Deep Learning-Based Prediction and Analysis of Highway Traffic Flow near Airports

Boyuan Jiang
College of Computer Science, Sichuan University, Sichuan, 610207, China

Keywords: Highway, LSTM, Large Airport, Prediction, Deep Learning.

Abstract: The surrounding highways of large airports play a crucial role in traffic, making it essential to accurately predict traffic flow. In this study, the Long Short-Term Memory (LSTM) model was employed as the data prediction model to forecast data from six stations on the M25 highway near London Heathrow Airport in August 2019. The LSTM model utilized a prediction interval with a time slot length of 5, and error analysis was conducted. The final predictions revealed a bimodal pattern in daily traffic volume on the highway, with a unimodal pattern in average vehicle speed. On highway ramps, daily traffic volume exhibited a multimodal pattern, and although average vehicle speed displayed slight fluctuations, it remained relatively stable overall. Furthermore, error analysis indicated that the LSTM model demonstrated a good fit and produced satisfactory prediction results. This paper has the potential to greatly contribute to the improvement and enhanced management of highway traffic surrounding large airports.

1 INTRODUCTION

With the rapid development of globalization and urbanization, smart transportation has become a significant issue in modern society. Highways, as crucial components of urban transportation networks, play an especially vital role near airports. For instance, on August 6, 2023, Beijing Daxing International Airport witnessed a daily passenger count exceeding 155,000 (Liu 2023). Moreover, from January 1 to October 20, 2023, Guangzhou Baiyun Airport served 50.0889 million passengers, marking a 111.95% year-on-year increase (Qian 2023). Consequently, rational prediction and efficient management of traffic flow on highways near airports are of utmost importance for ensuring urban traffic safety and smoothness.

Traffic flow prediction, a foundational technology in intelligent transportation systems, is crucial for traffic control and guidance (Pang et al 2019). Traditional traffic flow prediction methods usually rely on vehicle speed and trajectory data. However, these methods are not suitable for urban roads due to high population density and complex traffic conditions, making it impractical to deploy sensors at scale for collecting necessary traffic data (Li et al 2020). Furthermore, due to the non-stationary, non-periodic nature of traffic flow sequences, coupled with the influence of factors like holidays, prediction becomes particularly challenging (Ding et al 2019). Thus, traditional traffic flow prediction methods often prove inadequate for highways near airports.

Short-term traffic flow prediction is characterized by high uncertainty. To design highly accurate prediction methods, deep learning is the prevailing direction (Zhao et al 2019). The rapid advancement of deep learning has led to the utilization of various models for short-term traffic flow prediction, such as Autoregressive Integrated Moving Average (ARIMA), Artificial Neural Networks (ANN), Convolutional Neural Networks (CNN), and Long Short-Term Memory networks (LSTM). The introduction of Long Short-Term Memory neural networks has significantly enhanced the capability of traffic flow prediction. LSTM models, specifically designed for time series data, offer advantages in capturing time-dependent dependencies and non-linear relationships. They excel in extracting time series features, leading to higher prediction accuracy, and making them well-suited for short-term traffic flow prediction on highways (Zhang & Gong 2022). Previous research has compared ARIMA, LSTM, and Prophet time series forecasting algorithms, revealing that all three models perform well in traffic flow prediction. However, LSTM excels in terms of fitting,

Jiang, B.
Deep Learning-Based Prediction and Analysis of Highway Traffic Flow near Airports.
DOI: 10.5220/0012886200004547

In *Proceedings of the 1st International Conference on Data Science and Engineering (ICDSE 2024)*, pages 387-393
ISBN: 978-989-758-690-3

prediction accuracy, and generalization, while offering greater flexibility in the setting of influencing factors (Zhou & Xu 2021).

This paper aims to construct a short-term traffic flow prediction model based on LSTM within the framework of deep learning for the analysis and prediction of traffic flow on highways near major airports. This paper focuses on analysing and predicting historical traffic flow data from the M25 highway near London Heathrow Airport, utilizing the LSTM model. The objective is to enhance the accuracy and robustness of traffic flow prediction, better cope with fluctuations in traffic flow near airports, reduce traffic congestion, improve traffic efficiency, and enhance the travel experience for city residents and travelers, ultimately supporting better decision-making in traffic management. Additionally, this paper conducts an in-depth analysis of the model, exploring its performance and limitations in various scenarios. This deepens the understanding of the application of deep learning in the field of traffic, providing valuable insights and experiences that can contribute to the development of future traffic management and intelligent transportation systems.

2 METHODS

2.1 Data Sources

This paper utilized the UK highway dataset for analysis (https://webtris.highwaysenglan d.co.uk/). The dataset contains various information about the UK highway system, such as traffic flow data, road conditions, construction and maintenance projects, geographic information, and data time range. The focus of this research is traffic flow prediction and analysis, so data from six sites near Heathrow Airport on the M25 highway were selected for one month, spanning from August 1, 2019, to August 30, 2019. Site information is presented in Table 1.

Table 1: Site Information Table.

Legacy MIDAS ID	Site Name
30022731	MIDAS site at M25/4883A priority 1 on link 199131002; GPS Ref: 502104;172197; Clockwise
30025228	MIDAS site at M25/4909A priority 1 on link 200045638; GPS Ref: 503070;174470; Clockwise
30025227	MIDAS site at M25/4916A priority 1 on link 200045691; GPS Ref: 503510;175090; Clockwise
30027351	MIDAS site at M25/4926K priority 1 on link 200045818; GPS Ref: 503898;176048; Clockwise
30032052	MIDAS site at M25/7108B priority 1 on link 200045820; GPS Ref: 503800;176100; Clockwise
30025505	MIDAS site at M25/4936A priority 1 on link 200045641; GPS Ref: 504127;177021; Clockwise

To ensure the accuracy and completeness of the data, it is essential to conduct data preprocessing in the research process. The presence of a large amount of redundant data can increase memory consumption during subsequent model training, incurring unnecessary costs while diminishing model quality. Missing or improperly processed data can lead to program failures and inaccurate model predictions. Therefore, for the algorithm to be effective, it is necessary to use accurate data without missing values for forecasting and analysis.

In this study, data preprocessing was performed on a dataset containing traffic flow data from six sites over one month. Given the research focus on traffic flow prediction near a major airport on the highway, the study retained only time data, the number of vehicles, and average vehicle speed within 15-minute intervals. Regarding handling missing values, this paper employed either the value from the previous time point or the subsequent time point for filling in the gaps. Table 2 provides explanations of relevant variables.

Table 2: Explanation of relevant variables.

variable name	Type	explain
datetime	string	datetime
total_flow	float64	The number of vehicles passing through this station within a 15-minute interval.
speed	float64	The average vehicle speed passing through this station within a 15-minute interval.
month	string	month
day	string	day
hour	string	hour
minute	string	minute

2.2 Preliminary Analysis of Data

After preprocessing a one-month traffic volume dataset from six sites, the first step involves calculating the average number of vehicles passing through the site within a 15-minute interval and the average vehicle speed passing through the site during the same 15-minute interval. The results are shown in Table 3.

Table 3: The average number of vehicles and the average vehicle speed.

Legacy MIDAS ID	Flow	Speed
30022731	899.68	92.48
30025228	1074.71	81.27
30025227	841.22	76.84
30027351	321.85	68.14
30032052	97.65	69.50
30025505	1087.48	78.92

The second step is to extract and analyze temporal features on the M25 motorway for the respective sites. Taking the data from the "M25/4883A priority 1 on link 199131002" site as an example, this paper designates the data from the previous 25 days as the training set. Figure 1 depicts a line graph illustrating the fluctuations in the number of vehicles passing through the site at 15-minute intervals within the training set. This graph displays how the number of vehicles passing through the site at 15-minute intervals fluctuates over time. These 25 days can be roughly divided into 25 cycles, with peaks and troughs periodically alternating. The daily trends and patterns are generally consistent, although the numerical values of the peaks and troughs may sometimes exhibit significant differences.

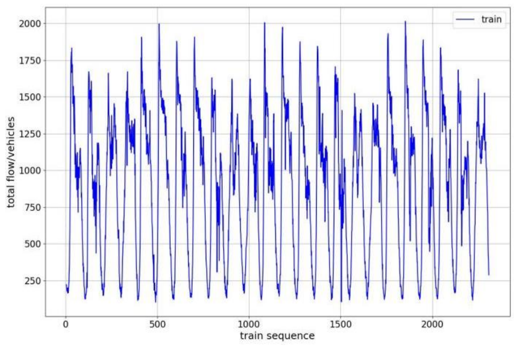

Figure 1: Sequence diagram of vehicle count fluctuations (Original).

Similarly, Figure 2 depicts a line graph of fluctuations in vehicle speed within the training set. This graph illustrates the temporal fluctuations in the average vehicle speed of vehicles passing through the station at 15-minute intervals. These 25 days can be roughly divided into 25 cycles, with peaks and troughs alternating periodically. The daily trends and patterns are generally consistent, but the numerical values of the peaks and troughs may sometimes exhibit significant variations.

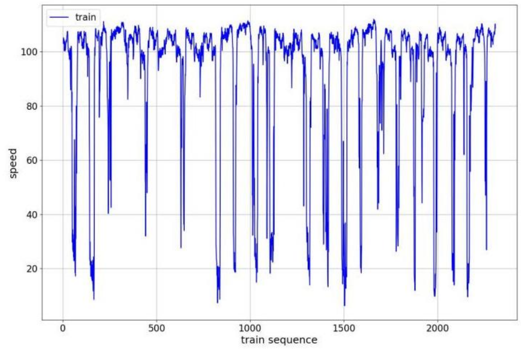

Figure 2: Sequence diagram of speed fluctuations (Original).

2.3 LSTM-Based Model

LSTM is a type of time-recursive neural network, which is an improvement upon recurrent neural networks (Recurrent Neural Network, RNN) (Zhao & Zhang 2018). LSTM addresses the issues of gradient explosion and long-term data dependencies that exist in RNNs (Yang et al 2017). The internal structure of LSTM is illustrated in Fig.3.

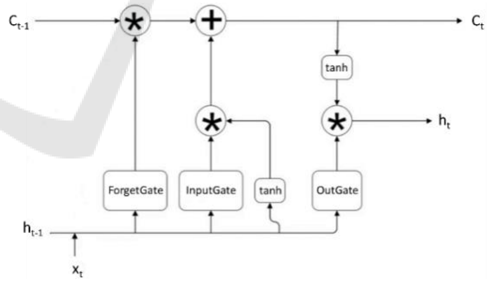

Figure 3: LSTM Flowchart (Picture credit: Original)

The input gate selectively stores new information and replaces forgotten information from the forget gate. The output gate determines which information can be outputted in the current state. The forget gate is responsible for discarding information that is no longer needed. LSTM is particularly suitable for processing and predicting time series data because it can handle uncertain time lags between significant events in the sequence.

The formula for the forget gate in an LSTM:

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

The formula for the input gate in an LSTM:

$$i_t = \sigma \ (W_i \ [h_{t-1}], x_t + b_i \qquad (2)$$

$$\tilde{C}_t = \tanh \ W_c \ [h_{t-1}, X_t] + \ b_c \qquad (3)$$

The formula for the output gate in an LSTM:

$$o_{t=} \ \sigma \ (W_O \ [h_{t-1}] + b_o \qquad (4)$$

$$h_t = \ o_t \tanh \ (c_t) \qquad (5)$$

Where, f_t is the forget gate unit, it is the input gate unit, it is the input gate unit, O_t is the output gate unit, and h_{t-1} is the hidden layer state. W_f, W_i, W_C, W_o are weight matrices, and b_f, b_i, b_c, b_o are bias vectors. The sigmoid function and tanh function are used in the equations. These equations describe the computational process within an LSTM cell, allowing it to capture and process long-term dependencies in sequential data (Liang et al 2020).

3 RESULTS AND DISCUSSION

3.1 Site Selection

Through the analysis of Table 3, it can be observed that the average number of vehicles passing through the "M25/4936A priority 1 on link 200045641" site within 15 minutes is the highest. On the other hand, the "M25/7108B priority 1 on link 200045820" site has the lowest average number of vehicles passing through it within 15 minutes. Additionally, the "M25/4883A priority 1 on link 199131002" site has the highest average vehicle speed for vehicles passing through it within 5 minutes, while the "M25/4926K priority 1 on link 200045818" site has the lowest average vehicle speed for vehicles passing through it within 15 minutes.

The data for the site "M25/7108B priority 1 on link 200045820" is notably unique, with both the average number of vehicles passing through the site within 15 minutes and the average speed of vehicles passing through within 15 minutes being relatively low. This peculiarity is attributed to the location of the site at the highway ramp, necessitating a separate predictive analysis. For the remaining five sites located along the highway, the site "M25/4936A priority 1 on link 200045641" with the highest traffic volume, and the site "M25/4883A priority 1 on link 199131002" with the fastest average vehicle speed are selected for predictive analysis.

3.2 Prediction Results and Real Results

Firstly, predictive analysis was conducted on the "M25/4936A priority 1 on link 200045641" site with the highest traffic volume. In this study, a two-layer LSTM with 80 neurons in each layer was implemented. The "Dropout" function was added to prevent overfitting. The neural network utilized the "adam" activation function and "mse" as the loss error. Subsequently, input data for the LSTM was created with a prediction interval of 5-time slots, meaning data from the previous 5 time periods were used to predict data for the next period. Finally, calculate the Mean Squared Error (MSE), Root Mean Squared Error (RMSE), and Mean Absolute Error (MAE) for the predicted results. Fig. 4 and Fig. 5 represent graphical illustrations of the predicted results for traffic volume and average vehicle speed.

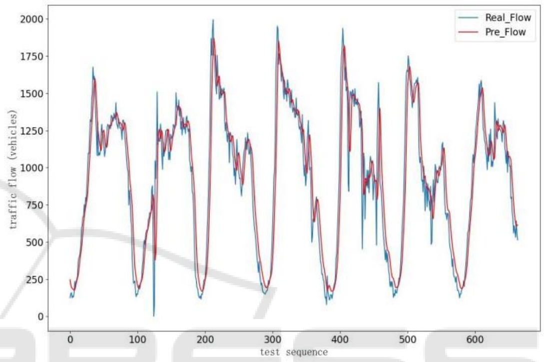

Figure 4: The traffic flow prediction results for the "M25/4936A priority 1 on link 200045641" (Original).

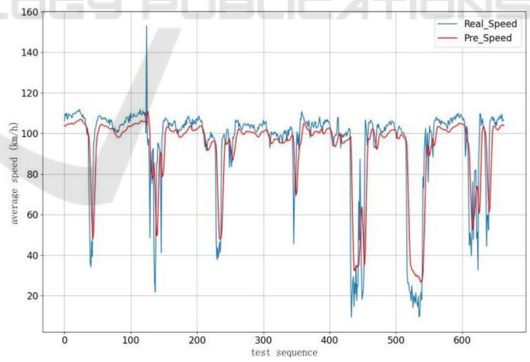

Figure 5: The speed prediction results for the "M25/4936A priority 1 on link 200045641" (Original).

Following the predictive results of the LSTM model, an error analysis was conducted, and Table 4 presents the results of this analysis. According to the error analysis, the MSE, RMSE, and MAE values for both traffic volume and average vehicle speed predictions were relatively small. The high degree of overlap between the predicted curves in Fig. 6 and Fig.7 and the true curves from the test set indicates a good fit and overall satisfactory predictive results.

Table 4: Error analysis for “M25/4936A priority 1 on link 200045641”.

Statistic	Traffic flow	Average speed
MSE	41066.96	131.59
RMSE	202.65	11.47
MAE	149.27	7.85

From the predictive results, it is observed that the traffic volume exhibits roughly bimodal peaks daily, while the average vehicle speed shows a unimodal minimum daily. The traffic volume is generally higher from 6:00 AM to 9:00 PM, while the average vehicle speed is lower from 7:00 AM to 8:00 PM, suggesting a correlation between higher traffic volume and slower average vehicle speed during these periods.

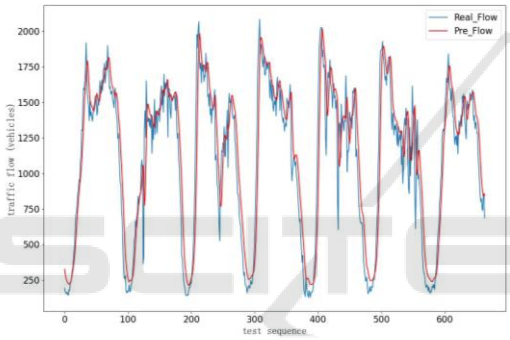

Figure 6: The traffic flow prediction results for the " M25/4883A priority 1 on link 199131002" (Original).

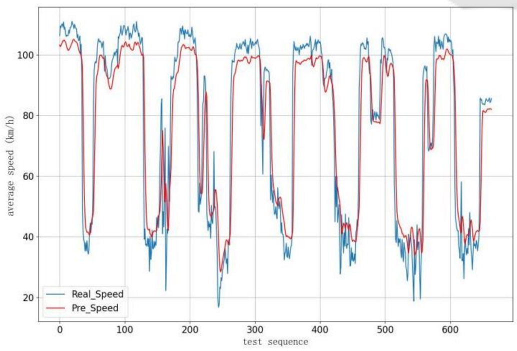

Figure 7: The speed prediction results for the " M25/4883A priority 1 on link 199131002" (Original).

Next, predictive analysis was conducted for the site with the highest average vehicle speed, "M25/4883A priority 1 on link 199131002". The forecasting method employed for this site was identical to the one used for the aforementioned site. Fig. 6 and Fig.7 illustrate the graphical representation of the predicted results for traffic volume and average vehicle speed.

Following the predictive results of the LSTM model, an error analysis was performed, and Table 5 presents the results of this analysis. According to the error analysis, the MSE, RMSE, and MAE values for both traffic volume and average vehicle speed predictions were relatively small. The high degree of overlap between the predicted curves in Fig. 7 and Fig. 8 and the true curves from the test set indicates a good fit and overall satisfactory predictive results.

Table 5: Error analysis for" M25/4883A priority 1 on link 199131002".

Statistic	Traffic flow	Average speed
MSE	29524.20	206.28
RMSE	171.83	14.36
MAE	121.73	8.50

From the predictive results, it is observed that the traffic volume exhibits roughly bimodal peaks daily, while the average vehicle speed shows a unimodal minimum daily. The traffic volume is generally higher from 6:00 AM to 9:00 PM, while the average vehicle speed is lower from 7:00 AM to 8:00 PM, suggesting a correlation between higher traffic volume and slower average vehicle speed during these periods.

Finally, predictive analysis was conducted for the site located at the highway ramp, "M25/7108B priority 1 on link 200045820". The forecasting method employed for this site was the same as the one used for the aforementioned sites. Fig. 8and Fig. 9 depict the graphical representation of the predicted results for traffic volume and average vehicle speed.

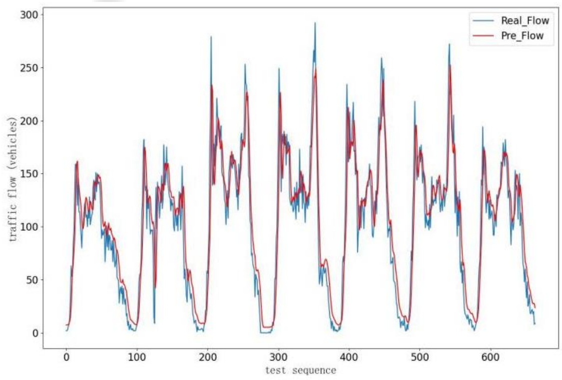

Figure 8: The traffic flow prediction results for the " M25/7108B priority 1 on link 200045820" (Original).

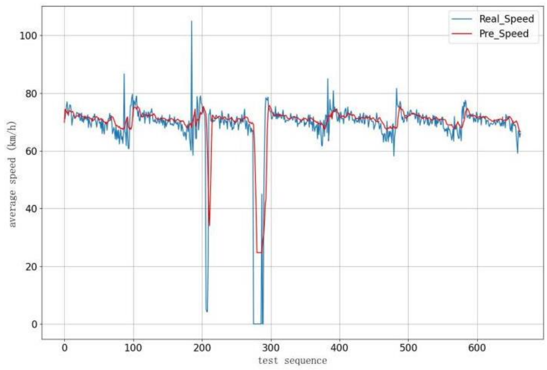

Figure 9: The speed prediction results for the " M25/7108B priority 1 on link 200045820" (Original).

Following the predictive results of the LSTM model, an error analysis was performed, and Table 6 presents the results of this analysis. According to the error analysis, the MSE, RMSE, and MAE values for both traffic volume and average vehicle speed predictions were relatively small. However, an anomaly was observed in the testing sequence from 200 to 300 in Fig. 8 and Fig. 9, where the average vehicle speed dropped to 0. This anomaly may be related to the actual road conditions. Excluding this abnormal portion, the fit was generally good in the remaining intervals, indicating satisfactory predictive results.

Table 6: Error analysis for" M25/7108B priority 1 on link 200045820".

Statistic	Traffic flow	Average speed
MSE	597.59	77.66
RMSE	24.45	8.81
MAE	18.24	3.80

From the predictive results, it is observed that the traffic volume exhibits roughly multi-modal peaks daily, with overall higher values from 6:00 AM to 9:00 PM. The average vehicle speed, except for the abnormal interval, remains relatively stable, fluctuating around 70 km/h daily.

From the above predictive analyses of the three sites, it can be observed that the sites with the maximum traffic volume and the fastest average vehicle speed exhibit similar trends in their predictions. Both traffic volume and average vehicle speed show a daily pattern of roughly bimodal peaks and unimodal minima, likely influenced by peak commuting hours. The overall higher traffic volume from 6:00 AM to 9:00 PM and lower average vehicle speed from 7:00 AM to 8:00 PM may be associated with increased daytime airport operations, with more flights departing and arriving, and fewer runway maintenance activities during the night.

For the site located at the highway ramp, daily traffic volume displays a multi-modal peak pattern, with overall higher values from 6:00 AM to 9:00 PM. This pattern is likely influenced by both peak commuting hours and the airport flight schedule. Apart from the abnormal interval where the vehicle speed is zero, the average vehicle speed remains relatively stable throughout the day, approximately at 70 km/h. This may be related to the speed limits specified for the highway ramp.

These findings suggest that while there are subtle differences in the daily variations of traffic volume and average vehicle speed, the overall trend of higher traffic volume corresponding to slower average vehicle speed is consistent.

The fact that the model's predictive results align with the observed patterns indicates its effectiveness in capturing the general traffic flow around Heathrow Airport. This reflects the preliminary correctness of choosing the LSTM model for passenger flow prediction. However, the selection of sites is still not sufficiently representative. In the follow-up, it is advisable to combine the Random Forest model to assess the representativeness of each site.

4 CONCLUSION

This research involves the prediction and analysis of traffic flow and average vehicle speeds on seven sites along the M25 motorway near Heathrow Airport in August 2019, using the Long Short-Term Memory (LSTM) model in deep learning. The data preprocessing phase includes the removal of irrelevant data and the handling of missing values. In the initial analysis stage, features in the temporal dimension at respective sites on the M25 motorway were extracted and analyzed. It was observed that both traffic volume and average vehicle speed exhibited periodic peaks and troughs, with daily trends and patterns remaining generally consistent. Three representative sites, namely "M25/4936A priority 1 on link 200045641," "M25/4883A priority 1 on link 199131002," and "M25/7108B priority 1 on link 200045820," were selected for LSTM prediction analysis. Following preprocessing and initial analysis, a 2-layer LSTM model with 80 neurons per layer was created, using a prediction interval of 5 5-time steps. According to the results of the prediction analysis, excluding road anomalies, the LSTM model effectively predicts traffic flow on the highway adjacent to the major airport. This capability can assist relevant authorities in better addressing fluctuations in traffic flow near the airport, reducing congestion, improving traffic efficiency, and enhancing the travel experience for urban residents and travelers. This research aims to provide better support for traffic management decision-making.

REFERENCES

W.W. Liu, Guangming Net. The passenger throughput of Beijing Daxing International Airport exceeded 30 million this year. (2023) https://baijiahao.baidu.com/s?id=1779439192585390131&wfr=spider&for=pc

Y. qian, Guangdong Provincial State-owned Assets Supervision and Administration Commission. The passenger throughput of Guangzhou Baiyun Airport exceeded 50 million in 2023. (2023) http://www.sasac.gov.cn/n2588025/n2588129/c29222603/content.html

Y. Pang, W. Zhao, Y. N. Zhao, & H. K. Xu. Microcontrollers & Embedded Systems, 3, 72-75 (2019).

H. Li, S. Zhang, B. Cao, & J. Fan, Journal of Chongqing University, 43(11): 29-40 (2020).

H. W. Ding, L. Wan, Z. Xin, &X. K. Deng, Computer Engineering and Applications, 55(2): 228-235 (2019).

H. Zhao, D. M. Zhao & C. H. Shi.Urban Rapid Rail Transit, 32(4): 50-54 (2019).

X. X. Zhang & Y. Gong. China Intelligent Transportation System, (09) :133-137, (2022).

T. Zhou & Y. J. Xu, Journal of Shanghai Maritime Transportation Institute, 3, 36-42 (2021).

Z. Zhao, Y. Y. Zhang, *A traffic flow prediction approach: LSTM with detrending* in Proceedings of 2018 IEEE International Conference on Progress in Informatics and Computing (PIC). 101-105(2018).

Y. Y. Yang, Q. Fu & D. S. Wan, Computer Technology and Development, 27(3), 35-38 (2017).

D. Liang, J. Xu, S. Y. Li, C. K. Sun, (2020). *Short-term passenger flow prediction of rail transit based on VMD-LSTM neural network combination model* in Proceedings of the 32nd Chinese Control and Decision Conference (CCDC) 897-902.