Machine Learning Models for Predicting Maritime Vessel Fuel Consumption in Offshore Environments

Sang Thanh Do
University of Essex
sd24476@essex.ac.uk

Sirui Zhu
University of Essex
sirui.zhu@essex.ac.uk

Issam Boukhennoufa
University of Essex
ib20472@essex.ac.uk

Shengyang Huang
University of Essex
sh24437@essex.ac.uk

Huaizhi Zhang
University of Essex
hz24245@essex.ac.uk

Amit Kumar Singh
University of Essex
a.k.singh@essex.ac.uk

Xiaojun Zhai
University of Essex
xzhai@essex.ac.uk

Hugo Spinola
Njord Offshore
hspinola@njordoffshore.com

Tom Mehew
Njord Offshore
tmehew@njordoffshore.com

***Abstract*—Fuel expenditure constitutes a substantial proportion of the operational costs of the vessel. Environmental factors, including meteorological conditions and oceanic currents, significantly impact fuel efficiency. Consequently, fuel costs often represent 50-60% of the total cost, which requires maritime companies to pursue methodologies and technologies that mitigate energy consumption. This research focused on developing a robust fuel consumption prediction model for vessels operating under a leading Crew Transfer vessel (CTV) operator Njord offshore [24], using real-time data from engine sensors. To achieve this, machine learning techniques, including linear regression, random forest [19], gradient boosting, extreme gradient boosting (XGBoost), support vector regression (SVR), and multilayer perceptron (MLP), were used. The input variables encompassed the characteristics of the vessel and the prevailing meteorological conditions at the time of transit. The efficacy of the six algorithms was evaluated using metrics such as mean absolute error (MAE), mean squared error (MSE), root mean squared error (RMSE), and the coefficient of determination (R^2). In particular, the gradient boosting algorithm demonstrated outstanding performance and the most suitability for the dataset, achieving a R^2 value of approximately 94% in predicting fuel consumption for vessels operating within the specified offshore environment.**

***Index Terms*—fuel prediction, machine learning; predictive model; vessel routing**

I. INTRODUCTION

Maritime fuel expenditure, particularly for Crew Transfer Vessels (CTVs) supporting Offshore Wind Farm (OWF) operations, is determined by a multifaceted set of variables encompassing vessel design, operational procedures, and ambient environmental conditions, as depicted in Figure 1. Historically, the impetus for minimizing fuel consumption has primarily been economic. Initial strategies centered on refining hull geometry and enhancing engine performance [1]. However, with technological advancements, there has been a transition towards integrating real-time data analysis and predictive modeling. Effective fuel management remains a critical concern, contingent upon factors such as vessel velocity, engine efficiency, meteorological conditions, and hull bio-fouling [1] - [2]. Maritime enterprises are actively pursuing efficient and sustainable energy utilization to curtail operational costs and adhere to international regulatory standards, including those set by the International Maritime Organization (IMO) and the International Convention for the Safety of Life at Sea (SOLAS) [3].

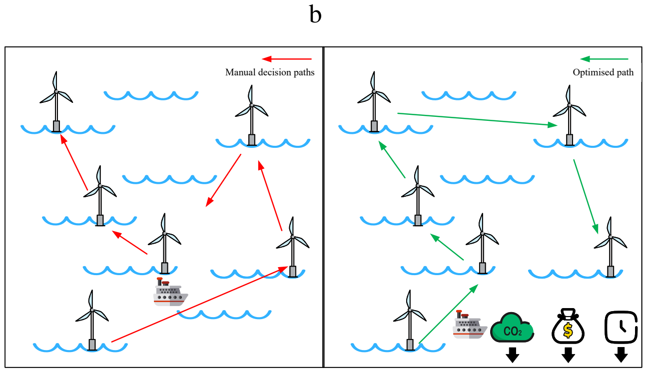

Fig. 1. Benefits of using the optimised path for crew transfer in offshore wind farm

The maritime industry has dedicated significant research efforts toward curbing fuel consumption and mitigating its environmental impact. The collection of works encompasses investigations into alternative propulsion technologies, such as wind, solar, and wave energy [5] - [7], as well as the potential of fuel cells [8]. However, it is a critical area that warrants further exploration: the development of holistic route optimization strategies that account for the interplay between fuel consumption, voyage time, and cargo considerations.

A primary challenge in optimizing vessel operations for fuel efficiency lies in the complex and intertwined nature of the factors that influence fuel consumption, such as vessel speed or sensor quality [9]. This complexity often obscures the relative importance of individual parameters, making it difficult for operators to prioritize effective fuel-saving measures. Therefore, robust predictive models are essential to identify the key drivers of fuel use and enable data-driven decision-making for optimized operations [4]. Constructing such models necessitates a sophisticated framework that captures the intricate interdependencies among these factors [10].

There is a rich body of literature that explores the multifaceted factors that influence vessel resistance and fuel consumption, offering various theoretical frameworks and analytical methodologies [11]. Researchers have investigated the effectiveness of machine learning algorithms in predicting fuel usage in analogous domains, such as public transportation, where Random Forest has exhibited strong performance [12]. Artificial neural networks (ANNs) have also been employed to predict vessel fuel consumption [13]. Comparative studies have evaluated different regression-based approaches, highlighting the potential of algorithms like Extra Trees [14]. Other studies have found Multiple Linear Regression to provide the most accurate estimations [15]. Furthermore, the application of machine learning extends to predicting port stay durations using CatBoost [16] and developing integrated models for fuel consumption forecasting [17]. However, they have not been considered for maritime vessels operating in real environments.

This research builds upon this foundation by evaluating six machine learning algorithms - Linear Regression [18], Random Forest [19], Gradient Boosting [20], XGBoost [21], SVR [22], and MLP [23] - that have demonstrated promising results in other fields. Drawing inspiration from semi-supervised learning, we leverage operational data from one of the UK's leading CTV operators, Njord Offshore [24], to predict fuel costs associated with vessel transit between wind turbines. Njord Offshore [24] specializes in managing Crew Transfer Vessels for the offshore wind farm sector, providing services with well-maintained vessels throughout Northern Europe. Through rigorous testing on a standardized dataset, we demonstrate that Gradient Boosting consistently achieves lower RMSE and MAE compared to other models, indicating superior predictive accuracy. A key contribution of this work lies in the comprehensive analysis of vessel-specific features relevant to fuel consumption prediction. Starting with a broad set of parameters, including vessel dimensions, inter-turbine distances, wave height, wind speed, and current velocity, we refine the feature set by eliminating irrelevant or noisy variables.

To guide the reader, this paper is organized as follows: The "Our framework" section elucidates the workflow of our prediction model design. The "Prediction" section provides a concise overview of prevalent machine learning techniques for predictive modeling. In the "Evaluation" section, we assess the performance of the selected algorithms using their default parameters. The "Optimization" section details the hyperparameter tuning strategies employed for each algorithm. Experimental results, evaluated using metrics such as MAE, MSE, R^2, and RMSE, are presented to showcase the effectiveness of our chosen algorithm. Finally, the "Conclusion" section summarizes the key findings of this study and discusses potential avenues for future research.

II. OUR FRAMEWORK

A. Workflow

The analysis uses operational data collected from Njord Offshore [24] to estimate fuel utilization. The recorded data tracks offshore processes through stations, monitoring vessel movements starting from departure until reaching each turbine destination and finally returning to the port. The research also includes information about technician and cargo transfers because they reveal the additional weight that affects fuel consumption. The analysis gained strength through the integration of weather data obtained from the Copernicus database [25] since weather conditions shape the vessel's fuel efficiency. Daily fuel usage readings came from vessel monitoring systems that operators installed for precise measurement accuracy.

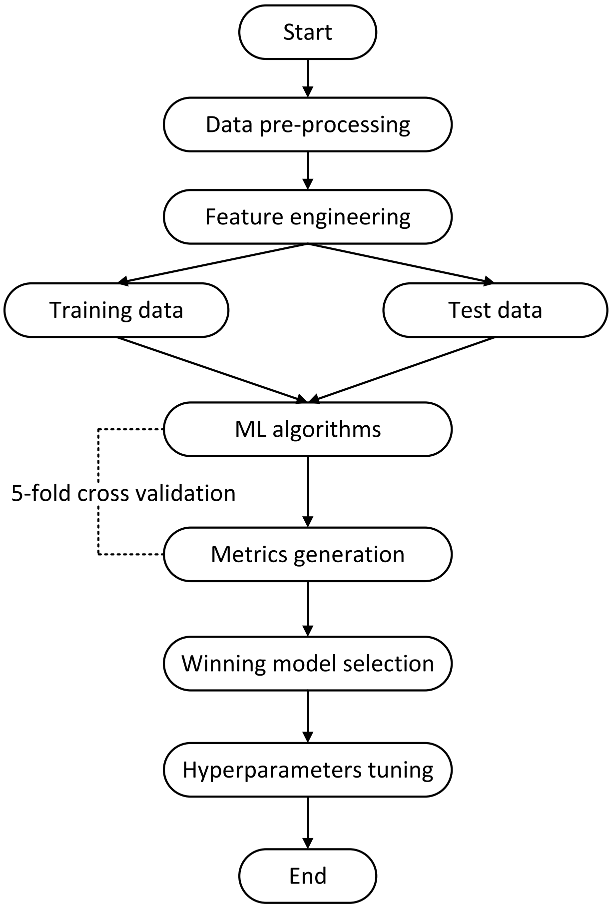

Fig. 2. The selection workflow of our prediction model

The complete model selection methodology is visually represented in the flowchart provided in Figure 2. It outlines the development and evaluation of a fuel prediction model. Initially, a comprehensive feature engineering process was conducted, involving the transformation of existing dataset variables and the derivation of novel features. Subsequently, six distinct machine learning algorithms were implemented and rigorously assessed using established performance metrics. The algorithm demonstrated superior predictive accuracy and was then subjected to hyperparameter tuning to enhance the precision of its predictions.

TABLE I
SAMPLE ROWS IN THE ORIGINAL DATASET.
(*) THIS VARIABLE IS THE OUTPUT OF OUR ALGORITHMS.

Variable	Row 1	Row 2
Distance abs	0	31.26799
Transfers Date	06/12/2021	06/12/2021
Transfers : Time Question	06:55:00	08:20:00
Transfers : Turbine	Lowestoft	EA1 C13
Transfers : No. of Pax	9	3
No. of Cargo L/D	0	3
To/From Where	Shore	Tp/Substation
Transfer Up/Down	Down(On to CTV)	Up(Leaving CTV)
Lat	52.474199	52.22616
Lon	1.736557	2.486233
distance	0	31.87548
Vessel	Vessel A	Vessel A
VHM0 (Wave Height)		1.304
Eastward Wind	1.69	0.95
Northward Wind	3.57	8.94
Eastward Current	-0.097	-0.065
Northward Current	-0.629	-0.305
Total fuel used L (*)	0	673.88756

B. Data pre-processing and feature engineering

Crew members of 32 CTVs have logged their daily progress reports manually starting from 03/12/2018 to serve as the research base. These records were then digitalized and contained departure time stamps together with vessel names and stations, turbine locations, and information about technician and cargo exchanges. Exact timestamps and geolocation matches were used to harmonize trip data with environmental parameters such as eastward and northward winds as well as ocean currents and wave height obtained from the Copernicus database [25]. The preprocessing actions served to create a stable foundation for dataset analysis regarding the operational settings and performance of CTVs. The original dataset that emanated from the company shows these examples of data records in Table I.

In this section, we also describe the processing steps applied to the input data to extract relevant features for our analysis. The data contained human errors such as inconsistent formats, typos, and missing entries, which were removed during data cleaning. Standardization was applied to timestamps, vessel IDs, and location names to ensure uniformity across datasets. The input data is transformed and manipulated to derive meaningful features that capture the dynamics of the vessel's journey. Below, we outline the specific transformations applied to some special input features:

1) Transfers: Time Question: The time-related input is converted into hours as the unit of measurement to represent the duration of each segment of the journey. Specifically, we calculate the time difference between two consecutive points in the journey. For example, as shown in Table I, if the arrival time at the first point is 08:20:00 and the departure time from the previous point is 06:55:00, the time difference is approximately 1.42 hours.

2) Transfers: No. of Pax: The number of passengers (No. of Pax) is processed to reflect the number of people on board during each segment of the journey. For instance, if 9 passengers board the vessel at the starting point and 3 passengers disembark at the first stop, the number of passengers on board for the segment from the starting point to the first stop is recorded as 9. For the next segment, the number of passengers on board is updated to 6, and so on.

3) Transfers: No. of Cargo Lifts/Drops: The number of cargo lifts/drops is processed similarly to the number of passengers but consider the direction of cargo movement on the "Transfer Up/Down" data. If the data indicates "Down (On to CTV)," we multiply the number of cargo items by 1, indicating cargo is being loaded onto the vessel. Conversely, for "Up (Leaving CTV)," we multiply by -1, indicating cargo is unloaded. For instance, as shown in Table I, if the number of cargo lifts/drops at two consecutive points is 0 and 3, respectively, the number of cargo items is 0 and -3.

4) Vessel Type: We classify the vessels into three distinct types provided by Njord Offshore [24], namely 21 m vessels, 24 m, and 26 m and more.

5) VHM0 (Wave Height), Eastward Wind, Northward Wind, Eastward Current, Northward Current: For these environmental features, we use the values measured at the previous point to represent the conditions for the current segment of the journey. For example, as shown in Table I, the Eastward Wind, Northward Wind, Eastward Current, and Northward Current values for the segment from Lowestoft to EA1 C13 are based on the measurements taken at Lowestoft, rather than at EA1 C13. For VHM0 (Wave Height), since the starting point typically lacks this measurement, we use the value from the arrival point.

6) Speed: The average speed of the vessel is calculated using the traveled distance divided by the time taken for each segment. This feature measures the vessel's speed in nautical miles per hour (nm/hr).

7) Consumed Fuel: The fuel consumption is derived from the vessel's engine data, which provides the fuel usage rate (in liters per hour, l/hr) at different speeds. For instance, if the calculated speed for a segment is 15 nm/hr, we use the fuel usage rate corresponding to 15 nm/hr, which might be 302 l/hr. This feature captures the fuel consumption rate based on the vessel's speed.

8) Cumulative Fuel: The cumulative fuel consumption is calculated by summing the total fuel used for all completed segments of the journey. This feature is important because it reflects the gradual weight reduction of the vessel due to fuel consumption, which can impact the vessel's performance and dynamics over time.

III. PREDICTION

Using Python and the scikit-learn library [18], we trained six machine learning algorithms: Linear Regression [18], Random Forest [19], Gradient Boosting [20], XGBoost [21], SVR [22], and MLP [23].

1) Linear Regression:

Linear regression [18], a foundational supervised learning technique, establishes a linear relationship between one dependent variable and one independent variable. Multiple linear regression extends this principle to scenarios involving multiple independent variables and a single dependent variable. The algorithm learns a function from observed data, represented by pairs of independent (X) and dependent (Y) values, enabling the prediction of Y for unseen values of X. In essence, the objective of regression analysis is to approximate a continuous-valued function that accurately predicts the dependent variable based on the provided independent features.

The multiple linear regression formula is given by:

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_n X_n + \epsilon \tag{1}$$

where Y is the dependent variable, $X_1, X_2, \ldots, X_n$ are the independent variables, β_0 is the intercept, $\beta_1, \beta_2, \ldots, \beta_n$ are the coefficients representing the change in Y for a unit change in the corresponding X variable, ϵ is the error term representing the difference between the observed and predicted values of Y.

2) Random Forest:

The Random Forest algorithm [19] is an ensemble learning method used for both classification and regression tasks. It operates by constructing a multitude of decision trees at training time and outputting the class that is the mode of the classes (classification) or mean prediction (regression) of the individual trees.

3) Gradient Boosting:

Gradient Boosting [20] is an ensemble learning method that sequentially builds a model by combining weak learners, typically decision trees. The algorithm minimizes a loss function by iteratively adding trees that correct the errors made by the previous ensemble. The core idea behind Gradient Boosting is to optimize a differentiable loss function $L(y, F(x))$, where y is the true target value and $F(x)$ is the prediction. The algorithm iteratively adds trees $h_m(x)$ to the model $F_m(x)$ to minimize the loss.

Specifically:

$$h_m(x) \approx -\frac{\partial L(y, F_{m-1}(x))}{\partial F_{m-1}(x)} \tag{2}$$

4) Extreme Gradient Boosting:

XGBoost [21] is an optimized distributed gradient boosting library designed to be highly efficient, flexible, and portable. XGBoost incorporates several optimizations for performance and accuracy. The objective function is defined as:

$$\mathcal{L}(\theta) = \sum_{i=1}^{n} l(y_i, \hat{y}_i) + \sum_{k=1}^{K} \Omega(f_k) \tag{3}$$

where:

- $l(y_i, \hat{y}_i)$ is the loss function.
- $\Omega(f_k) = \gamma T + \frac{1}{2}\lambda \sum_{j=1}^{T} w_j^2$ is the regularization term.
- T is the number of leaves.
- w_j is the leaf weight.
- γ and λ are regularization parameters.

5) Support Vector Regression:

In the realm of regression analysis, SVR [22] emerges as a powerful supervised learning technique. Building upon the foundational principles of Support Vector Machines (SVM), SVR adapts these concepts to accurately predict continuous-valued outputs. The core objective of SVR is to construct a predictive function that closely matches the target data, allowing for a controlled degree of deviation within a defined tolerance range.

SVR seeks to find a function $f(x)$ that deviates from the actual targets y_i by no more than ϵ for all training data. The objective is to minimize the following regularized risk function:

$$\frac{1}{2}\|w\|^2 + C \sum_{i=1}^{n} (\xi_i + \xi_i^*) \tag{4}$$

Subject to:

$$|y_i - f(x_i)| \leq \epsilon + \xi_i + \xi_i^* \tag{5}$$

$$\xi_i, \xi_i^* \geq 0 \tag{6}$$

where: w is the weight vector, C is the regularization parameter, ξ_i and ξ_i^* are slack variables, ϵ is the margin of tolerance, $f(x) = w^T \phi(x) + b$ is the regression function, where $\phi(x)$ is a feature mapping.

The ϵ-insensitive loss function is defined as:

$$L_\epsilon(y, f(x)) = \begin{cases} 0, & \text{if } |y - f(x)| \leq \epsilon \\ |y - f(x)| - \epsilon, & \text{otherwise} \end{cases} \tag{7}$$

6) Multilayer Perception:

MLP is a feedforward neural network trained using backpropagation.

The output of each neuron is computed as:

$$a_j^{(l)} = \sigma \left(\sum_i w_{ji}^{(l)} a_i^{(l-1)} + b_j^{(l)} \right) \tag{8}$$

where σ is the activation function.

Backpropagation

Weights are updated using gradient descent:

$$w_{ji}^{(l)} = w_{ji}^{(l)} - \alpha \frac{\partial L}{\partial w_{ji}^{(l)}} \tag{9}$$

where α is the learning rate and L is the loss function.

Activation Functions

Common activation functions include:

- Sigmoid: $\sigma(z) = \frac{1}{1+e^{-z}}$
- ReLU: $\sigma(z) = \max(0, z)$

To demonstrate the practical application of the algorithms discussed, Python code was developed. This code was executed on a personal computer, utilizing a standard desktop environment, to simulate real-world application scenarios. The implementation employed common libraries such as scikit-learn and pandas, demonstrating the feasibility of these algorithms in readily accessible computing environments. The

TABLE II
SELECTED FEATURES FOR TRAINING

No.	Feature	Unit or datatype
1	Distance abs	Nautical Miles
2	Transfers : Time Question	hh:mm:ss
3	Transfers : No. of Pax	integer
4	Transfers : No. of Cargo Lifts/Drops	integer
5	distance	Nautical Miles
6	VHM0 (Wave Height)	Float
7	Eastward Wind	Float
8	Northward Wind	Float
9	Eastward Current	Float
10	Northward Current	Float
11	Speed	Float
12	Month	integer
13	Season	integer
14	Consumed Fuel	Liters
15	Vessel type	integer
16	Cumulative Fuel	Liters

results presented in this paper were therefore generated from a typical user setup.

IV. EVALUATION

A comprehensive evaluation strategy was designed to test the predictive models' ability to perform robustly on the diverse dataset and yield reliable estimates of the target variable. The assessment aimed to capture a broad spectrum of model performance characteristics, guided by both established evaluation practices and the specific objectives of this study. To this end, we employed metrics that included MAE, MSE, RMSE, and the dependent variable (R^2). This collection of metrics enabled a thorough examination of the models' predictive capabilities, providing insights into both the accuracy and the overall fit of the models to the data. The training dataset consists of 80% of the dataset, and the testing dataset is the remaining.

In Table II, all variables serve as input features. Among these, `Distance_abs` and `distance` are both parameters related to the vessel's travel distance. The key difference between these two features is that `distance` represents the **actual sailing distance** recorded by the vessel's sensor, while `Distance_abs` represents the **theoretical distance** calculated from the coordinates of the two points. These two input variables are not used simultaneously in the model; instead, they are tested separately alongside the other variables listed in Table II.

Table III displays the results obtained using two different input features. The **upper section** of the table illustrates the metrics of algorithms using `distance` along with the other variables as inputs. The overall performance is significantly better compared to using `Distance_abs` as an input. Among the models, Random Forest and Gradient Boosting demonstrated superior performance. Consequently, we proceeded to fine-tune the hyperparameters for these two algorithms.

TABLE III
METRICS COMPARISON USING `DISTANCE` AND `DISTANCE_ABS` AS INPUTS

Using `distance` as an input feature

Algorithm	MAE	MSE	R^2	RMSE
Linear Regression	56.6624	6955.5453	0.9173	83.3999
Random Forest	45.1959	5080.9921	0.9396	71.2811
Gradient Boosting	46.1015	4869.4790	0.9421	69.7817
XGBoost	48.1239	6296.4515	0.9251	79.3502
SVR	137.5953	47026.5363	0.4409	216.8560
MLP	48.1607	5381.0183	0.9360	73.3554

Using `Distance_abs` as an input feature

Algorithm	MAE	MSE	R^2	RMSE
Linear Regression	71.5011	8751.7272	0.8844	93.5506
Random Forest	57.9833	6914.6126	0.9087	83.1541
Gradient Boosting	61.1288	6934.892	0.9084	83.2759
XGBoost	62.3155	7703.2699	0.8982	87.7682
SVR	139.5351	36291.07	0.5208	190.5021
MLP	56.8261	6230.483	0.9177	78.9334

TABLE IV
BEST HYPERPARAMETERS IN RANDOM FOREST AND GRADIENT BOOSTING

Parameter	Random Forest	Gradient Boosting
n_estimators	100	500
max_depth	20	5
min_samples_split	2	1
min_samples_leaf	2	2
bootstrap	True	-
subsample	-	0.9
learning_rate	-	0.05
max_features	-	1.0

We employed the **GridSearchCV** method [18] to identify the optimal hyperparameters for the two algorithms. The best hyperparameters for each algorithm are presented in Table IV. After re-tuning the models with these parameters, the test results, as shown in Table V, indicate that both models perform similarly. However, if we were to highlight a slight advantage, **Gradient Boosting** appears to perform marginally better across the evaluation metrics.

V. CONCLUSION AND FUTURE WORKS

We explore the potential of different machine learning algorithms to predict vessel fuel consumption in real-world operational settings. The study follows a three-phased approach: (1) We analyze the variables affecting fuel consumption and generate new key factors that impact predictive accuracy. (2) We train machine learning models on operational data, incorporating dynamic weather patterns to reflect real-time conditions. (3) We fine-tune model parameters to optimize predictive performance. To survey the viability of our technique, we utilize a three-year dataset (2021-2024) from Njord Offshore [24]. We at that point benchmark our approach be-

TABLE V
METRICS AFTER TUNING (GRADIENT BOOSTING AND RANDOM FOREST)

Algorithm	MAE	MSE	R^2	RMSE
Random Forest	44.6749	4970.6460	0.9409	70.5028
Gradient Boosting	43.6126	4754.9308	0.9435	68.9560

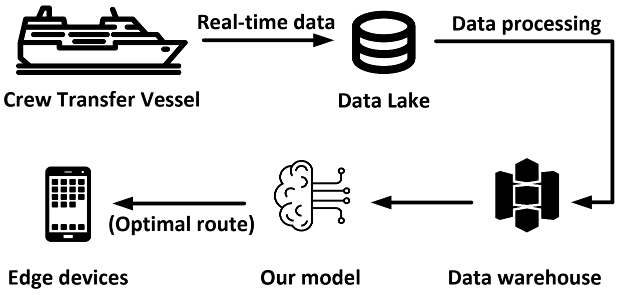

Fig. 3. The workflow of the real-time application

tween existing predictive models, illustrating the predominant execution of the Gradient Boosting algorithm with optimized hyperparameters. By leveraging strong, authentic records, we can achieve significant advancements in voyage optimization, fuel productivity, and emission reduction.

This application - as illustrated in Figure 3 - will serve as a decision support tool, providing captains with timely and actionable insights for optimized vessel operation.

ACKNOWLEDGMENT

This work is supported by UK Research and Innovation (UKRI) Innovate UK 10132557 and 10049632. For the purpose of open access, the author has applied a Creative Commons Attribution (CC BY) licence to any Author Accepted Manuscript version arising.

REFERENCES

[1] Ang, J. H., Goh, C., Li, Y., “Hull form design optimisation for improved efficiency and hydrodynamic performance of ’ship-shaped’ offshore vessels,” Royal Institution of Naval Architects - International Conference on Computer Applications in Shipbuilding, Vol. 2, pp. 71-80, 2015.

[2] Y. Wu, J. Hua, D.L. Wu, “Economic analysis of ship operation using a new antifouling strategy”, Ocean Engineering, Volume 266, Part 4, 2022.

[3] Joseph, A., Dalaklis, D., “The international convention for the safety of life at sea: highlighting interrelations of measures towards effective risk mitigation,” Journal of International Maritime Safety, Environmental Affairs, and Shipping, 5(1), 1–11, 2021.

[4] Beşikçi, E. B., Arslan, O., Turan, O., Olçer, A. I. “An artificial neural network based decision support system for energy efficient ship operations,” Computers & Operations Research, 66, 393–401, 2016.

[5] Jingru Xing, Liang Yang, “Wave devouring propulsion: An overview of flapping foil propulsion technology,” Renewable and Sustainable Energy Reviews, 2023.

[6] Yiping Zhang, Li Xu, Zhengqiang Ding, Mingchao Hu, “Wave propulsion and sea-keeping enhancement for ships in rough sea condition by flapping foils,” Ocean Engineering, Volume 266, Part 2, 2022.

[7] Chan ChunYin, Wang Junxian, Yang Liang, Zang Jun., “Wave-assisted propulsion: An experimental study on traveling ships,” Physics of Fluids. 36. 10.1063/5.0190630, 2024

[8] Maja P., Nikola Vladimir, Ivana J., Marija K., “Application of fuel cells with zero-carbon fuels in short-sea shipping,” Applied Energy, Volume 309, 2022.

[9] Yi Z., Kayvan P., Alan J. Murphy, Zigor U., Igor G., I. Quincoces, Jose A. Fernandes-Salvador, “Predicting ship fuel consumption using a combination of metocean and on-board data,” Ocean Engineering, vol. 285, 2023.

[10] Bhushan Taskar, Poul Andersen, “Benefit of speed reduction for ships in different weather conditions,” Transportation Research Part D: Transport and Environment, Volume 85, 2020.

[11] Ran Yan, Shuaian Wang, Harilaos N. Psaraftis, “Data analytics for fuel consumption management in maritime transportation: Status and perspectives,” Transportation Research Part E: Logistics and Transportation Review, Volume 155, 2021.

[12] Zhong Shuo Chen, Jasmine Siu Lee Lam, Zengqi Xiao, “Prediction of harbour vessel fuel consumption based on machine learning approach,” Ocean Engineering, Volume 278, 2023.

[13] Theodoropoulos, P., Spandonidis, C. C., Themelis, N., Giordamlis, C., Fassois, S., “Evaluation of Different Deep-Learning Models for the Prediction of a Ship’s Propulsion Power,” Journal of Marine Science and Engineering, 9(116), 2021.

[14] Gkerekos, C., Lazakis, I., Theotokatos, G., “Machine Learning Models for Predicting Ship Main Engine Fuel Oil Consumption: A Comparative Study,” Ocean Engineering, 188, 106282, 2019.

[15] Uyanık, T., Karatug, C., Arslanoglu, Y., “Machine Learning Approach to Ship Fuel Consumption: A Case of Container Vessel,” Transportation Research Part D: Transport and Environment, 84, 102389, 2020.

[16] Stepec, D., Martincic, T., Klein, F., Vladusic, D., Costa, J. P. “Machine Learning Based System for Vessel Turnaround Time Prediction,” In Proceedings of the 2020 21st IEEE International Conference on Mobile Data Management (MDM) (pp. 258–263).

[17] Hu, Z., Zhou, T., Osman, M. T., Li, X., Jin, Y., Zhen, R. , “A Novel Hybrid Fuel Consumption Prediction Model for Ocean-Going Container Ships Based on Sensor Data,” Journal of Marine Science and Engineering, 9(449), 2021.

[18] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay, “Scikit-learn: Machine learning in Python,” Journal of Machine Learning Research, vol. 12, pp. 2825-2830, 2011.

[19] L. Breiman, “Random forests,” Machine Learning, vol. 45, no. 1, pp. 5–32, 2001.

[20] J. H. Friedman, “Greedy function approximation: a gradient boosting machine,” Annals of Statistics, vol. 29, no. 5, pp. 1189-1232, 2001.

[21] T. Chen and C. Guestrin, “XGBoost: A scalable tree boosting system,” Proc. 22nd ACM SIGKDD Int. Conf. Knowledge Discovery and Data Mining, San Francisco, CA, USA, pp. 785–794, 2019.

[22] V. N. Vapnik, “The nature of statistical learning theory,” New York: Springer-Verlag, 1995.

[23] Ghorbanian, K., Gholamrezaei, M., “An artificial neural network approach to compressor performance prediction,” Applied Energy, pp. 1210–1221, 2009.

[24] https://njordoffshore.com/.

[25] https://www.copernicus.eu/en.