This is a peer-reviewed, accepted author manuscript of the following conference paper:
Velasco-Gallego, C, Lazakis, I & Polaki, V 2024, 'Analysis of attention mechanisms for the prediction of ship fuel oil consumption', Paper presented at 63rd International Congress of Naval Architecture Marine Technology and Maritime Industry, Madrid, Spain, 24/04/24 - 26/04/24.

Analysis of attention mechanisms for the prediction of ship fuel oil consumption

Christian Velasco-Gallego, Nebrija University

Iraklis Lazakis, University of Strathclyde

Vendanjali Polaki, University of Strathclyde

ABSTRACT

Carbon Dioxide (CO2) remains the dominant contributor to climate change in shipping with Heavy Fuel Oil (HFO) prevailing as the most significant fuel utilised in maritime transportation globally. Thus, while several technologies, including the consideration of renewable energies and alternative fuels, are being explored to contribute towards the Net Zero goal, the consumption of Fuel Oil (FO) continues to be of a substantial concern. Moreover, the optimal use of FO can lead to minimising CO2 emissions as well. This necessitates the development of more sophisticated tools to optimise onboard consumption, thereby facilitating the reduction of emissions and the associated operational costs. Accordingly, this paper analyses the use of an attention mechanism-based deep learning model for the prediction of FO consumption. A case study on a tanker vessel is conducted to assess the performance of this type of model, aiming to develop a decision-making tool for optimising ship FO consumption.

1. INTRODUCTION

Sea-based transportation accounts for 80% of international trade of goods while the shipping sector has experienced a 20% rise in GHG emissions over the last decade as of 2023 [1]. The COP28 conference had set out the goal of net zero emissions by 2050. Additionally, the fourth IMO GHG study 2020 approximated that the GHG emissions from shipping in 2018 accounted for 2.89% of total global man-made GHG emissions [2]. There are several maritime emissions' regulatory bodies worldwide such as the European Union Emissions' Trading System (EU ETS), the International Maritime Organisation (IMO), etc. As a result, ship operators must comply with policies and regulations laid down on emissions depending on the regions of operation. Moreover, fuel costs account for 50-60% of operating costs of the vessels [3]. Hence, it is increasingly imperative to strive towards cost-effective solutions reducing emissions through various methods such as integrating carbon capture solutions, voyage planning, optimising ship speed and fuel consumption, etc. Thus, the focus of this paper will be on forecasting ship fuel consumption.

In this respect, there have been various efforts to provide a review on forecasting ship fuel consumption [4]. The authors demarcated the models covered in these papers into two categories – statistical and machine learning methods. They found that machine learning methods performed better for high-dimensional datasets in comparison to the statistical techniques. However, the former suffered in terms of the interpretability of its predictions due to the complexity of the inner workings of the model. A lack of interpretability could indicate potential biases in these models. However, other researchers presented an interpretable machine learning model that addresses the trade-off between interpretability and predictive accuracy [5]. This paper is primarily focused on the predictive accuracy of the model which has been proposed.

The paper has been organised into 5 sections. Section 2 covers related previous work; Section 3 elucidates the methods used to design the deep learning model. Section 4 provides the application of the suggested methodology in the case of a tanker ship while Section 5 presents conclusions based on the discussed results in the previous section.

2. LITERATURE REVIEW

Attention mechanism allows for dynamic assignment of weights of input features based on the outcome expected of the machine learning model. Abadi et al. had first introduced the transformer architecture which relies on self-attention mechanisms for use in Natural Language Processing (NLP) [6]. This approach overcomes the limitations of sequence-to-sequence models, such as Convolutional Neural Networks(CNNs) and Recurrent Neural Networks (RNNs), by effectively capturing long-range dependencies and improving training efficiency.

Deep learning models are greatly accessible today through several open-source packages and frameworks such as Tensorflow [7], PyTorch [8], and Keras [9]. However, they have a nominal presence in literature regarding their exploitation for the prediction of ship fuel consumption. Zhang et al. proposed a deep learning model combining a Bidirectional – Long Short-Term Memory (Bi-LSTM) model with attention mechanism [10]. They obtained big data records from 8 international voyages from the Kamsarmax Bulk Carrier of Laskaridis Shipping Co. Ltd. They compared it with Decision Tree models with attention mechanism and found that the Bi-LSTM model performed well especially for high-frequency data. The leveraging of big data records will allow for an improvement in reduction of carbon emissions. Additionally, it provides a foundation for efficient fuel management for real operating conditions.

Lei et al made use of the LSTM model to forecast fuel consumption for data derived from inland watercrafts [11]. They obtained data from the Juhang 777 vessel over a period of 5 months. They found that the deep learning model performs better than the machine learning model, back propagation neural network, in terms of average relative error (in percentage). Wang et al. proposed two models: Physics-informed neural network (PI-NN), Mixed-Integer Quadratic Optimisation (MIO) [12]. The PI-NN model is cognizant of physics constraints which allows for the model to be highly explainable. However, as the number of constraints increase, the model becomes less flexible yet more accurate. The interpretability of the MIO-BF model is lower in comparison since the model alters the original values of the feature variables.

Tran et al. proposed a combination of Monte Carlo simulations and Artificial Neural Networks (ANN) for predicting the fuel consumption [13]. The data has been collected from a bulk carrier over a period of 2 years. Additionally, the model has been run for each of the three loading conditions (65%, 85% and 100% MCR). The performance of this model was compared with other models incorporating Artificial Neural Networks with other regression models and concluded that ANNs performed better. They also mentioned that the accuracy of the model was higher for 65% and 85% MCR loading conditions, which are more commonly observed in real-world maritime transportation.

Su et al. proposed an AI-based energy efficiency decision system using a prediction and optimisation technique [14]. They used a 2-stage method utilising the XGBoost model for prediction, which was chosen after comparison with 15 other machine learning models, followed by the employment of the Particle Swarm Optimisation technique to predict optimal ship speed to further improve the prediction of fuel consumption. They used a multi-source dataset including car and trucking data for the duration spanning the years 2012 to 2021. They emphasised the model's ability to manage high-dimensional data, to make more accurate predictions and to account for a large number of irregularities. Further, they mention the improvement in fuel-usage efficiency upon the employment of an optimisation technique.

Mekkaoui et al. proposed a data-driven solution based on deep learning sequence methods and historical ship voyage data to predict ship speed in real-time [15]. The data is obtained from ship trips in the area of St. Lawrence Seaway for a period of 2 years. The prediction of ship speed would allow for a ship operator to design the voyage in advance, taking various factors into account such as ship trajectory planning and weather to consequently predict trends in ship fuel consumption. The study throws light on three sequence models – LSTM, Bi-LSTM and Transformer. LSTM model performed poorly in comparison to the other two models in terms of metrics such as mean absolute error (MAE) and mean squared error (MSE). On the other hand, Theodoropoulos et al. 2021 made use of ANN and RNN models to predict the ship's propulsion power. The data used was collected over a period of nineteen months from a 165000-DWT tanker.

Panapakidis et al. highlighted ship fuel consumption prediction for passenger vessels. They proposed the combination of LSTM – RNN and Elman NN models for the forecasting of fuel consumption [16]. The data was obtained from 322 voyage reports over a period of 10 months from the Ro/Pax vessel operating in the Adriatic route spanning between Greece and Italy. They mentioned the use of statistical techniques to precede the selection of the deep learning model based on the linearity of the relationship between the parameters under study.

Wu et al. proposed a route optimisation method using a deep learning architecture consisting of CNN and the transformer architecture by the conversion of discrete weather data to a continuous format [17]. The weather observation and weather forecast data was obtained from the Integrated Marine Observing System (IMOS) and the European Centre for Medium-Range Weather Forecasts (ECMWF) for over a period of one year. Such a method allows for prior route planning which in turn helps ship operators plan and optimise fuel usage. A continuous prediction could allow for high-resolution prediction of future fuel usage. Similarly, Bui Duy et al. performed optimal route selection for container ships using a deep ANN architecture in conjunction with the Asymmetric Travelling Salesman Problem (ATSP) algorithm [18].

Within the current paper, the authors propose a novel method to predict fuel oil consumption onboard ships using an attention mechanism–based deep learning model. The concerned data has been obtained from a cargo vessel consisting of system-monitored parameters from the main engine and diesel generators of the vessel.

3. METHODOLOGY

By having explored the main contributions of this conference paper, the proposed methodology is graphically represented in Fig.3.1. Special attention is being given to the data pre-processing stage to guarantee that data possesses the necessary quality to be further processed, adheres to the required format, and aligns with the specific requisites of the data-driven task to be performed. Consequently, a total of seven data pre-processing stages are initially performed. The second phase relates to the development and training of the model to be analysed in this study which is an attention-based neural network. Finally, to analyse the performance of the developed model, an evaluation stage is introduced to ensure an adequate analysis of the obtained results.

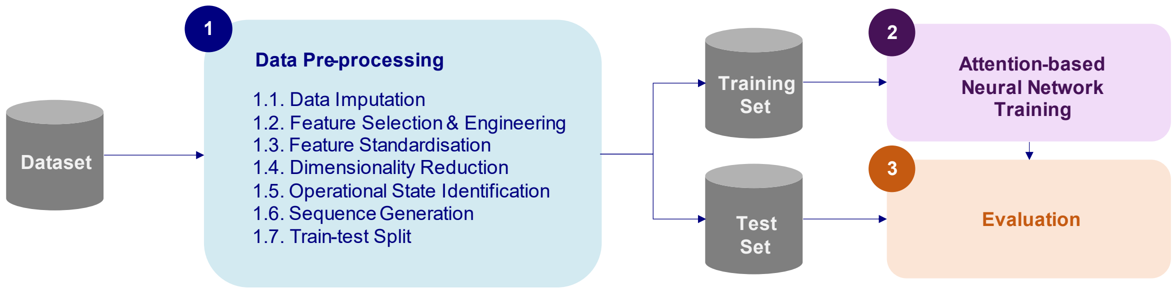

Figure 2.1. Graphical representation of the proposed methodology.

3.1. Data Pre-processing

As stated precedingly, a total of seven data pre-processing stages are applied to guarantee the adequate pre-processing of raw data.

To deal with potential missing values within the raw dataset, forward filling and backward filling is selected as the data imputation models. These two models are subject of analysis in this study, as they have been widely applied within the maritime domain when dealing with missing values [19], and univariate data imputation methods have presented promising results when dealing with short-term sensor data and Missing Completely at Random (MCAR) contexts [20]. Forward filling is applied when missing values are imputed with the preceding data point, whereas backward filling, by contrast, is considered when missing values are imputed with the subsequent data point.

To avert the utilisation of either irrelevant or redundant features available in the raw dataset when training the model, feature selection is performed. Consequently, correlation analysis is performed to select suitable features based on the task to be performed based on the relationship between features. To evaluate the linear relationship between features, the Pearson’s correlation coefficient is determined, which satisfies the equation hereunder.

$$\rho = \frac{1}{n-1} \sum_{i=1}^{n} \frac{x_i - \bar{x}}{s_x} \frac{y_i - \bar{y}}{s_y}, \tag{3.1.2}$$

where $\boldsymbol{x_i}$, $\boldsymbol{y_i}$ are individual data points, $\boldsymbol{\bar{x}}$, $\boldsymbol{\bar{y}}$ are the mean of the respective features being analysed, and $\boldsymbol{s_x}$, $\boldsymbol{s_y}$ relate to the standard deviation of the features. Additionally, to assess any potential non-linear relationship between features, the Spearman's rank correlation coefficient is also estimated. This coefficient analyses the potential relationship between features based on the rank of the data. Both coefficients always lie between -1 and +1. A resulting coefficient of either -1 or +1 indicates the features are perfectly correlated. By contrast, a coefficient of 0 establishes that no relationship exists between the analysed features. Thus, the strength of the relationship is obtained from the analysis of the magnitude obtained.

Feature standardisation is performed to guarantee that features contribute equally to the training of the model. When applying standardisation, all features are centred by subtracting the mean from all values. Results are then scaled by dividing their respective standard deviation. Thus, equation 3.1.3 is considered for standardising the features:

$$z = \frac{x - \bar{x}}{s} \tag{3.1.2}$$

The resulting standardised features will present a mean of zero and a standard deviation of one.

Dimensionality reduction is implemented to reduce the number of features to be considered for the training of the model while attempting to preserve as much information as possible that is presented within the dataset. Of all potential methods that can be applied for performing dimensionality reduction, the most common approach is the Principal Component Analysis (PCA), which is the method implemented in this study. This is an unsupervised learning technique that is mathematically defined as an orthogonal linear transformation. The main aim of this approach is to transform the data into a new coordinate system forming a new set of uncorrelated variables, which are named principal components.

To determine these principal components, a total of five phases are implemented. The first phase relates to the estimation of the covariance matrix to capture the fluctuations of a variable concerning the other features within the dataset. Then, the second phase is applied, which refer to the computation of both the eigenvectors and eigenvalues of the covariance matrix to identify the direction in which the data varies the most and the magnitude of such variations. The eigenvectors are then sorted based on the eigenvalues in ascending order to obtain the principal components, thus starting from the direction of maximum variance to minimum. The fourth phase select the number of principal components to be considered for the training of the model based on the resulting variance. For this study, the number of components will be selected when the accumulative variance of the retained principal components reaches 95%. The fifth and last step refer to the creation of a new matrix where lie the selected principal components based on their accumulative variance.

When analysing operational states of marine machinery, fluctuations may occur due to either environmental conditions or variations in the operating conditions. Accordingly, to only consider steady-state operational conditions, these fluctuations need to be adequately identified and discarded. In this study, Hierarchical Density-Based Spatial Clustering of Applications with Noise (HDBSCAN) is considered, as this approach require less parameters to be determined for adequately implementing the model if compared to DBSCAN. In fact, the only parameter that need to be estimated is the parameter MinPts. HDBSCAN is often regarded as an enhanced version of the DBSCAN, which facilitates a clustering hierarchy from which a tree of relevant clusters is constructed. These constructions are then utilised to derive a “flat” clustering based on cluster stability.

Once non-operational states have been removed from the data, sequences are generated through the application of the time series sliding window algorithm. This algorithm aims to section the overall time series into sequential and overlapping segments, a.k.a. windows, of a fixed size. Finally, the resulting sequences are split into a training set and a testing set in order to evaluate the model with unseen data (test set) that has not been used for training the model (training set).

3.2. Attention-based Neural Network Training

The primary contribution of this paper lies in the analysis of an attention-based neural network, which is introduced in this section. A neural network is commonly defined as an artificial intelligence model designed to emulate the human brain in a simplified manner. Accordingly, various types of neural networks have been developed to capture specific functionalities within the human brain. Of all these types, special attention is given in this study to those types of neural networks that consider temporal dependencies, as these are crucial for time series forecasting. Specifically, attention mechanisms are considered, as they can capture sequential dependencies and focus on relevant parts of the input sequences. Attention mechanisms have recently been subject of study due to their ability of focusing on specific segments of the sequence, thus identifying which parts of the input sequences are more relevant for the task being implemented. This is achieved by assigning weights to each part of the sequences, where higher weights indicate increased attention to that segment.

When integrated into the architecture of a neural network, attention mechanisms dynamically adjust their focus based on the information provided in the input sequence in order to generate the output. Thus, by incorporating attention mechanisms within the architecture of the neural network, an enhancement in the precision of the predictions is expected through the provision of special attention to specific segments of the input sequences. Even though several variants of the attention mechanism can be considered, the one introduced in this study is the scaled dot-product attention, which is comprised of the following sections:

- **Query, key, and value vectors.** As the name suggests, a total of three vectors are initially considered. The query vector (Q) represents the current input, the key vector (K) relates to the vector that emphasises the segments that require higher focus, and, finally, the value vector (V) indicates the values associated with each segment of the input sequence.

- **Weighted scores calculation.** A measure of similarity is introduced to assess how analogous the Q and K vectors are for each element in the sequence. The dot product is usually applied to assess this similarity. The result of this operation will indicate the relevance of each element in the sequence by a given query.

- ***Softmax* activation.** Once the weighted scores have been calculated, these are passed through a *softmax* activation function so that the resulting scores can be transformed into attention weights. As *softmax* is implemented, the sum of all weights equals 1, which enables its interpretation as probabilities. The resulting weights indicate whether an element of the input sequence require higher or lower focus, thus prioritising the identified relevant segments within the input sequences.

- **Output.** The value vector and the resulting attention weights are combined through the computation of a weighted sum to facilitate the output of the attention layer.

A diagram of the scaled dot-product attention is given in Figure 3.2.1.

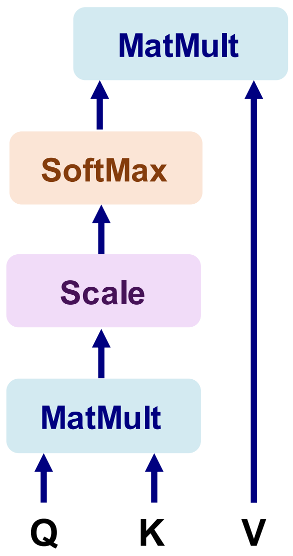

Figure 3.2.1. Graphical representation of the Scaled Dot-Product Attention.

Once considered the scaled dot-product attention, it is integrated to a Long Short-Term Memory (LSTM) neural network. LSTM has been selected for this study due to its capability to capture temporal dependencies. The main component of these networks is the memory cell. This consists of a cell state vector and gating units. The gating units are utilised to regulate the information flow into and out of the memory to maintain the state over time. A graphical representation of the developed architecture of the neural network is presented in Figure 3.2.2. As perceived, the neural network is comprised of blocks of LSTM with Attention Mechanism. The final part of the developed neural network is comprised by fully connected layers.

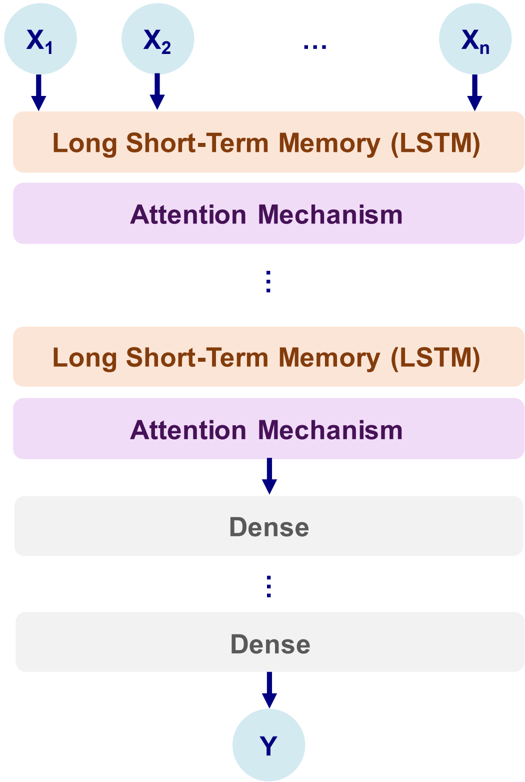

Figure 3.2.2. Graphical representation of the architecture of the developed neural network.

3.3. Evaluation

An evaluation phase is required to evaluate the results obtained from the training of the developed neural network that is analysed in this study. Consequently, the following metrics have been estimated: Mean Squared Error (MSE), Root Mean Squared Error (RMSE), Median Absolute Error (MedAE), and Max Error. MSE is estimated as it is probably the most utilised loss function for regression. It is computed by determining the mean of the sum of the squared errors. If the squared root of MSE if estimated, the RMSE is obtained. This is a scale-dependent error. MedAE is computed by determining the median of all absolute differences between the observed and the predicted occurrences. Finally, the max error is determined by considering the worst-case error.

4. CASE STUDY AND RESULTS

To analyse the performance of the attention mechanism-based neural network, the case study of a diesel generator of a tanker vessel is considered in order to predict its fuel consumption. In total, 26 ship machinery and weather condition parameters are considered.. The parameters considered in this study as well as their respective descriptive statistics are presented in Table A.1. More than 40,000 instances are analysed, collected in a 1-minute frequency basis. Additionally, missing values can be perceived, as the number of values per column differ from parameter to parameter. Thus, the firs stage of the data pre-processing phase needs to be implemented, which is the data imputation phase. Firstly, the percentage of missing values per feature has been estimated to assess their respective missingness criticality. A total of four parameters with high criticality have been identified, the percentage of missing values of which exceeds the 20%. The identified parameters are the following: DG1 Turbocharger Exhaust Gas Inlet Temperature (22.4%), Stern Tube Aft Bearing Temperature (41.9%), Stern Tube Fore Bearing Temperature (41.3%), and Water Depth (53.7%). Thus, as this is a preliminary study and the data imputation method considered for this analysis is not adequate when the percentage of missing values is large, the identified four parameters with high criticality are discarded from the analysis. However, more sophisticated data imputation methos should be considered to include relevant information from the discarded parameters in future studies. The remaining parameters present a low criticality, where the percentage of missing values is lower than 1.5%, except for the parameter to be predicted, which is the Fuel Oil Consumption (FOC) parameter, that presents a percentage of missing values of approximately 9%. Consequently, the forward filling data imputation method is implemented to impute the missing values. In those cases where forward filling could not be implemented backward filling was implemented instead. The resulting dataset after the implementation of missing values is a dataset with 22 parameters with no missing values.

The second step to be performed within the data pre-processing step is the Feature Selection & Engineering. Accordingly, to determine the relevant features subject to study in this analysis, both the Pearson’s correlation coefficient and Spearman’s rank correlation coefficient are computed. As analogous results are presented in both resulting matrices, the discussion of the correlation analysis results is provided altogether. Thus, to provide a better understanding of the results and relationships identified, the Pearson’s correlation matrix is graphically represented in Figure 4.1. Overall, the dependent variable (FOC parameter) is highly correlated with the diesel generator parameters, such as the exhaust gas outlet temperature of each of the six cylinders (Pearson’s correlation coefficient of 0.96) and the diesel generator power (Pearson’s correlation coefficient of 0.99). By contrast, the dependent variable is not correlated with weather conditions parameters, such as the speed over ground (Pearson’s correlation coefficient of 0.12) and wind speed (Pearson’s correlation coefficient of -0.04). In fact, an initial study has been conducted where weather conditions non-correlated parameters were included in the analysis. However, results outlined that including such conditions may lead to a decrease in the precision of the model. For this reason, those parameters that present a Pearson’s correlation coefficient lower than 0.8 were discarded from the analysis. Thus, even though weather conditions may be critical for the adequate prediction of FOC, it was decided in this study to remove them due to interference with the prediction. Accordingly, a total of 15 parameters remains.

If these 15 parameters are further analysed, it can be observed that mainly all of them are correlated with the remaining parameters. Thus, if these correlations are not addressed, it can yield multicollinearity challenges, model instability, and redundancy addition. Accordingly, these parameters need to be processed to obtain uncorrelated predictors. For this reason, Principal Component Analysis is performed in the dimensionality reduction phase, having performed feature standardisation prior to its implementation. If the cumulative variance presented in Figure 4.2. is then analysed, it can be observed that the adequate number of components is two. Accordingly, a total of two principal components are obtained, which are represented in Figure 4.3.

From the analysis of the resulting principal components, it can be observed that there are distinct transient and idle states that need to be adequately identified and discarded. To do so, the fifth stage of the data pre-processing phase, which is the operational state identification, is implemented. Accordingly, HDBSCAN technique is applied. After heuristic evaluation, the hyperparameter minPts is se to 5. If Figure 4.4. is observed, it can be determined that the shape of the three main clusters is well identified. However, this identification will not avert including potential transient states that lie within these three clusters. For this reason, domain knowledge is also required in this task to discard the transient states that the method could not detect, thus enabling a semi-automatic operational state identification framework. Firstly, based on domain knowledge, the two left clusters were discarded from the analysis, as these related to either idle or transient states. Regarding the third and final cluster, visual inspections were performed to discard any potential transient state that was undetected by the HDBSCAN algorithm.

Once all the identified idle and transient states were discarded, sequences could be generated for the training of the attention mechanism-based neural network. In total, 9000 sequences were generated. These resulting sequences were split into two sets so that the generalisation capabilities of the neural network could be assessed. The 70% of all sequences refer to the training set, whereas the remaining 30% relate to the test set.

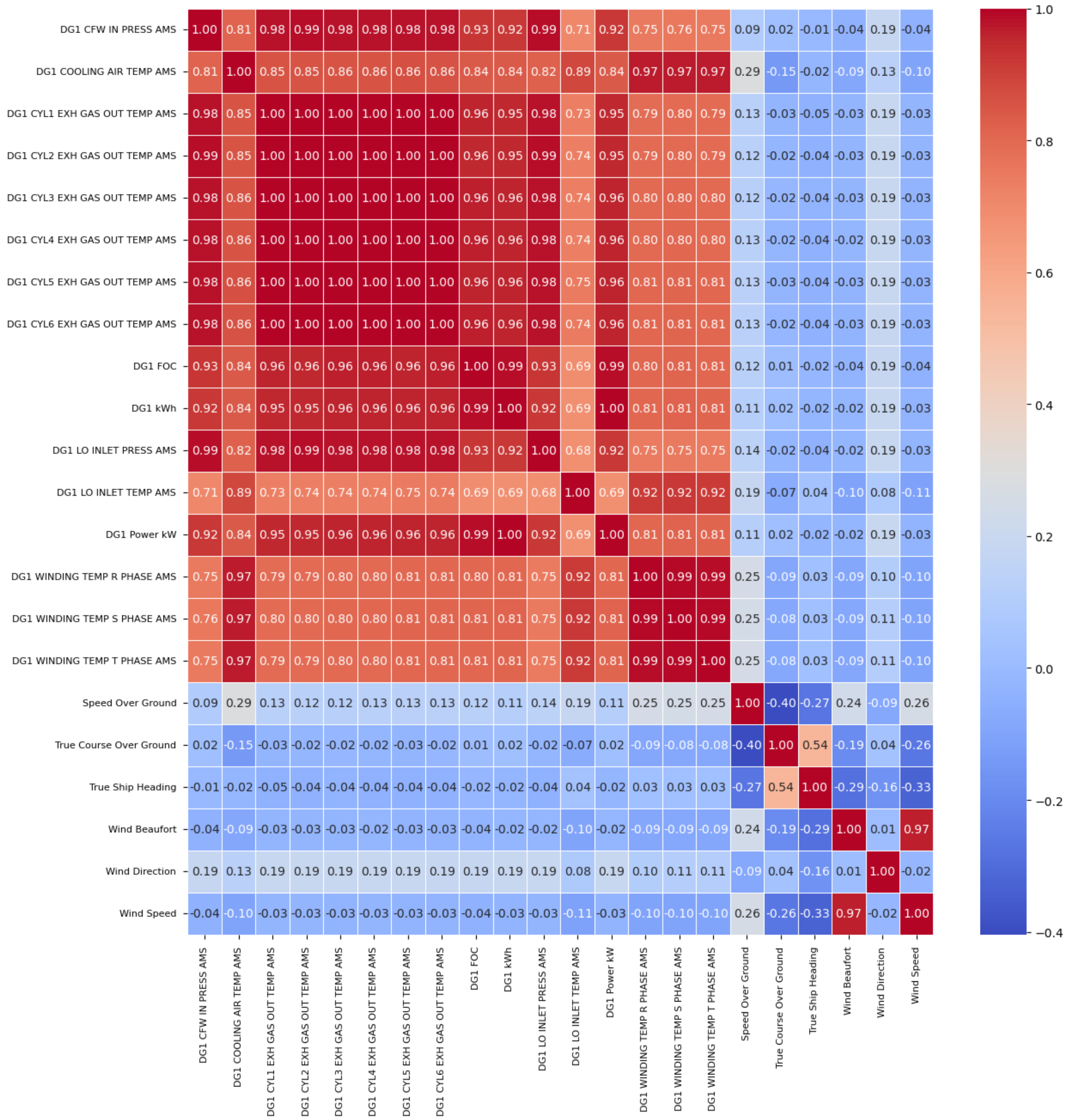

Figure 4.1. Pearson's correlation coefficient matrix.

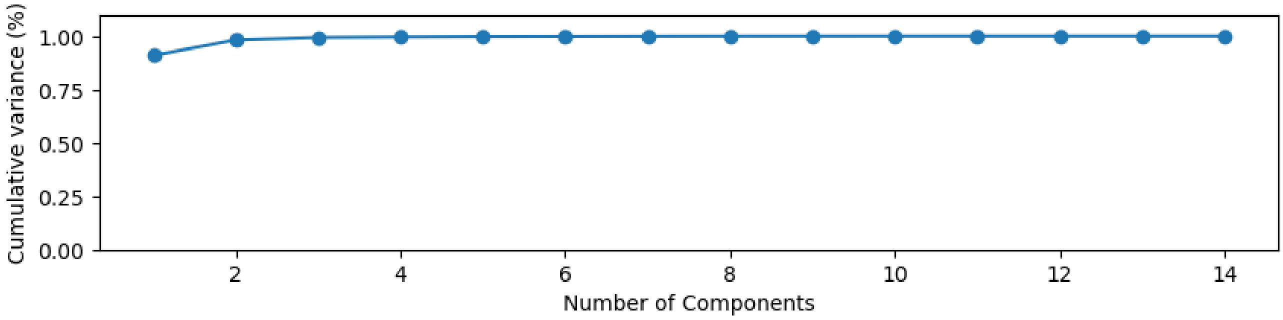

Figure 4.2. Number of components and their respective cumulative variance.

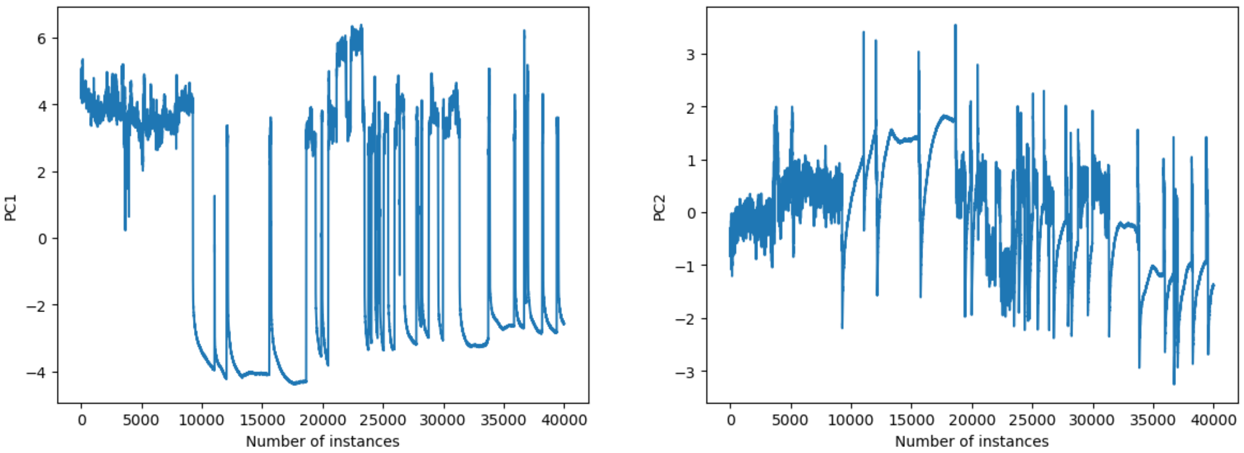

Figure 4.3. Graphical representation of the first and second principal component.

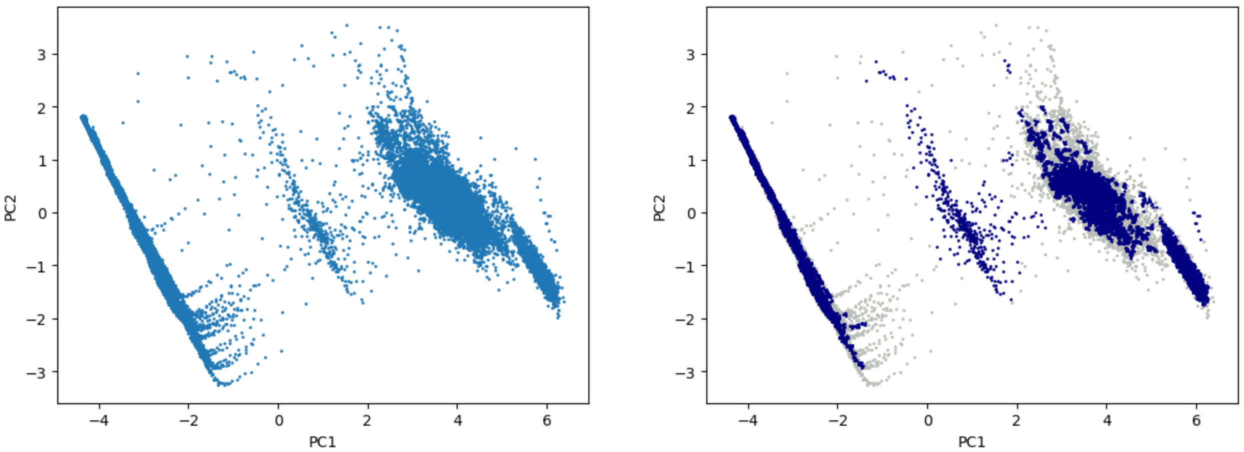

Figure 4.4. (a) Scatter plot of principal component 1 and 2 prior to the application of HDBSCAN (b) Scatter plot of principal component 1 and 2 subsequent to the application of HDBSCAN.

By splitting the dataset into two sets, the data pre-processing stage is finalised, and the developed neural network can then be trained. To do so, first the architecture of the neural network and the optimisation of the hyperparameters need to be determined. Consequently, hyperband optimisation was applied. Of all the analysed configurations, the one indicated in Table 4.1. was the one selected for the prediction of FOC.

Table 4.1. Configuration of the developed neural network.

Layer Type	**Output Shape**
Input	(None, 240, 2)
LSTM	(None, 240, 104)
Attention	(None, 240, 104)
LSTM	(None, 24)
Attention	(None, 120)
Dense	(None, 104)
Dense	(None, 104)
Dense	(None, 1)

The main results after the estimation of the evaluation metrics are presented in 4.2. As observed, promising results are obtained if, for instance, the RMSE is consider, as the models achieves a RMSE of 0.14 t/hr. However, a max error of 1.28 t/hr could be perceived. If this max error is comprehensively analysed, it can be observed that the instance relates to an anomaly in the sort-term context. Thus, this methodology could also be analysed for the detection of anomalies in short-term context in future work.

Table 4.2. Configuration of the developed neural network.

Evaluation Metric	**Attention-based Neural Network Result**
MSE	0.02
RMSE	0.14
MAE	0.07
Max. Error	1.28

It should be noted that these are initial results and further analysis needs to be conducted including a comparative study to adequately evaluate the results obtained. Moreover, self-attention has been considered through the implementation of scaled dot product attention. However, other attention mechanisms should also be considered, such as multi-head attention. Finally, include specific ship trips should be considered to have a more accurate evaluation of the predictive capabilities of the suggested model.

5. CONCLUSIONS

This paper presented an attention-based neural network to analyse the potential of attention mechanisms for the prediction of fuel oil consumption. Specifically, attention mechanisms were implemented subsequent to the implementation of long short-term memory layer in an attempt to capture relevant temporal dependencies that will facilitate the adequate prediction of fuel oil consumption. To assess the precision of the developed network, a case study on a total of 26 marine machinery and weather condition parameters is introduced. Results highlight the potential that this type of neural networks have when predicting fuel oil consumption. Nevertheless, further analysis needs to be performed to advance this area of knowledge through the application of artificial intelligence. Accordingly, the analysis of other attention mechanisms and the performance of a thorough case study is expected to be performed as part of the future work.

6. REFERENCES

[1] UNCTAD. (n.d.). Review of Maritime Transport 2023. [online] Available at: https://unctad.org/publication/review-maritime-transport-2023/ [Accessed 12 Mar. 2024].

[2] Imo.org. (2023). Marine Environment Protection Committee (MEPC 80), 3-7 July 2023. [online] Available at: https://www.imo.org/en/MediaCentre/MeetingSummaries/Pages/MEPC-80.aspx.

[3] Eide, M.S., Longva, T., Hoffmann, P., Endresen, Ø. and Dalsøren, S.B. (2011). Future cost scenarios for reduction of ship CO2emissions. Maritime Policy & Management, 38(1), pp.11–37. doi: https://doi.org/10.1080/03088839.2010.533711.

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

[5] Bertsimas, D., Delarue, A., Jaillet, P. and Martin, S. (2019). The Price of Interpretability. ArXiv. [online] Available at:

https://www.semanticscholar.org/reader/fda0f64e864ab98305c4d3ac08794c6ed019ff94 [Accessed 12 Mar. 2024].

[6] Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A.N., Kaiser, L. and Polosukhin, I. (2017). Attention Is All You Need. [online] arXiv.org. doi: https://doi.org/10.48550/arXiv.1706.03762.

[7] Abadi, M., Agarwal, A., Barham, P., Brevdo, E., Chen, Z., Citro, C., Corrado, G., Davis, A., Dean, J., Devin, M., Ghemawat, S., Goodfellow, I., Harp, A., Irving, G., Isard, M., Jia, Y., Jozefowicz, R., Kaiser, L., Kudlur, M. and Levenberg, J. (2015). TensorFlow: Large-Scale Machine Learning on Heterogeneous Distributed Systems. [online] Available at: https://static.googleusercontent.com/media/research.google.com/en//pubs/archive/45166.pdf.

[8] Paszke, A., Gross, S., Soumith Chintala, Chanan, G., Yang, E., DeVito, Z., Lin, Z., Alban Desmaison, Luca Antiga and Lerer, A. (2017). Automatic differentiation in PyTorch. [online] Openreview.net. Available at: https://openreview.net/forum?id=BJJsrmfCZ.

[9] Chollet, F., et al. (2015). Keras.

[10] Zhang, M., Nikolaos Tsoulakos, Kujala, P. and Spyros Hirdaris (2024). A deep learning method for the prediction of ship fuel consumption in real operational conditions. Engineering Applications of Artificial Intelligence, 130, pp.107425–107425. doi: https://doi.org/10.1016/j.engappai.2023.107425.

[11] Lei, L., Wen, Z. and Peng, Z. (2021). Prediction of Main Engine Speed and Fuel Consumption of Inland Ships Based on Deep Learning. Journal of Physics: Conference Series, 2025(1), p.012012. doi: https://doi.org/10.1088/1742-6596/2025/1/012012.

[12] Wang, H., Yan, R., Wang, S. and Zhen, L. (2023). Innovative approaches to addressing the tradeoff between interpretability and accuracy in ship fuel consumption prediction. Transportation Research Part C: Emerging Technologies, 157, pp.104361–104361. doi:

https://doi.org/10.1016/j.trc.2023.104361.

[13] Anh Tran, T. (2021). Comparative analysis on the fuel consumption prediction model for bulk carriers from ship launching to current states based on sea trial data and machine learning technique. Journal of Ocean Engineering and Science, 6(4), pp.317–339. doi: https://doi.org/10.1016/j.joes.2021.02.005.

[14] Su, M., Su, Z., Cao, S., Park, K.-S. and Bae, S.-H. (2023). Fuel Consumption Prediction and Optimization Model for Pure Car/Truck Transport Ships. Journal of Marine Science and Engineering, [online] 11(6), p.1231. doi: https://doi.org/10.3390/jmse11061231.

[15] El Mekkaoui, S., Benabbou, L., Caron, S. and Berrado, A. (2023). Deep Learning-Based Ship Speed Prediction for Intelligent Maritime Traffic Management. Journal of Marine Science and Engineering, 11(1), p.191. doi: https://doi.org/10.3390/jmse11010191.

[16] Panapakidis, I., Sourtzi, V.-M. and Dagoumas, A. (2020). Forecasting the Fuel Consumption of Passenger Ships with a Combination of Shallow and Deep Learning. Electronics, [online] 9(5), p.776. doi: https://doi.org/10.3390/electronics9050776.

[17] Wu, Z., Wang, S., Yuan, Q., Lou, N., Qiu, S., Bo, L. and Chen, X. (2023). Application of a deep learning-based discrete weather data continuousization model in ship route optimization. Ocean Engineering, 285, pp.115435–115435. doi: https://doi.org/10.1016/j.oceaneng.2023.115435.

[18] Bui-Duy, L. and Vu-Thi-Minh, N. (2020). Utilization of a deep learning-based fuel consumption model in choosing a liner shipping route for container ships in Asia. The Asian Journal of Shipping and Logistics. doi: https://doi.org/10.1016/j.ajsl.2020.04.003.

[19] Makridis G., Kyriazis D., Plitsos S., (2020). Predictive maintenance leveraging machine learning for time-series forecasting in the maritime industry. 2020 IEEE 23rd International Conference on Intelligent Transportation Systems (ITSC), pp.1-8, doi:

https://doi.org/10.1109/ITSC45102.2020.9294450.

[20] Velasco-Gallego C., Lazakis I., (2020). Real-time data-driven missing data imputation for short-term sensor data of marine systems. A comparative study. Ocean Engineering 218, pp. 1-23, doi:

https://doi.org/10.1016/j.oceaneng.2020.108261.

7. APPENDIX A. DESCRIPTIVE STATISTIC OF THE MONITORED PARAMETERS

Table A.1. Descriptive statistics of the monitored parameters

	DG1 CFW In Press. (bar)	DG1 Cooling Air Temp. (ºC)	DG1 Cyl1 Exh. Gas Out. Temp. (ºC)	DG1 Cyl2 Exh. Gas Out. Temp. (ºC)	DG1 Cyl3 Exh. Gas Out. Temp. (ºC)	DG1 Cyl4 Exh. Gas Out. Temp. (ºC)	DG1 Cyl5 Exh. Gas Out. Temp. (ºC)	DG1 Cyl6 Exh. Gas Out. Temp. (ºC)	DG1 Fuel Oil Consumption (t/hr)
count	41750	41443	41439	41439	41439	41439	41439	41440	37982
mean	2.81	37.66	196.82	210.56	210.52	210.90	191.10	203.07	0.90
std	1.18	6.68	140.84	154.30	154.66	156.43	142.15	150.04	1.02
min	1.50	25.50	55.40	57.00	57.20	58.10	47.60	49.20	0.00
25%	1.70	32.50	67.30	67.30	67.00	65.70	57.90	64.30	0.00
50%	1.80	37.70	73.70	76.00	76.90	76.00	70.70	76.00	0.00
75%	4.10	43.10	348.30	375.70	375.10	378.90	343.50	360.60	1.85
max	4.50	52.60	411.70	448.50	442.30	451.30	421.10	438.70	5.82

	DG1 kWh	DG1 Lubrication Oil Inlet Press (bar)	DG1 Lubrication Oil Inlet Temp (ºC)	DG1 Power (kW)	Inlet Temp. (ºC)	DG1 Winding Temp R Phase (ºC)	DG1 Winding Temp S Phase (ºC)	DG1 Winding Temp T Phase (ºC)	Speed Over Ground (m/s)
count	41487	41706	41382	41487	32387	41443	41443	41442	41336
mean	2.07	2.27	56.40	124.47	294.18	42.38	43.89	42.85	6.32
std	2.42	1.82	7.02	145.07	174.50	8.60	8.84	8.33	5.96
min	0.00	0.30	40.60	0.00	45.10	25.50	26.70	26.60	0.00
25%	0.00	0.60	50.20	0.00	82.40	36.00	37.30	36.80	0.02
50%	0.00	0.80	60.90	0.00	413.30	43.60	45.00	43.70	6.98
75%	4.16	4.20	62.10	249.87	439.60	48.70	50.30	48.90	12.51
max	9.27	5.00	68.20	555.93	510.10	67.60	68.50	67.30	15.32

	Stern Tube Aft Brg Temp (ºC)	Stern Tube Fore Brg Temp (ºC)	True Course Over Ground (º)	True Ship Heading (º)	Water Depth	Wind Beaufort	Wind Direction (º)	Wind Speed (m/s)
count	24258	24524	41335	41258	19344	41162	41334	41334
mean	28.55	28.99	165.10	147.72	39.51	3.61	189.16	6.80
std	4.96	4.97	88.84	72.81	90.55	2.42	120.69	5.64
min	20.00	20.00	0.00	0.00	1.70	0.00	0.00	0.00
25%	24.50	24.80	100.70	110.12	4.10	2.00	72.00	2.06
50%	27.40	28.10	164.80	148.12	9.70	4.00	218.00	6.17
75%	33.30	33.60	242.00	172.89	48.20	5.00	300.00	9.77
max	37.60	37.90	359.90	359.97	830.60	10.00	359.00	27.27