

## Research Article

# Prediction of Ship Traffic Flow and Congestion Based on Extreme Learning Machine with Whale Optimization Algorithm and Fuzzy *c*-Means Clustering

Yongjun Chen,<sup>1</sup> Ming Huang ,<sup>2,3,4</sup> Kaixuan Song,<sup>1</sup> and Tengfei Wang <sup>4</sup>

<sup>1</sup>Airport College, Binzhou University, Binzhou 256600, China

<sup>2</sup>Intelligent Transportation Systems Research Center, Wuhan University of Technology, Wuhan 430063, China

<sup>3</sup>National Engineering Research Center for Water Transport Safety (WTSC), Wuhan 430063, China

<sup>4</sup>School of Transportation and Logistics Engineering, Wuhan University of Technology, Wuhan 430063, China

Correspondence should be addressed to Ming Huang; hm@whut.edu.cn

Received 20 June 2022; Revised 30 October 2022; Accepted 26 November 2022; Published 20 May 2023

Academic Editor: Jing Zhao

Copyright © 2023 Yongjun Chen et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Accurately predicting short-term congestions in ship traffic flow is important for water traffic safety and intelligent shipping. We propose a method for predicting the traffic flow of ships by applying the whale optimization algorithm to an extreme learning machine. The method considers external environmental uncertainty and complexity of ships navigating in traffic-intensive waters. First, the parameters of ship traffic flow are divided into multiple modal components using variational mode decomposition and extreme learning machine. The machine and the whale optimization algorithm constitute a hybrid modelling approach for predicting individual modal components and integrating the results of individual components. Considering a map between ship traffic flow parameters and congestion, fuzzy *c*-means clustering is used to predict the level of ship traffic congestion. To verify the effectiveness of the proposed method, ship traffic flow data of the Yangtze River estuary were selected for evaluation. Results from the proposed method for predicting ship traffic flow parameters are consistent with measurements. Specifically, the prediction accuracy of the ship traffic congestion reaches 76.04%, which is reasonable and practical for predicting ship traffic congestion.

## 1. Introduction

The booming growth in trade has led to the development of the waterway transportation sector, and ship traffic flow prediction has become extremely important with the ever-increasing waterway transport demand. The Yangtze River, as the largest river that runs through the east and west of China, is the main channel to the sea for 11 inland provinces and autonomous regions of China and the main transportation artery connecting the three economic zones in southwest, central, and east part of China. The Yangtze River plays a pivotal role in the country's economic and social development because of its superior navigable conditions and huge transit capacity. Once the ships in the Yangtze River are blocked, the cost loss and environmental pollution will be a great loss. Therefore, it is a goal of researchers,

practitioners, and government managers to develop an approach for predicting ship traffic flow. The status of expected ship traffic flow can be used to identify congestion early, improve canal traffic, and prevent accident.

Methods for predicting traffic flow, which were first only employed for predicting automobile traffic flow, have found widespread use now. Xiao et al. [1] proposed and built differential information from the car-following inertia gray model and gray system based on the classic car-following model and traffic status to establish differential equations. Yan et al. [2] constructed a comprehensive traffic flow indication system based on the robust L1 model minimum two-multiplier twin support vector regression. Based on a recurrent neural network, Cui et al. [3] used a two-way/one-way stacked long short-term memory (LSTM) network for traffic prediction. They combined a recurrent neural

network and its variants into a traffic prediction method. Li et al. [4] used the dense geometric correspondence network for multistep traffic prediction. By applying dynamic relations between traffic stations in space and time, Peng [5] proposed a prediction framework based on a neural network to dynamically map urban traffic flow. Cai et al. [6] derived the posttest estimation of the maximum correlation in the fixed embedding algorithm update such that a Kalman filter provides suitable traffic flow prediction. Wang et al. [7] combined traffic flow data and considered weather conditions to build a short-term traffic flow prediction model based on an attention mechanism and a one-dimensional convolutional neural network with LSTM. Fang et al. [8] found that a typical LSTM network is prone to small fluctuations and can achieve high-accuracy traffic flow prediction. Shen et al. [9] performed traffic flow prediction considering multi-intersection perception. Lu et al. [10] performed isolated-point short-term traffic flow prediction based on a time-aware convolutional context block LSTM network and a novel loss switching mechanism. Li et al. [11] performed deep feature learning by applying an advanced multitarget particle group optimization algorithm to the parameters of a deep belief network and by applying supervised learning to predict short-term traffic flow. Zhang et al. [12–14] proposed convolutional neural networks based on dynamic feature coding. Considering the selection of historical data, Zhang [13] proposed a selection method to collect appropriate historical data for daily traffic flow prediction.

Salamanis [14] proposed an efficient large-scale multistep traffic prediction model with fast inference. Almeida [15] used statistical algorithms and neural networks to describe and predict traffic flow. Lee and Rhee [16] identified two basic spatial dependencies in traffic, used distance, direction, and location diagram convolution networks, and transformed the three spatial relations into deep neural networks for traffic speed prediction. Pavlyuk [17, 18] applied integrated learning techniques in space-time structure learning and applied them to predict short-term traffic flow in cities. In addition, the researcher proposed a spatiotemporal cross-validation method to evaluate the model performance. Alves and Cordeiro [19] developed an adaptive algorithm to accurately predict traffic flow and monitor highways connected to a complex network using local traffic measurements. Carpio [20] confirmed that LSTM-based traffic prediction can effectively reduce the complexity of traffic prediction. To overcome the limitation of traffic prediction on a specific road, Jin [21] adapted the bidirectional encoder representations from transformers to traffic modelling, suitably predicting traffic flow in various routes.

With the rapid growth in transport demand worldwide, traffic congestion is gradually increasing owing to natural and artificial factors. Traffic congestion not only causes unnecessary waste of time and human and financial resources for those affected but also causes delays and lags in world trade [12]. Therefore, it is critical to study and mitigate traffic congestion. To solve traffic congestion in specific scenarios, Jin [22] formulated a dual-target cash transport

vehicle path problem and designed a local search second algorithm for a particular terrain. Mei [23] applied an efficient clustering algorithm for the k-core decomposition of large networks. Guo [24] applied a method for early warning of traffic congestion areas based on dynamic identification, which can effectively track the target and detect congestion warning areas throughout the congestion evolution. Gao [25] applied a method to quantify the degree of traffic congestion, constructed an image-based traffic congestion assessment framework, and integrated a traffic parameter layer into a basic convolutional neural network that accelerates processing and avoids complex postprocessing. Zhou [26] proposed a clustering integration method based on structured hypermap learning to improve the clustering efficiency, stability, and robustness.

Nguyen [27] used clustering to automatically label data, qualitatively assess the significance of the labels, and quantify the effect of label separation results in the feature space. Costa [28] proposed an unsupervised approach for relating topic modelling and document clustering, seamlessly unifying and jointly performing both tasks using Bayesian generative modelling and post-ambiguous reasoning. As a result, a method was derived to estimate the congestion index using a speed transfer matrix and the corresponding center of mass. The congestion index is estimated using fuzzy reasoning optimized by genetic algorithms. The index is evaluated using data from a receiver of global satellite navigation system, and traffic status estimates can be obtained for most evaluated roads. Peixoto [29] devised clustering analysis for reducing traffic information at the edge of a vehicle network. Peter [30] applied the RSRU\_TM method for improving traffic management. Based on mobile crowd sensing for dynamic traffic efficiency estimation, Ali [31] applied an algorithm for traffic congestion control. Huertas [32] combined an unsupervised clustering algorithm with a predictive model based on multiple logistic regression for scalable prediction and analysis of traffic accidents. Chiabaut [33] proposed a method for real-time assessment of traffic conditions. Bhatia [34] applied a data-driven approach to build artificial intelligence models for vehicle traffic behavior prediction.

The aforementioned research findings offer a solid foundation for the study of the congestion risk of ship traffic flow, but given the state of the research, there is no comprehensive research on the variables affecting ship traffic flow and congestion risk [35]. Wang et al. [36] designed a ship traffic flow model based on a multiple hexagon-based convolutional neural network (mh-CNN) method. However, the large spatial resolution of the method makes it difficult to be used in inland waterway navigation studies. Xu and Zhang [37] proposed a RNN-based method to predict the ship traffic flow of Yangtze River. The method is simple and effective; however, the spatiotemporal dependence of ship traffic flow was not considered in the method. Ship traffic flow statistics, on the other hand, are highly spatially articulated and temporally correlated and are influenced by a range of factors. From the standpoint of water traffic spatial correlation [38, 39], adjacent sections or channels are impacted by nearby vessel traffic flow. Ship traffic flow exhibits

a particular time correlation in the close time distance [40]. So, it makes sense to look at how to create a suitable congestion risk model for ship traffic flow.

Considering the advantages and drawbacks of the abovementioned developments, we use variational mode decomposition (VMD), an extreme learning machine (ELM), the whale optimization algorithm (WOA), and fuzzy  $c$ -means (FCM) clustering to accurately predict ship traffic flow and congestion. In this paper, the VMD-ELM-WOA method is applied to predict ship traffic flow, which is convenient for planning in water areas and time periods prone to blockage and reduces the queuing time. The three parts of VMD-ELM-WOA are innovatively combined to reduce the time of traffic volume prediction and further improve the calculation accuracy.

The remainder of this paper is organized as follows. The traffic flow prediction model using VMD-ELM-WOA is introduced in Section 2. Section 3 details the prediction model for ship traffic congestion. In Section 4, a case study based on measured ship traffic data is reported along with its results in different scenarios. Finally, we draw conclusions in Section 5.

## 2. VMD-ELM-WOA Method for Predicting Ship Traffic Flow

**2.1. VMD Method of Ship Traffic Flow Analysis.** VMD was proposed in 2014 as a nonrecursive method. Since then, this decomposition method has been widely used to treat nonlinear problems, being suitable for processing traffic flow time series that have strong nonlinearity and high complexity. In fact, accurate prediction of ship traffic flow requires the decomposition of the corresponding time series.

VMD establishes a variable problem that is constantly updated by constructing a constrained model center and bandwidth to find the optimal solution. Suppose that a signal is set to be band-decomposed into  $k$  intrinsic mode function components by a variable fractional modality. Then,

$$\min_{\{u_k\}, \{\omega_k\}} \left\{ \sum_k \left\| \partial_t \left[ \left( \delta(t) + \frac{j}{\pi t} \right) u_k(t) \right] e^{-j\omega_k t} \right\|_2^2 \right\}. \quad (1)$$

The corresponding constrained model can be expressed as

$$s.t. \sum_k u_k = f, \quad (2)$$

where  $\{u_k\} = \{u_1, \dots, u_k\}$  and  $\{\omega_k\} = \{\omega_1, \dots, \omega_k\}$  are the modal numbers and center frequencies, respectively. Equation (1) can be solved by introducing Lagrangian operator  $\lambda$  and secondary penalty factor  $\alpha$ , which are translated into the following unconstrained problem:

$$L(\{u_k\}, \{\omega_k\}, \lambda) = \alpha \sum_k \left\| \partial_t \left[ \left( \delta(t) + \frac{j}{\pi t} \right) u_k(t) \right] e^{-j\omega_k t} \right\|_2^2. \quad (3)$$

The update of  $u_k^{n+1}$  and, according to the alternating direction method of multipliers, the update of  $\omega_k^{n+1}$  and  $\lambda_k^{n+1}$  can be expressed as

$$\hat{u}_k^{n+1}(\omega) = \frac{(\hat{f}(\omega) - \sum_{i \neq k} \hat{u}_i(\omega) + 0.5\hat{\lambda}(\omega))}{(1 + 2\alpha(\omega + \omega_k)^2)}, \quad (4)$$

$$\omega_k^{n+1} = \left( \frac{\int_0^\infty \omega |\hat{u}_k(\omega)|^2 d\omega}{\int_0^\infty |\hat{u}_k(\omega)|^2 d\omega} \right), \quad (5)$$

$$\hat{\lambda}^{n+1}(\omega) = \hat{\lambda}^n(\omega) + \tau \left[ \hat{f}(\omega) - \sum_k \hat{u}_k^{n+1}(\omega) \right]. \quad (6)$$

Finally, convergence is established for constant  $e > 0$  as follows:

$$\sum_{k=1}^k \frac{\|\hat{u}_k^{n+1} - \hat{u}_k^n\|_2^2}{\|\hat{u}_k^n\|_2^2} < e. \quad (7)$$

The steps for VMD are summarized as follows:

- (1) Set initialization parameters  $\{u_k^1\}$ ,  $\{\omega_k^1\}$ ,  $\lambda^1$  and  $\{\hat{u}_k^{n+1}\}$ ,  $\{\omega_k^{n+1}\}$ ,  $\hat{\lambda}^1$ .
- (2) Update equation (6) based on the results of equations (3)–(5).
- (3) Update until the convergence condition in equation (6) is satisfied.
- (4) Return the corresponding modal component based on the number of modes.

**2.2. ELM Method of Ship Traffic Flow Prediction.** A conventional neural network shows a slow convergence and high computation burden, impeding accurate predictions over a time series. Huang et al. proposed the ELM, which is basically a feedforward neural network, whose input weights and thresholds can be randomly initialized to improve the training efficiency for use in many cases. A diagram of the ELM structure is shown in Figure 1.

Consider  $N$  samples  $(X_i, X_j)$  that describe the flow and density of ship traffic. The input layer is mapped by the active function as follows:

$$\sum_{i=1}^L \beta_i f(\omega_i x_j + b_i) = o_j, \quad (8)$$

where  $L$  is the number of hidden layer nodes,  $x_1, x_2, \dots, x_i, \omega_i$  is the input weight,  $\beta_i$  is the output weight, and  $b_i$  is the bias of the hidden node.

After network training achieves learning with zero error,

$$\sum_{j=1}^N \|o_j - t_j\| = 0, \quad (9)$$

and there are  $\omega_i, \beta_i$ , and  $b_i$  that make the following condition hold:

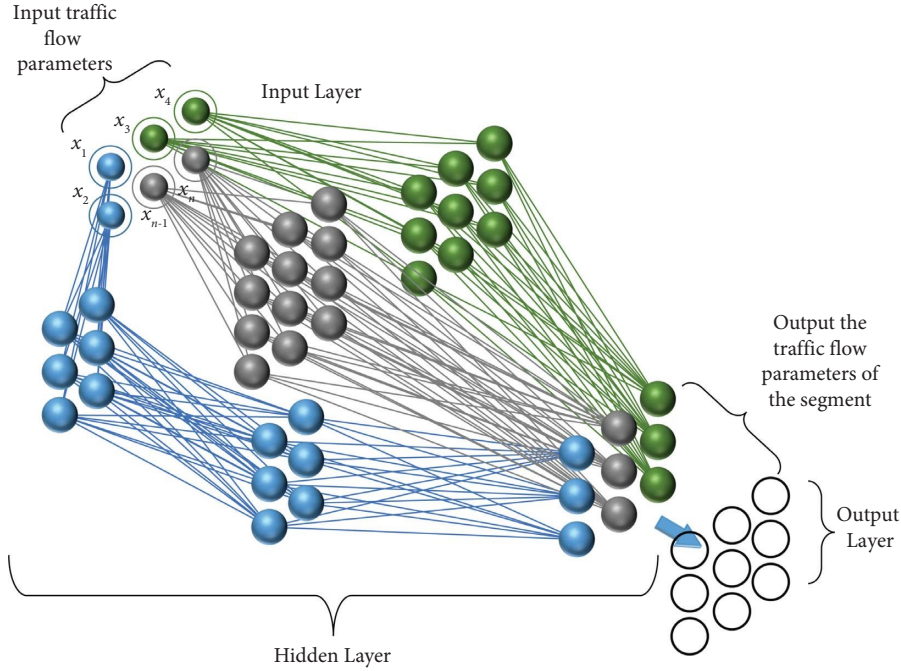


FIGURE 1: ELM diagram.

$$\sum_{i=1}^N \beta_i f(\omega_i X_j + b_j) = t_j, \quad (10)$$

where  $\beta_i$  is the output weight,  $b_i$  is the threshold between implicit layers, and  $t_j$  is the output value. Equation (9) can be written in the following matrix form:

$$H\beta = T, \quad (11)$$

$$H = \begin{bmatrix} f(\omega_1 X_1 + b_1) & \cdots & f(\omega_N X_1 + b_L) \\ \vdots & \ddots & \vdots \\ f(\omega_1 X_N + b_1) & \cdots & f(\omega_L X_N + b_L) \end{bmatrix}.$$

Considering the ELM principle, the weights and thresholds of the input can be randomly assigned, whereas the weights between the hidden layer and output layer can be obtained using the solution to the system of equations:

$$\beta^* = H^* \cdot T, \quad (12)$$

where  $*$  denotes the Moore–Penrose pseudoinverse of the corresponding matrix.

**2.3. WOA of Parameter Determination.** Owing to the randomness of ELM weights and thresholds, the predictions may be biased. Therefore, we apply the WOA to the internal parameters of the ship traffic flow model based on ELM. The WOA is a heuristic algorithm that resembles the behavior of humpback whales. In brief, foraging is described by a spiral bubble net for catching a prey, eventually leading to food swallowing. This behavior is expressed in a target model as detailed in the following.

### 2.3.1. Surrounding Prey

$$\begin{cases} \vec{D} = |\vec{C}\vec{X}^*(t) - \vec{X}(t)|, \\ \vec{X}(t+1) = \vec{X}^*(t) - \vec{A} \cdot \vec{D}, \end{cases} \quad (13)$$

where the vectors for the whale's current position and optimal position are used while the whale constantly updates its position during foraging. Vectors  $\vec{A}$  and  $\vec{C}$  are defined as

$$\begin{cases} \vec{A} = 2\vec{a} \times \vec{r} - \vec{a}, \\ \vec{C} = 2 \times \vec{r}. \end{cases} \quad (14)$$

### 2.3.2. Bubble Net Attack

- (1) Shrink surroundings: the bubble net is shrunk by adjusting the die value of vector  $\vec{a}$ , which drops linearly from 2 to 0 throughout the bubble net attack.
- (2) Update spiral: The humpback whales can also switch positions in a spiral during hunting. In addition to the shrinking surroundings to enclose the prey, the following update is defined:

$$\begin{cases} \vec{X}(t+1) = \vec{D}^l \cdot e^{bl} \cdot \cos(2\pi l) + \vec{X}^*(t), \\ \vec{D}^l = |\vec{X}^*(t) - \vec{X}(t)|, \end{cases} \quad (15)$$

where  $l$  is a random value affecting the distance between the current and optimal positions.

- (3) Search for prey: In addition to tracking individual whales in optimal locations, humpback whales also

use random searches to track their prey. This process can be described as follows:

$$\begin{aligned} \vec{D} &= |\vec{C} \cdot \vec{X}_{\text{rand}} - \vec{X}|, \\ \vec{X}(t+1) &= |\vec{X}_{\text{rand}} - \vec{A} \cdot \vec{D}|. \end{aligned} \quad (16)$$

The update of population locations depends on the die value of A, which is randomly set for a value greater than 1. When the die value is less than 1, the population searches along the direction of the optimal individual location.

**2.4. The Hybrid Method of VMD-ELM-WOA.** In view of the complexity and strong stochasticity of ship traffic flow, a combinatorial approach is proposed to perform time series of ship traffic flow prediction, as shown in the structure in Figure 2. The time series is first decomposed into multiple intrinsic mode function components using VMD. Then, prediction on every component is performed to finally superimpose all the predictions.

- (1) Reduce the randomness and complexity of the original traffic signal by splitting it into multiple intrinsic mode function components through VMD.
- (2) Build a combined prediction method. The weights and thresholds of the ELM for prediction are randomly initialized. Then, the WOA is applied to optimize the internal ship traffic flow parameters of the network, forming the hybrid ELM-WOA model.
- (3) Build a composite method based on the components obtained from step 1 to obtain multiple predictions and combine them for the final prediction.

### 3. Prediction of Ship Traffic Congestion

Predicting ship traffic flow relies on determining its main parameters. In addition, ship traffic congestion is reflected according to the corresponding ship traffic flow parameter. However, ships are often affected by external environmental disturbances, and their flow is complex and uncertain. Hence, we use FCM clustering to divide the ship traffic flow into different parameter mappings and then classify congestion in segments. Given the lack of quantitative criteria to classify ship traffic congestion, we analyze the classification of road vehicle congestion, comprehensively assess the characteristics of ship traffic flow in the waters of the Yangtze estuary, and finally classify ship congestion in navigable waters into four levels according to the congestion levels listed in Table 1.

The flow diagram for predicting ship traffic congestion is shown in Figure 3. Using the ship automatic identification system to build a base station, we can obtain dynamic data from ships navigating in the channel, reject invalid and abnormal data, determine the ship traffic flow and density, and predict ship traffic flow based on the VMD-ELM-WOA method. Then, FCM clustering is applied to obtain the level of ship traffic congestion.

FCM clustering is applied to the test set containing ship traffic flow information. As a result, the test set is divided into different clusters, and the interaction between clustering and a test sample can be used to quickly identify the  $k$ -nearest neighbors for classification. Then, the congestion level is determined according to the weight of the ship traffic flow and density sample by weighting the neighborhood.

The calculation based on FCM clustering proceeds as follows:

- (1) Apply FCM clustering to divide the test set into  $N$  clusters:  $X_1, X_2, \dots, X_N$ , with cluster collection centers  $u_1, u_2, \dots, u_N$ . Use any cluster  $X_k$  ( $1 \leq k \leq N$ ) to create a sample of the ship traffic flow and density of  $m_1, m_2, \dots, m_c$  ( $1 \leq c \leq l$ ). Define cluster radius  $R_k$  for the test sample. The maximum distance between the test sample and cluster center contained in  $X_k$  is given by

$$R_k = \max(\text{dist}(m_1, u_k), \text{dist}(m_2, u_k), \dots, \text{dist}(m_c, u_k)). \quad (17)$$

- (2) Define the predicted sample for randomly selected cluster  $X_k$  ( $1 \leq k \leq N$ ) and the target sample for prediction. The distance between the sample and cluster  $X_k$  is  $d_k$  with the following boundary distance:

$$d_k = \text{dist}(\bar{m}, u_k) - R_k. \quad (18)$$

If  $d_k > 0$ , the sample of ship traffic flow parameters to be predicted does not fall within the clustering level. If  $d_k = 0$ , the sample is at the clustering boundary.

- (3) Set distance  $d_{KNN_i}$  between the predicted ship traffic flow parameter sample and training set of nearest adjacent samples (1, 2, ...,  $z$ ). The weight corresponding to the sample of the nearest adjacent ship traffic flow parameters for  $k$  is given by

$$w_k = \frac{(\sum_{k=1}^l d_{KNN_k} - d_{KNN_k})}{((l-1)\sum_{k=1}^l d_{KNN_k})}. \quad (19)$$

### 4. Experimental Case Study

**4.1. Data and Evaluated Methods.** The data for this study were obtained from 4-month ship traffic flow recordings from September to December 2020. Data were collected every 15 min, 30 min, and 1 h for segments B, A, and C, respectively. The dynamic data from September 6 to December 29, 2020, were used as the training set with a prediction part from segment B. The data from December 30, 2020, were used for prediction. The samples for training were 11,256, 5768, and 2904, and those for prediction were 96, 48, and 24, respectively.

To verify the effectiveness of the proposed VMD-WOA-ELM method to predict ship traffic flow, we compared our complete method with combinations WOA-ELM and EMD (empirical mode decomposition)-WOA-ELM and prediction models ELM and backpropagation (BP). To fully evaluate the proposed VMD-WOA-ELM method, the

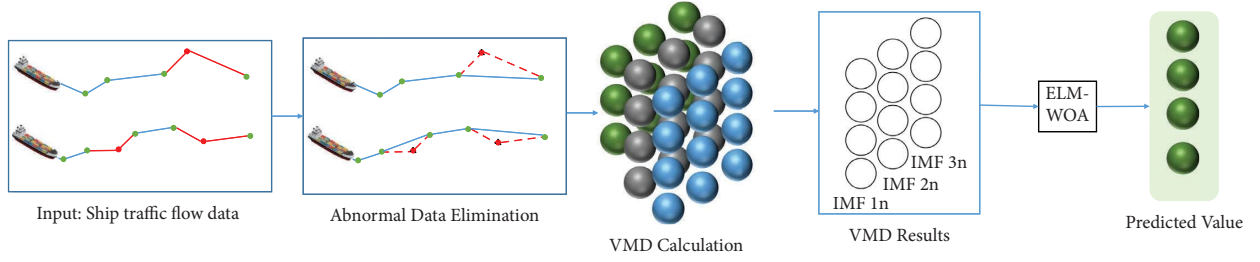


FIGURE 2: Diagram of VMD-ELM-WOA method for predicting ship traffic flow (IMF, intrinsic mode function).

TABLE 1: Classification of ship congestion adopted in this study.

Congestion level	I	II	III	IV
Description	Smooth flow	Mild congestion	Moderate congestion	Severe congestion
Congestion index range	(0, 0.3]	(0.3, 0.6]	(0.6, 0.8]	(0.8, 1]

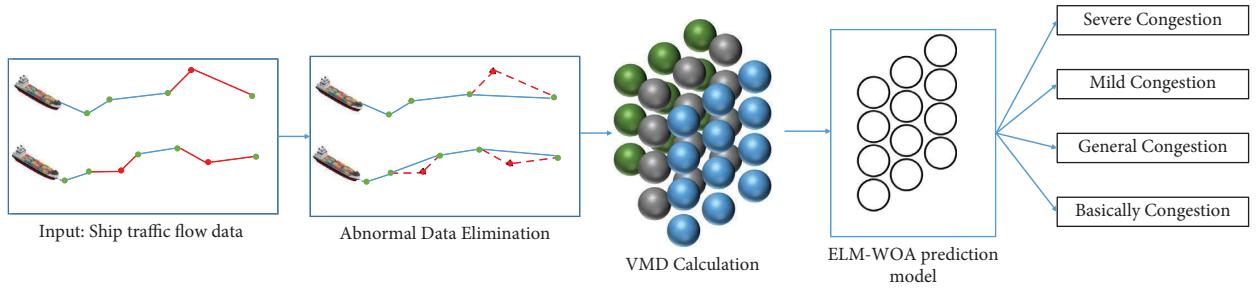


FIGURE 3: Flow diagram for predicting ship traffic congestion.

training set and sample set of the same traffic flow and density were selected as inputs, and the prediction was compared with the corresponding measurement. The computational platform used in this study was MathWorks MATLAB 2017.

**4.2. Evaluation Indicators.** To evaluate the accuracy of the prediction methods for ship traffic flow parameters, we used the mean absolute error (MAE), root-mean-square error (RMSE), and mean absolute percentage error (MAPE):

$$d_k = \text{dist}(\bar{m}, u_k) - R_k,$$

$$\text{RMSE} = \sqrt{\frac{1}{n} \sum (y_t - \hat{y}_t)^2}, \quad (20)$$

$$\text{MAPE} = \frac{1}{n} \sum_{i=1}^n \left| \frac{y_t - \hat{y}_t}{\hat{y}_t} \right|,$$

where  $y_t$  is the measured value,  $\hat{y}$  is the predicted value, and  $n$  is the number of samples.

**4.3. Spatiotemporal Evaluation.** Using the proposed and comparison methods for ship traffic flow prediction, we obtained the traffic flow and density of segment B with data from December 30, 2018. The traffic flow prediction over

time is shown in Figure 4 for intervals of 15 min, 30 min, and 1 h, and the corresponding MAE, RMSE, and MAPE are listed in Table 2.

Figure 5 and Table 3 show the corresponding prediction results for ship traffic density.

**4.4. Traffic Congestion Level.** Using FCM clustering, traffic flow and density measurements for the first 3 months of segment B were collected and combined with traffic flow and density measurements for December 30. Depending on the traffic flow parameters, traffic congestion is defined into four levels according to the navigation conditions, as shown in Figure 6. The prediction in segment B considering 15 min intervals and the measurements for December 31 are shown in Figure 7. The predicted traffic congestion is similar to the measurements, with the prediction accuracy reaching 76.04%, which indicates a suitable prediction of ship traffic congestion.

**4.5. Discussion.** In order to quantitatively verify the applicability of the model, this paper uses five models to calculate the values of MAE, RMSE, and MAPE for all 15 min, 30 min, and 60 min in December 2018, as shown in Tables 2 and 3. From the tables, it can be seen that VMD-WOA-ELM maintains a lower value in different time periods, as well as ship traffic flow and density, and on 3D colormap surface

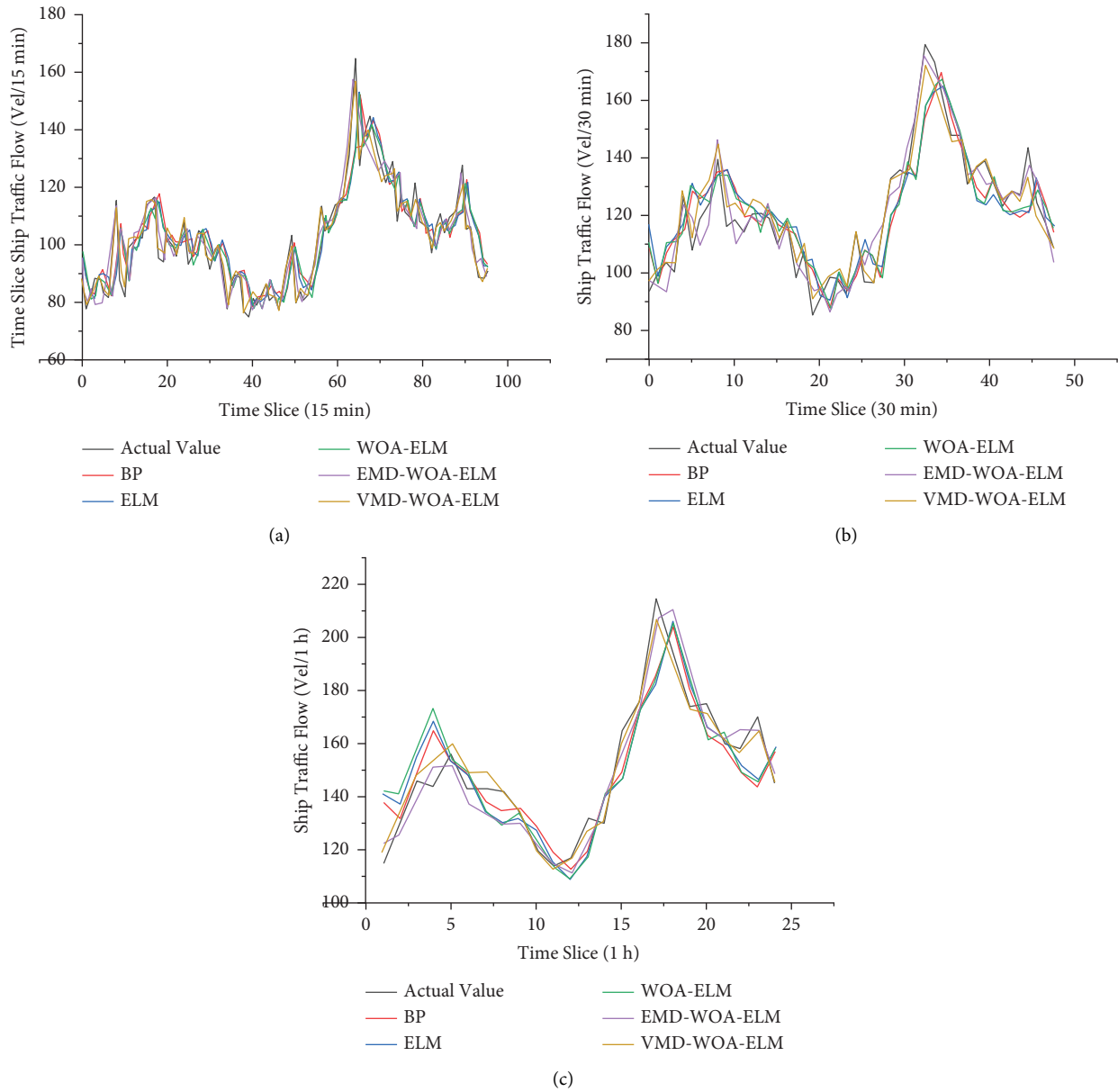


FIGURE 4: Results from proposed and comparison methods for prediction of ship traffic flow. (a) Traffic flow forecast of section B in 15 min on December 30, 2018. (b) Traffic flow forecast for section B in 30 minutes on December 30, 2018. (c) 1-hour traffic flow forecast for section B on December 30, 2018.

TABLE 2: Comparison of prediction of ship traffic flow between proposed and comparison methods.

Time slice	15 min			30 min			1 h		
	MAE	RMSE	MAPE (%)	MAE	RMSE	MAPE (%)	MAE	RMSE	MAPE (%)
WOA-ELM	6.988 0	9.111 0	6.994 0	8.974 8	10.931 4	7.103 5	11.833	14.735 5	8.044 8
EMD-WOA-ELM	4.560 8	5.898 2	4.558 3	5.509 0	6.632 2	4.324 9	6.634 1	8.239 7	4.481 5
VMD-WOA-ELM	2.099 4	2.640 4	2.114 0	3.393 2	4.236 5	2.810 9	3.143 5	4.112 0	2.135 6
ELM	7.570 2	9.840 5	7.523 8	9.842 7	11.695 2	8.091 5	12.751 9	15.351	8.555 6
BP	7.420 5	9.869 5	7.402 1	9.288 6	11.922 2	7.733 5	12.454 1	15.909	8.493 3

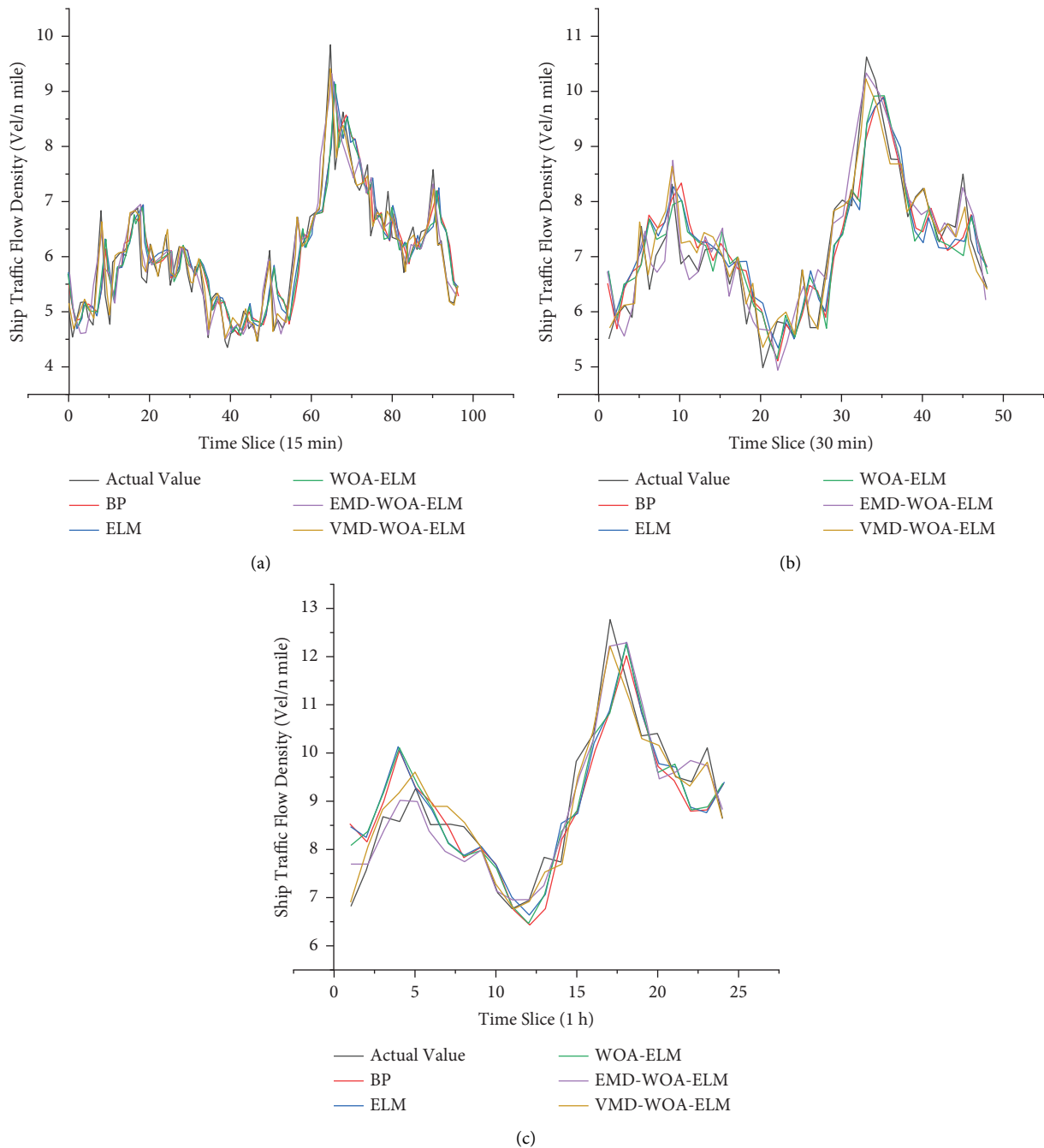


FIGURE 5: (a) Traffic flow prediction and error results of traffic flow prediction model for 15 min in section B on December 30, 2018. (b) Traffic flow prediction and error results of traffic flow prediction model for 30 min in section B on December 30, 2018. (c) Traffic flow prediction and error results of the 1-hour traffic flow prediction model in section B on December 30, 2018.

plot with projection, the mapping value is lower than that of the comparison model.

The prediction trend in the characteristic parameters of ship traffic flow is similar to the trend in the corresponding measurements, and FCM clustering allows to determine the level of ship traffic congestion over time and space, reflecting spatiotemporal heterogeneity. In addition, the proposed prediction method outperforms the comparison methods, as indicated by its lowest prediction errors. FCM clustering defines the congestion level according to the degree of

membership, thus performing a quantitative analysis of the ship traffic flow parameters.

The MAE, RMSE, and MAPE values of the VMD-WOA-ELM-based ship traffic flow prediction method proposed in this work are smaller and more stable generally, as illustrated in Figures 8–10. This demonstrates that the proposed method has a higher prediction accuracy since the prediction values achieved by using this method are closer to the measured values and can depict the application of ship traffic flow in terms of spatiotemporal variability. The technique



TABLE 3: Comparison of prediction of ship traffic density between proposed and comparison methods.

Time slice	15 min			30 min			1 h		
Method	MAE	RMSE	MAPE (%)	MAE	RMSE	MAPE (%)	MAE	RMSE	MAPE (%)
WOA-ELM	3.427 1	0.563 6	7.0071	0.529 8	0.656 7	7.156 7	0.617 9	0.786 8	7.444 3
EMD-WOA-ELM	0.263 1	0.343 9	4.446 0	0.342 1	0.408 7	4.468 6	0.400 6	0.473 1	4.518 8
VMD-WOA-ELM	0.126 5	0.158 9	2.132 2	0.196 2	0.248 4	2.744 1	0.174 5	0.234 0	2.002 3
ELM	0.447 3	0.585 4	7.460 1	0.532 7	0.677 1	7.321 7	0.705 9	0.9109	8.054 8
BP	0.4346	0.586 6	7.308 5	0.557 1	0.683 5	8.071 8	0.687 6	0.832 2	7.861 6

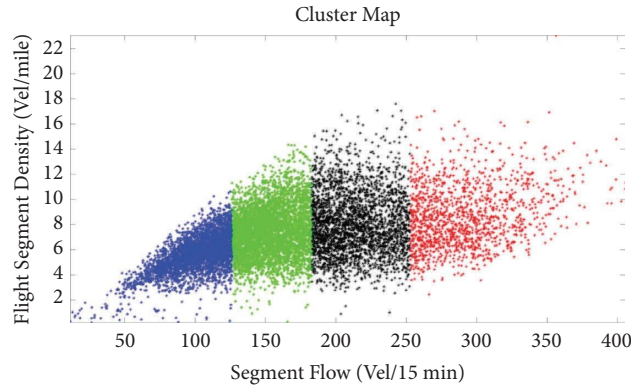


FIGURE 6: Cluster of ship traffic flow parameters and prediction parameters for first 3 months of segment B.

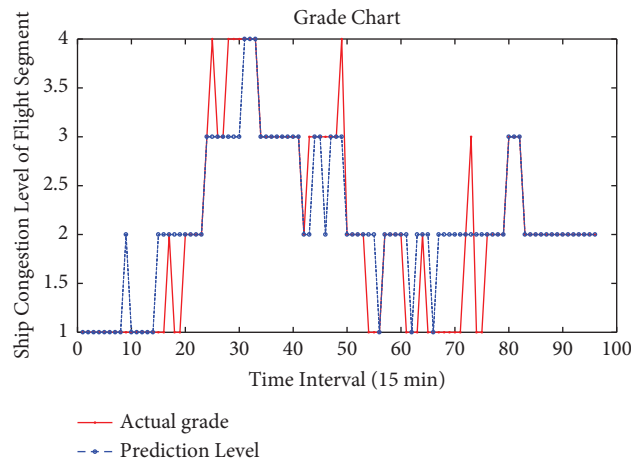


FIGURE 7: Measured and predicted levels of ship traffic congestion.

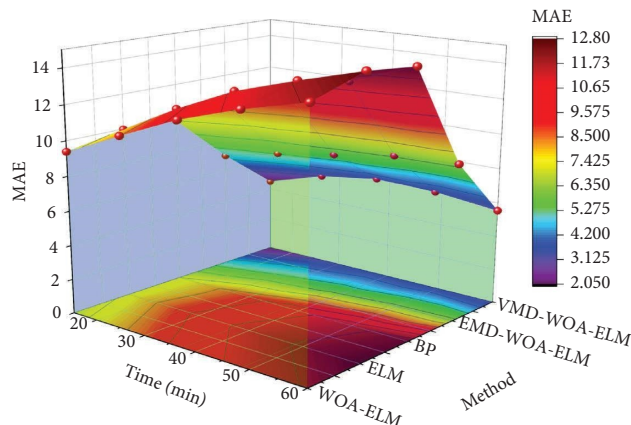


FIGURE 8: MAE of ship traffic flow.

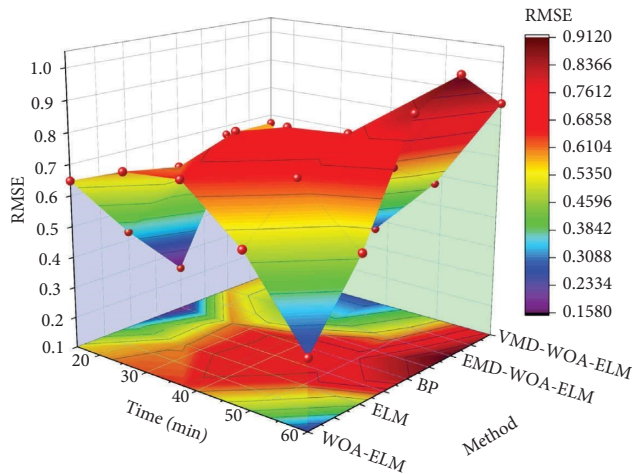


FIGURE 9: RMSE of ship traffic flow.

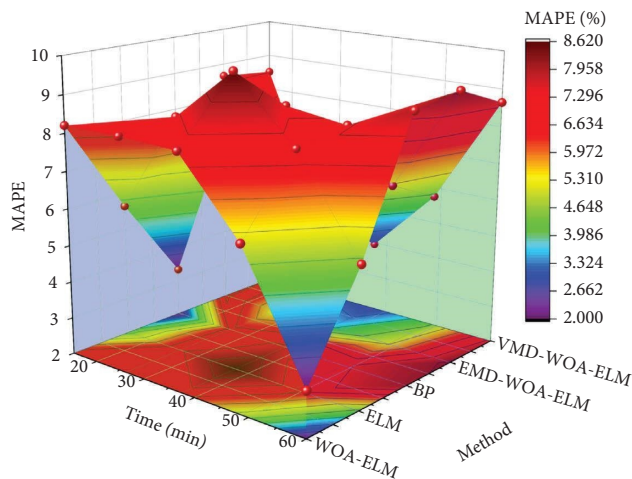


FIGURE 10: MAPE of ship traffic flow.

efficiently addresses the issue of redundant ship traffic flow data in areas with high ship traffic and lays the groundwork for developing a deep learning-based ship traffic flow prediction model.

## 5. Conclusions

A method is proposed in this paper to predict ship traffic flow using VMD and a hybrid ELM-WOA model. The ship traffic flow parameters are decomposed into multiple modes, and the ELM with whale optimization predicts each component, extracting spatiotemporal characteristics of ship traffic flow. Then, ship traffic congestion is predicted using FCM clustering, obtaining the parameter clustering graph and congestion level map. Experimental results show that the proposed method has high accuracy, with the predicted and measured values maintaining a suitable agreement. In addition, FCM clustering can accurately predict ship traffic congestion at four levels of severity.

The prediction scenario for this study included the traffic-intensive waters of the Yangtze estuary, but the

proposed method does not consider the complexity of the network formed by the navigating ships. In future work, we will explore complex network theory to study the complexity of the waterway navigation network and improve the generalization ability of the proposed prediction method.

It should be noted that the studied areas in this paper are somewhat regional in nature. Although the Yangtze River estuary is highly representative of the dense traffic flow waters, in order to make the corresponding methodological model applicable to all dense traffic flow waters, the subsequent research work should be extended to other watersheds based on the expansion of dense traffic flow waters. At the same time, this paper does not further deepen the research on the prediction and control of ship congestion before the occurrence of congestion based on the judgment of congestion risk level, and this research has extremely important theoretical and application values.

## Data Availability

The data used in this paper were obtained from the collection of AIS data of ships in the Yangtze estuary from the maritime regulatory authority website (<https://www.shipxy.com/>).

## Conflicts of Interest

The authors declare that there are no conflicts of interest regarding the publication of this paper.

## Acknowledgments

This research was funded by the National Key R&D Program of China (grant no. 2021YFB1600400) and the National Natural Science Foundation of China (grant no. 52101403).

## References

- [1] X. Xiao, H. Duan, and J. Wen, "A novel car-following inertia gray model and its application in forecasting short-term traffic flow," *Applied Mathematical Modelling*, vol. 87, pp. 546–570, 2020.
- [2] H. Yan, T. Zhang, Y. Qi, and D.-J. Yu, "Short-term traffic flow prediction based on a hybrid optimization algorithm," *Applied Mathematical Modelling*, vol. 102, pp. 385–404, 2022.
- [3] Z. Cui, R. Ke, Z. Pu, and Y. Wang, "Stacked bidirectional and unidirectional LSTM recurrent neural network for forecasting network-wide traffic state with missing values," *Transportation Research Part C: Emerging Technologies*, vol. 118, Article ID 102674, 2020.
- [4] G. Li, V. L. Knoop, and H. van Lint, "Multistep traffic forecasting by dynamic graph convolution: interpretations of real-time spatial correlations," *Transportation Research Part C: Emerging Technologies*, vol. 128, Article ID 103185, 2021.
- [5] H. Peng, "Spatial temporal incidence dynamic graph neural networks for traffic flow forecasting," *Information Sciences*, vol. 521, pp. 277–290, 2020.
- [6] L. Cai, Z. Zhang, J. Yang, Y. Yu, T. Zhou, and J. Qin, "A noise-immune Kalman filter for short-term traffic flow forecasting," *Physica A: Statistical Mechanics and Its Applications*, vol. 536, Article ID 122601, 2019.
- [7] K. Wang, C. Ma, Y. Qiao, X. Lu, W. Hao, and S. Dong, "A hybrid deep learning model with 1DCNN-LSTM-Attention networks for short-term traffic flow prediction," *Physica A*:

- Statistical Mechanics and Its Applications*, vol. 583, Article ID 126293, 2021.
- [8] W. Fang, W. Zhuo, J. Yan, Y. Song, D. Jiang, and T. Zhou, "Attention meets long short-term memory: a deep learning network for traffic flow forecasting," *Physica A: Statistical Mechanics and Its Applications*, vol. 587, Article ID 126485, 2022.
  - [9] Z. Shen, W. Wang, Q. Shen, S. Zhu, H. M. Fardoun, and J. Lou, "A novel learning method for multi-intersections aware traffic flow forecasting," *Neurocomputing*, vol. 398, pp. 477–484, 2020.
  - [10] H. Lu, Z. Ge, Y. Song, D. Jiang, T. Zhou, and J. Qin, "A temporal-aware LSTM enhanced by loss-switch mechanism for traffic flow forecasting," *Neurocomputing*, vol. 427, pp. 169–178, 2021.
  - [11] L. Li, L. Qin, X. Qu, J. Zhang, Y. Wang, and B. Ran, "Day-ahead traffic flow forecasting based on a deep belief network optimized by the multi-objective particle swarm algorithm," *Knowledge-Based Systems*, vol. 172, pp. 1–14, 2019.
  - [12] M. Zhang, F. Conti, H. Le Sourne et al., "A method for the direct assessment of ship collision damage and flooding risk in real conditions," *Ocean Engineering*, vol. 237, Article ID 109605, 2021.
  - [13] M. Zhang, J. Montewka, T. Manderbacka, P. Kujala, and S. Hirdaris, "A big data analytics method for the evaluation of ship-ship collision risk reflecting hydrometeorological conditions," *Reliability Engineering and System Safety*, vol. 213, Article ID 107674, 2021.
  - [14] M. Zhang, G. Taimuri, J. Zhang, and S. Hirdaris, "A deep learning method for the prediction of 6-DoF ship motions in real conditions," *Proceedings of the Institution of Mechanical Engineers - Part M: Journal of Engineering for the Maritime Environment*, Article ID 14750902231157852, 2023.
  - [15] D. Ma, X. B. Song, J. Zhu, and W. Ma, "Input data selection for daily traffic flow forecasting through contextual mining and intra-day pattern recognition," *Expert Systems with Applications*, vol. 176, Article ID 114902, 2021.
  - [16] A. I. Salamanis, A.-D. Lipitakis, G. A. Gravvanis, S. Kotsiantis, and D. Anagnostopoulos, "An adaptive cluster-based sparse autoregressive model for large-scale multi-step traffic forecasting," *Expert Systems with Applications*, vol. 180, Article ID 115093, 2021.
  - [17] A. Almeida, S. Brás, I. Oliveira, and S. Sargento, "Vehicular traffic flow prediction using deployed traffic counters in a city," *Future Generation Computer Systems*, vol. 128, pp. 429–442, 2022.
  - [18] K. Lee and W. Rhee, "DDP-GCN: multi-graph convolutional network for spatiotemporal traffic forecasting," *Transportation Research Part C: Emerging Technologies*, vol. 134, Article ID 103466, 2022.
  - [19] D. Pavlyuk, "Towards ensemble learning of traffic flows' spatiotemporal structure," *Transportation Research Procedia*, vol. 47, pp. 361–368, 2020.
  - [20] D. Pavlyuk, "Spatiotemporal cross-validation of urban traffic forecasting models," *Transportation Research Procedia*, vol. 52, pp. 179–186, 2021.
  - [21] M. A. C. Alves and R. L. F. Cordeiro, "Effective and unburdensome forecast of highway traffic flow with adaptive computing," *Knowledge-Based Systems*, vol. 212, Article ID 106603, 2021.
  - [22] F. Carpio, W. Bziuk, and A. Jukan, "Scaling migrations and replications of Virtual Network Functions based on network traffic forecasting," *Computer Networks*, vol. 203, Article ID 108582, 2021.
  - [23] K. Jin, J. Wi, E. Lee, S. Kang, S. Kim, and Y. Kim, "Traffic-BERT: pre-trained model with large-scale data for long-range traffic flow forecasting," *Expert Systems with Applications*, vol. 186, Article ID 115738, 2021.
  - [24] Y. Jin, X. Ge, L. Zhang, and J. Ren, "A two-stage algorithm for bi-objective logistics model of cash-in-transit vehicle routing problems with economic and environmental optimization based on real-time traffic data," *Journal of Industrial Information Integration*, vol. 26, Article ID 100273, 2022.
  - [25] G. Mei, J. Tu, L. Xiao, and F. Piccialli, "An efficient graph clustering algorithm by exploiting k-core decomposition and motifs," *Computers and Electrical Engineering*, vol. 96, Article ID 107564, 2021.
  - [26] Y. Guo, L. Yang, S. Hao, and J. Gao, "Dynamic identification of urban traffic congestion warning communities in heterogeneous networks," *Physica A: Statistical Mechanics and Its Applications*, vol. 522, pp. 98–111, 2019.
  - [27] Y. Gao, J. Li, Z. Xu, Z. Liu, X. Zhao, and J. Chen, "A novel image-based convolutional neural network approach for traffic congestion estimation," *Expert Systems with Applications*, vol. 180, Article ID 115037, 2021.
  - [28] P. Zhou, X. Wang, L. Du, and X. Li, "Clustering ensemble via structured hypergraph learning," *Information Fusion*, vol. 78, pp. 171–179, 2022.
  - [29] T. T. Nguyen, P. Krishnakumari, S. C. Calvert, H. L. Vu, and H. van Lint, "Feature extraction and clustering analysis of highway congestion," *Transportation Research Part C: Emerging Technologies*, vol. 100, pp. 238–258, 2019.
  - [30] G. Costa and R. Ortale, "Effective interrelation of Bayesian nonparametric document clustering and embedded-topic modeling," *Knowledge-Based Systems*, vol. 234, Article ID 107591, 2021.
  - [31] M. L. M. Peixoto, A. Maia, E. Mota et al., "A traffic data clustering framework based on fog computing for VANETs," *Vehicular Communications*, vol. 31, Article ID 100370, 2021.
  - [32] M. N. Peter and M. Pushpa Rani, "Smart city traffic based on traffic density using RSRU\_TM," *Materials Today: Proceedings*, vol. 56, pp. 3335–3342, 2022.
  - [33] A. Ali, M. A. Qureshi, M. Shiraz, and A. Shamim, "Mobile crowd sensing based dynamic traffic efficiency framework for urban traffic congestion control," *Sustainable Computing: Informatics and Systems*, vol. 32, Article ID 100608, 2021.
  - [34] J. A. Huertas, "Level of traffic stress-based classification: a clustering approach for Bogotá, Colombia," *Transportation Research Part D: Transport and Environment*, vol. 85, Article ID 102420, 2020.
  - [35] Y. Zhu and H. Tang, "Automatic damage detection and diagnosis for hydraulic structures using drones and artificial intelligence techniques," *Remote Sensing*, vol. 15, no. 3, p. 615, 2023.
  - [36] X. Wang, J. Li, and T. Zhang, "A machine-learning model for zonal ship flow prediction using AIS data: a case study in the south atlantic states region," *Journal of Marine Science and Engineering*, vol. 7, no. 12, p. 463, 2019.
  - [37] T. Xu and Q. Zhang, "Ship traffic flow prediction in wind farms water area based on spatiotemporal dependence,"

- Journal of Marine Science and Engineering*, vol. 10, no. 2, p. 295, 2022.
- [38] C. Zhao, A. Song, Y. Du, and B. Yang, "TrajGAT: a map-embedded graph attention network for real-time vehicle trajectory imputation of roadside perception," *Transportation Research Part C: Emerging Technologies*, vol. 142, Article ID 103787, 2022.
- [39] Y. Ji, L. Ni, C. Zhao, C. Lei, Y. Du, and W. Wang, "TriPField: a 3D potential field model and its applications to local path planning of autonomous vehicles," *IEEE Transactions on Intelligent Transportation Systems*, vol. 24, no. 3, pp. 3541–3554, 2023.
- [40] H. Itoh, "Method for prediction of ship traffic behaviour and encounter frequency," *Journal of Navigation*, vol. 75, no. 1, pp. 106–123, 2022.