Short-Term Inbound Passenger Flow Prediction of Urban Rail Transit Based on RF-BiLSTM

Bo Zhang, Xinfeng Yang, Yongqing Zhang, and Dongzhi Li

Abstract—**Short-term passenger flow prediction of urban rail transit based on historical data mining can guide the work of station organization, train scheduling, and passenger flow induction effectively and dynamically. We propose the RF-BiLSTM prediction model, which combines the random forest algorithm (RF) and bi-directional long short-term memory neural network (BiLSTM). Firstly, the time series features of the flow are obtained by clustering algorithm and correlation analysis. Secondly, RF is used to obtain the importance of the features. Finally, we compare the prediction performance of 7 models and investigate the impact of feature selection on deep learning models. Through case analysis, the prediction accuracy of RF is the highest when using a single model for prediction, and the MAPE is 0.102. When using the combined prediction model, the MAPE of RF-BiLSTM is 0.074. The accuracy of RF-BiLSTM is better compared with the results of a single model. The performance of BiLSTM is improved by 46.6% after the features are optimized using feature selection methods. The findings demonstrate the suitability of the combined prediction model RF-BiLSTM for predicting short-term inbound passenger flow.**

Index Terms—**urban rail transit, passenger flow prediction, feature selection, Bi-directional LSTM.**

I. Introduction

URBAN rail transit is now a crucial component of the high-capacity public transportation system in many large cities. The dynamic adjustment transportation organization scheme has become an important development direction of urban rail transit operation management to effectively control the operating costs and improve the level of passenger transport service based on a brief variation in passenger flow. Short-term passenger flow prediction can optimize routes and organize the flow for the rail transit operations department to provide early warning and decision support. It is the premise of operational organization optimization of rail transit enterprises.

Manuscript received November 12, 2022; revised April 22, 2023.

This work was supported in part by Double-First Class Major Research Programs, Educational Department of Gansu Province (No. GSSYLXM-04), Projects Funded by the Central Government to Guide Local Scientific and Technological Development (No. 22ZY1QA005).

Bo Zhang is a postgraduate student at School of Traffic and Transportation, Lanzhou Jiaotong University, Lanzhou 730070, China. (e-mail: bo_zhang0108@163.com).

Xinfeng Yang is a professor at School of Traffic and Transportation, Lanzhou Jiaotong University, Lanzhou 730070, China. (Corresponding author, e-mail: xinfengyang@mail.lzjtu.cn).

Yongqing Zhang is an engineer of Hangzhou Metro Operation Co., LTD, Hangzhou 310000, China. (e-mail: zhangyongqing@hzmetro.com).

Dongzhi Li is a postgraduate student at School of Traffic and Transportation, Lanzhou Jiaotong University, Lanzhou 730070, China. (e-mail: 1051986992@qq.com).

At present, there are a wealth of forecasting methods for short-term passenger flow, mainly including autoregressive integrated moving average (ARIMA) [1], random forest (RF) [2], support vector machine (SVM) [3], Gated Recurrent Unit (GRU) [4], long short-term memory (LSTM) [5] and graph neural network (GNN) [6].

In addition, many scholars used combinatorial methods to study time series problems and achieve excellent predictions. For passenger flow and bus speed prediction, Che et al. [7] and Ali et al. [8] combined wavelet transform and empirical modal decomposition with ARIMA, respectively, and proposed different combined models. Li et al. [9] suggested a model that merges the SVM with radial basis function neural networks to predicate the passenger flow. Sun et al. [10] introduced a unique Wavelet-SVM prediction model, which uses SVM to learn and predict passenger flow by decomposing the flow into high and low-frequency vectors. Xie et al. [11] combined temporal convolutional networks and LSTM to address the issue that rail transit flow is variable and challenging to predict. In order to predict bus passenger flow, LSTM, and locally weighted regression were merged in [12], and a modified STL-LSTM prediction model was presented. Huang et al. [13] used five prediction algorithms as the basic model, and three ensemble models were then built utilizing a variety of ensemble approaches, such as RF, AdaBoost, and Linear Regression, and then predicted the bus running time.

These studies of [7]-[13] have decomposed the passenger flow data, obtained different decomposition vectors, and made predictions according to the decomposition vectors. These studies cannot characterize the specific variables associated with changes in passenger flow, lacks the interaction between the flow and the related variables, and cannot obtain the importance of associated features. At the same time, the studies of [14]-[16] have not studied the features of passenger flow sufficiently and have not analyzed the feature importance effectively, which affects the model's accuracy.

In order to obtain the relevant features of passenger flow, analyze the importance of the features. The relevant features are extracted using the clustering and correlation analysis methods, and the feature selection function of the RF and the prediction function of Bi-direction LSTM are combined to propose RF-BiLSTM to predicate the short-term inbound passenger flow.

The remainder of this article is structured as follows. In Section II, the extraction method of passenger flow features is introduced. In Section III, the RF-based feature selection method is introduced, and the model of RF-BiLSTM is constructed. In section IV, a case is introduced, using the feature extraction method to extract the features, and using the RF-BiLSTM prediction model to select the features and predicate future passenger flow. In Section V, the conclusions are presented.

II. Passenger Flow Features Extraction Method

In passenger flow prediction, inputting too few features will reduce the prediction accuracy, and inputting too many will increase the model's computational complexity and cause the waste of computer resources. Using the features extraction method based on clustering and Spearman rank correlation coefficient to extract passenger flow features can improve the validity of the input features and reduce the prediction time.

A. Clustering Method

Urban rail transit's distribution of hourly passenger flow can be categorized into five types: unidirectional peak type, bidirectional peak type, full peak type, sudden peak type, and no-peak type [17]. The unidirectional peak type occurs in areas with obvious tidal passenger flow, such as residential and office areas. The bidirectional peak type occurs in the comprehensive functional areas. The full peak type occurs in the transportation hub areas. The sudden peak type occurs in large public facilities areas. The no-peak type occurs in underdeveloped areas.

The passenger flow distribution characteristics of each station are extracted by the clustering method, and the stations with similar passenger flow trends are categorized. Among clustering algorithms, the K-means++ algorithm has a simple clustering principle, easy implementation, and highly efficient operation. The steps are as follows.

Step 1. A random selection is made from the data X of an initial cluster center C_l.

Step 2. Determine the Euclidean distance $D(x)$ from the sample to the cluster center.

Step 3. After calculating the probability $P(x)$ that each sample point would be chosen as the cluster center, a sample point was randomly chosen as the following cluster center based on the probability distribution.

$$P\left(x\right)=\frac{D\left(x\right)^2}{\sum_{x\in X}D\left(x\right)^2} \tag{1}$$

Step 4. Repeat Step 2 and Step 3 until k cluster centers are obtained.

Step 5. Measure the $D(x)$ from each sample to the cluster center, divide the samples to the nearest centroid, and get k class clusters $\left\{S_1,S_2,...,S_k\right\}$.

Step 6. Determine the mean value of each cluster's sample features, and use that value to determine where each class's new clustering center C_l should be

$$C_l=\frac{\sum_{X_i\in S_l}X_i}{\left|S_l\right|}, \tag{2}$$

where X_i and $|S_l|$ denote the number of objects and the i-th object, respectively, in the l-th class cluster, $1\le i\le\left|S_l\right|$, $1\le l\le k$.

Step 7. To obtain the clustering result, repeat Step 5 and Step 6 until the clustering center stops changing.

B. Correlation Coefficient Method

Passenger flow data is a type of time series data, which has a certain correlation with historical data. The historical and current passenger flow correlation is examined using Spearman's rank correlation coefficient ρ, then the time series feature is determined. The ρ is defined as

$$\rho=\frac{\sum_{i=1}^{n}(rx_i-\overline{rx}_i)(ry_i-\overline{ry}_i)}{\sqrt{\sum_{i=1}^{n}(rx_i-\overline{rx}_i)^2}\sqrt{\sum_{i=1}^{n}(ry_i-\overline{ry}_i)^2}}, \tag{3}$$

where, rx_i and ry_i are the ranks of the random variables X and Y, respectively, $\overline{rx}_i$ and $\overline{ry}_i$ are the means of rx_i and ry_i respectively.

III. Prediction Model

A. Feature Selection

The random forest algorithm is capable of measuring the significance of related variables in addition to having good prediction accuracy. Wei [18] and Yassine et al. [19] used the RF to explain the importance of variables and used the importance ranking for feature selection. The RF algorithm calculates the degree of feature importance in this paper, and the feature selection and prediction are realized by combining the RF algorithm and BiLSTM neural network.

The feature importance indicates the degree of influence of data features on the prediction results. The accuracy of prediction can be increased, and the optimization time can be decreased by selecting features with higher importance, constructing feature sets, and inputting the feature sets into the model. Let *VIM* represent the feature importance, and *Gini* represent the gini index. The calculation process of *VIM* is (4) - (7).

$Gini_q^i$ is the gini index of node q of the i-th tree, C is the number of categories, and p_{qc}^i is the proportion of category c in q.

$$Gini_q^i=\sum_{c=1}^{C}\sum_{c'\ne c}p_{qc}^i p_{qc'}^i=1-\sum_{c=1}^{C}\left(p_{qc}^i\right)^2 \tag{4}$$

VIM_{jq}^i, $Gini_l^i$ and $Gini_r^i$ represent the change in the gini index before and after branching, as well as the gini index of the two new nodes.

$$VIM_{jq}^i=Gini_q^i-Gini_l^i-Gini_r^i \tag{5}$$

VIM_j is the importance of feature X in the forest of I trees, and Q is the set of nodes where feature X_j appears in i trees.

$$VIM_j=\sum_{i=1}^{I}\sum_{q\in Q}VIM_{jq}^i \tag{6}$$

Finally, the importance VIM_j of feature X_j is obtained by normalization.

$$VIM_j=\frac{VIM_j}{\sum_{k=1}^{K}VIM_k} \tag{7}$$

When the importance *VIM* of each feature is obtained, the features with minimum importance are successively removed to obtain different feature sets, and the different feature subsets are input into BiLSTM to acquire the prediction error. When the feature set is empty, end the cycle, and the feature set with the minimum prediction error is output. The features in the feature set at this time are the selected features. Fig. 1 displays the process of feature selection.

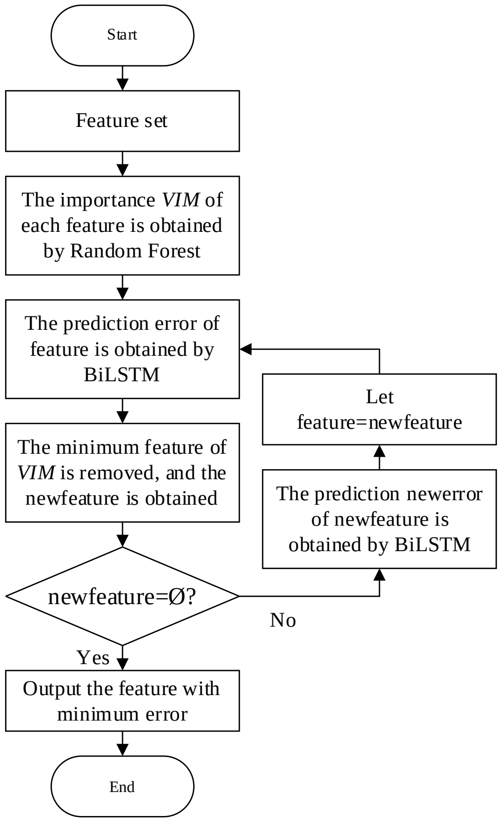


Fig. 1. The feature selection process

B. RF-BiLSTM Prediction Model

Long short-term memory neural network has an excellent predictive effect on time series problems. Input, forget, and output gates are used to introduce the sigmoid and tanh functions, which lessens the likelihood of gradient expansion and gradient disappearance. The gates and cell states in the LSTM are shown in Fig. 2, and the expressions for the gates and memory cells [20] are (8) – (13).

Input gate:

$$i_t = \sigma(W_i \cdot [h_{t-1}, x_t] + b_i), \tag{8}$$

$$\tilde{C}_t = \tanh(W_C \cdot [h_{t-1}, x_t] + b_C). \tag{9}$$

Output gate:

$$o_t = \sigma(W_o \cdot [h_{t-1}, x_t] + b_o), \tag{10}$$

$$h_t = o_t * \tanh(C_t). \tag{11}$$

Forget gate:

$$f_t = \sigma(W_f \cdot [h_{t-1}, x_t] + b_f). \tag{12}$$

Cell state:

$$C_t = f_t * C_{t-1} + i_t * \tilde{C}_t, \tag{13}$$

where b_i, b_C, b_o, *and* b_f are the bias, $\sigma(x) = 1/(1+\exp(-x))$, x_t is the input vector, $[h_{t-1}, x_t]$ represents connecting h_{t-1} and x_t into a vector, and the weight matrix in the forget gate, input gate, and output gate is represented by W_f, W_i, W_C and W_o.

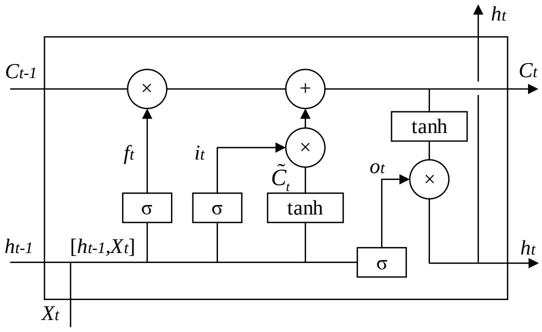


Fig. 2. The LSTM cell structure

Bi-directional LSTM [20] consists of forward LSTM and backward LSTM neural networks, as shown in Fig. 3, which overcomes the problem that the LSTM network has a poor memory effect for earlier information and does not consider the interdependence of current information and future information.

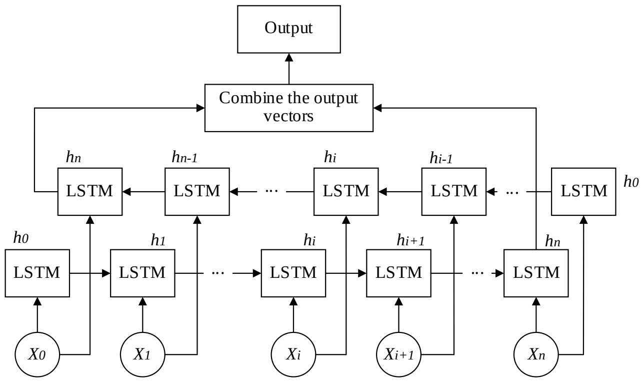


Fig. 3. The BiLSTM model

Based on the feature selection method and combined with BiLSTM, the RF-BiLSTM prediction model is proposed. The model can better to select passenger flow features, learn and predict the flow, and enhance the correlation between the model's input features and prediction precision. The single-step prediction method is used, and the flow in the next period is predicted according to the previously known data. Since the input data of the future is unknown, the sliding window prediction method is added to bring the data after each prediction into RF-BiLSTM as the input data until the prediction of all periods is completed. Fig. 4 shows the prediction process of the RF-BiLSTM.

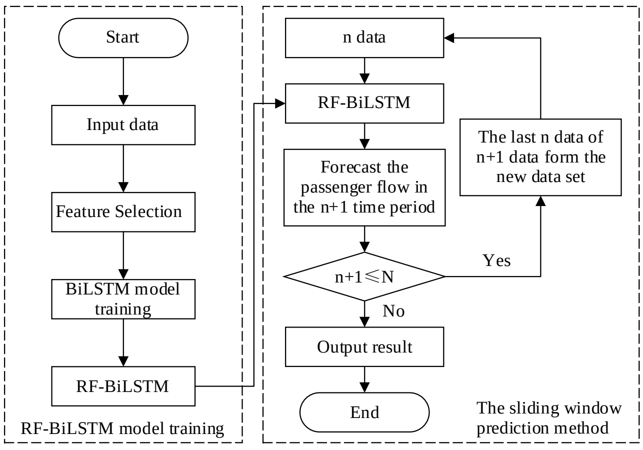


Fig. 4. The prediction process of RF-BiLSTM

C. *Evaluation Index*

Three error indicators, mean absolute error (MAE), mean absolute percentage error (MAPE), and root mean square error (RMSE), are introduced to evaluate the results. The formula is shown in (14) – (16). The closer the three error indicators are to zero, the more accurate the prediction result is.

$$MAE = \frac{1}{n}\sum_{i=1}^{n}\left|\hat{p}_i - p_i\right|, \tag{14}$$

$$MAPE = \frac{1}{n}\sum_{i=1}^{n}\left|\frac{\hat{p}_i - p_i}{p_i}\right|. \tag{15}$$

RMSE is different from MAE in that RMSE is more sensitive to high amount of deviation.

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(p_i - \hat{p}_i)^2}, \tag{16}$$

where p_i and $\hat{p}_i$ represent the values of actual and predicted, respectively, and n is the predicted time periods' number.

IV. Instance Analysis

The experimental data are the inbound passenger flow data from January 1, 2019, to January 27, 2019, from 6:00 to 22:00, and the data are counted in 30-minute intervals. There are 32*31*27 inbound passenger flow data for 31 stations, including 18 working days. The first 16 days' flow serves as the training set, while the final 2 days' flow serves as the validation set. The passenger flows in the last two days are predicted and analyzed using the evaluation index.

A. *Passenger Flow Features*

Initially, the features of passenger flow are analyzed by the clustering method. We consider that there are five typical types of the flow distribution, so the range of the number of clusters is set from 2 to 5. After clustering, the Calinski-Harabaz (CH) index and the Silhouette Coefficient (SC) results are shown in Fig. 5. The more significant the CH index and the SC are, the better the clustering effect is. The best clustering effect is achieved when the number of clusters is 3. Then we classify the passenger flow into three types. Table I displays the station number of the three types, and Fig. 6 displays the clustering centers for the three types.

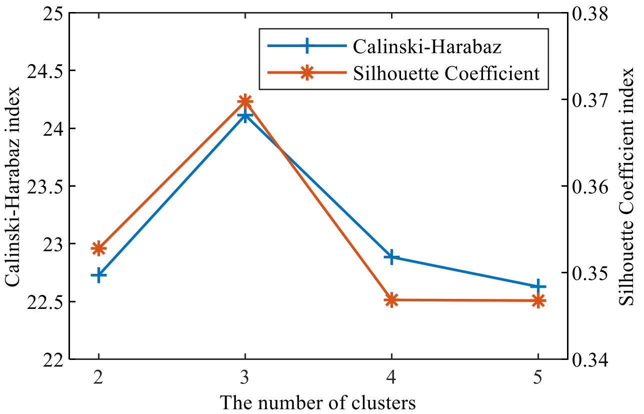


Fig. 5. CH index and Silhouette Coefficient

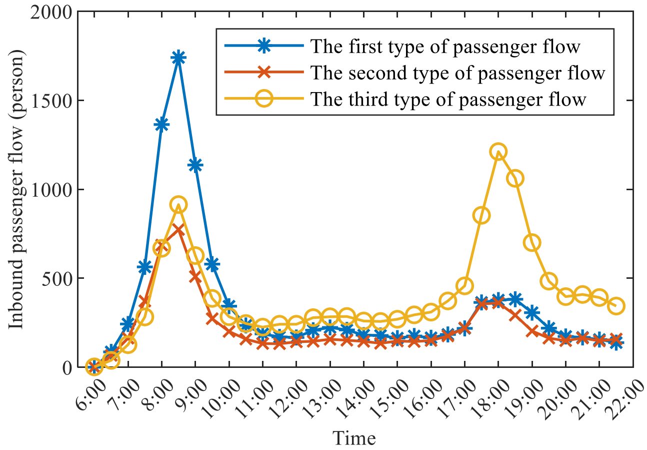


Fig. 6. The clustering centers

TABLE I
THE STATIONS OF THE THREE TYPES

Passenger flow type	Station number
The first type	8、9、22、23、25、26、27、28
The second type	0、1、2、3、4、5、6、7、14、17、25、29、30
The third type	10、11、12、13、15、16、18、19、20、21、24

From Fig. 6, the first type of passenger flow has a prominent morning peak, while the flow is less during the evening peak, and these stations belong to the residential area. The second type also has a pronounced morning peak, but the flow is lower than that of the first type, and these stations belong to the residential area with a low degree of development. The evening peak of the third type is larger than the morning peak, and the peak of morning and evening is obvious, and these stations belong to the comprehensive functional area.

Passenger flow has peak and off-peak features within a day. The peak features show the degree of the flow at a certain time, which can provide a basis for prediction. K-means ++ algorithm is used to extract peak features of three types of passenger flow, and Table Ⅱ shows the clustering results.

TABLE II
CLUSTERING RESULTS OF PASSENGER FLOW PERIOD

Passenger flow type	Peak features	Peak hours
The first type	Morning off-peak, Noon off-peak	6:00 - 7:30、9:00 - 14:30
	Morning peak	7:30 - 9:00
	Evening peak	14:30 - 22:00
The second type	Morning off-peak, Noon off-peak	6:00 - 7:00、9:00 - 15:00
	Morning peak	7:00 - 9:00
	Evening off-peak	15:00 - 22:00
The third type	Morning off-peak	6:00 - 7:30、9:00 - 12:00
	Morning peak	7:30 - 9:00
	Noon off-peak	12:00 - 17:00
	Evening peak	17:00 - 19:00
	Evening off-peak	19:00 - 22:00

Spearman's rank correlation coefficient ρ is used to analyze the closeness between the current passenger flow and the flow one day ago, the flow 30 minutes ago, the flow 60 minutes ago, and the flow 90 minutes ago. Where the significance (two-tailed) of the Spearman rank correlation coefficient is denoted by S, and the correlation analysis results are shown in Table III. Spearman's rank correlation coefficient indicates that the current passenger flow is significantly correlated with the flow one day ago, the flow 30 minutes ago, and the flow 60 minutes ago, so these three times can be used as passenger flow time series features.

TABLE III
SPEARMAN'S CORRELATION COEFFICIENT BETWEEN CURRENT PASSENGER FLOW AND HISTORICAL PASSENGER FLOW

		The current	One day ago	30 mins ago	60 mins ago	90 mins ago
The current	ρ	1.000	.961**	.768**	.440*	0.105
	S		<0.001	<0.001	0.012	0.566
One day ago	ρ	.961**	1.000	.785**	.418*	0.054
	S	<0.001		<0.001	0.017	0.770
30 mins ago	ρ	.768**	.785**	1.000	.766**	.440*
	S	<0.001	<0.001		<0.001	0.012
60 mins ago	ρ	.440*	.418*	.766**	1.000	.765**
	S	0.012	0.017	<0.001		<0.001
90 mins ago	ρ	0.105	0.054	.440*	.765**	1.000
	S	0.566	0.770	0.012	<0.001	

Note: **. and *. indicate significant correlation (two-tailed) at the levels of 0.01 and 0.05, respectively.

In a certain residential area, the fluctuation of passenger travel demand is slight, and the total change of the inbound flow of adjacent stations is small. Fig. 7 shows the inbound flow at station 2 concerning the inbound flow at station 1 and station 3 on the previous day. The flow of station 2 is about 1/2 of the flow of neighboring stations on the previous day. The sum of neighboring stations' flow on the previous day is used as a feature of the flow.

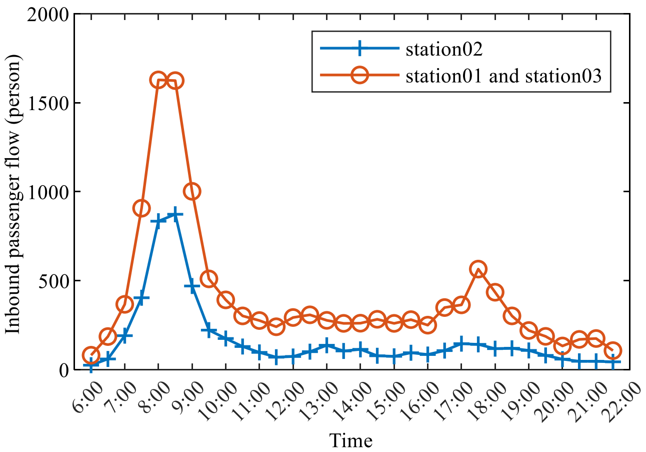


Fig. 7. The inbound flow at neighboring stations

There is a nonlinear relationship between weather indicators and passenger flow [21], so it is necessary to take weather indicators as the features of passenger flow. In this paper, we take the temperature, weather conditions, and wind level as relevant characteristics that may affect residents riding urban rail transit. Temperature is numerical data and can be used directly. Weather conditions and wind levels are character data, which need to be manually calibrated and converted into numerical data q and f before use. Table IV displays the mapping association between the weather condition and wind level.

TABLE IV
THE MAPPING RELATIONSHIP BETWEEN THE WEATHER CONDITION AND THE WIND LEVEL

Weather Condition	q	Wind Level	f
Sunny	0	0~2	0
Partly Cloudy	1	2~4	1
Cloudy	2	4~6	2
Shower	3	6~8	3
Light Rain	4		
Moderate Rain	5		
Heavy Rain	6		

B. Model Parameter Setting

The settings of batch and epoch parameters greatly impact prediction accuracy, and unreasonable settings can easily lead to underfitting or overfitting of the model. Huang et al. [22] investigate the effect of epoch on RMSE and training loss. We further study the effect of epoch and batch parameters on prediction accuracy. In order to identify better parameter settings to enhance the model's learning and prediction impacts, we conduct an experimental analysis of the batch and epoch parameters of the model. Taking station 9 as an example, the experimental situation of the batch parameter is [8, 16, 32, 64, 128], and the experimental situation of the epoch parameter is [50, 100, 150, 200, 250, 300]. The experimental results are fitted by MATLAB. Fig. 8. to Fig. 10. shows the error changes of RMSE, MAPE, and MAE.

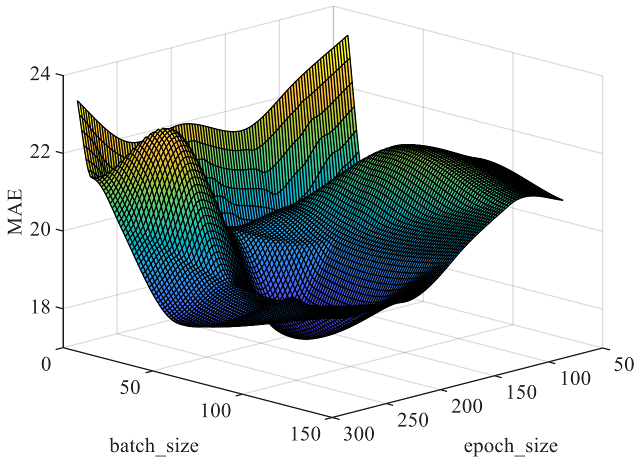


Fig. 8. The change in the MAE

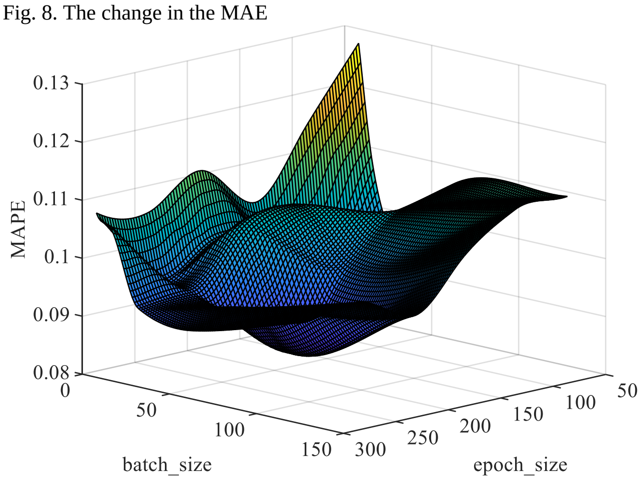


Fig. 9. The change in the MAPE

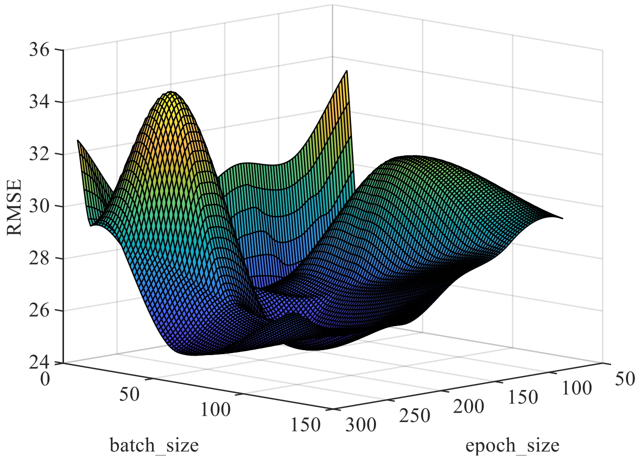


Fig. 10. The change in the RMSE

From Fig. 9 – 11, under the condition that the epoch is fixed and the batch is gradually increased, the evaluation index displays a trend of first lowering and then increasing. Because the batch is small, the training data of one batch of the model is insufficient, the learning effect is poor, and the model is under-fitting. Under the condition that the epoch gradually increases and the batch is fixed, the evaluation index shows a rising trend, then falling, and then rising again. The model training effect is under-fitting, fitting, and over-fitting. The evaluation index shows an upward trend when the epoch and batch are close to 300 and 150. The model is overfitted. The model training effect is the best when batch and epoch are set to 64 and 200, respectively.

Table V shows the parameter settings for the RF and BiLSTM.

TABLE V
PARAMETER SETTING

Model	Paramater	Value
RF	n_estimators	10
	max_depth	50
BiLSTM	batch_size	64
	epoch_size	200
	hidden layer	2
	cells number of hidden layer	64/32
	optimizer	adam
	loss function	mse
	activation function	tanh

C. Optimization Results and Evaluation

1) Feature Selection Result

According to the analysis of passenger flow features, a total of 8 passenger flow features are obtained: the passenger flow one day ago, the flow 30 minutes ago, the flow 60 minutes ago, the sum of neighboring stations' passenger flow on the previous day, temperature, wind, weather, and the peak feature. The *VIM* is used to express the importance of the features of the three types of passenger flow, and Table VI shows the results.

As shown in Table VI, weather conditions contribute the least to changes in passenger flow, indicating that passengers are not very sensitive to weather changes within the weekday range. The passenger flow one day ago and the peak feature are the key features.

TABLE VI
THE *VIM* OF THREE PASSENGER FLOW TYPES

Feature	*VIM* (%)		
	The first type	The second type	The third type
One day ago	62.26	55.08	79.65
30 mins ago	0.19	0.64	0.69
60 mins ago	0.16	0.55	0.68
The sum of neighboring stations' passenger flows	0.15	0.40	0.46
Temperature	0.03	0.15	0.16
Wind	0.01	0.01	0.02
Weather	0.03	0.10	0.19
Peak feature	35.17	43.07	18.15

According to the feature selection method, one feature with minimum feature importance is removed from the feature set each time, and obtain seven feature subsets. Each feature set is employed as input for the prediction model, and the prediction error of station 9 is shown in Fig. 11.

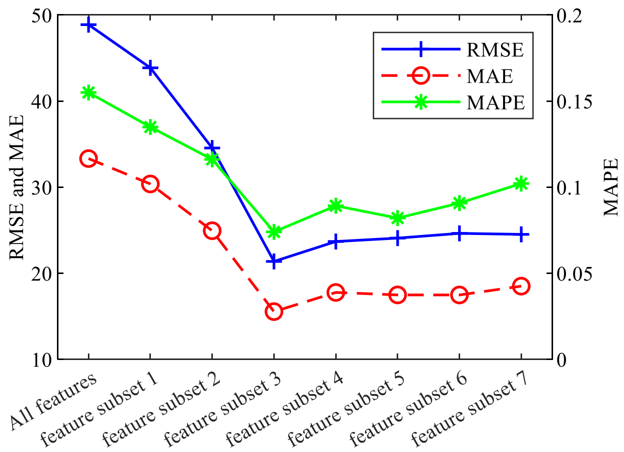


Fig. 11. Error of each feature set

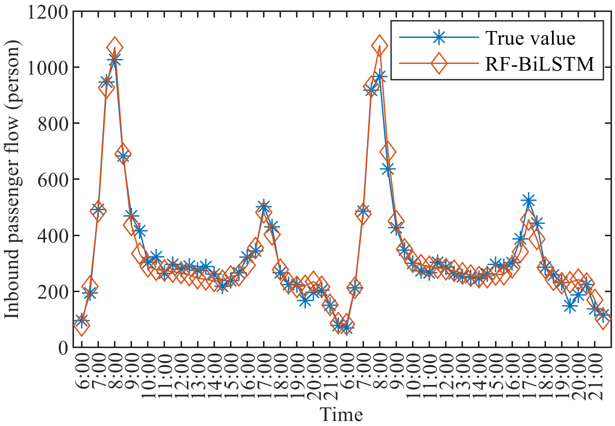
The model predicts best when the passenger flow features are set to feature subset 3. The features in feature subset 3 are the passenger flow one day ago, the flow 30 minutes ago, the flow 60 minutes ago, the sum of neighboring stations' passenger flow on the previous day, and the peak feature. It is considered that the five features of feature subset 3 are the critical features that affect the prediction accuracy. Compared to all features, subset 3 has three fewer features, which helps to reduce the computing time, Table VII shows the computing time of RF-BiLSTM. With the guarantee of prediction accuracy, it can be found that the average computing time per 5 times is reduced by 18.26%, which is mainly due to the reduction of epoch parameter from 250 to 200. The RF-BiLSTM uses the five features as its input features in station 9.

TABLE VII
THE COMPUTING TIME OF RF-BILSTM

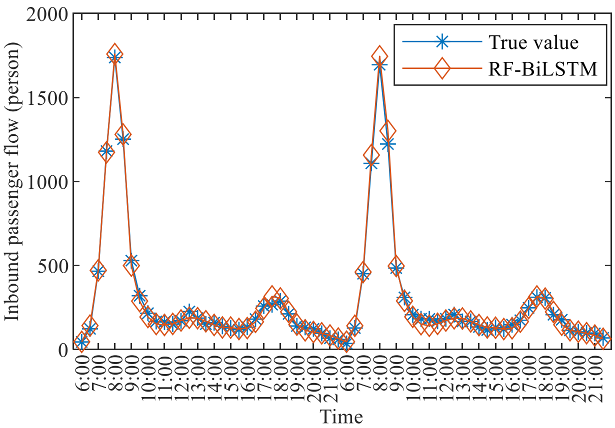
	Number of features		Reduction rate of computing time
	8	5	
Computing time of 5 times (s)	207.32	169.45	18.26%
Average computing time (s)	41.46	33.89	

2) Prediction Results

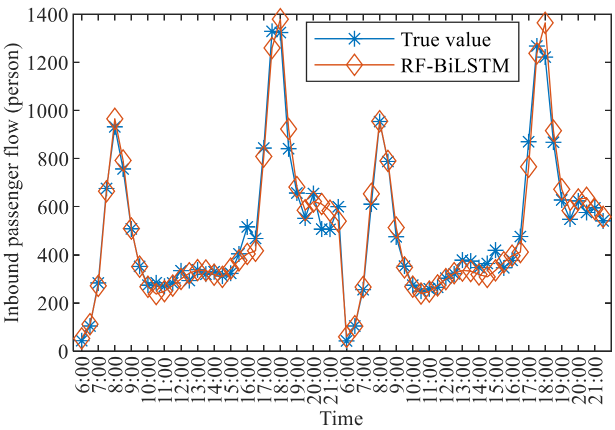
Limited by space, the prediction results of only one station in each passenger flow type are shown below, as shown in Fig. 12.



(a). Station 3



(b). Station 9



(c). Station 12

Fig. 12. Prediction results for three stations

From Fig. 12, the RF-BiLSTM model has excellent accuracy for three types of passenger flow. On the first day, the results of the three types' flow are very close to the true values. However, on the second day, the prediction effect of the three types is decreased. Because the prediction results of the second day are predicated on the previous day's data, the error will be gradually amplified, leading to a decrease in accuracy.

The inbound passenger flow at station 9 is predicted using the prediction models of Back Propagation (BP) neural network, SVM, RF, GRU, LSTM, BiLSTM, and RF-BILSTM in order to test the RF-BILSTM model's predictive ability. Table VIII displays the prediction error findings for each model, and Fig. 13 displays the prediction performance of 7 models on January 24.

TABLE VIII
THE PREDICTION ERROR

Model	MAE	MAPE	RMSE
BP	22.552	0.105	32.617
SVM	25.189	0.194	32.403
RF	19.470	0.102	27.807
GRU	22.768	0.121	29.882
LSTM	26.658	0.126	39.676
BiLSTM	27.210	0.149	40.061
RF-BiLSTM	15.539	0.074	21.367

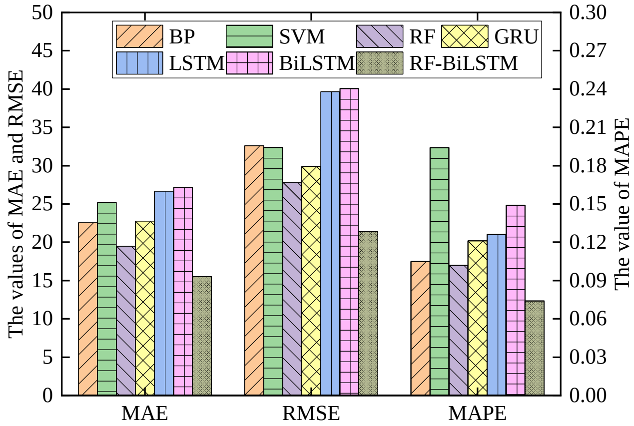
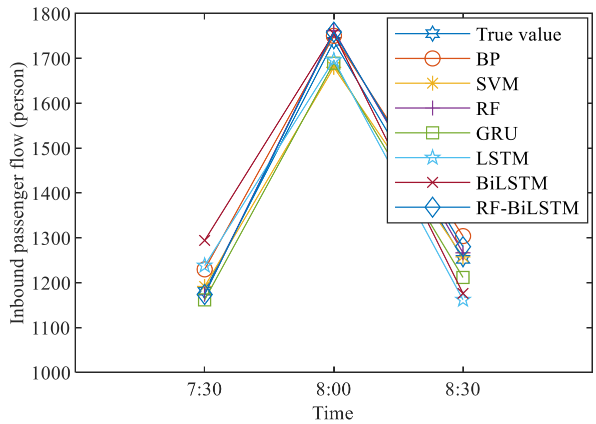


Fig. 13. Prediction performance of 7 models

From Table VIII and Fig. 13, when using a single model, the traditional machine learning method RF outperforms the time series deep learning methods GRU, LSTM, and BiLSTM in terms of prediction accuracy, and its MAPE is 0.102. After the feature selection method is added, the prediction of the deep learning method is significantly improved, and MAPE can be reduced to 0.074. RF makes random selection of feature subsets, and it is highly resistant to noise data. BiLSTM has memory and bidirectional learning mechanisms and will be more sensitive to noisy data. BiLSTM learns well for the optimized feature data.

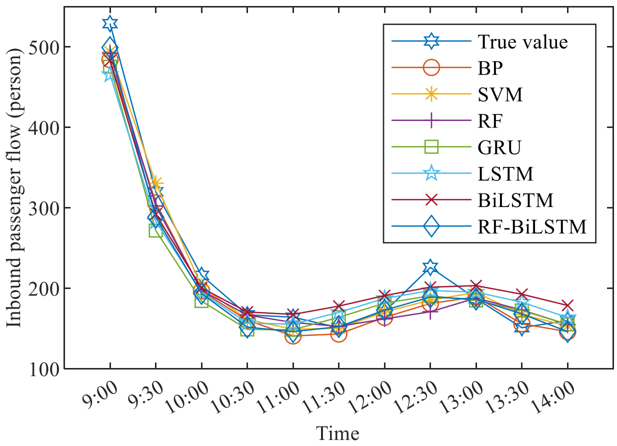
To show the difference in accuracy and stability of the seven prediction models intuitively, we predict the inbound passenger flow at station 9 for each peak hour, as shown in Figure 14.



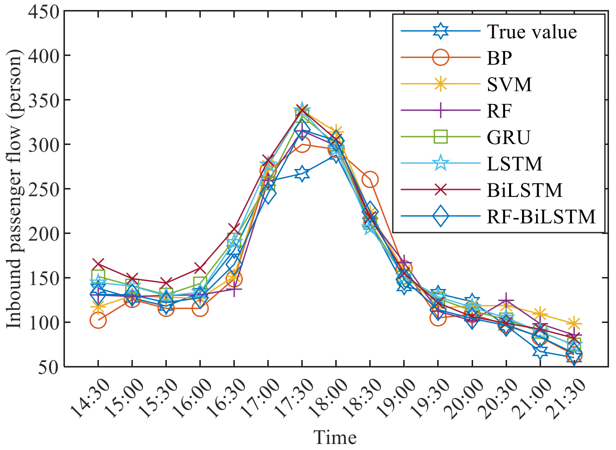
(a). Morning peak

TABLE IX
MODEL PREDICTION RESULTS FOR ALL FEATURES AND THE FEATURES AFTER FEATURE SELECTION

Model	Error			Performance Improvement			Average Improvement
	MAE	MAPE	RMSE	MAE	MAPE	RMSE	
BP	22.552	0.105	32.617	0.092	0.019	0.131	0.081
RF-BP	20.472	0.103	28.351				
SVM	25.189	0.194	32.403	0.209	0.247	0.188	0.215
RF-SVM	19.921	0.146	26.318				
LSTM	26.658	0.126	39.676	0.292	0.317	0.327	0.312
RF-LSTM	18.867	0.086	26.700				
GRU	22.768	0.121	29.882	0.154	0.301	0.050	0.168
RF-GRU	19.254	0.085	28.397				
BiLSTM	27.210	0.149	40.061	0.429	0.503	0.467	0.466
RF-BiLSTM	15.539	0.074	21.367				



(b). Noon off-peak



(c). Evening off-peak

Fig. 14. Comparison of prediction results

From Fig. 14(a), the results of RF-BiLSTM are closer to the true value than the other six models, which is more evident in the morning peak. From Fig. 14(b) - 14(c), the prediction of the seven models does not deviate too much from the true value, and the performance of RF-BiLSTM is more stable than the other six models. Peak passenger flow places strict requirements on the organization of train operations, and companies pay more attention to peak passenger flow during train operating hours. Accurate and stable peak passenger data can guide the work of station organization, train scheduling, and passenger flow induction effectively and dynamically.

We use all features and the features after feature selection to predict inbound passenger flow and compare the results of the five models to confirm the effectiveness of the feature selection for improving the performance of the single model. Table IX shows the performance improvements of the five models.

Table IX demonstrates that the combined model with feature selection outperforms the single prediction model in terms of accuracy. The performance of BiLSTM and LSTM is greatly improved after the features are optimized using feature selection methods, and the performance of BiLSTM is improved by 46.6%, which proves the improvement effect of the feature selection function on the single model.

The prediction results of all stations are counted. The RMSE, MAE, and MAPE of each station are shown in Fig. 15.

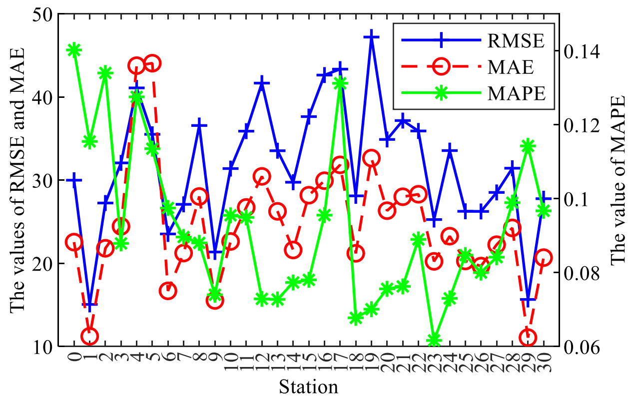


Fig. 15. The prediction errors of all stations

From Fig. 15, the prediction errors are at a low level for the majority of stations. The prediction errors indicate that the RF-BiLSTM has good applicability. The RF-BiLSTM can make accurate predictions for different types of inbound passenger flows. The MAPE of 80% of stations is below 10%, and the MAPE of station 23 can reach 6.16%. The RMSE of 73.3% of the stations is between 25 and 40, the RMSE of four stations is lower than 25, and the RMSE of 5 stations is higher than 40, indicating the deviation between the prediction and the true data is small. The average variation between the expected findings and the true data is not very large, and the MAE of 86.7% of stations is less than 30.

V. CONCLUSION

The RF-BiLSTM model is proposed to solve urban rail transit's short-term passenger flow prediction problem. Firstly, the clustering algorithm and correlation analysis are used to capture passenger flow features. After that, the random forest algorithm and BiLSTM neural network are combined to obtain the importance of the features and predict the flow. Finally, the accuracy of the RF-BiLSTM and other models is compared by an instance. The specific conclusions are as follows.

(1) The RF-BiLSTM model has good accuracy, stability, and applicability compared with other models in the paper.

(2) The feature selection method based on RF can greatly improve the prediction performance of a single model and reduce computing time.

Overall, the RF-BiLSTM model is suitable for short-term inbound passenger flow prediction and can provide the companies with accurate prediction data. The next research can focus on the features of holiday and rail transit networks to further strengthen the validity of the features and the prediction accuracy.

REFERENCES

[1] Chuan Ding, Jinxiao Duan, Yanru Zhang, Xinkai Wu, and Guizhen Yu, "Using an ARIMA-GARCH Modeling Approach to Improve Subway Short-Term Ridership Forecasting Accounting for Dynamic Volatility," *IEEE Transactions on Intelligent Transportation Systems*, vol. 19, no. 4, pp. 1054-1064, April 2018.

[2] Wei Li, Liying Sui, Zhou Min, Hairong Dong, "Short-term passenger flow forecast for urban rail transit based on multi-source data," *EURASIP Journal on Wireless Communications and Networking*, vol. 2021, no. 1, 2021.

[3] Zhuangbin Shi, Ning Zhang, Paul M. Schonfeld, and Jian Zhang, "Short-term metro passenger flow forecasting using ensemble-chaos support vector regression," *Transportmetrica A: Transport Science*, vol. 16, no. 2, pp. 194-212, 2020.

[4] Xijun Zhang, Guangjie Yu, Jiyang Shang, and Baoqi Zhang, "Short-term Traffic Flow Prediction With Residual Graph Attention Network," *Engineering Letters*, vol. 30, no. 4, pp. 1230-1236, 2022.

[5] Thandassery Sajanraj, Jaison Mulerikkal, S. Raghavendra, Vinith Rejathalal, and Vit Fabera, "Passenger flow prediction from AFC data using station memorizing LSTM for metro rail systems," *Neural Network World*, vol. 31, no. 3, pp. 173-189, 2021.

[6] Jinlei Zhang, Feng Chen, Yinan Guo, and Xiaohong Li, "Multi-graph convolutional network for short-term passenger flow forecasting in urban rail transit," *IET Intelligent Transport Systems*, vol. 14, no. 10, pp. 1210-1217, 2020.

[7] Jinxing Che, and Huachao Zhai, "WT-ARIMA Combination Modelling for Short-term Load Forecasting," *IAENG International Journal of Computer Science*, vol. 49, no. 2, pp542-549, 2022

[8] Kianifar Mohammad Ali, Motallebi Hassan, and Khatibi Bardsiri Vahid, "A stacked ensemble learning method for traffic speed forecasting using empirical mode decomposition," *Journal of the Chinese Institute of Engineers*, vol. 45, no. 3, pp. 282-291, 2022.

[9] Dewei Li, Chen Zhang, and Jinming Cao, "Short-Term Passenger Flow Prediction of a Passageway in a Subway Station Using Time Space Correlations Between Multi Sites," *IEEE Access*, vol. 8, pp. 72471-72484, 2020.

[10] Yuxing Sun, Biao Leng, and Wei Guan, "A novel wavelet-SVM short-time passenger flow prediction in Beijing subway system," *Neurocomputing*, vol. 166, pp. 109-121, 2015.

[11] Binglei Xie, Yu Sun, Xiaolong Huang, Le Yu, and Gangyan Xu, "Travel Characteristics Analysis and Passenger Flow Prediction of Intercity Shuttles in the Pearl River Delta on Holidays," *Sustainability*, vol. 12, no. 18, pp. 7249-7249, 2020.

[12] Feng Jiao, Lei Huang, Rongjia Song, and Haifeng Huang, "An Improved STL-LSTM Model for Daily Bus Passenger Flow Prediction during the COVID-19 Pandemic," *Sensors*, vol. 21, no. 17, pp. 5950-5950, 2021.

[13] Haifeng Huang, Lei Huang, Rongjia Song, Feng Jiao, and Tao Ai, "Bus Single-Trip Time Prediction Based on Ensemble Learning," *Computational Intelligence and Neuroscience*, vol. 2022, pp. 6831167-6831167, 2022.

[14] Zhongwei Hou, Zixue Du, Guang Yang, and Zhen Yang, "Short-Term Passenger Flow Prediction of Urban Rail Transit Based on a Combined Deep Learning Model," *Applied Sciences*, vol. 12, no. 15, pp. 7597-7597, 2022.

[15] Fei Xue, Enjian Yao, Ning Huan, Binbin Li, and Shasha Liu, "Prediction of Urban Rail Transit Ridership under Rainfall Weather Conditions," *Journal of Transportation Engineering, Part A: Systems*, vol. 146, no. 7, pp. 04020061, 2020.

[16] Ciyun Lin, Kang Wang, Dayong Wu, and Bowen Gong, "Passenger Flow Prediction Based on Land Use around Metro Stations: A Case Study," *Sustainability*, vol. 12, no. 17, pp. 6844-6844, 2020.

[17] Cheng Zhang, "Analysis on The Characteristics of Urban Rail Transit Passenger Flow," Southwest Jiaotong University, 2006.

[18] Li Wei. "Optimization and Application of Random Forest Algorithm for Applied Mathematics Specialty," *Security and Communication Networks*, vol. 2022, 2022.

[19] Akhiat Yassine, Manzali Youness, Chahhou Mohamed, Zinedine Ahmed, "A New Noisy Random Forest Based Method for Feature Selection," *Cybernetics and Information Technologies*, vol. 21, no. 2, pp. 10-28, 2021.

[20] XuDong Li, JieSheng Wang, WenKuo Hao, and XiaoRui Zhao, "Bi-directional Long-term Short-term Memory Classification Method of Diabetes Datasets," *Engineering Letters*, vol. 30, no.4, pp1603-1611, 2022.

[21] Jingwen Wu, and Hua Liao, "Weather, travel mode choice, and impacts on subway ridership in Beijing," *Transportation Research Part A*, vol. 135, no. C, pp. 264-279, 2020.

[22] Huang Sen, "Time Series Prediction based on Improved Deep Learning," *IAENG International Journal of Computer Science*, vol. 49, no.4, pp1133-1138, 2022.