*Research Article*

# Short-Term Traffic State Prediction Based on the Critical Road Selection Optimization in Transportation Networks

**Tian Ma ⓘ,[1] Guanghong Gong ⓘ,[1] and Yilong Ren ⓘ[2]**

[1]*School of Automation Science and Electrical Engineering, Beihang University, Beijing 100191, China*
[2]*School of Transportation Science and Engineering, Beihang University, Beijing 100191, China*

Correspondence should be addressed to Yilong Ren; yilongren@buaa.edu.cn

Short-term traffic prediction under corrupted or missing data for large-scale transportation networks has become an important and challenging topic in recent decades. Since the critical roads have predictive power on their adjacent roads, this paper proposes a novel hybrid short-term traffic state prediction method based on critical road selection optimization. First, the utility function of the quality of service (QoS) for the critical roads in a large-scale road network is proposed based on the coverage and the data score. Then, the critical road selection optimization model in the transportation networks is presented by selecting an appropriate set of critical roads with the maximum proportion of the total calculation resources to maximize the utility value of the QoS. Also, an innovative critical road selection method is introduced, which is considering the topological structure and the mobility of the urban road network. Subsequently, the traffic speed of the critical roads is regarded as the input of the convolutional long short-term memory neural network to predict the future traffic states of the entire network. Experiment results on the Beijing traffic network indicate that the proposed method outperforms prevailing DL approaches in the case of considering critical road sections.

## 1. Introduction

Real-time traffic state prediction plays a vital role in traffic management and public service. By predicting the evolution of traffic timely and accurately, governments and travelers could react to the traffic congestion ahead of time. For instance, intelligent transportation systems, advanced traffic management systems, and traveler information systems depend on real-time traffic state prediction.

In the past decades, there have been numerous studies on this topic [1]. These studies could be divided into mathematical models, statistical models, and data-driven methods [2]. Compared with data-driven methods, mathematical or statistical models derived from macroscopic and microscopic theories of traffic flow are difficult to handle unstable traffic conditions and complex road settings due to strong hypotheses and assumptions [3].

Data-driven methods have achieved promising results due to more potential in processing complex nonlinear problems

[4]. These methods include support vector machine (SVM), Bayesian network, and neural network. Among all these data-driven methods, deep learning approaches have proven effective in traffic state prediction. This method could exploit much deeper architectures and process the high-dimensional set of explanatory variables [5].

However, most of the deep learning approaches are constructed based on the entire dataset. The attributes of datasets influence the prediction performance. In other words, the deep learning approaches have highlighted the data quality in short-term traffic state prediction [5]. Due to the corrupted or missing data problem, there is limited high-quality real-time or historical data obtained, indicating that the data has low predictive quality [6]. These setbacks weaken the predictive accuracy and efficiency, limiting the capacity for providing a practical and reliable forecasting result.

Recently, researchers find that some critical roads significantly affect the traffic states of their adjacent roads at a specific road network [7, 8]. This conclusion indicates that

the traffic state of one road could be predicted based on its neighbors' state [6].

There have been numerous studies on this topic, but only a tiny percentage of them have paid attention to prediction based on critical road selection optimization. Therefore, if we extract the traffic state of the critical road sections with the most dominant predictive power, we could characterize the spatiotemporal features of traffic flow and predict the future traffic state of the overall network.

In this paper, we propose a novel hybrid short-term traffic state prediction method based on critical road selection optimization. Different scores were given to data collected on different road segments. The higher the criticality of the road segment, the higher the score of the data. The critical road selection problem is abstracted as a multiobjective optimization problem, maximizing the sensing coverage and data scores. In other words, our objective is to select the most suitable critical roads to maximize the quality of the service (QoS) with the limited consumption of network resources.

Then, a novel critical road selection method is proposed, considering the topological structure and the mobility of the urban road network. Subsequently, the traffic speed of the critical roads is regarded as the input of the convolutional long short-term memory neural network to predict the future traffic states of the entire network. Finally, to demonstrate the effectiveness of the proposed method, the numerical experiments using the traffic states depicted from GPS trajectory data in Beijing. In addition, other traditional machine-learning models are compared to demonstrate the advantage of the proposed method. Practical experiment results showed that the proposed method could precisely predict future traffic network states.

The rest of our paper is organized as follows. Section 2 discusses the literature on short-term traffic state prediction and the existing approaches to prediction based on critical roads. In Section 3, the critical road selection optimization model is proposed by selecting an appropriate set of critical roads with the maximum proportion of the total calculation resources to maximize the utility value of the QoS and also an innovative critical road selection method is introduced. Section 4 elaborates on the numerical experiments using the traffic states depicted from GPS trajectory data in Beijing. The last section concludes the study and discusses future work.

## 2. Literature Review

Over the past decades, considerable short-term traffic forecast models have emerged to handle the prediction of future traffic states ranging from a few seconds to few hours. These approaches could be generally categorized into mathematical approaches, statistical approaches, and data-driven approaches.

Mathematical approaches focus on predetermining the model structure by theoretical assumption. The evolution of the traffic state could be simulated by the theoretical mathematical models [9]. Because of the strong dependence on theoretical considerations, these mathematical

approaches do not match the actual traffic state condition [10]. Mathematical approaches depend on theoretical mathematical models to simulate the traffic evolution. There are many famous mathematical models in the past decades, such as the Greenshields model, car-following model, Van Lighthill–Whitham–Richards (LWR) model, and so on. The future traffic state of the road network could be calculated by these models directly [11, 12]. Most traffic simulation simulators, such as Q-Paramics, VISSIM, DynaMIT, and DynaSMART-X, predict traffic states based on the traffic flow models, car-following models, and dynamic traffic assignment models [13–16].

Statistical approaches usually rely on statistical assumptions. The autoregressive integrated moving average (ARIMA) and its variants are the most common mathematical approaches. Hamed et al. [17] proposed a simple ARIMA to predict the traffic state of the road network. Williams et al. developed a seasonal ARIMA (S-ARIMA) to predict the traffic state of the urban freeway. The results showed that the S-ARIMA model could obtain better properties [18, 19]. Ding et al. proposed a space-time ARIMA to predict the traffic state of the road network [20].

In addition, the Markov chain, Kalman filter (KF), KF-based approaches, and other approaches also have important applications in short-term traffic prediction [21–24]. However, these approaches fail to produce favorable results under unstable traffic conditions, such as unexpected events [25].

Different from mathematical approaches and statistical approaches, data-driven approaches rely on a sufficient mass of traffic data. In recent years, due to abundant data attached to extensive traffic sensors and advanced big data technology, data-driven approaches have developed rapidly. Numerous data-driven approaches were put forward for the short time traffic state prediction, such as SVM [26–28], neural network [29–32], and hybrid methods [33, 34].

Among these data-driven approaches, deep learning methods have become extremely popular and successful because of their powerful ability to process nonlinear high-dimensional problems. Huang et al. employed a deep belief network (DBN) with multitask learning for traffic flow prediction [4]. Lv et al. proposed an SAE model to predict the traffic flow, and the performance is superior to other methods at different prediction horizons [35]. Furthermore, numerous efforts have been devoted to emphasizing the temporal characteristics and spatial dependencies on prediction [2]. Ma et al. predicted the traffic speed of a large-scale transportation network using an LSTM neural network and CNN [36, 37]. Wu and Tan extracted spatial and temporal features by CNN and LSTM, respectively, and predicted traffic volume combined these two approaches [38]. Wang et al. proposed an eRCNN model then trained the recurrent CNN by reducing predictive feedback errors [39].

In summary, among all these traffic state prediction approaches, deep learning methods stand out as the most potent alternative. However, few prediction methods could achieve satisfactory accuracy because of the missing data problem in reality. For the critical roads that have a

significant effect on the traffic states of their respectively adjacent roads at a specific road network, we could extract the traffic state of the critical road sections and predict the future traffic state of the overall network.

## 3. Methodology

This research focuses on utilizing data from critical road sections to predict future traffic conditions of the overall urban transportation network. In this section, the utility function of the QoS for the critical roads in a large-scale road network is proposed based on the coverage and the data score. Then, the critical road selection optimization in the transportation networks is presented by selecting an appropriate set of critical roads with the maximum proportion of the total calculation resources to maximize the utility value of the QoS. Specifically, the critical road sections are selected by an innovative critical road selection method. Finally, the traffic speed of the critical roads is regarded as the input of the convolutional long short-term memory neural network to predict the future traffic states of the entire network.

*3.1. Road Network Topology Analysis.* In the practical urban road network, road sections are connected at intersections and eventually form a complex network. When performing sensing tasks in urban road network areas, the topological structure characteristics of the road network need to be considered. The results of numerous studies have shown that the topology of urban road networks is complex and diverse. The failure of connectivity in some sections of the network can impact the traffic operation in the whole network. However, it is inevitable that road connectivity in the network will sometimes be interrupted due to bad weather, technical failures, or serious accidents. This will lead to local traffic congestion and paralysis in the network, which will have harmful economic and social impacts. It means that the higher importance of road sections in the network under different spatial and temporal conditions indicates the more severe damage to the integrity of the road network connectivity brought by their failure. In order to avoid serious impacts on road network connectivity and vehicle movements, priority should be given to the use of high, critically important road sections for future traffic state prediction.

In order to study the structural properties of real road networks, it is necessary to reasonably abstract the road networks into topological structure diagrams, consisting of points and lines by suitable methods. The primal approach and the dual approach are two methods to abstract the road network into a complex network model. The primal approach uses the roads in the road network as edges in the abstract network and the intersections as nodes in the abstract network. The dual approach models the intersections as edges in the network and the roads in the urban road network as nodes, thus showing the interconnections between roads in a highly abstract way. The primal approach is intuitive and simple and can retain the layout characteristics of the road network. The dual approach ignores some

geographic significance of network entities, such as the geographic location of roads, length, and width, and therefore, it is more suitable for analyzing abstract network structures. In this paper, we pay more attention to the characteristics of road connectivity, so we use the primal approach to model the road network.

*3.2. QoS for the Critical Roads.* The QoS is the standard for measuring the performance of the critical roads in a prediction task. The greater the value of QoS, the better the prediction performance of the critical roads. Here, we propose a utility function to calculate the QoS for the critical roads.

The coverage of the selected roads is mainly concerned with people when predicting the traffic state of the whole road network. So, we choose coverage as our first metric. Coverage indicates the coverage of the selected vehicle to the entire road network. For better calculating the coverage $C$, we divide the road network into small grids, and the grid is used as the basic measuring element for coverage calculation. As shown in Figure 1, the pink curves represent roads. The grid with roads passing through is painted green. In this paper, the size of the grid is $0.0005° \times 0.0005°$, based on longitude and latitude, where 0.0005 longitude (latitude) is about 50 m.

Define $g_j$ as the single grid in the urban road network. Let $f(g_j)$ represent the coverage state of $g_j$, and $f(g_j) = 1$ indicates that grid $j$ is selected during the predicting task time; otherwise, $f(g_j) = 0$. In this paper, we tried to predict the whole network traffic state. If $g_j$ belongs to road section $l$ and road section $l$ has $\varphi$ grids, then the road factors could be calculated as $\partial(g_j) = 1/\varphi$. The following equations could calculate the coverage of the selected grids:

$$f(g_j) = \begin{cases} 1, & \text{if } g_j \text{ is selected in predicting task,} \\ 0, & \text{otherwise,} \end{cases}$$

$$C = \frac{\sum_{i=1}^{n} f(g_j) * \partial(g_j)}{n}. \tag{1}$$

On the other hand, we introduce the concept of data score to represent the criticality level of the roads. Specially, we give a higher score to the data collected from the critical roads than the data collected from the ordinary roads. Here, we use the correlations among road sections on both space and time to calculate the data score.

The spatial weights matrix represents the spatial dependency among road sections in traffic networks. According to graph theory, the local connectivity of a node can be calculated based on the connection between the node and its adjacent node, which is represented by the degree of node $k$. Suppose there exists a network, which can be represented as $G = (V', E)$, where $V'$ is the set of nodes in the network and $E$ is the set of edges in the network, representing the connections between nodes in the network. The network $G = (V', E)$ has $|V'| = N$ nodes and $|E| = M$ edges. Then, this network can be represented by the adjacency matrix $A_{M \times N}$. If nodes $i$ and $j$ are connected directly or

(a)                                                                                         (b)
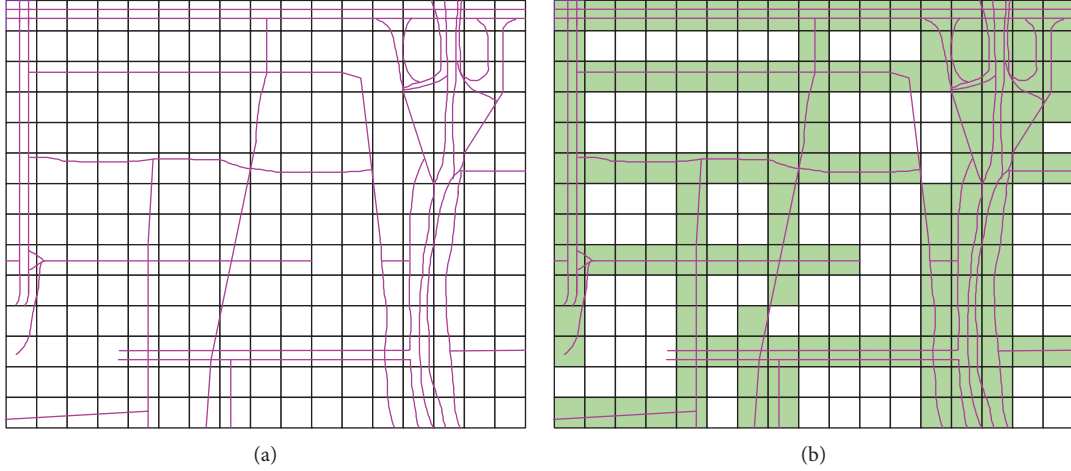
FIGURE 1: An area of the urban road network in Beijing, China. (a) The primal graph of the road network and (b) the gridding of the road network.

indirectly, we call these two nodes "$k$th-order neighbors," and their adjacency relationship could be expressed as equation (2). The spatial weights matrix $P$ could be defined as the sum of the $k$th spatial weights matrix; element $p_{ij}$ in $P$ is calculated as shown in equation (3).

$$A_{ij}^k = \begin{cases} 1, & \text{when nodes } i \text{ and } j \text{ are directly connected,} \\ 0, & \text{otherwise,} \end{cases}$$

(2)

$$p_{ij} = \sum_{i=1}^{K} \sum_{j \neq i}^{n} A_{ij}^k,$$

(3)

where $K$ is the highest order of the spatial weights matrix.

And, to comprehensively consider the spatial correlation of the road network with the temporal correlation, the spatial weights matrix is introduced as a spatial indicator to improve the initial correlation distance. Let $x_i^t$ be traffic speed in the road section $i$ at time $t$, where $x_i^t = (x_i^1, x_i^2, \ldots, x_i^T)$. Then, the integrated speed values of the adjacent road sections $RX_i^t$ could be computed by equation (4). Then, the data score $D_i(s)$ between road section $i$ and its adjacent road sections could be calculated by equation (5):

$$RX_i^t = \sum_{j=1}^{R} P_{ij} x_j^t,$$

(4)

$$D_i(s) = 1 - \frac{\sum_{t=1}^{T} \left(x_i^t - \overline{x_i}\right)\left(RX_i^{t+s} - \overline{RX_i}\right)}{\sqrt{\sum_{t=1}^{T} \left(x_i^t - \overline{x_i}\right)^2} \sqrt{\sum_{t=1}^{T} \left(RX_i^{t+s} - \overline{RX_i}\right)^2}},$$

(5)

where $s = (1, 2, \ldots, S)$ $(s > 0)$ is the time lag between the speeds of road sections $i$ and its adjacent road sections; $\overline{x_i}$ is the mean value of traffic speed $x_i^t$ in road sections $i$ during the time duration of $T$. $\overline{RX_i}$ is the mean value of integrated speed $RX_i^t$. Based on the previous analysis, mathematical expressions of data score $D$ are defined as follows:

$$D = \sum_{i=1}^{n} D_i(s).$$

(6)

The utility function of QoS, $U$, is defined as

$$U = \alpha \cdot \log_2 C + (1 - \alpha) \cdot \ln D,$$

(7)

where $\alpha \in [0, 1]$ is a parameter for tuning the weight of data coverage and data score. To make $\alpha$ more sensitive in equation (7), we apply $\log_2 C$ instead of $C$ and $\ln D$ instead of $D$. Since logarithm is a common way to aggregate data, this process can reduce the heteroscedasticity of data in the QoS function.

### 3.3. The Critical Road Selection Optimization

*3.3.1. Definition of the Critical Road Selection Problem.* The goal of the critical road selection model is to maximize the QoS of the urban traffic prediction system with limited critical roads, so the mathematical expression of the critical roads selection problem can be defined as

$$\max U = \alpha \cdot \log_2 C + (1 - \alpha) \cdot \ln \sum_{j=1}^{n} D_i(s)$$

$$\text{s.t.} \begin{cases} C = \dfrac{\sum_{i=1}^{n} f(g_j) * \partial(g_j)}{n}, \\[2ex] f(g_j) = \begin{cases} 1, & \text{if } g_j \text{ is selected in predicting task}, \\[1ex] 0, & \text{otherwise}, \end{cases} \\[2ex] F = \{f(g_j) | i = 1, 2, \ldots, n\}, \\[1ex] 0 \le \alpha \le 1, \\[1ex] D_i(s) = 1 - \dfrac{\sum_{t=1}^{T}(x_i^t - \overline{x_i})(\text{RX}_i^{t+s} - \overline{\text{RX}_i})}{\sqrt{\sum_{t=1}^{T}(x_i^t - \overline{x_i})^2} \sqrt{\sum_{t=1}^{T}(\text{RX}_i^{t+s} - \overline{\text{RX}_i})}}, \\[2ex] x_i^t = (x_i^1, x_i^2, \ldots, x_i^T), \\[1ex] \partial(g_j) = \dfrac{1}{\varphi}, \\[1ex] C < \mu, \end{cases} \tag{8}$$

where function $U$ is used to calculate the utility value of the QoS; $f(g_j)$ represents whether grid $j$ is selected, with $f(g_j) = 1$ indicating that grid $j$ has been selected; otherwise, the value of $f(g_j)$ is 0. $F$ is the set of critical roads, and $C$ is the coverage. $D_i(s)$ is the data score between road section $i$ and its adjacent road sections. $\mu$ is the threshold for the selection.

*3.3.2. The Critical Road Selection Method.* Based on the above analysis, a critical road selection model for traffic state prediction is constructed. In this paper, due to the spatiotemporal state changes of the road network, the selected roads need to be updated at intervals $T$. An improved greedy algorithm is proposed to solve the critical road selection problems. A greedy algorithm means that we always make the best choice for the current situation when solving an optimization problem. In other words, such an algorithm does not consider the overall situation but rather considers only the local situation. The pseudocode of the proposed algorithm is shown in Algorithm 1.

*3.4. Traffic State Prediction Using the Deep Learning Approach.* In this paper, a spatiotemporal recurrent convolutional network is proposed for the prediction (STRCN). The proposed STRCN inherits the advantages of deep convolutional neural networks (DCNN) and long short-term memory (LSTM) neural networks. The spatial dependencies of network-wide traffic can be captured by CNN, and the temporal dynamics can be learned by LSTM.

*3.4.1. Capturing Spatial Features by CNN.* CNN has been successfully applied to traffic prediction for its great potential in extracting features using multiple layers. A typical CNN mainly comprises multiple convolution layers and pooling layers. The former contributes to mine spatial dependencies of road sections since every layer retrieves a distinct feature using different filters. In comparison, the latter assists in reducing the number of parameters required for training CNN under the premise of ensuring prediction accuracy. Given that the input for CNN could be intuitively regarded as an image with each pixel value associating one kind of traffic state during a certain time, 2D CNN is naturally utilized to abstract spatial features between road sections. Figure 2 illustrates the structure of CNN, including the input layer, convolution layer, pooling layer, fully connected layer, and output layer. Each part plays a unique and vital role for CNN, and the details are briefly explained below.

Suppose that we need to predict the future traffic speed of the network $V^{t+a} = \{v_i^{t+a}\}_{i=1}^{m}$, where $a$ is the prediction horizon and $m$ is the number of road sections. The input of the CNN is the historical traffic speed of critical road sections $\{U^{t-n}, U^{t-1}, U^t\}$, where $U^{t-n} = \{v_i^{t-n}\}_{i=1}^{P_\alpha}$ represents the traffic state of critical road sections at time $(t-n)$, $n$ is the look back step, and $P_\alpha$ is the number of critical road sections. Then, the spatial features of input are captured by convolutional and pooling layers. Let $O_r^l$ be the output of $l$th convolutional and pooling layers with $r$ filters and the weights and bias of $l$th layers be $(W_r^l, b_r^l)$. Then, the $O_r^l$ could be calculated by the following equation:

```
Input: number of roads: n
        Number of prediction tasks: m
        Set of critical roads: F
        QoS for the critical roads: U
        Output: set of finally selected roads: R
(1) U ← ∅, U(F) ← 0
(2) for j from 1 to m do
(3)    i = F_i ⟶ head
(4)    for i from 1 to n do
(5)        U(F_i) = α · log₂ C + (1 − α) · ln ∑ⁿⱼ₌₁ D_i(s)
(6)        ΔU = U(F_{i+1}) − U(F_i)
(7)            if ΔU > 0 then
(8)                R←R∩F_i
(9)            else
(10)               i = F_i ⟶ next
(11)           end if
(12)   end for
(13) end for
(14) return R;
```

ALGORITHM 1: Greedy-based critical road selection method (pseudocode of the proposed algorithm).
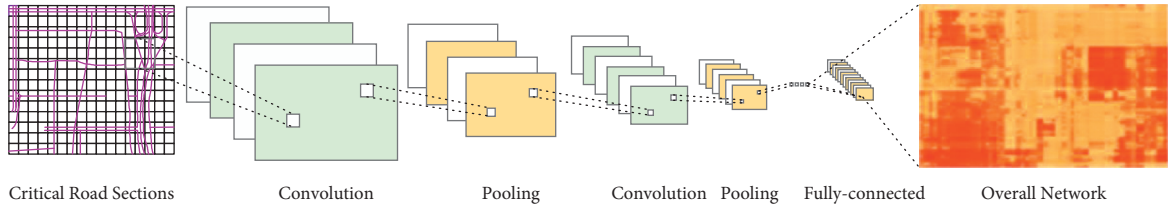


FIGURE 2: The structure of the CNN.

$$O_r^l = \text{pool}\left( f\left( \sum_r W_r^l * O_r^{l-1} + b_r^l \right) \right), \tag{9}$$

where $O_r^{l-1}$ means the output of the previous layer and $O_r^1$ is exactly the input layer. $f$ is a nonlinear activation function, and pool denotes the pooling procedure.

*3.4.2. Capturing Temporal Features by LSTM.* Intuitively, traffic states at each moment have a strict sequential relationship in time dimension rather than isolated from each other, which is especially suitable for RNN to capture the temporal evolution of traffic flow. However, it is difficult for traditional RNN to capture temporal dependency if two time intervals are remote. Then, LSTM, one of the specific forms of RNNs, is proposed to tackle these issues by adding memory cells in hidden layers. As shown in Figure 3, four main parts, an input gate, a neuron with a self-recurrent connection, a forget gate, and an output gate, are collaborated to alleviate the problems of traditional RNN caused by the gradient vanish and explosion.

In our model, next to CNN, the LSTM naturally takes the output of CNN $V^t = \{x_i^t\}_{i=1}^R$ as its input to predict the future traffic states $H_t = \{h_i^t\}_{i=1}^q$, namely, its output, where $q$ is the number of hidden units of the output layer. For a memory
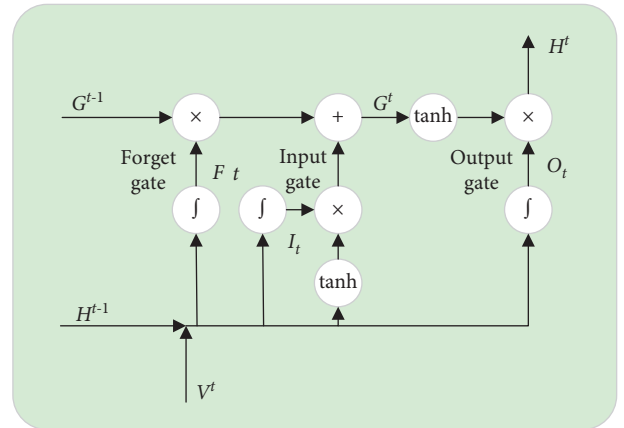


FIGURE 3: The architecture of LSTM NN.

cell, the input states are $G_{t-1}$ while the output is $G_t$. Meanwhile, the states of input, forget, and output gates are $I_t, F_t,$ and $O_t$, respectively. The temporal features could be iteratively calculated by the following equations:

$$\text{input gate: } I_t = \sigma\left(W_v^i V^t + W_h^i H^{t-1} + b_i\right), \tag{10}$$

$$\text{forget gate: } F_t = \sigma\left(W_v^f V^t + W_h^f H^{t-1} + b_f\right), \tag{11}$$

output gate: $O_t = \sigma\left(W_v^O V^t + W_h^o H^{t-1} + b_o\right),$  (12)

cell input: $G_{t-1} = \tanh\left(W_v^c V^t + W_h^c H^{t-1} + b_c\right),$  (13)

cell output: $G_t = I_t \odot G_t + F_t \odot G_{t-1},$  (14)

hidden layer output: $H_t = O_t \odot \tanh(G_t),$  (15)

where weights matrices $W$ and bias vectors $b$ are constructed to connect input layer, output layer, and the memory cell, $\odot$ denotes the scalar product of two vectors, and $\sigma(\cdot)$ represents the standard logistics sigmoid function defined as follows:

$$\sigma(x) = \frac{1}{1 + e^{-x}}.$$  (16)

*3.4.3. Training with STRCN.* Integrated with the advantages of CNN and LSTM, the STRCNs is utilized to predict future traffic states by sufficiently exploiting the spatiotemporal characteristics of the data. Eventually, a fully connected layer is employed to predict the future speed by taking the output of LSTM as input. The future speed could be calculated by the following equation:

$$Y^{t+1} = W_y H_t + b_y,$$  (17)

where $W_y$ and $b_y$ are weight and bias related to the hidden layer. Conclusively, the model is trained from end to end, and the values of $Y^{t+1}$ are prediction results, which are the output of the entire mode. Several hyperparameters within the model will be set and elaborated in the experiment section. Additionally, it is significant to note that the input size will alter as the number of critical road sections changes due to different extracting rate $\alpha$, and hence several hyperparameters will change too.

## 4. Case Study

*4.1. Data Used.* In this section, a case study is conducted to evaluate the performance of the proposed critical road selection optimization model and the traffic prediction method. The main urban road of a subtransportation network of Beijing near West Second Ring Road is selected as the research objective, as shown in Figure 4. The network comprises 278 road sections, including several kinds of hierarchies of roads, such as freeways, arterials, secondary roads, and collectors. The total length of all the roads is approximately 24.53 km, and the network covers around 0.6 km$^2$ areas.

Data collected by taxis equipped with GPS devices from June 1st, 2015, to August 31st, 2015 (92 days) is utilized for training the proposed model and predicting the future traffic speed of the network. The updating frequency of data is 2 minutes, and a time period ranging from 6 : 00 : 00 to 23 : 00 : 00 is concerned for high travel demand is repeatedly observed. Accounting for the traffic state varies every time interval, we could observe 511 traffic states per day.
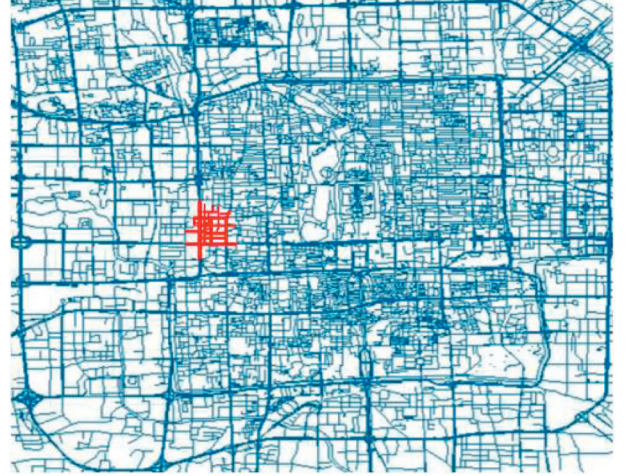


FIGURE 4: The layout of the subtransportation network in Beijing.

*4.2. Critical Road Selections.* The proposed QoS-based critical road selection method and other two road selection methods (random method and coverage-based method) were used to select the road sections for prediction in the road network. In a random method, the road is selected randomly during the prediction process. In the coverage-based method, the road is selected based on the coverage, which means that roads with higher coverage are preferred for selection. The three critical road selection methods are applied under different roads extracting rate $\mu$ (i.e., the proportion of critical road sections to all roads, $\mu \in [10\%, 20\%, 30\%, 40\%, 50\%, 60\%, 70\%, 80\%, 90\%, 100\%]$).

Eventually, the correspondence of extracting rate and the number of critical road sections are listed in Table 1. For example, $\mu = 0.50$ means that we will select 139 roads as critical road sections and subsequently use them to predict the traffic states of 278 roads.

*4.3. Results and Comparison*

*4.3.1. Performance between Different Critical Road Selection Methods.* The root means squared error (RMSE) and root mean squared error proportional (RMSEP) are employed to evaluate the performance of all the models, which could be calculated as in the following equations:

$$\text{RMSE} = \frac{1}{n_\mu} \sqrt{n_\mu \sum_{i=1}^{N_\mu} (\hat{y}_i - y_i)^2},$$  (18)

$$\text{RMSEP} = \frac{100}{\hat{y}_i^{\text{ave}}} \sqrt{\frac{1}{n_\mu} \sum_{i=1}^{N_\mu} (\hat{y}_i - y_i)^2},$$  (19)

where $\hat{y}_i$ is the $i$th ground-truth value and $y_i$ is the $i$th predicted value. The value of $n_\mu$ denotes the number of critical road sections at extracting rate $\mu$ and $N_\mu$ is the total number of traffic states.

TABLE 1: Correspondence of extracting rate and the number of critical road sections.

| Parameters | Values | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| Extracting rate $\mu$ | 0.60 | 0.65 | 0.70 | 0.75 | 0.80 | 0.85 | 0.90 | 0.95 | 1.00 |
| Number of roads | 167 | 181 | 195 | 209 | 222 | 236 | 250 | 264 | 278 |

TABLE 2: Performance of the STRCN using different critical road selection methods.

| Extracting rate $\mu$ | QoS-based method | | Random method | | Coverage-based method | |
|---|---|---|---|---|---|---|
| | RMSE | RMSEP (%) | RMSE | RMSEP (%) | RMSE | RMSEP (%) |
| 1.0 | 6.767 | 21.019 | 6.766 | 21.013 | 6.768 | 21.022 |
| 0.9 | 6.911 | 21.572 | 6.901 | 22.021 | 6.875 | 21.531 |
| 0.8 | 7.143 | 22.698 | 7.175 | 22.781 | 7.144 | 22.700 |
| 0.7 | 7.144 | 22.693 | 7.231 | 23.276 | 7.174 | 23.091 |
| 0.6 | 7.027 | 21.425 | 7.285 | 23.763 | 7.186 | 23.113 |
| 0.5 | 7.089 | 21.889 | 7.301 | 24.017 | 7.197 | 23.150 |
| 0.4 | 7.179 | 22.802 | 7.376 | 24.172 | 7.223 | 23.350 |
| 0.3 | 7.235 | 23.763 | 7.405 | 24.508 | 7.331 | 24.003 |
| 0.2 | 7.471 | 24.078 | 7.479 | 24.967 | 7.474 | 24.095 |
| 0.1 | 7.701 | 25.102 | 7.705 | 25.109 | 7.703 | 25.108 |
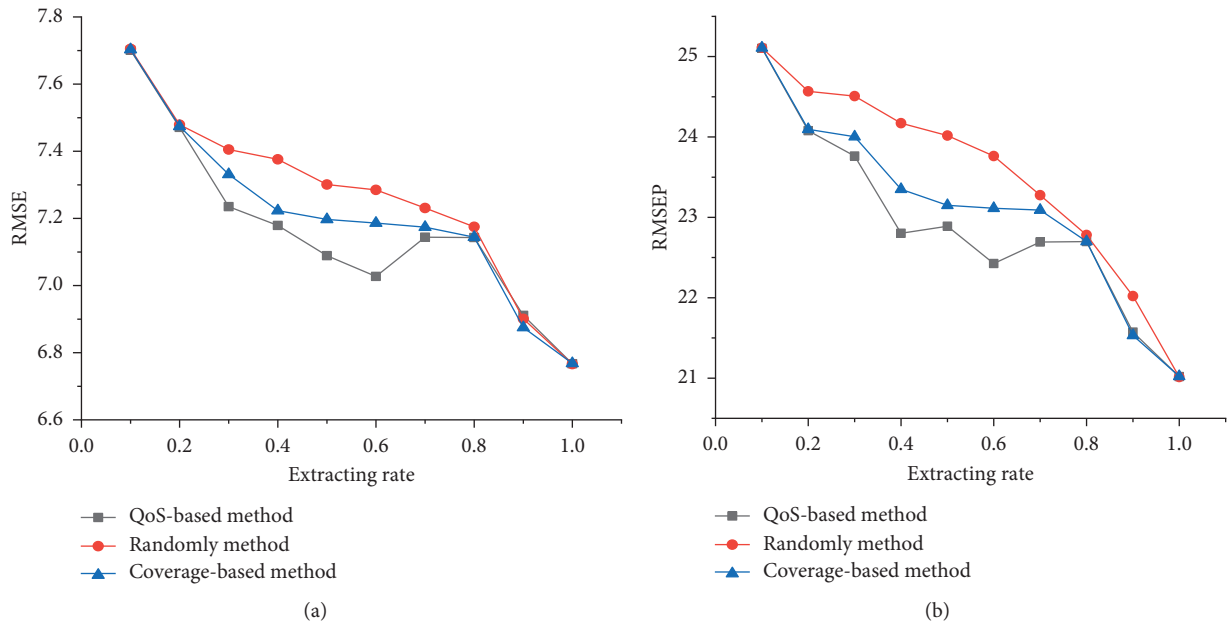


FIGURE 5: The trend of prediction accuracy using different critical road selection methods. (a) The RMSE of different methods and (b) the RMSEP of different methods.

### 4.3.2. Performance Using the Different Critical Road Selection Method.

In order to evaluate the performance of the proposed critical road selection optimization model, we train and test the STRCN model using different road selection methods. The RMSEs and RMSEPs in the context of different extracting rates $\mu$ are listed in Table 2.

It could be found that in the range of 0.8 to 1.0, the performance of the STRCN model using different road selection methods is almost the same. The reason is that the finally selected road sections by different methods have little

difference under the condition of high extracting rate. However, when the extracting rate is between 0.5 and 0.8, the performance of the QoS-based selection method is a little superior to the overall prediction model. In general, the decrease of accuracy is reasonable and within the acceptable limits, which demonstrates the validation and generalization of the approach and the fact that some road sections indeed have less contribution towards prediction.

Additionally, when the extracting rate comes to 0.5, the predictive performance gradually tends to be unstable,

TABLE 3: Performance between several DL algorithms.

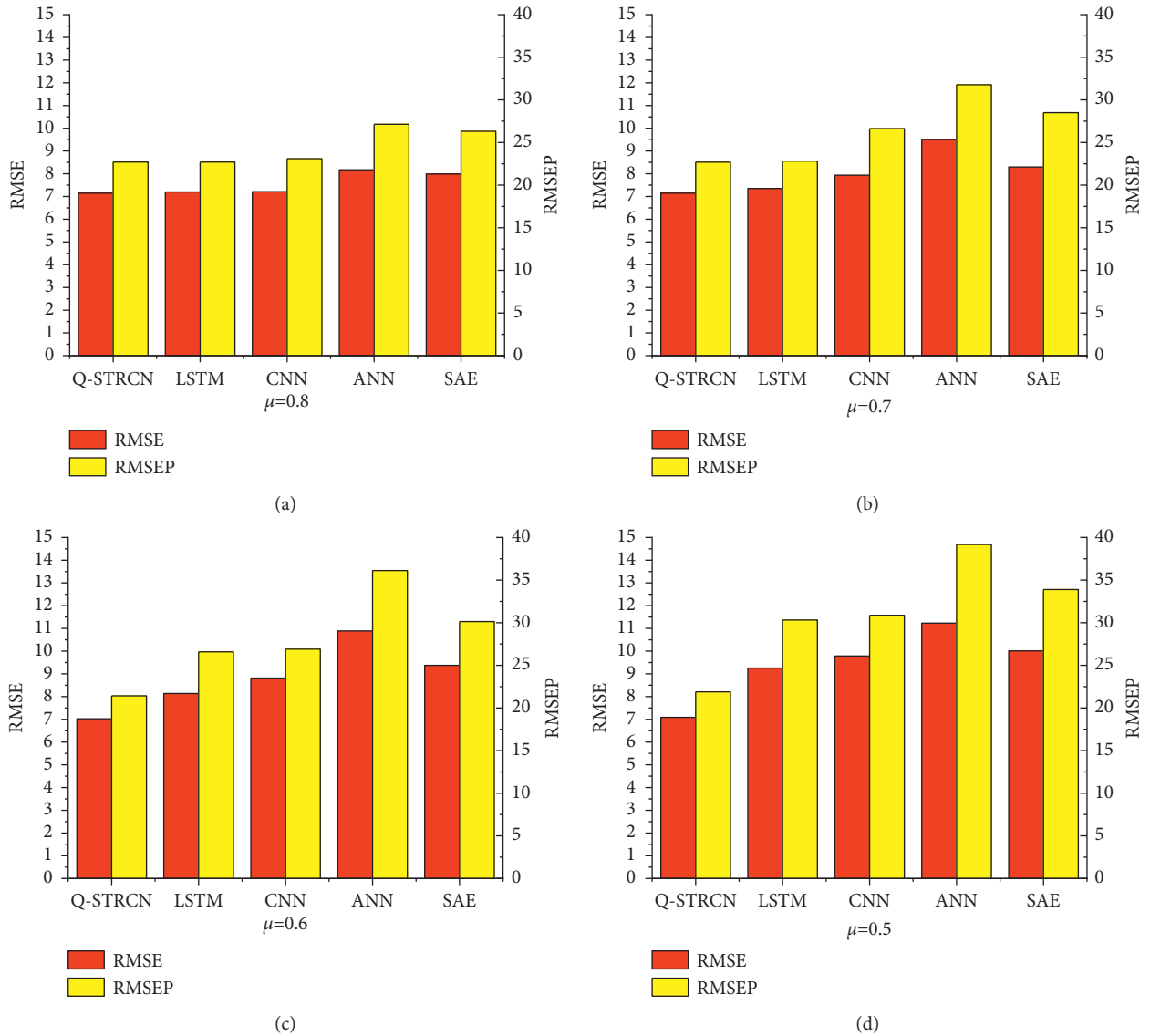| $\mu$ | Q-STRCN | | LSTM | | CNN | | ANN | | SAE | |
|---|---|---|---|---|---|---|---|---|---|---|
| | RMSE | RMSEP | RMSE | RMSEP | RMSE | RMSEP | RMSE | RMSEP | RMSE | RMSEP |
| 0.8 | 7.143 | 22.698 | 7.198 | 22.701 | 7.213 | 23.101 | 8.173 | 27.134 | 7.989 | 26.311 |
| 0.7 | 7.144 | 22.693 | 7.351 | 22.816 | 7.939 | 26.631 | 9.512 | 31.781 | 8.296 | 28.501 |
| 0.6 | 7.027 | 21.425 | 8.136 | 26.589 | 8.813 | 26.897 | 10.891 | 36.123 | 9.375 | 30.134 |
| 0.5 | 7.089 | 21.889 | 9.257 | 30.317 | 9.785 | 30.871 | 11.231 | 39.177 | 10.012 | 33.891 |



FIGURE 6: The quantitative results among different approaches under different extracting rates. (a) $\mu = 0.8$; (b) $\mu = 0.7$; (c) $\mu = 0.6$; (d) $\mu = 0.5$.

probably because too many roads are omitted. Particularly when the extracting rate is between 0.0 and 0.2, the performance is basically the same. That is because no matter what method is used, there are too many missing road sections to predict. Figure 5 shows that prediction accuracy generally declines as the extracting rate decreases with different critical road selection methods.

*4.3.3. Performance between Several DL Algorithms.* As we introduced before, corrupted or missing data generally exist on account of the monitoring equipment failure, extreme weather, data transmission error etc., which weakens the effectiveness of the prediction model or even disables the model. To test the performance of our model under random structural missing data, we stochastically extract a part of

road sections of the rate of $\mu$ where the value remains the same as those mentioned above.

Four popular deep learning-based algorithms are selected for comparison, including ANN, CNN, LSTM, and SAE. ANN adopts a plain and shallow structure to process multidimension and nonlinear problems. The parameters of the SAE are set according to [35], which achieves high accuracy in predicting traffic flow. We take 30 min historical traffic speed as the input to predict overall traffic states after 2 min.

Table 3 presents the quantitative results of Q-STRCN, ANN, CNN, and LSTM. It could be observed that the result of Q-STRCN outperforms other models, indicating that our model could precisely mine the spatiotemporal features of the data and make a relatively accurate prediction. Among all the rival algorithms, LSTM has the best performance, probably resulting from that the temporal features of the time-series data are essentially prominent. The results of ANN and SAE demonstrate that these two models fail to extract spatiotemporal characteristics that have vital impacts on prediction. Figure 6 shows the quantitative results among different approaches under different extracting rates.

## 5. Summary and Conclusions

Structural missing data usually has a massive negative effect on short-term traffic state prediction. In this paper, a novel hybrid short-term traffic state prediction method based on critical road selection optimization is proposed. First, the utility function of the quality of service (QoS) for the critical roads in a large-scale road network is proposed based on the coverage and the data score. Then, the critical road selection optimization model in the transportation networks is presented by selecting an appropriate set of critical roads with the maximum proportion of the total calculation resources to maximize the utility value of the QoS. Also, an innovative critical road selection method, which is considering the topological structure and the mobility of the urban road network, is introduced. Subsequently, the traffic speed of the critical roads is regarded as the input of the convolutional long short-term memory neural network to predict the future traffic states of the entire network. Experiment results on the Beijing traffic network indicate that the proposed method outperforms prevailing DL approaches in the cases of considering critical road sections.

However, even the case study showed that the proposed method could significantly improve the QoS of the traffic prediction, there is still a long way to go from practical application for the method. For future studies, the inherent attributes of the road should be taken into account when calculating the QoS for the road network. Besides, the traffic accident, temperature, weather, and other external factors affect the traffic prediction accuracy. All these will be left for our future research.

## Data Availability

All data, models, and code generated or used during the study are available in the submitted article.

## Conflicts of Interest

The authors declare that they have no conflicts of interest regarding the publication of this study.

## Acknowledgments

## References

[1] C. Antoniou, H. N. Koutsopoulos, and G. Yannis, "Dynamic data-driven local traffic state estimation and prediction," *Transportation Research Part C: Emerging Technologies*, vol. 34, no. 9, pp. 89–107, 2013.

[2] G. Yang, Y. Wang, H. Yu, Y. Ren, and J. Xie, "Short-term traffic state prediction based on the spatiotemporal features of critical road sections," *Sensors*, vol. 18, p. 2287, 2018.

[3] M. Elhenawy and H. Rakha, "Spatiotemporal traffic state prediction based on discriminatively pre-trained deep neural networks," *Advances in Science Technology and Engineering Systems Journal*, vol. 2, no. 3, pp. 678–686, 2017.

[4] W. Huang, G. Song, H. Hong, and K. Xie, "Deep architecture for traffic flow prediction: deep belief networks with multitask learning," *IEEE Transactions on Intelligent Transportation Systems*, vol. 15, pp. 2191–2201, 2014.

[5] E. I. Vlahogianni, M. G. Karlaftis, and J. C. Golias, "Short-term traffic forecasting: where we are and where we're going," *Transportation Research Part C: Emerging Technologies*, vol. 43, pp. 3–19, 2014.

[6] W. Min and L. Wynter, "Real-time road traffic prediction with spatio-temporal correlations," *Transportation Research Part C: Emerging Technologies*, vol. 19, no. 4, pp. 606–616, 2011.

[7] P. I. Richards, "Shock waves on highway," *Operations Research*, vol. 4, pp. 42–51, 1956.

[8] S. R. Chandra and H. Al-Deek, "Cross-correlation analysis and multivariate prediction of spatial time series of freeway traffic speeds," *Transportation Research Record Journal of the Transportation Research Board*, vol. 2061, pp. 64–76, 2008.

[9] H. Van Lint and C. Van Hinsbergen, "Short-term traffic and travel time prediction models," *Transportation Research E-Circular*, vol. 173, 2012.

[10] Y. Haiyang, W. Zhihai, W. Shuqin, Y. Wang, and X. Ma, "Spatiotemporal recurrent convolutional networks for traffic prediction in transportation networks," *Sensors*, vol. 17, no. 17, p. 1501, 2017.

[11] M. J. Lighthill and G. B. Whitham, "On kinematic waves I flood movement in long rivers. II a theory of traffic flow on long crowded roads," *Proceedings of the Royal Society of London. Series A. Mathematical and Physical Sciences*, vol. 229, pp. 281–345, 1955.

[12] H. Rakha and B. Crowther, "Comparison of greenshields, pipes, and van aerde car-following and traffic stream models," *Transportation Research Record Journal of the Transportation Research Board*, vol. 1802, no. 1, 2002.

[13] M. Ben-Akiva, M. Bierlaire, H. Koutsopoulos, and R. Mishalani, "DynaMIT: a simulation-based system for traffic prediction," in *Proceedings of the DACCORD Short Term Forecasting Workshop*, pp. 1–12, Delft, Netherlands, February 1998.

[14] H. S. Mahmassani, "Dynamic network traffic assignment and simulation methodology for advanced system management applications," *Networks and Spatial Economics*, vol. 1, no. 3, pp. 267–292, 2001.

[15] L. Smith, R. Beckman, and K. Baggerly, "TRANSIMS: transportation analysis and simulation system," Technical Report, Office of Scientific and Technical Information, Washington, DC, USA, 1995.

[16] J. Barros, M. Araujo, and R. J. F. Rossetti, "Short-term real-time traffic prediction methods: a survey," in *Proceedings of the 2015 International Conference on Models and Technologies for Intelligent Transportation Systems (MT-ITS)*, pp. 132–139, Budapest, Hungary, June 2015.

[17] M. M. Hamed, H. R. Al-Masaeid, and Z. M. B. Said, "Short-term prediction of traffic volume in urban arterials," *Journal of Transportation Engineering*, vol. 121, no. 3, pp. 249–254, 1995.

[18] B. M. Williams, P. K. Durvasula, and D. E. Brown, "Urban freeway traffic flow prediction: application of seasonal autoregressive integrated moving average and exponential smoothing models," *Transportation Research Record Journal of the Transportation Research Board*, vol. 1644, no. 1, pp. 132–141, 1998.

[19] B. M. Williams and L. A. Hoel, "Modeling and forecasting vehicular traffic flow as a seasonal ARIMA process: theoretical basis and empirical results," *Journal of Transportation Engineering*, vol. 129, no. 6, pp. 664–672, 2003.

[20] Q. Y. Ding, X. F. Wang, X. Y. Zhang, and Z. Q. Sun, "Forecasting traffic volume with space-time ARIMA model," *Advanced Materials Research*, vol. 156, pp. 979–983, 2010.

[21] Y. Qi and S. Ishak, "A hidden markov model for short term prediction of traffic conditions on freeways," *Transportation Research Part C: Emerging Technologies*, vol. 43, pp. 95–111, 2014.

[22] J. Guo, W. Huang, and B. M. Williams, "Adaptive kalman filter approach for stochastic short-term traffic flow rate prediction and uncertainty quantification," *Transportation Research Part C: Emerging Technologies*, vol. 43, pp. 50–64, 2014.

[23] Y. Wang and M. Papageorgiou, "Real-time freeway traffic state estimation based on extended kalman filter: a general approach," *Transportation Research Part B: Methodological*, vol. 39, no. 2, pp. 141–167, 2005.

[24] I. Okutani and Y. J. Stephanedes, "Dynamic prediction of traffic volume through kalman filtering theory," *Transportation Research Part B: Methodological*, vol. 18, no. 1, pp. 1–11, 1984.

[25] H. Yu, N. Ji, Y. Ren, and C. Yang, "A special event-based *K*-nearest neighbor model for short-term traffic state prediction," *IEEE Access*, vol. 7, no. 99, pp. 81717–81729, 2019.

[26] X. Liu, X. Fang, Z. Qin, C. Ye, and M. Xie, "A short-term forecasting algorithm for network traffic based on chaos theory and svm," *Journal of Network & Systems Management*, vol. 19, pp. 427–447, 2011.

[27] Y. Sun, B. Leng, and W. Guan, "A novel wavelet-svm short-time passenger flow prediction in Beijing subway system," *Neurocomputing*, vol. 166, pp. 109–121, 2015.

[28] C. H. Wu, J. M. Ho, and D. T. Lee, "Travel-time prediction with support vector regression," *IEEE Transactions on Intelligent Transportation Systems*, vol. 5, pp. 276–281, 2004.

[29] M. G. Karlaftis and E. I. Vlahogianni, "Statistical methods versus neural networks in transportation research: differences, similarities and some insights," *Transportation Research Part C*, vol. 19, pp. 387–399, 2011.

[30] B. L. Smith, *Forecasting Freeway Traffic Flow for Intelligent Transportation Systems Application*, University of Virginia, Charlottesville, Virginia, 1995.

[31] B. L. Smith and M. J. Demetsky, "Short-term traffic flow prediction: neural network approach," *Transportation Research Record*, vol. 1453, pp. 98–104, 1994.

[32] J. Wang, W. Deng, and Y. Guo, "New bayesian combination method for short-term traffic flow forecasting," *Transportation Research Part C: Emerging Technologies*, vol. 43, pp. 79–94, 2014.

[33] X. Jiang, L. Zhang, and X. Chen, "Short-term forecasting of high-speed rail demand: a hybrid approach combining ensemble empirical mode decomposition and gray support vector machine with real-world applications in China," *Transportation Research Part C: Emerging Technologies*, vol. 44, pp. 110–127, 2014.

[34] S. Sun, C. Zhang, and G. Yu, "A bayesian network approach to traffic flow forecasting," *IEEE Transactions on Intelligent Transportation Systems*, vol. 7, pp. 124–132, 2006.

[35] Y. Lv, Y. Duan, W. Kang, Z. Li, and F. Y. Wang, "Traffic flow prediction with big data: a deep learning approach," *IEEE Transactions on Intelligent Transportation Systems*, vol. 16, pp. 865–873, 2015.

[36] X. Ma, Z. Dai, Z. He, J. Ma, Y. Wang, and Y. Wang, "Learning traffic as images: a deep convolutional neural network for large-scale transportation network speed prediction," *Sensors*, vol. 17, no. 4, p. 818, 2017.

[37] 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," *Transportation Research Part C: Emerging Technologies*, vol. 54, pp. 187–197, 2015.

[38] Y. Wu and H. Tan, "Short-term traffic flow forecasting with spatial-temporal correlation in a hybrid deep learning framework," 2016.

[39] J. Wang, Q. Gu, J. Wu, G. Liu, and Z. Xiong, "In Traffic speed prediction and congestion source exploration: a deep learning method," in *Proceedings of the IEEE International Conference on Data Mining*, pp. 499–508, New Orleans, LA, USA, November 2017.