Proceedings of CONF-FMCE 2025 Symposium: Semantic Communication for Media Compression and Transmission
DOI: 10.54254/2755-2721/2025.GL24060

A Multi-Scale Traffic Flow Prediction Model Based on LSTM

Jianyi Cui

Ocean University of China, Qingdao, China
1269915834@qq.com

Abstract: With the accelerating pace of urbanization, traffic flow prediction faces challenges in adapting to dynamic environments. Existing studies exhibit significant limitations in integrating external factors and extending to multi-scale prediction scenarios. Targeting the domain of traffic flow forecasting, this paper focuses on optimizing the Long Short-Term Memory (LSTM) network model, with particular attention to the impact of weather conditions on prediction accuracy and the performance differences between hourly and daily multi-scale forecasting. While mainstream approaches have improved LSTM performance through algorithmic optimization (e.g., Particle Swarm Optimization, Bayesian Optimization), a systematic solution to the integration of external factors and adaptation to different time granularities is still lacking. This study proposes an LSTM architecture that incorporates temperature as an embedded parameter, constructing a multi-factor input model and designing a dual-scale prediction framework at both the hourly (24-hour window) and daily levels (7/30/90-day windows). Using traffic flow and meteorological data from Interstate 94 in Minnesota, USA (2012–2018), the research explores the trade-off between external factors and time scales in LSTM modeling. The results provide a refined optimization path for traffic flow forecasting under complex scenarios.

Keywords: Short-term traffic flow prediction, time series forecasting, Long Short-Term Memory (LSTM), deep learning

1. Introduction

With the acceleration of urbanization and the continuous growth of transportation demand, traffic congestion has become a common challenge faced by major cities worldwide. Accurate traffic flow prediction is of great significance for optimizing traffic management, alleviating congestion, and improving road efficiency. In recent years, the application of deep learning models to traffic flow forecasting has emerged as a research hotspot. As a neural network model capable of effectively handling sequential data, the Long Short-Term Memory (LSTM) network has shown tremendous potential in this field. Traffic flow prediction not only supports transportation authorities in formulating more scientific traffic control strategies but also provides real-time traffic information for travelers, thereby optimizing route planning. However, traffic flow data is characterized by high nonlinearity and dynamic variability, making it difficult for traditional forecasting methods to capture its complex spatiotemporal patterns. As such, improving the accuracy and robustness of traffic flow prediction using deep learning models—particularly LSTM—has become a crucial research focus. This paper centers on LSTM-based modeling for traffic flow prediction. Owing to its capacity to

DOI: 10.54254/2755-2721/2025.GL24060

capture long-term dependencies, the LSTM model effectively extracts temporal features in traffic flow data. We conduct an in-depth exploration of data preprocessing, model construction, and feature selection, and validate the LSTM model's effectiveness through empirical experiments. Compared with traditional statistical and machine learning methods, LSTM models exhibit significant advantages in handling large-scale, high-dimensional traffic data, better addressing the complexities and uncertainties inherent in traffic forecasting. Following a structured research framework—“Problem Identification – Model Improvement – Experimental Validation – Conclusion and Summary”—this paper begins by presenting the research background and significance, and outlines the current limitations of LSTM models in integrating external factors and adapting to multi-scale prediction tasks. A literature review follows, summarizing recent progress and ongoing challenges. We then elaborate on the research methodology, including the construction of a dataset combining traffic and meteorological data from Minnesota, USA, the design of a multi-scale LSTM model with embedded temperature features, and three comparative experiments using hourly (24-hour window) and daily (7-day, 30-day, 90-day windows) time scales. The experimental results section highlights the limited effect of temperature on predictive accuracy (a mere 4.6% reduction in MSE), as well as the notable performance differences across time scales (hourly model MSE = 0.0141, significantly outperforming the daily models). Finally, we summarize key findings, emphasize the critical role of time granularity in model performance, and propose future research directions including the integration of additional external factors and the exploration of hybrid modeling strategies.

2. Literature review

The application of Long Short-Term Memory (LSTM) networks in traffic flow prediction has evolved alongside the growing demand for responsive modeling in dynamic urban environments. As a result, research has shifted from merely validating LSTM’s effectiveness to enhancing the model through algorithmic optimization and exploring how to better integrate it into real-world prediction systems.

Wang et al. [1], using 5-minute interval traffic data from detector station ID 403349 provided by California’s PeMS system, applied LSTM for short-term traffic flow prediction and demonstrated its capability in capturing temporal dependencies at 5- and 10-minute intervals. By incorporating Particle Swarm Optimization (PSO) to fine-tune hyperparameters, their PSO-LSTM model outperformed conventional LSTM in terms of Root Mean Square Error (RMSE) and Mean Absolute Error (MAE), establishing LSTM’s foundational role in time series-based traffic forecasting.

Cao et al. [2] analyzed hourly traffic data from Interstate 94 in Minnesota (UCI dataset, 2015–2018), comparing the performance of LSTM, Gated Recurrent Units (GRU), and Temporal Convolutional Networks (TCN) in hourly traffic prediction. The results showed that LSTM had superior capability in capturing long-term dependencies compared to GRU, while TCN achieved slightly better results at the cost of higher computational complexity. This suggests that LSTM strikes a favorable balance between accuracy and efficiency in medium-to-short-term forecasting tasks. Subsequent studies have focused on enhancing LSTM performance through various algorithmic optimizations.

Shen et al. [3] used Bayesian optimization to automatically tune hyperparameters for freeway traffic forecasting based on 15-minute interval flow data from UK highways (2018 dataset), achieving a 6.3% reduction in RMSE. Duan et al. [4] applied Grey Wolf Optimization (GWO) to radar-based hourly traffic data (2021–2022) from the Chatiaoling Tunnel in Shaanxi Province, improving model adaptability in real-time and reducing energy consumption, demonstrating the benefits of bio-inspired algorithms in traffic modeling.

These studies highlight the significant potential of algorithmic optimization in enhancing the generalization ability and practical utility of LSTM models. Jia et al. [5] proposed an innovative approach by gridding traffic trajectory data. Using DiDi ride-hailing data collected within Chengdu’s

DOI: 10.54254/2755-2721/2025.GL24060

Second Ring Road (November 2022, with sampling intervals of 2–4 seconds), they constructed traffic grid clusters via Gaussian Mixture Model (GMM) clustering and used Bi-LSTM to capture spatiotemporal correlations. The experimental results revealed that their model achieved a MAE of 3.0687 during morning rush hours, outperforming traditional multivariate linear regression by 4.14%, thereby showcasing LSTM's potential for spatial dimension expansion. Xue et al. [6] explored a traffic flow forecasting method based on Graph Neural Networks (GNNs), emphasizing the spatiotemporal characteristics of traffic flow. The study employed Graph Convolutional Networks (GCN) to extract spatial features and integrated a time-varying GRU to process road network sequences. Using a sequence-to-sequence architecture, they applied the model to traffic flow data from METR-LA (Los Angeles) and taxi flow data from Luohu District in Shenzhen. The model demonstrated high accuracy and was applicable to short-term (5–30 minutes), mid-term (30–60 minutes), and long-term (1-hour) predictions, highlighting the potential of data-driven approaches in traffic flow forecasting.

Yang et al. [7] introduced the MCNN-ABiLSTM model, which incorporates Multiscale Convolutional Neural Networks and an attention-based Bidirectional LSTM mechanism to enhance the temporal and spatial sensitivity of traffic flow series. They used Pearson correlation coefficients to quantify spatial dependencies between intersections and employed an improved PSO algorithm to label external factors. The experimental results showed that MCNN-ABiLSTM significantly outperformed baseline models in terms of RMSE, MAE, and MAPE. Tang et al. [8] and Zhou et al. [9] focused on urban traffic flow prediction. Tang proposed a CNN-LSTM-AM model for short-term traffic forecasting using PeMS data from California, while Zhou developed an attention-based CNN-LSTM model to predict taxi flow using trajectory data from Beijing. The datasets used in both studies were comprehensive and representative, providing robust support for analyzing the spatiotemporal characteristics of urban traffic.

However, limitations persist in current research: (1) Insufficient integration of external factors such as weather and events; (2) Limited exploration of the scalability of multi-scale forecasting, particularly in handling different temporal granularities (e.g., hourly vs. daily flow). Building upon the above findings, the present study proposes an improved LSTM architecture that incorporates weather variables, aiming to predict traffic flow at both the hourly and daily levels with enhanced adaptability and precision.

3. Principle of the LSTM algorithm

Long Short-Term Memory (LSTM) networks are a special type of Recurrent Neural Network (RNN) first proposed by Hochreiter and Schmidhuber [10] in 1997. The core idea of LSTM is to address the gradient vanishing problem in training long sequences by introducing a gating mechanism and a cell state. The following section presents a detailed explanation of the mathematical model and working principles of LSTM.

Each LSTM unit contains three gates and two state vectors:

Forget Gate: Controls the extent to which the previous cell state is retained.

Input Gate: Regulates the amount of new information to be stored.

Output Gate: Determines how much of the cell state is output.

Cell State: Acts as the carrier of long-term memory.

Hidden State: Represents short-term memory.

The computation of an LSTM unit at time step t can be described by the following equations:

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

Where:

σ denotes the sigmoid activation function,

DOI: 10.54254/2755-2721/2025.GL24060

$W_f \in \mathbb{R}^{h\times(d+h)}$ denotes the weight matrix of the forget gate.,
b_o are their corresponding biases,
h_{t-1} is the hidden state from the previous time step,
x_t is the input at the current time step.
Input gate computation:

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

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

Cell state update:

$$C_t = f_t \odot C_{t-1} + i_t \odot \tilde{C}_t \tag{4}$$

Output gate computation:

$$o_t = \sigma(W_o \cdot [h_{t-1}, x_t] + b_o) \tag{5}$$

$$h_t = o_t \odot \tanh\,(C_t) \tag{6}$$

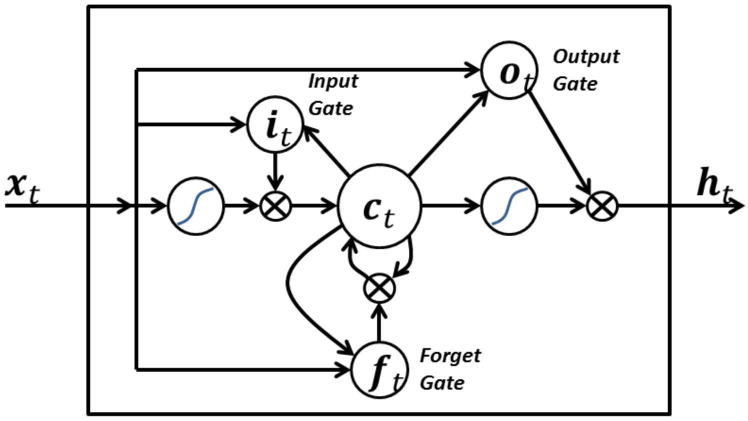

Figure 1: Graphical explanation of the cell state of LSTM

In Figure 1,it shows the graphical explanation of the cell state of LSTM.In this study, the LSTM model utilizes 128 neurons in each hidden layer, which directly influences the model's capacity to learn temporal patterns in traffic flow. A two-layer LSTM architecture is adopted to extract hierarchical features, enhancing the model's ability to capture complex traffic dynamics. During training, 20% of the neuron outputs are randomly dropped (dropout) to effectively prevent overfitting on traffic data.

The model construction process is as follows:

Step 1: Data preparation.

Step 2: Dataset partitioning.

Step 3: Model training. The LSTM model is trained using the training dataset.

Step 4: Model testing. The trained LSTM model is applied to the test dataset to assess its performance.

Step 5: Model evaluation.

4. Results

The data used in this study originate from hourly westbound traffic flow records on Interstate 94 in the Minneapolis–Saint Paul metropolitan area, Minnesota [11], spanning a total of seven years from

DOI: 10.54254/2755-2721/2025.GL24060

2012 to 2018. Each entry contains hourly traffic volume and corresponding temperature information. By summing hourly volumes for each day, a new dataset of daily traffic volumes was constructed.

The following experimental cases were designed:

Case 1. Hourly Prediction

LSTM was applied to the entire dataset of hourly traffic flow and temperature from 2012 to 2018, with a timestep of 24—i.e., the previous 24 hours were used to predict the next hour. The dataset was split into 98% training, 2% testing, and 20% validation (note: the validation proportion is relative to training data).

Case 2. Hourly + Temperature Prediction

This setup was identical to Case 1, except temperature was explicitly included as a feature alongside traffic flow data. The timestep was again set to 24, and the data split remained the same.

Case 3. Daily Prediction

This experiment used daily aggregated traffic flow data. Three sub-cases were conducted based on different timesteps: Case 3a: Using the past 7 days to predict the next day. Case 3b: Using the past 30 days to predict the next day. Case 3c: Using the past 90 days to predict the next day. To better highlight differences in performance, the dataset was divided into 90% training, 10% testing, and 20% validation (relative to training).

For all experiments, the Stochastic Gradient Descent (SGD) optimizer was selected for its ability to enhance generalization through mini-batch randomness, maintain stable gradient updates, and adapt efficiently to dynamic data. The Mean Squared Error (MSE) was chosen as the loss function for its suitability in penalizing large deviations, its differentiability for gradient-based optimization, and its ability to directly quantify overall deviation between predicted and actual values. The difference between predicted and actual values was used as the primary evaluation metric.

The following figures illustrate the experimental outcomes:

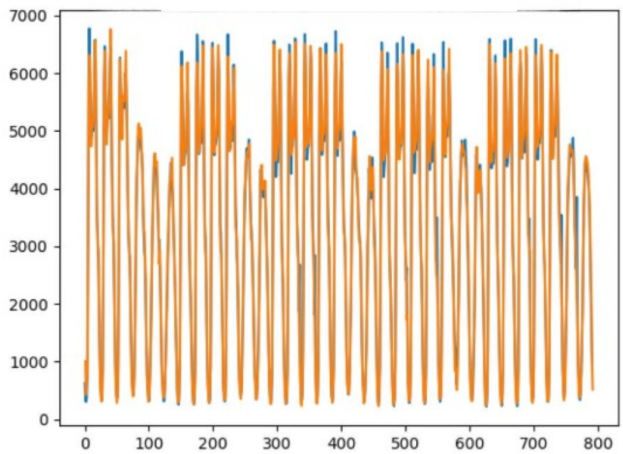

Figure 2: Hourly prediction in Case1

In Figure 2, a consistent pattern of five peaks and two valleys in hourly traffic volume per week is observed. Each "peak" corresponds to either the morning or evening rush hours, while the valleys reflect lower nighttime traffic. This five-high-two-low pattern aligns with reduced traffic volumes on weekends compared to weekdays.

DOI: 10.54254/2755-2721/2025.GL24060

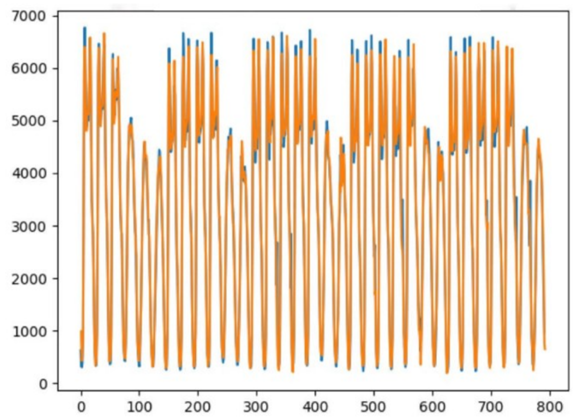

Figure 3: Hourly prediction in Case2 considering temperature

In Figure3, when temperature is included as an input feature, the prediction performance shows little improvement compared to Case 1. This suggests that temperature has minimal impact on LSTM's performance in this context.

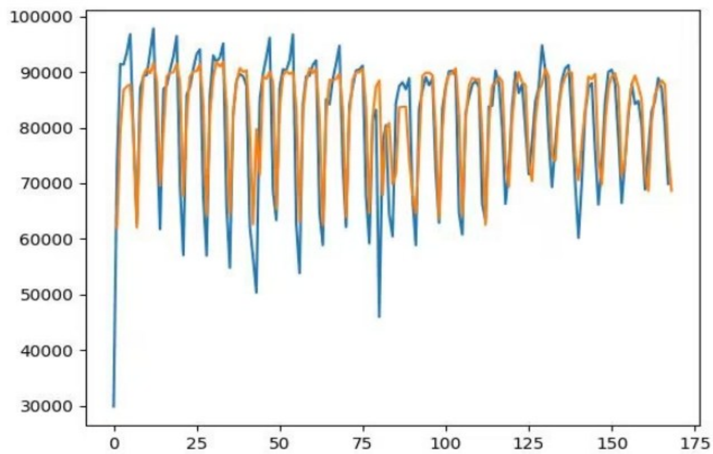

Figure 4: Daily prediction based on a 7-day period in Case3a

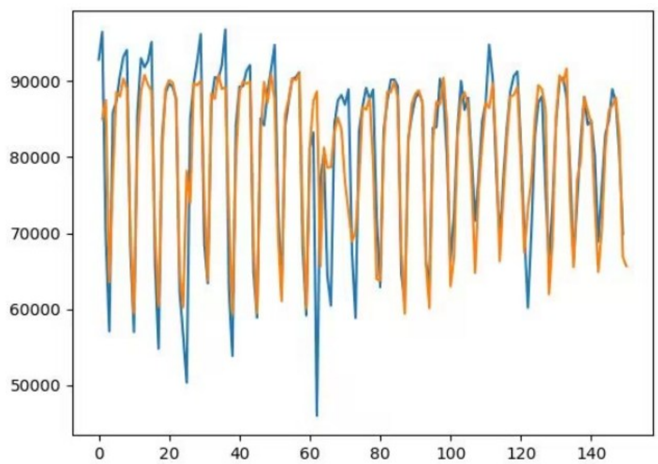

Figure 5: Daily prediction based on a 30-day period in Case3b

DOI: 10.54254/2755-2721/2025.GL24060

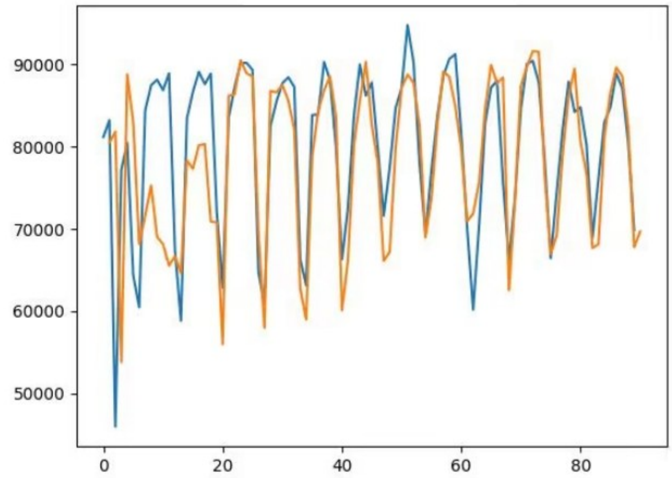

Figure 6: Daily prediction based on a 90-day period in Case3b

Figure 4 Figure 5 and Figure 6 illustrate daily predictions using 7-day, 30-day, and 90-day timesteps, respectively. Compared with hourly predictions, daily forecasts are less accurate. However, the general five-high-two-low weekly pattern remains. Comparison of Case 1 (hourly) and Case 3 (daily) under different time granularities shows that model performance varies significantly with timestep size and data aggregation level. The following Table 1 summarizes the MSE values for each experimental setup:

Table 1: MSE of all cases

	MSE
Case1	0.01408546
Case2	0.01343374
Case3a	0.50603939
Case3b	0.22317
Case3c	0.35447066

5. Discussion

In the task of traffic flow prediction, incorporating temperature as an external factor did not lead to significant improvements in model performance. Comparing Case 1 (using only traffic flow data) and Case 2 (including temperature), the latter achieved a marginally better MSE—0.01343374 compared to 0.01408546 in Case 1. This slight improvement suggests that temperature has a limited effect on LSTM-based traffic prediction. One possible reason is that traffic flow itself already captures the key dynamic patterns, rendering the additional temperature feature redundant or weakly correlated in this context.

Case 1 adopted an hourly prediction granularity, while Case 3 relied on daily data. This change in temporal resolution had a substantial impact on model performance. In Case 3, which aimed to predict the next day’s traffic flow using the past 7, 30, or 90 days, the MSE values were significantly higher: Case 3a (7-day window): 0.50603939; Case 3b (30-day window): 0.22317; Case 3c (90-day window): 0.35447066. These results suggest that longer input windows may make it harder for the model to capture short-term fluctuations in traffic, increasing prediction uncertainty and reducing accuracy. Furthermore, the use of daily data reduces the number of training samples, which can negatively affect the model's learning and generalization capabilities.

DOI: 10.54254/2755-2721/2025.GL24060

Notable differences were also observed within the Case 3 sub-experiments. The length of the temporal window played a key role in determining what kind of features the model was able to learn. Shorter windows (e.g., 7 days) were better suited for capturing abrupt shifts and short-term dynamics, whereas longer windows smoothed out the data too much, potentially masking seasonal or weekly variations. For instance, a flood event that occurred on day 0 of the test set in Case 3a significantly affected prediction accuracy, highlighting the model's sensitivity to extreme weather events. Longer windows (e.g., 90 days) tend to average out such anomalies, reducing the model's responsiveness to sudden changes in traffic conditions.

6. Conclusion

This study addresses traffic flow prediction by proposing a multi-scale LSTM architecture that incorporates external temperature data, aiming to enhance prediction accuracy across different time granularities. Using real-world data from Interstate 94 in Minnesota (2012–2018), three sets of comparative experiments were conducted at the hourly level (24-hour window) and daily level (7-day, 30-day, and 90-day windows). The results indicate that: Including temperature (Case 2) yields only a 4.6% reduction in MSE compared to using traffic flow alone (Case 1), suggesting limited predictive contribution of temperature in this context. Hourly predictions (MSE = 0.0141) consistently outperform daily predictions (e.g., Case 3a MSE = 0.5060). Longer input windows (e.g., 90 days) tend to smooth out fluctuations, leading to poorer detection of seasonal patterns. Shorter input windows (e.g., 7 days) are more responsive to sudden events, such as natural disasters, but may suffer from overfitting or noise sensitivity. This research highlights the critical role of temporal granularity in LSTM performance for traffic forecasting. The findings offer valuable theoretical insights and practical guidance for building more refined and adaptive predictive models under complex traffic scenarios.

References

[1] *Wang, N., Cheng, L., Zhen, J., & Duan, X. (2024). Website design for short-term traffic flow prediction based on PSO-LSTM. Journal of Langfang Normal University (Natural Science Edition), 24(1), 29–32.*

[2] *Cao, Y., Chang, Y., & Chen, M. (2024). Comparative study of traffic flow prediction methods based on neural networks. Science and Innovation, (13), 42–44. https://doi.org/10.15913/j.cnki.kjycx.2024.13.009*

[3] *Shen, K., Zhu, Y., Meng, X., & Li, H. (2022). Expressway traffic flow prediction based on Bayesian-optimized LSTM. Modern Computer, 28(11), 51–55.*

[4] *Duan, Z., & Du, W. (2024). Tunnel traffic flow prediction and lighting adjustment research based on GWO-LSTM model. Modern Tunneling Technology, 61(3), 157–165. https://doi.org/10.13807/j.cnki.mtt.2024.03.017*

[5] *Jia, X., Feng, C., Su, Z., Qian, Z., Song, T., Liu, H., & Lü, Y. (2023). Traffic flow prediction using Bi-LSTM for urban traffic grid clusters. Journal of Chongqing University, 46(9), 130–141.*

[6] *Xue, Y. (2024). Research on traffic flow prediction methods based on graph neural networks. Logistics Technology, 47(10), 118–123, 129. https://doi.org/10.13714/j.cnki.1002-3100.2024.10.021*

[7] *Yang, G., Chen, J., Zhang, Z., & Wang, W. (2024). MCNN-ABiLSTM traffic flow prediction model consideri ng external factors. Journal of Tianjin University of Technology and Education, 34(1), 66–73. https://doi.org /10.19573/j.issn2095-0926.202401012*

[8] *Tang, Z., Zhao, D., & Wang, S. (2024). Short-term traffic flow prediction based on CNN-LSTM-AM. Science Technology and Engineering, 24(31), 13562–13567.*

[9] *Zhou, X., Jin, J., Bao, N., Yuan, T., & Cui, Y. (2024). Urban taxi traffic flow prediction based on attention convolutional long short-term memory model. Journal of Safety Science and Technology, 34(7), 153–162. https://doi.org/10.16265/j.cnki.issn1003-3033.2024.07.0089*

[10] *Hochreiter, S., & Schmidhuber, J. (1996). LSTM can solve hard long time lag problems. In Proceedings of the 10th International Conference on Neural Information Processing Systems (pp. 473–479). MIT Press.*

[11] *Hogue, J. (2019). Metro Interstate Traffic Volume [Dataset]. UCI Machine Learning Repository. https://doi.o rg/10.24432/C5X60B*