Proceedings of CONF-MLA 2025 Symposium: Applied Artificial Intelligence Research
DOI: 10.54254/2755-2721/2025.BJ24681

Construction of a Short-Term Traffic Flow Prediction Model Based on Improved LSTM and Performance Evaluation Across Multiple Time Granularities

Nuo Chen

Tiangong University, Tianjin, China
C181557911N@126.com

Abstract. As urban traffic congestion continues to intensify, predicting short-term traffic flow has become essential to enabling real-time control in intelligent transportation systems (ITS). However, traditional models face significant limitations in capturing the spatiotemporal and nonlinear characteristics of traffic data. Long Short-Term Memory (LSTM) networks, with their gated mechanisms, can effectively model long-term dependencies and periodic patterns in traffic flow. The accuracy of these predictions directly influences decision-making in scenarios such as traffic guidance and emergency management, offering substantial practical value for improving road network efficiency. This study constructs an optimized LSTM model to evaluate its effectiveness in short-term traffic prediction and to compare predictive performance across different time granularities. A dual-layer LSTM architecture is employed, incorporating the Adam optimizer, Dropout, and early stopping as regularization strategies. Using urban traffic monitoring data from the United States, both hourly and daily prediction models are developed for experimental validation. Results indicate that the hourly prediction model (MSE = 0.0709) markedly surpasses the daily model (MSE = 0.2987), effectively identifying recurring patterns like rush periods in the morning and evening. These outcomes offer a practical solution for adaptive traffic regulation.

Keywords: Intelligent Transportation Systems, Short-Term Traffic Flow Prediction, LSTM, Deep Learning

1. Introduction

Amid swift growth in the economy and society, along with the widespread use of affordable new energy vehicles, daily traffic volume has grown markedly, while congestion continues to be a serious concern. In most developed and developing regions, intelligent transportation systems (ITS) are the preferred solution, relying on accurate and efficient traffic flow prediction for optimal operation. Among the various components of ITS, short-term traffic flow prediction serves as the foundation for traffic guidance, travel route planning, and manpower scheduling. Consequently, it has become a key factor in mitigating traffic congestion.

This study addresses the challenge of short-term traffic flow prediction by developing a sophisticated predictive model and analyzing its strengths and limitations. A variety of approaches have been introduced for short-term traffic forecasting, with neural network-based models becoming increasingly

DOI: 10.54254/2755-2721/2025.BJ24681

popular because they can capture complex nonlinear patterns and provide reliable predictions. In this regard, Long Short-Term Memory (LSTM) networks demonstrate distinct strengths in handling time series data. In contrast to traditional back-propagation (BP) neural networks, LSTM—an enhanced type of recurrent neural network (RNN)—is more effective in dealing with sequential data over time. The architecture of LSTM is composed of three essential gates: a forget gate that decides what information should be removed, an input gate that determines which new data to incorporate, and an output gate that chooses what to convey to the subsequent layer. These components empower LSTM to model intricate temporal dependencies and perform well when dealing with data featuring regular cycles, making it particularly effective for short-term traffic prediction.

The remainder of this paper is structured as follows. First, a concise literature review summarizes prior research on short-term traffic flow prediction. The discussion then turns to deep learning-based methods, with a particular focus on LSTM networks and their variants. Subsequently, the paper analyzes the distinctive features and performance characteristics of each method presented. Finally, the conclusion summarizes the key findings and implications of the study.

2. Literature review

In recent years, neural networks have demonstrated significant advantages in traffic flow prediction. The models have evolved from traditional backpropagation (BP) neural networks to Long Short-Term Memory (LSTM) networks capable of capturing spatiotemporal features, as well as to further enhanced variants.

BP neural networks were once the dominant approach in early traffic forecasting due to their simple structure and ease of implementation. However, their sensitivity to initial parameters and tendency to fall into local optima limited their effectiveness. In 2017, Siddiquee et al. [1] applied a BP neural network to address the challenge of sparse rural road data in Bangladesh. Their model successfully captured hourly traffic patterns and predicted daily volumes while filtering out abnormal fluctuations such as strikes, demonstrating adaptability in complex traffic environments. In 2021, Olayode et al. [2] proposed a hybrid ANN-PSO model that combined artificial neural networks with a particle swarm optimization algorithm. Applied to traffic prediction at a four-way signalized intersection, the model significantly outperformed traditional ANN approaches, with the R^2 value increasing from 0.99169 to 0.9971. This result validated the role of optimization algorithms in enhancing global search capabilities. As traditional BP networks struggled to handle non-stationary data, researchers pursued further improvements. In 2024, Wang et al. [3] introduced an improved sparrow search algorithm (ISSA) to optimize BP networks, resulting in the ISSA-BP model. This advancement enhanced global optimization capabilities and reduced prediction error. In the same year, Kong et al. [4] developed a model that integrated variational mode decomposition (VMD) with a hybrid sparrow search algorithm-based BP network (HSSA-BP), effectively addressing traffic flow non-stationarity. Their model outperformed benchmark models across various error metrics, including extended mean absolute error (eMAE) and extended root mean square error (eRMSE).

LSTM has gradually become a core method in traffic flow prediction due to its strength in modeling long-term dependencies in time series data. In 2018, Luo et al. [5] proposed a KNN-LSTM model that employed the k-nearest neighbors algorithm to filter relevant station data as inputs to the LSTM. This approach improved prediction accuracy by an average of 12.28%, laying the groundwork for future research on spatiotemporal correlation. In mid-2019, Wei et al. [6] introduced the AE-LSTM model, which used an autoencoder to extract spatial features and integrated them with LSTM for dynamic spatiotemporal modeling. On the California PeMS dataset, the model achieved a root mean square error (RMSE) as low as 26.32 and reached a peak-period matching accuracy of 90%. Later that year, two noteworthy studies were published concurrently. Weng et al. [7] utilized a genetic algorithm to optimize

DOI: 10.54254/2755-2721/2025.BJ24681

LSTM parameters, achieving lower prediction errors than traditional models on the Guangzhou–Shenzhen Expressway dataset, thereby highlighting the potential of intelligent optimization. Simultaneously, Yan et al. [8] proposed a CNN-LSTM model that used convolutional layers to extract spatial features from adjacent intersections. This approach reduced RMSE by 9.8% and systematically validated the necessity of integrating spatial and temporal modeling. At the end of 2019, Yang et al. [9] introduced an attention-enhanced LSTM model, referred to as LSTM+, which significantly improved performance metrics such as mean absolute error (MAE) and RMSE by more than 7% compared to the standard LSTM, expanding the theoretical capabilities of the model.

Compared with BP neural networks, LSTM and its variants offer several notable advantages. First, the gated mechanisms in LSTM effectively mitigate the vanishing gradient problem, enabling the model to capture both nonlinearity and long-term dependencies in traffic flow. For example, LSTM+ [9] reduced error accumulation when processing extended sequences. Second, LSTM is highly adaptable for integration with spatial feature extraction techniques such as CNN and KNN, supporting joint spatiotemporal forecasting. The CNN-LSTM model [8] reduced prediction errors by nearly 10% on California highway datasets. Third, LSTM demonstrates greater robustness in handling complex external factors, such as holidays and unexpected events. In early 2025, a model based on hierarchical clustering and fluctuation coefficients [10] achieved over 90% prediction accuracy on Runyang Bridge traffic data. By contrast, although optimization algorithms such as ISSA and PSO can enhance BP neural networks, their shallow architectures and static weight updating mechanisms limit their ability to handle high-dimensional, dynamic spatiotemporal data effectively.

Overall, existing research indicates that LSTM and its enhanced variants provide clear advantages in traffic flow prediction, particularly in capturing spatiotemporal nonlinearity, modeling long-term dependencies, and managing external disturbances. Future studies may further explore the integration of multimodal data, lightweight model architectures, and online learning mechanisms to enhance real-time forecasting accuracy and improve model generalizability.

3. Deep learning models

Deep learning is a machine learning methodology that achieves high-level feature extraction by constructing nonlinear transformation models with multiple hidden layers. Compared with traditional shallow models, its core advantage lies in the hierarchical abstraction of data features through stacked hidden layers and the optimization of network parameters using the backpropagation algorithm. This hierarchical feature learning mechanism enables deep learning models to demonstrate superior performance in areas such as image recognition, speech processing, and time series prediction. Among these methods, architectures such as RNNs and their enhanced form—LSTM networks—are widely recognized for their effectiveness in handling sequence data, as they are capable of capturing time-based dependencies.

3.1. Recurrent Neural Network

Recurrent Neural Networks (RNNs) introduce temporal feedback connections within the hidden layers (as illustrated in Figure 1), thereby overcoming the static modeling limitations of traditional feedforward neural networks. The core concept of RNNs is to capture dynamic sequence characteristics by recursively transmitting hidden states over time:

$$s_t = \sigma(Ux_t + Ws_{t-1} + b_h) \tag{1}$$

$$o_t = g(Vs_t) \tag{2}$$

DOI: 10.54254/2755-2721/2025.BJ24681

In the above equations, x_t denotes the input vector at time step t, o_t is the output vector at the current time step, and s_{t-1} is the hidden state from the previous time step. U, W, and V are the weight matrices the input-hidden layer, the hidden-hidden layers, and the hidden-output layer, respectively. The functions $\sigma(\cdot)$ and $g(\cdot)$ represent activation functions, commonly chosen as tanh or ReLU. RNNs enable the unified modeling of variable-length sequences through shared parameters across time steps. However, these models often suffer from gradient vanishing or explosion: as the input sequence becomes longer, the gradients propagated backward may diminish or surge uncontrollably, impairing the network's capacity to retain information over extended sequences. To overcome this issue, advanced RNN variants—especially the Long Short-Term Memory (LSTM) network—have been introduced.

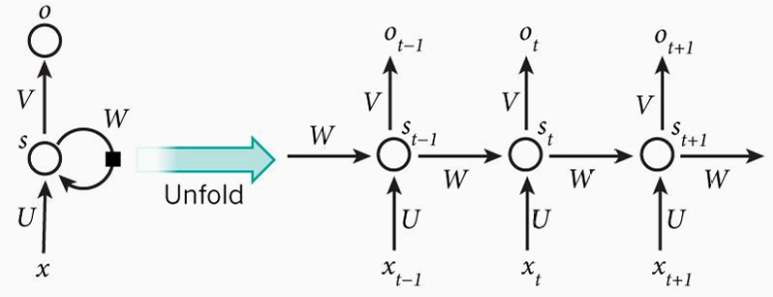

Figure 1: Basic Recurrent Neural Network

3.2. Long Short-Term Memory

Long Short-Term Memory (LSTM) networks [11] represent an enhanced form of Recurrent Neural Networks (RNNs), specifically built to handle and forecast sequence-based data. As depicted in Figure 2, LSTM integrates memory cells along with gate mechanisms into its structure, efficiently mitigating the issues of gradient vanishing and explosion that often arise in traditional RNNs when processing extended input sequences. Each LSTM cell comprises three primary gates: the forget gate, which determines what past data should be removed; the input gate, which selects new content to be written into the memory cell; and the output gate, which produces the current output based on the present cell state. This structure allows LSTM to maintain and transmit essential information across long temporal spans, thereby improving its capacity to learn long-range dependencies. As a result, LSTM has become a foundational model in applications such as NLP, time series prediction, and speech analysis. The architecture of the LSTM unit is illustrated in Figure 3. The computation carried out by a memory block at a specific time step t is outlined below:

$$f_t = \sigma\left(W_f \cdot [h_{t-1}, x_t] + b_f\right) \tag{3}$$

$$i_t = \sigma(W_i \cdot [h_{t-1}, x_t] + b_i) \tag{4}$$

$$\widetilde{C}_t = \tanh\left(W_c \cdot \left[h_{t-1}, x_t\right] + b_c\right) \tag{5}$$

$$C_t = f_t \circ C_{t-1} + i_t \circ \widetilde{C_t} \tag{6}$$

$$o_t = \sigma(W_o \cdot [h_{t-1}, x_t] + b_o) \tag{7}$$

$$h_t = o_t \circ \tanh(C_t) \tag{8}$$

DOI: 10.54254/2755-2721/2025.BJ24681

In the above equations, $\widetilde{C_t}$ represents the updated state of the memory cell at time t; i_t, f_t, o_t, C_t, and h_t denote the input gate, forget gate, output gate, memory cell, and the hidden state output at time t, respectively; x_t denotes the input at time t; h_{t-1} and C_{t-1} are the hidden state and memory cell output from the previous time step $t-1$. $\sigma(\cdot)$ and $\tanh(\cdot)$ represent the sigmoid and tanh activation functions, respectively.

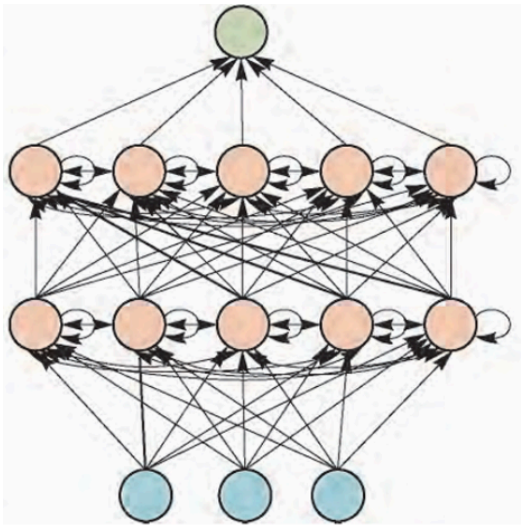

Figure 2: LSTM network structure

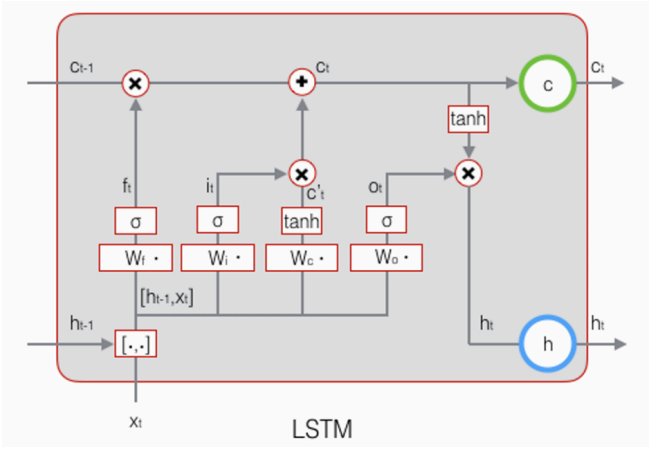

Figure 3: Internal structure of a memory block

4. Case study

4.1. Data source

In order to demonstrate the validity of the proposed approach, this study employs urban traffic flow data obtained from a monitoring system in a U.S. city. The dataset, made publicly available by Kaggle user fedesoriano and provided by an anonymous traffic research organization for educational and research purposes only, comprises traffic flow observations from four urban traffic junctions, totaling 48,120 records. At each junction, vehicle counts were recorded hourly via sensors over different time periods. The dataset includes the following main fields: record time (DateTime), junction ID (Junction), vehicle count (Vehicles), and record ID (ID). While the data coverage varies across junctions, the overall period spans from November 1, 2015, to July 1, 2017. Junctions 1 through 3 each contain 14,592 records,

DOI: 10.54254/2755-2721/2025.BJ24681

whereas Junction 4 is incomplete, with only 4,344 records. Hourly vehicle counts range from 1 to 180, with an average of approximately 22.8 and a standard deviation of 20.7. Each record is uniquely identified, and the dataset contains no missing values or formatting errors, ensuring data integrity for experimental analysis.

4.2. Data processing

For the experiment, traffic flow data from Junction 1, spanning from November 1, 2015, to July 1, 2017, was selected. Two datasets were constructed with different time granularities. Dataset 1 retained the original 1-hour granularity for hourly traffic flow prediction, resulting in a sample size of $n = 14592$. Dataset 2 aggregated traffic volume over 24-hour windows for daily prediction, with a time granularity of one day and a sample size of $n = 608$. To validate the model's ability to generalize, the two datasets were divided into training and testing portions using a 9:1 ratio, where 90% was used for training the model and the remaining 10% for assessing its performance. A sequential split strategy was adopted to maintain the temporal order and prevent any leakage of future information into the past.

4.3. Model structure and parameter settings

The LSTM-based traffic flow prediction model was implemented using a stacked network architecture in PyTorch. The input feature dimension is 1 (representing hourly/daily traffic volume), and the output dimension is also 1. A two-layer LSTM structure was employed to improve temporal feature extraction, with each layer containing 128 memory units. During training, the model processes 64 samples per parameter update (batch size = 64). The model is optimized using the Adam optimizer with a learning rate of 0.001 for 20 epochs. An early stopping strategy set with a patience value of 5 is applied; it interrupts training when the validation loss fails to decrease for five epochs in a row, effectively avoiding overfitting. To further mitigate overfitting, a dropout mechanism (dropout rate = 0.2) randomly deactivates 20% of neurons between layers. The weight matrices are initialized using a uniform distribution, and the bias terms are initialized as zero vectors. This setup enables the model to effectively capture traffic flow fluctuations while enhancing generalization through dynamic gradient updates and regularization techniques.

To evaluate prediction performance, the mean squared error (MSE) is used as the evaluation metric. The formula is defined as:

$$\mathrm{MSE} = \frac{1}{N}\sum_{i=1}^{N}\left(y_i - \hat{y}_i\right)^2 \tag{9}$$

where y_i represents the actual traffic flow value, $\hat{y}_i$ is the predicted value, and N denotes the number of prediction samples.

4.4. Experimental results and analysis

To predict 1-hour traffic flow using 24-hour data, Dataset 1 was employed with a time window of 24 and a prediction output size of 1. The model demonstrated good convergence within 20 training epochs, as the mean squared error (MSE) on the validation set steadily decreased from an initial value of 0.0540 to 0.0341. This indicates stable learning performance during training, with no signs of overfitting. During the prediction phase, the normalized outputs were converted back to the original scale, yielding a predicted traffic volume of 69.68 vehicles for the next hour. The predicted trend closely matched the actual traffic pattern, as shown in Figure 4. The final standardized MSE on the test set was 0.0709, highlighting the model's ability to effectively capture temporal features and demonstrating strong generalization, making it suitable for short-term traffic flow prediction tasks.

DOI: 10.54254/2755-2721/2025.BJ24681

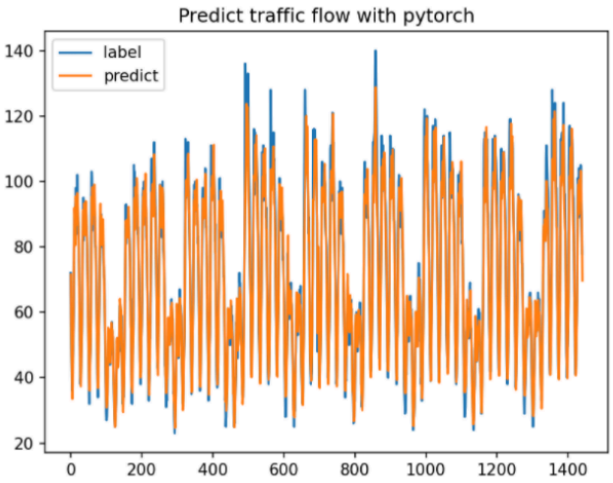

Figure 4: Results of hourly prediction

For predicting 1-day traffic flow using 7 days of data, Dataset 2 was utilized with a time window of 7 and an output size of 1. After 20 training epochs, the model exhibited a consistent decrease in both training and validation loss, indicating favorable convergence. Specifically, the training loss decreased from 0.78 to 0.15, while the validation loss dropped from 0.64 to 0.13. These results demonstrate that the model has strong learning capabilities, good fitting performance, and did not suffer from overfitting. In terms of prediction accuracy, the final MSE on the test set was 0.2987, suggesting reasonable fitting and prediction capabilities for traffic flow trends. Based on the historical data of 7 consecutive days, the model predicted a traffic volume of 1757.41 vehicles for the next day. This prediction was consistent with the original data in both magnitude and trend, as illustrated in Figure 5, effectively reflecting the underlying traffic flow patterns.

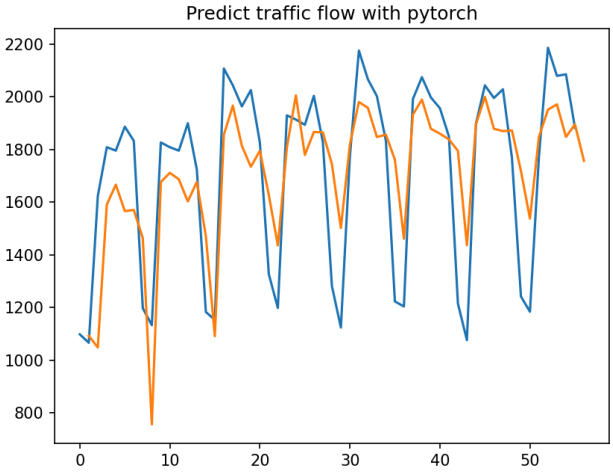

Figure 5: Results of daily prediction

5. Discussion

In this study on traffic flow prediction, both hourly and daily prediction tasks produced notable results, with the model demonstrating strong predictive capabilities on test sets across different temporal scales. Despite differences in prediction accuracy between the two tasks, both approaches effectively captured short-term variations in traffic flow, offering valuable insights for traffic analysis and forecasting.

DOI: 10.54254/2755-2721/2025.BJ24681

The hourly prediction task exhibited several key advantages. With an MSE of 0.0709 on the test set, the model demonstrated a strong ability to capture traffic flow changes over short time scales, resulting in lower prediction errors. This performance is primarily attributed to the nature of hourly data, which tends to exhibit smaller fluctuations and more pronounced patterns and periodicity. As shown in Figure 4, the cyclical variation in traffic flow is clearly visible, with the model accurately identifying peak periods corresponding to real-world morning and evening rush hours, in line with typical commuter behavior. Additionally, the model captures weekday-weekend traffic differences, with a clear up-and-down pattern reflecting actual traffic behavior. Hourly predictions benefit from a larger sample size and higher variability, enabling the model to learn more detailed short-term temporal features. This leads to a more stable training process, faster convergence, and a lower MSE.

On the other hand, daily prediction presented its own set of characteristics. The test set showed an MSE of 0.2987, indicating lower precision compared to the hourly predictions. As shown in Figure 5, daily predictions followed the general trend of actual traffic flow and captured the weekly periodic pattern. However, the predicted curve lagged behind during peak and trough periods, exhibiting a smoother amplitude than the actual flow. This suggests that daily prediction struggles to capture finer details, resulting in some loss of accuracy. Since the number of daily samples is only 1/24th of that in the hourly dataset, the model has less data to represent various traffic flow patterns. The reduced data variance further limits the model's ability to learn detailed features, leading to higher prediction errors. Additionally, the daily prediction curve is smoother compared to the hourly prediction, lacking the sensitivity to sudden or sharp changes. This reflects the limitations arising from a smaller sample size and lower data variability.

Compared to daily prediction, hourly prediction offers clear advantages. In terms of timeliness and granularity, it more accurately captures subtle intra-day variations in traffic flow, while daily prediction primarily identifies broader trends and fails to detect short-term fluctuations. The significantly lower MSE in hourly prediction demonstrates its superior performance in capturing traffic dynamics and minimizing error. With a larger sample size and greater data variation, hourly prediction enables the model to learn more short-term temporal features, leading to a more stable training process and faster convergence. In contrast, daily prediction is hindered by a smaller sample size and narrower data range, resulting in higher prediction errors.

In practical applications, hourly prediction provides precise support for dynamic traffic control, real-time route planning, and short-term dispatching decisions, significantly enhancing the responsiveness and regulatory capacity of urban traffic systems. While daily prediction offers a broader overview of traffic trends, hourly prediction excels in capturing short-term changes, making it more valuable for intelligent transportation and emergency management scenarios.

6. Conclusion

This research introduces a traffic flow forecasting approach based on deep learning, employing an enhanced Long Short-Term Memory (LSTM) network within the intelligent transportation domain. By designing a two-layer stacked LSTM structure and applying regularization methods including the Adam optimizer, Dropout, and early stopping, the model's capability to capture temporal patterns and its overall generalization were greatly improved. Experiments using both hourly and daily traffic flow datasets from a U.S. city demonstrated the model's effectiveness across different temporal scales. Results show that hourly prediction achieved outstanding performance with a standardized Mean Squared Error (MSE) of 0.0709, accurately capturing short-term periodic patterns, such as morning and evening peak traffic. Although daily prediction followed the overall trends (MSE = 0.2987), it was less effective at capturing finer details due to a limited sample size and reduced data variability. This study demonstrates that the dual-layer LSTM structure improves temporal feature learning, while the

DOI: 10.54254/2755-2721/2025.BJ24681

integrated regularization techniques mitigate overfitting, confirming the model's strong advantages in short-term traffic prediction. These results provide a reliable foundation for dynamic traffic control and real-time route planning. Future work could explore multimodal data fusion and lightweight model designs to further enhance real-time performance.

References

[1] Siddiquee, M. S. A., & Hoque, S. (2017). Predicting the daily traffic volume from hourly traffic data using artificial neural network. Neural Network World, (3).
[2] Olayode, I. O., Tartibu, L. K., Okwu, M. O., & Severino, A. (2021). Comparative traffic flow prediction of a heuristic ANN model and a hybrid ANN-PSO model in the traffic flow modelling of vehicles at a four-way signalized road intersection. Sustainability, 13(19), 10704.
[3] Wang, S., Li, X., Zhan, J., & Lü, T. (2024). Improved sparrow search algorithm optimized BP neural network for short-term traffic flow prediction. Journal of Qingdao University of Technology, 45(01), 126-133+140.
[4] Kong, S. (2024). Short-term traffic flow prediction based on the VMD-HSSA-BP neural network model. Electronic Design Engineering, 32(10), 1-7. https: //doi.org/10.14022/j.issn1674-6236.2024.10.001
[5] Luo, X., Li, D., Yang, Y., & Zhang, S. (2018). Short-term traffic flow prediction based on KNN-LSTM. Journal of Beijing University of Technology, 44(12), 1521-1527.
[6] Wei, W., Wu, H., & Ma, H. (2019). An autoencoder and LSTM-based traffic flow prediction method. Sensors, 19(13), 2946.
[7] Wen, H., Zhang, D., & Lu, S. (2019). Application of the GA-LSTM model in highway traffic flow prediction. Journal of Harbin Institute of Technology, 51(09), 81-87+95.
[8] Yan, Z., Yu, C., Han, L., Su, W., & Liu, P. (2019). Short-term traffic flow prediction based on CNN+LSTM. Computer Engineering and Design, 40(09), 2620-2624+2659. https: //doi.org/10.16208/j.issn1000-7024.2019.09.038
[9] Yang, B., Sun, S., Li, J., Lin, X., & Tian, Y. (2019). Traffic flow prediction using LSTM with feature enhancement. Neurocomputing, 332, 320-327.
[10] Lei, Y., Wang, L., Qian, P., Wang, B., & Zhang, J. (2025). Short-term traffic flow prediction for cross-river bridges during holidays based on LSTM neural network. Modern Transportation and Metallurgical Materials, 5(01), 76-84.
[11] Hochreiter, S., & Schmidhuber, J. (1997). Long short-term memory. Neural Computation, 9(8), 1735–1780.