Passenger Flow Prediction Based on Composite Model

Shuying Liu, Xiuqian Xia, Yaoyao Liu, He Zhang, Xiqin Ao*
College of Big Data and Artificial Intelligence, AnHui Xin Hua University, Hefei, China
**Corresponding Author.*

Abstract:
The article aims to investigate the performance combination model and traditional single prediction model in traffic flow prediction, providing reliable and accurate traffic flow prediction data for intelligent transportation systems. In this article, a fusion of these two models urban rail transit passenger flow prediction model (TCN-BiLSTM) is proposed based on temporal convolutional network (TCN) and bi-directional long and short term memory network (BiLSTM). Firstly, using TCN to extract hidden information and temporal relationships of urban rail passenger flow, capturing the dependency relationships between long-distance features. Secondly, the BiLSTM model further obtains the before and after connections between the passenger flow characteristics of urban rail transit. Finally, the output results of BiLSTM are incorporated into an attention mechanism to allocate different permissions, increase the weight of important features, and reduce the impact of redundant information on the prediction accuracy of TCN-BiLSTM. This article evaluates the model's predictions using MAE (Mean Absolute Error), RMSE (Root Mean Square Error), R2 (Coefficient of Determination), and MAPE (Mean Absolute Percentage Error).The results of the ablation experiment showed that the TCN-BiLSTM prediction model reduced RMSE, MAE, and MAPE by 10.3, 6.6, and 1.9, respectively. Meanwhile, R2 also reached 0.904. When using the TCN-BiLSTM prediction model for prediction at different sites, the R2 values are all above 0.9. By comparison, the TCN-BiLSTM prediction model has the smallest error, highest fit, and high portability in predicting metro rail passenger traffic passenger flow.

Keywords: TCN, BiLSTM, metro rail passenger traffic, passenger flow

INTRODUCTION

With the rapid development of social economy, the size of cities is growing increasingly large, and the phenomenon of traffic congestion in the city is also more serious. To relieve the pressure of urban road traffic and reduce people's travel time, urban rail transit with subway as the mainstream came into being. With the continuous expansion of the scale of metro rail passenger traffic, predicting the future passenger flow according to the dynamic change of passenger flow can not only help travelers choose reasonable routes and avoid congestion, but also help operators formulate reasonable operation plans, deploy security, security and other measures at stations in advance, and conduct scientific passenger flow guidance, management and control [1-3].

The prediction of metro rail passenger traffic also belongs to the forecasting category of traffic flow. Experts and scholars have achieved fruitful results in passenger flow forecasting for urban rail transportation. Duan et al. used the spatio-temporal topology diagram to fill in the missing data, and used the spatio-temporal autoregressive moving average model to prediction traffic flow, reducing the number of filling points to reduce the prediction time and improve efficiency of model [4]. Ye et al. introduced an improved sparrow search algorithm to optimize support vector regression prediction model, which solved the problems of slow convergence and local optimality of traditional sparrow search algorithm, and built HMSSSA-SVR prediction model to achieve accurate prediction of short-term traffic flow [5]. Zhang et al. introduced a timely passenger flow prediction model of urban rail transit considering multi-time scale features. This method is composed of two modules, Attention-GRU and improved Transformer, which can explore the continuity and periodicity characteristics of historical passenger flow respectively to prediction passenger flow [6]. Kuang et al. used adaptive spatiotemporal graph convolutional cyclic network to predict traffic flow. In this thesis, the spatial network structure diagram of traffic network was constructed, and the spatial features of traffic flow were extracted by adaptive graph convolutional neural network. Gated unit GRU is used to extract the time characteristics of traffic flow, which improves the accuracy of prediction [7]. In their paper, Dai et al. introduced a traffic flow prediction model that combines spatiotemporal analysis with gated recursive unit (GRU). The Characteristics of time and space of traffic flow are converted into a two-dimensional matrix and processed by GRU to achieve the purpose of short-term traffic flow prediction [8]. Lian et al. introduced a traffic flow prediction model based on the attention mechanism by stacking long short-term memory network (LSTM) neural network to capture the time series features of traffic flow data and introduce the attention mechanism to learn the global features, and improve the prediction accuracy [9].

In the above studies, researchers such as Zhang and Kuang used graph convolutional recurrent networks for prediction. Although they successfully identified temporal features in traffic flow, they only focused on the impact of speed as a single dimension and did not discover multidimensional impact features; however, although researchers such as Duan and Lian have improved the basic prediction model, the shortcomings of a single prediction model have not been resolved.

This thesis introducs a prediction methodology TCN-BiLSTM. First, the TCN is used to extract the hidden information and time relationship of metro rail passenger traffic passenger flow, and to capture the dependence relationship between long-distance features. Secondly, BiLSTM model further acquires the relationship between passenger flow characteristics of metro rail passenger traffic. Finally, the output results of BiLSTM are added to the attention mechanism, and the weight of important features is increased by assigning different permissions to reduce the influence of redundant information on the prediction accuracy of TCN-BiLSTM.

The prediction methodology introduced overcomes the defects of single model which is difficult to learn multidimensional feature parameters, poor generalization ability and low prediction accuracy, and improves the prediction accuracy. Finally, authors use real passenger flow data of Metro Line 1 to conduct simulation experiments, and compare the TCN-BiLSTM with the TCN, LSTM, BiLSTM, and TCN-LSTM to further verify the effectiveness of this model.

RELATED WORK

TCN

TCN is a kind of neural network based on convolutions [10]. It was introduced by Lea et al in 2016 and applied to video segmentation, and has gradually been extended to speech recognition, motion detection and time series-based prediction and classification [11]. TCN overcomes the problem that the traditional modeling methods based on time series data (such as RNN, LSTM) need to calculate each time serial, and the model parameters are too large, and realizes parallel calculation of time series data, and has advantages fewer model parameters. The structure of the TCN is shown in Figure 1.

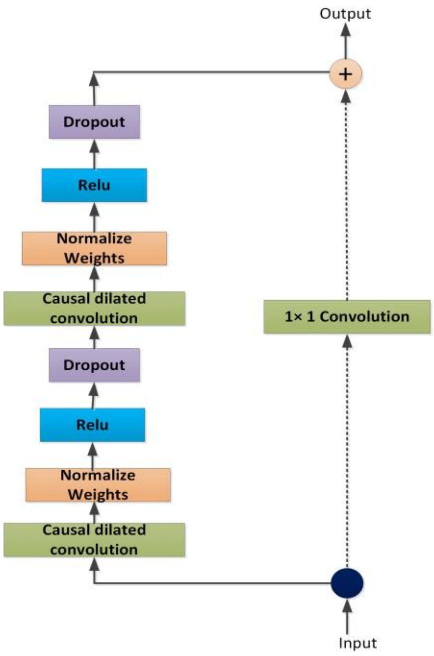

Figure 1. TCN structure diagram

TCN is composed of two parts. The left part is composed of four parts: dilated convolutions, normalization, ReLU and Dropout, and the right part is a residual block, that is, a one-dimensional convolution.

Dilated Convolutions, also known as extended convolution [12], expand the sampling interval of convolutional nuclei to make the output features of the model have a wider sensory field and higher expressiveness. Extended convolution is similar to empty convolution in CNN.

Extended convolution improves the scalability of causal convolutional networks, but the receptive field of TCN is still affected by network depth, convolution kernel size and expansion factor. As the depth of the network increases, problems such as vanishing or exploding gradients and degradation of network performance will arise. TCN normalizes each weight vector in the network, simultaneously uses ReLU nonlinear excitation function and Dropout regularization [13,14], and adds residual connections. Among them, the ReLU function and Dropout regularization can speed up network convergence, improve model expressiveness, prevent overfitting, and avoid gradient problems.

BiLSTM

LSTM [15,16] overcomes the problem of long-term dependence, which cannot be handled by RNN(Recurrent Neural Netword), and has a better ability to learn long-term information.It is a improved RNN [17,18] whose memory unit is mainly composed of three gates: input , output, and forget. They symbolize the gate of information, controlling the transmission of neuronal information [19]. At the same time, they interact to control the operation of the module. In the LSTM model, the opening and closing of doors determine whether historical information is retained or deleted. The three gates, under the action of the activation function, will produce a number between [0, 1], 0 indicates that all information is deleted, 1 indicates that all information is retained. The structure of the LSTM is shown in Figure 2.

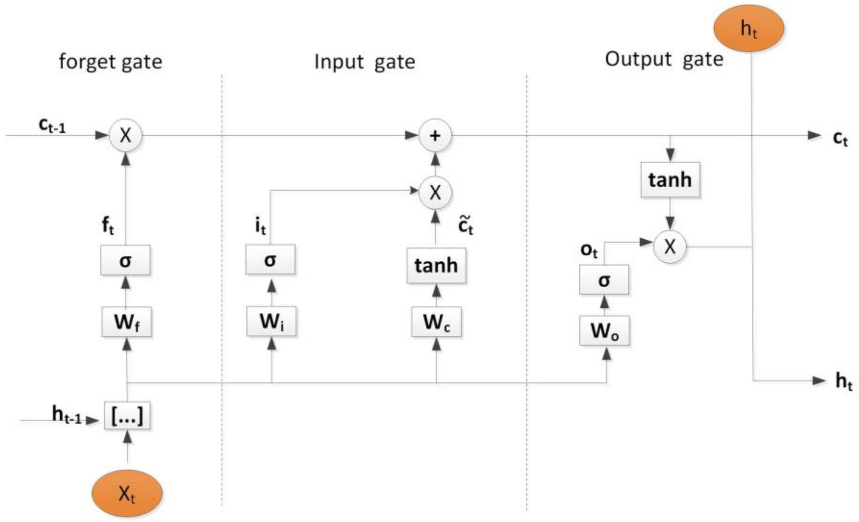

Figure 2. The structure of of LSTM

The forget gate determines which information is deleted, which is crucial for separating important features and reducing noise interference. The forgetting gate is mathematically described by the following formula:

$$f_t = \sigma(W_f \cdot (h_{t-1}, X_t) + b_f) \quad (1)$$

The input gate controls how much information in the network input vector X_t at the current moment is retained in the cell state C_t. In this way, the input gate can effectively filter out important information and prevent irrelevant content from entering the memory module. The the input gate is mathematically described by the following formula:

$$i_t = \sigma(W_i \cdot (h_{t-1}, X_t) + b_i) \quad (2)$$

$$\widetilde{C_t} = \tanh(W_c \cdot (h_{t-1}, X_t) + b_c) \quad (3)$$

$$C_t = f_t * C_{t-1} + i_t * \widetilde{C_t} \quad (4)$$

C_{t-1} is the state of a previous moment, $\widetilde{C_t}$ is the current state, $f_t * C_{t-1}$ represents the part expected to be forgotten, C_t is the new cell state, and is updated in the memory module.

The Output gate is responsible for regulating how much information from the unit state Ct is passed to the output value ht of the LSTM at the current moment. This ensures that only relevant and important memory information can affect the output result. The input gate is mathematically described by the following formula:

$$O_t = \sigma\left(W_o \cdot (h_{t-1}, X_t)\right) + b_o \tag{5}$$

$$h_t = O_t * \tanh(C_t) \tag{6}$$

i_t is the input gate vector, O_t is the output gate vector, h_t is the output vector, and X_t is the input vector. W is the weight and b is the bias.

The sigmoid function is mathematically described by the following formula:

$$S(x) = \frac{1}{1+e^{-x}} \tag{7}$$

The tanh function is mathematically described by the following formula:

$$\tanh(x) = \frac{e^x - e^{-x}}{e^x + e^{-x}} \tag{8}$$

BiLSTM [20-22] contains a forward LSTM and an inverse LSTM, and the results of the two LSTM processing are spliced together as the final output of BiLSTM. By stacking two layers of LSTM, BiLSTM not only overcomes the limitation that the model can only predict the next time information based on the previous time information [23]. At the same time, it can also learn the before-and-after information between features more effectively, and use the positive and negative direction information of the current position to carry out two-way learning. Figure 3 shows the structure of BiLSTM.

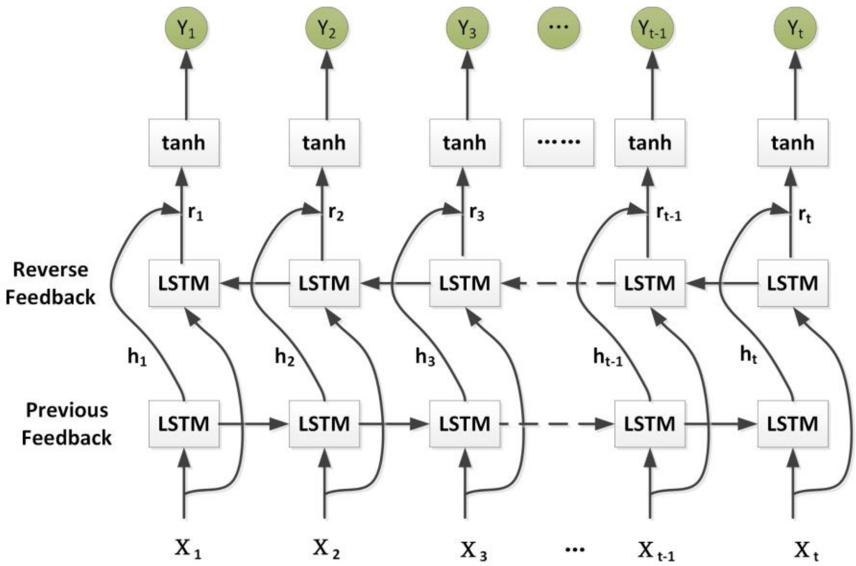

Figure 3. BiLSTM structure diagram

Where, the formula for h_t can be expressed as:

$$h_t = \omega_1 x_t + \omega_2 h_{t+1} \tag{9}$$

The formula for r_t can be expressed as:

$$r_t = \omega_3 x_t + \omega_4 r_{t-1} \tag{10}$$

The formula for Y_t can be expressed as:

$$Y_t = tanh(\omega_5 r_t + \omega_6 h_t) \tag{11}$$

BUILD PREDICTION MODEL BASED ON TCN-BILSTM

In order to provide accurate passenger flow prediction data for metro rail passenger traffic operators, the article designed the TCN-BiLSTM. The model makes prediction with three components. The first component is the pre-processing of data. This module focuses on accomplishing the collection and processing of data and splitting the processed dataset, proportionally, into two parts: the training set and the test set; the second component is the training of the model, which is mainly responsible for the training of the parameters, and finds the optimal parameter combination through optimization of loss function. The third part is model prediction and evaluation, mainly through the training of the optimal prediction model to predict the passenger flow, through the comparison between the predicted value and the real value, calculate the evaluation index, and finally evaluate the model and verify the effectiveness and accuracy of the prediction model introduced in the article. Figure 4 shows the prediction process.

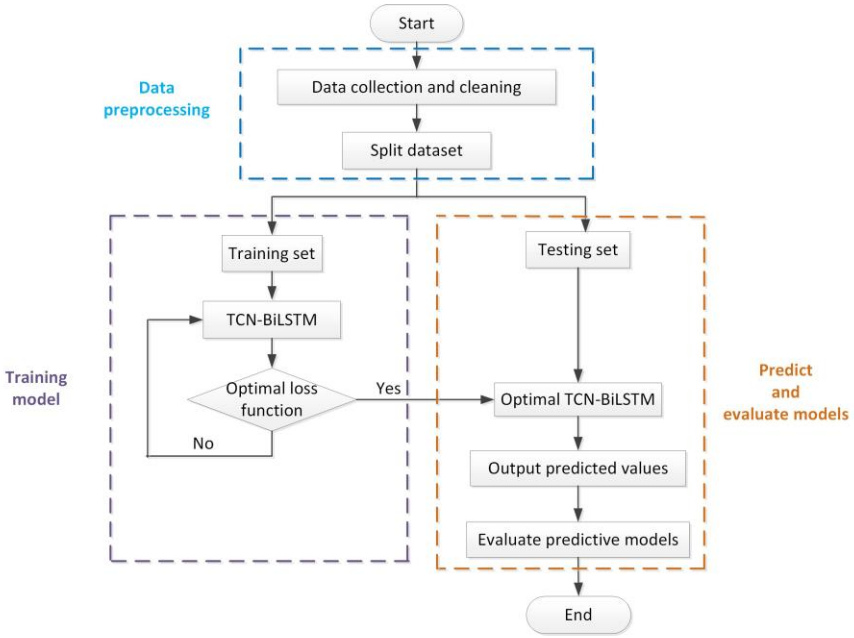

Figure 4. Model prediction flow diagrams

TCN-BiLSTM consists of three parts: TCN, BiLSTM and fully connected layer. Figure 5 shows its structure.

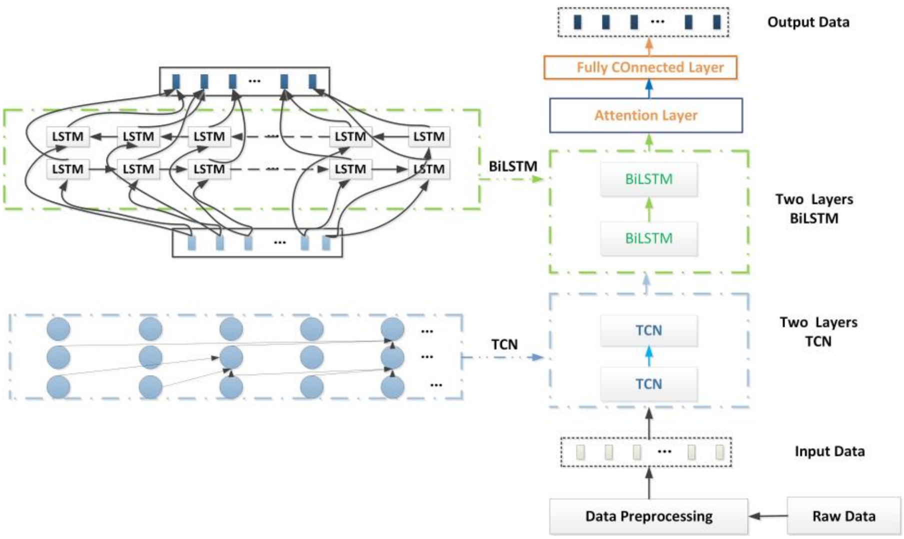

Figure 5. Structure diagram of TCN-BiLSTM

TCN-BiLSTM's process is as follows: First, the historical data affecting the passenger flow is normalized and input into the TCN layer. The multi-layer convolution of different feature parameters is carried out by TCN to extract the local deep features of the metro rail passenger traffic data and obtain the feature vector of the relationship between the parameters. Then, the relationship feature vector extracted from the TCN layer is input into the BiLSTM to further enhance the information of key time points and improve the prediction effect in a long period of time. The output results of the last layer of BiLSTM are spliced and fused into the Attention layer, and the weight of important features is increased by assigning different permissions. Finally, after the full connectivity layer, the passenger flow of the station is predicted.

ANALYSIS OF APPLICATION EXAMPLES

Data Preprocessing

Data source

To verify the validity of the TCN-BiLSTM, a total of 349 days of the automatic fare collection system (AFC) data of Line 1 in a city in China from January 2, 2022 to December 17, 2022 were selected. Automated ticketing system data is the relevant travel information collected as passengers pass through the gates as they enter and exit the station. The original data set contains 42 fields of information such as card number and card device number. The field information needed for passenger flow prediction is extracted from the original data set, and 7 fields such as card number, station number are retained. The field information is shown in Table 1.

Table 1. List of reserved fields

Field	Illustrate
Card number	Card number issued by All-in-one card
Inbound coding	Card number issued by All-in-one card
Inbound site name	Site name after matching inbound codes
Outbound encoding	Station code when leaving the station
Outbound site name	Site name after matching outbound encoding
Check-in card time	Specific time for passengers use their cards to access the station
Check-out card time	Exact time for passengers use their cards to exit the station

Data preprocessing

The cleaning of the original data set is mainly carried out from two aspects: one is to delete the records that are not within the operating time range of the line (6:00 ~ 22:00). The second is to fill the missing data, this experiment adopts the mean value method, that is, the average value of the previous 7 days of historical data of the site at the same time is used to fill in. The formula is:

$$X_{dt} = \frac{\mathrm{X}_{(d-7)t}+\mathrm{X}_{(d-6)t}+\cdots+\mathrm{X}_{(d-2)t}+\mathrm{X}_{(d-1)t}}{7} \quad (12)$$

X_{dt} is the filling value, and $\mathrm{X}_{(d-i)t}$ is the passenger traffic of this site at t time i days before.

This experiment uses hour as the time granularity for passenger flow prediction. Therefore, the pre-processed data need to be summarized by hour. Table 2 shows the passenger flow data for some stations from 6am to 10pm on September 16, 2022.

Table 2. Partial raw data

Date	…	West Third Ring	Erqi Square	…	Zijing Mountain	…
2022-09-16-06	…	349	1542	…	864	…
2022-09-16-07	…	1488	2846	…	2001	…
2022-09-16-08	…	990	3076	…	2350	…
…	…	…	…	…	…	…
2022-09-16-20	…	385	2056	…	1395	…
2022-09-16-21	…	296	1712	…	1521	…
2022-09-16-22	…	117	326	…	536	…

Station passenger flow has the characteristics of diversity, the Erqi Square station is one of the most prosperous commercial areas, and there are many residential neighborhoods around. Therefore, this station is a comprehensive station with a variety of passenger flow distribution, which has a high research value. At the same time, to test the generalization ability of the TCN-BiLSTM, passenger flow prediction is also carried out for other 29 stations on the whole route.

Standardized data processing

Data standardization is a preprocessing method that ensures that all data is converted to unitless, uniform scale values by scaling data values [24]. Among the many standardization techniques, data normalization is particularly common, this method remaps data into the interval [-1, 1]. In this way, the deviation caused by different magnitude or units of the original data can be eliminated, so that the data can be analyzed under a unified comparison basis.

The experiments in this thesis use the Max-Min method to normalize the data, and the formula is:

$$\mathrm{x}' = \frac{\mathrm{x}-\mathrm{Min(x)}}{\mathrm{Max(x)}-\mathrm{Min(x)}} \tag{13}$$

Model Evaluation

To verify the prediction performance and the degree of superiority of this model, MAE, MAPE, RMSE and R2 are selected as evaluation indexes [25,26]. Among them, The value of MAE is a non-negative value, which is the average absolute error between the predicted value and the true value, and can better reflect the reality of the prediction error. MAPE is a dimensionless value, which can reflect the reliability and error level of the prediction model. Theoretically, the smaller the value of MAPE, the better the model fits and the higher the accuracy [27]. RMSE is used to detect the deviation between the predicted value and the true value of the model and reflect the dispersion of the deviation distribution [28]. The lower the values of the aforementioned three evaluation metrics, the more accurate the prediction. However, for data sets of different dimensions, they will have different evaluation criteria because of the size of the data value. Therefore, R2 can provide the same measurement standard for different data sets in view of the above situation [29,30]. The value range of R2 is [0,1], when R2=1, that the predicted value is exactly equal to the true value, without any error, and it is the best prediction model. Of course, this is the ideal situation. In general, the larger the R2, thebetter the model fit. The formulas is:

$$MAE = \frac{\sum_{t=1}^{n}|y_p(t)-y_a(t)|}{n} \tag{14}$$

$$MAPE = \frac{\sum_{t=1}^{n}\left|\frac{y_a(t)-y_p(t)}{y_a(t)}\right|}{n} \tag{15}$$

$$RMSE = \sqrt{\frac{\sum_{t=1}^{n}\left(y_a(t)-y_p(t)\right)^2}{n}} \tag{16}$$

$$R^2 = 1-\frac{\sum_{t=1}^{n}\left(y_p(t)-y_a(t)\right)^2}{\sum_{t=1}^{n}\left(y_p(t)-\bar{y}\right)^2} \tag{17}$$

Where, $y_p(t)$ and $y_a(t)$ are the predicted and true values at time t, $\bar{y}$ and n are the mean true value and sample size of the predicted sample, respectively.

Model Hyperparametric Training

The purpose of training the model is to obtain the optimal parameter combination by adjusting the model parameters multiple times, achieve the highest learning and prediction performance of the model. The parameters can be divided into two categories: One type automatically learns from data during model training and ultimately obtains the optimal parameter values; the other category cannot be obtained by automatic learning from the data and is to be sought in the training process of the model. These parameters are usually referred to as hyperparameters, such as batch size, hidden layers, learning rate, and other parameters in LSTM models. These hyperparameters require experience or multiple experiments to obtain optimal values. The following is the optimization training of the main hyperparameters.

BiLSTM parameter settings

The parameters in LSTM play a crucial role in prediction accuracy. Therefore, this thesis uses grid search to optimize the model parameters to find the optimal parameter configuration. Due to the limited space of this thesis, this thesis only describes the number of hidden layer neurons, input step size and the optimization process of hidden layer number of LSTM model.

(1) Number of hidden layer neurons

The number of hidden layer neurons in the LSTM model has different effects on the performance and learning ability of the model. Increasing the number of neurons increases the expressiveness and complexity of the model. Conversely, if the number of neurons is too few, the model may not adequately capture the complex relationships between the input data, limiting its fitting ability and predictive accuracy. In order to verify the influence of the number of neurons in the hidden layer on the prediction accuracy of the model, this thesis designed different hidden layer numbers for model training. Table 3 is the training results. Figure 6 illustrates the visualized prediction results of this model under different numbers of neurons.

Table 3. Training results of different numbers of neurons

Number of neurons	RMSE	MAE	MAPE(%)	R^2
16	126.622	78.997	13.244	0.862
32	**108.101**	**64.511**	**10.975**	**0.900**
64	110.750	65.267	10.255	0.884
128	116.464	66.776	10.399	0.895

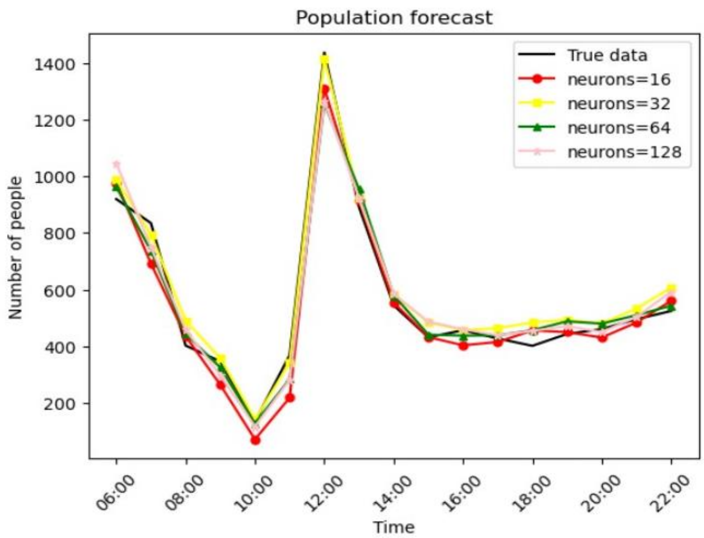

Figure 6. Comparison chart of prediction results with varying numbers of neurons

After analyzing the training results, the error of the model is minimized when input step size is 32.

(2) Input step size

Table 4. Prediction results for different input step sizes

Input step size	RMSE	MAE	MAPE(%)	R^2
2	128.589	75.684	12.128	0.857
3	**107.176**	**64.258**	**10.987**	**0.901**
5	120.765	70.501	11.479	0.875
7	120.314	70.990	12.309	0.876

The input step size is also the input nodes number. When LSTM performs data prediction, the number of input layers determines the dependency length of forecast output data on time series. If the number of nodes is too small, the model will be unable to learn more historical information and reduce the prediction granularity. If the number of nodes is too large, the calculation amount of the model will be increased, and the performance of the model will be reduced. In this thesis, multiple groups of input node points are set for optimization training. Table 4 is the training results.

After analyzing the training results, the error of the model is minimized when input step size is 3.

(3) Number of hidden layers

The hidden layers in LSTM have stronger nonlinear processing ability, and the more hidden layers, the deeper the model can learn and capture the more complex and long-distance dependencies. But the more hidden layers, it can lead to overfitting phenomenon, but also increase the training time of the model. To verify the influence of hidden layers on model prediction accuracy, this thesis designed three different hidden layers for model training. Table 5 is the training results. Figure 7 illustrates the visualized prediction results of this model under varying numbers of hidden layers.

Table 5. Training results of different number hidden layers

Hidden Layers	RMSE	MAE	MAPE(%)	R^2
1	115.123	68.360	10.840	0.886
2	**105.883**	**63.197**	**10.705**	**0.904**
3	114.500	67.052	10.132	0.887

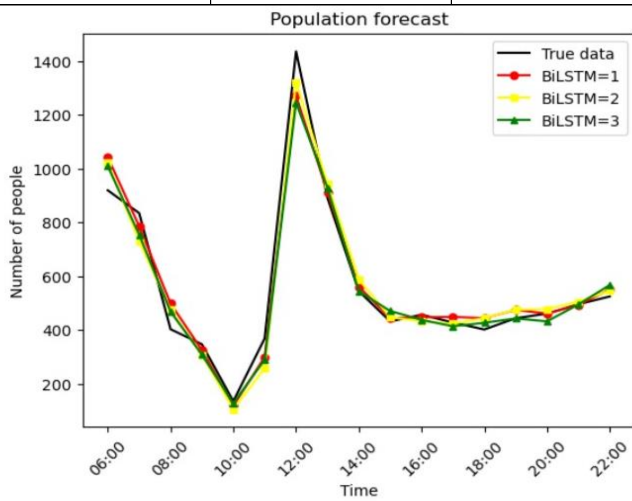

Figure 7. Comparison of prediction results of different hidden layers

After analyzing the training results, the error of the model is minimized when the hidden layers is 2.

Configuring TCN model parameters

(1) Atrous coefficient

Table 6. Training results of different atrous coefficients

Atrous Coefficient	RMSE	MAE	MAPE(%)	R^2
1	117.611	74.831	12.440	0.881
2	114.835	67.674	10.731	0.887
4	**108.641**	**62.394**	**10.128**	**0.899**
8	137.108	79.318	11.899	0.839

Atrous Convolution is introduced into the TCN model in order to obtain longer dependencies. Atrous convolution extends the receptive field of TCN by sampling the input of the convolution at intervals, where the sampling rate is determined by the cavity coefficient d. The value of d determines the length of the history information that the output data can depend on. If the value of d is too small and the historical information on which the output data depends is too short, more data dependencies cannot be obtained. If it is too long, it will increase the noise of the data and reduce the prediction accuracy because it relies on too much historical information. In this thesis, several groups of data are selected for training. Table 6 is the training results. Figure 8 illustrates visualized prediction results of TCN under different dilation rates.

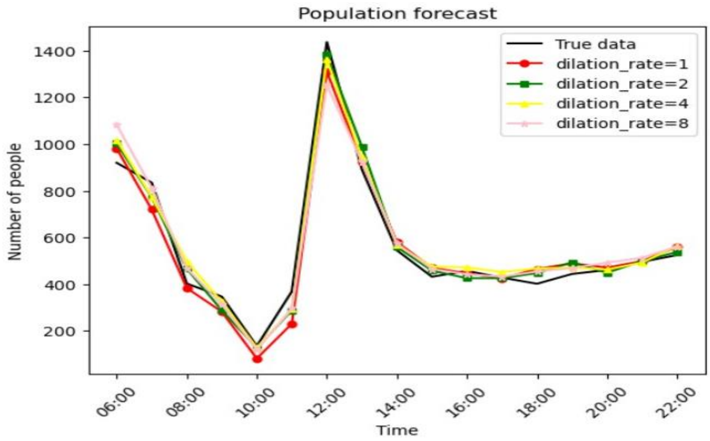

Figure 8. Comparison of prediction results of different void coefficients

(2) Convolution kernel size

Theoretically speaking, the larger the convolution kernel of TCN, the more comprehensive the features extracted, but it increases the calculation amount of the model and reduce the efficiency. The smaller the convolution kernel, although more detailed features can be extracted, some global information will also be lost. Therefore, selecting the appropriate convolution kernel size is very important to improve the accuracy and efficiency. Different convolution kernels are selected for model training. Table 7 is the training results. Figure 9 illustrates the visualized prediction results of this model under different kernel sizes.

Table 7. Training results of different convolutional kernels

Convolutional kernel	RMSE	MAE	MAPE(%)	R^2
16	120.468	71.437	10.582	0.875
32	108.908	64.240	10.383	0.898
64	**103.105**	**61.824**	**10.195**	**0.909**
128	112.265	67.260	11.214	0.893

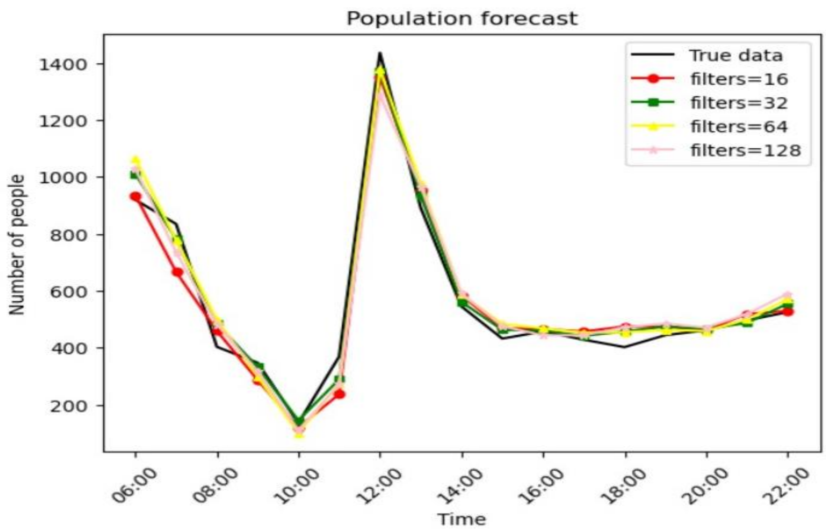

Figure 9. Comparison of prediction results of different convolution kernel sizes

After analyzing the training results, the error of the model is minimized when the convolution kernel size is 64.

Analysis of Metro passenger flow forecast results

To validate the performance of TCN-BiLSTM, this thesis selected AFC data of Metro Line 1 as the basis, and selected the credit card data from January 22, 2022 to December 17, 2022 for simulation experiment, with the data as the time interval of hours. In the simulation experiment, the first 70% of the data set is selected as the training set to train the model and find the optimal parameter combination. The last 30% is a test set that evaluates the performance of the model. To reduce the potential impact of random errors on the accuracy, five tests were carried out for each parameter configuration in the simulation experiment, and the average prediction error of the test was taken as the final error value under the parameter configuration. Figure 10 illustrates of the visualized prediction results of TCN-BiLSTM.

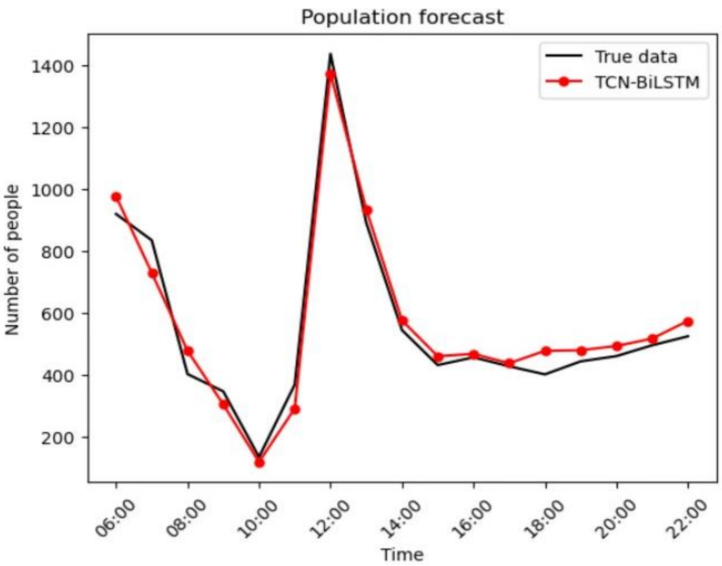

Figure 10. Prediction Results of TCN-BiLSTM

To further verify the validity of TCN-BiLSTM, this thesis also selected the traditional single model TCN, LSTM, BiLTSM and the combined model TCN-LSTM for ablation experiments, Figure 11 shows the results.

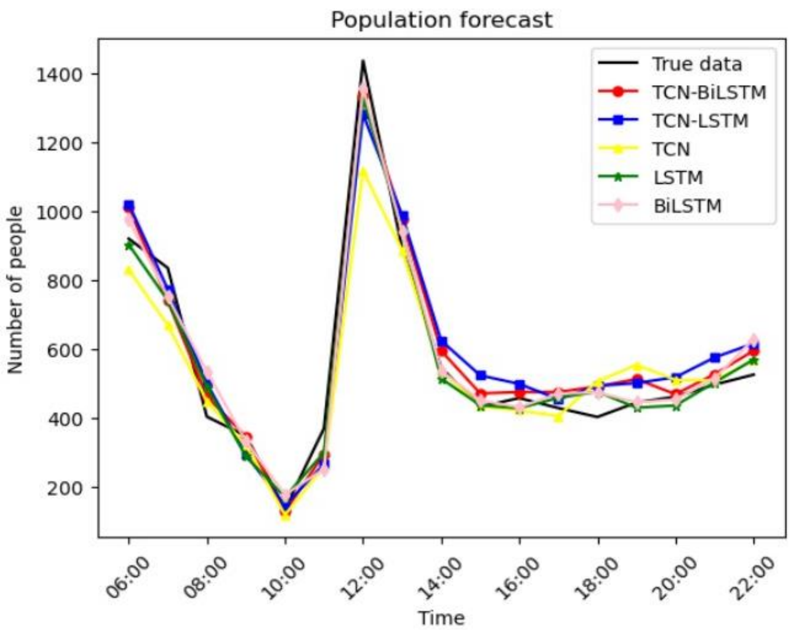

Figure 11. Comparison chart of multi-model prediction results

Table 8 is the performance indicators of the multiple models. As can be seen from Table 8, the error of TCN-BiLSTM introduced in this thesis is the smallest, which is 10.3 lower in RMSE, 6.6 lower in MAE, and 1.9 lower in MAPE than that of the second-best combined model TCN-LSTM.

Table 8. Prediction performance index values of ablation experiment model

Prediction Model	RMSE	MAE	MAPE(%)	R^2
TCN	139.028	92.067	15.934	0.834
LSTM	122.705	74.997	13.074	0.851
BiLSTM	123.257	73.555	12.938	0.869
TCN-LSTM	115.802	70.678	12.743	0.885
TCN-BiLSTM	**105.505**	**64.083**	**10.880**	**0.904**

After analyzing the data in Table 8, that the R2 corresponding to the TCN-BiLSTM model is 0.904, which is closest to 1. This further indicates that the introduced TCN-BiLSTM has the best prediction effect and the highest degree of fit, which also reflects the effectiveness and accuracy of the TCN- BiLSTM in predicting urban rail transit passenger flow.

To verify the predictive performance of the introduced model at different stations, simulation experiments were conducted on all 30 stations on Metro Line 1. Table 9 shows the predictive results of some stations. From the table, it can be seen that the TCN-BiLSTM has good predictive performance in predicting passenger flow data at different stations. This also indicates that the model can be applied to different scenarios and has good portability.

Table 9. Predictive performance index values for different sites

Site Name	prediction model	RMSE	MAE	MAPE(%)	R^2
May Day Park	TCN	145.489	94.125	12.392	0.859
	LSTM	128.258	85.458	14.988	0.890
	BiLSTM	128.577	82.065	12.712	0.890
	TCN-LSTM	116.612	73.441	9.893	0.909
	TCN-BiLSTM	**109.382**	**72.755**	**9.607**	**0.920**
Dongfeng South Road	TCN	149.054	101.031	12.317	0.878
	LSTM	147.306	100.991	13.438	0.906
	BiLSTM	147.164	99.258	13.070	0.906
	TCN-LSTM	145.364	91.525	10.693	0.919
	TCN-BiLSTM	**132.233**	**80.181**	**9.873**	**0.923**

CONCLUSIONS

The TCN-BiLSTM proposed in this thesis, and predicts the passenger flow of metro rail passenger traffic, while considering the nonlinear and periodic characteristics of passenger flow. First, the time sequence characteristic information of passenger traffic is extracted using the time convolutional network. Secondly, using the advantages of BiLSTM, the relationship between the front and back of the passenger flow sequence is extracted. Finally, the attention mechanism is introduced to assign weight to different features to improve the predictive performance of model.

(1) The TCN-BiLSTM introduced in this thesis fully explores the nonlinear and periodic characteristics of metro rail passenger traffic passenger flow, and achieves good prediction results under different application scenarios, verifying the effectiveness and portability.

(2) Comparing the TCN-BiLSTM with the traditional TCN, LSTM, BiLTSM and combined model TCN-LSTM, the TCN-BiLSTM prediction model showed significant advantages in predicting subway passenger flow, with the RMSE decreased by at least 7.23, the MAE decreased by at least 0.68, MAPE decreased by at least 0.86% and R2 increased by at least 0.011. This result fully proves the effectiveness of the forecast model designed in the subway passenger flow prediction, showing its excellent forecasting performance. The TCN-BiLSTM has the lowest prediction error and the best fit.

Although this model has achieved good prediction results, there are many further improvements to be made. It is still a long way to predict the passenger flow of metro rail passenger traffic. In the later stage, we can also be carried out from the following aspects:

(1) Suggestions for improving the model and algorithm: Due to sample data limitations, we can currently only obtain passenger flow information for each station on a line, and do not include data for all lines in the city. Therefore, we cannot obtain the overall passenger flow information for the city. On the basis of obtaining the overall historical data, the model can be adjusted later to consider the impact between lines and stations, because the passengers flow of a certain station will be affected by the flow of passengers getting on and off the station. At the same time, the passenger flow on ordinary weekdays, weekends and holidays should be analyzed to establish different prediction models. The improved model can better improve prediction accuracy and provide travelers with more accurate route options.

(2) Improvements on factors affecting passenger flow: When making predictions in this article, the impact of weather factors is based on days as the time unit. However, in real life, the weather of a day is not fixed and is likely to change. When collecting weather information in the later stage, weather data with a smaller time granularity is supplemented to better reflect the impact of weather on passenger flow in the model. This article only considers the impact of weather on passenger flow, but large-scale activities, whether the site is located in a residential area, whether there are schools or factories nearby, whether it is in a commercial district, etc. will all have an impact on the passenger flow of the site. It is recommended that various influencing factors such as geography and surrounding business values be added to achieve better forecasting results.

(3) On the basis of improving the predictive performance of the model, we plan to design and develop a set predict the passenger flow of metro rail passenger traffic system. In this system, real-time or non timed passenger flow monitoring is implemented for each station and route based on the data transmitted by the AFC system, and the passenger flow information of each station is visualized and displayed. Using the predictive data provided by this system, it can also provide data support for operators to schedule vehicles, adjust operation plans, and allocate security personnel, thereby improving subway transportation efficiency and service quality, as well as enhancing the comfort of passengers.

ACKNOWLEDGMENTS

1. Natural science research project of colleges and universities in Anhui Province. NO: KJ2021A1159

2. 2023 Anhui Province College Student Innovation and Entrepreneurship Project. NO: 202312216027

3. 2024Anhui Province College Student Innovation and Entrepreneurship Project. NO: 202412216030

4.2021National first-class undergraduate major construction point: Software Engineering Major. NO: 2020ylzy01

5. Intelligent Information Processing Research and Innovation Team of Anhui Provincial Department of Education. NO: 2024AH0012

REFERENCES

[1] E. Jenelius, "Data-Driven Metro Train Crowding Prediction Based on Real-Time Load Data," in IEEE Transactions on Intelligent Transportation Systems, vol. 21, no. 6, pp. 2254-2265, June 2020, doi: 10.1109/TITS.2019.2914729.

[2] M. Ni, Q. He and J. Gao, "Forecasting the Subway Passenger Flow Under Event Occurrences With Social Media," in IEEE Transactions on Intelligent Transportation Systems, vol. 18, no. 6, pp. 1623-1632, June 2017, doi: 10.1109/TITS.2016.2611644.

[3] Vlahogianni E I, Karlaftis M G, Golias J C, "Short-term traffic forecasting: Where we are and where we're going," Transportation Research Part C Emerging Technologies, 2014, 43(1).

[4] W. C. Duan and Z. Gong, "Research and application of urban real-time traffic flow prediction based on STARIMA," 2022 IEEE International Conference on Advances in Electrical Engineering and Computer Applications (AEECA), Dalian, China, 2022, pp. 1521-1525.

[5] D.X Ye, R.B Han, L.H Yan., "ISSA-SVR Model for Short-term Traffic Flow Prediction", Computer Engineering and Design, 2019, 45(02):608-617.
[6] W.J. Zhang, H.Z.H. Yang, B. Zhang, et al., "Short-term passenger flow prediction model of urban rail transit considering multi-time scale characteristics", Transportation Systems Engineering and Information, 2022, 22(06):212-223.
[7] X.Y. Kuang, Y.M. Xu, "Traffic flow prediction based on adaptive spatiotemporal graph convolutional cyclic network", Journal of Wuhan University of Technology (Traffic Science and Engineering Edition): 2024, 05(13):1-8.
[8] G.W. Dai, C. Ma and X. Xu, "Short-Term Traffic Flow Prediction Method for Urban Road Sections Based on Space–Time Analysis and GRU," in IEEE Access, vol. 7, pp. 143025-143035, 2019.
[9] Q.Y Lian, W. Sun, R. S. Li, "Stacked LSTM short-time ship traffic flow prediction model based on attention mechanism", Journal of Dalian Maritime University, 2024, 50(01):57-65.
[10] D.M Sun, CH, ZH Chen, Y, P Sun, "Life Prediction of Rolling Bearing Based on CNN and TCN Neural Network", Machinery Design & Manufacture: 2024, 05(17):1-10.
[11] T. Selim, I. Elkabani and M. A. Abdou, "Students Engagement Level Detection in Online e-Learning Using Hybrid EfficientNetB7 Together With TCN, LSTM, and Bi-LSTM", IEEE Access, vol. 10, pp. 99573-99583, 2022, doi: 10.1109/ACCESS.2022.3206779.
[12] K. M. Karthick Raghunath et al., "Redefining Urban Traffic Dynamics With TCN-FL Driven Traffic Prediction and Control Strategies", IEEE Access, vol. 12, pp. 115386-115399, 2024, doi: 10.1109/ACCESS.2024.3443298.
[13] Zhao Yu, Chen Lixia, LiANG Mengjiao, "Time probability prediction and meteorological warning modeling of rain-type landslide based on LSTM_TCN model", Bulletin of Geological Science and Technology, 2019, 43(02):201-214.
[14] Zhang Tao, "Research on short-term passenger flow prediction of urban rail transit based on TCN-LSTM combination model", Beijing University of Chemical Technology, 2024.
[15] D. Ma, X. Song and P. Li, "Daily Traffic Flow Forecasting Through a Contextual Convolutional Recurrent Neural Network Modeling Inter- and Intra-Day Traffic Patterns," IEEE Transactions on Intelligent Transportation Systems, vol. 22, no. 5, pp. 2627-2636, May 2021.
[16] Wang Y, Wu H, Zhang J, et al., "PredRNN: a recurrent neural network for spatiotemporal predictive learning", IEEE Transactions on Pattern Analysis and Machine Intelligence, 2023, 45(2): 2208-2225.
[17] P. Ray, B. Ganguli and A. Chakrabarti, "A Hybrid Approach of Bayesian Structural Time Series With LSTM to Identify the Influence of News Sentiment on Short-Term Forecasting of Stock Price," IEEE Transactions on Computational Social Systems, vol. 8, no. 5, pp. 1153-1162, Oct. 2021, doi: 10.1109/TCSS.2021.3073964.
[18] C.-W. Tsai, C.-H. Hsia, S.-J. Yang, S.-J. Liu and Z.-Y. Fang, "Optimizing hyperparameters of deep learning in predicting bus passengers based on simulated annealing", Appl. Soft Comput., vol. 88, Mar. 2020.
[19] R. Joshi, J. Ghosh, N. Kalani and R. L. Tanna, "Assessment of Stacked LSTM, Bidirectional LSTM, ConvLSTM2D, and Auto Encoders LSTM Time Series Regression Analysis at ADITYA-U Tokamak," IEEE Transactions on Plasma Science, vol. 52, no. 7, pp. 2403-2409, July 2024.
[20] Chen Li, Zheng Lin-jiang and Yang Jie, et al., "Short-term traffic flow prediction: from the perspective of traffic flow decomposition", Neurocomputing, 2020,413, 444-456.
[21] P. Noursalehi, H. N. Koutsopoulos and J. Zhao, "Dynamic Origin-Destination Prediction in Urban Rail Systems: A Multi-Resolution Spatio-Temporal Deep Learning Approach," IEEE Transactions on Intelligent Transportation Systems, vol. 23, no. 6, pp. 5106-5115, June 2022.
[22] A. Parmar, D. K. A, A. Chouhan and K. Captain, "Dual-Stream CNN-BiLSTM Model with Attention Layer for Automatic Modulation Classification," 2023 15th International Conference on COMmunication Systems & NETworkS (COMSNETS), Bangalore, India, 2023, pp. 603-608.
[23] S Subbiah, S Paramasivan, K Arockiasamy, S Senthivel and M Thangavel, "Deep Learning for Wind Speed Forecasting Using Bi-LSTM with Selected Features", Intell Autom Soft Comput, vol. 35, pp. 3829-44, 2022.

[24] R. Sutopo, J. M. -Y. Lim and V. M. Baskaran, "Efficient Long-Term Dependencies Learning for Passenger Flow Prediction With Selective Feedback Mechanism," in IEEE Transactions on Intelligent Transportation Systems, vol. 23, no. 12, pp. 24020-24030, Dec. 2022.

[25] Y. -S. Jeong, Y. -J. Byon, M. M. Castro-Neto and S. M. Easa, "Supervised Weighting-Online Learning Algorithm for Short-Term Traffic Flow Prediction," in IEEE Transactions on Intelligent Transportation Systems, vol. 14, no. 4, pp. 1700-1707, Dec. 2013, doi: 10.1109/TITS.2013.2267735.

[26] Wang Xiao-quan, Shao Chun-fu, Yin Chao-ying, et al., "Short term traffic flow forecasting method based on ARIMA- GARCH-M model", Journal of Beijing Jiaotong University, 2018, 42(4): 79-84.

[27] Li M Q, Hong Z X, Chen L, et al., "Temporal multi-graph convolutional network for traffic flow prediction", IEEE Transactions on Intelligent Transportation Systems, 2021, 22(6): 3337-3348.

[28] Ke Rui-min, Li Wan, Cui Zhi-yong, et al., "Two-stream multi-channel convolutional neural network for multi-lane traffic speed prediction considering traffic volume impact", Transportation Research Record, 2020, 2674(4): 459-470.

[29] Cui Z Y, Henrickson K, Ke R M, et al., "Traffic graph convolutional recurrent neural network: a deep learning framework for network-scale traffic learning and forecasting", IEEE Transactions on Intelligent Transportation Systems, 2020, 21(11):4883-4894.

[30] Q. Zhaowei, L. Haitao, L. Zhihui and Z. Tao, "Short-Term Traffic Flow Forecasting Method With M-B-LSTM Hybrid Network," in IEEE Transactions on Intelligent Transportation Systems, vol. 23, no. 1, pp. 225-235, Jan. 2022.