exhibit interdependencies, see Fig. 13. Hence, they are not included as input parameters (see Fig. 15).

To train the DLM, the paper employed an architecture consisting of an input layer, three bi-LSTM layers, three attention layers, and an output layer. The Bi-LSTM layer had 128 hidden units. To optimize the hyperparameters of the Bi-LSTM with an attention mechanism model.

(e.g., regularization parameter, dropout rate, and learning rate), a grid search was conducted using the grid search cross Validation function with a 5-fold cross-validation scheme (Ranjan et al., 2019; Krstajic et al., 2014; Zhang et al., 2023). The best parameters were then determined based on the lowest MSE obtained during validation. The model characteristics and the selected hyperparameters are summarised in Table 4.

The dataset was divided using 5-fold cross-validation. The training and validation losses were computed based on the basis of training 80%

```

Save model: model.save('Ship fuel consumption model.h5')
Load model: loaded_model = load_model('Ship fuel consumption model.h5')
Use the model: Predictions = loaded_model.predict(New_inputs)
    
```

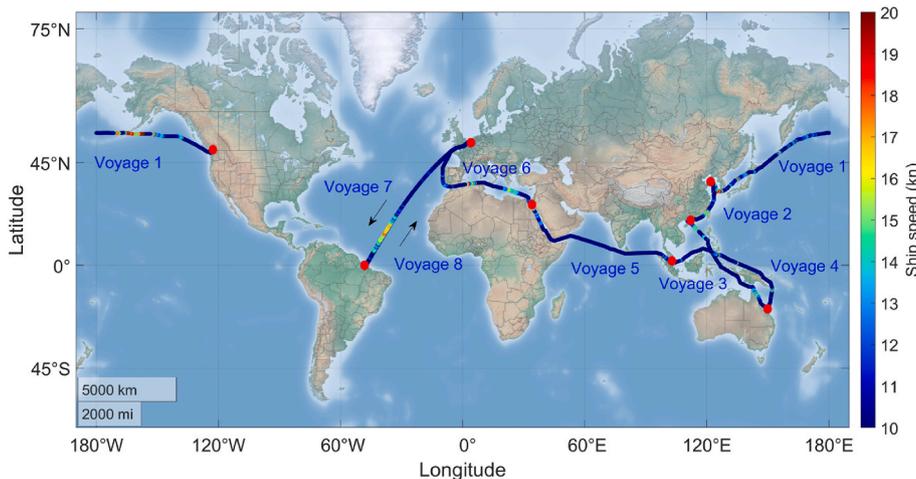


Fig. 19. Ship fuel consumption prediction by calling the trained model.. (Ship trajectories of new data of a bulk carrier from 01.2023 to 06.2023.)

**Table 6**  
The general information of 8 whole voyages.

	Condition	Departure/ destination	Voyage lengths	Number of data
Voyage 1	Laden condition	Vancouver/Yantai	7223.90 nm	28,479
Voyage 2	Ballast condition	Yantai/Hong Kong	980.22 nm	6081
Voyage 3	Laden condition	Hong Kong/Sarina	3035.639 nm	17,347
Voyage 4	Ballast condition	Sarina/Singapore	3764.36 nm	19,545
Voyage 5	Laden condition	Singapore/Pouad	4724.23 nm	19,749
Voyage 6	Laden condition	Pouad/Rotterdam	3954.41 nm	8299
Voyage 7	Ballast condition	Rotterdam/Belem	4230.67 nm	19,930
Voyage 8	Laden condition	Belem/Rotterdam	4197.65 nm	21,206

and 20% of the dataset records respectively. The curves displayed in Fig. 16 demonstrate that both the training and validation losses decreased and stabilized around the 178th epoch. Consequently, if no improvement in the validation performance is observed beyond this point, the training process can be terminated early to mitigate overfitting. For the case study considered upon the completion of 178th epochs, the deep learning model achieved a state of optimal fitting. This suggests balanced convergence between the model’s performance and the training data. Optimal fitting also implies that the model neither exhibits symptoms of overfitting, nor is characterized by excessive complexity. In this sense, it attains a desirable equilibrium between capturing any big data intricacies.

As shown in Fig. 16, the training process suggests that the DLM may achieve an optimal fit, by effectively mitigating both over- and underfitting. Moreover, the MSE obtained by 5-fold cross-validation was 2.04e-2. To further validate the model, new inputs were selected from the testing database illustrated in Fig. 15 and error rates were computed by using Eq. (19), see Fig. 17. These findings signify that the trained DLM can effectively capture the ship energy system’s characteristics under real operational conditions. Additionally, errors are depicted in

Fig. 18. The analysis revealed that over 90% of the prediction errors are below 4%, and the average error rate was 0.98%. Perhaps it is crucial to highlight that the adopted model bears some limitations in terms of effectively capturing abnormal fluctuations present in sensor collected data, see the peak values on the blue line outlined in Fig. 17. Ideally, such onerous effects should be removed during data collection.

### 3.3. Validation and comparisons

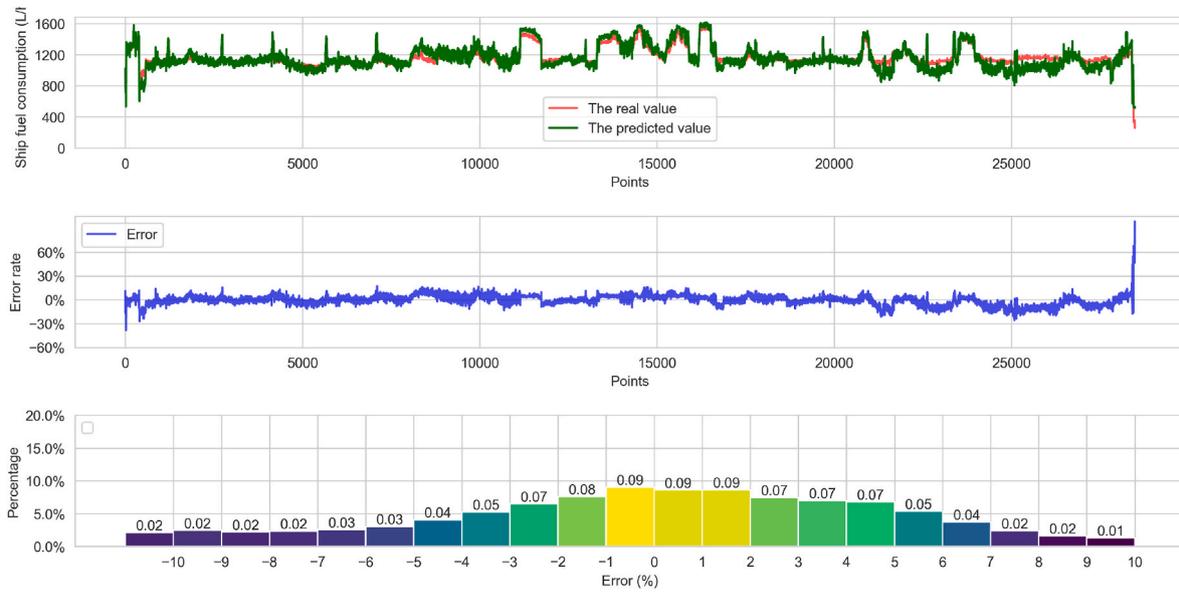
A comparison of the method with existing models is presented in Table 5. The average MSE was calculated for the prediction of SFC using the methods, i.e., Ada boosting, DT, GPR, XG Boost, LASSO, GRU, MLR, RF, SVR, GRU, LSTM, Bi-LSTM, ANN, LSTM with attention mechanism, see Table 1. Whereas most DLMs demonstrate superior performance in predicting SFC as compared to traditional MLMs, and the errors of the new method appear to be the lowest. Possible reasons behind this observation are.

- The big data collection system onboard the vessel provided high quality data streams over long-term sea trials, while the importance evaluation model of influencing factors selects the key features for the development of deep learning model.
- The utilization of the attention mechanism and Bi-LSTM enhanced the DLM learning capability. This is because it allows for the long term bi-directional and simultaneous processing of multiple data streams.
- By effectively capturing the dependencies and important information available in data streams, the adopted method can be used to better understand and represent the complex relationships involved in ship energy systems.

### 3.4. Generalization assessment

To generalise the applicability of the method, the trained model was tested by big data streams from operational data corresponding to the period from February 2023 to June 2023. The new data encompassed of 8 worldwide voyages, see Fig. 19. The longest and shortest trajectories were 7223.9 nautical miles and 980 nautical miles respectively, see Table 6.

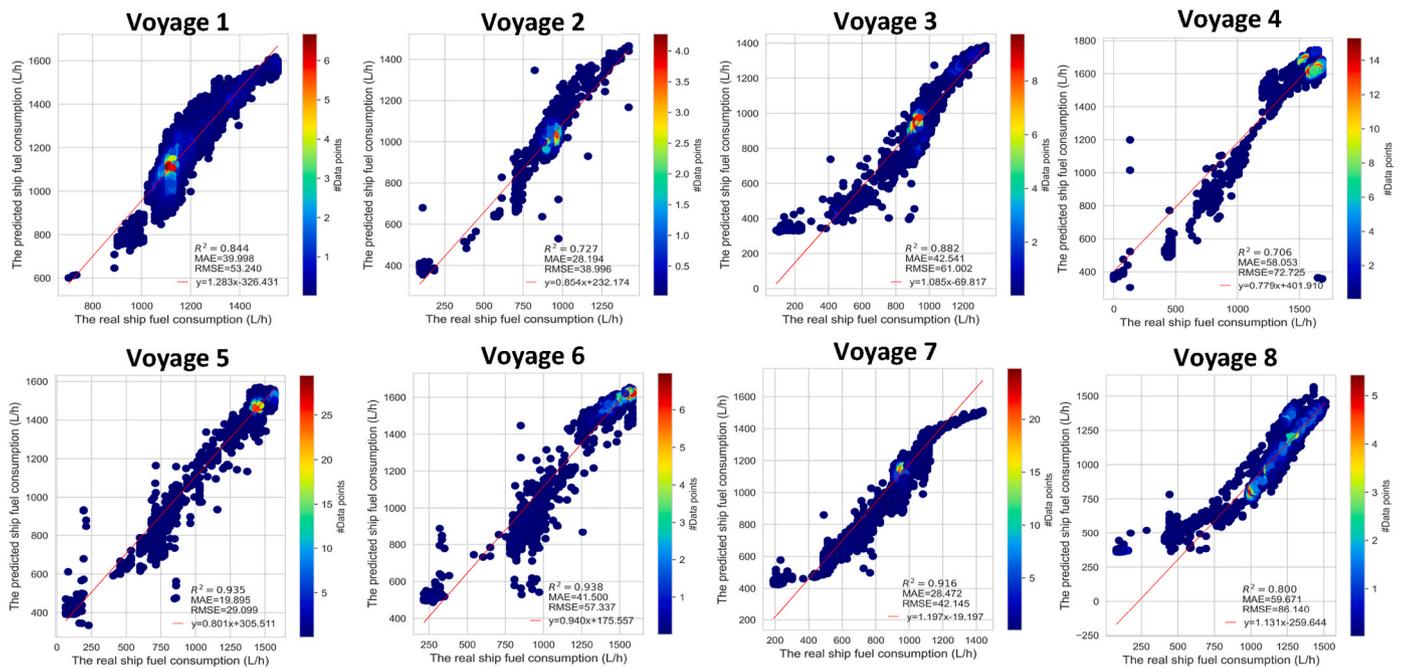
As shown in Fig. 19, the SFC over the total number of voyages was predicted by using the profile of the trained model in the time domain.



**Fig. 20.** The error analysis of ship fuel consumption prediction for a whole voyage 1 (In the upper figure, the red line represents the real values of ship fuel consumption, while the green line represents the predicted results. In the middle figure, the blue line represents the error rate in the time domain. The bottom figure presents the prediction error distributions.).

**Table 7**  
Generalization ability evaluation for different global voyages.

	Voyage 1	Voyage 2	Voyage 3	Voyage 4	Voyage 5	Voyage 6	Voyage 7	Voyage 8
$R^2$	0.84	0.73	0.88	0.71	0.94	0.94	0.92	0.80
RMSE (L/h)	53.24	38.99	61.02	72.72	29.09	57.33	42.14	86.14
MAE (L/h)	39.99	28.19	42.54	58.05	19.89	41.50	28.47	59.67
Average error rate	-0.51%	2.64%	-2.68%	5.56%	2.53%	1.36%	2.48%	-3.59%



**Fig. 21.** The comparison of the real and the predicted ship fuel consumption.

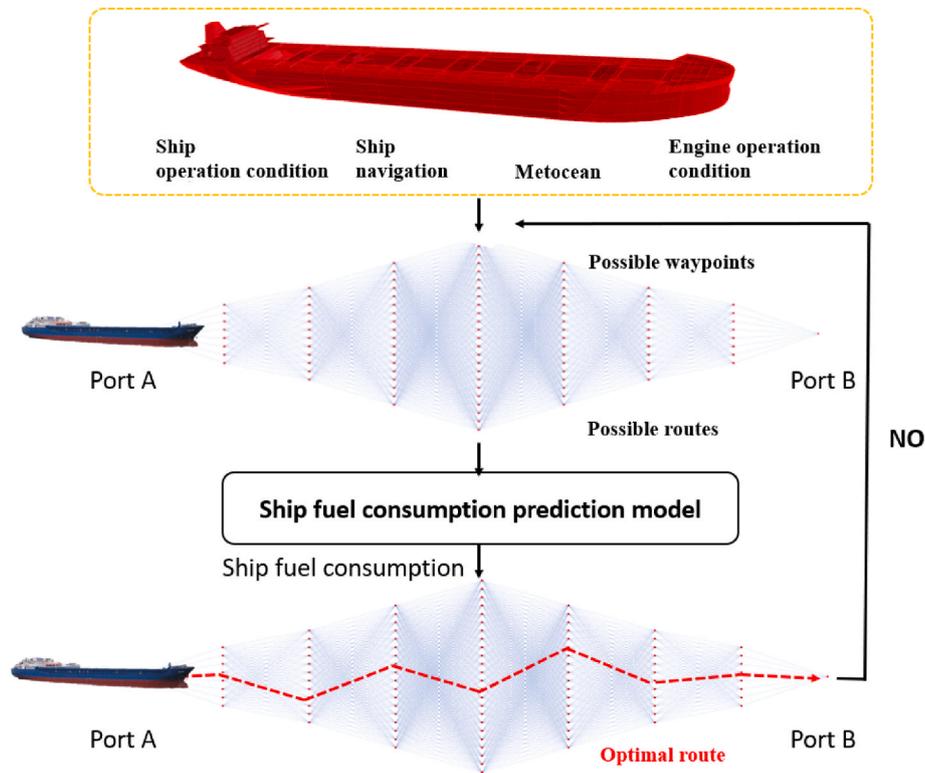


Fig. 22. A flowchart of multi-objective optimization method for ship fuel consumption reduction based on prediction model.

Fig. 20 highlights that more than 90 % of the prediction errors were below 10 %, and the average error for voyage 1 was 0.51%. The evaluation metrics  $R^2$ , RMSE, MAE and the average error rate were used to present the differences between the real and the predicted SFC values, see Table 7, Fig. 21 and Fig. B1 in Appendix B. The analysis indicated that the  $R^2$  values range from 0.71 to 0.94. Voyage 2 exhibits the smallest MAE at 28.19 L/h, and the RMSE for voyage 2 is also the lowest at 38.99 L/h. The average error rate values range from  $-0.51\%$  to  $5.56\%$  and, the smallest average error rate of  $-0.51\%$  is observed in voyage 1, see Fig. B1.

These findings affirm the viability of deploying the trained model as an efficient tool for forecasting SFC during comparable voyages. Consequently, the application of the trained model holds the potential to furnish valuable insights and facilitate efficient fuel management and optimization endeavours in real operational conditions.

#### 4. Future works

The SFC prediction model presented in this paper may be further developed to inform intelligent decision support tools for use in fuel efficient shipping operations of new build or retrofitted ships (Lu et al., 2015; Wang et al., 2021; Vitali et al., 2020). SFC is influenced by various interconnected factors, including ship operational conditions, navigation information, engine performance, and weather conditions (Vinayak et al., 2021; Wang et al., 2021). Hence, optimization when the analysis accounts for only an individual ship or limited operational factors may

be hard to achieve (see Section 3.1). It is envisaged that future work should aim to (i) determine optimal ship operation commands, (ii) identify the optimal routing and (iii) suggest decision support criteria for realistic SFC when different green ship technologies are retrofitted or introduced at new building stage (see Section 3 and Fig. 22). Within this context a multi-objective optimization process could be developed to ensure unification of criteria and methods for use in advanced decision support systems.

#### 5. Conclusions

This paper presented a DLM that may be used for the prediction of SFC in real operational conditions. The proposed framework considered (1) evaluating the importance of influencing factors and identifying key factors that may impact SFC, (2) the training of deep learning neural networks to effectively capture the energy system characteristics of ships operating in real operational conditions, and (3) long-term shipping operations. To validate the method, big data records of two years of operation for a Kamsarmax bulk carrier of Laskaridis Shipping Co. Ltd were used. Subsequently, a comprehensive comparison with existing methods was conducted to demonstrate the effectiveness of the DLM using a Bi-LSTM with attention mechanism. A more recent dataset encompassing data from 8 international voyages over February to June 2023 was also employed to assess the generalization potential of the model. Key conclusions can be summarised as follows.

- The innovative use of a high frequency data driven method that combines DT and Bi-LSTM models with attention mechanism for predicting SFC in the time domain is promising.
- DT methods are valuable in terms of capturing potentially nonlinear relationships between influencing factors and SFC. Notwithstanding this, it is critical to appreciate that shipping operations may involve uncertainties associated with the quality and availability of navigation data, operational conditions, engine parameters, and other unknown external factors.
- A comprehensive comparison of existing methods suggested that, in most cases, DLM may outperform traditional machine learning methods, see Table 5. The latter could be attributed to the complex and nonlinear nature of the relationships of data that can be processed by modern methods.
- The Bi-LSTM model with attention mechanisms stands out as the best choice for predicting SFC. This is because it can effectively capture ship energy system characteristics, while accounting for real operational conditions over long term operations, see Fig. 21 and B1 and Table 7.

Leveraging of big data records and DLM may assist with ship voyage optimization, fuel savings and emission reductions. This may benefit the environment but also contribute to the environmentally sustainable

operations of new-build vessels or ships retrofitted with green and renew-able technologies.

**Declaration of competing interest**

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

**Data availability**

The authors do not have permission to share data.

**Acknowledgment**

All authors acknowledge research funding received funding from the Horizons Europe project” RETROFIT solutions to achieve 55% GHG reduction by 2030 (RETROFIT55) – Project No.: 101096068”. Mingyang Zhang and Spyros Hirdaris express their gratitude for selected sponsorships received from Merenkulun Säätiö. Special thanks of appreciation go to CSC Finland for the provision of their parallel computing facilities. The views set out in this paper are those of the authors and do not necessarily reflect the views of their sponsors.

**Appendix A. Word cloud of 266 parameters of the collected data streams**



Fig. A1. Word cloud of 266 parameters of the collected data streams from sea trials.

**Appendix B. The error analysis of ship fuel consumption prediction for 8 whole voyages**

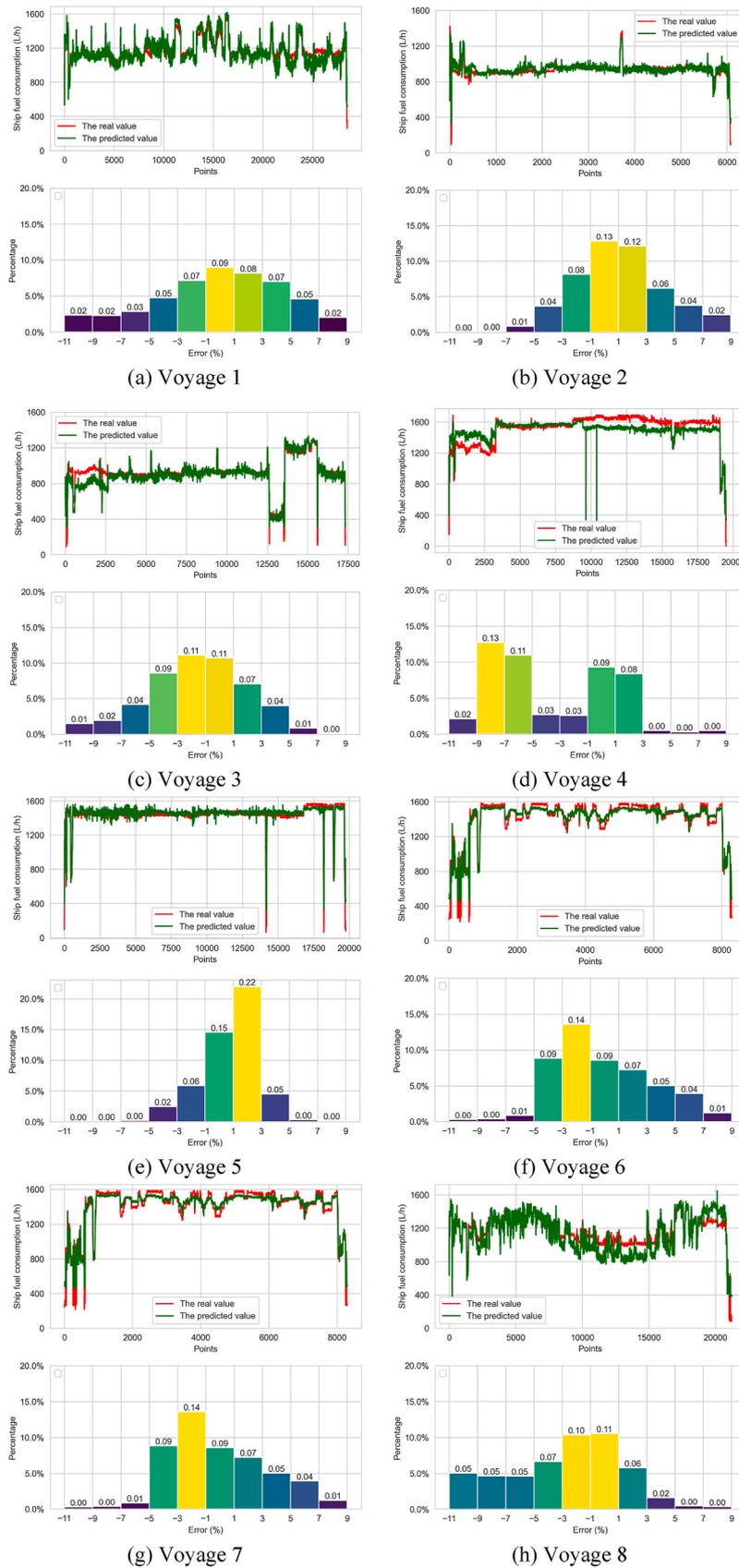


Fig. B1. The error analysis of ship fuel consumption prediction for 8 whole voyages.

## References

- Abreu, L.R., Maciel, I.S., Alves, J.S., Braga, L.C., Pontes, H.L., 2023. A decision tree model for the prediction of the stay time of ships in Brazilian ports. *Eng. Appl. Artif. Intell.* 117, 105634.
- Bocchetti, D., Lepore, A., Palumbo, B., Vitiello, L., 2013. A statistical control of the ship fuel consumption. In: *Proceedings, International Conference on the Design, Construction and Operation of Passenger Ships*. royal institution of naval architects, November, pp. 20–21.
- Cai, J., Luo, J., Wang, S., Yang, S., 2018. Feature selection in machine learning: a new perspective. *Neurocomputing* 300, 70–79.
- Campbell, R., Terziev, M., Tezdogan, T., Incecik, A., 2022. Computational fluid dynamics predictions of draught and trim variations on ship resistance in confined waters. *Appl. Ocean Res.* 126, 103301.
- Carlton, J., 2018. *Marine Propellers and Propulsion*. Butterworth-Heinemann.
- Chen, Z.S., Lam, J.S.L., Xiao, Z., 2023. Prediction of harbour vessel fuel consumption based on machine learning approach. *Ocean Eng.* 278, 114483.
- Cho, J.H., Lee, S.H., Oh, D., Paik, K.J., 2023. A numerical study on the added resistance and motion of a ship in bow quartering waves using a soft spring system. *Ocean Eng.* 280, 114620.
- Choi, J.E., Min, K.S., Kim, J.H., Lee, S.B., Seo, H.W., 2010. Resistance and propulsion characteristics of various commercial ships based on CFD results. *Ocean Eng.* 37 (7), 549–566.
- Chorowski, J.K., Bahdanau, D., Serdyuk, D., Cho, K., Bengio, Y., 2015. Attention-based models for speech recognition. *Adv. Neural Inf. Process. Syst.* 28.
- Coraddu, A., Oneto, L., Baldi, F., Anguita, D., 2017. Vessels fuel consumption forecast and trim optimisation: a data analytics perspective. *Ocean Eng.* 130, 351–370.
- Dai, K., Li, Y., Gong, J., Fu, Z., Li, A., Zhang, D., 2022. Numerical study on propulsive factors in regular head and oblique waves. *Brodogradnja: Teorija i praksa brodogradnje i pomorske tehnike* 73 (1), 37–56.
- Du, Y., Chen, Y., Li, X., Schönborn, A., Sun, Z., 2022a. Data fusion and machine learning for ship fuel efficiency modeling: Part II—Voyage report data, AIS data and meteorological data. *Communications in Transportation Research* 2, 100073.
- Du, Y., Chen, Y., Li, X., Schönborn, A., Sun, Z., 2022b. Data fusion and machine learning for ship fuel efficiency modeling: Part III—Sensor data and meteorological data. *Communications in Transportation Research* 2, 100072.
- Elkafas, A.G., Elgohary, M.M., Zeid, A.E., 2019. Numerical study on the hydrodynamic drag force of a container ship model. *Alex. Eng. J.* 58 (3), 849–859.
- Fan, A., Yan, X., Bucknall, R., Yin, Q., Ji, S., Liu, Y., et al., 2020. A novel ship energy efficiency model considering random environmental parameters. *Journal of Marine Engineering & Technology* 19 (4), 215–228.
- Fan, A., Yang, J., Yang, L., Wu, D., Vladimir, N., 2022. A review of ship fuel consumption models. *Ocean Eng.* 264, 112405.
- Farkas, A., Degiuli, N., Martić, I., Ančić, I., 2020. Performance prediction method for fouled surfaces. *Appl. Ocean Res.* 99, 102151.
- Fiskin, R., Cakir, E., Sevgili, C., 2021. Decision tree and logistic regression analysis to explore factors contributing to harbour tugboat accidents. *J. Navig.* 74 (1), 79–104.
- Fujii, H., Takahashi, T., 1975. Experimental study on the resistance increase of a large full ship in regular oblique waves. *J. Soc. Nav. Archit. Jpn.* 1975 (137), 132–137.
- Gao, R., Du, L., Yuen, K.F., 2020. Robust empirical wavelet fuzzy cognitive map for time series forecasting. *Eng. Appl. Artif. Intell.* 96, 103978.
- Gao, R., Li, R., Hu, M., Suganthan, P.N., Yuen, K.F., 2023. Significant wave height forecasting using hybrid ensemble deep randomized networks with neurons pruning. *Eng. Appl. Artif. Intell.* 117, 105535.
- Grlij, C.G., Degiuli, N., Tuković, Z., Farkas, A., Martić, I., 2023. The effect of loading conditions and ship speed on the wind and air resistance of a containership. *Ocean Eng.* 273, 113991.
- Haase, M., Zurcher, K., Davidson, G., Binns, J.R., Thomas, G., Bose, N., 2016. Novel CFD-based full-scale resistance prediction for large medium-speed catamarans. *Ocean Eng.* 111, 198–208.
- Hall, M.A., 1999. *Correlation-based Feature Selection for Machine Learning*. Doctoral dissertation, The University of Waikato.
- Haranen, M., Pakkanen, P., Kariranta, R., Salo, J., 2016. White, grey and black-box modelling in ship performance evaluation. In: *1st Hull Performance & Insight Conference (HullPIC)*, pp. 115–127.
- Hasselmann, K., Barnett, T.P., Bouws, E., Carlson, H., Cartwright, D.E., Enke, K., et al., 1973. Measurements of wind-wave growth and swell decay during the joint north sea wave project (JONSWAP). *Ergaenzungsheft zur Deutschen Hydrographischen Zeitschrift. Reihe A*.
- Holtrop, J., Mennen, G.G.J., 1982. An approximate power prediction method. *Int. Shipbuild. Prog.* 29 (335), 166–170.
- Hu, Z., Jin, Y., Hu, Q., Sen, S., Zhou, T., Osman, M.T., 2019. Prediction of fuel consumption for enroute ship based on machine learning. *IEEE Access* 7, 119497–119505.
- Huang, L., Pena, B., Liu, Y., Anderlini, E., 2022. Machine learning in sustainable ship design and operation: a review. *Ocean Eng.* 266, 112907.
- IMO, 2020a. *Fourth Greenhouse Gas Study 2020*.
- IMO, 2020b. *Just in time arrival guide barriers and potential solutions*. <https://greenvoyage2050.imo.org/wp-content/uploads/2021/01/GIA-just-in-time-hires.pdf>. (Accessed 15 June 2023).
- ISO, 2015. *Marine Technology—Guidelines for the Assessment of Speed and Power Performance by Analysis of Speed Trial Data*. ISO, Geneva, Switzerland, 15016.
- ITTC, 2002. *Resistance uncertainty analysis, example for resistance test*. In: *Venice, Italy: Proceedings of the 23th International Towing Tank Conference (ITTC)*.
- ITTC, 2014. *Analysis of speed/power trial data. Recommended procedures and guidelines* 7, 5-04-01-01.2.
- ITTC, 2017. *Seakeeping Experiments. Recommended Procedures and Guidelines*, vol. 7, 5-02-07-02.1.
- Jeon, M., Noh, Y., Shin, Y., Lim, O.K., Lee, I., Cho, D., 2018. Prediction of ship fuel consumption by using an artificial neural network. *J. Mech. Sci. Technol.* 32, 5785–5796.
- Jinkine, V., Ferdinande, V., 1974. A method for predicting the added resistance of fast cargo ships in head waves. *Int. Shipbuild. Prog.* 21 (238), 149–167.
- Julianto, R.I., Prabowo, A.R., Muhayat, N., Putranto, T., Adiputra, R., 2021. Investigation of hull design to quantify resistance criteria using Holtrop's regression based method and Savitsky's mathematical model: a study case of fishing vessels. *J. Eng. Sci. Technol.* 16 (2), 1426–1443.
- Karabadjji, N.E.I., Seridi, H., Khelif, I., Azizi, N., Boulkroune, R., 2014. Improved decision tree construction based on attribute selection and data sampling for fault diagnosis in rotating machines. *Eng. Appl. Artif. Intell.* 35, 71–83.
- Kazemitabar, J., Amini, A., Bloniarz, A., Talwalkar, A.S., 2017. Variable importance using decision trees. *Adv. Neural Inf. Process. Syst.* 30.
- Kevin, N.U.S.A., Kodak, G., 2023. Comparison of maritime and road transportations in emissions perspective: a review article. *Int. J. Electron. Govern.* 10 (2), 48–60.
- Kim, M., Hizir, O., Turan, O., Day, S., Incecik, A., 2017. Estimation of added resistance and ship speed loss in a seaway. *Ocean Eng.* 141, 465–476.
- Kim, Y.R., Jung, M., Park, J.B., 2021. Development of a fuel consumption prediction model based on machine learning using ship in-service data. *J. Mar. Sci. Eng.* 9 (2), 137.
- Kim, Y.R., Steen, S., Kramel, D., Muri, H., Strømman, A.H., 2023. Modelling of ship resistance and power consumption for the global fleet: the MariTEAM model. *Ocean Eng.* 281, 114758.
- Kotsiantis, S.B., 2013. Decision trees: a recent overview. *Artif. Intell. Rev.* 39, 261–283.
- Krstajic, D., Buturovic, L.J., Leahy, D.E., Thomas, S., 2014. Cross-validation pitfalls when selecting and assessing regression and classification models. *J. Cheminf.* 6, 1–15.
- Lang, X., 2023. *Data-driven Ship Performance Models—Emphasis on Energy Efficiency and Fatigue Safety* (Doctoral Dissertation, Chalmers Tekniska Högskola (Sweden)).
- Lang, X., Mao, W., 2020. A semi-empirical model for ship speed loss prediction at head sea and its validation by full-scale measurements. *Ocean Eng.* 209, 107494.
- Lewis, E.V., 1988. *Principles of naval architecture second revision*. Jersey: Sname 2, 152–157.
- Li, H., Yang, Z., 2023. Incorporation of AIS data-based machine learning into unsupervised route planning for maritime autonomous surface ships. *Transport. Res. E Logist. Transport. Rev.* 176, 103171.
- Li, H., Jiao, H., Yang, Z., 2023a. AIS data-driven ship trajectory prediction modelling and analysis based on machine learning and deep learning methods. *Transport. Res. E Logist. Transport. Rev.* 175, 103152.
- Li, L., Jin, F., Huang, D., Wang, G., 2023b. Soil seismic response modeling of KiK-net downhole array sites with CNN and LSTM networks. *Eng. Appl. Artif. Intell.* 121, 105990.
- 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.
- Lin, J.C.W., Shao, Y., Zhou, Y., Pirouz, M., Chen, H.C., 2019. A Bi-LSTM mention hypergraph model with encoding schema for mention extraction. *Eng. Appl. Artif. Intell.* 85, 175–181.
- Liu, S., Papanikolaou, A., 2020. Regression analysis of experimental data for added resistance in waves of arbitrary heading and development of a semi-empirical formula. *Ocean Eng.* 206.
- Liu, S., Shang, B., Papanikolaou, A., Bolbot, V., 2016. Improved formula for estimating added resistance of ships in engineering applications. *J. Mar. Sci. Appl.* 15 (4), 442–451.
- Liu, Z., Liu, W., Chen, Q., Luo, F., Zhai, S., 2020. Resistance reduction technology research of high speed ships based on a new type of bow appendage. *Ocean Eng.* 206, 107246.
- Lu, R., Turan, O., Boulougouris, E., Banks, C., Incecik, A., 2015. A semi-empirical ship operational performance prediction model for voyage optimization towards energy efficient shipping. *Ocean Eng.* 110, 18–28.
- Luo, S., Ma, N., Hirakawa, Y., 2016. Evaluation of resistance increase and speed loss of a ship in wind and waves. *J. Ocean Eng. Sci.* 1 (3), 212–218.
- !count(/sb:host[1]/child::\*//sb:date)">Ma, J., Jia, C., Yang, X., Cheng, X., Li, W., Zhang, C., . A data-driven approach for collision risk early warning in vessel encounter situations using attention-BiLSTM. *IEEE Access* 8, 188771–188783.
- Ma, J., Xia, D., Wang, Y., Niu, X., Jiang, S., Liu, Z., Guo, H., 2022. A comprehensive comparison among metaheuristics (MHs) for geohazard modeling using machine learning: insights from a case study of landslide displacement prediction. *Eng. Appl. Artif. Intell.* 114, 105150.
- Min, K.S., Kang, S.H., 2010. Study on the form factor and full-scale ship resistance prediction method. *J. Mar. Sci. Technol.* 15, 108–118.
- Molland, A.F., Turnock, S.R., Hudson, D.A., 2017. *Ship Resistance and Propulsion*. Cambridge university press.
- Papandreou, C., Ziakopoulos, A., 2022. Predicting VLCC fuel consumption with machine learning using operationally available sensor data. *Ocean Eng.* 243, 110321.
- Perner, P., Apte, C., 2004. Empirical evaluation of feature subset selection based on a real-world data set. *Eng. Appl. Artif. Intell.* 17 (3), 285–288.
- Probha, N.A., Hoque, M.S., 2018. A study on transport safety perspectives in Bangladesh through comparative analysis of roadway, railway and waterway accidents. In: *Proceedings of the Asia-Pacific Conference on Intelligent Medical 2018 & International Conference on Transportation and Traffic Engineering 2018*, pp. 81–85.
- Raghavan, B., 2010. *Neyman-Pearson and Minimax Cost Complexity Pruning of Classification Trees*. Purdue University.

- Rajagopal, S., Hareesha, K.S., Kundapur, P.P., 2020. Performance analysis of binary and multiclass models using azure machine learning. *Int. J. Electr. Comput. Eng.* 10 (1), 2088–8708.
- Ranjan, G.S.K., Verma, A.K., Radhika, S., 2019. K-nearest neighbors and grid search cv based real time fault monitoring system for industries. In: 2019 IEEE 5th International Conference for Convergence in Technology (I2CT). IEEE, pp. 1–5.
- Sadat-Hosseini, H., Wu, P.C., Carrica, P.M., Kim, H., Toda, Y., Stern, F., 2013. CFD verification and validation of added resistance and motions of KVLC2 with fixed and free surge in short and long head waves. *Ocean Eng.* 59, 240–273.
- Sagi, O., Rokach, L., 2021. Approximating XGBoost with an interpretable decision tree. *Inf. Sci.* 572, 522–542.
- Saydam, A.Z., Küçükü, G.N., İnel, M., Gökçay, S., 2022. Uncertainty quantification of self-propulsion analyses with RANS-CFD and comparison with full-scale ship trials. *Brodogradnja: Teorija i praksa brodogradnje i pomorske tehnike* 73 (4), 107–129.
- Shang, G., Xu, L., Tian, J., Cai, D., Xu, Z., Zhou, Z., 2023. A real-time green construction optimization strategy for engineering vessels considering fuel consumption and productivity: a case study on a cutter suction dredger. *Energy* 274, 127326.
- Staudemeyer, R.C., Morris, E.R., 2019. Understanding LSTM—a Tutorial into Long Short-Term Memory Recurrent Neural Networks arXiv preprint arXiv:1909.09586.
- Tilling, F., Ringsberg, J.W., 2019. A 4 DOF simulation model developed for fuel consumption prediction of ships at sea. *Ships Offshore Struct.* 14 (Suppl. 1), 112–120.
- Tsujimoto, M., Shibata, K., Kuroda, M., Takagi, K., 2008. A practical correction method for added resistance in waves. *J. Jpn. Soc. Nav. Archit. Ocean Eng.* 8, 177–184.
- UNCTAD, 2022. Review of Maritime Transport 2022.
- UNFCCC, 2022. Long-term low-emission development strategies. Synthesis report by the secretariat. Sharm El-Sheikh Climate Change Conference.
- Uyanik, T., Karatug, Ç., Arslanoğlu, Y., 2020. Machine learning approach to ship fuel consumption: a case of container vessel. *Transport. Res. Transport Environ.* 84, 102389.
- Valanto, P., Hong, Y., 2015. Experimental investigation on ship wave added resistance in regular head, oblique, beam, and following waves. In: The Twenty-Fifth International Ocean and Polar Engineering Conference. OnePetro.
- Vinayak, P.P., Prabu, C.S.K., Vishwanath, N., Prakash, S.O., 2021. Numerical simulation of ship navigation in rough seas based on ECMWF data. *Brodogradnja: Teorija i praksa brodogradnje i pomorske tehnike* 72 (1), 19–58.
- Vitali, N., Ppčić-Oršić, J., Soares, C.G., 2020. Coupling voyage and weather data to estimate speed loss of container ships in realistic conditions. *Ocean Eng.* 210, 106758.
- Wang, H., 2020. Development of Voyage Optimization Algorithms for Sustainable Shipping and Their Impact to Ship Design. Doctoral dissertation, Chalmers Tekniska Hogskola (Sweden).
- Wang, H., Lang, X., Mao, W., 2021. Voyage optimization combining genetic algorithm and dynamic programming for fuel/emissions reduction. *Transport. Res. Transport Environ.* 90, 102670.
- Wang, H., Lang, X., Mao, W., Zhang, D., Storhaug, G., 2020. Effectiveness of 2D optimization algorithms considering voluntary speed reduction under uncertain meteocean conditions. *Ocean Eng.* 200, 107063.
- Wang, S., Ji, B., Zhao, J., Liu, W., Xu, T., 2018. Predicting ship fuel consumption based on LASSO regression. *Transport. Res. Transport Environ.* 65, 817–824.
- Wu, L., Gao, G., Yu, J., Zhou, F., Yang, Y., Wang, T., 2023. PDD: partitioning DAG-topology DNNs for streaming tasks. *IEEE Internet Things* 32, 109232.
- Xu, G., Zhang, Z., Zhang, T., Yu, S., Meng, Y., Chen, S., 2022. Aspect-level sentiment classification based on attention-BiLSTM model and transfer learning. *Knowl. Base Syst.* 245, 108586.
- Yaakob, O., Ahmed, Y.M., Rashid, M.F.A., Elbatran, A.H., 2015. Determining ship resistance using computational fluid dynamics (CFD). *Journal of Transport System Engineering* 2 (1), 20–25.
- Yan, X., He, Y.Q., Fan, A., 2023. Carbon footprint prediction considering the evolution of alternative fuels and cargo: A case study of Yangtze river ships. *Renew. Sustain. Energy Rev.* 173, 113068.
- Yan, R., Wang, S., Du, Y., 2020. Development of a two-stage ship fuel consumption prediction and reduction model for a dry bulk ship. *Transport. Res. E Logist. Transport. Rev.* 138, 101930.
- Yan, R., Wang, S., Psarafitis, H.N., 2021. Data analytics for fuel consumption management in maritime transportation: status and perspectives. *Transport. Res. E Logist. Transport. Rev.* 155, 102489.
- Yuan, Z., Liu, J., Liu, Y., Yuan, Y., Zhang, Q., Li, Z., 2020a. Fitting analysis of inland ship fuel consumption considering navigation status and environmental factors. *IEEE Access* 8, 187441–187454.
- Yuan, Z., Liu, J., Liu, Y., Zhang, Q., Liu, R.W., 2020b. A multi-task analysis and modelling paradigm using LSTM for multi-source monitoring data of inland vessels. *Ocean Eng.* 213, 107604.
- Zhang, M., Conti, F., Le Sourne, H., Vassalos, D., Kujala, P., Lindroth, D., Hirdaris, S., 2021. A method for the direct assessment of ship collision damage and flooding risk in real conditions. *Ocean Eng.* 237, 109605.
- Zhang, M., Taimuri, G., Zhang, J., Hirdaris, S., 2023a. A deep learning method for the prediction of 6-DoF ship motions in real conditions. *Proc. IME M J. Eng. Marit. Environ.* 237 (4), 887–905, 14750902231157852.
- Zhang, M., Kujala, P., Musharraf, M., Zhang, J., Hirdaris, S., 2023b. A machine learning method for the prediction of ship motion trajectories in real operational conditions. *Ocean Eng.* 283 (1), 114905.
- Zhou, H., Zhang, J., Zhou, Y., Guo, X., Ma, Y., 2021. A feature selection algorithm of decision tree based on feature weight. *Expert Syst. Appl.* 164, 113842.
- Zhou, T., Hu, Q., Hu, Z., Zhen, R., 2022. An adaptive hyper parameter tuning model for ship fuel consumption prediction under complex maritime environments. *J. Ocean Eng. Sci.* 7 (3), 255–263.
- 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 meteocean and on-board data. *Ocean Eng.* 285, 115509.