Data-driven ship berthing forecasting for cold ironing in maritime transportation

Bakar, Nur Najihah Binti Abu; Guerrero, Josep M.; Bazmohammadi, Najmeh; Vasquez, Juan C.; Uyanik, Tayfun; Çimen, Halil

Published in:
Applied Energy

DOI (link to publication from Publisher):
https://doi.org/10.1016/j.apenergy.2022.119947

Publication date:
2022

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

Data-driven ship berthing forecasting for cold ironing in maritime transportation

Nur Najihah Abu Bakar[1,2,*], Najmeh Bazmohammadi[1], Halil Çimen[3], Tayfun Uyanik[4], Juan C. Vasquez[1], Josep M. Guerrero[1]

[1]Center for Research on Microgrids (CROM), AAU Energy, Aalborg University, 9220 Aalborg East, Denmark; nurnbab@energy.aau.dk; naj@energy.aau.dk; juq@energy.aau.dk; joz@energy.aau.dk.

[2]Faculty of Electrical Engineering Technology, University Malaysia Perlis (UniMAP), Kampus Pauh Putra, 02600 Arau, Perlis, Malaysia; nurnajihah@unimap.edu.my.

[3]Department of Electrical-Electronics Engineering, Konya Technical University, Konya, Turkey (e-mail; hcimen@ktun.edu.tr.

[4]Istanbul Technical University, Maritime Faculty, 34940, Tuzla, İstanbul, Turkey; uyanikt@itu.edu.tr.

* Correspondence: nurnbab@energy.aau.dk

Abstract

Cold ironing (CI) is an electrification alternative in the maritime sector used to reduce shipborne emissions by switching from fuel to electricity when a ship docks at a port. During the ship's berthing mode of operation, accurately estimating the berthing duration could assist the port operator to manage the berth allocation and energy scheduling optimally. However, the involvement of multiple input parameters with a large dataset requires a suitable handling method. Thus, this paper proposed a data-driven approach for ship berthing forecasting of cold ironing with various models such as artificial neural networks, multiple linear regression, random forest, decision tree, and extreme gradient boosting. Meanwhile, RMSE and MAE are two main indicators applied to assess forecasting accuracy. The simulation-based result shows that the artificial neural network outperforms all other models with the lowest error performance of RMSE (3.1343) and MAE (0.2548), suggesting its capability to handle nonlinearities in complex forecasting problems of port activity. The high accuracy of forecasting output in this study, which is berthing duration contributes to close estimation of two info: 1) CI power consumption and 2) departure time of the ship. This information is vital to the port operator to be used in the energy management system (EMS) as well as in the berth allocation problem (BAP).

Keywords: Cold ironing, data-driven, electrification, emission, forecasting, ship transportation.

Nomenclature

Acronyms/abbreviations

AI	Artificial intelligence
ANN	Artificial neural network
AIS	Automatic identification system
AES	All-electric-ship
BAP	Berth allocation problem
BPNN	Back-propagation neural network
CI	Cold ironing
CISF	Cold ironing ship berthing forecasting
CO2	Carbon dioxide
EMS	Energy management system
ESS	Energy storage system
ECA	Emission control area
GBR	Gradient boosting regression
IMO	International maritime organization
KNN	K-nearest neighbor regression
KPI	Key performance indicator
LTLF	Long-term load forecasting
LR	Linear regression
LSTM	Long short-term memory
ML	Machine learning
MLR	Multiple linear regression
MTLF	Medium-term load forecasting
MAE	Mean absolute error
MSE	Mean square error
RBFNN	Radial basis function neural network
RES	Renewable energy source
RF	Random forest regression
RMSE	Root mean square error
R^2	Coefficient of determination
STLF	Short-term load forecasting
VSTLF	Very short-term load forecasting
XG Boost	Extreme gradient boosting

Variables/parameters/sets

a_n	MLR coefficient
b	Scalar parameter of ANN
$f_k(x_i)$	The function of input k-th of the decision tree
x_i	Input samples of ANN
w_i	Weight elements of ANN
$\hat{y}_i$	The predicted value for XG Boost

1. Introduction

Shipborne emissions from maritime transportation contribute to significant climate change, with shipping activities accounting for 2.5% of global CO2 emissions [1]. On the other hand, the health effects caused by ships' emissions are considerable. As the alternative, cold ironing (CI) is implemented on the shoreside to provide electricity to the ship, replacing the polluted fuel-based auxiliary engines. During the idle state (berthing) of the visiting ship at port, some of the onboard devices and hoteling loads need the power to be maintained. Thus, only the main propulsion engines are turned off, while the auxiliary engines are kept on supplying the essential load, such as lighting, communication devices, alarm system, and crew living space. These auxiliaries' generator mainly burns heavy diesel oil and emits hazardous emissions into the atmosphere during their stay [2]. Due to the strict sulfur control by IMO and ECA, shore-to-ship power supply or known as CI gaining attention from the port operators and shipping lines to solve the emissions problem during the berthing mode of operation. This electrification technology supplies the onboard loads by connecting the ship to the shore power system allowing the auxiliary engines to be turned off and preventing emissions [3]. Existing installations of the CI infrastructure, such as those at the Port of Gothenburg, Port of Los Angeles, and Port of Stockholm, provide insight into the market acceptability of this shore-to-ship technology [4]. It is estimated to reduce carbon emissions by 800,000 tons if all European ports deploy CI facilities [5]. Thus, CI implementation at ports becomes one of the progressive strategies toward climate neutrality by 2050 complying with the Paris Agreement's legal convention on climate change in 2015 [6].

Despite its contribution that reducing emissions at the port during the berthing mode of operation, the CI system requires a substantial amount of energy from the shore power, putting pressure on the main grid [7]. Besides that, the ports' inadequate electric power supply capacity may limit its potential for providing CI services to numerous ships that arrive at the same time. Considering the insufficiency of the sole reliance on the utility grid, the concept of a seaport microgrid to satisfy load requirements from CI and shoreside activities has emerged. The framework of a typical seaport microgrid for CI is illustrated in Fig. 1.

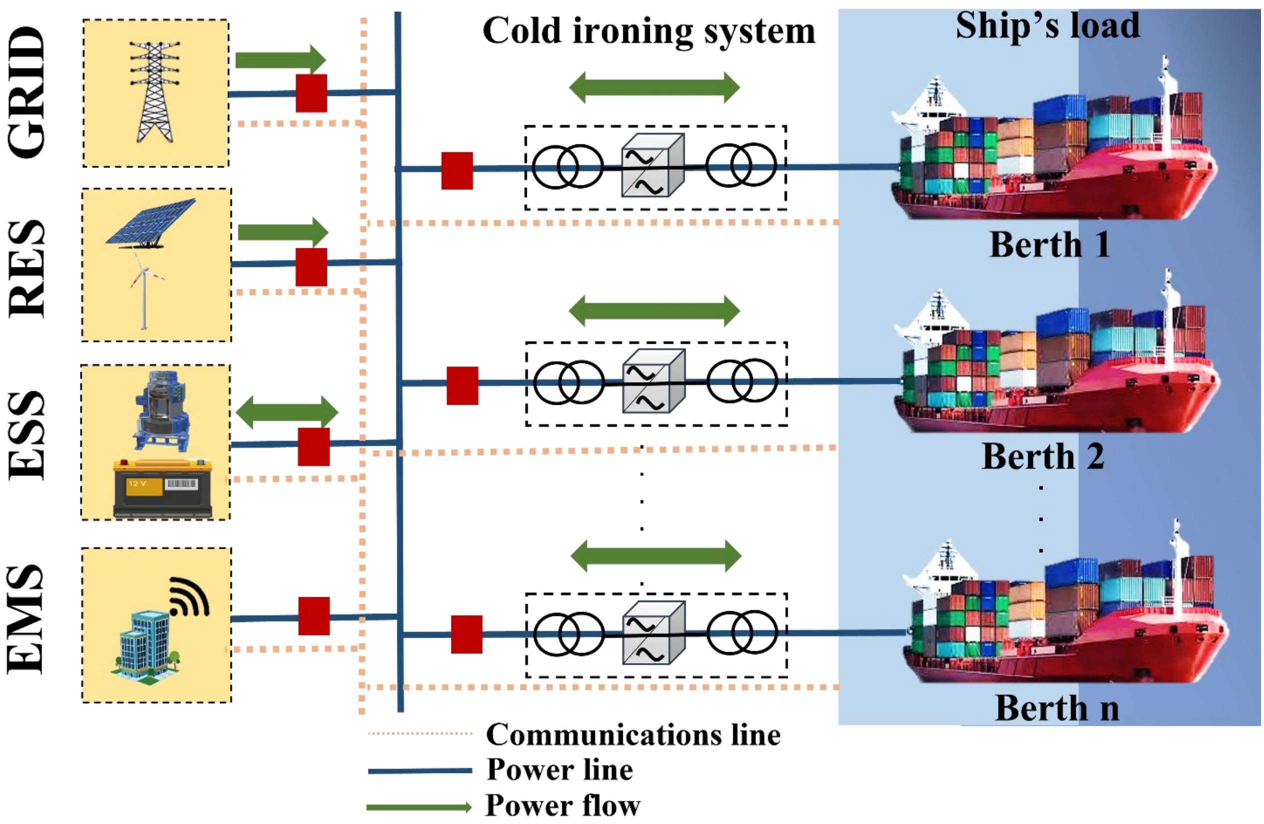


Fig. 1. Seaport microgrid with cold ironing system. (Acronyms: renewable energy sources (RES), energy storage system (ESS), energy management system (EMS))

Microgrid and CI are two of the most prominent maritime electrification techniques showing the evolution of energy from diesel to electricity. The well manages coordination between the seaport microgrid and CI is essential to ensure the power balance and maximize resource utilization. Interruption in power delivery might cause a huge disruption in ports operation and result in considerable economical losses for related parties. To prevent any power failure, efficient energy management is of vital importance for multifaceted decision-making to maintain a reliable energy supply. Furthermore, understanding the load behavior is an essential requirement for the planning and operation management of seaport microgrids. However, several factors influence the dynamic load of CI in seaport microgrids, necessitating the deployment of advanced forecasting techniques.

Existing studies on CI forecasting such as in [8], perform the forecasting technique by considering the aspect of electricity price as the forecasting output. However, forecasting CI's electricity price only helps the ship owner during the voyage mode of operation for them to decide the best route to go for minimal operation cost. None of the available research focuses on the berth terminal itself, where CI is allocated and consumed. During the berthing mode of operation at the port terminal, the visiting ship is connected to the CI power supply on the shoreside to continuously supply the onboard load while waiting for the completion of loading/unloading operations. The berthing duration of the ship at the berth terminal during this state is important for better management of CI particularly for the port operator to assign the best berth allocation of the incoming ship. It is also crucial for the energy-related authorities in the port area to get an accurate estimation for them to face the increased energy demand from port activities. This is supported by the recent publication in [9] that integrate bilevel port microgrid scheduling incorporating cold ironing and berth allocation problem (BAP). In that bilevel's methodology, both stages (BAP and microgrid's EMS) require the ship's berthing duration information. However, due to the lack of technique to acquire the info, the data used is only based on observation and assumption. This research gap highlighted the significance of ship berthing forecasting in port EMS.

It is to be noted that demand for better management in cold ironing's operation is strongly related to the behavior of its main consumer, which is ship transportation. The berthing duration of the incoming ship is a crucial aspect, as a longer stay at the port consumes more energy from CI and may cause traffic congestion as well as prolonged waiting times for other ships. Duration of berthing also can provide a projection of how much pollution can be prevented with the help of CI electrification.

From this perspective, this paper aims to forecast the berthing duration of the ship at the port terminal.

However, accurately estimating the ship's berthing duration is a challenging task as it is reliant on several factors. Understanding all the possibilities of ship-related factors is beneficial for formulating a good forecasting algorithm with the lowest error deviation. Among possible associated parameters are the type of the visiting ship, the frequency of calls per ship type, the capacity of the ship, the hour of arrival, the size of the ship, and the ship's mode of operation. This initial assumption for the input-output relationship in forecasting needs verification to clarify the logic behind all the initial assumptions. The main issue is a huge quantity of data and varying data for each category make it a complex process for accurate forecasting without appropriate approaches. Thus, data-driven integration is needed as it is capable to extract meaningful insights (e.g. patterns, complex relations, correlation variables) over huge quantities of data and generate novel outcomes with the help of artificial intelligence (AI). Yet, some studies are still considering non-data-driven due to the requirement for human resources, such as, [10], manually splitting the engine loads into five unique operating circumstances to execute normalized value. The drawbacks of human interference in the process are time-consuming and increased risk of human error during prediction [11]. With the arising concern mentioned above, the objective of this study is to forecast ship berthing duration for cold ironing by incorporating a data-driven approach to achieve the lowest error performance.

Accordingly, this paper presents a data-driven based for cold ironing ship berthing forecasting (CISF) with the following contributions:

1) This paper focuses mainly on a forecasting technique for CI systems at the berth terminal during a ship's berthing mode of operation, in contrast to existing research works that incorporate a forecasting module as a minor component in their framework model due to CI's indirect engagement with other applications, which is ship's voyage scheduling.
2) To address the uncertain number of influence factors that affect CI, this paper provides insight into the correlation between potential inputs and their effect on the expected output. This input-output formulation is significant to capture the vital trend and create a high-accuracy forecasting model that is capable to predict future activities and suggesting necessary action in advance for optimal outcomes.
3) Several data-driven forecasting methods were applied namely, artificial neural network (ANN), multiple linear regression (MLR), decision tree, random forest, and extreme gradient boosting (XG Boost). The performance analysis of these comparative methods will provide the best forecasting model for CISF with minimum error.

The rest of this paper is structured as follows. In section 2, the related research literature is presented. Section 3 is dedicated to describing the forecasting methodology including data pre-processing, correlation analysis between variables, and algorithms used in the data-driven model. Meanwhile, section 4 analyzed the forecasting performance for each algorithm and the finding from the simulation result is discussed in section 5. Finally, the significant findings of the paper are summarized in Section 6.

2. Literature review

Many forecasting studies in the maritime sector have been performed on various applications with different methods including data-driven strategy. Data-driven is the powerful method of statistical pattern recognition paradigm from a set of data and generates high accuracy forecasting output [12]. Research publication from [13] implements data-driven machine learning (ML) to predict the fuel consumption of the ship. A set of data is trained with different ML methods including bayesian ridge, kernel ridge, multiple linear, and ridge regressions. An algorithm with the lowest error performance is used for the future forecasting of fuel consumption.

In the case of CI, an all-electric ship (AES) is among its main consumers. A few research studies conduct a forecasting technique that involves CI and AES cooperation. This electrification of the ship generally utilized an energy storage system (ESS) to minimize both costs of operation and emission [14]. The ESS of the AES can be charged by using the CI facility when the ship docks at the port. Due to this, [15] performed the ship propulsion load forecasting with two inputs (sea state and distance path of the ship) to manage the ship's route and speed with the help of CI. The ship's load deviation from the forecasting output is then sub to available units of the generators/CI/ESS. This robust optimization technique is to ensure efficiency in AES's energy management and provides optimal voyage scheduling. In comparison, despite performing an uncertainty analysis on the ship's load, Zhao et al. [8] on the other hand evaluate the impact of the CI on the voyage and generation scheduling under stochastic CI electricity prices at a different port. More specifically, information on the CI electricity will assist the AES control administrator to determine the best berthing location and time for charging the ESS with economical practices.

Wen S et al. [16] integrate a deep learning-based forecasting technique with four different methods (BPNN/RBFNN/Elman NN/LSTM) to predict day-ahead electricity prices at three different ports with a 12-month dataset considering different factors including various seasons, working days, and holidays. The simulation result shows that with the help of CI electricity price estimation, scheduling the AES when the price is relatively low can greatly reduce the operational cost, sailing time, and emissions. Besides, some of the research studies utilize shore electricity price forecasting for the berth allocation problem (BAP) to minimize the total handling time, ship waiting time, and operational cost [9]. BAP case study in [17] integrates the synergy between seaport microgrid and CI where problem formulation involves shore electricity price. Meanwhile, considering the volatile nature of the renewable energy sources (RES) generation in port microgrids, Conte F et al. [18] use forecasting to estimate the RES production and compensate for the forecasting error to the available ESS. The compensation is applied when the forecast value exceeds the actual generation of RES, the error deviation is corrected by discharging energy from ESS. Similarly, when the forecasting value is less than the actual generation of RES, the error deviation is resolved by storing the excessive generation into the battery. Reliability of the generation sources from the port microgrid is significant to providing sustainable supply to the CI facilities, ensuring a smooth operation at the port terminal. Table 1 shows other research publications related to the CI case study that executes forecasting techniques.

Table 1

Forecasting strategy in the cold ironing case study.

Year/citation	Method	Input data	Output forecasting
2022 [8]	Not available	Not available	Electricity price
2022 [9]	Not available	Not available	-Fix and critical ship's load -Renewable energy source (RES)
2021 [19]	Deep learning	-24 hours electricity prices -Load demand	Electricity price
2021 [20]	-Environmental and Protection Agency (EPA) formula -The monitoring, Reporting, and Verification (MRV) formula	-Different types of the ships -Load factor while hoteling -Auxiliary to main engine ratio	Load of the ship
2021 [21]	Feed-forward Artificial	-Hourly load	Load profile

	Neural Network (ANN)	-Day lagged load -Week lagged load	
2020 [22]	-Gradient Boosting Regression (GBR) -Random Forest Regression (RF) -BP Network (BP) -Linear Regression (LR) -K-Nearest Neighbor Regression (KNN)	-Net tonnage -Deadweight tonnage -Actual weight -Efficiency of facilities	Load of the ship

Even though there have been several studies on the forecasting strategies associated with CI application, which are listed in Table 1, none of it solely focuses on the CI system itself. Mostly due to the indirect involvement of the CI and its implications for other maritime applications, particularly in the case of voyage schedules. However, it is only beneficial to the ship owner for decision-making during voyage mode of operation. In addition, the forecasting technique presented in the research publication is not mentioned in detail and no further discussion on forecasting part. Meanwhile, the CI system is located and consumed at the port terminal suggesting that action during the berthing mode of operation cannot be ignored. A suitable berth station with an adequate CI's power capacity must be assigned to the incoming ship as soon as it enters port. It is thus important to forecast ship berthing duration for further implementation in port management affairs. The following subsection describes the detailed methodology used for the CISF:

3. Methodology

3.1. Problem description

In this subsection, the forecasting models of the CISF are explained. All the forecasting models were executed by using Spider (Phyton 3.9) and Matlab interface. The model is expected to be capable of imitating the volatile behavior of the ship's berthing duration with minimum error based on several variables. Forecasting berthing duration is beneficial to the CI's terminal operator in two ways, whereas the forecasting data can be used to estimate the ship's departure time and CI power consumption. Thus, any necessary action can be planned to improve ship management at the port and prevent unwanted event. Fig. 2. illustrates the forecasting model in the case of CISF.

Berthing duration has a significant impact on CI applications as it implies longer berthing hours consuming higher shoreside power. Identifying CI power consumption in advance gives the advantage to the port in scheduling the generation with economical practices. Hein K et al. [23] carry out day-ahead operation planning in a seaport microgrid for scheduling low-cost clean energy and incorporating CI to optimize port performance. The formulation to estimate the power demand of berthed ships by using departure time and the influence of berthing hour on CI consumption is explained in [24]. Besides, estimation of the ship's departure time from CISF is necessary for optimal management of ship allocation to avoid traffic congestion, longer waiting time, and minimizing the handling duration at the terminal. Ship waiting time issues and their problem formulation with a different approach are discussed in [25][26][27]. Meanwhile, [28][29] formulate 'expected departure ship' as one of the important data in the BAP-solving mathematical algorithm. The proposed approach assists the port operator in determining the best berthing time, berthing location, and quay/yard equipment allocation for the arriving ship. Low accuracy of ship departure prediction could impact negatively on port efficiency. Thus, the CISF model with high precision assists minimal error in the estimation of CI consumption and ship departure time, necessitating the deployment of forecasting techniques as described in the following subtopic.

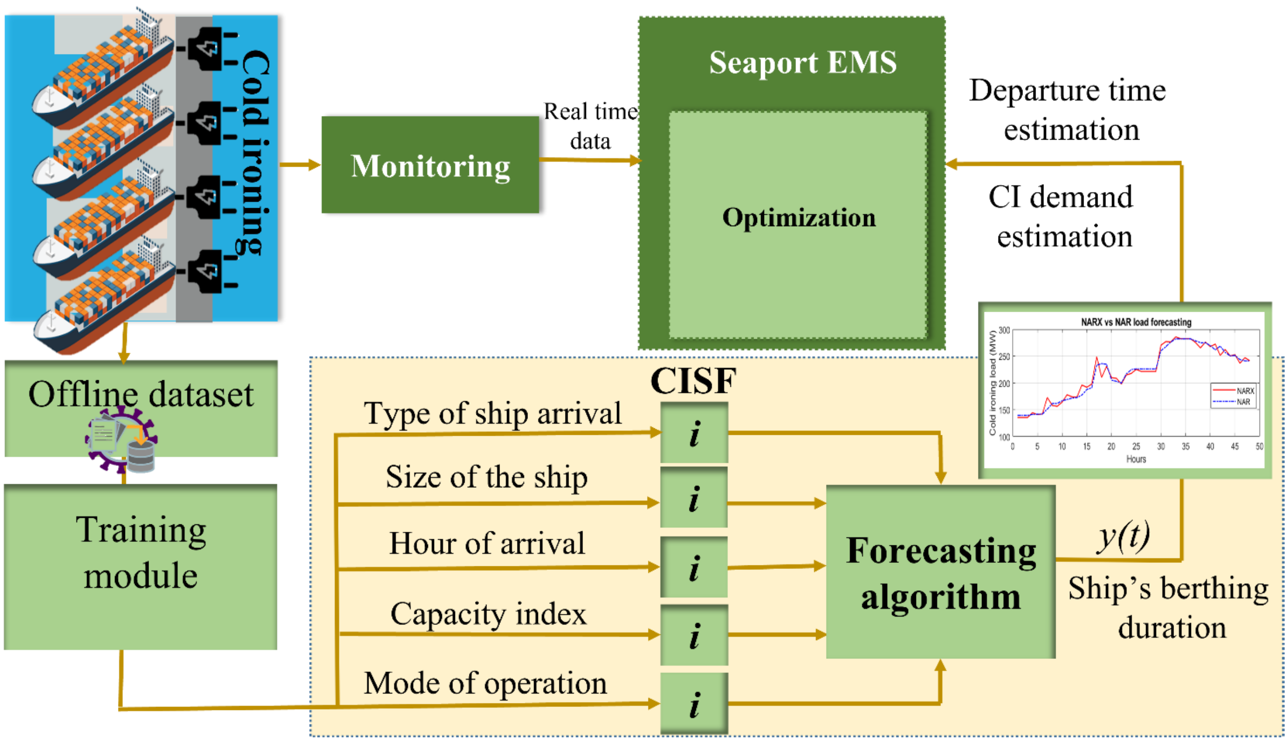


Fig. 2. CISF forecasting inputs and output.

3.2. Data pre-processing

The maritime dataset for the CISF case study is obtained from Port of Aalborg's ship tracking [30]. Fig. 3. shows the pattern of ship arrival at Port of Aalborg for the one-year duration from February 2021 until Jan 2022. There are different types of visiting ships, and the frequency of arrivals spikes during the summer season, especially for leisure boats. However, CISF only considers cargo and tanker as they consume long berthing hours, have a huge size, and make regular port calls. Prousalidis J et al. [31] emphasize that CI is most advantageous when applied to ships that visit the same port frequently. In these cases, cargo and tanker have a great potential for CI implementation due to their consistency of ship calling throughout the year because their supplies must be delivered regardless of the seasons. Fig. 3 shows the ship frequency arrival at Port of Aalborg for all types of ships in a one-year duration.

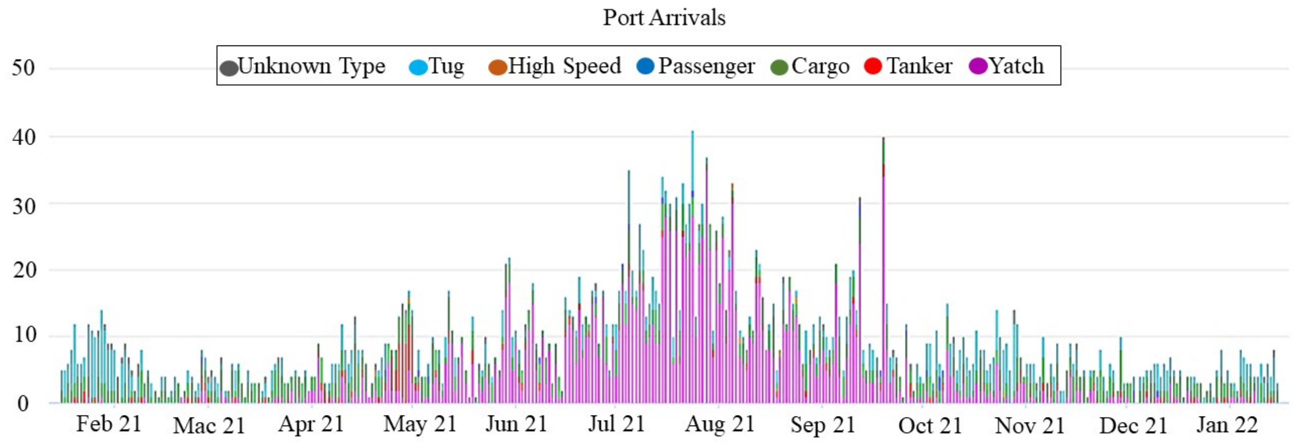


Fig. 3. Ship frequency arrival at Port of Aalborg in one-year duration.

Apart from the type of the ship, a nonlinear relationship is detected between the varying number of arrivals and departures of ships per hour, ships berthing duration, and power demand per ship as shown in Fig. 4. Each of these aspects stimulates changes in the shore-to-ship power consumption resulting in a dynamic load behavior. Thus, considering these factors is necessary to ensure that the training module can imitate the desired output with minimal error. Also, the different mode of ship operation such as loading/unloading, maintenance, or transit, has a substantial influence on the berthing duration. Each mode of operation consumes a different berthing duration where loading/unloading activities require a longer berthing time compared to the ship's maintenance and transit activities.

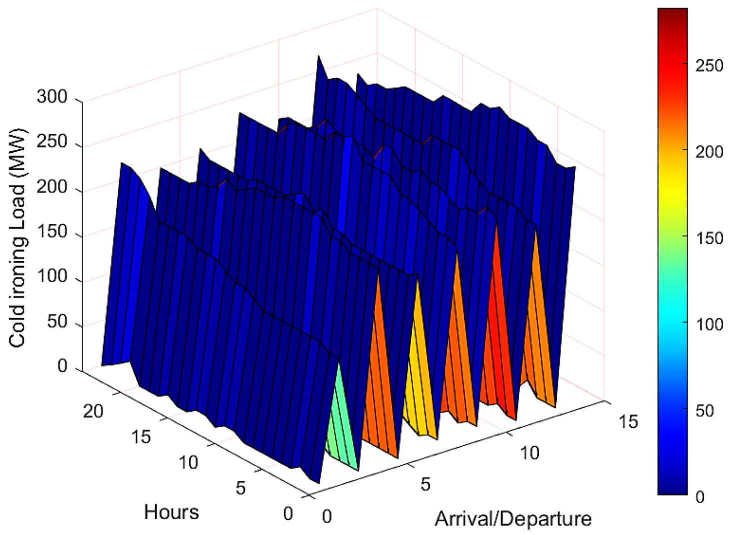


Fig. 4. Relationship between the number of ship's arrival/departure at each hour, berthing hours, and cold ironing consumption.

Another important step in load forecasting is to identify an appropriate time horizon before selecting a load forecasting model. The choice is also influenced by the forecasting application and purpose. It can be divided into four different types of time horizons as follows [32][33]:

1) Very short-term load forecasting (VSTLF) - A few seconds to several minutes.

2) Short-term load forecasting (STLF) - An hour to one week.

3) Medium-term load forecasting (MTLF) - A week to one year.

4) Long-term load forecasting (LTLF) - More than one year.

The time horizon used in the CISF training module is medium-term duration by using a month dataset where $t \in 1,2,3 \ldots 672\ (h)$. Considering the ship-berthing nature that varies from a few hours to a few days, 672 hours timeframe is ideal for observing the pattern of CISF. The dataset has several input parameters (time of arrival, type of the ship, size of the ship, mode of operation, ship index capacity) and berthing duration is the output parameter, all of which are statistically examined as summarized in Table 2. The statistical analysis is used in the data cleaning process to detect the redundancies, outliers, null values, and missing data that might hinder the training progression. Fig. 5. illustrates the overall process of CISF model development.

Table 2

Statistical variables of the data set.

	Arrival time (a.m/p.m)	Cargo arrival	Tanker arrival	Cargo size (m^2)	Tanker size (m^2)	Cargo mode of operation	Tanker mode of operation	Cargo index capacity	Tanker index capacity	Cargo berthing (h)	Tanker berthing (h)
Count	672	672	672	672	672	672	672	672	672	672	672
Mean	11.5	0.079	0.03	118.26	39.63	0.077	0.013	8.444	5.01	2.8476	0.311
Std	6.93	0.27	0.17	498.06	544.63	0.27	0.12	59.41	65.72	15.84	2.58
Min	0	0	0	0	0	0	0	0	0	0	0
Max	23	1	1	5510	13152	1	1	1056	1644	264	37

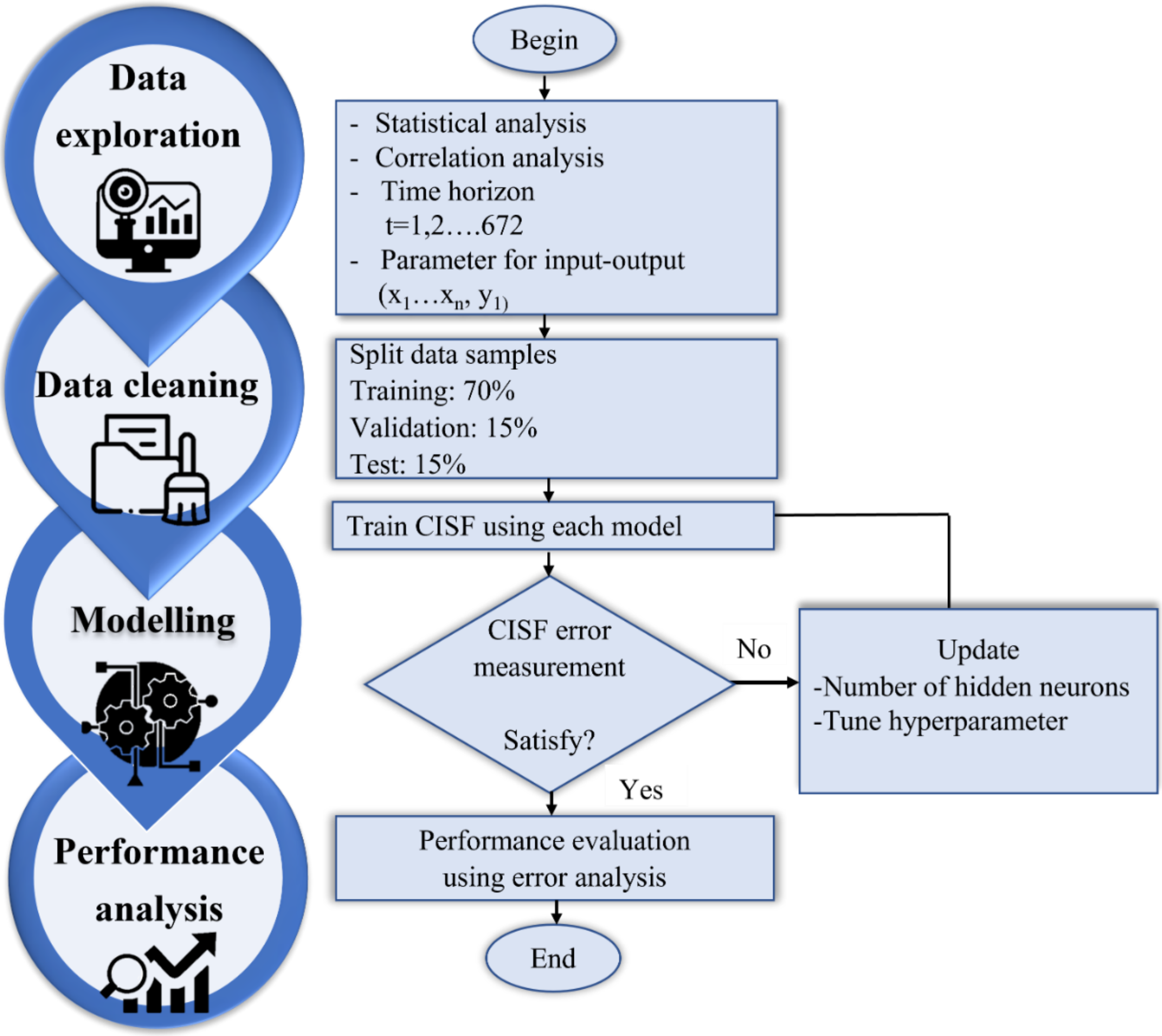


Fig. 5. CISF flow chart.

3.3. Correlation analysis

The relationship between any two variables can be assessed by using correlation analysis. Fig. 6. shows the mapping of Pearson correlation between input-output variables in CISF. It indicates true correlation exists between variables when the coefficient approaches one. Fig. 7. provides a more detailed statistical analysis of the association between variables including;

- Time of arrival
- Type of the ship
- Size of the ship
- Mode of operation
- Index capacity
- Berthing duration

It can be observed that higher index capacity and late arrival time result in increasing the berthing duration. Furthermore, a linear relationship is formed between the size of the ship and the berthing hour.

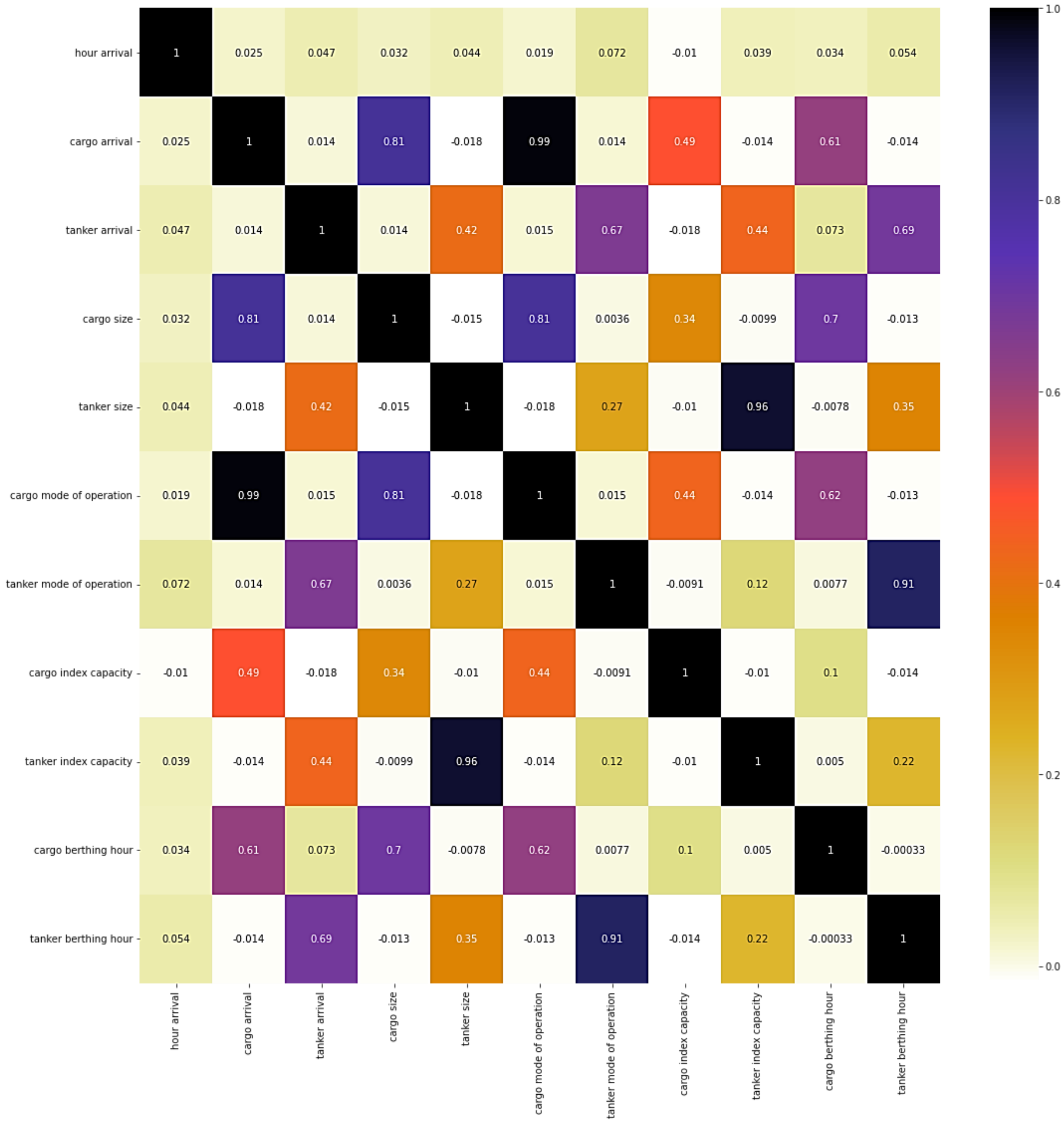


Fig. 6. Pearson correlation matrix of the CISF dataset.

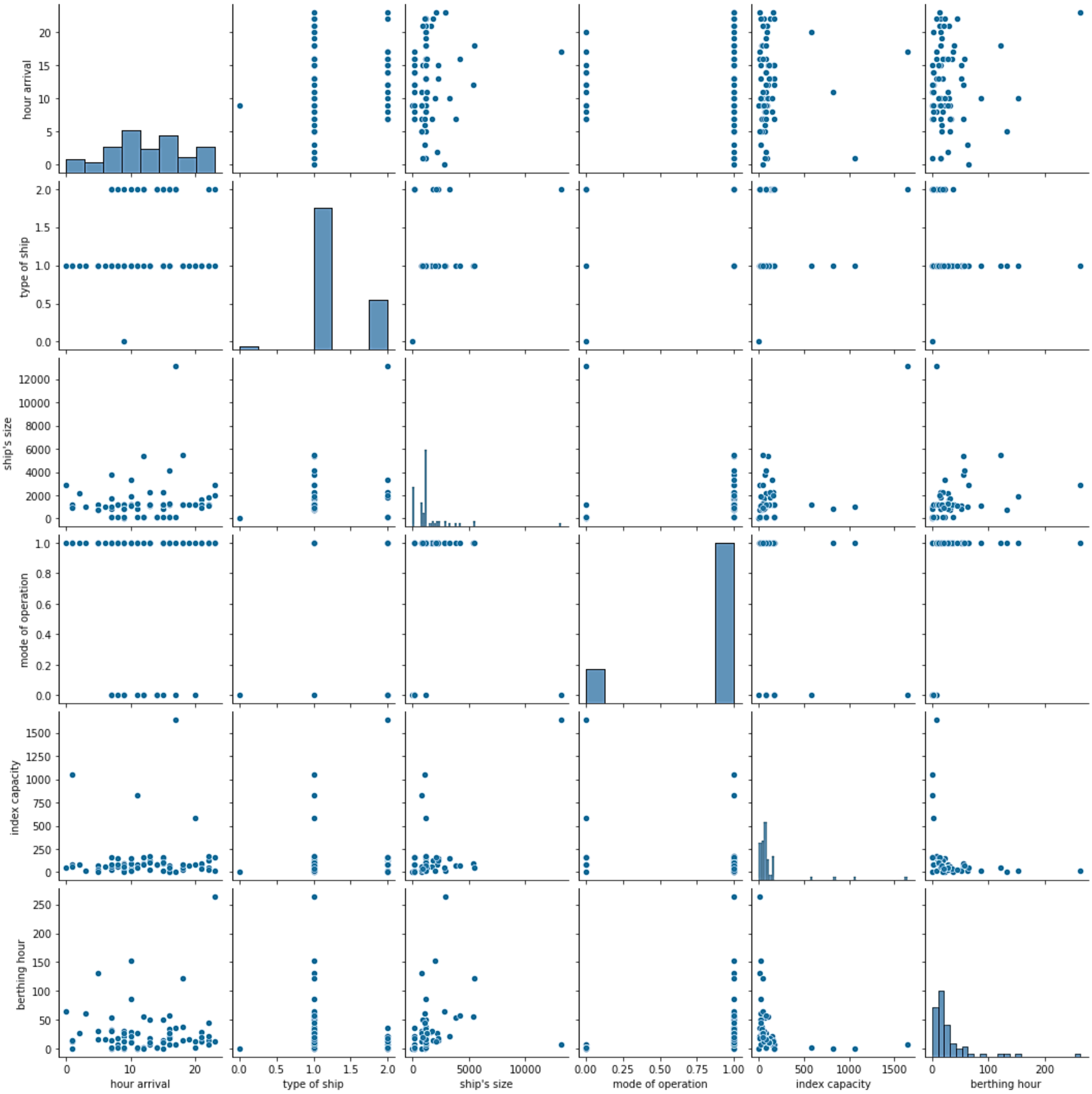


Fig. 7. The pair plot diagram of the dataset.

3.4. Model algorithm

3.4.1. Artificial Neural Network

ANN is among the artificial intelligence (AI) techniques that have received extensive attention and are regarded as one of the most powerful computational tools ever developed [34]. A neural network is a structural system made up of three layers: an input layer, a hidden layer, and an output layer [35]. Meanwhile, deep learning system architectures consist of more than one hidden layer [36]. They have the ability to model and process nonlinear input-output relationships by analyzing historical data [37].

They can also handle large and complex systems with multiple parameters, which makes them suitable for solving the CISF problem. One of the major concerns of these techniques is the issue of over-fitting, which particularly arises from the need for a large number of layers for precise output prediction [32]. There are n input samples denoted as $x = [x_1, x_2, \ldots, x_n]$ which are assigned to the corresponding weights of w_1 to w_n and the biases vector of b. Weight elements of w_i and b are scalar parameters that can be adjusted. The central idea underneath a neural network is that such parameters can be tuned to have the desired results in the output. These inputs are directed through the m hidden layers. Therefore, the net output function is calculated as:

$$y = \sum_{i=1}^{n} x_i w_i + b \tag{1}$$

In this proposed method, ANN networks are designed with 10 number of hidden layers feeding with 9 inputs as illustrated in Fig. 8. As for the algorithm, Levenberg–Marquardt (LM) backpropagation is employed for training the network. Levenberg–Marquardt is the fastest algorithm that takes less computational time to train the model [33].

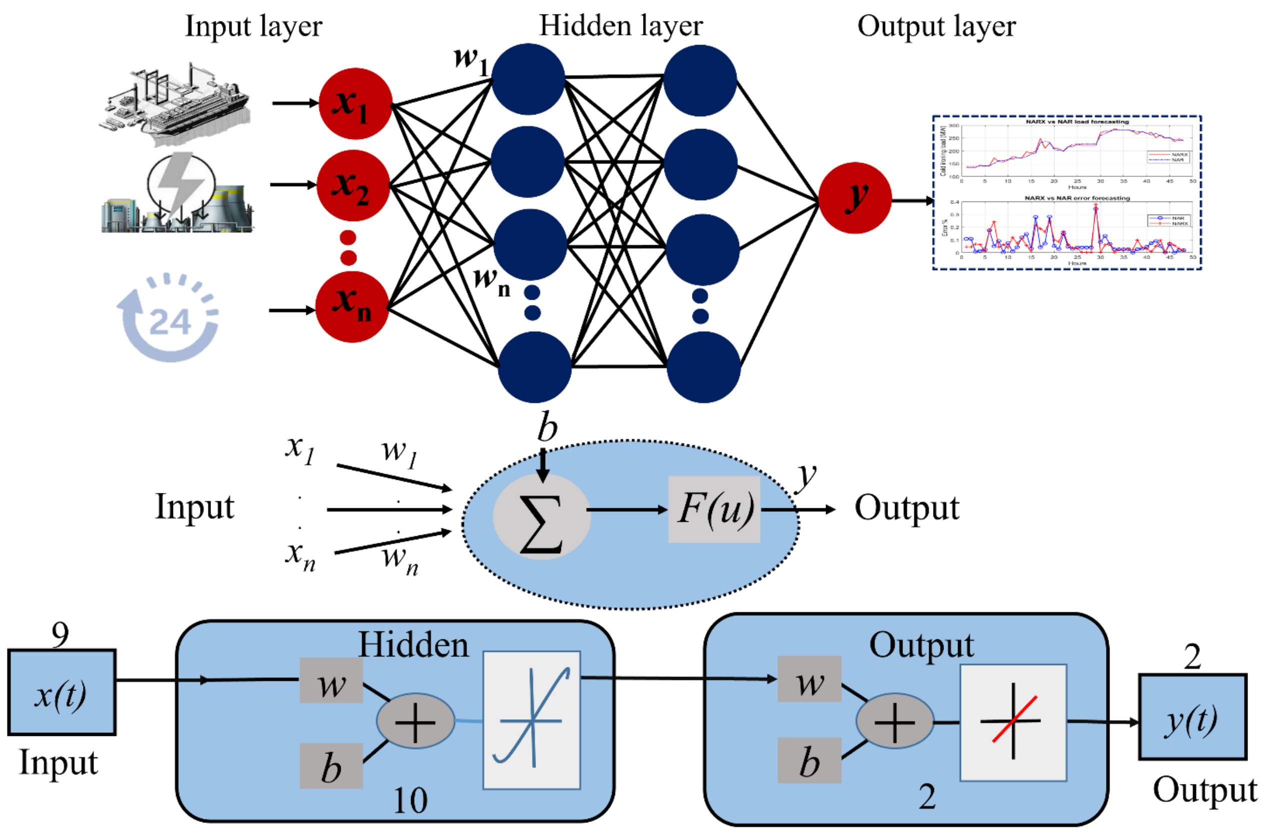


Fig. 8. Proposed ANN structure.

3.4.2. Multiple Linear Regression

Multiple linear regression (MLR) is a statistical technique that uses to investigate the relationship between two or more variables [38]. Equation (2) is used for MLR:

$$y = a_0 + a_1 x_1 + \cdots + a_n x_n \tag{2}$$

where $a_0, a_1, \ldots, a_n$ are coefficients, y is the dependent variable, and $x_1, x_2, \ldots, x_n$ are independent variables. In this method, a_n (coefficients) are calculated as;

$$a_n = argmin_{(a)}^{\left(\sum_{i=1}^{n}\left(y_i - a_0 - \sum_{j=1}^{n} a_j x_{ij}\right)^2\right)} \tag{3}$$

3.4.3. Decision Tree

Decision Tree is a decision-making tool that works with both continuous and categorical variables and uses a flowchart-like tree structure. Mean square error is used to form the dividing sub-node in most cases. The branches represent either the conditions (decision nodes) or the outcome (end nodes).

3.4.4. Random Forest

The random forest builds numerous subgroup decision trees from the dataset. Then, for each subgroup, a new tree is formed, and the process is repeated until the final prediction is made [39]. Each tree's predictions are collected, and the total value is averaged.

3.4.5. Extreme Gradient Boosting

Machine learning algorithms called extreme gradient boosting (XG Boost) can be used to solve regression predictive modeling problems. It is based on the concept of staging the forecast, with the second stage focusing on minimizing the previous stage's inaccuracy. The main goal is to acquire the desired result with the least amount of error possible for the entire dataset. The XG Boost model is derived from equation (4) [40].

$$\hat{y}_i = \sum_{k=1}^{K} f_k(x_i), f_k \in F \tag{4}$$

where k is the number of decision-tree, $f_k(x_i)$ is the function of input in k-th decision-tree, $\hat{y}_i$ is the predicted value.

3.5. Performance indicator

The performance and accuracy of the forecasting model can be measured by using the forecast's key performance indicators (KPI). The most commonly used KPIs are mean absolute error (MAE), mean square error (MSE), Root Mean Square Error (RMSE), and coefficient of determination (R^2) [41], [42]. The formulations of the KPIs are given below:

$$\mathrm{MAE} = \frac{1}{n}\sum_{n=1}^{n}|y - y'| \tag{5}$$

$$\mathrm{MSE} = \frac{1}{n}\sum_{n=1}^{n}(y - y')^2 \tag{6}$$

$$\mathrm{RMSE} = \sqrt{\frac{1}{n}\sum_{n=1}^{n}(y - y')^2} \tag{7}$$

$$R^2 = 1 - \frac{\sum_{n=1}^{n}(y-y')^2}{\sum_{n=1}^{n}(y-k)^2} \tag{8}$$

where n is the number of data samples, y is the actual desired value, y' is the predicted value, and k is the mean of the actual value.

4. Simulation result

The main objective of the proposed training module is to get the target value as close as possible to the actual output. Simulation by using multiple methods attempts to find the algorithm that performs the best on the dataset for the CISF case study. A forecasting model with minimum error represents the high accuracy of the prediction and is capable to provide a trustworthy outcome for the new data. Fig. 9. shows the comparative result of the forecasting value versus the actual value from all proposed methods which are, ANN, MLR, random forest, XG Boost, and decision tree. The red dots on the graph represent the actual berthing duration of the ship, while the other colors represent the prediction output generated from different methods. ANN and decision tree successfully imitate the actual berthing duration from most of the ship's data sample with negligible error and are ready to be used for forecasting. Meanwhile, it can be observed that MLR prediction is far from the actual value in all data samples. In the remaining techniques, a substantial proportion of the output is close to the real value, while some of the targets are distant from the actual value.

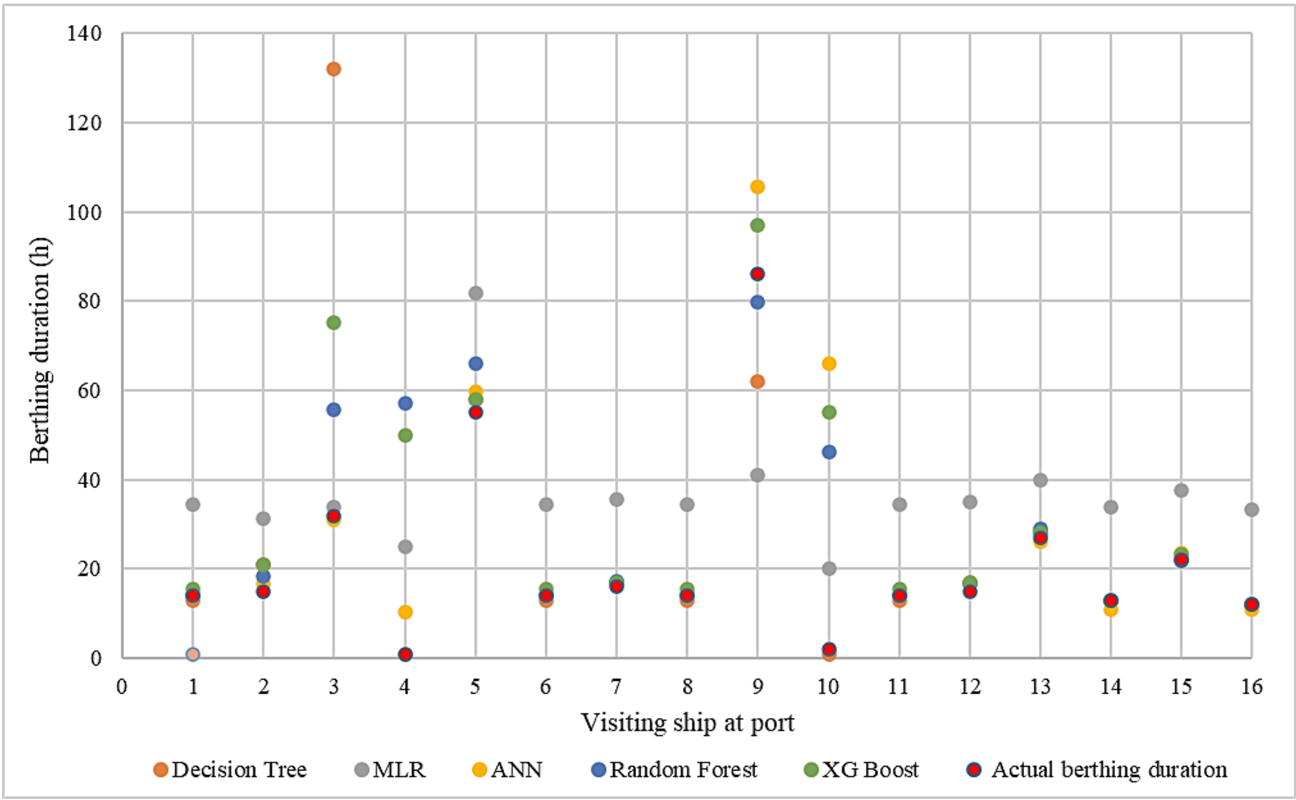


Fig. 9. Comparison CISF of the actual and prediction value for all methods.

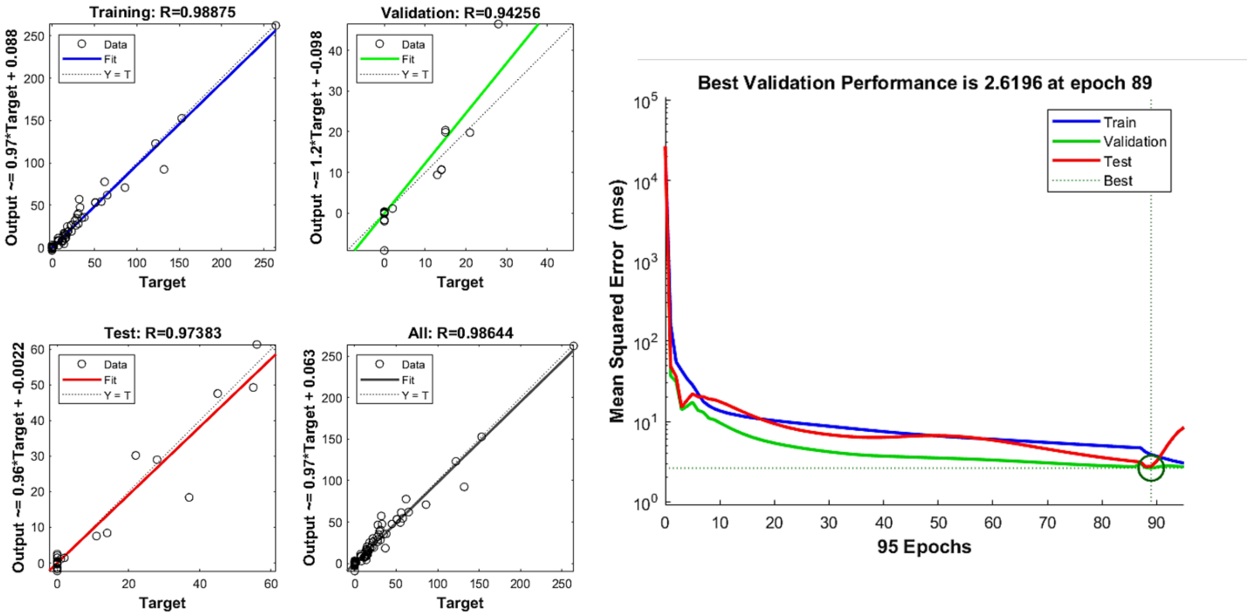
To measure the performance validity, a few validation features are analyzed. Table 3 summarizes the accuracy level for each algorithm by using RMSE, MAE, and R2. Given the average value of these indicators, the ANN model has the lowest error of RMSE and MAE with 3.1343 and 0.2548, respectively. On the other hand, random forest and decision tree also show a good performance with the error measurement slightly higher than ANN. Meanwhile, the MLR technique has worse performance with the highest error validation, 55.43, 2.0825, and 11.51% for RMSE, MAE, and R2. This could be due to its nature to model the relationship between a continuous response variable. However, data input in CISF is more complex with a mix of the continuous, binary number, and categorical variables. In addition, a nonlinear relationship is detected between the input parameters and output parameters as illustrated in Fig. 4., which results in a dynamic pattern of berthing duration for each visiting ship. ANN framework allows the model to learn deep dimensions from the input and has the advantage of having higher accuracy in the prediction that is driven by another external uncertainty and involves disturbance. Because of these strengths, ANN outperforms other algorithms and becomes a good model for CISF in the first place. Nevertheless, these conclusions are not general and in other applications, the other model might show better performance.

Table 3

The error performance of the forecasting approaches.

Algorithm	Type of error		
	RMSE	R^2 (%)	MAE
ANN	3.1343	98.64	0.2548
MLR	55.434	11.51	2.0825
Random Forest	5.5473	91.14	0.3346
XG Boost	9.2918	85.16	0.3661
Decision tree	3.9369	93.71	0.2972

Since ANN is selected for the best approaches in the case of CISF, the final evaluation of the detailed regression performance is necessary. Fig. 10. shows the regression values for training, validation, and testing of the CISF data set. The findings reveal that ANN's overall performance in training, validation, and test is 0.98875, 0.94256, and 0.97383 respectively. Given the average value of 98.644% for overall performance suggest that the model is good for forecasting. As the training run, the error must decrease as much as possible, and training can keep continuing. In the case of ANN, the MSE is decreasing along with epochs, and training stop at epoch 89 with 95 iterations. Forcing the training data to keep running on that zone might cause to overfitting problem.



(a) Overall performance (b) MSE validation performance

Fig. 10. ANN performance for training, test, validation, and overall.

The final training module with the lowest error performance can be used to forecast a new dataset. Fig. 11. shows the forecasting result from two different types of ships, which are cargo and tanker by using the ANN training module. It shows that most of the ships calling frequency at the port successfully imitates the actual value of berthing duration. It suggests that selected input variables are strongly related to the varying pattern of berthing duration for each visiting ship. The highest error performance in the case of cargo CISF is 40% meanwhile the highest error forecasting for tanker CISF is 38%, which might be caused by null variables that hinder the target output. In addition, the forecasting capability of data-driven models essentially depends on the quality of data trained. A large error deviation might occur at the region where the parameter's value changes sharply.

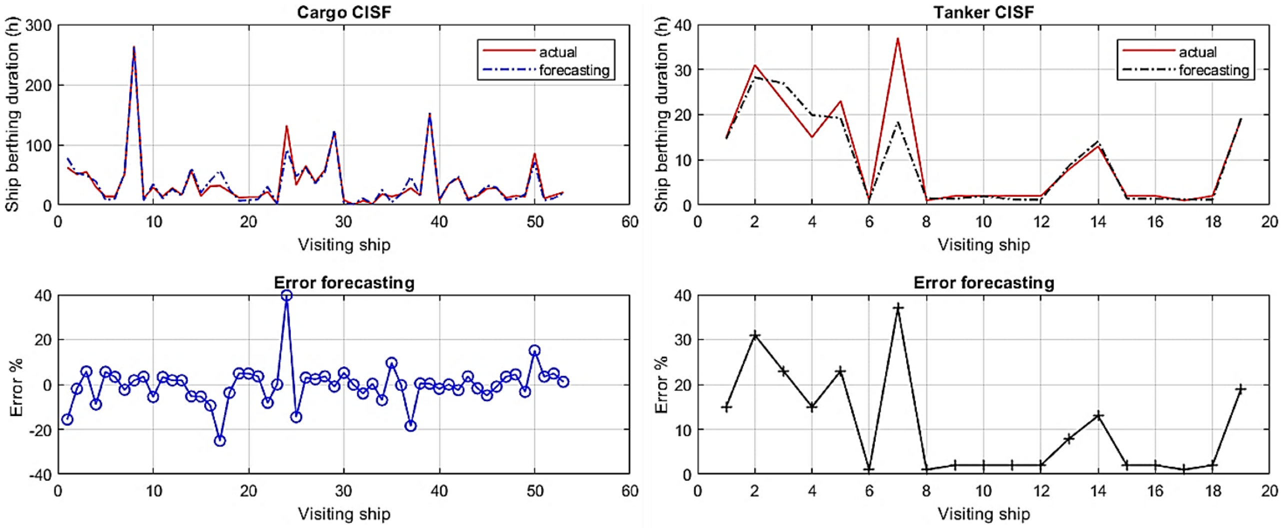


Fig. 11. CISF for different types of ships by using ANN.

5. Discussion

The selection input variable in this case study is based on a different perspective and may influence the berthing period in a different aspect. For instance, different type of ships has different characteristic, function, and requirement to fulfill. One of the great challenges to getting an accurate output is related to the uncertainty variables. The traffic of the ship calling at the port can suddenly change. It is undeniable that the port has the ship's arrival information in advance with the help of an automatic identification system (AIS). However, there is the possibility of the sudden change in arrival tentative due to uncontrol events caused by weather conditions, technical problems, and special control requests for the fast arrival of the ship. This uncontrol event may result in a delay or early arrival to the port compared to the actual planning time arrival. Hour of arrival also varies the berthing period in which arrival at peak time might cause a longer waiting time to the ship in que while arriving late at night will acquire long handling operation due to the lack of manpower. Another aspect in consideration is the varying size of the ship. The bigger size of the ships is capable to carry a larger capacity of goods hence consuming more time for loading and unloading activities. However, some of the datasets suggest that the big size of ship berthing in a short duration. In this sense, it might relate to the ship's mode of operation whereas loading/unloading activities, transit, refueling, and visiting port for maintenance consume different berthing hours depending on the mode of operation. Regardless of the multiple input parameters, most of the calling ships at the port can imitate the actual value of berthing duration with minimal error. One of the novelties in this study is by demonstrating the high correlation between chosen input variables to the varying pattern of berthing duration for each visiting ship.

The above-mentioned input-output relationship with a large and different number of datasets leads to the complexity of the forecasting process. To solve this issue, the integration of advanced forecasting techniques such as a data-driven approach can efficiently process the huge quantity of data, capture the correlation pattern, and draw an accurate conclusion. It is proven in this study whereby, out of five data-driven algorithms, four of them (ANN, random forest, XG Boost, and decision tree) showed good performance with ANN outperforming the others. In comparison to one of the available research studies in [20] that estimates port power demand for cold ironing application by using a non-data-driven method, which is environmental and protection agency (EPA)-based. This EPA method is based on mathematical formulation where several parameters need to be calculated manually. This method is beneficial when detailed measured data are not available, and thus can be calculated with the provided mathematical formula. Despite the promising result shown in the analysis, this approach is not suitable for forecasting problems with a large volume of datasets such as ship berthing forecasting. It will be time-consuming and have a high possibility for human error during calculation. In this case, data-driven is a quicker and more effective solution in clustering, classifying, and interpreting the given input data for high accuracy output. The comparative numerical results in this paper with a different model of data-driven is to find the best forecasting model for ship berthing duration problems. This illustrates the aim of this paper on how to take advantage of using data-driven for real port forecasting problems.

Forecasting output in this study which is ship berthing duration indicating the duration of the ship will be at the port, the amount of CI's power needed, and the amount of emission that can be prevented. Moreover, accurately forecasting the berthing duration has implications for other port control such as EMS and BAP. This can be explained by berthing duration will give information on the expected departure of the ship. The port operator uses the data as an input in the optimization algorithm to optimally allocate the incoming and awaiting ship to suitable berth allocation, namely BAP. Thus, the ship owner can reduce the waiting time at the port and get assigned to a berth terminal with an adequate CI capacity suitable to their ship's power requirement. On the other hand, the duration of berthing also suggests how much power is needed for the berthed ship. This information will help energy authorities at the port to efficiently manage the EMS. The outcome from the CISF is not only beneficial to the port operator but also to the ship owner. In the existing research study such as in [8] and [19], both of them performed forecasted electricity price of the CI at a different port. This forecasting info assists ship owners to select the optimal voyage route to minimize the operation cost and navigation time by scheduling the ship to the nearest port when the CI electricity price is relatively low. However, it is useful for the ship owner only during the voyage mode of operation. None of the existing research performs forecasting techniques at the port terminal where the CI system is allocated. They ignore the need for forecasting when CI is practically consumed which is during the berthing mode of operation. Additionally, the available CI forecasting study is only beneficial for decision-making from the ship owner's perspective and neglects the port operator's viewpoint. Thus, this CISF fulfills this research gap by providing forecasting output of CI during berthing mode of operation and is useful for both parties, ship owner and port operator.

6. Conclusion

CISF is one of the important frameworks for port operators, especially in management affairs. In this paper, a data-driven approach for CISF has been proposed and validated by simulation. In the first stage, statistical analysis is examined and performing cleaning process that detects the null values and outliers. Since the berthing duration of the ship is associated with a lot of factors, correlation analysis between potential inputs helps to validate the true correlation of the chosen input for CISF. The refined data is then fed to the five different models namely, ANN, MLR, random forest, XG Boost, and decision tree. The results of simulations reveal that ANN, random forest, XG Boost, and decision three show a good performance with RMSE 3.1343, 5.5473, 9.2918, and 3.9369 respectively. This suggests that ANN has the highest accuracy and lowest error performance of all forecasting models, thus becoming the best forecasting model for CIFS. The finding from the numerical result not only proves that the data-driven approach is applicable for the case of CISF, but also demonstrates the strong relationship between selected input variables toward forecasting output. As a limitation of the proposed method, it requires substantial amounts of data during the training process to ensure they are trained with high accuracy and imitate the real value as close as possible. However, obtaining a large quantity of data may not always be practical. To mitigate this problem, new approaches trainable with limited data should be developed.

CRediT authorship contribution statement

N.N.A.B.; Conceptualization, Methodology, Software, Data curation, Writing - original draft, Visualization, Investigation, Validation. N.B.; Conceptualization, Validation, Supervision, Writing – review & editing. H.C.; Conceptualization, Validation. T.U.; Conceptualization, Software. J.M.G.; Supervision, Project administration. J.C.V.; Supervision, Project administration.

Acknowledgment

This research work was supported by a Villum Investigator grant (no. 25920) from the Villum Fonden; University Malaysia Perlis (UniMAP); and the Ministry of Education Malaysia.

Reference

[1] J. Yuan and V. Nian, "Ship energy consumption prediction with Gaussian process metamodel," *Energy Procedia*, vol. 152, pp. 655–660, 2018, doi: 10.1016/j.egypro.2018.09.226.

[2] Y. A. Bakar, N.N.A., Guerrero, J.M., C. Vasquez, J., Bazmohammadi, N., Othman, M., Rasmussen, B.D. and Al-Turki, "Optimal Configuration and Sizing of Seaport Microgrids including Renewable Energy and Cold Ironing – The Port of Aalborg Case Study," *Energies*, vol. 15(2), p. 431, 2021.

[3] Y. A. Bakar, N.N.A., Guerrero, J.M., Vasquez, J.C., Bazmohammadi, N., Yu, Y., Abusorrah, A. and Al-Turki, "A Review of the Conceptualization and Operational Management of Seaport Microgrids on the Shore and Seaside," *Energies*, vol. 14, no. 7941, pp. 1–31, 2021, doi: https://doi.org/10.3390/en14237941.

[4] J. G. Saenz, "Energy analysis and costs estimation of an On-shore Power Supply system in the Port of Gävle," University of Gavle, 2019.

[5] D. Colarossi and P. Principi, "Technical analysis and economic evaluation of a complex shore-to-ship power supply system," *Appl. Therm. Eng.*, vol. 181, p. 115988, 2020, doi: 10.1016/j.applthermaleng.2020.115988.

[6] R. Ssali, "Ship-port interface : analysis of the cost- effectiveness of cold ironing at Mombasa Port," World Maritime University, 2018.

[7] M. U. Mutarraf *et al.*, "A Decentralized Control Scheme for Adaptive Power-Sharing in Ships based Seaport Microgrid.," *IECON Proc. (Industrial Electron. Conf.*, vol. 2020-Octob, pp. 3126–3131, 2020, doi: 10.1109/IECON43393.2020.9255344.

[8] M. Z. Tianyang Zhao, Jiayu Qiu, Shuli Wen, "Efficient Onboard Energy Storage System Sizing for Navigation Routing Under Onshore Uncertainties," *IEEE Trans. Ind. Appl.*, vol. 58, no. 2, pp. 1664–1674, 2022.

[9] Y. Zhang, C. Liang, J. Shi, G. Lim, and Y. Wu, "Optimal Port Microgrid Scheduling Incorporating Onshore Power Supply and Berth Allocation Under Uncertainty," *Appl. Energy*, vol. 313, p. 118856, 2022, doi: 10.1016/j.apenergy.2022.118856.

[10] H. Z. Ellefsen, André Listou, Peihua Han, Xu Cheng, Finn Tore Holmeset, Vilmar Æsøy,

"Online Fault Detection in Autonomous Ferries : Using Fault-Type Independent Spectral Anomaly Detection," *IEEE Trans. Instrum. Meas.*, vol. 69, no. 10, pp. 8216–8225, 2020, doi: 10.1109/TIM.2020.2994012.

[11] C. Velasco-gallego and I. Lazakis, "A real-time data-driven framework for the identification of steady states of marine machinery," *Appl. Ocean Res.*, vol. 121, no. October 2021, p. 103052, 2022, doi: 10.1016/j.apor.2022.103052.

[12] S. Sultana, S. Akter, and E. Kyriazis, "Technological Forecasting & Social Change How data-driven innovation capability is shaping the future of market agility and competitive performance," *Technol. Forecast. Soc. Chang.*, vol. 174, p. 121260, 2022, doi: 10.1016/j.techfore.2021.121260.

[13] T. Uyan and Y. Arslano, "Machine learning approach to ship fuel consumption : A case of container vessel," *Transp. Res. Part D*, vol. 84, p. 102389, 2020, doi: 10.1016/j.trd.2020.102389.

[14] Y. Huang, H. Lan, Y. Y. Hong, S. Wen, and S. Fang, "Joint voyage scheduling and economic dispatch for all-electric ships with virtual energy storage systems," *Energy*, vol. 190, p. 116268, 2020, doi: 10.1016/j.energy.2019.116268.

[15] K. Hein, Y. Xu, G. Wilson, A. K. Gupta, and G. W. and A. K. G. K. Hein, Y. Xu, "Coordinated Optimal Voyage Planning and Energy Management of All-Electric Ship With Hybrid Energy Storage System," *IEEE Trans. Power Syst.*, vol. 36, no. 3, pp. 2355–2365, 2021, doi: 10.1109/TPWRS.2020.3029331.

[16] S. Wen *et al.*, "Joint Energy Management and Voyage Scheduling for All-Electric Ships Using Dynamic Real-Time Electricity Price of Onshore Power," *Conf. Rec. - Ind. Commer. Power Syst. Tech. Conf.*, vol. 2020-June, 2020, doi: 10.1109/ICPS48389.2020.9176793.

[17] X. Sun, J. Qiu, Y. Tao, Y. Yi, and J. Zhao, "Distributed Optimal Voltage Control and Berth Allocation of All-Electric Ships in Seaport Microgrids," *IEEE Trans. Smart Grid*, vol. 3053, no. c, pp. 1–1, 2022, doi: 10.1109/tsg.2022.3161647.

[18] F. Conte, F. D'Agostino, D. Kaza, S. Massucco, G. Natrella, and F. Silvestro, "Optimal Management of a Smart Port with Shore-Connection and Hydrogen Supplying by Stochastic Model Predictive Control," 2022, [Online]. Available: http://arxiv.org/abs/2204.00453.

[19] S. Wen, T. Zhao, Y. Tang, and S. Member, "Coordinated Optimal Energy Management and Voyage Scheduling for All-Electric Ships Based on Predicted Shore-Side Electricity Price," *IEEE Trans. Ind. Appl.*, vol. 57(1), pp. 139–148, 2021.

[20] F. D'Agostino, G. P. Schiapparelli, S. Dallas, D. Spathis, V. Georgiou, and J. Prousalidis, "On Estimating the Port Power Demands for Cold Ironing Applications," *2021 IEEE Electr. Sh. Technol. Symp. ESTS 2021*, pp. 1–5, 2021, doi: 10.1109/ESTS49166.2021.9512359.

[21] and P. D. Alikhani, Parnian, Lina Bertling Tjernberg, Linda Astner, "Forecasting the Electrical Demand at the Port of Gävle Container Terminal," *IEEE PES Innov. Smart Grid Technol. Eur. (ISGT Eur.*, pp. 1–6, 2021.

[22] Y. Peng, H. Liu, X. Li, J. Huang, and W. Wang, "Machine learning method for energy consumption prediction of ships in port considering green ports," *J. Clean. Prod.*, vol. 264, p. 121564, 2020, doi: 10.1016/j.jclepro.2020.121564.

[23] K. Hein, Y. Xu, W. Gary, and A. K. Gupta, "Robustly coordinated operational scheduling of a grid-connected seaport microgrid under uncertainties," *IET Gener. Transm. Distrib.*, vol. 15, no. 2, pp. 347–358, 2021, doi: 10.1049/gtd2.12025.

[24] A. Rolan *et al.*, "Integration of Cold Ironing and Renewable Sources in the Barcelona Smart Port," *IEEE Trans. Ind. Appl.*, vol. 55, no. 6, pp. 7198–7206, 2019, doi: 10.1109/TIA.2019.2910781.

[25] X. Li *et al.*, "A method for optimizing installation capacity and operation strategy of a hybrid renewable energy system with offshore wind energy for a green container terminal," *Ocean Eng.*, vol. 186, no. May, p. 106125, 2019, doi: 10.1016/j.oceaneng.2019.106125.

[26] Z. Wan *et al.*, "Evaluation of emission reduction strategies for berthing containerships : A case study of the Shekou Container Terminal," *J. Clean. Prod.*, vol. 299, p. 126820, 2021, doi: 10.1016/j.jclepro.2021.126820.

[27] W. W. Yun Peng , Meng Dong , Xiangda Li , Huakun Liu, "Cooperative optimization of shore power allocation and berth allocation : A balance between cost and environmental benefit," *J. Clean. Prod.*, vol. 279, p. 123816, 2021, doi: 10.1016/j.jclepro.2020.123816.

[28] Ç. Iris and J. S. L. Lam, "Optimal energy management and operations planning in seaports with smart grid while harnessing renewable energy under uncertainty," *Omega (United Kingdom)*, vol. 103, p. 102445, 2021, doi: 10.1016/j.omega.2021.102445.

[29] N. Cheimanoff, F. Fontane, M. N. Kitri, and N. Tchernev, "Exact and heuristic methods for the integrated berth allocation and specific time-invariant quay crane assignment problems," *Comput. Oper. Res.*, vol. 141, no. January, 2022, doi: 10.1016/j.cor.2022.105695.

[30] "Port of Aalborg," 2022. https://www.myshiptracking.com/ports/port-of-aalborg-in-dk-denmark-id-80 (accessed Apr. 02, 2022).

[31] J. Prousalidis, D. Lyridis, and S. Dallas, “Ship to Shore Electric Interconnection : from adolescence to maturity,” in *IEEE Electric Ship Technologies Symposium (ESTS)*, 2017, no. October 2018, pp. 200–206, doi: 10.1109/ESTS.2017.8069281.

[32] A. Al Mamun, M. Sohel, N. Mohammad, M. S. Haque Sunny, D. R. Dipta, and E. Hossain, “A Comprehensive Review of the Load Forecasting Techniques Using Single and Hybrid Predictive Models,” *IEEE Access*, vol. 8, pp. 134911–134939, 2020, doi: 10.1109/ACCESS.2020.3010702.

[33] N. Ammar, M. Sulaiman, and A. F. M. Nor, “Analysis load forecasting of power system using fuzzy logic and artificial neural network,” *J. Telecommun. Electron. Comput. Eng.*, vol. 9, no. 3, pp. 181–192, 2017.

[34] C.-Y. C. Hsu, Che-chiang, “Regional load forecasting in Taiwan — applications of artificial neural networks,” *Energy Convers. Manag.*, vol. 44, pp. 1941–1949, 2003, doi: 10.1016/S0196-8904(02)00225-X.

[35] A. Khalid, S. Aslam, K. Aurangzeb, S. I. Haider, M. Ashraf, and N. Javaid, “An efficient energy management approach using fog-as-a-service for sharing economy in a smart grid,” *Energies*, vol. 11, no. 12, pp. 1–17, 2018, doi: 10.3390/en11123500.

[36] S. Hu and Y. Zuo, “A Review about Building Hidden Layer Methods of Deep Learning,” *J. Adv. Inf. Technol.*, vol. 7, no. 1, pp. 13–22, 2016, doi: 10.12720/jait.7.1.13-22.

[37] M. Q. Raza and A. Khosravi, “A review on artificial intelligence based load demand forecasting techniques for smart grid and buildings,” *Renew. Sustain. Energy Rev.*, vol. 50, no. June 2015, pp. 1352–1372, 2015, doi: 10.1016/j.rser.2015.04.065.

[38] O. B. Öztürk and E. Başar, “Multiple linear regression analysis and artificial neural networks based decision support system for energy efficiency in shipping,” *Ocean Eng.*, vol. 243, no. November 2021, 2022, doi: 10.1016/j.oceaneng.2021.110209.

[39] T. Uyanık, Ç. Karatuğ, and Y. Arslanoğlu, “Machine learning based visibility estimation to ensure safer navigation in strait of Istanbul,” *Appl. Ocean Res.*, vol. 112, no. May, 2021, doi: 10.1016/j.apor.2021.102693.

[40] R. Guo, Z. Zhao, T. Wang, G. Liu, J. Zhao, and D. Gao, “Degradation state recognition of piston pump based on ICEEMDAN and XGBoost,” *Appl. Sci.*, vol. 10, no. 18, pp. 1–17, 2020, doi: 10.3390/APP10186593.

[41] A. Naz, M. U. Javed, N. Javaid, T. Saba, M. Alhussein, and K. Aurangzeb, “Short-term electric load and price forecasting using enhanced extreme learning machine optimization in smart

grids," *Energies*, vol. 12, no. 5, 2019, doi: 10.3390/en12050866.

[42] F. M. Butt, L. Hussain, A. Mahmood, and K. J. Lone, "Artificial Intelligence based accurately load forecasting system to forecast short and medium-term load demands," *Math. Biosci. Eng.*, vol. 18, no. 1, pp. 400–425, 2021, doi: 10.3934/MBE.2021022.