Hindawi
Mobile Information Systems
Volume 2022, Article ID 7846630, 17 pages
https://doi.org/10.1155/2022/7846630

Research Article

GMM Clustering Based on WOA Optimization and Space-Time Coupled Urban Rail Traffic Flow Prediction by CEEMD-SE-BiGRU-AM

Qiong Jiang

East China Normal University, Statistics, Shanghai 200241, China

Correspondence should be addressed to Qiong Jiang; 10195000441@stu.ecnu.edu.cn

Received 26 May 2022; Revised 17 June 2022; Accepted 29 June 2022; Published 21 August 2022

Academic Editor: Praveen Kumar Reddy Maddikunta

Copyright © 2022 Qiong Jiang. 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 the short-term passenger flow of urban subways is very important for urban subway stations to formulate passenger flow organization and evacuation plans effectively and rationally plan passenger travel routes. This paper establishes a novel framework to predict the hourly inbound and outbound passenger flow of subway stations based on WOA-GMM station classifiers, CEEMD-SE noise reduction, and BiGRU optimized by attention. Firstly, this paper classifies subway stations using the improved Gaussian mixture model (GMM) with Whale Optimization Algorithm (WOA) to realize the feature extraction of different types of subway stations. Secondly, this paper uses the Complementary Ensemble Empirical Mode Decomposition (CEEMD) to decompose the noise reduction of each station's hourly inbound and outbound passenger flow. It combines the empirical modal components by calculating the sample entropy (SE), which makes the time series stable and reduces the time cost of forecasting. Finally, a Bi-directional Gated Recurrent Unit (BiGRU) model improved by attention mechanism (AM) is established for each station's inbound and outbound passenger flow. The prediction model established in this paper is verified by subway passenger flow data in Shanghai, China, within 24 days. Finally, it is concluded that the model can predict the passenger flow of subway stations. Compared with the traditional Backpropagation Neural Network (BP), the Long Short-Term Memory (LSTM), and the normal BiGRU model, the model proposed in this paper has an average reduction of 65.90%, 64.54%, and 49.06% in Mean Absolute Percentage Error (MAPE) in the prediction of the hourly inbound and outbound passenger flow of each type of station, respectively.

1. Introduction

The ability to accurately forecast the short-term passenger flow of city subways is critical to inhabitants' quality of life. With the rapid advancement of global urbanization, many urban rail transit systems have achieved explosive growth [1]. The number of newly opened transportation systems has surged during the past decade. In 2019 alone, the total number of passengers transported by subway and light rail globally reached 70.794 billion. By the end of 2020, 538 cities in 77 nations and regions have inaugurated urban rail transportation, with a total operating mileage of over 33,346 km and over 34,220 stations [2]. The rapid expansion of the rail transit network has also caused passenger flow organization and train scheduling problems. Excessive passenger flow on the subway and excessive inbound and outbound passenger flow of the station may cause safety problems such as stampede incidents and are not conducive to passengers reaching their destination according to the target time. However, due to the different external characteristics such as spatial layout and commercial effect of each station, excessive attention to these characteristics will complicate the prediction of passenger flow. Therefore, making the best use of the time and space features of subway stations and accurately anticipating the passenger flow of rail transit stations is a pressing issue that must be addressed.

2. Literature Review

Considering that the geographical location of different stations and the surrounding economic development conditions are different, the passenger flow between adjacent

stations greatly correlates. Therefore, it is necessary to perform cluster analysis on stations from time and space perspectives. Zhou et al. [2] thoroughly considered each station's topology structure and passenger flow information, which used complex network evaluation and the k-means method to cluster the stations. It proposed a method to balance and control passenger flow. However, in practical applications, the k-means algorithm [3], Affinity Propagation (AP) algorithm [4], Spectral Clustering (SC) algorithm [5], and other methods are more suitable for processing spherical data, and the processing effect for data with irregular shapes is not satisfactory. In contrast, the GMM algorithm assumes that a combination of multiple Gaussian distributions forms the data points, and there is no restriction on the shape of the data. Multiple Gaussian distribution functions can be used to fit the shape of the data points to the greatest extent. However, since the GMM algorithm is based on the Expectation-Maximum (EM) algorithm, it is simple to attain a local optimum and it is extremely sensitive to the initial value. For such a problem, Guo et al. [7] presented utilizing the Particle Swarm Optimization (PSO) optimized GMM as a correction approach for the parameters of the electromagnetic transient model. It also verifies the effectiveness of optimizing the initial parameters of the GMM algorithm through the heuristic algorithm. WOA [8] is a new variant of the heuristic optimization algorithm that outperforms the PSO and Grey Wolf Optimizer (GWO) algorithm [9, 10] in terms of performance. Abbas et al. [11] used the WOA to extract essential features in the breast cancer dataset, which greatly improved the prediction accuracy. Gadekallu et al. [7] used the combination of Principal Component Analysis (PCA) and WOA to accurately extract the important features of tomato disease data and used a deep neural network to identify the disease types. Among them, using the WOA improves the model's accuracy by 4% on the test set. Yan et al. [13] used WOA to improve the extreme gradient boosting (XGB) model and found that the optimized model predicted daily reference evapotranspiration considerably better than the basic FAO-56 PM model. To overcome the problems, the parameters of GMM based on the EM algorithm attain the local optimum and rely too much on the selection of the initial value. The WOA is used to optimize the initial parameters of GMM in this study. At the same time, it also solves the problem of low fitting accuracy and robustness of the k-means algorithm.

In addition, changes in the external environment will have a particular impact on passenger flow data. The data noise created in this manner is unpredictable, and too much random fluctuation will diminish the predictability of the results. Therefore, it is necessary to decompose and denoise the original data. In recent years, researchers have developed various processing methods to identify the characteristics of nonlinear sequences to maximize the universality of the prediction results and reduce the impact of external environmental factors on the results. These include wavelet transform (WT) [14], Singular Spectrum Analysis (SSA) [15], Empirical Mode Decomposition (EMD) [16], Ensemble Empirical Mode Decomposition (EEMD) [17], etc. Some researchers have also proved the higher accuracy of the hybrid prediction model with decomposition and noise reduction processing. Zhu et al. [18] used the Autoregressive Integrated Moving Average model (ARIMA) to predict the passenger flow and combined WT to decompose and denoise the data, which achieved higher accuracy than the ARIMA model alone. However, the wavelet transform needs to specify the wavelet basis function artificially, and the decomposition function lacks directionality [19]. In contrast, the empirical mode decomposition method is more adaptive. Zhao et al. [20] used the empirical mode decomposition method to convert the time series into a series of eigenmode functions and residuals. They used a long and short-term memory neural network to estimate subway passenger flow, and the results were more accurate than previous predictions. However, EMD is prone to modal aliasing, and CEEMD [21] solves this problem by introducing complementary noise. Therefore, this paper uses CEEMD with high computational efficiency and high reconstruction accuracy of the original signal to denoise the original data.

Parametric and nonparametric models are the two types of traditional short-term passenger flow prediction models. Therefore, the traditional short-term passenger flow prediction models are mainly divided into parametric and nonparametric models. One of the representative parametric models is ARIMA [22]. ARIMA considers the periodic characteristics of time, and the prediction method is direct and does not require too much preprocessing. However, ARIMA ignores the influence of randomness on the overall forecasting results, and the forecasting accuracy of time series with strong randomness is not high. Compared with parametric models, nonparametric models are more flexible and can deal with more complex situations, including Support Vector Machine (SVM) [23], LSTM [24], and Gated Recurrent Unit (GRU) [25]. The WT-SVM model proposed by Sun et al. [26] complemented the advantages of WT and SVM to predict the subway passenger flow. Therefore, SVM still has a big limitation in kernel selection. Yang et al. [27] proposed the Wave-LSTM model, which used the method of controlling variables to determine the model parameters and verified the effectiveness and prediction accuracy of the hybrid model by comparing it with traditional models. In contrast, Deep Neural Networks (DNNs) [28] have a greater advantage in sequence learning, so the improved LSTM model captures the long-term time dependence of sequences mainly through gated recursive units. [29], used LSTM and GRU to forecast the short-term passenger flow of urban rail transit and found that 5 minutes was the best temporal granularity for both models to predict short-term passenger flow. Under this time granularity, the overall performance of GRU is better than that of LSTM. However, it is often difficult for recurrent neural network models to capture local features accurately. In this regard, Wu and Tan [30] used a combination of CNN and LSTM to predict traffic flow, but this method would increase the difficulty of training and be less efficient for long-term sequence prediction. The AM [31] method extracts valuable features globally. Reference [32] improved the CNN model mainly used for local feature perception by using the features with a strong global

perception of AM and effectively extracting global features. Therefore, this paper adopts the BiGRU model improved by AM to make the final passenger flow prediction, which greatly improves information utilization and accuracy.

3. Paper Contribution

This paper establishes the WOA-GMM-CEEMD-SE-BiGRU-AM model to predict the short-term passenger flow of urban rail transit and compares it with other models to verify its accuracy and effectiveness. The contributions of this paper are as follows:

(1) Since different stations have different spatiotemporal characteristics, this paper employs GMM to group stations with comparable characteristics. However, the EM algorithm is the core part of the GMM, and its clustering effect largely depends on the initial value of the EM algorithm. Therefore, the initial parameters of the GMM are optimized by WOA. The stations are finally classified according to their spatiotemporal characteristics.

(2) The passenger flow data of rail transit fluctuates greatly, and the regularity is poor. Therefore, this paper decomposes and denoises the passenger flow data of each type of station through CEEMD and uses a more regular component as the target variable of prediction to increase the prediction accuracy. However, considering the time cost of prediction, similar variables are combined by calculating SE.

(3) In order to predict the passenger flow of rail transit, this paper firstly establishes the BiGRU, but to extract the local features of the time series more accurately, AM is used to improve the BiGRU. Finally, compared with other models, it can be verified that the model established in this paper has high accuracy.

The rest of the chapters are arranged as follows: Chapter 4 introduces the model, 4.1 introduces the processing flow of the station clustering algorithm, 4.2 introduces the CEEMD-SE method for decomposing noise reduction, and 4.3 introduces the building process of the BiGRU-AM model. The third chapter carries on the calculation of the example and the validation of the model, and the fourth chapter concludes this paper 4.

4. Model Introduction

4.1. Clustering Model Based on WOA-GMM

4.1.1. Whale Optimization Algorithm. Since GMM is classified based on the EM algorithm, this model can easily achieve local convergence [33]. Therefore, this paper uses WOA to solve this problem. The whale optimization algorithm is a metaheuristic algorithm based on swarm intelligence. By simulating the predation behavior of humpback whales, the random learning strategy is used to achieve global shrinkage, and the convergence rate is fast, which is suitable for solving global optimization problems. The predation process of the humpback whale is as follows:

Step1: Determine the prey location and select the hunting method.

(1) Shrinkage and encirclement mechanism:

When the humpback whale surrounds the prey, it will choose to swim toward the humpback whale with the best position, identify the position of the prey, and surround it. Such hunting is realized by reducing the value of a.

$$\begin{aligned} D' &= \left|CX_p(t) - X(t)\right| \\ X(t+1) &= X_p(t) - AD', r < 0.5 \\ A &= 2ar - a, C = 2r, a = 2 - 2\left(t/t_{\max}\right), |A| < 1 \end{aligned} \tag{1}$$

where $X(t)$ is the position of the t-th iteration of the humpback whale. $X_p(t)$ is the optimal position currently searched; t is the current iteration number; D is the distance from the humpback whale to the prey. $t_{\max}$ is the maximum number of iterations; a is a value linearly decreasing from 2 to 0; r is a random vector in the range 0 to 1.

(2) Spiral spit mechanism

Calculate the distance between itself and the optimal individual in the current population, and spit out bubbles while swimming in a spiral.

$$\begin{aligned} D'' &= \left|X_p(t) - X(t)\right| \\ X(t+1) &= D''e^{bl}\cos(2\pi l) + X_p(t), r \geq 0.5 \end{aligned} \tag{2}$$

b is the logarithmic spiral shape constant; l is a random number in [−1, 1].

Step 2: Search for prey.

When humpback whales search for prey, they will randomly swim according to the position of the same species to achieve a global search.

$$\begin{aligned} D &= \left|CX_{\text{rand}}(t) - X(t)\right| \\ X(t+1) &= X_{\text{rand}}(t) - A\ D, |A| \geq 1. \end{aligned} \tag{3}$$

$X_{\text{rand}}(t)$ is a random position.

4.1.2. Gaussian Mixture Model. The Gaussian mixture model (GMM) [34] assumes that all data points come from different distributions, and by finding a mixture of multi-dimensional Gaussian probability distributions, it eventually generates a distribution that can satisfy any shape. Data points are judged to be from the same distribution in one group. Compared with the k-means algorithm to calculate the Euclidean distance between data points, the GMM algorithm can measure clusters' distribution probability, which makes the fitting accuracy higher. In addition, GMM has lower time complexity than traditional clustering algorithms and is more in line with the central limit theorem under the condition of large samples.

In the multidimensional case, the objective function of GMM is

$$L = \log(p(X)) = \sum_{i=1}^{N} \log\left(\sum_{k=1}^{K} p\left(x_i|\mu_k, \sum_k\right) p\left(\mu_k, \sum_k\right)\right)$$
$$p\left(x_i|\mu_k, \sum_k\right) = \frac{1}{(2\pi)^{n/2}\sum_k 1/2} \exp\left(-\frac{1}{2}(x_i - \mu_k)^T \sum_k^{-1}(x_i - \mu_k)\right) \sim N\left(\mu_k, \sum_k\right). \tag{4}$$

$p(X)$ is the product of the likelihood functions of all Gaussian distributions. x_i is the i-th sample (dimension p); $p(x_i|\mu_k, \sum_k)$ is the conditional distribution under the specified parameters, which obeys the expectation μ_k (dimension p), and the covariance matrix is $\sum_k$ (dimension $p \times p$) which is usually distributed; L is the log-likelihood function; K is the number of clusters; N is the number of samples.

The calculation steps of GMM are as follows:

Step 1: Set the number of sample clusters to K, and the samples obey a Gaussian distribution:

$$p(x_i) = \sum_{i=1}^{K} \alpha_i p\left(x_i|\mu_i, \sum_i\right), \tag{5}$$

where $p(x_i)$ is a multivariate Gaussian distribution; α_i is the probability of the i-th mixture component.

Step 2: Calculate the probability of x_j:

$$\gamma_{ij} = \frac{\alpha_i p\left(x_j|\mu_i, \Sigma_i\right)}{\sum_{l=1}^{k} \alpha_l p\left(x_j|\mu_l, \Sigma_l\right)}, \tag{6}$$

where γ_{ij} represents the probability that the j-th sample belongs to the i-th class.

Step 3: Compute the maximum likelihood estimate for each cluster to update the model parameters, including the expectation, variance, and probability of each data belonging to class i of Gaussian mixture distribution.

$$\mu_i' = \frac{\sum_{j=1}^{m} \gamma_{ij} x_j}{\sum_{j=1}^{m} \gamma_{ij}},$$
$$\Sigma_i' = \frac{\sum_{j=1}^{m} \gamma_{ij}\left(x_j - \mu_i'\right)\left(x_j - \mu_i'\right)^T}{\sum_{j=1}^{m} \gamma_{ij}}, \tag{7}$$
$$\alpha_i' = \frac{\sum_{j=1}^{m} \gamma_{ij}}{m}.$$

Step 4: Repeat steps 2 and 3 until the objective function converges.

Step 5: The j-th template is divided into the corresponding category according to the rule of taking the maximum value of γ_{ij}.

4.1.3. WOA-GMM. GMM is established based on the EM algorithm. The EM algorithm is an expectation-maximization algorithm, but it is easy to fall into a local optimum, and the result strongly depends on the initial parameters. Therefore, this paper uses the WOA to initialize the parameters of the EM algorithm and then establishes the GMM.

Step 1: Determine the number of clusters.

Step 2: Construct a Gaussian mixture model objective function.

Step 3: WOA is used to solve the maximum likelihood estimation of the Gaussian mixture model constructed in Step 2.

Step 4: The optimization parameters obtained in Step 3 are used as the initial parameters of the EM algorithm.

Step 5: Iterate until optimal.

The model flow is shown in Figure 1. $t1$ and $itermax1$ are the current and maximum iterations of the WOA. $t2$ and $itermax2$ are the current number of iterations and the maximum number of iterations of the GMM algorithm.

4.2. Data Noise Reduction Model Based on CEEMD-SE

4.2.1. Complementary Ensemble Empirical Mode Decomposition. EMD is a decomposition noise reduction method used to smooth and linearize data. However, due to the discontinuity of the Intrinsic Mode Function (IMF) components, the method will produce modal aliasing during the decomposition process, and the improved EEMD model solves this problem. At the same time, due to the white noise introduced by EEMD, its residual parts will affect the original data to a large extent [35]. As a result, this work uses the CEEMD model, including complimentary noise to improve the computational efficiency and signal reconstruction accuracy. The calculation process of the CEEMD model is as follows:

Step 1: A set of positive and negative Gaussian white noise sequences $\epsilon_i^+(t)$ and $\epsilon_i^-(t)$ is added, which obtains a new passenger flow sequence based on the original passenger flow sequence $x(t)$.

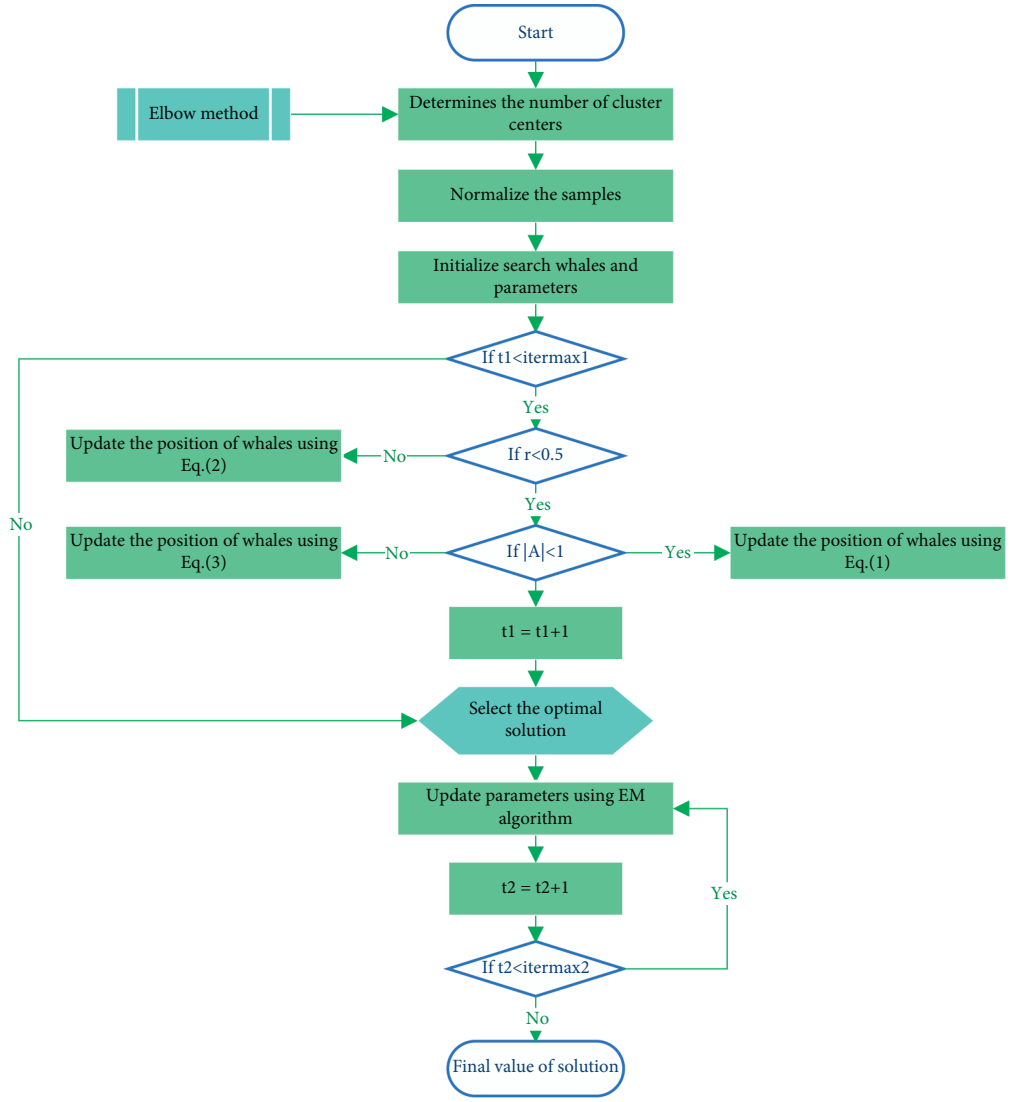

Figure 1: Flowchart of station clustering model.

$$x_i^+(t) = x(t) + \epsilon_i^+(t),$$
$$x_i^-(t) = x(t) + \epsilon_i^-(t). \tag{8}$$

$x_i^+(t)$ and $x_i^-(t)$ represent the result of adding positive and negative white noise to the original passenger flow sequence.

Step 2: Perform empirical mode decomposition on the sequences in Step 1 separately to obtain n-1 IMF components and one residual component.

$$x_i^+(t) = \sum_{j=1}^{n} c_{ij}^+(t),$$
$$x_i^-(t) = \sum_{j=1}^{n} c_{ij}^-(t), \tag{9}$$

$c_{ij}^+(t)$ and $c_{ij}^-(t)$ are, respectively, the j-th modal component obtained by empirical modal decomposition after adding Gaussian white noise for the i-th time; when $j = n$, it is the residual component.

Step 3: Replace the Gaussian white noise sequence and repeat Step 1 and Step 2 until N groups of eigenmode components are obtained.

Step 4: Calculate the average value of N groups of IMF components in Step 3 to obtain the final decomposition result.

$$c_i(t) = \frac{1}{2N} \sum_{j=1}^{N} \left(c_{ij}^+(t) + c_{ij}^-(t) \right). \tag{10}$$

$c_i(t)$ is the final i-th decomposition result.

4.2.2. Sample Entropy. After CEEMD decomposes the time series data, multiple IMF components can be obtained. SE [36] can measure the complexity of different IMF components, which improves the Approximate Entropy (AE), reduces the dependence on the length of the time series in the calculation process, and reduces the impact of the loss of original data on the complexity calculation. The sample entropy calculation steps are as follows:

Step1: Let a particular IMF component obtained by CEEMD decomposition be $X = (x(1), x(2), \ldots, x(N))$. Take all consecutive u-dimensional vectors $X_i = (x(i), x(i+1), \ldots, x(i+u-1)), i = 1, 2, \ldots, N-u+1$ in X.

Step 2: Define the distance between X_i and X_j as

$$d\left(X_i, X_j\right) = \max_{k\in(0,u-1)} |x(i+k) - x(j+k)|. \tag{11}$$

Step 3: Given a similarity tolerance r ($r > 0$), count the number of $d(X_i, X_j) < r$ for each vector X_i, denoted by G_i, and calculate its ratio to N-u expressed as $G_i^u(r)$.

$$G_i^u(r) = \frac{1}{N-u} G_i. \tag{12}$$

Step 4: Calculate the mean of all $G_i^u(r)$ expressed as $G^u(r)$.

$$G^u(r) = \frac{1}{N-u+1} \sum_{i=1}^{N-u+1} G_i^u(r). \tag{13}$$

Step 5: Increase the u dimension to $u+1$ dimension, repeat Step1~Step4, and get $G^{u+1}(r)$.

Step 6: Calculate SE.

$$\text{SampEn}(u, r) = \lim_{N\longrightarrow\infty} \left(-\ln\left[\frac{G^{u+1}(r)}{G^u(r)}\right]\right). \tag{14}$$

However, when SE is applied to actual data, N is always finite, thus giving an estimate of sample entropy as

$$\text{SampEn}(u, r, N) = -\ln\left[\frac{G^{u+1}(r)}{G^u(r)}\right]. \tag{15}$$

4.3. Predictive Model (BiGRU-AM)

4.3.1. Bidirectional Gated Recurrent Unit. GRU is a gated recurrent neural network model that can solve problems such as long-term memory and gradient in backpropagation and is suitable for processing time series data. Compared with LSTM, the GRU model replaces the input gate and forget gate functions with only the update gate. The formula of the hidden layer of the model is as follows:

$$\begin{aligned} z_t &= \text{sigmoid}\left(w_z\left[h_{t-1}, x_t\right] + b_z\right), \\ r_t &= \text{sigmoid}\left(w_r\left[h_{t-1}, x_t\right] + b_r\right), \\ \widetilde{h}_t &= \tanh\left(w_h\left[r_t h_{t-1}, x_t\right] + b_h\right), \\ h_t &= \left(1 - z_t\right)h_{t-1} + z_t\widetilde{h}_t, \end{aligned} \tag{16}$$

where z_t is the update gate. r_t is the reset gate; $\widetilde{h}_t$ is the newly generated information after processing the historical information through the update gate. w_z, w_r, and w_h are the weight parameters of the GRU network; b_z, b_r, b_h are the thresholds of the GRU network.

The GRU structure diagram is shown in Figure 2.

The hidden state of the GRU model layer only considers the impact of historical information on the current input, and the impact of future information is equally important, so this paper uses BiGRU to solve this problem. BiGRU consists of two unidirectional and opposite GRUs. The model integrates the information before and after the current state, and the final output is the weighted summation of the output results of the forward GRU and the backward GRU and is linearly superimposed with the threshold. The main calculation formula of BiGRU is

$$\begin{aligned} h_f^{(t)} &= GRU\left(h^{(t-1)}, x^{(t)}\right), \\ h_b^{(t)} &= GRU\left(h^{(t+1)}, x^{(t)}\right), \\ h^{(t)} &= \alpha h_f^{(t)} + \beta h_b^{(t)} + \gamma, \end{aligned} \tag{17}$$

where $h_f^{(t)}$ and $h_b^{(t)}$ are the forward and backward GRU output vectors for the current state, and β are the weight parameters of the forward and backward GRU; γ is the linear threshold. The BiGRU network structure is shown in Figure 3.

4.3.2. Attention Mechanism. Although BiGRU improves the efficiency of information utilization to a certain extent, it still has the problem that local features cannot be accurately extracted when dealing with long-term sequence data. Therefore, when building the model, this paper adds an attention layer to the hidden layer to solve this problem. The attention layer further extracts the key information of the time series by calculating the weight of the output result in the BiGRU layer, thereby reducing the length of the output result and improving the robustness of the model:

$$\begin{aligned} v_i^{(t)} &= u_s \tanh\left(w h_i^{(t)} + b\right), \\ \alpha_i^{(t)} &= \frac{e^{v_i^{(t)}}}{\sum_{i=1}^{n} e^{v_i^{(t)}}}, \\ c^{(t)} &= \sum_{i=1}^{n} \alpha_i^{(t)} h_i^{(t)}, \end{aligned} \tag{18}$$

$v_i^{(t)}$ is the unnormalized attention score. The higher the attention score, the better the match between the input vector and the target vector; u_s is the randomly initialized attention matrix; w is the weight coefficient; b is the bias coefficient; $c^{(t)}$ is the attention vector; $h_i^{(t)}$ is the value represented by the i-th output result at the t-th time.

5. Data Analysis

5.1. Data Sources and Descriptions. This paper selects Automatic Fare Collection (AFC) data of Shanghai Metro Lines 1, 11, 2, 8, 9, 12, and 16 from April 1, 2015, to April 24, 2015, to test the established subway station passenger flow prediction model. The dataset contains timestamps, station names, and passenger arrival and departure records.

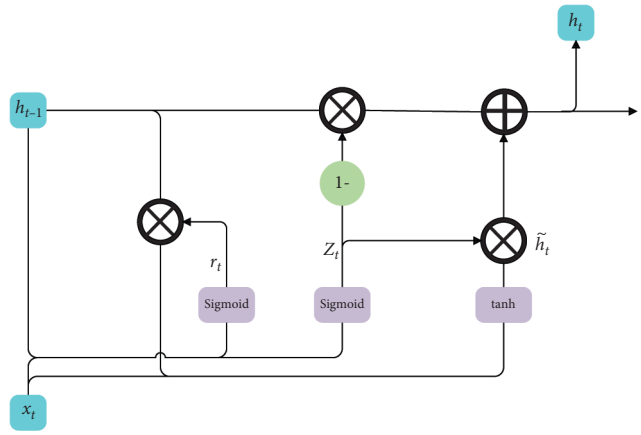

Figure 2: GRU structure diagram.

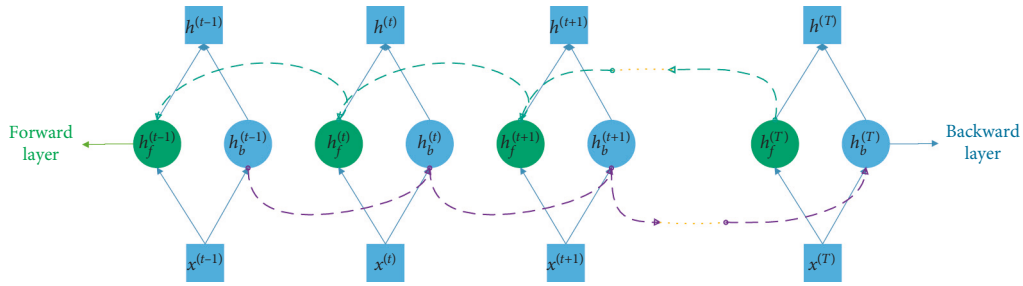

Figure 3: BiGRU network structure diagram.

5.2. Station Clustering. Considering the large differences in passenger flow data in different time periods, the original data needs to be standardized first. Second, the optimal number of clusters needs to be determined. Akaike Information Criterion (AIC) and Bayesian Information Criterion (BIC) are always used for trading off the complexity of the clustering model and the goodness of fitting the data:

$$AIC = -2\ln(L) + 2k, \tag{19}$$

$$BIC = -2\ln(L) + \ln(N)K. \tag{20}$$

L is the log-likelihood function in equation (4). K is the number of clusters; N is the sample size.

The second term in (19) and (20) is the penalty term. It can be seen that the penalty term of BIC considers the sample size, and the overall value is larger than that of AIC. Therefore, BIC can more effectively balance the model accuracy and complexity when the number of samples is too large, preventing the model from becoming too complex. Since 162 stations need to be clustered and due to the large sample size, only BIC is used to determine the optimal number of classifications. The smaller the BIC, the better the model classification effect. Set the number of clusters to a range from 1 to 20, and take the number of clusters corresponding to the minimum BIC value. From Figure 4, it can be determined that the optimal number of clusters is 5.

Secondly, use WOA to find GMM parameters for the aggregated passenger flow data of 5 types of stations: the number of humpback whales is 80, and the maximum number of iterations is 500. Finally, the optimal initial weights of each Gaussian distribution are 0.09, 0.35, 0.31, 0.09, and 0.16.

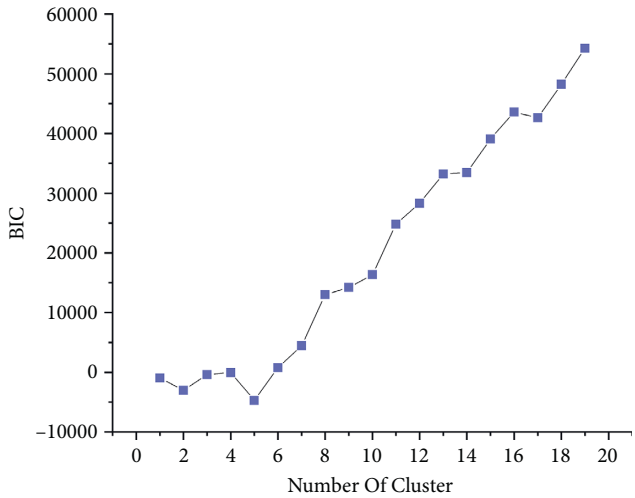

Figure 4: Bayesian Information Criterion to determine the optimal number of clusters.

Table 1: WOA-GMM vs. GMM.

Model	AIC	BIC
WOA-GMM	−16764.0204	−788.7970
GMM	−16562.1785	−586.9550

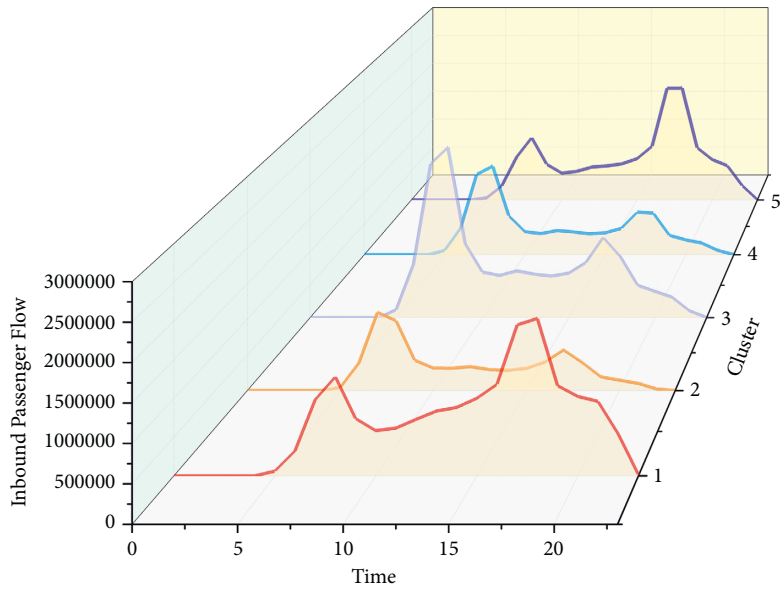

Figure 5: The inbound volume of 5 clusters.

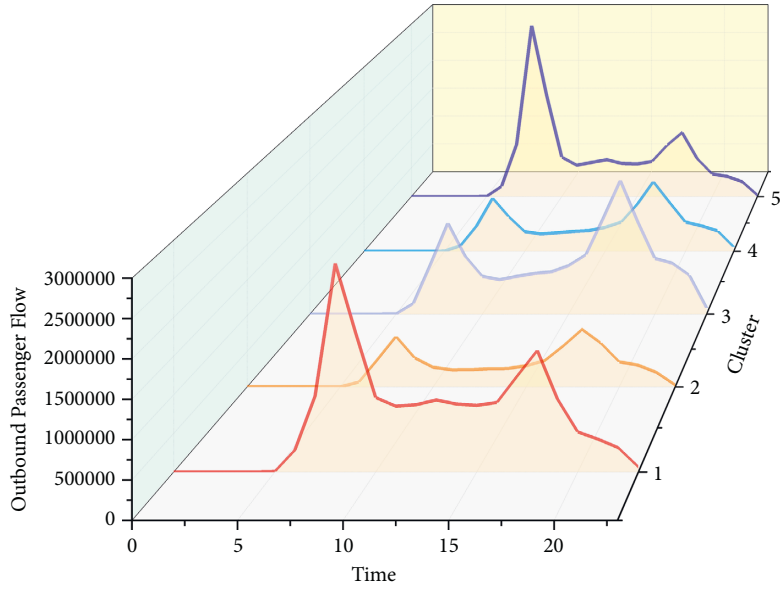

Figure 6: The outbound volume of 5 clusters.

To illustrate the advantages and efficiency of the WOA, we still use AIC and BIC as evaluation metrics to compare the GMM and WOA-GMM. As shown in Table 1, the model optimized by the WOA is smaller in both AIC and BIC indicators, which proves its excellent improvement effect.

Figures 5 and 6, respectively, represent the 24-hour cumulative passenger flow changes in 5 clusters within 24 days. The typical characteristics of these clusters of stations are shown in Table 2.

5.3. Decomposition and Noise Reduction of Station Passenger Flow Sequence. Next, this paper takes the inbound passenger flow of the first cluster of the station as an example to illustrate the process of data decomposition and noise reduction.

Using CEEMD, the empirical mode decomposition process of adding complementary white noise is carried out step by step for the daily hourly inbound and outbound passenger flow data of five clusters of subway stations. Figure 7 represents the decomposition of the corresponding data of the first cluster of subway station inbound passenger flow into 10 IMF components, the last of which represents the residual.

As shown in Figure 7, many of the components have similar complexities. IMF1~IMF5 have high complexity, IM6~IMF7 have lower complexity and a particular trend, and IMF8~R have obvious trends. However, conclusions based on

Table 2: Typical features of stations.

	Number of stations	Passenger flow characteristics	Representative stations	Spatial characteristics of the station
Cluster1	15	The number of such stations is small, but the passenger flow is large. The morning and evening peaks of passenger flow are obvious, and the inbound volume in the evening peak is higher than in the morning peak. The outbound volume in the morning peak reaches an obvious peak, and the inbound volume is still large after 20 : 00.	Qibao Station, Shanghai Railway Station, Zhongshan Park Station, People's Square Station, etc.	Most of these stations are subway transfer stations, railway stations, airport transportation hubs, or located in the city center, which belong to the city's core business district. After the evening peak, the passenger flow is still relatively large, which further proves that the commercial development around these stations is relatively good, and more people work overtime at night and participate in entertainment activities.
Cluster 2	57	The number of such sites is large, and the traffic is low. Compared with other categories, the morning and evening peaks of cluster 2 are not very prominent, and the overall passenger flow is relatively low.	Wild Animal Park Station, Donglu Road Station, Zhongchun Road Station, Lingang Avenue Station, etc.	Such stations are mostly located in underdeveloped areas, where buildings are mostly nonresidential, nonoffice, or in scenic spots.
Cluster 3	50	The passenger flow of such stations has apparent peaks in the morning and evening. The number of inbound volumes in the morning peak is the highest among several types of stations, and the number of outbound stations in the evening peak is also larger.	Sanlin Station, Shanghai Indoor Stadium Station, Century Park Station, Beixinjing Station, etc.	The areas where such stations are located are primarily residential, and some are located in the outer suburbs, with many houses and relatively low house prices. Therefore, many citizens go to work during the morning rush hour and leave work during the evening rush hour.
Cluster 4	14	The overall passenger flow of such stations is low, and the inbound volume in the morning peak is the most obvious.	Shanghainan Railway Station, Jiuting Station, Songjiang University Town Station, Shendu Highway Station, etc.	This station is also surrounded by residential areas, similar to the third cluster, but the overall passenger flow is low.
Cluster 5	26	The morning peak departure volume and evening peak arrival volume of such stations are extremely obvious.	Dongchang Road Station, China Art Museum Station, Hechuan Road Station, Shangcheng Road Station, etc.	There are many R&D and technology buildings around such stations and rich industrial parks.

observation alone are not accurate, so this paper uses sample entropy (SE) to determine the complexity of these components.

Figure 8 represents the changing trends of inbound passenger flow of the first cluster of the station, respectively. Next, to reduce the training time and improve the model prediction efficiency, this paper combines the IMF components of the inbound passenger flow of various types of stations into the combined components in Table 3 according to the IMF components with similar complexity as shown in Figure 8.

The trend of the combined components over time is shown in Figure 9. In this paper, the recombination components Comp1 and Comp2 with high frequency and randomness are regarded as high-frequency components; the recombination components Comp3 with low frequency and certain randomness are regarded as low-frequency components; the recombination components Comp4 with the lowest frequency are regarded as trend components. The assignment results are shown in Table 3.

5.4. Passenger Flow Forecasting Process. This paper optimizes BiGRU based on the attention mechanism, implements it in *Python*, and uses the TensorFlow framework to build the model. The experimental hardware device is configured as an 11th Gen Intel(R) Core(TM) i7-1165G7 with 16G memory. According to the prediction module in the model process, the model parameters are first initialized. The number of nodes in the input layer is set to three according to the number of auxiliary prediction variables (date, hour, subway station type), and the number of hidden layers of the BiGRU-AM network is selected to be two. Adam is used to continuously adjust the hyperparameters of the BiGRU-AM model in the iterative process until the error value is small.

In the process of inbound passenger flow prediction for the first station cluster, the Root Mean Square Error (RMSE) is used to evaluate the prediction accuracy, and its expression is shown in equation (21). The smaller the

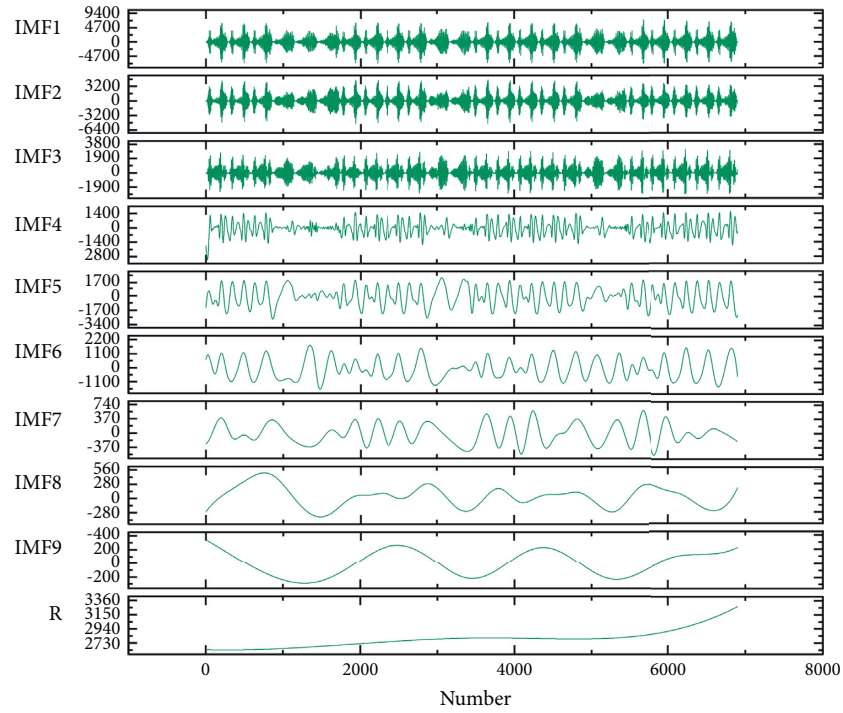

Figure 7: The first cluster of station inbound passenger flow IMF.

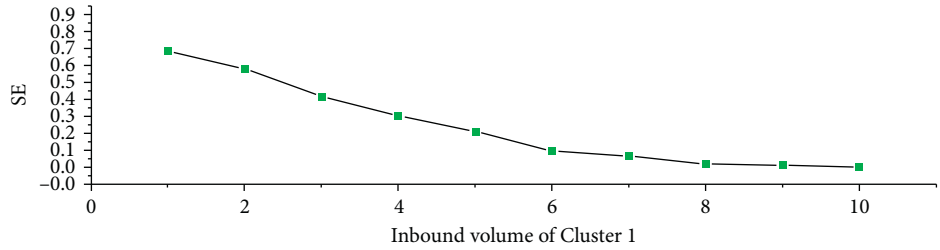

Figure 8: SE variation trend of inbound IMF for the first cluster of the station.

Table 3: Grouping of IMF components of inbound passenger flow at the first cluster of stations.

Components	IMF	Frequency and randomness
Component 1	IMF1, IMF2	High-frequency component
Component 2	IMF3, IMF4, IMF5	High-frequency component
Component 3	IMF6, IMF7	Low-frequency component
Component 4	IMF8, IMF9, R	Trend component

RMSE, the higher the prediction accuracy. Figure 10 shows the effect of setting the number of neurons in the two layers in BiGRU on the prediction accuracy of each combined component. It can be seen that when the number of neurons in the two layers is 13 and 11, respectively, the overall predicted RMSE reaches the minimum. The same method is used for the number of neurons in the BiGRU layer of other clusters of inbound and outbound passenger flow prediction models. Next, choose the optimal learning rate with the optimal number of neurons already determined. To simplify the model, a unified learning rate is determined separately in the prediction process of inbound passenger flow and outbound passenger flow. Taking the inbound passenger flow as an example, Figure 11 shows the variation of the inbound passenger flow prediction accuracy with the learning rate for each station cluster. It can be seen that the RMSE is the smallest when the learning rate is 0.006. Similarly, it can be found that the optimal learning rate for inbound passenger flow prediction should also be set to 0.006.

This paper sets the step size to 20 and the batch size to 110. According to experience, the number of neurons in the attention layer is set to 64. In this process, the optimal number of iterations for inbound and outbound passenger flow and the number of neurons in the two GRU layers in BiGRU are shown in Tables 4 and 5, respectively. Among them, the number of iterations is the average number of iterative predictions for each IMF combination, and the ratio of the training set and test set is 9 : 1.

In order to verify the validity of the BiGRU-AM model in the prediction of subway passenger flow, experiments are

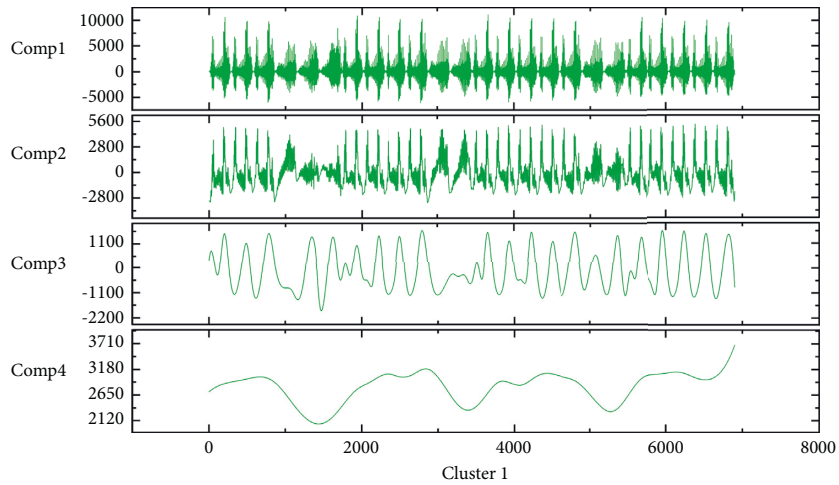

Figure 9: The trend of the combined components over time through CEEMD-SE.

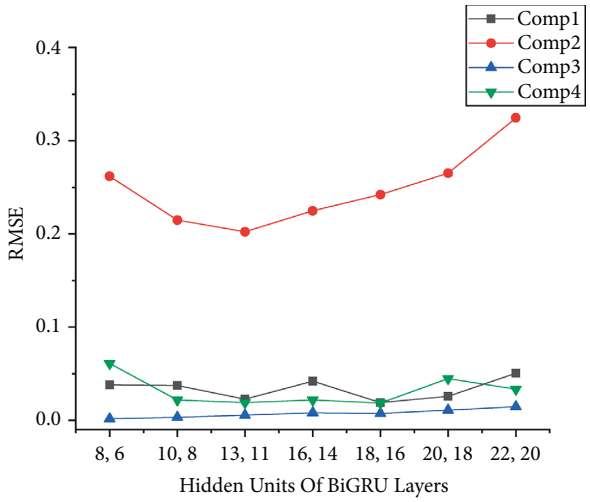

Figure 10: Hidden units determination.

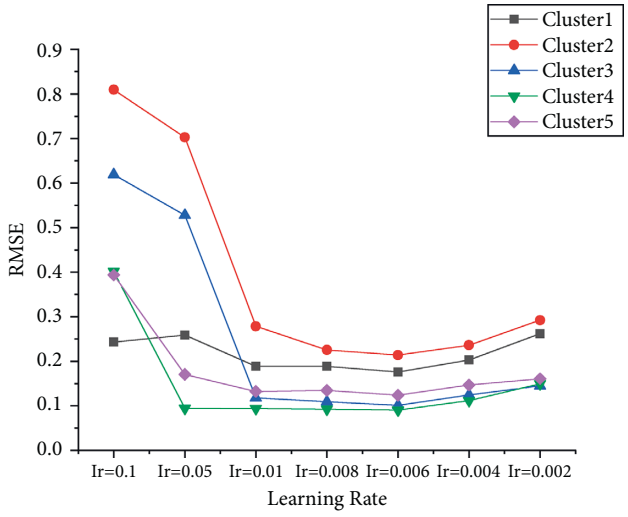

Figure 11: Learning rate determination.

carried out based on the above data. The prediction results of the combined components prediction results are shown in Figure 12. Figure 12 shows that the model's prediction results for Comp1~Comp3 are close to the true value, while the prediction for Comp4 is slightly different from the true value. It further illustrates the advantages of decomposing the passenger flow series into more trending components. In later chapters, we will further evaluate the model with more accurate metrics.

5.5. Predictive Model Comparative Analysis. In order to further illustrate the effectiveness of the overall model, after classifying the stations, BPNN, LSTM, and BiGRU are first selected for comparison. Secondly, the BiGRU model and the BiGRU-AM model are selected to compare with the CEEMD-BiGRU-AM model finally constructed in this paper to prove the accuracy of the basic model and the effect of model improvement.

There are many error evaluation indicators in subway passenger flow forecasting. This paper adopts the following four commonly used error evaluation indicators: MAPE, RMSE, Mean Absolute Error (MAE), and Coefficient of Determination (R^2). The expression is as follows:

$$\begin{aligned} MAPE &= \frac{1}{n}\sum_{i=1}^{n}\left|\frac{\hat{y}_i - y_i}{y_i}\right| \times 100\%, \\ RMSE &= \sqrt{\frac{1}{n}\sum_{i=1}^{n}(\hat{y}_i - y_i)^2}, \\ MAE &= \frac{1}{n}\sum_{i=1}^{n}|\hat{y}_i - y_i|, \\ R^2 &= 1 - \frac{\sum_{i=1}^{n}(\hat{y}_i - y_i)^2}{\sum_{i=1}^{n}(\overline{y} - y_i)^2}, \end{aligned} \tag{21}$$

Table 4: Inbound passenger flow prediction hyperparameter settings.

	Cluster 1	Cluster 2	Cluster 3	Cluster 4	Cluster 5
Average number of iterations	499	421	437	322	323
BiGRU hidden units 1	13	16	22	16	18
BiGRU hidden units 2	11	14	21	14	17

Table 5: Outbound passenger flow prediction hyperparameter settings.

	Cluster 1	Cluster 2	Cluster 3	Cluster 4	Cluster 5
Average number of iterations	278	251	486	522	503
BiGRU hidden units 1	13	16	22	20	18
BiGRU hidden units 2	11	14	21	19	17

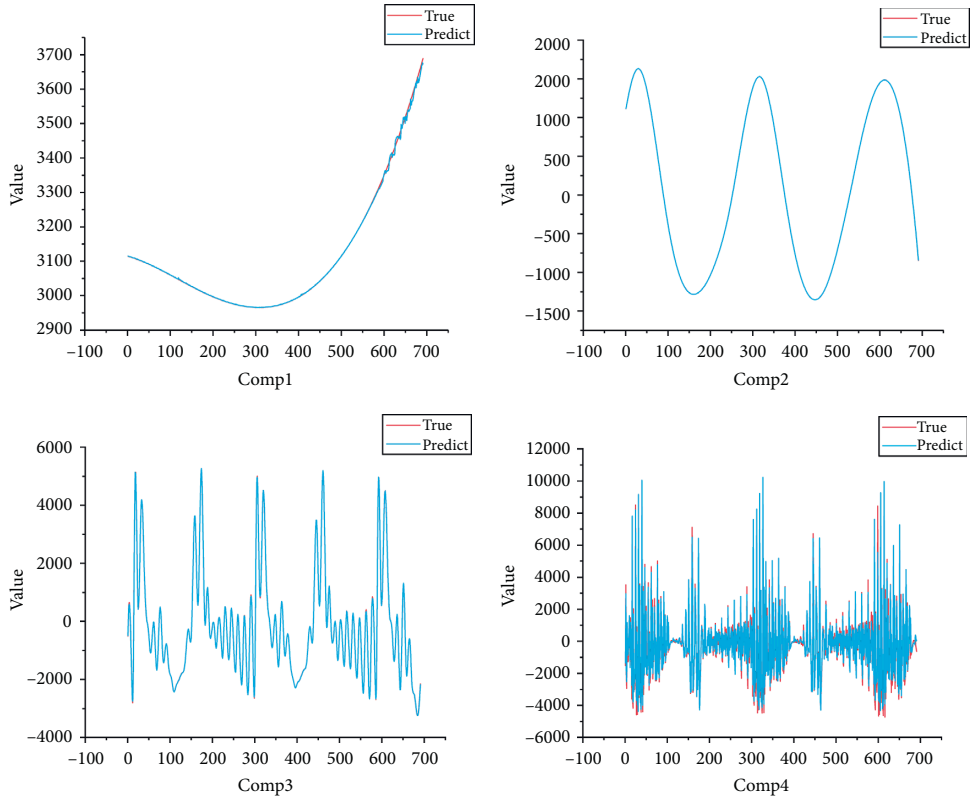

Figure 12: Prediction results of inbound traffic by the combined components of the first cluster of the station.

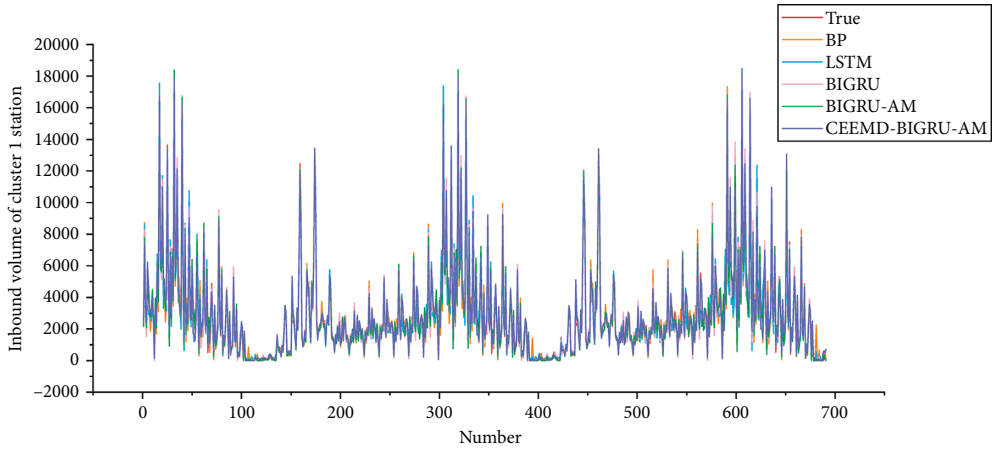

Figure 13: Prediction effect of different models on the inbound volume of the first cluster of the station.

where n is the number of test set data; y_i and $\widehat{y}_i$ are the actual and predicted values at a time, respectively; $\overline{y}$ is the sample mean.

Among them, MAPE represents the difference between the actual value and the predicted value as a percentage of the predicted value itself, RMSE represents the average difference between the predicted value and the actual value, and MAE represents the average absolute difference between the predicted value and the actual value. The smaller these three indicators are, the better the model prediction effect is. R^2 represents the proportion of the total variation of predicted passenger flow that can be explained by independent variables such as station location, entry, and exit time. The larger the R^2, the better the prediction effect of the model.

Figure 13 compares the forecasting effects of different models for the first cluster of subway station inbound passenger flow in a short period of time. It can be seen from the figure that, compared with other models, the CEEMD-BIGRE-AM model has the highest fitting degree to the passenger flow sequence and is closer to the actual value. In comparison, other models are more different from the actual value. The basic hyperparameters of the comparison models in this paper are the same as those of the main model, and the settings mentioned in Section 5.4 are the same.

Next, we will use the four model evaluation indicators MAPE, RMSE, $|R^2$, and MAE to comprehensively analyze the prediction accuracy of different models. The prediction errors of different models for inbound and outbound passenger flow are shown in Tables 6 and 7.

It can be seen from the comparison that the CEEMD-BiGRU-AM model outperforms the other five models on the four error evaluation indicators of MAPE, RMSE, MAE, and R^2. For predicting each type of passenger flow, the model established in this paper has reached the minimum in MAPE, RMSE, and MAE indicators and the maximum in R^2. Regardless of whether the overall passenger flow of the station is large, the morning and evening peak passenger flow is large, or the passenger flow is small.

(1) BP neural network easily falls into local optimum, resulting in problems such as gradient disappearance. Although LSTM solves this problem, its hidden layer only considers the impact of historical information on the current input, and GRU as a simplified version also has this problem. BiGRU considers the equally important future information by combining two GRUs in opposite directions. Tables 6 and 7 show that the standard BiGRU model outperforms the BP and LSTM models in most prediction situations and metrics. It verifies the superiority of using the BiGRU model as the base model.

(2) The model can extract local features more accurately after using the attention mechanism to improve BiGRU. The BiGRU-AM model has been far superior to the BiGRU, LSTM, and BP neural network models for each station's inbound and outbound passenger flow prediction cluster. For the prediction of the fourth type of station departures, the BiGRU-AM model reduced 31.19%, 52.18%, and 27.71% in MAPE indicators compared with BP neural network, LSTM, and BiGRU, respectively.

(3) The data becomes more stable after using the CEEMD-SE model to decompose and reconstruct the passenger flow time series data, and the prediction accuracy is further improved. In terms of the RMSE index, the station's entry and exit prediction accuracy are reduced by more than 40% on average compared to BP and LSTM and increased by more than 0.8% on average.

Table 6: Prediction results of inbound passenger flow by different models.

	Model	MAPE	RMSE	MAE	R^2
Cluster1	BP	0.5802	519.7456	329.7053	0.9685
	LSTM	0.9802	486.2718	294.5146	0.9725
	BiGRU	0.5616	347.8101	224.6894	0.9859
	BiGRU-AM	0.2508	278.0797	181.2615	0.9910
	CEEMD-SE-BiGRU-AM	0.1846	106.7617	52.9222	0.9945
Cluster 2	BP	1.4894	90.0056	64.1910	0.9367
	LSTM	1.3702	77.6200	58.8833	0.9529
	BiGRU	1.3609	79.4832	58.3151	0.9507
	BiGRU-AM	0.8673	76.7342	54.2109	0.9540
	CEEMD-SE-BiGRU-AM	0.5303	51.9745	35.6999	0.9789
Cluster 3	BP	0.5435	123.6597	89.2809	0.9763
	LSTM	0.3719	74.5092	54.4306	0.9914
	BiGRU	0.3369	74.4969	53.7010	0.9914
	BiGRU-AM	0.4445	72.3535	50.8886	0.9919
	CEEMD-SE-BiGRU-AM	0.1388	22.1308	14.6956	0.9992
Cluster 4	BP	1.377	207.373	141.936	0.9840
	LSTM	1.0932	151.2548	100.1705	0.9915
	BiGRU	0.7577	133.337	96.8934	0.9934
	BiGRU-AM	0.8194	116.7057	84.085	0.9949
	CEEMD-SE-BiGRU-AM	0.3983	113.671	57.7737	0.9952
Cluster 5	BP	1.213	245.1781	174.3473	0.9744
	LSTM	0.9732	161.1152	110.4429	0.9890
	BiGRU	1.2524	152.2723	103.4169	0.9901
	BiGRU-AM	1.7635	147.832	98.4795	0.9907
	CEEMD-SE-BiGRU-AM	0.9436	138.6708	89.9303	0.9918

5.6. Statistical Testing of Predictive Models. Considering that the above four evaluation indicators may have generalization and randomness, this paper further uses the statistical test method to verify the model's overall prediction effect. In order to verify the overall statistical significance of the combined model proposed in this paper, the Friedman test, Nemenyi test [37], and paired t-test [38] were used to verify the differences between different models.

The Friedman test can determine whether different algorithms perform significantly on multiple datasets without the need for normality assumptions. In this

Table 7: Prediction results of inbound passenger flow by different models.

	Model	MAPE	RMSE	MAE	R^2
Cluster1	BP	0.5411	529.8081	354.0068	0.9624
	LSTM	0.8076	580.7955	338.0679	0.9548
	BiGRU	0.2570	393.9422	252.4433	0.9841
	BiGRU-AM	0.6007	442.1145	207.9781	0.9843
	CEEMD-SE-BiGRU-AM	0.1971	330.0983	217.4731	0.9854
Cluster 2	BP	1.2569	84.7622	57.3832	0.9258
	LSTM	1.2017	87.9346	63.8113	0.9202
	BiGRU	0.7426	53.9097	40.0580	0.9700
	BiGRU-AM	0.5117	49.3602	36.0384	0.9749
	CEEMD-SE-BiGRU-AM	0.4114	48.0017	35.4340	0.9762
Cluster 3	BP	0.4730	143.0040	103.4939	0.9525
	LSTM	0.4425	92.7357	69.7211	0.9800
	BiGRU	0.2982	85.0159	61.7147	0.9832
	BiGRU-AM	0.2075	75.8107	54.9353	0.9867
	CEEMD-SE-BiGRU-AM	0.0631	20.1710	14.3087	0.9991
Cluster 4	BP	0.8992	145.4933	104.8215	0.9880
	LSTM	1.2938	132.9566	99.9509	0.9890
	BiGRU	0.8559	130.3113	87.7698	0.9904
	BiGRU-AM	0.6187	118.6154	84.5365	0.9922
	CEEMD-SE-BiGRU-AM	0.2563	68.5438	41.4484	0.9974
Cluster 5	BP	0.8703	220.3739	151.0948	0.9844
	LSTM	0.6068	140.2159	89.9998	0.9936
	BiGRU	0.8587	132.4883	88.1245	0.9943
	BiGRU-AM	0.5882	136.9004	88.4033	0.9940
	CEEMD-SE-BiGRU-AM	0.1644	53.1928	33.0957	0.9991

section, this paper uses the inbound and outbound traffic of each type of passenger flow as a whole to consider 10 sets of data sets and studies the differences between different models on the four measurement indicators MAPE, RMSE, MAE, and R^2, under the condition of a significance level $\alpha = 0.05$. The null hypothesis is that there is no significant difference in the performance of the five algorithms, and the alternative hypothesis is that the performance of the five models is significantly different. It is necessary to sort each algorithm and its corresponding evaluation index corresponding to each data, which set it as 1, 2, 3, 4, and 5 from the 1st to the 5th.

Due to the small sample size, it cannot be calculated directly using the chi-square statistic, but the F statistic is constructed from it:

$$\tau_{\chi^2} = \frac{k-1}{k} \cdot \frac{12N}{k^2-1} \sum_{i=1}^{k} \left(r_i - \frac{k+1}{2} \right)^2$$
$$\tau_F = \frac{(N-1)\tau_{\chi^2}}{N(k-1) - \tau_{\chi^2}}, \tag{22}$$

k is the number of algorithms to be compared, which is 5 here; N is the number of datasets, 10 here. r_i is the average ordinal value of the i-th algorithm. τ_{χ^2} is the chi-square statistic. τ_F is the F statistic and obeys the F distribution with k-1 and (k-1) (N-1) degrees of freedom.

The F-test here has a critical value of $F_{0.05}(4, 36) = 2.634$ and a significance level of 0.05. Through calculation, the p-values (probability that the test statistic is greater than the critical value) corresponding to different metrics are shown in Table 8.

It can be found that, for all indicators, the p-values of all indicators are much less than 0.01. Therefore, the null hypothesis is rejected, and the prediction effects among the five models are considered to be significantly different.

On this basis, it is necessary to continue to compare whether there are significant differences between the two models. In this regard, the Shapiro-Wilk test [39] is firstly performed on the evaluation index corresponding to each algorithm to judge whether it conforms to the normality assumption. The null hypothesis for this test is that the data follow a normal distribution. So if the p-value is less than 0.05, the null hypothesis is rejected, indicating that the data are not from a normal distribution. Its test statistic is

$$W = \frac{\left(\sum_{i=1}^{n} a_i x_{(i)} \right)^2}{\sum_{i=1}^{n} (x_i - \overline{x})^2}, \tag{23}$$

$x_{(i)}$ is the i-th smallest sample value in the data; $\overline{x}$ is the average value of the sample; x_i is the i-th sample value; a_i is a constant that meets certain conditions, which will not be repeated here.

Next, using the Shapiro-Wilk test, it can be seen that, among the four indicators, only the five models corresponding to the RMSE after logarithmic transformation do not reject the assumption of normal distribution. The p-values of the tests are shown in Table 9.

As can be seen in Table 9, most of the models have p-values less than 0.05 for MAPE, MAE, and metrics. Therefore, we performed a further test using Nemenyi's test, which does not require normality to be assumed. The null hypothesis is that there is no significant difference between the two models, and the alternative hypothesis is that there is a significant difference between the two models. The statistic is

$$C\ D = q_\alpha \sqrt{\frac{k(k+1)}{6N}}, \tag{24}$$

q_α is the critical value of α of the Tukey distribution, where $\alpha = 0.05$ is taken and the value is 2.728. k is the number of algorithms to be compared, which is 5. N is the number of datasets, 10 here.

The critical value is 1.9290 by calculation, and the average sequence value of the five models is shown in Table 10. The difference in mean ordinal values between the two pairs is shown in Figure 14.

In the case of a small amount of data, the Nemenyi test rejected the null hypothesis that the two models are inconsistent. When the average ordinal difference between the two models is only greater than the critical value of 1.929, the null hypothesis is rejected, and there is a significant

Table 8: p-value of the Friedman test.

	MAPE	MAE	RMSE	R^2
p-value	$1.745e-05$	$1.256e-07$	$2.394e-07$	$1.977e-07$

Table 9: p-value of the Shapiro-Wilk test.

	BP	LSTM	BiGRU	BiGRU-AM	CEEMD-SE-BiGRU-AM
MAPE	0.1665	0.5567	0.4062	0.0219	0.0478
MAE	0.0243	0.0008	0.0073	0.0293	0.0007
RMSE	0.0090	0.0006	0.0053	0.0050	0.0037
R^2	0.2390	0.0224	0.0060	0.0020	0.0370
ln(RMSE)	0.3250	0.0758	0.2861	0.5042	0.7861

Table 10: The average sequence value of five models.

	BP	LSTM	BiGRU	BiGRU-AM	CEEMD-SE-BiGRU-AM
MAPE	4.4	3.9	2.9	2.8	1
MAE	4.9	4.1	2.9	2.0	1.1
R^2	4.8	4.05	3.05	2.1	1

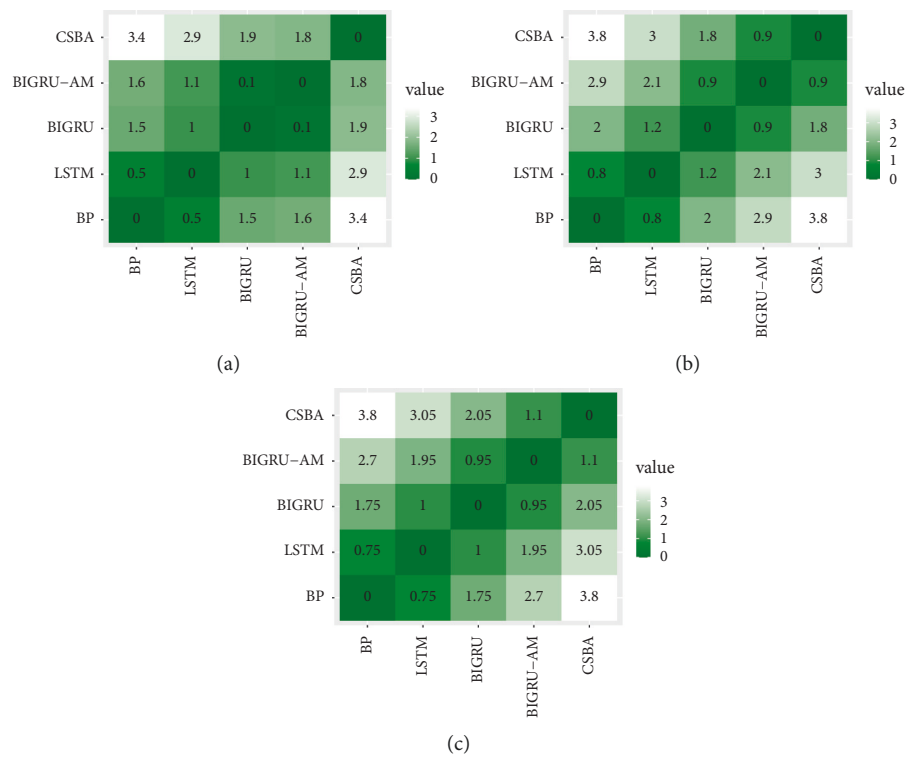

Figure 14: Nemenyi test of MAPE, MAE, and R^2 (CSBA replaces the CEEMD-SE-BiGRU-AM model in the figure), (a) MAPE, (b) MAE, (c) R^2.

difference between the two contrasting models. It can be seen from Figure 14 that the combined model established in this paper is significantly different from other models, and the differences between the BiGRU-AM model and BP and LSTM are also significant, while other models cannot directly reject the null hypothesis.

Then, for the ln(RMSE) that meets the normality assumption, a paired t-test with the same null hypothesis is performed between the two models, and its statistic is

$$t = \frac{\overline{d}}{s/\sqrt{N}}, \tag{25}$$

Table 11: p-value of the paired t-test.

	BP	LSTM	BiGRU	BiGRU-AM	CEEMD-SE-BiGRU-AM
BP	1	0.0092	<0.0001	<0.0001	0.0004
LSTM	0.0092	1	0.0256	0.0104	0.0007
BiGRU	<0.0001	0.0256	1	0.0736	0.0036
BiGRU-AM	<0.0001	0.0104	0.0736	1	0.0050
CEEMD-SE-BiGRU-AM	0.0004	0.0007	0.0036	0.0050	1

$\overline{d}$ is the mean of the sample differences; N is the sample size of 10; s is the standard deviation of the sample differences.

The statistic (25) follows a t-distribution with 9 degrees of freedom. It is obtained that the significance level is $\alpha = 0.05$ the p-value of the paired t-test between the models, as shown in Table 11.

At a significance level $\alpha = 0.05$, the above test can prove the following:

(1) The model after comprehensively using CEEMD and AM to improve BiGRU is significantly different from other comparable models.

(2) BiGRU already has a significant difference compared with BP and LSTM and has a significant difference with BiGRU-AM under the significance level $\alpha = 0.05$.

(3) Combining the results of Tables 9 and 10, it can be found that the model established in this paper is significantly better than other comparative models. AM and CEEMD have a noticeable improvement effect on the BiGRU model.

6. Conclusion

This paper proposes a clustering algorithm based on the GMM and WOA. On this basis, CEEMD decomposes the passenger flow sequence and then uses the AM-optimized BiGRU model to predict the short-term passenger flow. This paper uses the AFC data of some lines of the Shanghai Metro from April 1, 2015, to April 24, 2015, as a calculation example. This paper draws the following conclusions:

(1) Using WOA to optimize the parameters of the GMM, the stations are divided into 5 categories according to their spatiotemporal characteristics to save computational costs.

(2) Use the CEEMD method combined with SE to stabilize the time series and decompose and denoise the time series of inbound and outbound passenger flow in this method. The actual measurement example shows that, compared with the BiGRU-AM model without decomposition and noise reduction, the method improves the model's measurement of MAPE by an average of 49.92%.

(3) Using BiGRU as the prediction model's main body and the attention mechanism to capture local features of long-term sequences, the model reduces the MAPE indicators by 31.19%, 52.18%, and 27.71%, respectively, compared to BP, LSTM, and BiGRU.

This paper predicts the hourly inbound and outbound passenger flow of subway stations in different geographic locations.

6.1. Future Work. Some further research work needs to be considered for the current research framework. For example, a model is established based on the structure of the subway station network to form a more complex spatiotemporal system to improve the efficiency of this theoretical framework and broaden its application field, but it may increase the time cost. Of course, further verification of this theoretical approach is needed.

Data Availability

The data used to support the findings of this study are included within the article.

Conflicts of Interest

The author declares that there are no conflicts of interest.

References

[1] N. Shiwakoti, R. Tay, P. Stasinopoulos, and P. Woolley, "Passengers' perceived ability to get out safely from an underground train station in an emergency situation," *Cognition, Technology & Work*, vol. 20, no. 3, pp. 367–375, 2018.

[2] B. Han, Z. Yang, and H. Yu, "Statistical Analysis of Urban Rail Transit Operation in the World in 2020," *Urban Rapid Rail Transit*, vol. 34, pp. 5–11, 2021.

[3] Y. Zhou, S. Zheng, Z. Hu, and Y. Chen, "Metro station risk classification based on smart card data: a case study in Beijing," *Physica A: Statistical Mechanics and Its Applications*, vol. 594, Article ID 127019, 2022.

[4] W.-L. Zhao, C.-H. Deng, and C.-W. Ngo, "k -means: a revisit," *Neurocomputing*, vol. 291, pp. 195–206, 2018.

[5] X. Xu, S. Ding, L. Wang, and Y. Wang, "A robust density peaks clustering algorithm with density-sensitive similarity," *Knowledge-Based Systems*, vol. 200, Article ID 106028, 2020.

[6] B. J. Frey and D. Dueck, "Clustering by passing messages between data points," *Science*, vol. 315, no. 5814, pp. 972–976, 2007.

[7] Y. Guo, H. Jia, Y. Song, J. Kou, and C. Shen, "Parameter Correction for Electromagnetic Transient Simulation Model Based on GMM-PSO Hybrid Algorithm," *Power System Technology*, vol. 8, 2022.

[8] S. Mirjalili and A. Lewis, "The Whale Optimization Algorithm," *Advances in Engineering Software*, vol. 95, pp. 51–67, 2016.

[9] A.N. Jadhav and N. Gomathi, "WGC: Hybridization of exponential grey wolf optimizer with whale optimization for

data clustering," *Alexandria Engineering Journal*, vol. 57, pp. 1569–1584, 2018.

[10] S.C. Chu, H.C. Huang, J.F. Roddick, and J.S. Pan, "Overview of Algorithms for Swarm Intelligence," *Technologies and Applications*, vol. 6922, pp. 28–41, 2011.

[11] S. Abbas, Z. Jalil, A. R. Javed et al., "BCD-WERT: a novel approach for breast cancer detection using whale optimization based efficient features and extremely randomized tree algorithm," *PeerJ Computer Science*, vol. 7, Article ID e390, 2021.

[12] T. R. Gadekallu, D. S. Rajput, M. P. K. Reddy et al., "A novel PCA-whale optimization-based deep neural network model for classification of tomato plant diseases using GPU," *Journal of Real-Time Image Processing*, vol. 18, no. 4, pp. 1383–1396, 2021.

[13] S. Yan, L. Wu, J. Fan, F. Zhang, Y. Zou, and Y. Wu, "A novel hybrid WOA-XGB model for estimating daily reference evapotranspiration using local and external meteorological data: Applications in arid and humid regions of China," *Agricultural Water Management*, vol. 244, p. 2021, Article ID 106594.

[14] R. Liu, Y. Wang, H. Zhou, and Z. Qian, "Short-term passenger flow prediction based on wavelet transform and kernel extreme learning machine," *IEEE Access*, vol. 7, pp. 158025–158034, 2019.

[15] Y. Xiao, J. J. Liu, Y. Hu, Y. Wang, K. K. Lai, and S. Wang, "A neuro-fuzzy combination model based on singular spectrum analysis for air transport demand forecasting," *Journal of Air Transport Management*, vol. 39, pp. 1–11, 2014.

[16] A. M. Awajan, M. T. Ismail, and S. Al Wadi, "Improving forecasting accuracy for stock market data using EMD-HW bagging," *PLoS One*, vol. 13, no. 7, Article ID e0199582, 2018.

[17] Z. Wu and N. E. Huang, "Ensemble empirical mode decomposition: a noise-assisted data analysis method," *Advances in Adaptive Data Analysis*, vol. 01, no. 01, pp. 1–41, 2009.

[18] C. Zhu, X. Sun, P. Li, J. Zhang, and Y. Li, "Prediction of short-term urban rail transit flow incorporating station classification and data noise reduction," *Journal of Railway Science and Engineering*, pp. 1–10, 2022.

[19] A. Sovic and D. Sersic, "Signal decomposition methods for reducing drawbacks of the DWT," *Engineering Review*, vol. 32, 2012.

[20] Y. Zhao, L. Xia, and X. Jiang, "Short-term metro passenger flow prediction based on EMD-LSTM," *Journal of Traffic and Transportation Engineering*, vol. 20, pp. 194–204, 2020.

[21] J.-R. Yeh, J.-S. Shieh, and N. E. Huang, "Complementary Ensemble empirical mode decomposition: a novel noise enhanced data analysis method," *Advances in Adaptive Data Analysis*, vol. 02, no. 02, pp. 135–156, 2010.

[22] J. Tang, J. Liang, F. Liu, J. Hao, and Y. Wang, "Multi-community passenger demand prediction at region level based on spatio-temporal graph convolutional network," *Transportation Research Part C: Emerging Technologies*, vol. 124, Article ID 102951, 2021.

[23] L. Li, Y. Wang, and X. Li, "Tourists forecast lanzhou based on the baolan high-speed railway by the arima model," *Applied Mathematics and Nonlinear Sciences*, vol. 5, no. 1, pp. 55–60, 2020.

[24] X. Wang, N. Zhang, Y. Zhang, and Z. Shi, "Forecasting of short-term metro ridership with support vector machine online model," *Journal of Advanced Transportation*, vol. 2018, Article ID 3189238, 13 pages, 2018.

[25] Y. Wei and M.-C. Chen, "Forecasting the short-term metro passenger flow with empirical mode decomposition and neural networks," *Transportation Research Part C: Emerging Technologies*, vol. 21, no. 1, pp. 148–162, 2012.

[26] Y. Sun, B. Leng, and W. Guan, "A novel wavelet-SVM short-time passenger flow prediction in Beijing subway system," *Neurocomputing*, vol. 166, pp. 109–121, 2015.

[27] X. Yang, Q. Xue, X. Yang et al., "A novel prediction model for the inbound passenger flow of urban rail transit," *Information Sciences*, vol. 566, pp. 347–363, 2021.

[28] I. Sutskever, O. Vinyals, and Q. V. Le, "Sequence to sequence learning with neural networks," 2014, http://arXiv.org/abs/1409.3215.

[29] H. Zhang, Z. Gao, J. Li et al., "Short-term passenger flow forecasting of urban rail transit based on recurrent neural network," *Journal of Jilin University(Engineering and Technology Edition*, vol. 9, 2022.

[30] Y. Wu and H. Tan, "Short-term traffic flow forecasting with spatial-temporal correlation in a hybrid deep learning framework," arXiv:1612.01022, 2016.

[31] Z. Niu, G. Zhong, and H. Yu, "A review on the attention mechanism of deep learning," *Neurocomputing*, vol. 452, pp. 48–62, 2021.

[32] C. Fang, D. He, K. Li, Y. Liu, and F. Wang, "Image-based thickener mud layer height prediction with attention mechanism-based CNN," *ISA Transactions*, 2021.

[33] F. Xinghua, Z. Guo, and H. Ma, "An improved EM-based semi-supervised learning method," in *Proceedings of the 2009 International Joint Conference on Bioinformatics, Systems Biology and Intelligent Computingpp*, pp. 529–532, Article ID IEEE, Shanghai, China, 03-05 August 2009.

[34] D. Jia, "Shipborne non-intrusive load identification method based on hybrid GMM/SVM," *Periodical of Ocean University of China*, vol. 52, pp. 129–133, 2022.

[35] Y. Ding, Z. Chen, H. Zhang, X. Wang, and Y. Guo, "A short-term wind power prediction model based on CEEMD and WOA-KELM," *Renewable Energy*, vol. 189, pp. 188–198, 2022.

[36] J. S. Richman and J. R. Moorman, "Physiological time-series analysis using approximate entropy and sample entropy," *American Journal of Physiology - Heart and Circulatory Physiology*, vol. 278, no. 6, pp. H2039–H2049, 2000.

[37] K. Wen, G. Zhao, B. He, J. Ma, and H. Zhang, "A decomposition-based forecasting method with transfer learning for railway short-term passenger flow in holidays," *Expert Systems with Applications*, vol. 189, Article ID 116102, 2022.

[38] Z. Cheng, M. Trépanier, and L. Sun, "Incorporating travel behavior regularity into passenger flow forecasting," *Transportation Research Part C: Emerging Technologies*, vol. 128, Article ID 103200, 2021.

[39] S. S. Shapiro and R. S. Francia, "An approximate analysis of variance test for normality," *Journal of the American Statistical Association*, vol. 67, no. 337, pp. 215-216, 1972.