

RESEARCH ARTICLE

Short-Term Traffic Speed Forecasting Using a Deep Learning Method Based on Multitemporal Traffic Flow Volume

YACONG GAO¹, CHENJING ZHOU², JIAN RONG², YI WANG², AND SIYANG LIU¹¹Beijing Key Laboratory of Traffic Engineering, Beijing University of Technology, Beijing 100124, China²School of Civil Engineering, Guangzhou University, Guangzhou 510006, China

Corresponding author: Chenjing Zhou (13439099781@139.com)

This work was partially supported by the project “Guangxi Highway Traffic Congestion Prevention and Control Intelligent decision-making system” in Guangxi, China.

ABSTRACT Accurate traffic speed forecasting not only can help traffic management departments make better judgments and improve the efficacy of road monitoring but also can help drivers plan their driving routes and arrive safely and smoothly at their destination. This paper focuses on the lack of traffic speed data and proposes a method for traffic speed forecasting based on the multitemporal traffic flow volume of the previous and later moment states. First, according to traffic flow volume data, the different traffic patterns of previous and later moment states were extracted. Second, the performance of five forecasting models, namely, long short-term memory (LSTM), backpropagation (BP), classification and regression trees, k-nearest neighbor, and support vector regression, were compared. Finally, the model with the best prediction results was used to conduct sensitivity analysis experiments for different traffic patterns. Through a real-data case study, we found that the LSTM model has the highest prediction accuracy compared to other models in both time and space. This traffic pattern “previous = 3 and later = 3” can forecast traffic speed more accurately, and its forecasting ability is robust across a range of scenarios.

INDEX TERMS Traffic speed forecasting, traffic patterns, deep learning, traffic flow theory.

I. INTRODUCTION

Traffic speed forecasting is a critical component of the intelligent transportation system, which can dynamically capture the development trend of the traffic flow status [1]. Accurate traffic speed forecasting can be beneficial for both road traffic management and individual travelers. For example, a road traffic management department may identify and predict congestion sites according to traffic speed forecasting results, as well as analyze the causes of traffic congestion and propose solutions to relieve congestion [2]. Moreover, individual travelers can plan their travel times and trips using the forecasting results, which are available on various navigation apps, such as Waze and Google Maps [3].

Many existing road infrastructure have not been equipped with traffic flow monitoring equipment. To adapt with the

intelligent upgrading of urban transportation, some cities choose to use floating car data or increase traffic flow data monitoring equipment to sense the state of traffic operations [4]. However, the flow of speed integration monitoring equipment is more costly. Most cities choose only the flow detection function, the cost of relatively low-cost camera equipment. Whether it is floating car data or camera detection data, the core component is flow detection, resulting in the lack of speed data. Therefore, predicting speed based on traffic volume is valuable work. According to the basic diagram of traffic flow, there is a complex nonlinear relationship between traffic flow volume and speed [5]. We posit that speed and volume exist interactively, and the speed of the current moment is correlated with previous and later moment states. Thus, we proposed a novel method to carry out traffic speed prediction based on the flow volume of previous and later moment states. The main contributions of this study could be enlisted as follows:

The associate editor coordinating the review of this manuscript and approving it for publication was Yeliz Karaca.

1) We propose a method for short-term traffic speed forecasting based on the multitemporal traffic flow volume of the previous and later moment states. This method can real-time and accurately reflect the road operation state without traffic speed data.

2) We build an LSTM short-term traffic forecasting model that can accurately capture recent spatio-temporal features. We also conduct extensive experiments on actual traffic data. The results show that the LSTM model outperforms four baseline methods and achieves greatest result.

3) We determine the best traffic flow pattern for the short-term traffic speed forecasting model, based on the LSTM model parameter sensitivity analysis experiments.

The remainder of this work is structured as follows: Section 2 summarizes related research. The proposed method of developing the traffic feature pattern and related prediction methods is described in section 3. In section 4, the proposed method is evaluated and the optimal traffic pattern for previous and later moment states is determined with real-world traffic flow data. Finally, the main contributions of the paper are presented and discussed in section 5.

II. RELATED WORKS

We reviewed existing studies on traffic forecasting and highlight recent traffic speed forecasting research development.

Since the first highway traffic volume forecasting model was established in 1979, researchers from several disciplines have focused their attention and efforts on traffic forecasting [6]. With the development of a new generation of intelligent systems, new methods for addressing traffic forecasting problems have been developed. These approaches are broadly classified into two types: classical statistical approaches and data-driven methods [7], [8].

Classical statistical methods have a clear physical meaning and are easily explained. Academic researchers have developed a variety of such methods, ranging from linear regression [9] and support vector machine [10], [11] to the Autoregressive Integrated Moving Average (ARIMA) model [12]–[14]. Williams conducted a representative study on applying the ARIMA model, which considered the effect of seasons on traffic forecasts. He established ARIMA models with explanatory variables and a seasonal autoregressive integrated moving average (SARIMA) model as well as a SARIMA model with explanatory variables, and compared the performance of different methods for traffic prediction [15].

However, most traditional statistical approaches are incapable of considering aspects such as individual randomness and nonlinearity. These traditional approaches frequently use the mean feature of the data as the model's input, neglecting the extreme data of the individual, and the resulting model lacks generalization ability [16]. Because of differences in the driving behavior of individual vehicles at the micro-level, the traffic flow volume and speed at the medium and macro levels are also different in actual traffic

flow. Therefore, these statistical approaches are ineffective in describing the operational condition of real-world traffic flow [17], [18]. Deep learning technology has grown significantly in recent years, providing a new way to describe complicated, nonlinear data correlations for intelligent traffic management. They are now used extensively in a variety of areas [19]–[23]. A data-driven model is a type of model that represents complicated, nonlinear data relationships. The structures constructed using the data-driven model are trained by establishing the accuracy of the predicted target. The basic parameters of the model are constantly iterated and updated during the training process to reach convergence and ultimately address the traffic prediction problems [24]. Rather than using statistical methods to determine the functional relationship between independent variables and variables, Moreover, statistical models require strict assumptions and physical derivation processes, but data-driven models have powerful data learning capabilities and do not require as many model assumptions [25]. When dealing with complicated and extreme data, data-driven models outperform traditional statistical models [26].

In recent years, transportation researchers have begun to use deep learning methods in their studies. Ma *et al.* first introduced long short-term memory (LSTM) into the field of traffic research, and they used microwave sensor data to conduct traffic speed prediction research [26]. Zhang *et al.* proposed a new convolutional LSTM neural network structure in this context, which represents actual traffic conditions by constructing multiple features and considering the relationship between adjacent lanes, upstream and downstream, to achieve a multilane short-term traffic forecast [27]. Ma *et al.* used a specific convolutional neural network (CNN) to extract inter-and intraday traffic flow patterns. Then, to learn the intraday temporal evolution of traffic flow, the extracted features were fed into LSTM units [20]. Meng *et al.* combined the dynamic time planning algorithm into an LSTM neural network model, fine-tuned the time-series features, and tested the approach in the traffic speed prediction problem at different time intervals [24]. Based on the nonstationary spatiotemporal pattern of the road traffic, Cheng *et al.* developed a dynamic spatiotemporal k-nearest-n model for short-term traffic forecasting. The model was validated by data from Beijing's urban roads and California's highways [25].

In addition to the influence of the day of the week, other environmental factors such as weather, holidays, seasons, and important events may have a significant impact on traffic flow patterns [28]–[30]. Dunne *et al.* designed a neural wavelet predictive model that captures the effects of rainfall to forecast hourly traffic flow using the stationary wavelet transform (SWT) [31]. The results of the experiments show that using rainfall data as an independent factor, it is possible to reliably predict future traffic flow at different sites. Similarly, it was demonstrated that the neural wavelet model has a greater prediction performance than the traditional

artificial neural network model. Jia *et al.* proposed utilizing the deep belief network and LSTM to forecast urban road traffic flow in the event of rainfall [32]. When the extra rainfall factor was considered, the experimental results show that the deep learning prediction model outperforms the existing prediction model. Furthermore, LSTMs outperform deep belief networks in extracting traffic flow data time-series patterns. Existing traffic prediction algorithms are mainly useful for predicting traffic flow and speed along the entire road segment.

A more detailed lane-level traffic speed prediction was performed in a recent study. Gu *et al.* chose a road segment with a strong association with the road section to be predicted for data extraction and built a two-layer deep learning framework by combining the advantages of LSTM and GRU [19]. Ke *et al.* converted traffic speed and volume data into spatiotemporal multichannel matrices before training a two-stream multichannel CNN model to predict lane-level traffic speeds [33]. It reflected not only the relationship between individual lanes but also the correlation between the spatiotemporal characteristics of the input data. Inspired by this past work, we have proposed a new short-term, high-precision traffic speed prediction method based on traffic flow patterns.

We obtained specific traffic pattern data and selected five methods, namely LSTM, backpropagation (BP) network, classification and regression trees (CART), k-nearest neighbor (KNN), and support vector regression (SVR), to carry out the traffic speed prediction study. The biggest challenge of the daily traffic speed forecasting problem in this research is comparing the accuracy and robustness of the prediction results of these five algorithms. The next research work will be carried out based on the algorithm with the best prediction results.

In this study, focusing on the complex correlation between traffic flow volume and speed, we proposed a novel traffic speed prediction method that extracts time-series patterns from traffic flow data of previous and later moment states. Another challenge of neighbor traffic speed prediction is to determine the best traffic pattern. Existing short-term traffic speed forecasting studies have focused on the traffic flow patterns of previous states, but have not modeled the traffic flow patterns of later states. As a result, large errors in short-term traffic forecasts can result. It is particularly important to determine the best traffic pattern of previous and later moment states to improve the accuracy of the prediction results.

Following are the three challenges that we attempted to resolve in this study:

1. How to compare the traffic speed forecasting effect of all methods.
2. How to prove that the traffic speed in the current state is related to the traffic flow volume in both previous and later moment states.
3. How to determine the best traffic pattern for previous and later moment states.

III. METHODOLOGY

A. OVERVIEW

The proposed method in this study has two major components: (1) traffic flow pattern extraction and (2) an introduction to LSTM, BP, CART, KNN, and SVR. We next introduce these two components.

B. TRAFFIC FLOW PATTERN EXTRACTION

The traffic flow volume and speed data are collected at a given time interval (t), such as 2 minutes, 5 minutes, and 15 minutes. For a day, there are in total T ($T = 60 \times 24/t$) time intervals. Longer time intervals weaken the essential traffic characteristics within the traffic flow. To accurately characterize the operational features of traffic flow and improve the forecast effect of traffic speed, we chose 2 minutes as the time interval, which means that the value of T is 720 at one day.

According to the traffic flow basic diagram, a strong relationship exists between traffic flow and speed, and this relationship is not a simple linear one. For example, the same flow value corresponds to different speed values, and the same speed value also corresponds to different flow values. Therefore, it is unreasonable to only use current traffic flow data to predict the speed, and a specific method is needed to extract the traffic pattern. Hence, we attempted to construct a traffic flow pattern using the traffic flow volume data of previous and later moment states. We simulated and forecasted the current traffic speed to determine the relationship between traffic flow and speed. The detailed process is shown in Fig. 1. We could obtain time-series data of traffic flow and speed as the training data with the loop detector. Assume the current state is M , n represents the time step of previous states, and m represents the time step of later moment states. First, we extracted the continuous time-series data of the current states M and the n and m states. Second, we gradually constructed $M + 1, M + 2, \dots, M + K$ state traffic flow, and speed pattern vectors by the sliding window method. Finally, the traffic flow matrix and speed matrix were formed. In this study, we used the flow matrix as the input of the prediction algorithm and the speed matrix as the prediction target and applied different datasets to compare the performance of all algorithms. On this basis, the speed prediction study under different traffic patterns was carried out by varying the values of n and m .

C. LSTM

Ma *et al.* proposed the LSTM model as a special recurrent neural network model. It can address the gradient disappearance and gradient explosion problems caused by the difficulties in finding the weight matrix [34]. As shown in Fig. 2, LSTM networks incorporate a new structure known as memory cells.

Each memory cell is made of four main components: an input gate, a forget gate, an output gate, and a neuron with self-repetition. The input gate (f_t) determines the unit that needs to be updated, as shown in (1). Through the sigmoid

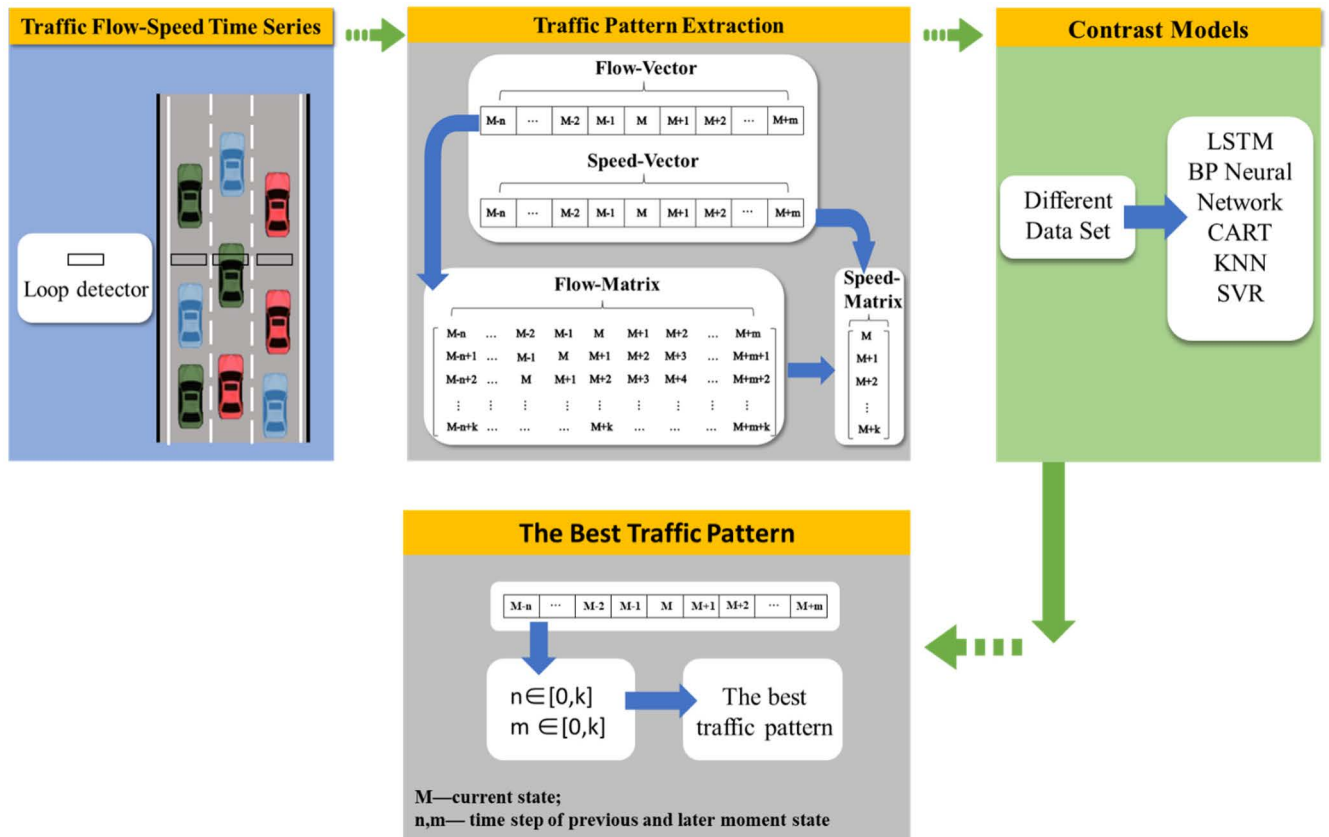


FIGURE 1. Research framework.

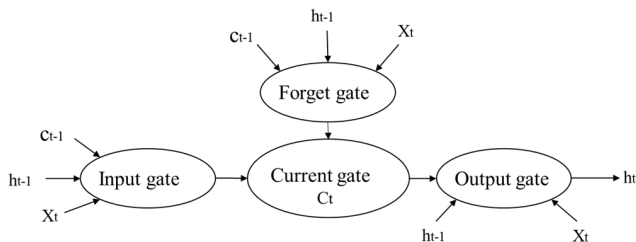


FIGURE 2. LSTM networks.

neural layer of the forget gate, the information in the united state at the previous moment is discarded or retained and updated, as shown in (2) and (3). Then, combine the input gate and the forget gate, as shown in (4). Finally, the output threshold layer decides how much information to pass to the next time unit and output it, as shown in (5), where the hidden state information is obtained as shown in (6).

$$f_t = \sigma_g(W_f x_t + U_f h_{t-1} + b_f) \quad (1)$$

$$i_t = \sigma_g(W_i x_t + U_i h_{t-1} + b_i) \quad (2)$$

$$s_t = \tanh(W_c \cdot [W_c x_t + U_c h_{t-1}] + b_c) \quad (3)$$

$$c_t = f_t \cdot c_{t-1} + i_t \cdot s_t \quad (4)$$

$$o_t = \sigma_g(W_o x_t + U_o h_{t-1} + b_o) \quad (5)$$

$$h_t = o_t \cdot \tanh(c_t) \quad (6)$$

where $W_f, b_f, W_i, b_i, W_c, b_c, W_o,$ and b_o are the weights and offsets of each threshold layer. $\sigma,$ and \tanh are the activation functions.

D. BP NEURAL NETWORK

The BP neural network consists of an input layer, hidden layer, and output layer. The hidden interlayer can be extended to multiple layers. Each neuron between adjacent layers is fully connected, and there is no connection between each neuron in each layer [35]. The main advantage of the BP neural network is its strong nonlinear mapping ability. In theory, a BP neural network with three or more layers can predict a nonlinear function with arbitrary precision so long as the number of neurons in the hidden layer is sufficient.

The BP neural network model can be described as follows:

$$\begin{cases} z_j = f \left(\sum_{i=1}^m W_{ij} x_i + b_j \right) \\ y_k = f \left(\sum_{j=1}^n W_{jk} z_j + b_k \right) \end{cases} \quad (7)$$

where x_i is the input layer; z_j is the hidden layer, and y_k is the output layer. W_{ij} and W_{jk} are the weights; W_{ij} represents the weight between the i neuron in the input layer and the j neuron in the hidden layer, and W_{jk} represents the weight

between the j neuron in the hidden layer and the k neuron in the output layer; f is the excitation function; b_j and b_k are the biases. The excitation function expression is as follows:

$$\text{Sigmoid}(x) = \frac{1}{1 + e^{-x}} \quad (8)$$

Because the neural network model has numerous weights and biases, algorithms are required to modify these weights. We used the BP algorithm to train the neural network based on the model. Forward propagation of the input signal and BP of the output error were also used in the training of the BP neural network. The input samples are transported from the input layer to the output layer after being processed by the neurons in the hidden layer during the forward propagation process. The error BP procedure is transmitted if the error between the measured output and the expected output of the output layer does not fulfill the requirements. The error information is passed back to each layer's neurons along the original connection path during BP, and the network changes the connection weights of neurons using the gradient descent method. Minimize the network's real output until it matches the desired result to achieve an ideal network [36].

E. CLASSIFICATION AND REGRESSION TREE

CART uses a binary discretization methodology to discretize continuous data, providing the model to be used for both classification and regression tasks [37]. The CART model utilizes Gini impurity as a decision-making index for choosing node feature variables, which solves the problem that information gain favors attributes with higher values.

Suppose the dataset $D = \{(x_1, y_1), (x_2, y_2), \dots, (x_N, y_N)\}$, where a feature variable X_j is in $X \subseteq R^m, j = 1, 2, \dots, m$ has q values; then, the Gini impurity is defined as follows:

$$\text{Gini}(X_j) = \sum_{i=1}^q p_i(1 - p_i) \quad (9)$$

where p_i is the probability that the characteristic variable value is i .

The CART model generates decision rules that do not conflict or have less disagreement with the decision space because it learns from the data of the training attribute space. As a result, there is more than one reasonable decision tree, and comparing the relative quality of multiple decision tree models based only on their degree of fit to the training set is impossible. Additionally, because a single decision tree selects feature variables only once at each node and there is no iterative feedback, it is simple to obtain the local optimal solution.

F. K-NEAREST NEIGHBOR

KNN is a distance-based nonparametric algorithm [38]. For an input (x_0) in the test set, the core idea of KNN is to establish a vector space model in the training set, and find the K points closest to the distance based on the metric, which is represented by the set R_K :

$$R_K = \{(x_1, y_1), (x_2, y_2), \dots, (x_K, y_K)\}^T \quad (10)$$

Then, the output of x can be expressed as the average of K output values:

$$\hat{y} = \frac{1}{k} \sum_{x_k \in R_K} y_k \quad (11)$$

KNN has a variety of distance metric functions; the most used is Euclidean distance, for which the calculation method is as follows:

$$\text{dist}(x_0, x_k) = \sqrt{\sum_{s=1}^S (x_0^{(s)} - x_k^{(s)})^2} \quad (12)$$

For high-dimensional input, the Gaussian kernel function is often used to assign weights to K nearest neighbors. The calculation method is as follows:

$$W_K(x_0, x_k) = \frac{1}{K} \exp\left\{-\frac{\|x_k - x_0\|^2}{2K}\right\} \quad (13)$$

During training, the fitting ability of the KNN algorithm can be improved by adjusting the K value.

G. SUPPORT VECTOR REGRESSION

SVR is an application of support vector machines to regression problems. The core idea of SVR is to find a hyperplane that maximizes the interval between classes, and the principle of SVR is similar [39]. In SVR, the mapping from X to Y is represented as follows:

$$\hat{Y} = \vartheta(X) = w^T X + b \quad (14)$$

where represents the weight of the input. b is the offset.

For the samples (x_i, y_i) in the training set, use ξ_i to represent the deviation between \hat{Y}_i and Y_i , and ε define as the maximum value of the deviation; then, the deviation function is expressed as follows:

$$\text{Loss}_\varepsilon(\xi) = \begin{cases} 0, & |\xi| < \varepsilon \\ |\xi| - \varepsilon, & \text{other} \end{cases} \quad (15)$$

To min bias, the following optimization problem is established to solve ω and b :

$$\begin{aligned} \text{Min} & \frac{1}{2} \|\omega\|^2 + C \sum_{i=1}^{N_{in}} (\xi_i + \xi_i^*) \\ \text{s.t.} & \begin{cases} y_i - \omega x_i - b \leq \varepsilon + \xi_i \\ \omega x_i - b - y_i \leq \xi_i \\ \xi_i, \xi_i^* \geq 0 \end{cases} \end{aligned} \quad (17)$$

where C is a regularization constant to avoid overfitting problems.

IV. DATA COLLECTION AND CASE STUDY

To evaluate the performance of the proposed prediction method, we conducted a real-data-based case study and compared the various scenarios to ensure robustness. To determine the best traffic pattern, based on the best prediction method, we changed the time step of previous and later moment states to compare the prediction results of different traffic patterns.

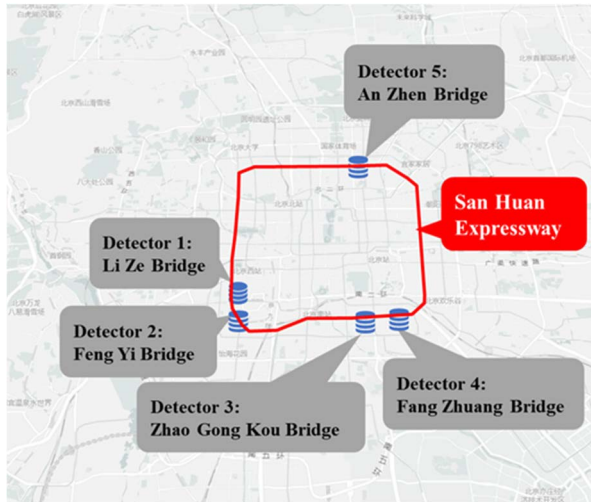


FIGURE 3. Distribution of five detectors on the third ring expressway.

A. OVERVIEW

The datasets used in this study are sampled from the road traffic management system. In this research, data are taken from five loop detectors on the Third Ring expressway: detector 1 (Li Ze Bridge), detector 2 (Feng Yi Bridge), detector 3 (Zhao Gong Kou Bridge), detector 4 (Fang Zhuang Bridge), and detector 5 (An Zhen Bridge). Traffic data were obtained from June 1 to 15, 2009. Each data detection point has 6 lanes with a speed limit of 80 km/h. The data includes time, traffic flow, lane number, traffic speed, and large vehicle scale. As shown in Fig. 3. We uploaded the dataset used in this study to GitHub (<https://github.com/gao0628/Dataset>).

Some abnormalities found in the dataset can be attributed to issues with the loop detectors or transmission interruptions between the local controller and the data center. Abnormal data may have a negative influence on the forecasting models and should be repaired before forecasting. The process of data cleaning is as follows:

1. **Abnormal data identification:** Simultaneously identify and estimate abnormal traffic volume and speed. For example, at a certain moment, the vehicle speed is an abnormal value, and the speed and traffic flow data at this moment need to be re-estimated. The abnormal data judgment rules are as follows: (i) The speed of the vehicle exceeds 80 km/h, and 80 km/h is the speed limit value of the road; and (ii) data exceptions due to interruptions in transmission between the loop detector and the local controller and data center. We used wavelet analysis to identify the location and extent of outliers [40]. By calculating the mean and variance of the detail coefficients in the wavelet analysis algorithm, we substituted the results into (18) to determine the location of abnormal data:

$$Thr(j, z) = \mu_j \pm z\sigma_j \tag{18}$$

where $Thr(j, z)$ is the boundary of abnormal data; μ_j is the average value of the detail coefficients at the j level ($j = 1$); σ_j

is the standard deviation of the detail coefficient at the j level ($j = 1$); and z is the value under the 95% confidence interval, where $z = 1.96$.

Here, five points around the abnormal point are determined as abnormal data [41]. Next, it is necessary to reestimate the abnormal data.

2. **Data reconstruction:** The Lagrange interpolation polynomial is a simple numerical estimation method that does not require the determination of the functional relationship between independent and dependent variables [42]. Therefore, we used Lagrange interpolation polynomial to reconstruct abnormal data. The abnormal data reconstruction process is shown in (19):

$$l_n(x) = \sum_{k=0}^n l_k(x)f(k) \tag{19}$$

where $l_n(x)$ is the Lagrangian polynomial interpolation, $f(k)$ is the interpolated function, and $l_k(x)$ is the n th degree polynomial, and its formula is as follows:

$$l_k(x) = \prod_{\substack{j=0 \\ j \neq k}}^n \frac{x - x_j}{x_k - x_j} \tag{20}$$

where j, k are known interpolation nodes.

3. **Data denoising:** After data reconstruction, there are some estimation errors. We assumed the errors are white noise and used signal filtering technology to filter and denoise the white noise. The Kalman filter is applied in various research fields because of its excellent filtering function [43]. Therefore, we chose the Kalman filter for data denoising. The Kalman filter has two main parts: the time update equation and the measurement update equation. The time update equation is responsible for the timely forward estimation of the current state variables and error covariance estimates. The measurement update equation is responsible for feedback; that is, it first combines the prior estimate with the new measurement variable to construct an improved posterior estimate. Fig. 4 shows the overall steps of the Kalman filter algorithm. Where “ $\hat{\cdot}$ ” is the estimated value; “ $-$ ” is the prior value; $y(t)$ is the vector of the state system; $y_o(t)$ is the vector of the observation system; $P(t)$ is the estimated error difference covariance; $K(t)$ is the Kalman gain; and \emptyset, T, H are coefficient matrices.

After completing the data identification and reconstruction algorithm, the data processing result shown in Fig. 5 can be obtained.

B. FORECASTING WITH THE DIFFERENT METHODS

As described in the previous section and shown in Fig. 1, this simulation was carried out for many existing methods to determine the optimal performance of the forecasting model. These methods included CART, SVR, KNN, BP neural network, and LSTM. All models in this study were built with TensorFlow, Keras, and Pandas package with Python. For CART, its core idea is to find the best node, the best branch,

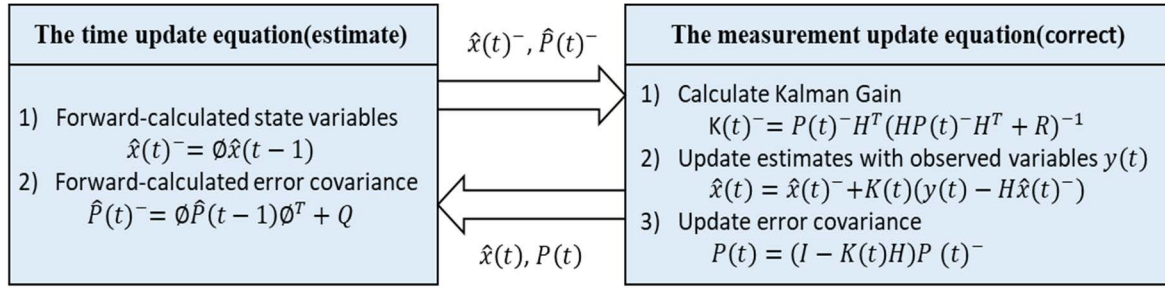


FIGURE 4. Kalman filter algorithm data processing.

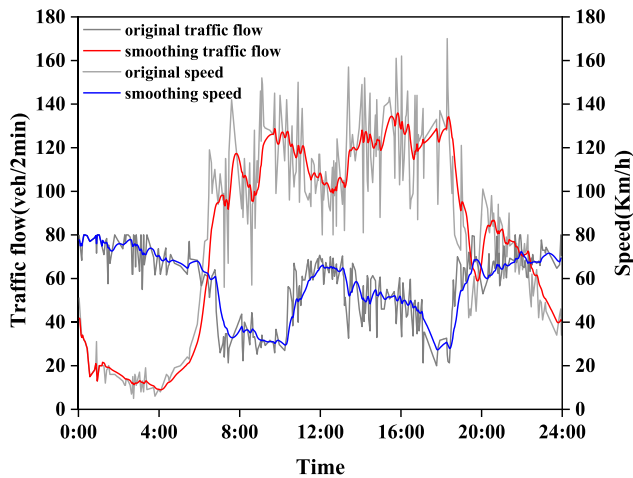


FIGURE 5. Results of original and smoothed data for traffic flow and speed.

and the key parameters to prevent the decision tree from overfitting. SVM uses the radial basis function (RBF) as the kernel function, and other parameters are determined by cross-validation, including penalty factor C, the nuclear parameter γ , and the upper error limit ϵ . For KNN, the Euclidean distance function is used as the distance between the calculated samples, and the parameter K value is also determined by cross-validation. For the BP neural network and LSTM, the number of hidden layer nodes, the maximum number of iterations, and the learning rate are used. All models use parameters that were obtained by trial and error through multiple experiments, which are widely used for problem solving, parameter tuning, or knowledge acquisition. In the experiment, the parameter setting of all models are shown in table 1.

Forecasting results contain T data points throughout the day. For each data point, we used APE to evaluate the forecasting accuracy, expressed as follows:

$$APE(t) = \frac{y(t) - \hat{y}(t)}{y(t)} \times 100\% \quad (21)$$

where $y(t)$ is the true value of the traffic speed in the t-th time interval, $\hat{y}(t)$ is the forecasted value of traffic speed in the t-th time interval, and $APE(t)$ denotes the absolute percentage error for traffic speed in the t-th time interval. For

TABLE 1. Model parameters.

Model	Parameter description
LSTM	Input size=7, Hidden size=64, Batch size=72, Step size=4, Epoch=250, Learning rate=0.001, Optimizer=Adam.
BP	Input size=7, Hidden size=64, Batch size=72, Hidden layer number=2, Epoch=250, Learning rate=0.01, Optimizer=Adam.
CART	Max depth=2, Min samples split=10, Min samples leaf=5, Max lead nodes=4.
KNN	Neighbors number=15.
SVR	Kernel='rbf', C=1, Epsilon=0.19.

the forecasting results of the whole day, the mean absolute percentage error (MAPE) is used to assess the performance, expressed as follows:

$$MAPE = \frac{1}{T} \sum_{t=1}^T |APE(t)| \quad (22)$$

The pattern to mine traffic patterns of previous and later moment states, that is, the value of n and m, is an important parameter that can influence the forecasting results. We chose the pattern of “n = 3 and m = 3” to compare the performance of all forecasting models. After determining the optimal forecasting model, we analyzed the results to determine the best traffic flow pattern (as discussed in the following sections). The forecasting results for a weekday (June 5) and a weekend day (June 7) are shown in Fig. 6. Fig. 6(a) shows that the MAPE of LSTM is about (3.83%), which is much lower than the results provided by BP (6.12%), CART (7.43%), KNN (15.67%), and SVR (16.06%). Fig. 6(b) shows that the MAPE of LSTM is about (2.05%), which is much lower than the results provided by BP (4.12%), CART (4.40%), and KNN (13.08%), and SVR (15.31%). However, from the APE of all methods, it can be seen that in some cases, simple machine learning methods can produce better forecasting results than deep learning methods. For example, the APE of CART may be lower than that of BP and LSTM in some cases. Therefore, it is necessary to compare the performance of forecasting

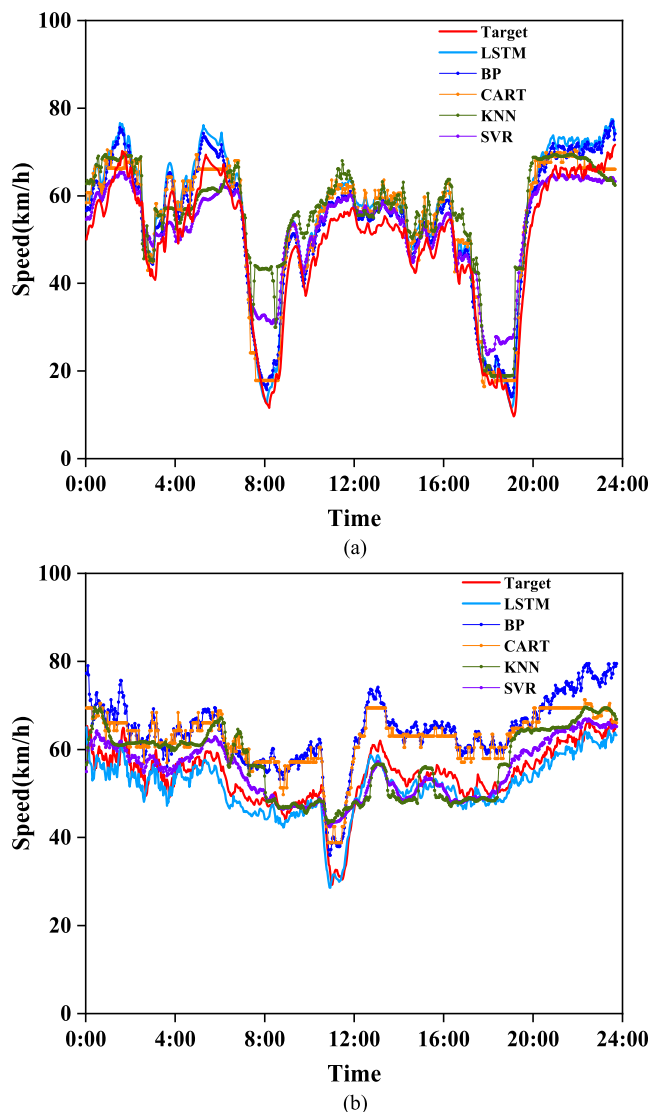


FIGURE 6. Forecasting results and real data on (a) June 5 (weekday); and (b) June 7 (weekend).

models in time and space dimensions to evaluate the performance of the model.

To further evaluate the performance of LSTM against existing methods, as shown in table 2, we presented forecasting performance across different days on detector 1 among all models. From the observation, it can be found that the average MAPE of the LSTM model is 3.29%, which is lower than the other models. It indicates that the LSTM model outperforms the other models and the fluctuation of MAPE is between 1% and 6% from June 1 to 15. This is due to LSTM can capture long-term memory features as the model input features to make the LSTM have a better prediction effect. In addition, a phenomenon worth considering is that the forecasting result of CART, SVR, and KNN is better than BP in some days. For example, CART’s MAPE is 3.68% and BP is 5.67% on June 7. However, their forecasting accuracy is not too stable.

To verify the robust performance of LSTM, as shown in table 3, we presented forecasting performance across different detectors on June 1 among all models. From the average MAPE of all models, we observed that the average MAPE of the LSTM is better than that of other methods, and the minimum MAPE is 1.32% at detector 2. It shows that the LSTM model has good stability. Therefore, we chose the LSTM model to determine the best traffic pattern.

C. DETERMINE THE BEST TRAFFIC PATTERN

To investigate the influence of the traffic pattern on the forecasting model, we compared the forecasting results with different flow features. The size of the traffic pattern should be moderate. If fewer data are input, there is less historical information for forecasting, making its prediction result MAPE poor. Moreover, the MAPEs are generally smaller for more data, which is expected because such forecasting tasks become relatively easier. Ma *et al.* explored the impact of different time intervals on the forecasting model and found that 15–20 minutes is the optimal time interval [20]. Therefore, we set the maximum interval to 20 minutes; that is, the maximum traffic pattern is “ $n = 5$ and $m = 5$.” On this basis, we carried out a random combination experiment of previous and later moment flow characteristics, with a total of 36 combination modes. We input 36 flow patterns to the LSTM forecasting model, and the speed model results are shown in Fig. 7. Fig. 7 shows the contour map of the flow patterns around Detector 1 and June 1, and the colors represent MAPE under different traffic patterns. Fig. 7 shows that the MAPE is smaller for this pattern “ $n = 2-5$ and $m = 2-5$.” However, the LSTM prediction results of “ $n = 0$ and $m = 0-5$ ” or “ $n = 0-5$ and $m = 0$ ” are generally poor, because only relying on the traffic flow of previous or later moment states cannot effectively characterize the current speed operation state. To predict the speed state more accurately, it is necessary to construct traffic patterns in conjunction with the flow of previous and later moment states.

To determine the optimal traffic pattern, we compared the traffic patterns of previous and later moment states from June 1 to 15, and the predicted results are shown in Fig. 8. Figure 8 shows that the prediction accuracy of this pattern of “ $n = 0$ and $m = 0-5$ ” or “ $n = 0-5$ and $m = 0$ ” is poor on all days, and the prediction results are also unstable. This result also verifies our previous conjecture that the previous or later moment states’ flow pattern alone cannot effectively predict the current speed operation state. We also see that the size of the “previous and later moment states” time step is not related to the predictive effect of the model. For example, “ $n = 5$ and $m = 5$ ” has a higher MAPE than “ $n = 5$ and $m = 4$ ” and “ $n = 4$ and $m = 5$,” which is expected because the size of ‘previous and later moment states’ is large, and may cross the peak of various traffic flows, resulting in a certain deviation between the predicted results and the actual values. By analyzing the average MAPE of all days, it is finally determined that the best traffic pattern is “ $n = 3$ and $m = 3$.”

TABLE 2. Performance comparison of different models on Detector 1, MAPE from June 1 to 15.

Days	LSTM	BP	CART	KNN	SVR
June 1	2.47%	5.68%	4.27%	6.19%	7.51%
June 2	4.03%	5.30%	9.97%	20.44%	17.48%
June 3	2.77%	4.68%	5.37%	6.87%	8.74%
June 4	2.63%	5.07%	4.72%	8.14%	7.97%
June 5	4.11%	3.31%	7.39%	14.26%	16.46%
June 6	2.20%	6.12%	3.95%	6.80%	6.73%
June 7	2.10%	5.67%	3.68%	10.06%	13.90%
June 8	5.58%	6.89%	8.35%	12.14%	11.68%
June 9	4.55%	7.48%	13.84%	17.77%	14.80%
June 10	3.33%	4.98%	10.01%	14.78%	14.53%
June 11	4.73%	6.84%	9.34%	15.19%	13.01%
June 12	1.75%	4.58%	5.42%	8.05%	12.73%
June 13	2.74%	4.40%	5.04%	6.26%	4.02%
June 14	3.29%	5.61%	6.53%	10.46%	12.48%
June 15	3.06%	3.08%	4.87%	9.63%	12.26%
Average	3.29%	5.31%	6.85%	11.13%	11.62%

TABLE 3. Performance comparison of different models on June 1 MAPE from Detector 1 to 5.

Detectors	LSTM	BP	CART	KNN	SVR
Detectors 1	1.43%	3.45%	4.01%	18.16%	20.03%
Detectors 2	1.32%	3.65%	3.12%	3.86%	5.34%
Detectors 3	3.95%	4.06%	7.31%	14.26%	16.46%
Detectors 4	2.83%	4.91%	6.68%	8.86%	9.23%
Detectors 5	5.77%	6.79%	15.51%	23.64%	24.97%
Average	3.06%	4.57%	7.33%	13.76%	15.21%

D. ROBUSTNESS OF THE BEST TRAFFIC PATTERN

To further illustrate the performance of the best traffic pattern, we conducted several robust analyses across different contexts. We evaluated the spatiotemporal robustness of the best traffic pattern.

The forecasting results using the LSTM algorithm for detector 1 across 15 target days for the best traffic pattern “ $n = 3$ and $m = 3$ ” are shown in Fig. 9. It can be observed that the prediction accuracy of all days is relatively robust, and most of the APEs are lower than 15%. The previous analyses focused on the data from one loop detector on the Third Ring Expressway. To show the robustness across different spatial contexts, the best traffic pattern was also applied with data from five additional detectors on the Third Ring Expressway. The APE distribution for the five additional detectors on June 1 is presented in Fig. 10. As can be seen from Fig. 10, the prediction accuracy of all detection points is relatively

good. However, the APE of detector 5 is larger than that of other detector points during the peak period. This is mainly because the point of detector 5 is distributed in the North Third Ring Road, and a large number of Expressway inlet and outlet ramps are located around this point. During peak hours, vehicles in this area often change lanes to complete driving tasks, resulting in unstable traffic flow. Therefore, the accuracy of the prediction model is low in terms of APE. In addition, there is no need to pay special attention to the low accuracy of the prediction model caused by complex road conditions, because the prediction results are also acceptable at these detection points.

In addition, we summarized the MAPE of June 1 to 15 and detectors 1 to 5, as shown in Fig. 11. The MAPE of all detectors and all days in Fig. 11 is lower than 9%, which further illustrates the robust performance of the best traffic pattern in time and space dimensions. Similarly, we can also

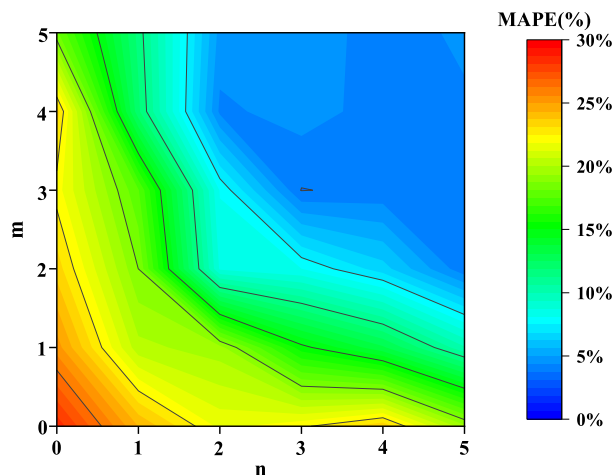


FIGURE 7. MAPE for different combinations of traffic flow patterns (n represents the time step of previous states and m represents the time step of later moment states).

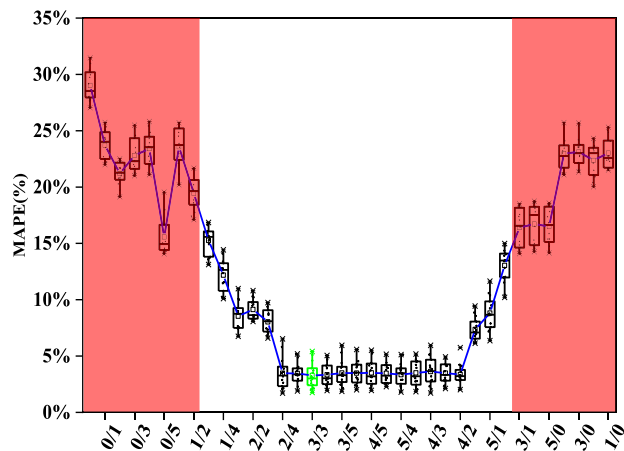


FIGURE 8. MAPE of different combinations of traffic flow patterns in 15 days (3/3 represents this pattern “ $n = 3$ and $m = 3$ ”).

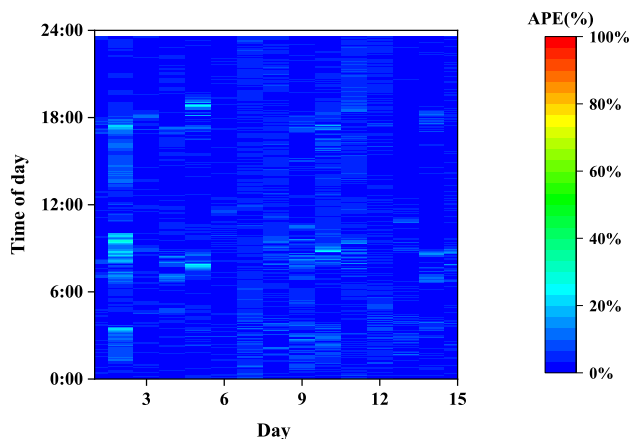


FIGURE 9. APE distribution across different days. (Detector 1, June 1-15, APE within 2 minutes.)

see that the MAPE of detector 5 is relatively large on all days, which further verifies our previous conjecture that the

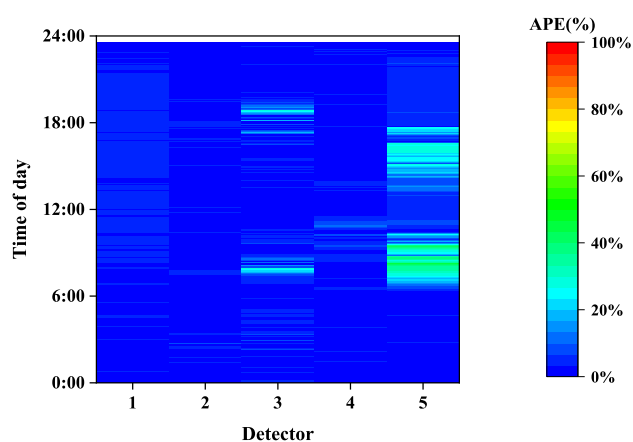


FIGURE 10. APE distribution across different detectors. (June 1, Detector 1-5, APE within 2 minutes.)

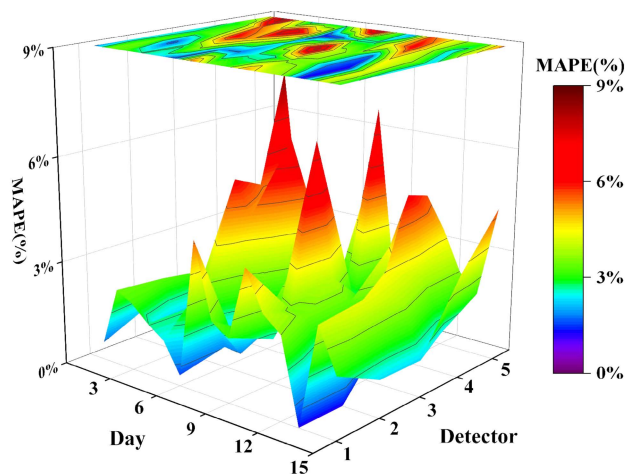


FIGURE 11. MAPE on different days and different detectors.

complex geometric conditions around the region also have a certain impact on the prediction results.

V. CONCLUSION

The running status of expressways or highways can be represented by traffic speed. We proposed a speed prediction methodology based on the traffic flow of previous and later moment states based on the relationship between flow and speed in traffic flow theory. First, we revised the dataset on the Beijing Third Ring Expressway. Second, we compared both traditional machine learning methods and deep learning methods in terms of prediction performance in the dimension of spatiotemporal features and concluded that the LSTM method had the best effect. Finally, we studied the time step of previous and later moment states and discovered that the pattern “ $n = 3$ and $m = 3$ ” had the best speed prediction effect, with robust forecast accuracy in diverse time and space.

The method proposed in this study can quickly sense the operational status of the road network by relying only on traffic flow data. It helps travelers plan their travel routes

and helps road management personnel identify congestion points and formulate congestion mitigation strategies. For example, the Electronic Toll Collection (ETC) system on the highway can acquire traffic flow data but not vehicle running speed information, and flow data alone cannot adequately reflect the highway's operating status. Based on the flow data from the ETC system, we can forecast the speed and further assess the highway's operating status.

Other limitations of the proposed method exist, and it can be improved in the following ways. First, the traffic flow collected by some neighboring detectors is correlated. Therefore, one of our future study areas may be to extend the proposed method to forecast traffic speed for multiple detectors at once or even a large-scale network. Second, the performance of traffic speed forecasting could be enhanced by further assessing the road conditions and traffic conditions at the detection location, such as the detector's upstream and downstream road conditions, road speed limit, and other parameters. Third, we can optimize the extraction method of the previous and later moment states traffic flow information to raise prediction performance. Fourth, only the flow characteristic factor is currently considered; if more factors such as external factors (e.g., the environment and weather) are incorporated, the results may improve.

REFERENCES

- [1] J. Zhang, F.-Y. Wang, K. Wang, W.-H. Lin, X. Xu, and C. Chen, "Data-driven intelligent transportation systems: A survey," *IEEE Trans. Intell. Transp. Syst.*, vol. 12, no. 4, pp. 1624–1639, Dec. 2011, doi: [10.1109/TITS.2011.2158001](https://doi.org/10.1109/TITS.2011.2158001).
- [2] F. G. Habtemichael and M. Cetin, "Short-term traffic flow rate forecasting based on identifying similar traffic patterns," *Transp. Res. C, Emerg. Technol.*, vol. 66, pp. 61–78, May 2015, doi: [10.1016/j.trc.2015.08.017](https://doi.org/10.1016/j.trc.2015.08.017).
- [3] P. Vovsha and S. Bekhor, "Link-nested logit model of route choice: Overcoming route overlapping problem," *Transp. Res. Rec., J. Transp. Res. Board*, vol. 1645, no. 1, pp. 133–142, Jan. 1998, doi: [10.3141/1645-17](https://doi.org/10.3141/1645-17).
- [4] F. Remppe, P. Franek, and K. Bogenberger, "On the estimation of traffic speeds with deep convolutional neural networks given probe data," *Transp. Res. C, Emerg. Technol.*, vol. 134, Jan. 2022, Art. no. 103448, doi: [10.1016/j.trc.2021.103448](https://doi.org/10.1016/j.trc.2021.103448).
- [5] C.-J. Jin, W. Wang, R. Jiang, H. M. Zhang, H. Wang, and M.-B. Hu, "Understanding the structure of hyper-congested traffic from empirical and experimental evidences," *Transp. Res. C, Emerg. Technol.*, vol. 60, pp. 324–338, Nov. 2015, doi: [10.1016/j.trc.2015.09.007](https://doi.org/10.1016/j.trc.2015.09.007).
- [6] M. Ahmed and A. Cook, "Analysis of freeway traffic time series data by using box-jenkins techniques," *Transp. Res. Rec.*, vol. 773, pp. 1–9, Jan. 1979.
- [7] E. I. Vlahogianni, M. G. Karlaftis, and J. C. Golias, "Short-term traffic forecasting: Where we are and where we're going," *Transp. Res. C, Emerg. Technol.*, vol. 43, pp. 3–19, Jun. 2014, doi: [10.1016/j.trc.2014.01.005](https://doi.org/10.1016/j.trc.2014.01.005).
- [8] M. Lippi, M. Bertini, and P. Frasconi, "Short-term traffic flow forecasting: An experimental comparison of time-series analysis and supervised learning," *IEEE Trans. Intell. Transp. Syst.*, vol. 14, no. 2, pp. 871–882, Jun. 2013, doi: [10.1109/TITS.2013.2247040](https://doi.org/10.1109/TITS.2013.2247040).
- [9] Y. Feng and S. Wang, "A forecast for bicycle rental demand based on random forests and multiple linear regression," in *Proc. IEEE/ACIS 16th Int. Conf. Comput. Inf. Sci. (ICIS)*, May 2017, pp. 101–105, doi: [10.1109/ICIS.2017.7959977](https://doi.org/10.1109/ICIS.2017.7959977).
- [10] M. A. Wani and H. F. Bhat, "Multiclass SVM algorithms for wind speed prediction," in *Proc. IEEE 6th Int. Conf. Renew. Energy Res. Appl. (ICRERA)*, Nov. 2017, pp. 1139–1143, doi: [10.1109/ICRERA.2017.8191231](https://doi.org/10.1109/ICRERA.2017.8191231).
- [11] M. Duo, Y. Qi, G. Lina, and E. Xu, "A short-term traffic flow prediction model based on EMD and GPSO-SVM," in *Proc. IEEE 2nd Adv. Inf. Technol., Electron. Autom. Control Conf. (IAEAC)*, Mar. 2017, pp. 2554–2558, doi: [10.1109/IAEAC.2017.8054485](https://doi.org/10.1109/IAEAC.2017.8054485).
- [12] M. Cools, E. Moons, and G. Wets, "Investigating the variability in daily traffic counts through use of ARIMAX and SARIMAX models," *Transp. Res. Rec., J. Transp. Res. Board*, vol. 2136, no. 1, pp. 57–66, Jan. 2009, doi: [10.3141/2136-07](https://doi.org/10.3141/2136-07).
- [13] S. Lee and D. B. Fambro, "Application of subset autoregressive integrated moving average model for short-term freeway traffic volume forecasting," *Transp. Res. Rec., J. Transp. Res. Board*, vol. 1678, no. 1, pp. 179–188, Jan. 1999, doi: [10.3141/1678-22](https://doi.org/10.3141/1678-22).
- [14] B. L. Smith, B. M. Williams, and R. K. Oswald, "Comparison of parametric and nonparametric models for traffic flow forecasting," *Transp. Res. C, Emerg. Technol.*, vol. 10, no. 4, pp. 303–321, Aug. 2002, doi: [10.1016/S0968-090X\(02\)00009-8](https://doi.org/10.1016/S0968-090X(02)00009-8).
- [15] B. Williams and L. Hoel, "Modeling and forecasting vehicular traffic flow as a seasonal ARIMA process: Theoretical basis and empirical results," *J. Transp. Eng.*, vol. 129, no. 6, pp. 664–672, Nov./Dec. 2003, doi: [10.1061/\(ASCE\)0733-947X\(2003\)129:6\(664\)](https://doi.org/10.1061/(ASCE)0733-947X(2003)129:6(664)).
- [16] E. I. Vlahogianni, J. C. Golias, and M. G. Karlaftis, "Short-term traffic forecasting: Overview of objectives and methods," *Transp. Res.*, vol. 24, no. 5, pp. 533–557, Sep. 2004, doi: [10.1080/0144164042000195072](https://doi.org/10.1080/0144164042000195072).
- [17] S. R. Chandra and H. Al-Deek, "Predictions of freeway traffic speeds and volumes using vector autoregressive models," *J. Intell. Transp. Syst.*, vol. 13, no. 2, pp. 53–72, May 2009, doi: [10.1080/15472450902858368](https://doi.org/10.1080/15472450902858368).
- [18] T. L. Pan, A. Sumalee, R. X. Zhong, and N. Indra-payoong, "Short-term traffic state prediction based on temporal-spatial correlation," *IEEE Trans. Intell. Transp. Syst.*, vol. 14, no. 3, pp. 1242–1254, Sep. 2013, doi: [10.1109/TITS.2013.2258916](https://doi.org/10.1109/TITS.2013.2258916).
- [19] Y. Gu, W. Lu, L. Qin, M. Li, and Z. Shao, "Short-term prediction of lane-level traffic speeds: A fusion deep learning model," *Transp. Res. C, Emerg. Technol.*, vol. 106, pp. 1–16, Sep. 2019, doi: [10.1016/j.trc.2019.07.003](https://doi.org/10.1016/j.trc.2019.07.003).
- [20] 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 Trans. Intell. Transp. Syst.*, vol. 22, no. 5, pp. 2627–2636, May 2021, doi: [10.1109/TITS.2020.2973279](https://doi.org/10.1109/TITS.2020.2973279).
- [21] H. Nguyen, L.-M. Kieu, T. Wen, and C. Cai, "Deep learning methods in transportation domain: A review," *IET Intell. Transp. Syst.*, vol. 12, no. 9, pp. 998–1004, 2018, doi: [10.1049/iet-its.2018.0064](https://doi.org/10.1049/iet-its.2018.0064).
- [22] J. Zheng and M. Huang, "Traffic flow forecast through time series analysis based on deep learning," *IEEE Access*, vol. 8, pp. 82562–82570, 2020, doi: [10.1109/ACCESS.2020.2990738](https://doi.org/10.1109/ACCESS.2020.2990738).
- [23] H.-F. Yang, T. S. Dillon, and Y.-P. P. Chen, "Optimized structure of the traffic flow forecasting model with a deep learning approach," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 28, no. 10, pp. 2371–2381, Oct. 2017, doi: [10.1109/TNNLS.2016.2574840](https://doi.org/10.1109/TNNLS.2016.2574840).
- [24] X. Meng, H. Fu, L. Peng, G. Liu, Y. Yu, Z. Wang, and E. Chen, "D-LSTM: Short-term road traffic speed prediction model based on GPS positioning data," *IEEE Trans. Intell. Transp. Syst.*, vol. 23, no. 3, pp. 2021–2030, Mar. 2022, doi: [10.1109/TITS.2020.3030546](https://doi.org/10.1109/TITS.2020.3030546).
- [25] S. Cheng, F. Lu, P. Peng, and S. Wu, "Short-term traffic forecasting: An adaptive ST-KNN model that considers spatial heterogeneity," *Comput., Environ. Urban Syst.*, vol. 71, pp. 186–198, Sep. 2018, doi: [10.1016/j.compenurbysys.2018.05.009](https://doi.org/10.1016/j.compenurbysys.2018.05.009).
- [26] X. Ma, Z. Tao, Y. Wang, H. Yu, and Y. Wang, "Long short-term memory neural network for traffic speed prediction using remote microwave sensor data," *Transp. Res. C, Emerg. Technol.*, vol. 54, pp. 187–197, May 2015, doi: [10.1016/j.trc.2015.03.014](https://doi.org/10.1016/j.trc.2015.03.014).
- [27] C. Zhang, S. Sun, and G. Yu, "A Bayesian network approach to time series forecasting of short-term traffic flows," in *Proc. 7th Int. IEEE Conf. Intell. Transp. Syst.*, Oct. 2004, pp. 216–221, doi: [10.1109/ITSC.2004.1398900](https://doi.org/10.1109/ITSC.2004.1398900).
- [28] J. Yun, S.-H. Paek, D. Lee, and H. Yoon, "Traffic volume forecasting model for cyber weather forecasting," in *Proc. Int. Conf. Hybrid Inf. Technol.*, Nov. 2006, pp. 274–276, doi: [10.1109/ICHIT.2006.253623](https://doi.org/10.1109/ICHIT.2006.253623).
- [29] H. Peng, S. U. Bobade, M. E. Cotterell, and J. A. Miller, "Forecasting traffic flow: Short term, long term, and when it rains," in *Big Data—Big Data (Lecture Notes in Computer Science)*, vol. 10968. Cham, Switzerland: Springer, 2018, doi: [10.1007/978-3-319-94301-5_5](https://doi.org/10.1007/978-3-319-94301-5_5).
- [30] E. I. Vlahogianni and M. G. Karlaftis, "Comparing traffic flow time-series under fine and adverse weather conditions using recurrence-based complexity measures," *Nonlinear Dyn.*, vol. 69, no. 4, pp. 1949–1963, Sep. 2012, doi: [10.1007/s11071-012-0399-x](https://doi.org/10.1007/s11071-012-0399-x).
- [31] S. Dunne and B. Ghosh, "Weather adaptive traffic prediction using neural network models," *IEEE Trans. Intell. Transp. Syst.*, vol. 14, no. 1, pp. 370–379, Mar. 2013, doi: [10.1109/TITS.2012.2225049](https://doi.org/10.1109/TITS.2012.2225049).

[32] Y. Jia, J. Wu, and M. Xu, "Traffic flow prediction with rainfall impact using a deep learning method," *J. Adv. Transp.*, vol. 2017, pp. 1–10, Aug. 2017, doi: [10.1155/2017/6575947](https://doi.org/10.1155/2017/6575947).

[33] R. Ke, W. Li, Z. Cui, and Y. Wang, "Two-stream multi-channel convolutional neural network for multi-lane traffic speed prediction considering traffic volume impact," *Transp. Res. Rec., J. Transp. Res. Board*, vol. 2674, no. 4, pp. 459–470, Apr. 2020, doi: [10.1177/0361198120911052](https://doi.org/10.1177/0361198120911052).

[34] J. Ma, H. Liu, C. Peng, and T. Qiu, "Unauthorized broadcasting identification: A deep LSTM recurrent learning approach," *IEEE Trans. Instrum. Meas.*, vol. 69, no. 9, pp. 5981–5983, Sep. 2020, doi: [10.1109/TIM.2020.3008988](https://doi.org/10.1109/TIM.2020.3008988).

[35] Q. Lu, R. Yang, M. Zhong, and Y. Wang, "An improved fault diagnosis method of rotating machinery using sensitive features and RLS-BP neural network," *IEEE Trans. Instrum. Meas.*, vol. 69, no. 4, pp. 1585–1593, Apr. 2020, doi: [10.1109/TIM.2019.2913057](https://doi.org/10.1109/TIM.2019.2913057).

[36] Z. Yu, L. Qin, Y. Chen, and M. D. Parmar, "Stock price forecasting based on LLE-BP neural network model," *Phys. A, Stat. Mech. Appl.*, vol. 553, Sep. 2020, Art. no. 124197, doi: [10.1016/j.physa.2020.124197](https://doi.org/10.1016/j.physa.2020.124197).

[37] R. Tang and X. Zhang, "CART decision tree combined with Boruta feature selection for medical data classification," in *Proc. 5th IEEE Int. Conf. Big Data Anal. (ICBDA)*, May 2020, pp. 80–84, doi: [10.1109/ICBDA49040.2020.9101199](https://doi.org/10.1109/ICBDA49040.2020.9101199).

[38] N. Ling, G. Aneiros, and P. Vieu, "KNN estimation in functional partial linear modeling," *Stat. Papers*, vol. 61, no. 1, pp. 423–444, Feb. 2020, doi: [10.1007/s00362-017-0946-0](https://doi.org/10.1007/s00362-017-0946-0).

[39] H. Zeynal, Z. Zakaria, and A. Kor, "Traffic flow prediction using SVR-ant colony optimization: A practical case of Tehran highway," in *Proc. IEEE Int. Conf. Power Eng. Appl. (ICPEA)*, Mar. 2021, pp. 155–160, doi: [10.1109/ICPEA51500.2021.9417845](https://doi.org/10.1109/ICPEA51500.2021.9417845).

[40] Y. Li and S. Zhou, "A new instantaneous wavelet bicoherence for local fault detection of rotating machinery," *IEEE Trans. Instrum. Meas.*, vol. 69, no. 1, pp. 135–143, Jan. 2020, doi: [10.1109/TIM.2019.2895483](https://doi.org/10.1109/TIM.2019.2895483).

[41] E. Cabrera-Navarrete, G. Ronquillo-Jarillo, and A. Markov, "Wavelet analysis for spectral inversion of seismic reflection data," *J. Appl. Geophys.*, vol. 177, Jun. 2020, Art. no. 104034, doi: [10.1016/j.jappgeo.2020.104034](https://doi.org/10.1016/j.jappgeo.2020.104034).

[42] M. Errachid, A. Essanhaji, and A. Messaoudi, "RMVPIA: A new algorithm for computing the Lagrange multivariate polynomial interpolation," *Numer. Algorithms*, vol. 84, no. 4, pp. 1507–1534, Aug. 2020, doi: [10.1007/s11075-020-00907-8](https://doi.org/10.1007/s11075-020-00907-8).

[43] K. Turgut and A. K. Hocaoglu, "Comparison of Kalman filter and converted measurements Kalman filter for interacting multiple model," in *Proc. 26th Signal Process. Commun. Appl. Conf.*, 2018, pp. 1–4,



CHENJING ZHOU received the B.S. degree in transport engineering from the Beijing Institute of Technology, Beijing, China, in 2008, and the Ph.D. degree in transportation planning and management from the Beijing University of Technology, Beijing, in 2015. He is currently a Professor in the School of Civil Engineering at Guangzhou University, China. His research interests include deep learning, traffic simulation, and intelligent transportation.



JIAN RONG received the B.S. degree in applied mechanics from the Beijing Institute of Technology, Beijing, China, in 1994, and the Ph.D. degree in transportation planning and management from the Beijing University of Technology, Beijing, in 1999. He is currently a Professor with the School of Civil Engineering, Guangzhou University, China. He has expertise and experience in intelligent transportation and machine learning for more than the last 20 years. He has published more than 100 papers in the field of intelligent transportation. His research interests include deep learning, highway capacity, traffic flow, and machine learning.



YI WANG received the B.S. degree from the Beijing Institute of Technology, Zhuhai, China, the master's degree in transportation engineering from Chongqing Jiaotong University, Chongqing, China, and the Ph.D. degree in transportation engineering from the Beijing University of Technology. He is currently an Assistant Lecturer with the School of Civil Engineering, Guangzhou University, Guangzhou, China. His research interests include deep learning, highway capacity, traffic flow, and machine learning.



YACONG GAO received the B.S. degree in transport engineering from the Hebei University of Science and Technology, Hebei, China, in 2018, and the M.S. degree in transportation engineering from the Beijing University of Technology, Beijing, China, in 2021, where he is currently pursuing the Ph.D. degree in transportation engineering. His research interests include the areas of deep learning and traffic flow prediction.



SIYANG LIU received the B.S. degree in transport engineering from the Beijing University of Civil Engineering and Architecture, Beijing, China, in 2017. He is currently pursuing the Ph.D. degree in transportation engineering with the Beijing University of Technology, Beijing. His research interests include deep learning, travel behavior, and transit-oriented development.

...