Linh Bui-Duy; Ngoc Vu-Thi-Minh

Article

Utilization of a deep learning-based fuel consumption model in choosing a liner shipping route for container ships in Asia

Asian Journal of Shipping and Logistics (AJSL)

Provided in Cooperation with:
Korean Association of Shipping and Logistics, Seoul

Suggested Citation: Linh Bui-Duy; Ngoc Vu-Thi-Minh (2021) : Utilization of a deep learning-based fuel consumption model in choosing a liner shipping route for container ships in Asia, Asian Journal of Shipping and Logistics (AJSL), ISSN 2352-4871, Elsevier, Amsterdam, Vol. 37, Iss. 1, pp. 1-11, https://doi.org/10.1016/j.ajsl.2020.04.003

This Version is available at:
https://hdl.handle.net/10419/329657

The Asian Journal of Shipping and Logistics 37 2021 1–11

HOSTED BY

Contents lists available at ScienceDirect

The Asian Journal of Shipping and Logistics

journal homepage: www.elsevier.com/locate/ajsl

Original Article

Utilization of a deep learning-based fuel consumption model in choosing a liner shipping route for container ships in Asia

Linh Bui-Duy, Ngoc Vu-Thi-Minh*

School of Economics and International Business, Foreign Trade University, Hanoi, Viet Nam

ARTICLE INFO

Article history:
Received 3 February 2020
Received in revised form 7 April 2020
Accepted 17 April 2020

Keywords:
Asymmetric traveling salesman problem
Ship fuel consumption model
Liner shipping
Route choice optimization
Deep-learning neural network
Container ships

ABSTRACT

Designating the ideal shipping route can spare expenses, enlarge profits and improve the competitiveness of shipping companies. Liner shipping route choice is mainly contingent on fuel cost, which always contributes the major proportion of the ship's operating cost. Although many studies on this topic have been carried out, none are based on the fuel consumption forecast model designed by the advanced machine learning method. This paper provides a platform idea for selecting the optimal operating route for container ships to minimize fuel cost by using an asymmetric traveling salesman problem (ATSP) algorithm solution, in which the fuel consumption model for the route is estimated based on the deep-machine learning method. Five input variables are given in the model including average velocity, sailing time, ship's capacity, wind speed, and wind direction. The mean absolute percentage error (MAPE) of the model is 5.89%, indicating that the predictive result obtains a very high accuracy, close to 95%. The optimal model is thus applied in combination with ATSP to address the optimal solution for a certain route.

1. Introduction

Ship operating costs consist of several components, such as crew costs, repair and maintenance costs, marine insurance and protection and indemnity (P&I), stores, spares, and fuel cost, where, fuel cost forms more than 60% (Golias et al., 2009). In addition, rising and volatile fuel prices constitute a serious concern for shipping companies (Bal Beşikçi et al., 2016), and selecting the optimum route to minimize fuel costs is an effective fundamental solution that shipping companies can utilize in their decisions when operating liner ships. When fuel cost prediction is handled by advanced methods based on learning actual data, the predictive error can be diminished, the proposed model can be used for a larger number of observations, and it can reflect the situation in the real world.

Several studies have made progress in optimizing the shipping route choice; however, extant studies have not yet incorporated the hybrid machine-learning method with the asymmetric traveling salesman problem (ATSP) algorithm, one of the fastest and most effective ways for optimization (Kim, Kim, & Shim, 2002). This paper aims to fill this research gap by applying a two-phase method. An estimating model of fuel consumption is constructed based on a deep-learning artificial neural network (ANN), and the ATSP solution for a liner shipping route is then handled out to minimize the fuel cost.

The rest of this paper is organized as follows. Relevant studies that relate to predicting fuel consumption and optimization by different methodologies are reviewed in Section 2. Section 3 presents the exploratory analysis of the dataset by random forest, one of the most popular and powerful decision tree-based ensemble techniques. The explored variables in Section 3 are then inserted into the input layer in the deep-learning ANN architecture to estimate fuel consumption in Section 4. Section 5 utilizes the predictive results from ANN as a platform of the ATSP solution for optimizing the route. Section 6 provides some implications of our work and the conclusion is carried out in Section 7.

2. Related works

The several studies that have researched fuel consumption prediction can be classified into two approaches: the classical statistical and automatic computing method. Except for the emission factor from the emission function, extant studies estimated fuel consumption utilizing the operational mode of each ship (sailing,

* Corresponding author.
E-mail addresses: duylinh@ftu.edu.vn (L. Bui-Duy), vuminhngoc@ftu.edu.vn, vuthiminhngocftu@gmail.com (N. Vu-Thi-Minh).

Peer review under responsibility of the Korean Association of Shipping and Logistics, Inc.

https://doi.org/10.1016/j.ajsl.2020.04.003

maneuvering, in-port). The utilized information consists of the load factor for the main engines and auxiliary engines as described in the expression (1) (Alver, Saraç, & Alver Şahin, 2018; Dragović et al., 2018; Maragkogianni & Papaefthimiou, 2015):

$$F_{j,k,l} = \sum_{1}^{n} P_j \times \mathrm{LF}_{j,l} \times T_{j,l} \quad (1)$$

where j, k, and l represent the engine type (either main engine, auxiliary engine or boiler), the fuel type (either residual oil (RO), marine distillate oil (MDO), marine gas oil (MGO) or general diesel oil (GDO)), and the operating mode (either normal cruising, slow-steaming, maneuvering, anchoring, or berthing), respectively. F is the estimated fuel consumption (Mt); P is the rated power (kW); LF is the load factor; and T is the operating time (h). Some studies performed a calculation of in-port ship emissions based on the activity mode as follows (Chang, Roh, & Park, 2014; Corbett, Wang, & Winebrake, 2009):

$$F_{ijk} = \left[\mathrm{MF}_k \left(\frac{S_{1k}}{S_{0k}} \right)^3 + \mathrm{AF}_k \right] \frac{D_{ij}}{24 S_{1k}} \quad (2)$$

where i and j are the departure and destination port; k represents an individual vessel; F_{ijk} denotes the fuel consumption of the vessel k traveling from port i to port j in metric tons (Mt); MF_k and AF_k denote the daily fuel consumption rate of the main engine and the auxiliary engine (Mt/day); S_{1k} and S_{0k} represent the operational speed (Knots) and the design speed (Knots) of vessel k, respectively; and D_{ij} is the distance from port i to port j (NM). None of the above studies considered a hybrid computing programming to estimate fuel, which is more accurate.

Recently, some studies have conducted an ANN for predicting engine performance (Ilangkumaran, Sakthivel, & Nagarajan, 2016; Niu et al., 2017; Parkes, Sobey, & Hudson, 2018; Zheng et al., 2019). However, they mostly focused on experimental data or recorded from a small number of ships (maximum of three ships). Furthermore, they only considered the accuracy of the proposed models in comparison with other classical statistical methods. None of them applied the results from predictive models to support decision-making in regards to shipping route choice.

Many studies have been conducted in the field of shipping route choice. However, Lin and Chang (2018) reviewed the studies on ship routing and scheduling problems and showed that liner shipping has not been widely investigated. Chen, Yip, and Mou (2018) investigated the impacts of emission control areas (ECAs) on aspects of global shipping by a discrete choice model (DCM). Kashiha, Thill, and Depken (2016) investigated the influence of geography and transportation costs on the decisions by shippers as to which port to export to by using a conditional logit model. An examination of incorporating the inventory costs of containerized cargoes into existing liner service planning models was carried out by Wang (2017). However, none of these above studies employed a traveling salesman problem (TSP) combined with a hybrid artificial neural network.

TSP is a problem in combinatorial optimization that has been studied in operations research, theoretical computer science, and engineering (Hui, 2012), and variants like the maximum benefit TSP or the price collecting TSP may have numerous economic and transportation applications (Rasmussen, 2011). Zhao, Hu, and Lin (2016) assessed the potential of the Northern Sea Route (NSR) based on designing a multi-port multi-trip liner service. The profit of the round-trip network was maximized by the TSP solution. Kim et al. (2002) managed to minimize the total transportation distance for product delivery over the planning horizon while satisfying the demands of the retailer. Nevertheless, relevant studies that employed TSP for shipping route choice have not taken advantage of the combination with machine-learning techniques.

Utilizing the strength of deep-learning ANN, which can approximate a large class of functions with a high degree of accuracy (Khashei & Bijari, 2010), incorporated with one of the fastest and most effective ways to find the best route (Kim et al., 2002), we employ ANN for prediction fuel of consumption and figure out the ATSP algorithm for liner shipping route choice.

3. Materials and methods

3.1. Overall framework

To solve the indicated research problem, in step 1, we conduct two-stage data collection and implement data filtering to make the data strong enough for later modeling. We also depend on the after-cleaning dataset to identify the potential inputs for the construction of the machine-learning platform. In step 2, a Bootstrapping aggregate non-parametric technique in a decision-tree is designed for the exploratory target of the inputs' importance. The deep-learning ANN is thus executed in step 3 by selecting the best performing hyper-parameters based on the principle of trial-and-error, and the trade-off between accuracy and executing time. In the final step, we tackle the major research problem by utilizing the estimated results of ANN models to construct the asymmetric matrix of fuel consumption for a certain number of ports in Asia. The optimum shipping route connecting all those ports is thus defined by the ATSP solution. The research framework can be seen in Fig. 1

3.2. Exploratory data analysis

3.2.1. Two-phase data collection and filtering

Firstly, we received the historical voyage data from a Vietnamese branch of a worldwide leading shipping company for 2 years from February 2017 to January 2019. The information in the dataset includes the ship name; position (longitude and latitude); departure and arrival port; departure and arrival time; distance (Nm), recorded operating velocity (knots), sailing time (hour), and fuel consumption (Mt) between two adjacent positions. The additional information, including the ship's capacity (TEU), average wind speed (knots converted from Beaufort scale), and wind direction, were conducted through the secondary data collection phase. The ship's capacity was obtained from Vesseltracking (2019). Wind-related weather condition details were gathered from a data package from the Copernicus Marine Environment monitoring service. All the positions in the dataset are located in Asia, which is the operating area of the fleet. As the wind direction was reported in cardinal directions, we converted it to azimuth degree to make the numeric dataset for further analysis.

The original data was filtered to remove all insufficient information rows and noise. We applied a Z-score (Grubbs, 1969) for treating univariate outliers, and Mahalanobis distance for multivariate outliers (Mahalanobis, 1930) as denoising techniques. From 10,853 data rows of the original dataset, 105 missing rows and 287 outliers were removed, and the remaining 10,461 data rows were further analyzed for modeling.

3.2.2. Variable importance by the rain forest algorithm

Before performing the prediction of fuel consumption by the deep-learning technique, the dataset needs to be classified to rank the importance of the input variables, thereby conducting a predicting algorithm to get a highly accurate result. There is a plethora of classification algorithms, in which tree-based learning algorithms are selected to be one of the most-used supervised learning methods. Unlike other machine-learning algorithms based on statistical techniques, a decision tree is a non-parametric model relying on

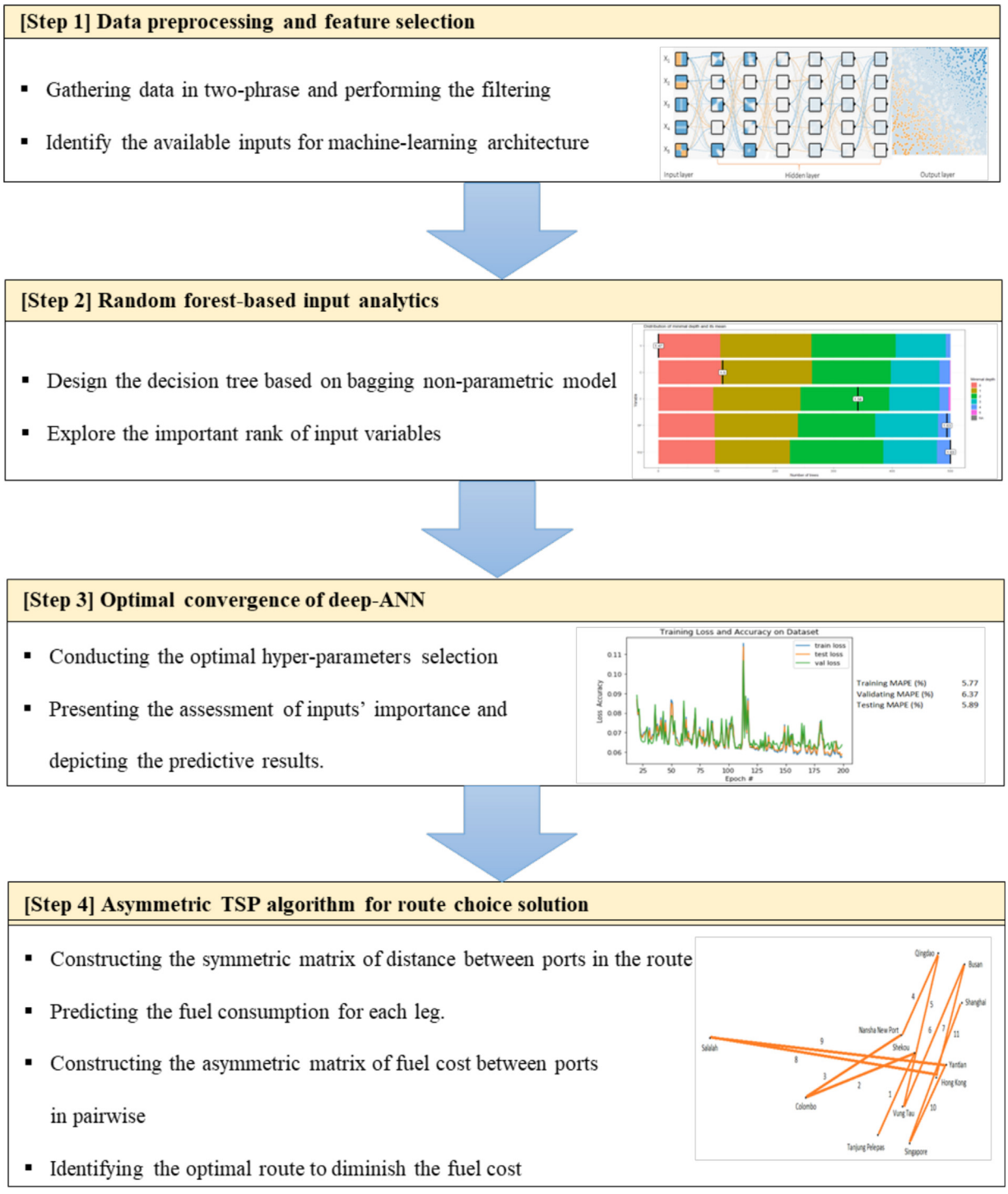


Fig. 1. Research framework.

continuous data and does not obligate basic assumptions for the model. Among decision tree algorithms, random forest (or decision tree forests), developed by Breiman (2001), is considered as one of the most common ensemble models. These models can obtain a higher accuracy in performance than several decision tree techniques. By deploying the Bootstrap aggregation and casually splitting at each input node of the decision trees, the algorithm can be implemented with randomness. The suitable multi-measures for classification forests to consider the importance of variables were selected by R programming by Paluszynska, Biecek, and Jiang (2019), including the mean minimal depth, the number of nodes, and the node purity.

For the available input variables, to estimate fuel consumption, we considered five available inputs in our dataset including sailing time (T), average operating velocity (V), wind speed (BF), wind direction (Wd), and ship's capacity (C). The output variable (F) is fuel

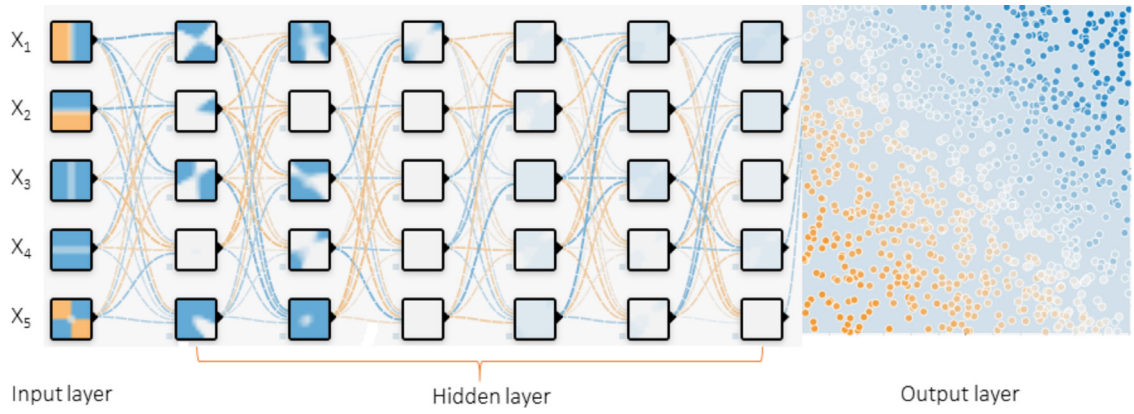


Fig. 2. A visualized example of ANN.

consumption. The ranked result of five available inputs for ANN by using the multi-measures random forest technique is provided in Section 4.2.

3.3. Deep learning ANN architecture

Theoretically, a neural network is an interconnected assembly of simple processing elements, units or nodes, whose functionality is loosely based on the animal neuron. The connection powers between nodes, or weights, which express the processing capability of the network are addressed and stored by a learning process of adaptation to a set of past training patterns. The usage scope of ANN models is often for data modeling, whereby they can be a preferred choice compared to traditional standard nonlinear regression, classifying, or clustering techniques. Thus, they are typically used in problems that may be couched in terms of forecasting (Kevin Gurney, 2005). A neural network includes a set of layers, with each layer including a number of neurons connected to each other as a network. Generally, there is one input layer, one output layer, and a diverse number of hidden layers. The number of nodes in the input and output layers is equivalent to the number of input and output variables, respectively. Starting from the input layer, the nodes in each layer are connected to nodes in the next layer by different weights that the ANN programming sets up each training time. The process is continuous and repetitive until the output has the lowest loss compared to the observation. Fig. 2 shows the instance diagram for an artificial neural network.

Several types of ANN exist that are suitable for numerical data mining, such as the single hidden layer feedforward ANN, shallow or deep multilayer perceptron (MLP) ANN. In particular, the MLP both shallow and deep usually give more accurate results than single hidden layer feedforward ANN because data is passed through more hidden layers than those in single hidden layer ANN. We set up testing for both ANN types based on the trial-and-error principle and accuracy–convergence time trade-off, and deep MLP-ANN with ten hidden layers was selected.

Based on the classification results conducted by the random forest technique in Section 3.2, five variables were identified as input nodes of the input layer including average velocity, capacity, time, wind speed, and wind direction. The output layer has 1 node, which is estimated fuel consumption. Due to the difference in scale of these inputs, we normalized the variables to make them comparable to each other and create a unique scale. We normalized all inputs and outputs between 0 and 1 by using min-max feature scaling according to Equation (3).

$$X_n = \frac{X - X_{\min}}{X_{\max} - X_{\min}} \tag{3}$$

where X_n is the normalized value of a variable X (actual value in a variable), $X_{\max}$ and $X_{\min}$ are the maximum and minimum values of X, respectively.

In this paper, the deep ANN architecture consists of ten hidden layers. A bias, x_0, which is an additional parameter in the ANN, is used to adjust the output along with the weighted sum of the inputs to the neuron. For a neuron with N inputs, $\{w_i\}$ is the weights on the connections, $\{x_i\}$ is the input to the neuron, where $i \in I = \{0,\ldots, N-1\}$, and y is the corresponding output. When data is obtained, a neuron receives input values from the neurons it is connected to and the activation of neuron a is computed (Parkes et al., 2018):

$$a = \sum w_i x_i \tag{4}$$

Considering the activation function of the output of the established neuron is $y = f(a)$, where f is the activation function. There are three popular activation functions: sigmoid, tanh, and the rectified linear unit (ReLU). These functions can be defined as follows (Liew, Khalil-Hani, & Bakhteri, 2016):

$$\text{Sigmoid}: f(x) = \frac{1}{1+e^{-x}} \tag{5}$$

$$\text{Tank}: f(x) = \tanh(x) = \frac{\sinh(x)}{\cosh(x)} = \frac{1-e^{-2x}}{1+e^{-2x}} \tag{6}$$

$$\text{ReLU}: f(x) = \max(0, x) \tag{7}$$

For sigmoid and tanh activation functions, when the absolute of the input value is large, the gradient of these functions will be very close to 0. Small gradients can worsen the overall training performance, as most learning algorithms rely on these gradients to perform optimization and parameter tuning. Alternatively, the convergence of training when using the ReLU is much faster, and faster learning has a significant influence on the performance of large models trained on large datasets. This acceleration results from the ReLU being calculated almost instantaneously; its gradient is also computed very quickly with a gradient equal to 1 if the input is greater than 0, and equal to 0 if the input is less than 0 (Krizhevsky, Sutskever, & Hinton, 2012). This paper employs ReLU as the activation function due to this strength.

In this paper, the dataset was randomly split into 80% for the training set and 20% for the testing set, in which the training set is continuously split into training and validation set with the ratio of approximately 7:1. For the training set, to prevent the overfitting situation in machine learning, we applied one of the most simple but powerful methods, namely the dropout technique (Nitish et al., 2014). This technique prevents overfitting and provides a way of approximately exponentially combining many different neural network architectures efficiently. We set up a dropout for each hidden

Table 1
Trial-and-error procedures for tuning hyper-parameters.

Step	Hidden layers	Hidden nodes	Epoch	CLR	Running time (seconds)	Training MAPE	Validating MAPE	Testing MAPE	Training MASE	Validating MASE	Testing MASE
	1				23.56	11.4	12.1	11.5	3.14	3.22	10
	2				39.16	11	11.8	11.2	3.01	2.89	3.02
	4				60.24	9.2	9.5	9.5	2.56	2.78	2.59
	6				100.98	8.5	8.8	8.7	1.14	1.37	1.2
	8				189.53	7.9	8.5	8.0	1.08	1.15	1.13
1	10	100	100	0.001	297.19	7.0	7.4	7.1	0.81	0.92	0.85
	11				486.23	6.8	7.2	7.0	0.85	1.01	0.92
	12				790.56	6.7	7.1	7.0	0.83	0.99	0.89
	13				1134.75	6.8	7.3	6.9	0.87	0.95	0.9
	14				1865.54	6.7	7.1	6.9	0.79	0.86	0.83
	15				2857.32	6.3	6.8	6.7	0.77	0.85	0.81
		50			256.14	7.1	8.1	7.7	0.95	1.04	0.99
		100			297.19	7.0	7.4	7.1	0.81	0.92	0.85
2	10	150	100	0.001	323.15	6.5	6.9	6.6	0.74	0.86	0.8
		200			402.1	6.5	7.0	6.9	0.74	0.89	0.84
			100		323.15	6.5	6.9	6.6	0.74	0.86	0.8
			200		488.62	6.1	6.6	6.2	0.69	0.80	0.77
3	10	150	500	0.001	890.13	6.4	7.0	6.6	0.71	0.90	0.86
			1000		2390.71	6.3	6.9	6.7	0.68	0.78	0.75
				0.001	488.62	6.1	6.6	6.2	0.69	0.80	0.77
				0.005	479.39	5.77	6.37	5.89	0.65	0.77	0.69
4	10	150	200	0.01	473.60	5.87	6.87	6.12	0.69	0.77	0.72
				0.05	467.20	6.09	6.6	6.38	0.74	0.79	0.77

layer by using the standard restricted Boltzmann machines algorithm (RBM). The typical range of dropout rates for hidden layers is from 0.5 to 0.8, smaller rate leads to higher accuracy and large rate may not produce enough dropout to prevent overfitting (Nitish et al., 2014). Hence, the dropout was selected at the rate of 0.5. In addition, for the training set, several algorithms are useful for training the ANN, such as the back-propagation algorithm (BPA), genetic algorithm (GA), simulating annealing algorithm (SAA), and particle swarm optimization algorithm (PSO). Among these, the most powerful training algorithm is the BPA (Zhang et al., 2007), which was selected for the training algorithm in this study. For measuring the performance of the deep-learning model, we used the mean absolute percentage error (MAPE) and mean absolute scaled error (MASE).

4. The predictive result of the deep-learning ANN

4.1. Hyper-parameters selection

The selection of hyper-parameters is crucial to the forecasting results of the model. The key hyper-parameters include the number of epochs, cyclical learning rate (CLR), batch size, and the number of nodes in each hidden layer. A CLR that is too small can increase the hazard of overfitting. Conversely, a CLR that is too large makes the training divergent. The number of epochs/iterations chosen for training should be large enough to maximize the final test performance but no larger. The total number of training examples present in a single batch and iterations is the number of batches needed to complete one epoch. Too many nodes in each hidden layer causing an extremely slow executing time, while with too few nodes, the algorithm cannot converge. The adjustment of these hyper-parameters was based on Smith (2018) combined with the trial-and-error principle. Choosing hyper-parameters must strike a balance between execution time and minimizing errors. With the numeric dataset, the usual range of hidden layers was from one to three (Niu et al., 2017; Parkes et al., 2018; Petersen, Jacobsen, & Winther, 2012; Zhang et al., 2019), additionally, more hidden layers and nodes can lead to slower execution time. The numbers of hidden layers and nodes were first tested. Based on the trial-and-error principle, we set up a certain number of hidden nodes, epochs, and CLR says 100 hidden nodes, 100 epochs, and 0.001 CLR as employed by several studies (Choi, Lee, & Kim, 2008; Parkes et al., 2018; Su, Wang, & Lai, 2012) for testing the number of hidden layers were tested from one to fifteen. Practically, the expansion from three to ten hidden layers lengthens the time by 3 min in our dataset, which is acceptable. However, the increased accuracy can be significantly improved by over 2%. The increase of hidden layers up to fifteen with the trial step of 1 presenting very long running time but not the impressive potential of accuracy as shown in Table 1. After selecting ten hidden layers, we conducted analogous procedures to choose the optimized number of hidden nodes from 50 to 200. For the remaining hyper-parameters, the programming was repetitively run, starting with 1000 epochs and a CLR of 0.001. When visualizing the loss function with 1000 epochs, the convergence was realized to locate under 500 epochs. Based on that, 100, 200, and 500 epochs were trial programmed and 200 epochs were selected in step 3. For CLR, when starting with a small value, the network begins to converge and, as the learning rate increases, it eventually becomes too large and causes the test/validation loss to increase and the accuracy to decrease was gradually increased until convergence occurred (Smith, 2018). Therefore, CLR was tested with levels of 0.001, 0.005, 0.01, 0.05 and the training and validating results are fairly converged at CLR of 0.005 as shown in Table 1. Based on those tuning procedures, Table 2 summarizes the selected list of hyper-parameters.

Table 2
Selected hyper-parameters for deep ANN.

Hyper-parameters	Value
Number of hidden layers	10
Number of epochs	200
CLR	0.005
Batch size	256
Activation function	ReLU
Training algorithm	BP, dropout
Performance criteria	MAPE; MASE

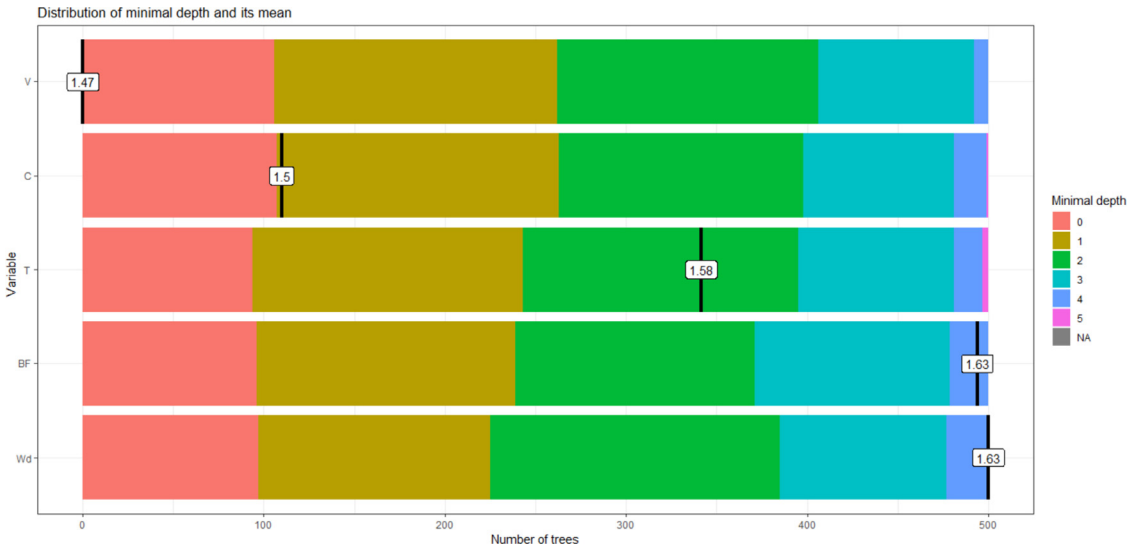


Fig. 3. Random forest mean minimal depth of variables.

Table 3
Multi-measures random forest for variables importance.

Inputs	Mean minimal depth	Number of nodes	Node purity	Number of trees
Velocity	1.471	152,830	487,705.21	500
Capacity	1.502	130,454	92,448.16	500
Time	1.583	100,226	28,128.10	500
Wind speed	1.631	101,735	19,100.37	500
Wind direction	1.638	120,398	27,982.31	500

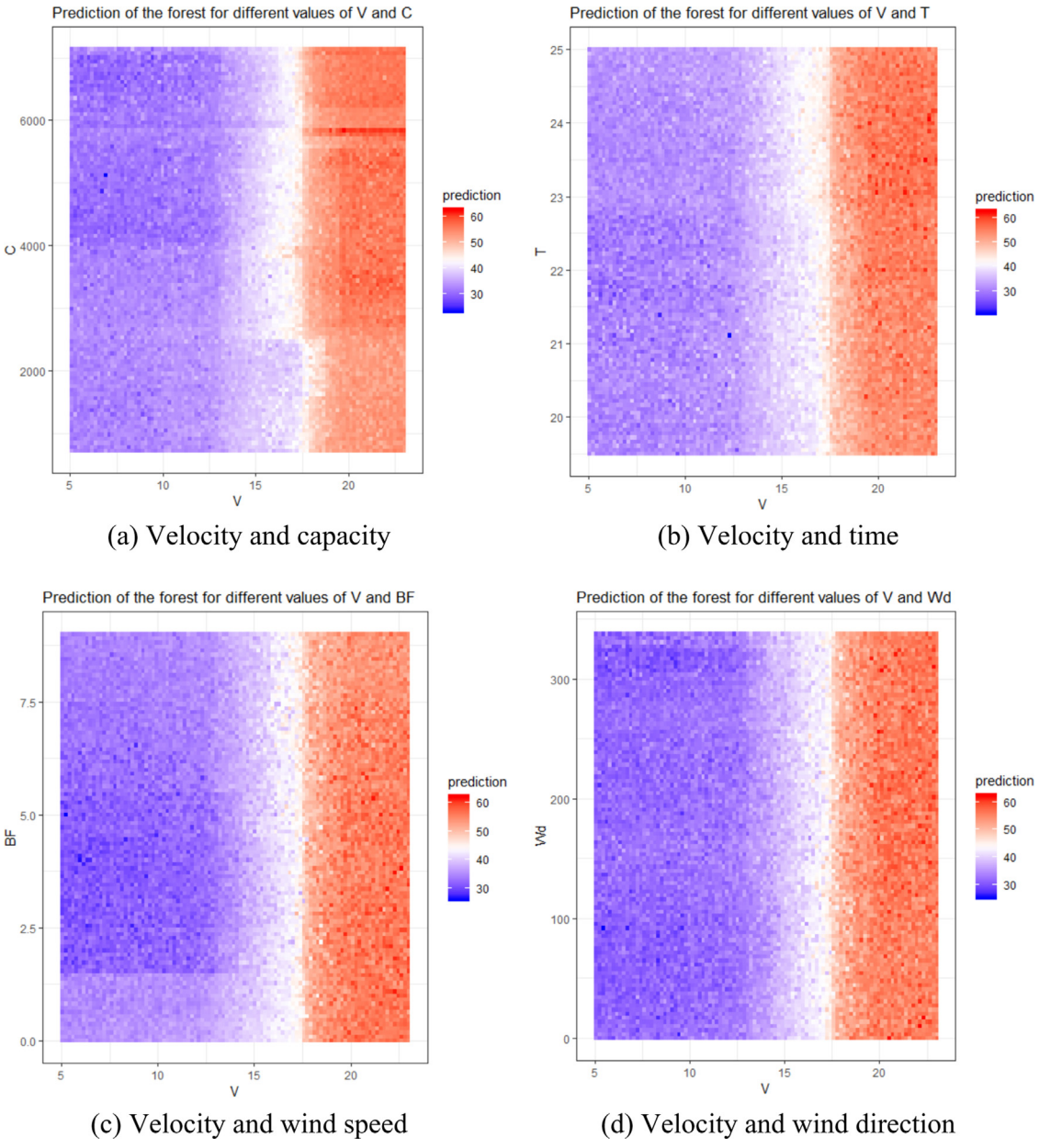
4.2. Predictive results

As mentioned in Section 3.2, this section shows the importance rank of five inputs and reveals the predictive result from the deep ANN architecture. Fig. 3 depicts the minimal depth distribution of five variables. The depth of a tree is equal to the length of the longest path from the root to the leaves in this tree. Afterwards, the programming calculated the mean minimal depth as only $\bar{B}$ out of B (number of trees) observations are considered, where $\bar{B}$ is equal to the maximum number of trees in which any variable was used for splitting. It can be easily seen that velocity is the most important because its mean minimal depth is 1.47, which is the smallest, indicating that it is very close to the root node, followed by capacity, sailing time, wind speed and wind direction. Other measures including node purity and the number of nodes also revealed identical results, as shown in Table 3.

To further investigate the most frequent interaction of variables in pairs, we plotted the prediction on a grid of values for the components of each interaction pair of velocity with another variable, as velocity has the most impact on fuel consumption. From Fig. 4, we see that higher velocity leads to higher predicted probabilities of recorded fuel consumption. The thresholds seem to exist at around 17–18 knots between the red area and the rest of the area, and around 14 knots between the blue area and the remaining area. This is consistent with studies suggesting that energy saving can be effectively obtained by slow-steaming (Psaraftis, 2019; Woo & Moon, 2014). The relationship between other components cannot be seen clearly because they are not dominant variables such as velocity. However, the main purpose of this research is to predict the output based on the input variables for shipping route choice, but not focus on the detailed relationship between output and inputs. Fig. 4(a) interestingly shows that the darkness increases from bottom to top, indicating two phenomena. Besides the logical trend at the red area that bigger ships consume more fuel than smaller ships, there is a converse phenomenon at the blue area, where bigger ships seem to consume less fuel than smaller ones. It can be easily perceived because the diagram only shows the interaction between variables in pairwise, while other predictors impact fuel consumption, and it does not reveal the general effect of capacity on energy consumption.

Although velocity is the most important variable in the forecast model, we still preserved all five variables in the deep-learning model. As the architecture is designed with optimum hyperparameters selected in Section 4.1, the deep-learning ANN is trained using the stochastic gradient descent optimization algorithm and programmed by Python 3.6.3. The programming employs the classification analysis from random forest and updates the weight of ANN for each executing time to optimize the objective loss function. Training, validating, and testing MAPE (%) is provided in Fig. 5. The results clearly show that with 200 epochs and a CLR of 0.005, the validating and testing errors yielded the minimum values. This was demonstrated by the principle that the number of epochs/iterations for training should be large enough to maximize the final test performance but no larger because, once convergence has been obtained, increasing the number of epochs results in a larger error (Smith, 2018). Comparing to the assessing criteria provided by Çelik, Teke, and Yildirim (2016) that is, MAPE $\leq$ 10% means high accuracy, 10% $\leq$ MAPE $\leq$ 20% means good prediction, 20% $\leq$ MAPE $\leq$ 50% means reasonable prediction and MAPE $\geq$ 50% means inaccurate prediction, the results also indicate that the predicted output only biased from the actual observation around 5–6%, which implies high accuracy.

Furthermore, MASE which is one of the most powerful for assessing predictive accuracy is also used in this study. MASE < 1 means that the proposed method performs smaller predictive errors than the naive method, and gets higher accuracy when MASE is smaller (Hyndman & Koehler, 2006). The MASE value of our proposed model with 200 epochs and CLR of 0.005 level is 0.77 for the



(a) Velocity and capacity

(b) Velocity and time

(c) Velocity and wind speed

(d) Velocity and wind direction

Fig. 4. Output predictive probabilities by input interaction pairwise.

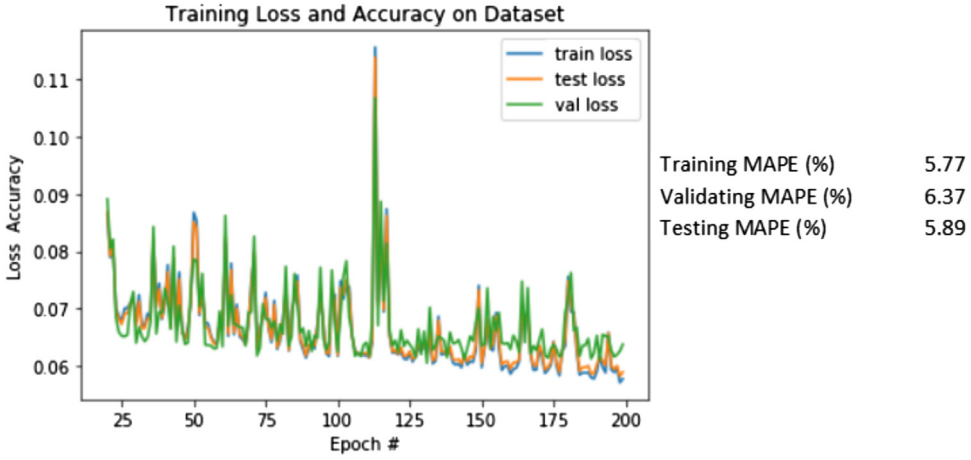


Fig. 5. Deep-learning ANN measure of performance.

validating set and 0.69 for the testing set which is the smallest. The optimal function of two variables (V, T) and three inputs (V, T, C) only obtained the MAPE 10.8% and 7.3%, respectively, while MASE values are 0.97 and 0.90, respectively. Obviously, insertion of ship's capacity into the model can significantly increase the accuracy by around 3.3% while the appearance of the wind-related weather condition only cuts down the predictive error by around 1.5%. The predictive results demonstrate the concurrence with the random forest analysis of variables importance.

5. The ATSP algorithm for liner shipping route optimization

The practical liner shipping operating problem for the shipping companies is derived from their demand to operate a roundtrip liner route. The shipping company provided us the information of the ship's capacity, name of all ports in the operating route, and the distances between them. This ship has to visit every port in the route once, and then sail back to the first departure port. The order of port arrival in the route is defined based on the minimal operating cost rule, in which the fuel cost is dominant. It can be assumed in this research and possibly in the real world that other costs are fixed for any choice. Utilizing our deep-ANN model to obtain the predicted fuel consumption amount can solve this problem by defining the optimal shipping route.

The first step is preparing all necessary parameters for the deep-learning ANN and the TSP algorithm. To do this, weather-related conditions were obtained from the Copernicus Marine Environment monitoring service, the ship's capacity is 8272 TEU, and port distance in pairs is given in Table 4. A speed window $[V_l; V_u]$ can be set up for planned speed and average speed $\bar{V}$ of the route, which is used for the model, where V_l, V_u is the lower bound and upper bound value of velocity, respectively. Sailing time is thus computed based on voyage distance and $\bar{V}$. For the fuel unit price (USD/ton), we assume that the ship purchases fuel for each leg of the route at each departure port. Therefore, we obtained the relative unit price at each port of the list from Ship and Bunker (2019).

In step 2, we used our optimal ANN model from Section 4.2 to predict the fuel consumption with all five parameters defined in step 1. To do this, we constructed one symmetric distance matrix as follows

$$[D_{ij}]_{n\times n} = \begin{bmatrix} 0 & D_{12} & \dots & D_{1n} \\ D_{21} & 0 & \dots & D_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ D_{n1} & D_{n2} & \dots & 0 \end{bmatrix} \tag{8}$$

where n is the total number of ports in the route; D_{ij} denotes the distance from port i to port j, and $\forall i,j = 1, 2, \dots, n$.

We thus computed the sailing time matrix as follows

$$[T_{ij}]_{n\times n} = [D_{ij}]_{n\times n}/\bar{V} = \begin{bmatrix} 0 & D_{12}/\bar{V} & \dots & D_{1n}/\bar{V} \\ D_{21}/\bar{V} & 0 & \dots & D_{2n}/\bar{V} \\ \vdots & \vdots & \ddots & \vdots \\ D_{n1}/\bar{V} & D_{n2}/\bar{V} & \dots & 0 \end{bmatrix}$$

$$= \begin{bmatrix} 0 & T_{12} & \dots & T_{1n} \\ T_{21} & 0 & \dots & T_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ T_{n1} & T_{n2} & \dots & 0 \end{bmatrix} \tag{9}$$

Table 4
Pairwise port distance for some Asian ports.

	Tanjung Pelepas	Hong Kong	Yantian	Vung Tau	Nansha New Port	Shekou	Singapore	Salalah	Colombo	Busan	Shanghai	Qingdao
Tanjung Pelepas	0	1851	1871	795	1842	1869	57	3557	1637	3229	2713	3066
Hong Kong	1851	0	20	934	100	18	1460	5410	3490	1475	955	1307
Yantian	1871	20	0	1085	80	25	1851	5429	3510	1488	968	1320
Vung Tau	795	934	1085	0	1056	1082	638	4353	2434	2523	1996	2348
Nansha New Port	1842	100	80	1056	0	82	1882	5400	3481	1568	1048	1400
Shekou	1869	18	25	1082	82	0	1849	5427	3507	1486	966	1318
Singapore	57	1460	1851	638	1882	1849	0	3614	1695	3208	2692	3044
Salalah	3557	5410	5429	4353	5400	5427	3614	0	1920	6787	6270	6623
Colombo	1637	3490	3510	2434	3481	3507	1695	1920	0	4867	4350	4703
Busan	3229	1475	1488	2523	1568	1486	3208	6787	4867	0	535	616
Shanghai	2713	955	968	1996	1048	966	2692	6270	4350	535	0	367
Qingdao	3066	1307	1320	2348	1400	1318	3044	6623	4703	616	367	0

Based on all computed inputs, the consumed fuel is predicted by the ANN model

$$[F_{ij}]_{n\times n} = f(\bar{V}, T, C, BF, Wd) = \begin{bmatrix} 0 & F_{12} & \dots & F_{1n} \\ F_{21} & 0 & \dots & F_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ F_{n1} & F_{n2} & \dots & 0 \end{bmatrix} \quad (10)$$

F_{ij} expresses the total fuel amount (ton) that the ship consumes for the voyage from port i to port j.

In step 3, we created an asymmetric matrix of total fuel cost by each voyage:

$$[C_{ij}]_{n\times n} = \begin{bmatrix} 0 & f_1/F_{12} & \dots & f_1 \times F_{1n} \\ f_2 \times F_{21} & 0 & \dots & f_2 \times F_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ f_n \times F_{n1} & f_n \times F_{n2} & \dots & 0 \end{bmatrix}$$

$$= \begin{bmatrix} 0 & C_{12} & \dots & C_{1n} \\ C_{21} & 0 & \dots & C_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ C_{n1} & C_{n2} & \dots & 0 \end{bmatrix} \quad (11)$$

where C_{ij} is the total fuel cost for the voyage from port i to port j (USD); f_i denotes the fuel unit price (USD/ton) at the port i, and $\forall i,j = 1, 2, \dots, n$.

Step 4 employs the asymmetric TSP algorithm to solve the following integer linear programming problem (Miller, Zemlin, & Tucker, 1960):

$$x_{ij} = \begin{cases} 1 & \text{the path goes from port } i \text{ to port } j \\ 0 & \text{otherwise} \end{cases} \quad (12)$$

For $i = 1, \dots, n$, let u_i be a dummy variable, which is finally taken to be the fuel cost from port i to port j.

$$\min \sum_{i=1}^{n} \sum_{j\neq i, j=1}^{n} C_{ij}x_{ij} : \quad (13)$$

$$x_{ij} \in \{0, 1\} \quad i, j = 1, \dots, n; \quad (14)$$

$$u_i \in Z \quad i = 2, \dots, n; \quad (15)$$

$$\sum_{j=i, j\neq 1}^{n} x_{ij} = 1 \quad j = 1, \dots, n; \quad (16)$$

$$\sum_{j=i, j\neq 1}^{n} x_{ij} = 1 \quad i = 1, \dots, n; \quad (17)$$

$$u_i - u_j + nx_{ij} \leq n - 1 \quad 2 \leq i \neq j \leq n; \quad (18)$$

$$0 \leq u_i \leq n - 1 \quad 2 \leq i \leq n. \quad (19)$$

The set of constraints (14)–(16) expresses that each port is arrived at from exactly one other port, and the set of equalities (14), (15), (17) requires that there is a departure from each port to exactly one other port. The constraints (18), (19) enforce that there is only a single tour covering all ports, and not two or more disjointed tours that only collectively cover all ports.

The ATSP integer linear programming problem that the shipping company provided us is solved by R programming. With the fuel unit price at all ports in the route given in Table 5, the assumed speed window [17.5, 18.5] (knots), $\bar{V} = 18$ knots, and the first departure port is Tanjung Pelepas, the result for the optimal route is defined as

Tanjung Pelepas → Shekou → Colombo → Nansha New Port → Qingdao → Vung Tau → Busan → Hong Kong → Salalah → Yantian → Singapore → Shanghai → Tanjung Pelepas, which is depicted in Fig. 6. The corresponding fuel cost is estimated at 98,769.6 USD for the optimal route.

Table 5
Fuel unit price at departure ports.

Port	Fuel unit price (USD/ton)
Tanjung Pelepas	358.1
Hong Kong	356.9
Yantian	401.4
Vung Tau	482.0
Nansha New Port	401.4
Shekou	401.4
Singapore	313.1
Salalah	572.0
Colombo	407.7
Busan	505.5
Shanghai	401.4
Qingdao	406.3

6. Discussion

With a given number of ports in a certain liner route, the distance of pairwise port components can easily be obtained. Combining this with planning average speed, ship's capacity, and wind-related conditions, shipping companies can optimize their route to minimize their fuel cost. This paper highlights the platform research for the liner shipping route choice based on a hybrid machine-learning method combined with an ATSP solution. In regards to planned velocity, although ship captains cannot guarantee that the real speed perfectly matches the optimal one, $\bar{V}$ expresses the operating of the ship itself. The captains can easily set this up as perfectly exact with the planned one. Additionally, the difference between V_l and V_u can suitably be 1 knot (Wang & Meng, 2015), which does not change the solution. This paper possibly provides a feasible platform to shipping companies for diminishing the fuel cost.

Besides the above impressive achievement, this paper itself has some limitations. Firstly, other ship operating costs are excluded, which can be comprehensively analyzed in the future. Some expenses may not increase when the shipping companies choose this optimal route compared to other alternatives such as the crew cost, or repair and maintenance costs, or spares. Some expenses such as marine insurance and P&I may considerably increase or decrease. Further studies should focus on detailed assessment of those expenses when the data is available. On the other hand, more comprehensive perception can be conducted by evaluating the trade-off between fuel cost curtailment and commodities flow. From the perspective of environmental protection, this study can be utilized to design a route for the minimization of exhausted emissions because fuel consumption and emissions are related to each other by the emission factor.

7. Conclusions

This paper tackled the research problem of designing the optimal route to minimize total fuel cost. A hybrid machine-learning technique was firstly constructed to predict fuel consumption for container ships based on learning the operating information of a fleet in the Asian area. With the given assumptions of a liner shipping route, the predictive results are then applied and combined with the ATSP algorithm to design an optimal route that meets the expectation of shipping companies wanting to minimize fuel cost. This is of practical significance as fuel cost contributes more

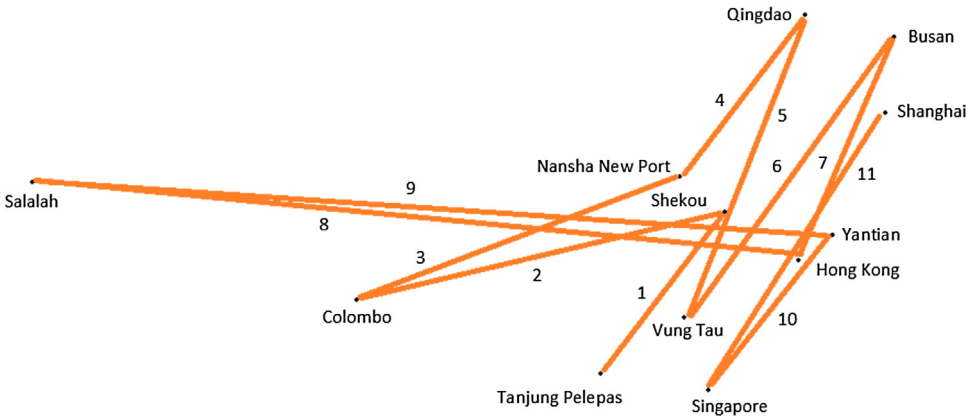


Fig. 6. Optimal shipping route solution by the ATSP algorithm.

than 60% of the ship's total operating costs. As deep-learning is a data-driven technique, it greatly depends on the quality of the learning dataset. When the recorded dataset has high convergence, this method gives high accuracy predictions. Conversely, when the data set is fragmented, it needs to be combined with other methods, such as theory-driven methods, to increase the efficiency of the model. As the dataset becomes more complete, future research directions can be extended to other types of ships or other areas with different weather conditions. More comprehensive studies of the accompanying supply chains can also be developed in the future based on this research's platform to meet a variety of economic and environmental protection goals.

Declaration of interest

The authors report no conflict of interest.

Acknowledgment

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

References

Alver, F., Saraç, B. A., & Alver Şahin, Ü. (2018). Estimating of shipping emissions in the Samsun Port from 2010 to 2015. *Atmospheric Pollution Research*, http://dx.doi.org/10.1016/j.apr.2018.02.003

Bal Beşikçi, E., Arslan, O., Turan, O., & Ölçer, A. I. (2016). An artificial neural network based decision support system for energy efficient ship operations. *Computers and Operations Research*, http://dx.doi.org/10.1016/j.cor.2015.04.004

Breiman, L. (2001). Random forests. *Machine Learning*, http://dx.doi.org/10.1023/A:1010933404324

Çelik, Ö., Teke, A., & Yildirim, H. B. (2016). The optimized artificial neural network model with Levenberg-Marquardt algorithm for global solar radiation estimation in Eastern Mediterranean Region of Turkey. *Journal of Cleaner Production*, http://dx.doi.org/10.1016/j.jclepro.2015.12.082

Chang, Y. T., Roh, Y., & Park, H. (2014). Assessing noxious gases of vessel operations in a potential Emission Control Area. *Transportation Research Part D: Transport and Environment*, http://dx.doi.org/10.1016/j.trd.2014.03.003

Chen, L., Yip, T. L., & Mou, J. (2018). Provision of Emission Control Area and the impact on shipping route choice and ship emissions. *Transportation Research Part D: Transport and Environment*, http://dx.doi.org/10.1016/j.trd.2017.07.003

Choi, B., Lee, J. H., & Kim, D. H. (2008). Solving local minima problem with large number of hidden nodes on two-layered feed-forward artificial neural networks. *Neurocomputing*, http://dx.doi.org/10.1016/j.neucom.2008.04.004

Corbett, J. J., Wang, H., & Winebrake, J. J. (2009). The effectiveness and costs of speed reductions on emissions from international shipping. *Transportation Research Part D: Transport and Environment*, http://dx.doi.org/10.1016/j.trd.2009.08.005

Dragovic, B., Tzannatos, E., Tselentis, V., Meštrovic, R., & Škuric, M. (2018). Ship emissions and their externalities in cruise ports. *Transportation Research Part D: Transport and Environment*, http://dx.doi.org/10.1016/j.trd.2015.11.007

Golias, M. M., Saharidis, G. K., Boile, M., Theofanis, S., & Ierapetritou, M. G. (2009). The berth allocation problem: Optimizing vessel arrival time. *Maritime Economics and Logistics*, http://dx.doi.org/10.1057/mel.2009.12

Grubbs, F. E. (1969). Procedures for detecting outlying observations in samples. *Technometrics*, http://dx.doi.org/10.1080/00401706.1969.10490657

Hui, W. (2012). Comparison of several intelligent algorithms for solving TSP problem in industrial engineering. *Systems Engineering Procedia*, http://dx.doi.org/10.1016/j.sepro.2011.11.070

Hyndman, R. J., & Koehler, A. B. (2006). Another look at measures of forecast accuracy. *International Journal of Forecasting*, http://dx.doi.org/10.1016/j.ijforecast.2006.03.001

Ilangkumaran, M., Sakthivel, G., & Nagarajan, G. (2016). Artificial neural network approach to predict the engine performance of fish oil biodiesel with diethyl ether using back propagation algorithm. *International Journal of Ambient Energy*, http://dx.doi.org/10.1080/01430750.2014.984082

Kashiha, M., Thill, J. C., & Depken, C. A. (2016). Shipping route choice across geographies: Coastal vs. landlocked countries. *Transportation Research Part E: Logistics and Transportation Review*, http://dx.doi.org/10.1016/j.tre.2016.03.012

Kevin Gurney. (2005). *An introduction to neural networks. Edited by Routledge*. London and New York: Taylor & Francis.

Khashei, M., & Bijari, M. (2010). An artificial neural network (p, d, q) model for time-series forecasting. *Expert Systems with Applications*, http://dx.doi.org/10.1016/j.eswa.2009.05.044

Kim, J. U., Kim, Y. D., & Shim, S. O. (2002). Heuristic algorithms for a multi-period multi-stop transportation planning problem. *Journal of the Operational Research Society*, http://dx.doi.org/10.1057/palgrave.jors.2601386

Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2012). ImageNet classification with deep convolutional neural networks. In *Advances in Neural Information Processing Systems*.

Liew, S. S., Khalil-Hani, M., & Bakhteri, R. (2016). Bounded activation functions for enhanced training stability of deep neural networks on visual pattern recognition problems. *Neurocomputing*, http://dx.doi.org/10.1016/j.neucom.2016.08.037

Lin, D. Y., & Chang, Y. T. (2018). Ship routing and freight assignment problem for liner shipping: Application to the Northern Sea Route planning problem. *Transportation Research Part E: Logistics and Transportation Review*, http://dx.doi.org/10.1016/j.tre.2017.12.003

Mahalanobis, P. C. (1930). On tests and measures of group divergence. *Journal and Proceedings, Asiatic Society of Bengal*.

Maragkogianni, A., & Papaefthimiou, S. (2015). Evaluating the social cost of cruise ships air emissions in major ports of Greece. *Transportation Research Part D: Transport and Environment*, http://dx.doi.org/10.1016/j.trd.2015.02.014

Miller, C. E., Zemlin, R. A., & Tucker, A. W. (1960). Integer programming formulation of traveling salesman problems. *Journal of the ACM (JACM)*, http://dx.doi.org/10.1145/321043.321046

Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I., & Salakhutdinov, R. (2014). Dropout: A simple way to prevent neural networks from overfitting. *Journal of Machine Learning Research*.

Niu, X., Yang, C., Wang, H., & Wang, Y. (2017). Investigation of ANN and SVM based on limited samples for performance and emissions prediction of a CRDI-assisted marine diesel engine. *Applied Thermal Engineering*, http://dx.doi.org/10.1016/j.applthermaleng.2016.10.042

Paluszynska, A., Biecek, P., & Jiang, Y. (2019). *RandomForestExplainer: Explaining and visualizing random forests in terms of variable importance*.

Parkes, A. I., Sobey, A. J., & Hudson, D. A. (2018). Physics-based shaft power prediction for large merchant ships using neural networks. *Ocean Engineering*, http://dx.doi.org/10.1016/j.oceaneng.2018.07.060

Petersen, J. P., Jacobsen, D. J., & Winther, O. (2012). Statistical modelling for ship propulsion efficiency. *Journal of Marine Science and Technology*, http://dx.doi.org/10.1007/s00773-011-0151-0

Psaraftis, H. N. (2019). Speed optimization vs speed reduction: The choice between speed limits and a Bunker Levy. *Sustainability (Switzerland)*, http://dx.doi.org/10.3390/su11082249

Rasmussen, R. (2011). TSP in spreadsheets – A guided tour. *International Review of Economics Education*, http://dx.doi.org/10.1016/S1477-3880(15)30037-2

Ship, & Bunker. (2019). *Bunker prices*.. Available from: https://shipandbunker.com/
Smith, L. N. (2018). *Disciplined approach to neural network*.
Su, R. Q., Wang, W. X., & Lai, Y. C. (2012). Detecting hidden nodes in complex networks from time series. *Physical Review E – Statistical, Nonlinear, and Soft Matter Physics*, http://dx.doi.org/10.1103/PhysRevE.85.065201
Accessed 03.09.19 Vesseltracking. (2019). *Vessel tracking* Accessed 03.09.19.. Available from: http://www.vesseltracking.net
Wang, S. (2017). Formulating cargo inventory costs for liner shipping network design. *Maritime Policy and Management*, http://dx.doi.org/10.1080/03088839.2016.1245879
Wang, S., & Meng, Q. (2015). Robust bunker management for liner shipping networks. *European Journal of Operational Research*, http://dx.doi.org/10.1016/j.ejor.2014.12.049
Woo, J. K., & Moon, D. S. H. (2014). The effects of slow steaming on the environmental performance in liner shipping. *Maritime Policy and Management*, http://dx.doi.org/10.1080/03088839.2013.819131
Zhang, J.-R., Zhang, J., Lok, T.-M., & Lyu, M. R. (2007). A hybrid particle swarm optimization-back-propagation algorithm for feedforward neural network training. *Applied Mathematics and Computation*, http://dx.doi.org/10.1016/j.amc.2006.07.025
Zhang, C., Zhang, D., Zhang, M., & Mao, W. (2019). Data-driven ship energy efficiency analysis and optimization model for route planning in ice-covered Arctic waters. *Ocean Engineering*, http://dx.doi.org/10.1016/j.oceaneng.2019.05.053
Zhao, H., Hu, H., & Lin, Y. (2016). Study on China-EU container shipping network in the context of Northern Sea Route. *Journal of Transport Geography*, http://dx.doi.org/10.1016/j.jtrangeo.2016.01.013
Zheng, J., Zheng, H., Yin, L., Liang, Y., Wang, B., Li, Z., et al. (2019). A voyage with minimal fuel consumption for cruise ships. *Journal of Cleaner Production*, http://dx.doi.org/10.1016/j.jclepro.2019.01.032