WILEY | Hindawi

Research Article

A Hybrid Model for Short-Term Traffic Flow Prediction Based on Variational Mode Decomposition, Wavelet Threshold Denoising, and Long Short-Term Memory Neural Network

Yang Yu ⬤, Qiang Shang ⬤, and Tian Xie ⬤

School of Transportation and Vehicle Engineering, Shandong University of Technology, Zibo, Shandong 255049, China

Correspondence should be addressed to Qiang Shang; shangqiangv587@163.com

Traffic flow prediction plays an important role in intelligent transportation system (ITS). However, due to the randomness and complex periodicity of traffic flow data, traditional prediction models often fail to achieve good results. On the other hand, external disturbances or abnormal detectors will cause the collected traffic flow data to contain noise components, resulting in a decrease in prediction accuracy. In order to improve the accuracy of traffic flow prediction, this study proposes a mixed traffic flow prediction model VMD-WD-LSTM using variational mode decomposition (VMD), wavelet threshold denoising (WD), and long short-term memory (LSTM) network. Firstly, we decompose the original traffic flow sequence into K components through VMD and determine the number of components K according to the sample entropy of different K values. Then, each component is denoised by wavelet threshold to obtain the denoised subsequence. Finally, LSTM is used to predict each subsequence, and the predicted values of each subsequence are combined into the final prediction results. In addition, the performance of the proposed model and the latest traffic flow prediction model is compared on the several well-known public datasets. The empirical analysis shows that the proposed model not only has good prediction accuracy but also has superior robustness.

1. Introduction

With the rapid development of cities and the rapid increase of urban population, the number of vehicles on urban roads is also increasing. Therefore, the increased traffic pressure on urban roads has caused more and more serious problems, such as traffic accidents and traffic pollution, and road congestion has become an important factor affecting the quality of daily life of residents. Faced with this situation, the development and application of ITS has been recognized as an effective way to solve or alleviate traffic problems. Therefore, on the basis of obtaining accurate future traffic data through historical data, the intelligent transportation system can perceive future traffic conditions and traffic conditions of each section. Then, the system can formulate effective traffic organization and guidance strategies to reduce the probability of road congestion, so as to achieve the purpose of improving road traffic efficiency [1]. However, due to the complexity of road traffic or the environment, different unexpected situations often occur, resulting in the traffic flow data measured by the detector to be interfered, which will affect the regularity of daily traffic flow and thus affect the traffic flow data. Data fluctuations caused by such interference factors are called noise.

Wavelet denoising is a commonly used denoising method in the field of traffic flow prediction. The Kalman filter model based on wavelet decomposition has been used for short-term traffic flow prediction. The empirical results show that the combination of wavelet decomposition and Kalman filter can reduce the impact of noise on prediction to a certain extent [2]. Peng and Xiang proposed a traffic flow prediction method based on phase space reconstruction and wavelet denoising, in which wavelet denoising was used to preprocess the original traffic flow data [3]. In the fuzzy neural network prediction model proposed by Xiao et al., wavelet decomposition was used to smooth historical traffic flow data, and the results show that wavelet denoising can significantly improve the prediction accuracy [4]. Tang et al.

compared the denoising performance of four wavelet functions, coif (coiflet), db (daubechies), haar, and sym (symlet), on the original traffic flow data. The analysis results show that the db wavelet function has the best denoising performance [5].

In order to further improve the denoising performance, an empirical mode decomposition (EMD) denoising method was proposed that has been widely used so far [6]. EMD decomposes a complex signal into a finite number of intrinsic mode functions (IMFs), and each of the decomposed IMF components contains local characteristic signals of different time scales of the original signal. The high-frequency IMF component contains noise, and the low-frequency IMF component contains the characteristics of the original signal, that is, denoising is achieved by processing high-frequency signal. EMD has the advantages of being simple, intuitive, and efficient, but the disadvantage is that it is prone to modal aliasing. In order to make up for the disadvantages of EMD, integrated empirical mode decomposition (EEMD) was proposed [7]. Because EEMD introduces white noise on the basis of EMD to supplement the missing scale, the phenomenon of modal aliasing can be overcome to a certain extent. In 2020, Chen et al. proposed a traffic flow prediction model called EEMD-ANN using EEMD and artificial neural network (ANN) [8]. In 2021, Chen et al. compared the performance of EMD, EEMD, and wavelet in traffic flow data denoising, and the results showed that EEMD has the best performance [9].

Variational mode decomposition (VMD) is a signal processing method proposed in recent years [10]. Different from the principle of EMD, VMD uses completely non-recursive modal variation to process the signal, and it determines the optimal center frequency and bandwidth of the component by solving the constrained variational problem, so it basically overcomes the end effect and modal aliasing of EMD. At present, VMD has been applied in many fields and achieved good results. Liu et al. proposed a wind speed prediction model using VMD and singular spectrum analysis (SSA) [10]. In this model, the original data were decomposed by VMD, and then SSA was used to extract the low-frequency components of the decomposed data for prediction. In [11], VMD was used to process the original streamflow data, and then LSTM was employed to predict the streamflow [12]. The comparison result illustrated that performance of VMD is better than that of EEMD and discrete wavelet transform (DWT). Shi et al. proposed a hybrid prediction model for network traffic based on VMD and extreme learning machine (ELM) [13], and empirical analysis results showed that VMD denoising can effectively improve prediction accuracy. Due to the good performance of VMD in other prediction fields, we have reason to believe that VMD also has great potential in traffic prediction.

After the original traffic flow data are denoised, the selection of the prediction model is very important. In order to improve the prediction accuracy, a large number of models with different data characteristics and calculation processes have been proposed for traffic flow prediction, including traditional statistical models, such as autoregressive integrated moving average (ARIMA) model [14, 15]

and Kalman filter model (Kalman filter) [16, 17], and machine learning-based models, such as support vector machine (SVM) [18–20] and artificial neural network (ANN) [21, 22]. In recent years, deep learning has attracted much attention in traffic flow prediction because of its superior performance. As a variant of recurrent neural network (RNN), LSTM improves the shortcomings of gradient disappearance and gradient explosion. At present, LSTM is widely used in many prediction fields, including traffic flow prediction. Tian et al. proposed a traffic flow prediction model based on LSTM, and empirical analysis showed that the prediction accuracy of LSTM is higher than that of SVM and feedforward neural network (FFNN) [23].

In [24], convolutional neural network (CNN) was first used to extract daily features of traffic flow, and then LSTM was used to predict traffic flow. Ma et al. pointed out that bidirectional long short-term memory (BiLSTM) is more effective in short-term traffic flow prediction [25]. The empirical results in [26] show that the performance of LSTM for traffic speed prediction is better than other comparative parametric and nonparametric methods. In [27], the attention mechanism was introduced in LSTM to improve the accuracy of the model predicting traffic speed, which can properly assign weights to distinguish the importance of traffic speed time sequences. In view of the excellent performance of LSTM in traffic flow prediction, LSTM is selected as the prediction model and its parameters are optimized.

Many related studies in the field of transportation show that the subsequence obtained by signal decomposition of the original measured signal data is more conducive to showing the irregular periodic variation characteristics of the signal than the original data. In the study of predicting the missing measurement signal data of SHM systems, Li et al. [28] decomposed the original signal data into multiple subsequences by the empirical mode decomposition (EMD) method and then used ARIMA, ANN, LSTM, and SVR models to predict different subsequences. The final prediction results show that the prediction performance of the hybrid model after signal decomposition is better than that of the original data directly, which proves the superiority of signal decomposition in the field of traffic data prediction. In 2021, Huang et al. [29] used EMD to extract the intrinsic mode function (IMF) in order to make full use of the time characteristics of traffic flow. The original traffic flow data were decomposed into three components according to their own characteristics: trend component, residual component, and residual component. These three components were analyzed and predicted, respectively, and the accuracy of the prediction results was higher than that of the single method for direct prediction of the original data. This study shows that the decomposed signals are more likely to show characteristics through the prediction model.

In addition, Li et al. [30] used the ensemble empirical mode decomposition method for travel time prediction in 2018. They first decomposed the original travel time data series into multiple functions with different characteristics through the ensemble empirical mode decomposition method and then expressed these functions with the random

vector function chain network. Finally, the output results of different networks were combined to obtain the final prediction results. The results show that the effect of ensemble empirical mode decomposition is better than that of empirical mode decomposition. The above studies show that the signal decomposition method for original data can improve the prediction performance of the model to a certain extent if it can overcome the mode mixing phenomenon in the empirical mode decomposition.

To sum up, traffic flow prediction mainly faces the following two problems. One is how to reduce the influence of noise contained in the original traffic flow data on the prediction results. The other is how to accurately show the irregular periodic variation characteristics of traffic flow data. In view of the obvious improvement of traffic flow data prediction performance by the data denoising method and signal decomposition method, this paper proposes a WD-VMD-LSTM hybrid model for traffic flow prediction. This method first decomposes the complex original traffic flow data into multiple subsequences containing more prominent features through the variational mode decomposition method and then performs wavelet denoising on several subsequences, respectively. Finally, the long short-term memory network model is used to predict the denoised subsequences, respectively, and the final prediction results are obtained by combining the results of different subsequences. Compared with the empirical mode decomposition, the variational mode decomposition method can effectively avoid the phenomenon of mode mixing and boundary effect. The decomposed subsequence contains the data characteristics in the original signal. At the same time, the VMD method also has the advantages of anti-noise interference, so it is not easy to be affected by noise in the process of signal decomposition. On the other hand, the above research shows that the wavelet denoising method can effectively reduce the influence of noise on traffic flow prediction. The data characteristics contained in the original traffic flow data are difficult to identify, and the denoising processing will affect the characteristics of the original signal. Therefore, the wavelet denoising of the subsequence obtained by the variational mode decomposition can highlight the characteristics of the original signal and avoid the denoising method to suppress the useful signal. In addition, this paper compares LSTM with artificial neural network in the part of result discussion. The long short-term memory network model is more suitable for predicting complex time series data as a deep learning method.

The main contributions of this paper are summarized as follows:

(1) A denoising method combining variational mode decomposition and wavelet threshold denoising is proposed to process the original traffic flow data. At the same time, this paper compares the prediction performance of different prediction models before and after data processing.

(2) In order to avoid the phenomenon of modal aliasing and the increase of data complexity, the number of components K is determined according to the sample entropy with different K values.

(3) This study compares the denoising effects of different signal decomposition methods combined with the wavelet threshold denoising model. The advantage of variational mode decomposition in dealing with traffic flow data is discussed.

(4) Adam (its name is derived from adaptive moment estimation) optimizer is used to obtain a better model when training LSTM.

(5) In this study, two different public datasets are used to comprehensively compare different prediction models, and it is proved that the proposed model has better prediction performance than other comparison models.

The rest of this paper is organized as follows. The methods of WD, VMD, and LSTM are briefly introduced, and the process of the proposed VMD-WD-LSTM model is listed in Section 2. Section 3 demonstrates experiments where the prediction results of the proposed model and comparison models are evaluated. Finally, Section 4 provides conclusion of this research and makes the next research plan.

2. Methodology

2.1. Variational Mode Decomposition. Variational mode decomposition (VMD) is a method of signal processing using completely nonrecursive modal variation. Compared with traditional empirical mode decomposition (EMD), this technology can artificially determine the number of modal decompositions and then realize the frequency-domain decomposition and effective separation of IMF according to the best center frequency and limited bandwidth of each component after decomposition. In this way, the effective decomposition component of the target signal is obtained, and the optimal solution of the variational problem is realized. Variational mode decomposition has a solid theoretical foundation, and there is no end effect of traditional empirical mode decomposition and the problem of modal component aliasing. This method can reduce the complexity and nonstationarity of nonlinear time series and can decompose multiple stationary subsequences with different frequency scales.

The first step of variational mode decomposition is to construct the variational problem and solve the constrained variational problem:

$$\min_{\{u_k\}\{w_k\}} \left\{ \sum_k \left\| \partial_t \left[\left(\delta(t) + \frac{j}{\pi t} \right) * u_k(t) \right] e^{-jw_k t} \right\|_2^2 \right\}$$

$$\text{s.t.} \quad \sum_k u_k(t) = f(t),$$

(1)

where $u_k(t)$ is input signal modal function, $\{u_k\}$ is the k-th modal component with limited bandwidth after decomposition, $\{w_k\}$ is the center frequency corresponding to the k-th modal component of the input signal, $\delta(t)$ is the Dirac function, $*$ represents the convolution operator, and $f(t)$ is the input signal.

Then, introduce the Lagrangian multiplication operator λ and the quadratic penalty factor α to rewrite formula (1) to transform the constrained variational problem into an unconstrained variational problem. The rewritten formula is as follows:

$$L(\{u_k\}, \{w_k\}, \lambda) = \alpha \sum_k \left\| \partial_t \left[\left(\delta(t) + \frac{j}{\pi t} \right) * u_k(t) \right] e^{-jw_k t} \right\|_2^2$$
$$+ \left\| f(t) - \sum_k u_k(t) \right\|_2^2 + \left[\lambda(t), f(t) - \sum_k u_k(t) \right]. \tag{2}$$

Use the alternating direction multiplication algorithm (ADMM) to solve equation (2) and obtain the optimal solution of the respective center frequencies of a group of modal components, that is, alternately update u_k, w_k, and λ to obtain the minimum point of the extended Lagrangian expression, and the formula is as follows:

$$\hat{u}_k^{n+1}(w) = \frac{\hat{f}(w) - \sum_{i=1}^{k-1} \hat{u}_k^{n+1}(w) - \sum_{i=k+1}^{k} \hat{u}_k^n(w) + \left(\hat{\lambda}^n(w)/2 \right)}{1 + 2\alpha (w - w_k^n)^2},$$

$$w_k^{n+1} = \frac{\int_0^\infty w |\hat{u}_k^{n+1}(w)|^2 dw}{\int_0^\infty |\hat{u}_k^{n+1}(w)|^2 dw},$$

$$\hat{\lambda}^{n+1}(w) = \hat{\lambda}^n(w) + \tau \left(\hat{f}(w) - \sum_k |\hat{u}_k^{n+1}(w)| \right). \tag{3}$$

After dividing the frequency band according to the characteristics of the original signal, continuously update the center frequency of each inherent modal component and the corresponding component, and finally, realize the adaptive decomposition of the target signal according to the constraints.

2.2. Wavelet Threshold Denoising.

The noise in the original data is usually a high-frequency signal, and the useful data are regarded as a low-frequency signal. Wavelet decomposition decomposes the signal into approximate components containing low-frequency signals and detailed components containing high-frequency signals. The part containing low-frequency signals can be further decomposed, as shown in Figure 1. Figure 1 is a three-layer wavelet decomposition diagram, cA1, cA2, and cA3 represent the low-frequency signal part of the original signal, while cD1, cD2, and cD3 represent the high-frequency signal part of the original signal. The cD1, cD2, and cD3 contain noise. In this study, cD1, cD2, and cD3 are processed by wavelet threshold denoising, and then we reconstruct the signal by wavelet transform. Finally, the denoising results are obtained.

Wavelet threshold denoising uses the continuity characteristics of the original signal in the time series, and the wavelet coefficient of noise is smaller than the wavelet coefficient of the useful signal. Select an appropriate threshold
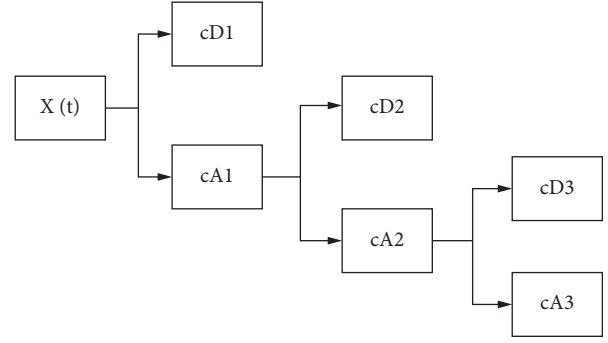

FIGURE 1: Wavelet decomposition process (cA is low-frequency information; cD is high-frequency information).

through this feature, quantize the wavelet coefficients, and then reconstruct the wavelet coefficients to obtain the denoised data.

Wavelet threshold denoising can be divided into hard threshold denoising and soft threshold denoising in selecting threshold function. In terms of the effect of signal denoising, the signal after soft threshold denoising is smoother, but it is easy to remove some useful signals. The signal after hard threshold denoising will oscillate and there will be jumping points, but the error should be lower than the soft threshold. Denoising does not affect the degree of approximation between the denoised signal and the original signal. Therefore, from the perspective of ensuring the accuracy of the prediction result, this study chooses to perform hard threshold denoising. Hard threshold denoising is when the wavelet coefficient is greater than the threshold, it is determined to be generated by the signal, and it is retained after processing. When the wavelet coefficient is less than the threshold, it is determined to be noise generated and replaced with 0, as shown in the following equation:

$$w_\lambda = \begin{cases} w, & |w| \geq \lambda, \\ 0, & |w| < \lambda, \end{cases} \tag{4}$$

where w_λ represents the new wavelet coefficient and λ represents the set threshold.

2.3. LSTM Network.

At present, the deep learning model has been widely used in the research of time series data. As a kind of neural network model, the deep learning model can extract the characteristics of the input signal and obtain the law of the complex signal. Among the deep learning models, the recurrent neural network (RNN) shows good adaptability when performing time series data analysis. Long short-term memory (LSTM) network is a variant of cyclic neural network, which improves the problems of gradient explosion and gradient disappearance in cyclic neural network and performs better in analyzing time series data.

The LSTM network is composed of an input layer, a hidden layer, and an output layer. Compared with the traditional RNN, the hidden layer of the LSTM is a unit with a unique memory mode. Figure 2 shows the hidden layer structure of the RNN and the hidden layer structure of the LSTM.
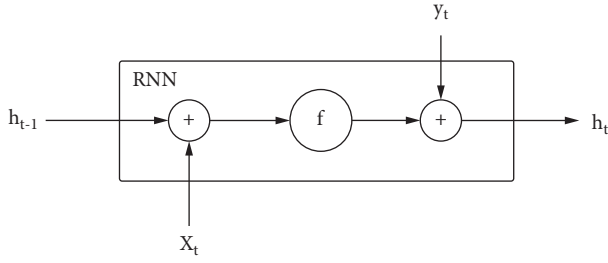
Figure 2: RNN architecture.

The memory unit is the core of the LSTM unit structure (see Figure 3). The memory unit at the current time t is marked as c_t. The memory unit can delete or add information through input gates, forget gates, and output gates. Specifically, the workflow of the LSTM unit is as follows:

(1) The LSTM unit receives the current state x_t, the hidden state h_{t-1} of the LSTM at the previous moment, and the state c_{t-1} of the internal memory unit through the input gate, forget gate, and output gate at each moment.

(2) After receiving the information, each gate performs operations on the information from different sources and decides whether to activate it.

(3) After the information received by the input gate is transformed by a nonlinear function, it is combined with the state of the internal memory unit processed by the forget gate to form a new memory unit state c_t, and then the newly formed memory unit state is formed by the dynamic control of the output gate. The output information h_t outputs the LSTM unit.

The calculation relationship between various variables is as follows:

$$
\begin{aligned}
i_t &= \sigma\left(W_{xi}x_t + W_{hi}h_{t-1} + W_{ci}c_{t-1} + b_i\right), \\
f_i &= \sigma\left(W_{xf}x_t + W_{hf}h_{t-1} + W_{cf}c_{t-1} + b_f\right), \\
c_t &= f_t c_{t-1} + i_t \tanh\left(W_{xc}x_t + W_{hc}h_{t-1} + b_c\right), \quad (5) \\
o_t &= \sigma\left(W_{xo}x_t + W_{ho}h_{t-1} + W_{co}c_t + b_o\right), \\
h_t &= o_t \tanh\left(c_t\right).
\end{aligned}
$$

In the above equation, i, f, c, and o are input gate, forget gate, cell state, and output gate, respectively. W_{xi}, W_{xf}, W_{xc}, and W_{xo} are all weight coefficient matrices linking the input signal x_t, and W_{hi}, W_{hf}, W_{hc}, W_{ho} are the weight coefficient matrices of the input signal h_t of the link hidden layer, and W_{ci}, W_{cf}, W_{co} are the diagonal matrices to link neuron activation function of the output vector c_t with the gate function. b_i, b_c, b_f, and b_o are bias vectors, σ is the sigmoid activation function, and tanh is the hyperbolic tangent activation function.

2.4. The Proposed Model (VMD-WD-LSTM). The framework of the VMD-WD-LSTM-based traffic flow prediction model is shown in Figure 4. The main steps of the VMD-WD-LSTM model are as follows:

(1) The original traffic flow data are decomposed into multiple eigenmode functions (IMFs) by VMD, and the number of IMFs is determined by the sample entropy of the reconstructed data under different K values.

(2) Each IMF is processed by the hard threshold function denoising method of wavelet threshold denoising, and the denoised subsequences are obtained.

(3) LSTM is employed to predict each subsequence, and the predicted value of each subsequence is synthesized into the final prediction result.

3. Experiments

This section provides a concise and precise description of the experimental results, their interpretation, and the experimental conclusions that can be drawn.

3.1. Data Description. The open-source data used are selected from the PeMS database, which collects traffic data from more than 39,000 individual detectors. The sensor layout covers the highway system in all metropolitan areas in California. Specifically, the experimental data of this paper are collected from the Kumeyaay Highway in California. This paper selects three detectors from many detectors (see Figure 5). We took the complete traffic flow data for five consecutive days from Monday to Friday from these three detection points for analysis. In addition, in order not to affect the accuracy of the prediction results, the dates we selected do not include holidays. The time period is from 14 September 2020 to 18 September 2020, the time interval of traffic flow is 5 minutes, and the number of samples is 1440 data. All the data obtained are divided into training set and testing set. From Monday to Thursday, 1152 (80% of all data) data are used as training set, and 288 (20% of all data) data on Friday are used as testing set. The three detectors are represented by A, B, and C, respectively. Figure 5 gives the location of each detector, and the detailed information of each detector is shown in Table 1. In this experiment, the traffic flow data of the three detectors are all used to test and analyze the performance of the proposed model. Limited to the length of the paper, the traffic flow data of detector A are taken as an example to illustrate the specific operations and results of each step of the proposed model.

The original traffic flow data of the three detectors are illustrated in Figure 6. It can be seen that the traffic flow from Monday to Friday has obvious periodicity, and the characteristics of the daily traffic volume change are obvious, but there is obvious nonlinearity and volatility. Part of the reason is due to the presence of noise.

As mentioned in the experiment, the traffic flow data obtained from the detector are easily affected by various unexpected factors, and there is a certain degree of abnormal fluctuations. Therefore, the abnormal value of the data is suppressed through the variational mode decomposition and wavelet threshold denoising, so as to obtain reliable traffic flow data.
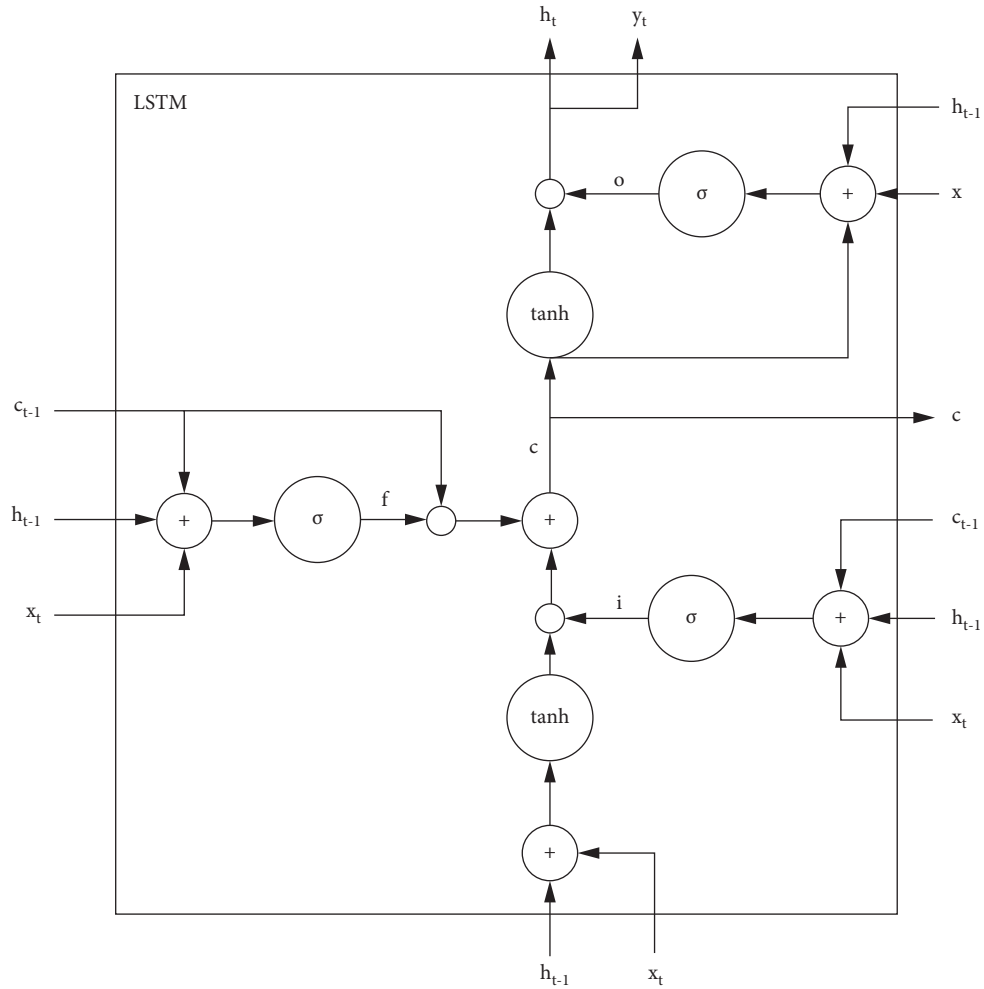
Figure 3: LSTM neural network architecture.

3.2. Evaluation Indexes. The experiment uses three commonly used standards to evaluate the advantages and disadvantages of the model. The three standards are the root mean square error (RMSE), the average absolute error (MAE), and the average absolute percentage error (MAPE), which are defined as follows:

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^{n} \left(f_i - \widehat{f}_i\right)^2},$$

$$MAE = \frac{1}{n} \sum_{i=1}^{n} \left(f_i - \widehat{f}_i\right), \qquad (6)$$

$$MAPE = \frac{1}{n} \sum_{i=1}^{n} \left|\frac{f_i - \widehat{f}_i}{f_i}\right| \times 100\%,$$

where f_i represents the real traffic flow data, $\widehat{f}_i$ represents the predicted data, and n is the number of samples.

3.3. VMD Results of Traffic Flow Data. The original traffic flow data of detector A are decomposed using VMD. And it is important to determine that original traffic flow data

should be decomposed to how many IMFs (each IMF corresponds to a reconstructed component). On the one hand, too few IMFs may not be able to extract the features hidden in the original data. On the other hand, too many IMFs may lead to a poor prediction result because of prediction error accumulation in the ensemble step. In this study, the optimal number of IMFs is determined according to the sample entropy values corresponding to the reconstructed component with different number of IMFs. The greater the sample entropy, the greater the complexity of the sample sequence, which makes data prediction more difficult. Therefore, the number of IMFs corresponding to the minimum sample entropy is the optimal number of decompositions. The greater the sample entropy, the greater the complexity of the sample sequence, which makes data prediction more difficult. Therefore, it is necessary to select the K value that minimizes the entropy of the sample as the decomposition number. The sample entropy values corresponding to the number K of IMFs are shown in Figure 7. It can be seen from Figure 7 that when $K = 3$, the obtained sample entropy is the smallest, and the optimal number of IMFs is 3. The three IMF components obtained by VMD are shown in Figure 8.
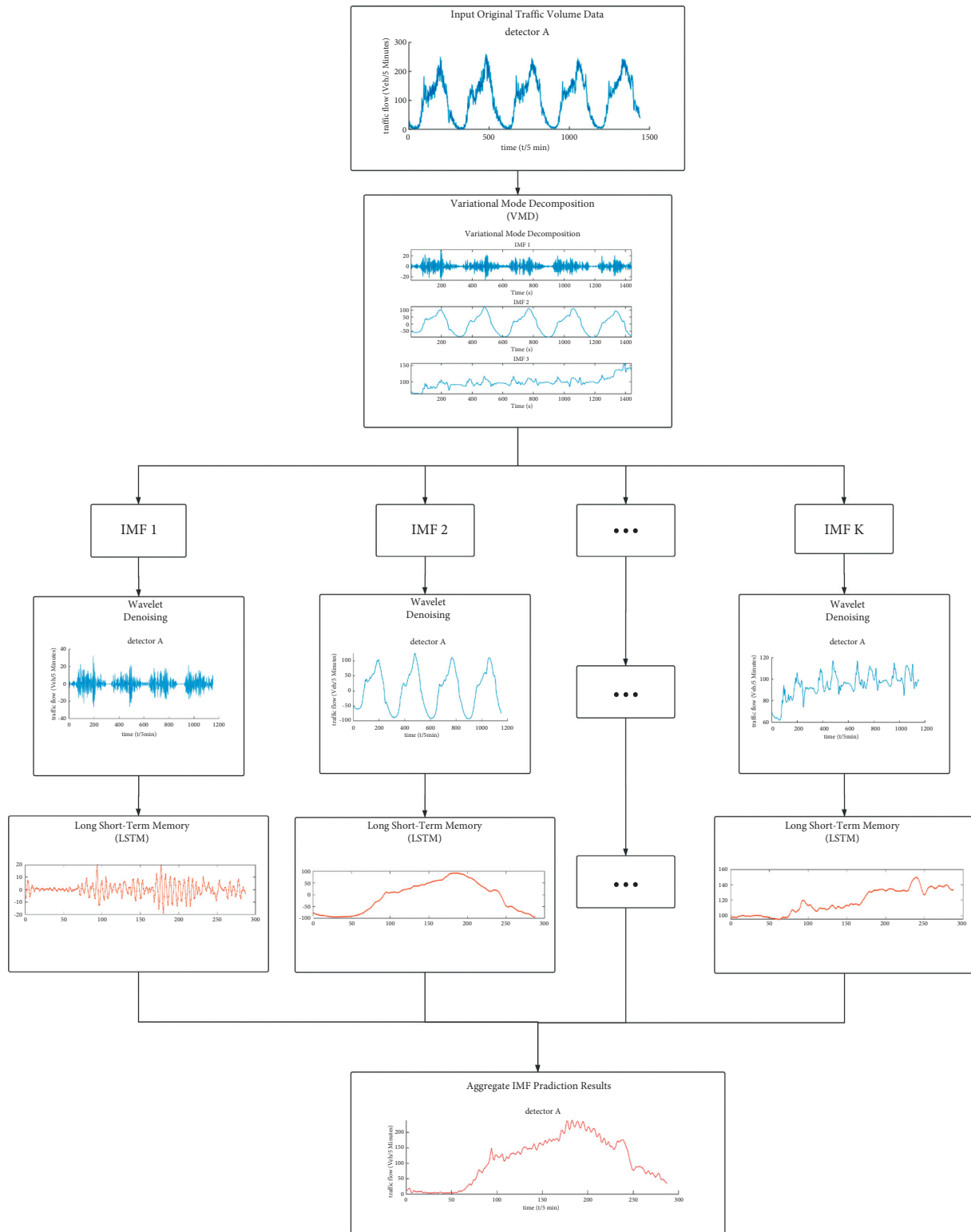
Figure 4: The framework of the VMD-WD-LSTM model.

3.4. Wavelet Denoising of Each IMF. In order to further reduce the influence of noise on the prediction results, wavelet transform is used to denoise each IMF. In order to ensure a more impressive result in terms of root mean square error, a hard threshold function is selected. The result of wavelet denoising can be seen from Figure 9. It can be seen that the data curve after wavelet denoising is smoother and the data features are more obvious.
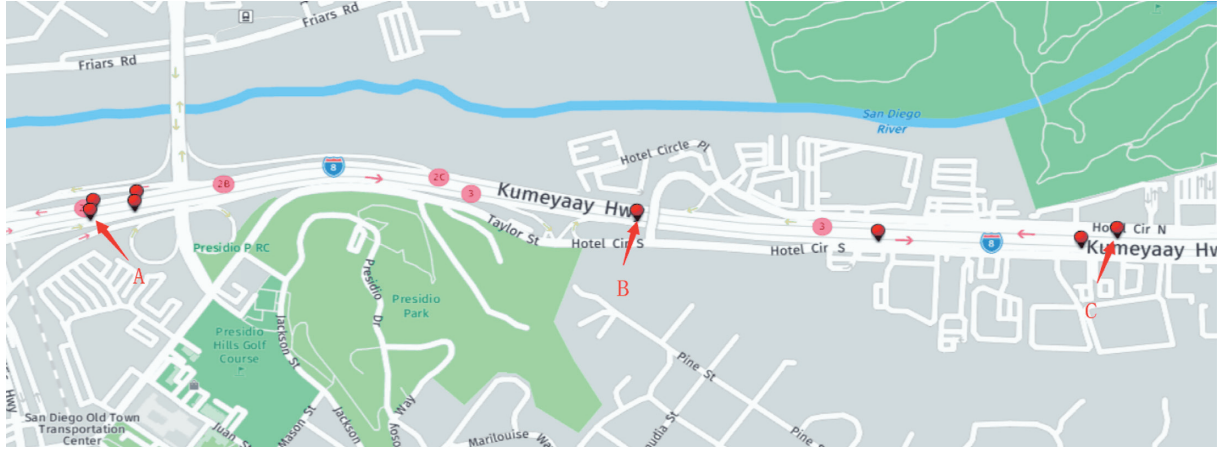
Figure 5: Location of the three detectors.

Table 1: Detailed information for the three detectors.

Location	Freeway	Detector number	Max cap	Lane point
A	I8	1113028	68.2	2
B	I8	1115407	161.0	5
C	I8	1115549	162.6	5

3.5. Results and Discussion. The LSTM model is employed to predict each subsequence obtained by VMD-WD, and then the predicted value of each subsequence is synthesized into the final prediction result. The LSTM model is composed of an input layer, a hidden layer, and an output layer. The hidden layer contains 200 neural units. The input feature dimension and the output feature dimension are both 1. In terms of options for training deep learning neural networks, the LSTM model uses the Adam optimizer. Specifically, the maximum number of training epochs is 250, and the initial learning rate is 0.005. In order to avoid the problem of gradient explosion, when the training reaches 125 epochs, the global learning rate is reduced by a multiplier factor (which is set to 0.2).

Figures 10–12 show the prediction results of IMF1 component, IMF2 component, and IMF3 component, respectively. It can be seen intuitively that the predicted curves and observed curves of IMF1, IMF2, and IMF3 components are highly fitted. Moreover, RMSE of the prediction results of the IMF1, IMF2, and IMF3 components is 1.8089 veh/5 min, 2.1161 veh/5 min, and 1.7235 veh/5 min, respectively, which also illustrate that the proposed model has superior performance. The characteristics of each IMF component can also be seen from these three figures. IMF1 represents the high-frequency component obtained from the original data, which shows the randomness of the original traffic flow data. IMF2 and IMF3 both represent low-frequency components, which show the regularity of the original traffic flow data. In summary, it is easier to capture the different features contained in the original signal by separately predicting each IMF component obtained by VMD. The final step is to accumulate the prediction results of each IMF component to get the final prediction result. Figure 13 shows the

cumulative prediction results. It can also be seen intuitively that the predicted curve and the observed curve are highly fitted.

In order to prove the influence of the VMD-WD hybrid denoising method on the final prediction results and to prove the prediction performance of LSTM model, this paper sets up a comparative study. Firstly, the prediction results of detector A are predicted by the traditional differential autoregressive integrated moving average (ARIMA) model, artificial neural network (ANN) model, and LSTM model, respectively, and the root mean square error and mean absolute error of the three models are compared (see Figure 14). Then, the original traffic flow data of detector A are denoised by the VMD-WD hybrid method in this study. Finally, the denoised data are predicted by ARIMA, ANN, and LSTM models, respectively, and RMSE and MAE are calculated for comparison. The final results are shown in Figures 14 and 15. It can be seen from the image that the prediction performance of LSTM model is significantly better than that of the ARIMA model and ANN model. Similarly, from the data denoised by the VMD-WD hybrid denoising method, the prediction results of the LSTM model are still the most accurate. On the other hand, the prediction performance of the hybrid prediction model combined with VMD-WD denoising method is better than that of the single prediction model. The prediction performance of the LSTM model is obviously improved. Compared with the LSTM model, the RMSE and MAE of the VMD-WD-LSTM model are reduced by 37.8% and 38.2%. The above results show that VMD-WD denoising method can improve the prediction performance of prediction model, and the prediction accuracy of the LSTM model based on deep learning is higher than that of traditional methods.
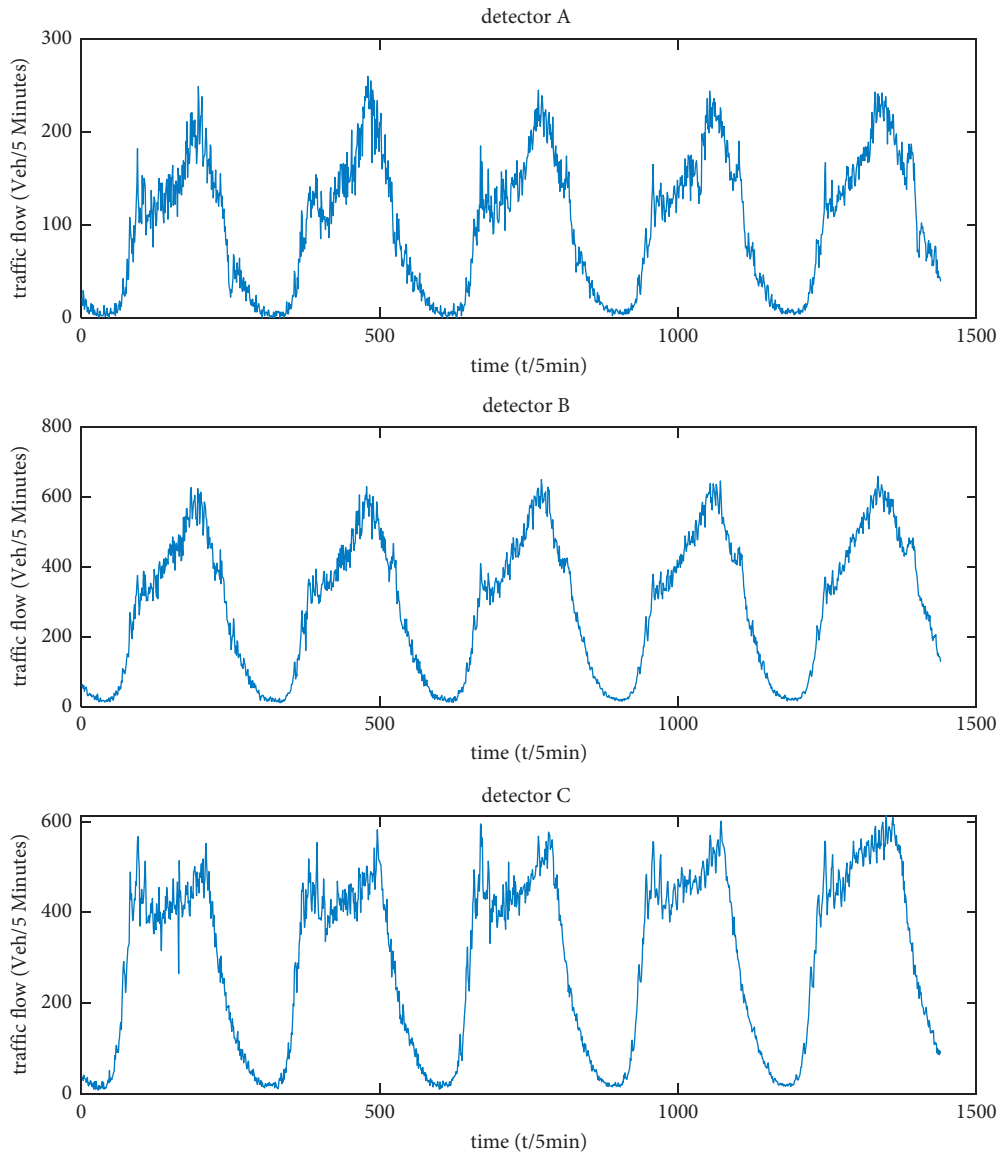
detector A

detector B

detector C

FIGURE 6: The original traffic flow data of the three detectors.
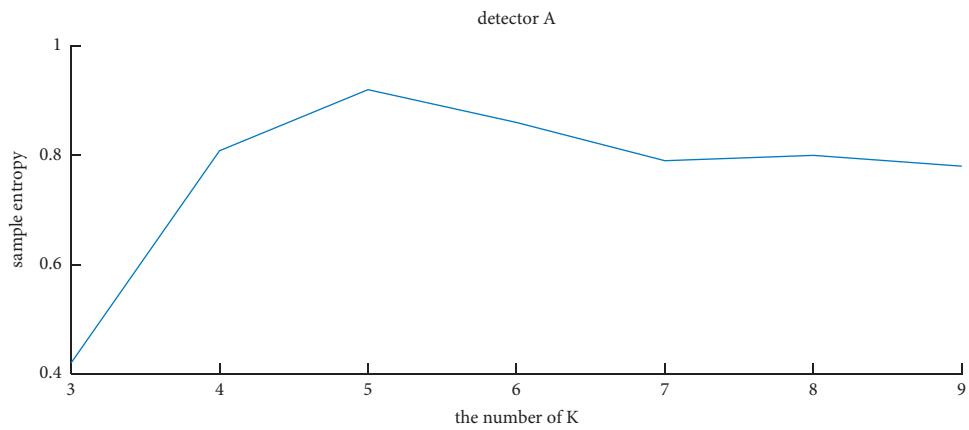
detector A

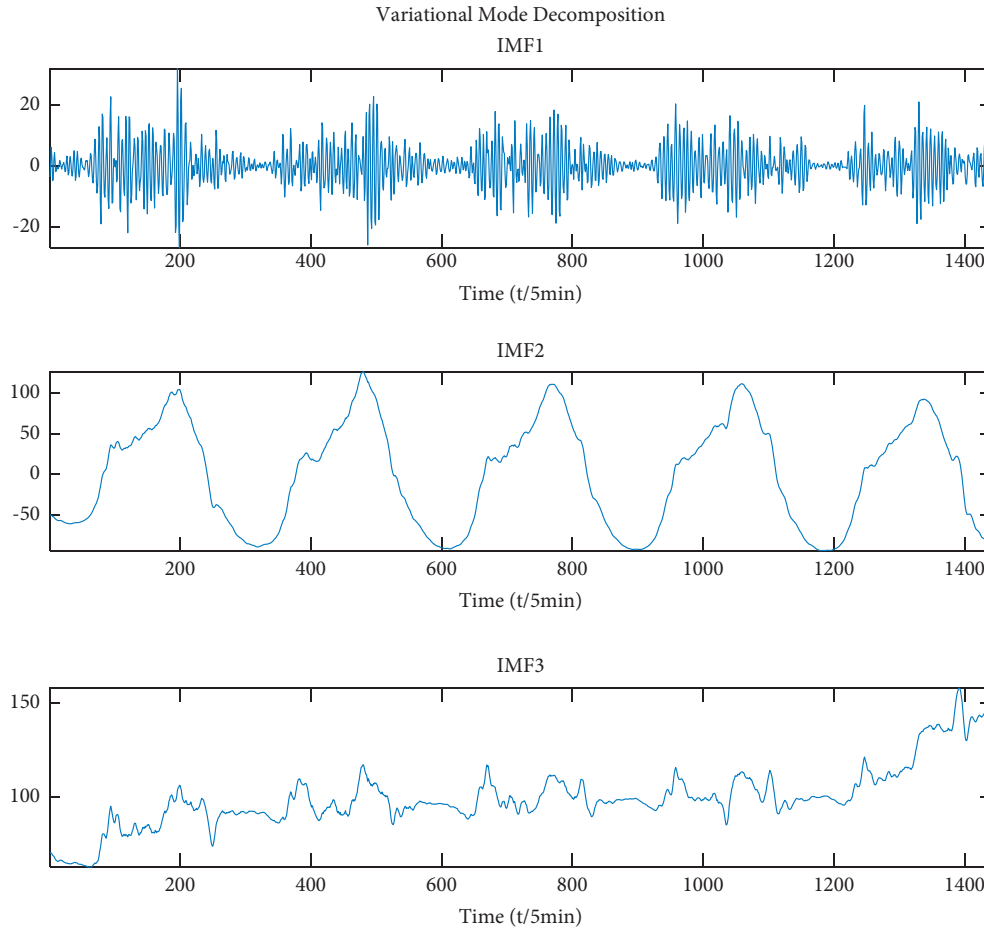FIGURE 7: Sample entropy with different K.

Figure 8: VMD result of detector A.

In order to further prove the prediction performance of the hybrid prediction model proposed in this study, seven different prediction models are introduced in this paper. The prediction results of the seven models are compared with those of the VMD-WD-LSTM method. The comparison methods include the prediction model without denoising steps, the model with signal decomposition by other methods, the models with different prediction methods, and the latest methods proposed in this research field in recent two years. The RMSE and MAE of the WD-LSTM method were 10.85 and 8.066, respectively, which were compared with the prediction results of the LSTM method. MAE and MAPE are reduced by 2%, 2.6%, and 28.9%, respectively. It can be seen that the prediction performance of the prediction model after wavelet threshold denoising is improved because the wavelet threshold denoising method can remove the noise signal in the original signal to a certain extent and retain the characteristics of the original signal, which reduces the interference of noise on traffic flow prediction. However, due to the complex time characteristics of the original signal, the denoising effect still has room for improvement. Then, this study compares the EMD-WD-LSTM method with the WD-LSTM method. It can be seen from the evaluation index that the RMSE and MAE of the EMD-

WD-LSTM method are further reduced compared with the latter. This is because the signal after the EMD method decomposes the original signal containing complex characteristics into simpler subsequences, which is not only conducive to the wavelet threshold method to identify high-frequency noise but also conducive to the LSTM model for prediction. However, the MAPE of EMD-WD-LSTM method is higher than that of the latter. This study believes that this is due to the limitation of EMD method itself. The IMF components obtained by EMD decomposition will have the phenomenon of modal mixing. When there are abnormal events and other disturbances in the signal, each IMF will contain more than one frequency component, which will affect the prediction performance of the prediction model to a certain extent. In the third step, this paper compares the prediction results of VMD-WD-LSTM method with the LSTM method, WD-LSTM method, and EMD-WD-LSTM method. It can be seen from the prediction indicators (Table 2 and Figure 16) that the prediction performance of the VMD-WD-LSTM method proposed in this study has been significantly improved compared with the above three methods. Compared with the EMD-WD-LSTM method, RMSE, MAE, and MAPE are increased by 32.5%, 32.9%, and 34.5%, respectively. This result is consistent with the description in the first part of

wavelet threshold denoising for IMF1

wavelet threshold denoising for IMF2
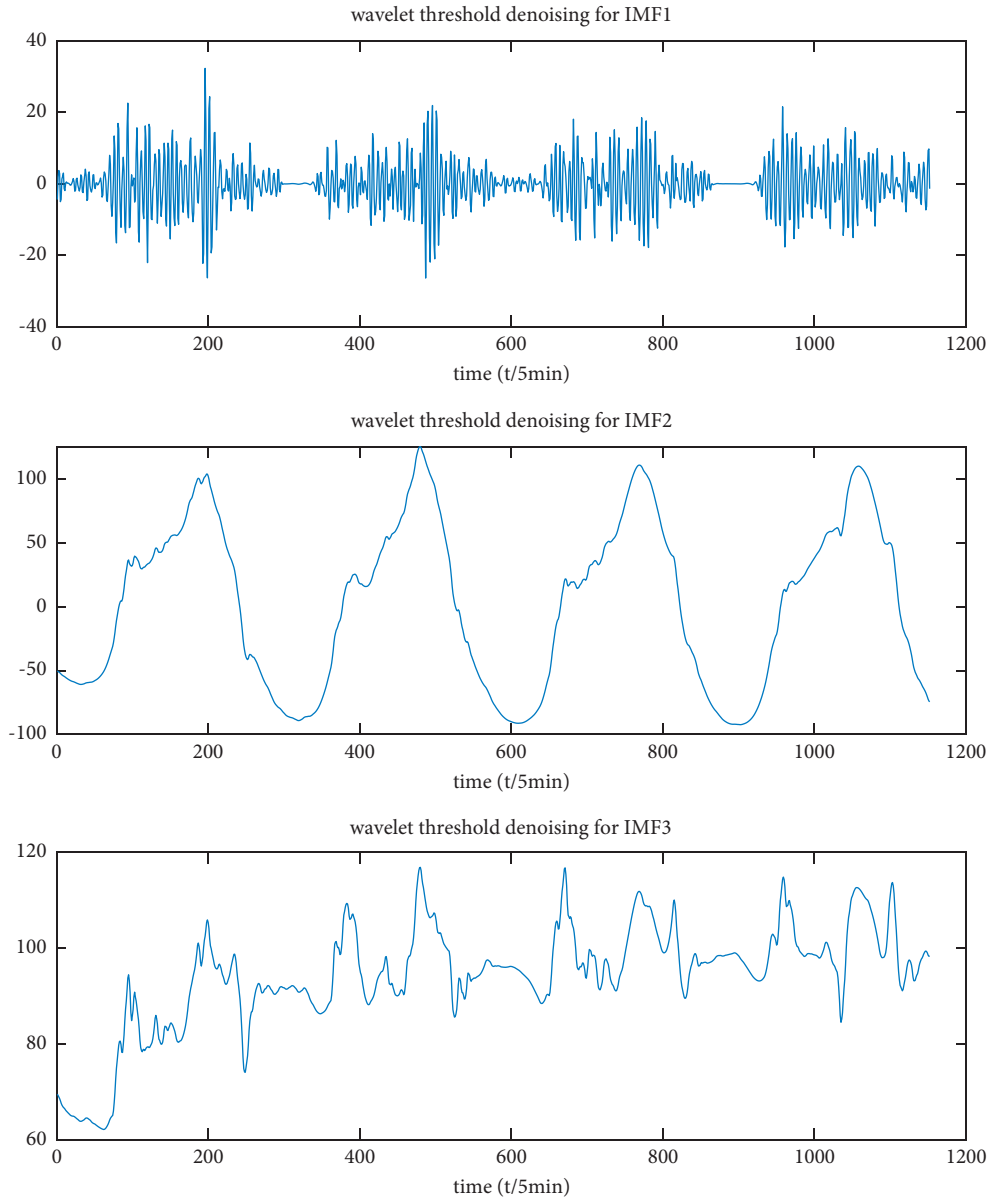
wavelet threshold denoising for IMF3

FIGURE 9: Curves of different IMF components based on WD processing.

this paper. Compared with EMD, the VMD method overcomes the problem of modal aliasing, so the decomposed subsequence is more conducive to denoising and prediction. On the other hand, the prediction performance of the VMD-WD-LSTM method is better than that of the VMD-WD-ANN method, which indicates that the LSTM model is more suitable for time series prediction than the traditional artificial neural network model. Finally, this paper selects the method proposed in the field of traffic flow prediction in the past two years to compare it with the VMD-WD-LSTM method proposed in this study. These methods are the EEMD-ANN model proposed by Chen et al. [8] in 2020, ARIMA-LSTM model proposed by Lu et al. [31], and TSD-BiLSTM model proposed by Huang et al. [29] in 2021. The operation steps and parameter

selection of the above model are strictly consistent with the literature, and the model parameters are shown in Table 3. Comparing the prediction results of the four models for detector A traffic flow data, it can be seen that the RMSE, MAE, and MAPE of the VMD-WD-LSTM method are the lowest among the four models. As shown in Figure 16, the prediction image of the VMD-WD-LSTM model is closest to the real data. This shows that the VMD-WD-LSTM model proposed in this study still has practical value compared with the new methods in this field.

The boxplots of the absolute errors of the different models are shown in Figure 17. For each boxplot, the central mark (red line) is the median; the edges of boxes are the 25th (Q1) and 75th (Q3) percentiles, and the interquartile range (IQR = Q3−Q1) is used for evaluating the degree of
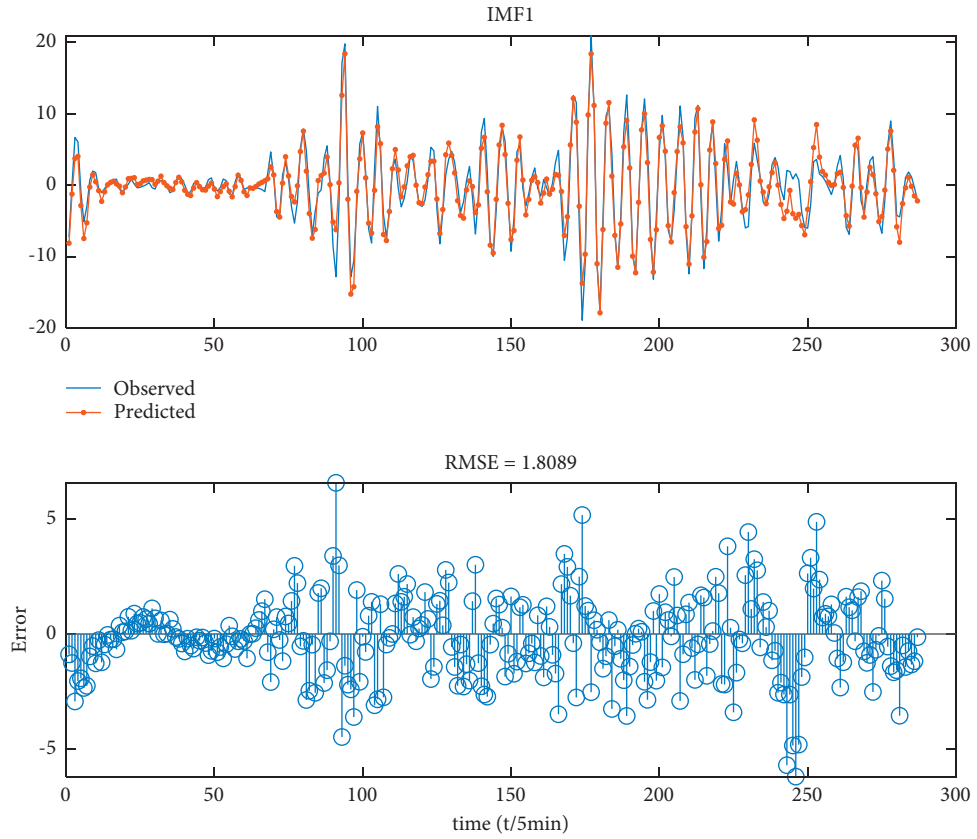
Figure 10: Prediction results of IMF1 (detector A).
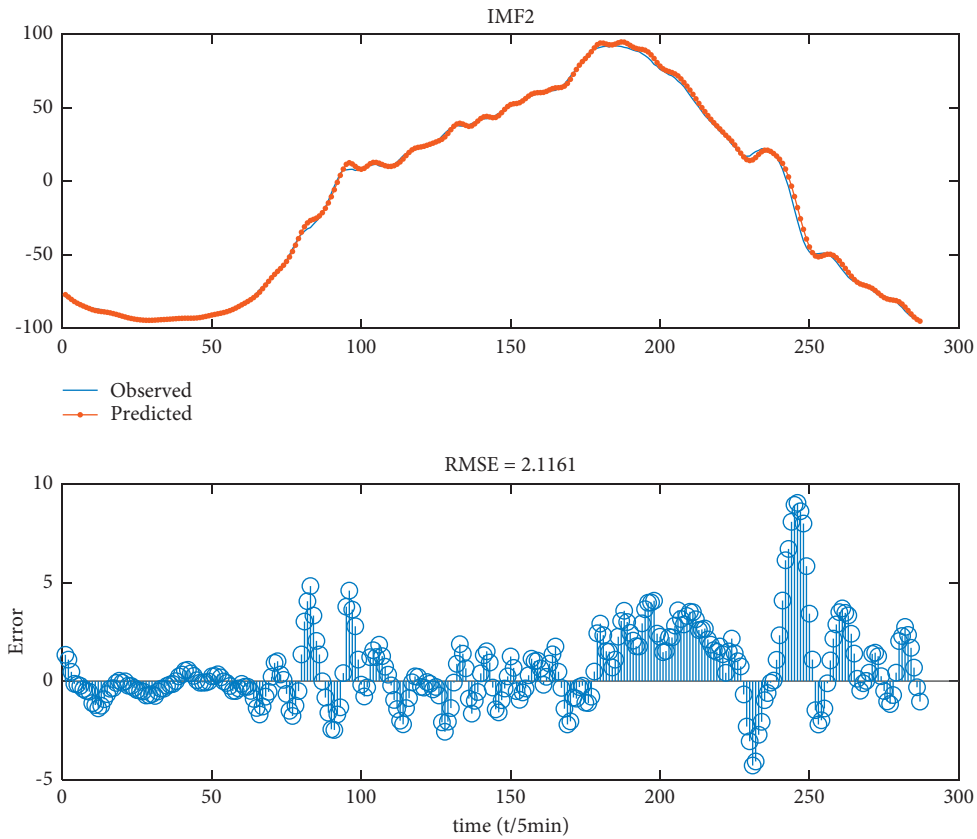

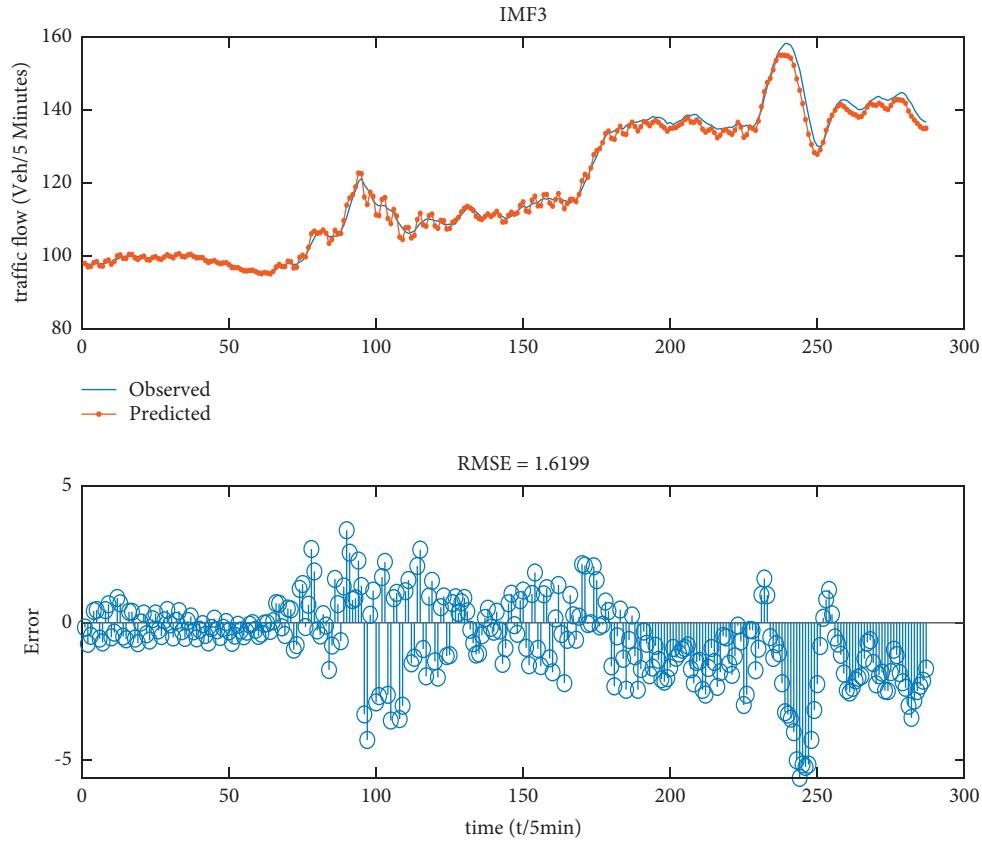Figure 11: Prediction results of IMF2 (detector A).

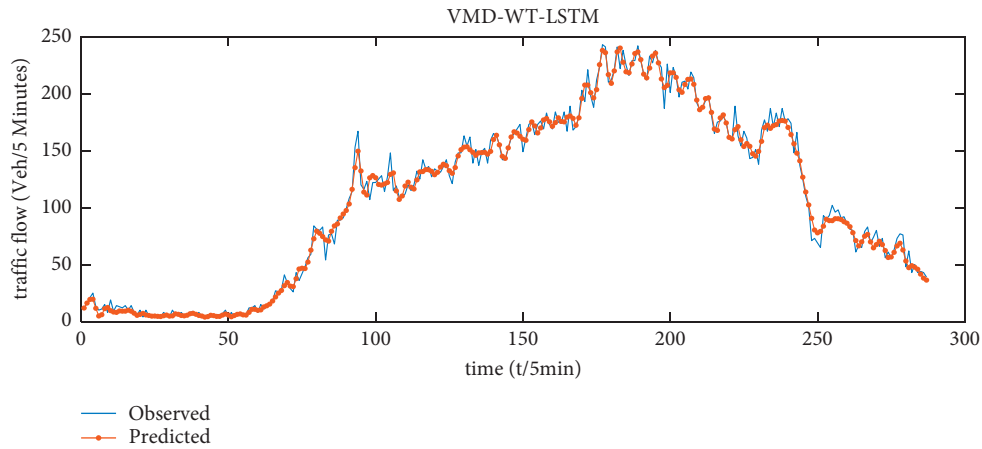FIGURE 12: Prediction results of IMF3 (detector A).

FIGURE 13: Prediction result of the VMD-WD-LSTM model (detector A).

concentration to median; the whiskers extending to the most extreme data points are not considered as outliers (abnormal data points). It can be seen from Figure 17 that the IQR of the absolute error of the VMD-WD-LSTM model is the smallest (that is, the fluctuation of the absolute error is the smallest), indicating the outstanding stability of this prediction model.

We continue to compare and analyze the performance of the proposed model on the traffic flow data of detectors B and C to analyze whether the model can maintain good

prediction performance on different traffic flow data. Figures 18 and 19 show the prediction curves of different models for detector B and detector C data, respectively. It can be clearly seen from Figures 18 and 19 that the predicted and observed values of the VMD-WD-LSTM model have achieved good fitting results on the traffic flow data of detector B and detector C. Tables 4 and 5 illustrate RMSE, MAE, and MAPE for different models with data from detectors B and C, respectively. Specifically, the RMSE,
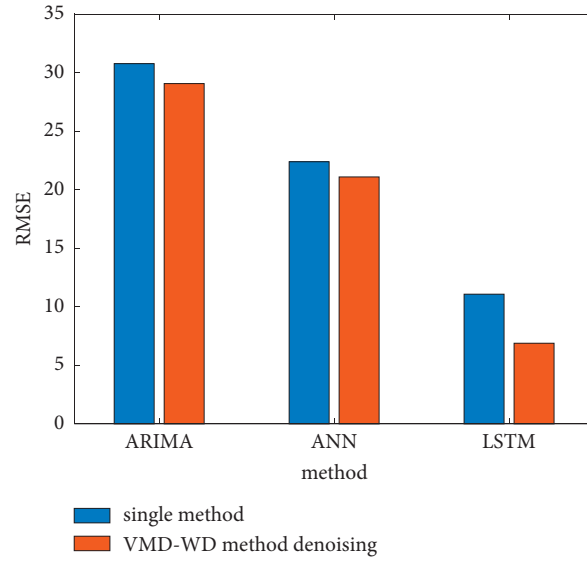
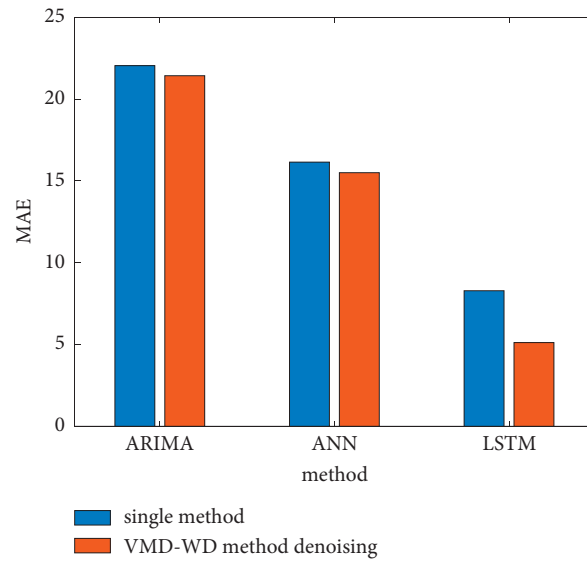Figure 14: RMSE of the proposed method and the three single methods.

Figure 15: Comparison of hybrid VMD-WD methods.

Table 2: The prediction errors of different models using the traffic flow data of detector A.

Prediction models	RMSE (veh/5 min)	MAE (veh/5 min)	MAPE (%)
LSTM	11.077	8.285	15.360
WD-LSTM	10.850	8.066	10.912
EMD-WD-LSTM	10.217	7.626	14.227
VMD-WD-ANN	21.092	15.498	20.030
EEMD-ANN	22.305	16.936	25.391
TSD-BiLSTM	16.448	11.877	17.247
ARIMA-LSTM	12.138	8.846	12.612
VMD-WD-LSTM	**6.888**	**5.116**	**9.313**

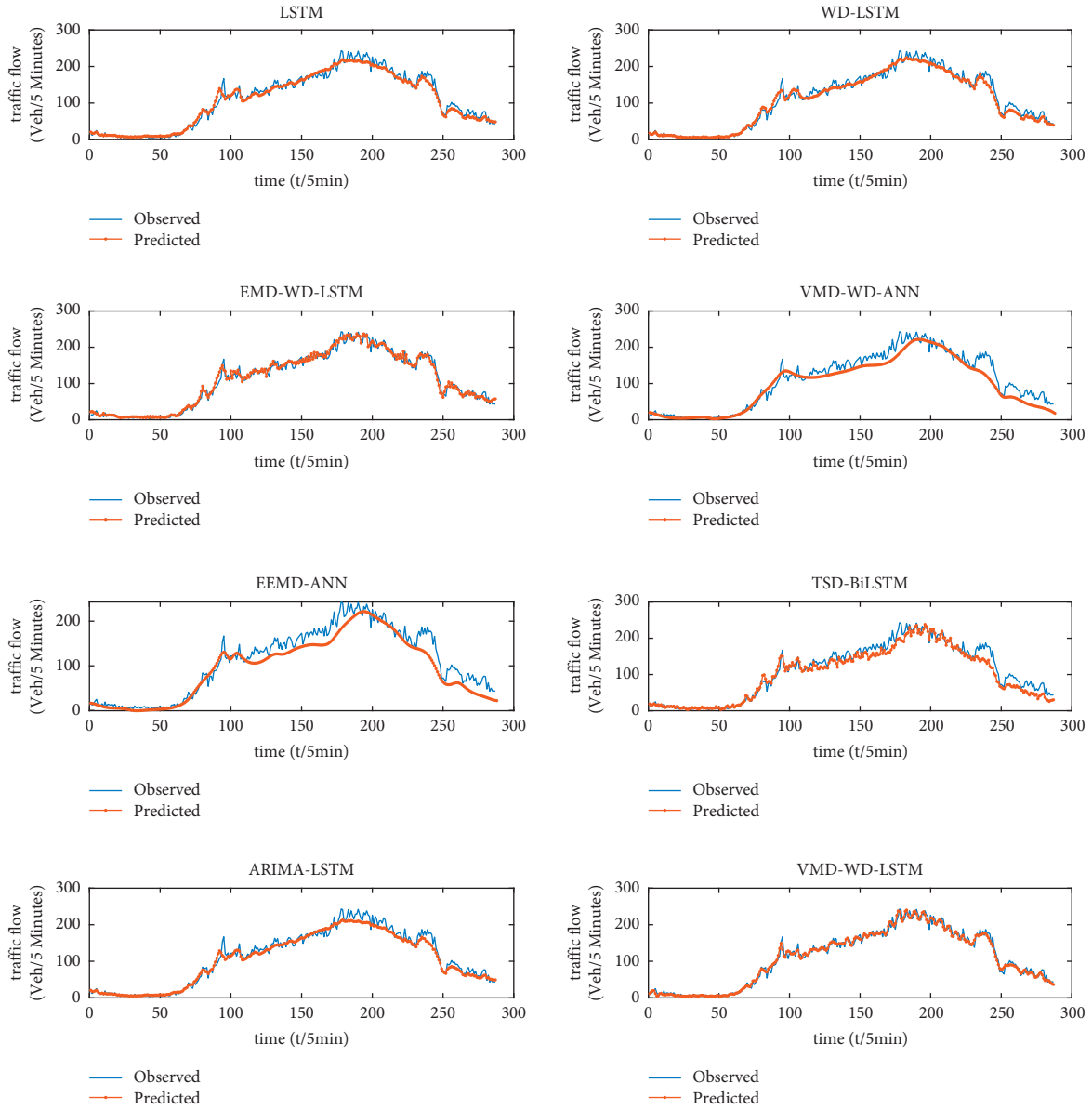Bold indicates the best value of each evaluation index.

Figure 16: Predicted curves of different models using the traffic flow data of detector A.

Table 3: Parameter configuration of different models.

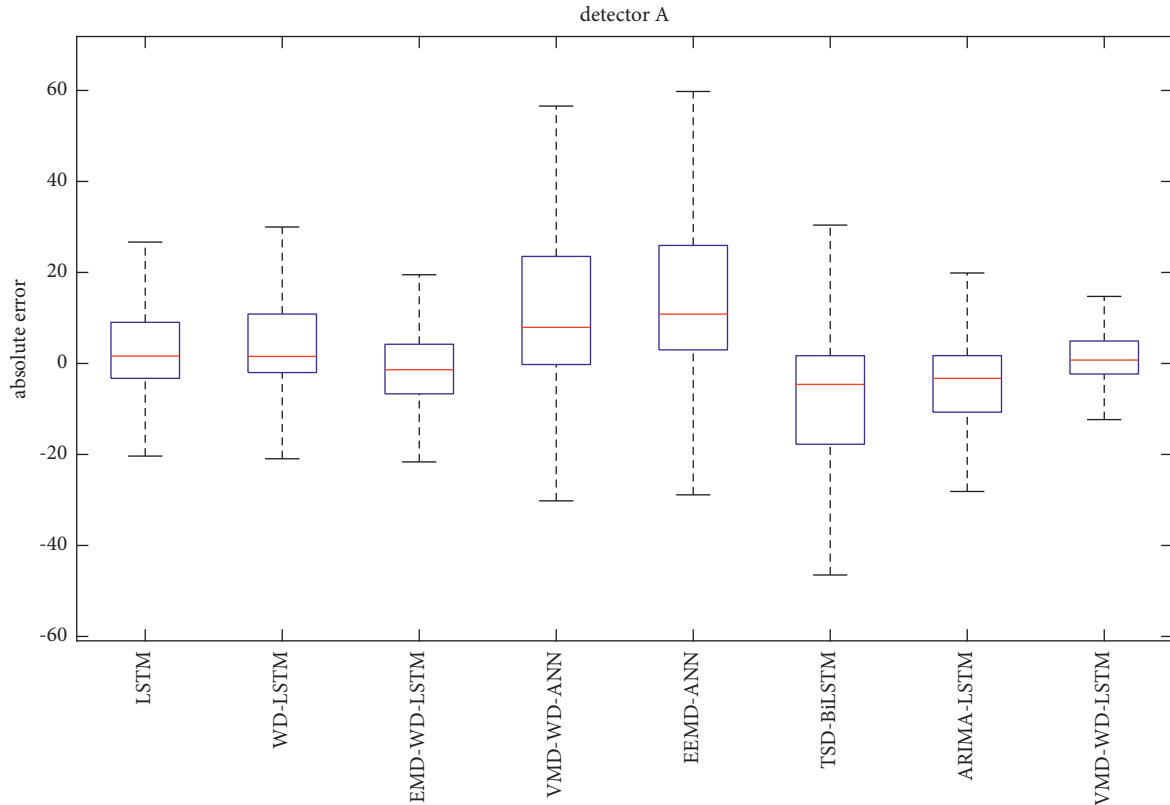| Model	Primary parameter	Setting value
ARIMA	p	2
	d	1
	q	4
EEMD	Added white noise	0.2
	The ensemble number	1000
LSTM	Learning rate	0.005
	The number of training epochs	250
	Number of hidden units	288
BiLSTM	Learning rate	0.005
	The number of training epochs	50
	Number of hidden units	32
ANN	Learning rate	0.005
	The number of training epochs	250
	Number of hidden units	288

FIGURE 17: Boxplots of absolute error of different models (detector A).

MAE, and MAPE of the VMD-WD-LSTM model on the detector B data are 10.304, 7.984, and 5.979%; the RMSE, MAE, and MAPE of the VMD-WD-LSTM model for detector C data are 13.980,10.965, and 5.268%. It can be seen that RMSE and MAE on detector B data of the VMD-WD-LSTM model and RMSE, MAE, and MAPE on detector C data are the smallest, while detection point B and detection point C have more lanes and larger traffic flow than detection point A, indicating that the method proposed in this study can maintain excellent performance in dealing with different types of road sections and different scales of traffic flow. We further analyze the comparison methods one by one. First, from the traffic flow prediction results of the two detection points, the prediction results of the method after signal decomposition and denoising are more accurate than the results obtained by a single prediction model. From the evaluation indicators RMSE and MAE, it can be seen that the VMD-WD-LSTM method proposed in this study still shows the best prediction performance, but the MAPE of the VMD-WD-LSTM model (5.979%) is slightly higher than that of the EMD-WD-LSTM model (5.208%). This shows that the error of the results predicted by the EMD-WD-LSTM method is smaller than that of the original data. In this case, we believe that the traffic flow has obvious cyclical characteristics and is affected by various external factors. However, the external factors are difficult to predict. The difference in the error between the predicted flow and the actual flow in a small enough range does not mean that there is a significant difference in the prediction

performance. On the other hand, this paper also recognizes that the VMD-WD-LSTM model still has room for improvement, and different types of external factors should be considered as important factors affecting traffic flow prediction in future research. The RMSE, MAE, and MAPE of the VMD-WD-LSTM model are much smaller than those of the VMD-WD-ANN model, which further proves that LSTM has better prediction accuracy due to its advantages in long-term dependence of capturing time series. The RMSE, MAE, and MAPE of the VMD-WD-LSTM model on the three detector data are smaller than those of the WD-LSTM model, indicating that the wavelet threshold denoising method combined with VMD can not only effectively improve the prediction accuracy but also show good robustness. In addition, the RMSE and MAE of the WD-LSTM model and the EMD-WD-LSTM model on the data of detector B and detector C are slightly smaller than those of the single LSTM model, while the MAPE is slightly larger, indicating that only the wavelet threshold denoising method cannot stably maintain the effect of reducing the noise data interference when dealing with the traffic flow data with complex characteristics. The number of IMF components obtained by EMD is uncertain. When removing high-frequency components for EMD-based denoising, the direction of the original signal will be affected. Therefore, it is difficult for the EMD-WD-LSTM model to maintain stable prediction accuracy on different datasets. The prediction errors of the VMD-WD-ANN model and EEMD-ANN model are generally large. The
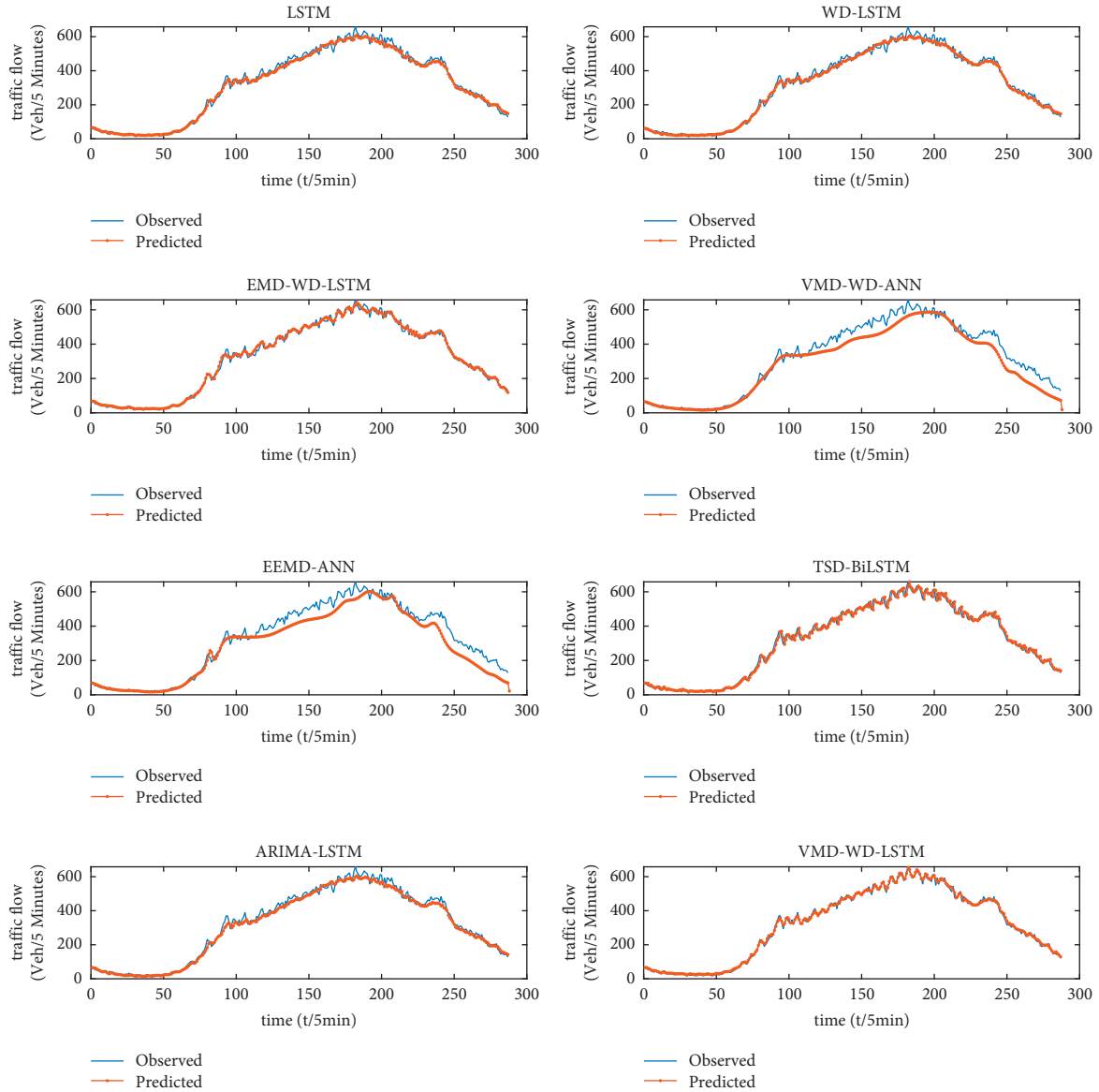
FIGURE 18: Predicted curves of different models using the traffic flow data of detector B.

reason may be that the ANN model can predict the trend of traffic flow sequence but cannot capture the time-varying characteristics of traffic flow (see Figures 16, 18, and 19). In addition, compared with several new methods, the VMD-WD-LSTM method has maintained a good prediction effect as always. In summary, the VMD-WD-LSTM model has the highest prediction accuracy and the strongest robustness in different comparison methods.

Figures 20 and 21 are boxplots of the absolute errors of different models using the traffic flow data of detector B and detector C, respectively. It can be seen that the VMD-WD-LSTM model has the smallest IQR on the detector B data and the detector C data, although it is not as obvious as on the detector A data. This result is sufficient to demonstrate that the VMD-WD-LSTM model not only has the smallest MAE but also has the smallest fluctuation range of the absolute error. In addition, compared with the absolute error of

detector A data, the absolute error of the VMD-WD-ANN model and the EEMD-ANN model on the detector B data and detector C data has increased significantly, which also shows that the performance of the ANN model is not stable enough. The reason for this phenomenon may be caused by the large traffic flow of detector B and detector C during peak hours. On the traffic flow data of the three detectors, the absolute error of the WD-LSTM model is more concentrated than that of the single LSTM model, which also proves the effectiveness of WD-based preprocessing of the traffic flow time series.

Based on the previous analysis results, it can be seen that the VMD-WD-LSTM model is able to predict the traffic flow collected from different detectors with a smaller absolute error than that of the comparison model. The results obtained are encouraging.
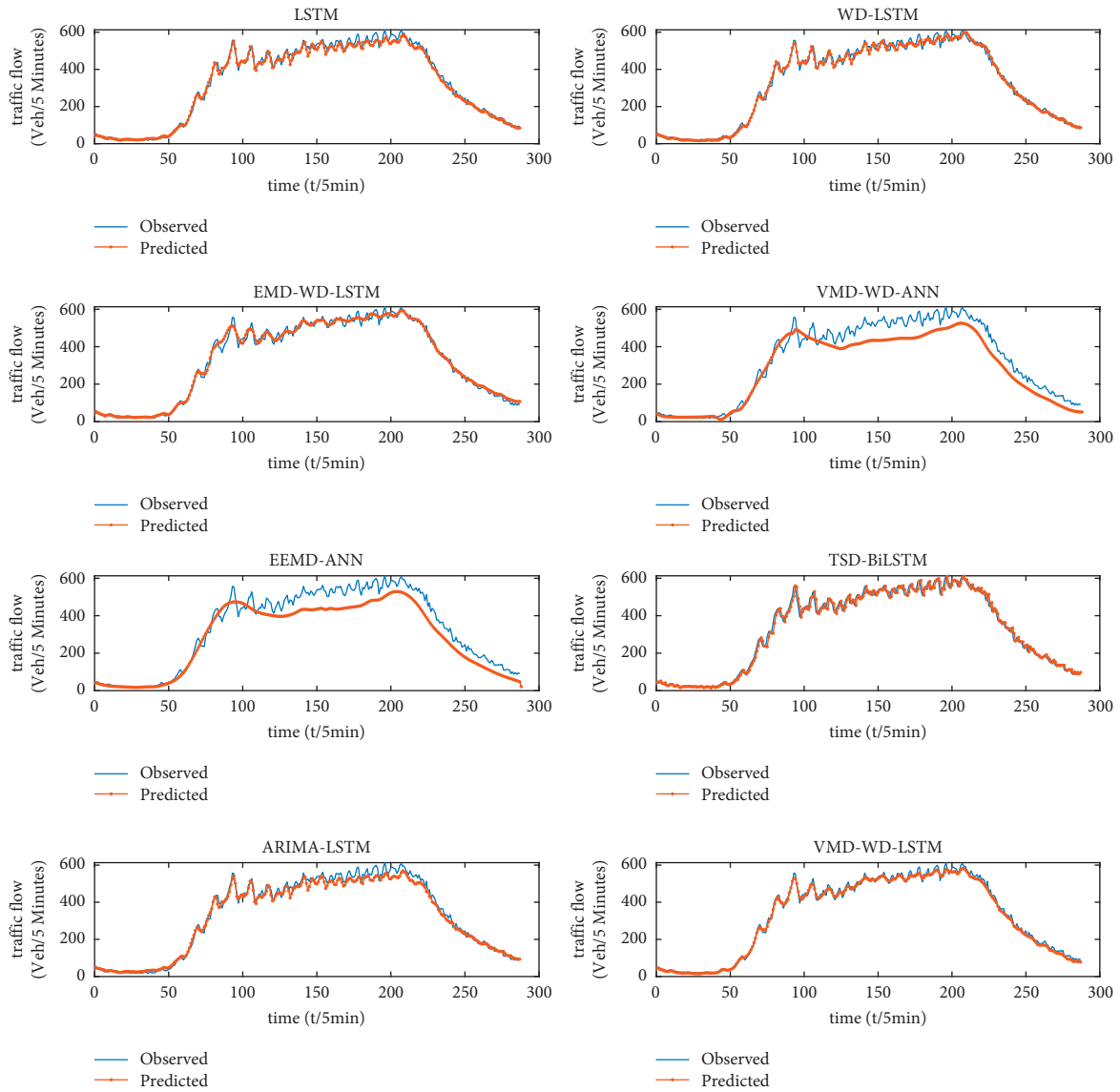
FIGURE 19: Predicted curves of different models using the traffic flow data of detector C.

TABLE 4: The prediction errors of different models using the traffic flow data of detector B.

Prediction models	RMSE (veh/5 min)	MAE (veh/5 min)	MAPE (%)
LSTM	18.711	14.094	6.320
WD-LSTM	17.969	13.538	6.457
EMD-WD-LSTM	13.220	9.817	**5.208**
VMD-WD-ANN	49.357	37.466	13.365
EEMD-ANN	51.022	38.303	13.557
TSD-BiLSTM	18.765	13.664	6.531
ARIMA-LSTM	20.234	15.220	7.173
VMD-WD-LSTM	**10.304**	**7.984**	5.979

Bold indicates the best value of each evaluation index.

TABLE 5: The prediction errors of different models using the traffic flow data of detector C.

Prediction models	RMSE (veh/5 min)	MAE (veh/5 min)	MAPE (%)
LSTM	18.248	13.444	6.137
WD-LSTM	16.939	12.442	5.885
EMD-WD-LSTM	17.157	13.071	6.620
VMD-WD-ANN	60.737	49.949	17.915
EEMD-ANN	61.205	51.208	15.970
TSD-BiLSTM	20.801	15.161	6.616

Table 5: Continued.

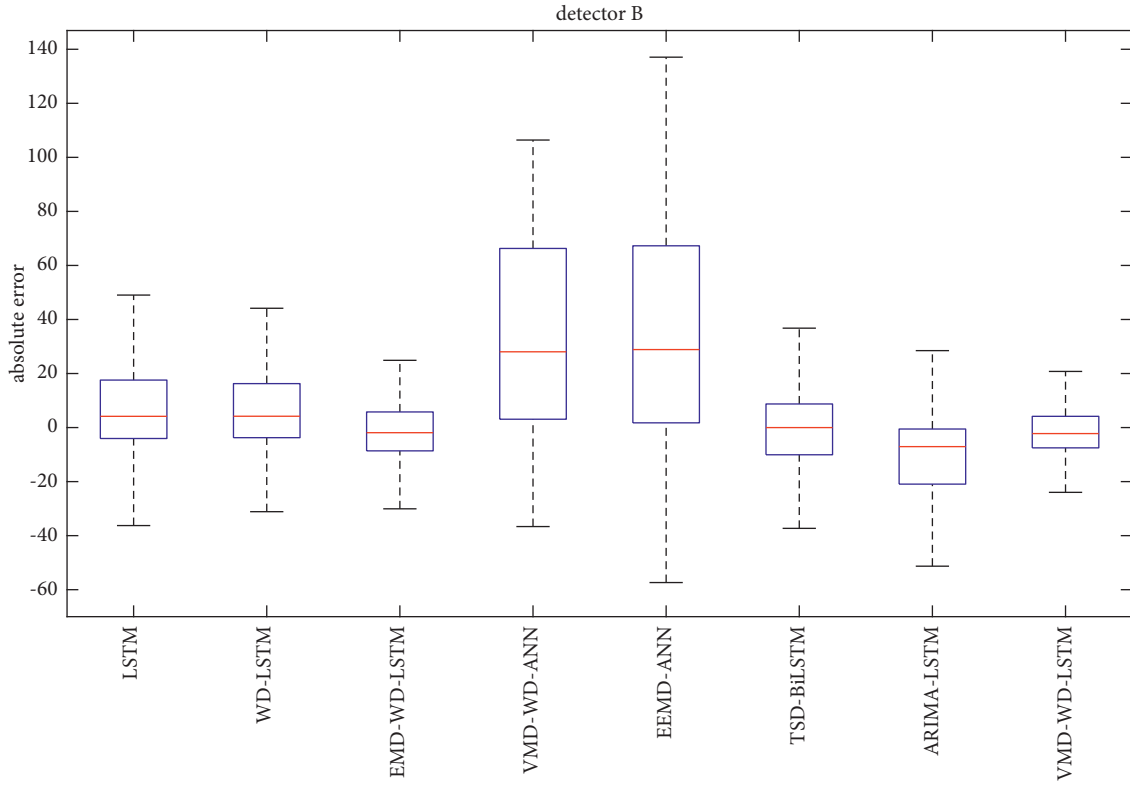| Prediction models	RMSE (veh/5 min)	MAE (veh/5 min)	MAPE (%)
ARIMA-LSTM	22.969	17.244	9.057
VMD-WD-LSTM	**13.980**	**10.965**	**5.268**

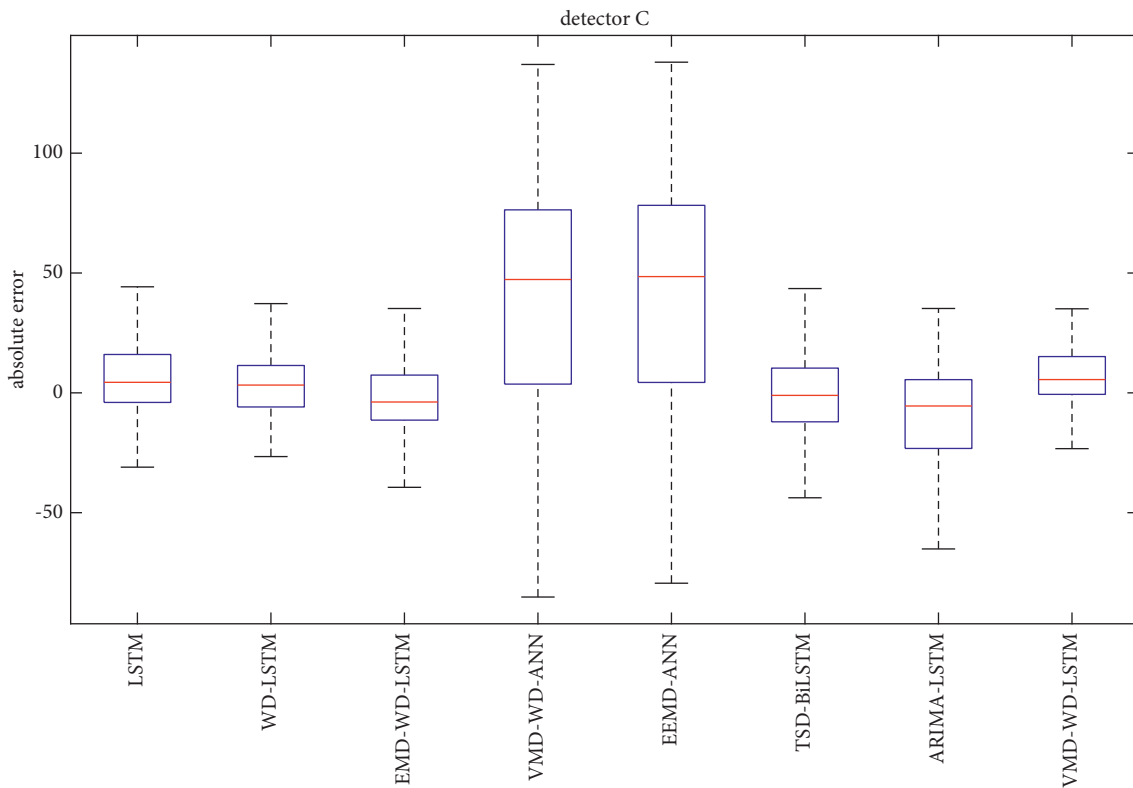Figure 20: Boxplots of absolute error of different models (detector B).

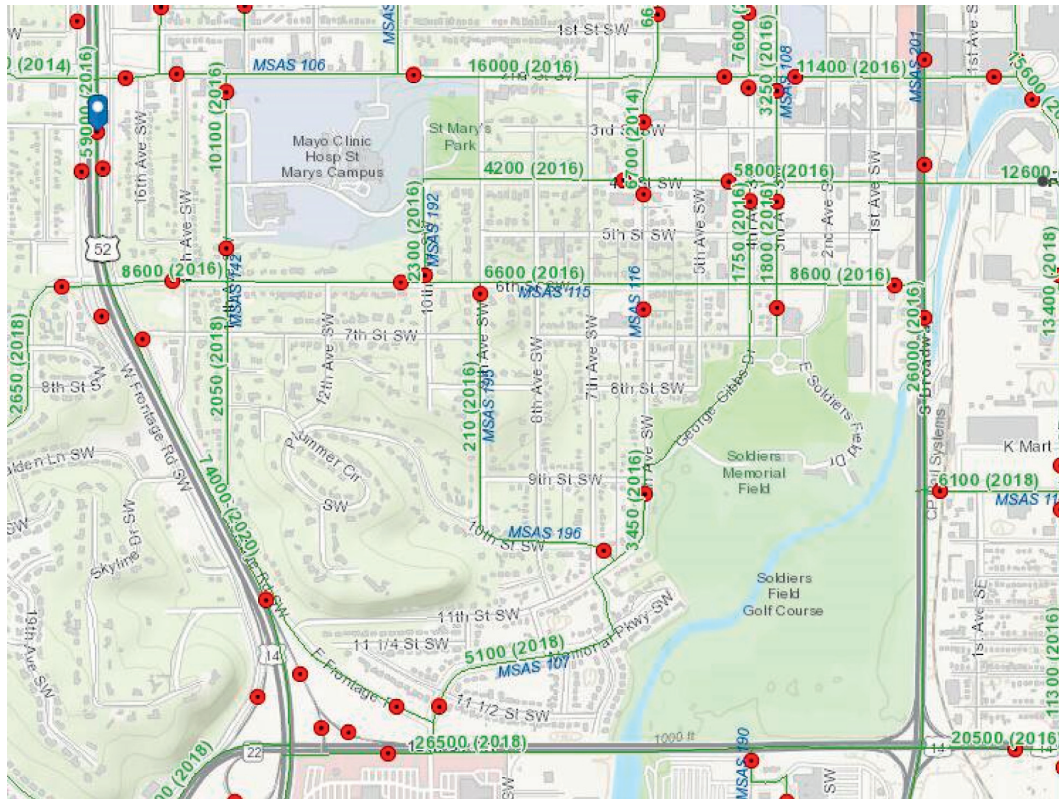Figure 21: Boxplots of absolute error of different models (detector C).

FIGURE 22: The location of the detection point (marked blue).

TABLE 6: Detailed information for detector D.

Detector	Freeway	Detector number	Number of main lanes
D	US-52	3891	3

Finally, in order to further verify the prediction performance of the proposed method in this study in the prediction of traffic flow and show the prediction effect of the VMD-WD-LSTM method on different datasets, this paper selects a section of traffic flow data from the public dataset provided by Minnesota Department of Transportation (Mn/DOT) and Transportation Research Data Lab (TDRL). The traffic flow data are obtained by a loop detector (denoted as detector D) of Minnesota Expressway. The traffic flow time interval of this data sample is also 5 minutes, and a detection point is selected in Rochester, Minnesota. The data collection period is from September 14, 2020, to September 18, 2020. The location of the detection point is shown by the blue marker in Figure 22. Table 6 shows the relevant information of the road section.

For the traffic flow data of detector D, we use the same method as detector A, detector B, and detector C to denoise and predict and then compare the prediction performance of different prediction models by evaluating RMSE, MAE, and MAPE. The prediction results of eight prediction models are shown in Figure 23. It can be seen from the predicted image that the prediction results of the EMD-WD-LSTM method and EMD-WD-LSTM method have obvious fluctuations. This is because the traffic flow of

detection point D is smaller than that of the three detection points in the PeMS dataset, and the uncertainty of traffic flow is enhanced. Any traffic flow change caused by external factors will interfere with the periodic change of traffic flow to a greater extent, which will make the process of identifying noise data more difficult. However, it can be seen from Table 7 that the RMSE, MAE, and MAPE of the VMD-WD-LSTM method are still reduced by 51%, 48%, and 45% compared with the single LSTM model, and it still has better prediction performance than the seven comparison methods. Compared with the data of several other detection points, the RMSE and MAE of the traffic flow prediction results of detection point D are significantly less than those of the other three detection points. This situation is due to the differences in road types and temporal and spatial correlation between different detection points. On the other hand, it shows that the proposed method will get more accurate prediction results when the traffic flow is small. In addition, in the image shown in Figure 24, the absolute error of the VMD-WD-LSTM method is the smallest, which is consistent with the previous analysis results. This shows that the prediction model proposed in this study also has excellent performance and good robustness on other datasets.
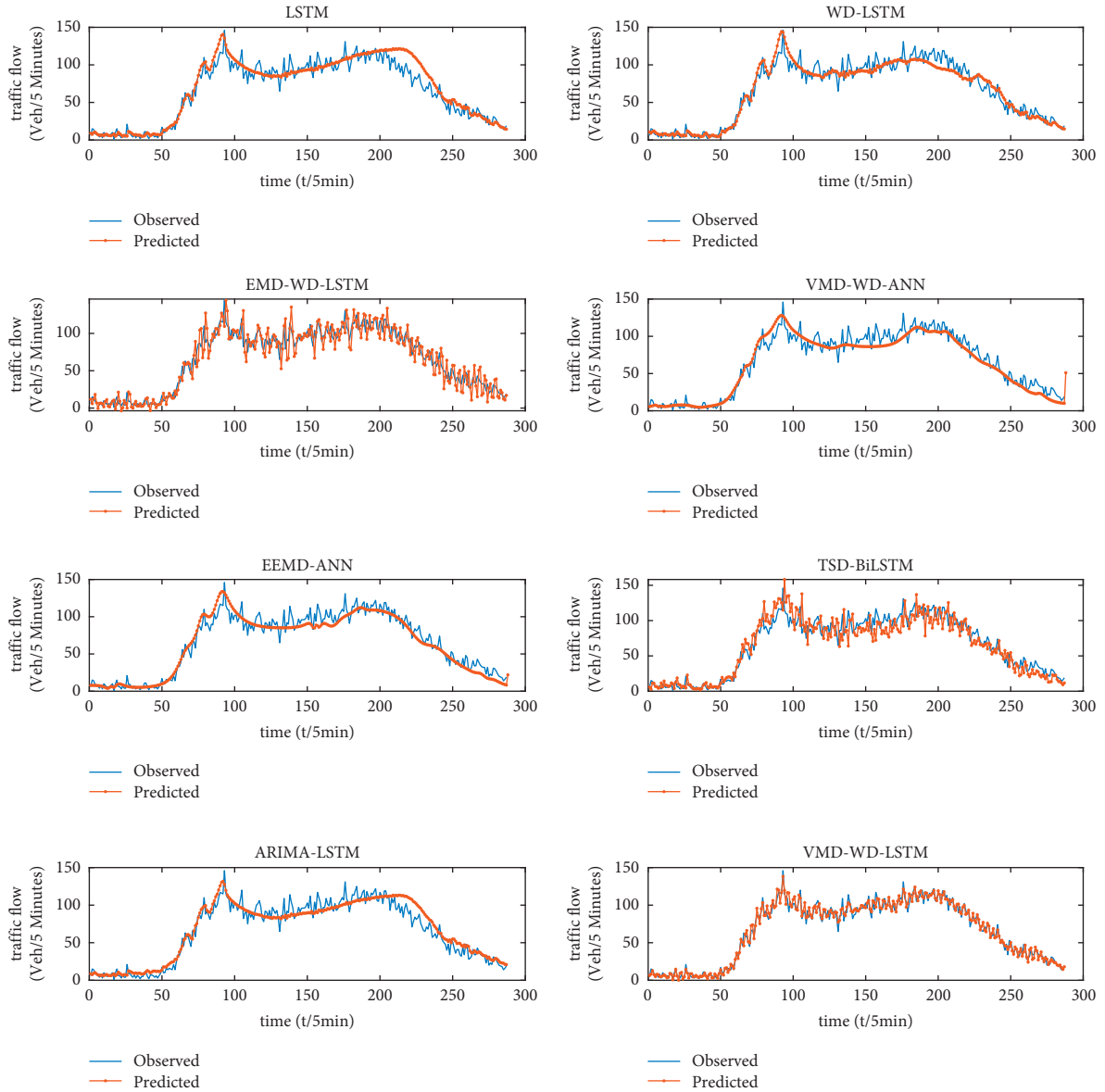
Figure 23: Predicted curves of different models using the traffic flow data of detector D.

Table 7: The prediction errors of different models using the traffic flow data of detector D.

Prediction models	RMSE (veh/5 min)	MAE (veh/5 min)	MAPE (%)
LSTM	11.462	8.290	22.69
WD-LSTM	9.795	7.382	21.316
EMD-WD-LSTM	13.976	10.824	32.833
VMD-WD-ANN	9.909	7.702	20.667
EEMD-ANN	9.688	7.546	20.461
TSD-BiLSTM	14.898	10.929	27.831
ARIMA-LSTM	10.269	7.936	28.417
VMD-WD-LSTM	**5.623**	**4.301**	**12.500**

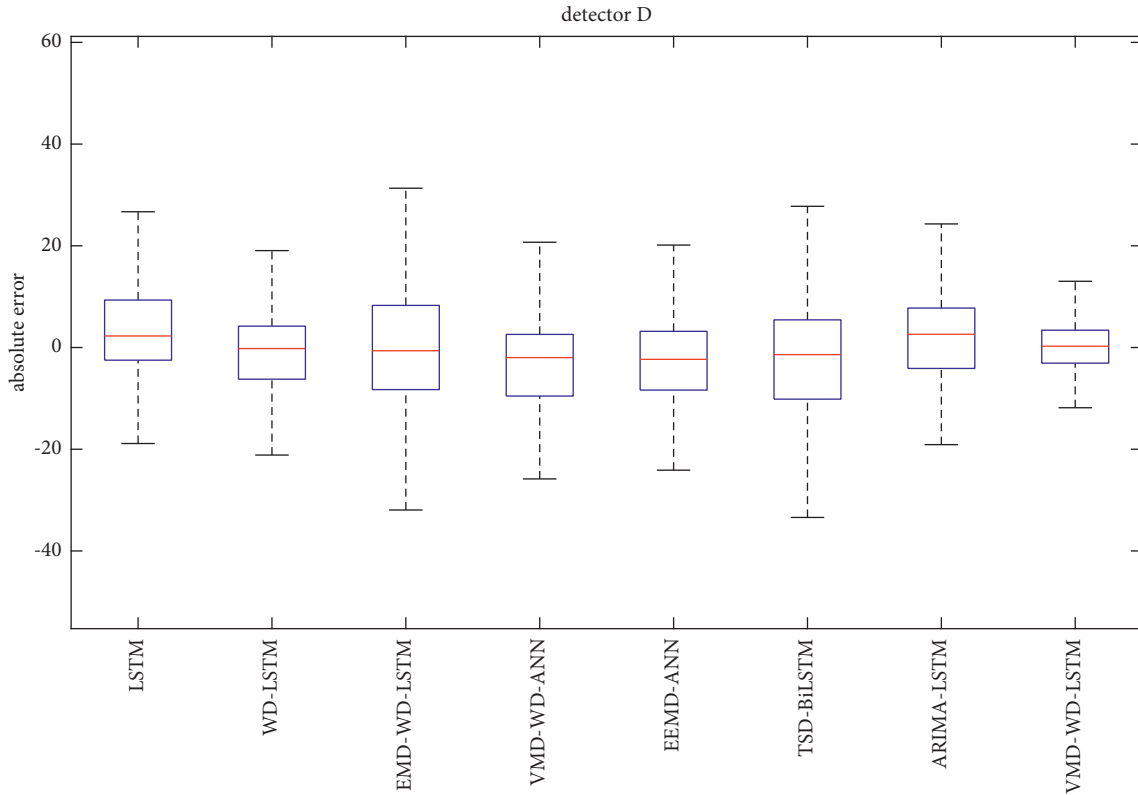Bold indicates the best value of each evaluation index.

FIGURE 24: Boxplots of absolute error of different models (detector D).

4. Conclusions

The randomness and complex periodic characteristics of traffic flow make it difficult to predict traffic flow. To solve this problem, this study proposes a VMD-WD-LSTM prediction method, which includes data denoising, signal decomposition, and data prediction. Specifically, this model first decomposes the original traffic flow data into multiple subsequences by the variational mode decomposition (VMD) method. Since the VMD method can control the number of decompositions, we determine the appropriate number of decompositions by sample entropy. The subsequences obtained by these decompositions can reflect different characteristics of the original signal. The second step is to conduct wavelet threshold denoising for each decomposed subsequence. Compared with the denoising method for original data, denoising multiple IMFs can better reduce the impact of noise on prediction results. Finally, the denoised IMF component is predicted by the LSTM model, and the final prediction results are obtained by combining the predicted values of each component.

In order to evaluate the denoising effect and prediction performance of the VMD-WD-LSTM model, this study first compares the results obtained by direct prediction of the original data with the results predicted by the VMD-WD method after denoising. From the prediction results of the three prediction models of ARIMA, ANN, and LSTM, the denoising of the original signal by the VMD-WD method can improve the prediction performance, and the improvement effect on the LSTM model is the most obvious. In

addition, the performance of the proposed model is also compared with LSTM, WD-LSTM, EMD-WD-LSTM, VMD-WD-ANN, EEMD-ANN, TSD-BiLSTM, and ARIMA-LSTM methods on four detectors in two different open-source datasets. The results show that the VMD-WD denoising method can better reduce noise pollution. On the basis of data denoising with VMD-WD method, the LSTM model can accurately predict the characteristics of traffic flow data and obtain excellent prediction results.

In summary, the VMD-WD-LSTM model proposed in this study can realize the feature decomposition of the original traffic flow data and the prediction of the traffic flow on the working day. Accurate prediction of traffic flow can effectively avoid traffic congestion. In the face of the upcoming congestion, we can make early warnings and take evacuation measures. At present, the method proposed in this paper still has some shortcomings. First, this study only analyzes the traffic flow of working days in terms of data selection and does not analyze the changes in traffic flow at weekends, holidays, and special periods. In addition, different weather conditions, road conditions, and spatial-temporal correlation will affect the prediction results. The method proposed in this study only analyzes the traffic flow changes during the nonholiday period and does not consider the special weather and road conditions. Obviously, it is not feasible to predict the traffic flow affected by other factors by traffic flow in general. Therefore, in the future work, this study will focus on the analysis of traffic flow changes in different scenarios and study the impact of different external factors on traffic flow. Another future research direction is

the analysis and prediction of traffic flow changes in specific periods of the day, especially in the morning or evening when traffic pressure increases. It is also important to predict traffic flow in this period alone.

Data Availability

The data used to support the findings of this study are available from the corresponding author upon request.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

References

[1] 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.

[2] Y. Xie, Y. Zhang, and Z. Ye, "Short-term traffic volume forecasting using kalman filter with discrete wavelet decomposition," *Computer-Aided Civil and Infrastructure Engineering*, vol. 22, no. 5, pp. 326–334, 2007.

[3] Y. Peng and W. Xiang, "Short-term traffic volume prediction using GA-BP based on wavelet denoising and phase space reconstruction," *Physica A: Statistical Mechanics and Its Applications*, vol. 549, 2020.

[4] H. Xiao, H. Sun, B. Ran, and Y. Oh, "Fuzzy-neural network traffic prediction framework with wavelet decomposition," *Transportation Research Record*, vol. 1836, 2003.

[5] J. Tang, X. Chen, Z. Hu, F. Zong, C. Han, and L. Li, "Traffic flow prediction based on combination of support vector machine and data denoising schemes," *Physica A: Statistical Mechanics and Its Applications*, vol. 534, Article ID 120642, 2019.

[6] N. E. Huang, Z. Shen, S. R. Long et al., "The empirical mode decomposition and the Hilbert spectrum for nonlinear and non-stationary time series analysis," *Proceedings of the Royal Society of London. Series A: Mathematical, Physical and Engineering Sciences*, vol. 454, no. 1971, pp. 903–995, 1998.

[7] X. Chen, H. Chen, Y. Yang et al., "Traffic flow prediction by an ensemble framework with data denoising and deep learning model," *Physica A: Statistical Mechanics and Its Applications*, vol. 565, Article ID 125574, 2021.

[8] X. Chen, J. Lu, J. Zhao, Z. Qu, Y. Yang, and J. Xian, "Traffic flow prediction at varied time scales via ensemble empirical mode decomposition and artificial neural network," *Sustainability*, vol. 12, no. 9, Article ID 3678, 2020.

[9] Z. Wu, J. Feng, F. Qiao, and Z.-M. Tan, "Fast multidimensional ensemble empirical mode decomposition for the analysis of big spatio-temporal datasets," *Philosophical Transactions of the Royal Society A*, vol. 374, no. 2065, 2016.

[10] K. Dragomiretskiy and D. Zosso, "Variational mode decomposition," *IEEE Transactions on Signal Processing*, vol. 62, no. 3, pp. 531–544, 2014.

[11] H. Liu, X. Mi, and Y. Li, "Smart multi-step deep learning model for wind speed forecasting based on variational mode decomposition, singular spectrum analysis, LSTM network and ELM," *Energy Conversion and Management*, vol. 159, pp. 54–64, 2018.

[12] G. Zuo, J. Luo, Ni Wang, Y. Lian, and X. He, "Decomposition ensemble model based on variational mode decomposition and long short-term memory for streamflow forecasting," *Journal of Hydrology*, vol. 585, Article ID 124776, 2020.

[13] J. Shi, Y.-B. Leau, K. Li, and H. Chen, "Optimal variational mode decomposition and integrated extreme learning machine for network traffic prediction," *IEEE Access*, vol. 9, pp. 51818–51831, 2021.

[14] M. Van Der Voort, M. Dougherty, S. Watson, and S. Watson, "Combining Kohonen maps with ARIMA time series models to forecast traffic flow," *Transportation Research Part C: Emerging Technologies*, vol. 4, no. 5, pp. 307–318, 1996.

[15] B. L. Smith, B. M. Williams, and R. Keith Oswald, "Comparison of parametric and nonparametric models for traffic flow forecasting," *Transportation Research Part C: Emerging Technologies*, vol. 10, no. 4, pp. 303–321, 2002.

[16] J. Guo, W. Huang, B. M. Williams, and 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.

[17] L. Cai, Z. Zhang, J. Yang, Y. Yu, T. Zhou, and J. Qin, "A noise-immune Kalman filter for short-term traffic flow forecasting," *Physica A: Statistical Mechanics and Its Applications*, vol. 536, Article ID 122601, 2019.

[18] N. Zhang, Y. Zhang, and H. Lu, "Seasonal autoregressive integrated moving average and support vector machine models: prediction of short-term traffic flow on freeways," *Transportation Research Record*, vol. 2215, no. 1, 2011.

[19] Z. Mingheng, Z. Yaobao, H. Ganglong, and C. Gang, "Accurate multisteps traffic flow prediction based on SVM," *Mathematical Problems in Engineering*, vol. 2013, Article ID 418303, 8 pages, 2013.

[20] T. D. Toan and V.-H. Truong, "Support vector machine for short-term traffic flow prediction and improvement of its model training using nearest neighbor approach," *Transportation Research Record: Journal of the Transportation Research Board*, vol. 2675, no. 4, pp. 362–373, 2021.

[21] M. Özuysal, S. P. Çalışkanelli, S. Tanyel, and T. Baran, "Capacity prediction for traffic circles: applicability of ANN," *Proceedings of the Institution of Civil Engineers-Transport*, vol. 162, no. 4, pp. 195–206, 2009.

[22] K. Kumar, M. Parida, and V. Kumar Katiyar, "Short term traffic flow prediction in heterogeneous condition using artificial neural network," *Transport*, vol. 30, no. 4, pp. 397–405, 2015.

[23] Y. Tian and L. Pan, "Predicting short-term traffic flow by long short-term memory recurrent neural network," in *Proceedings of the 2015 IEEE international conference on smart city/SocialCom/SustainCom (SmartCity)*, pp. 153–158, IEEE, Chengdu, China, December 2015.

[24] 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 Transactions on Intelligent Transportation Systems*, vol. 22, no. 5, pp. 2627–2636, 2021.

[25] C. Ma, G. Dai, and J. Zhou, "Short-term traffic flow prediction for urban road sections based on time series analysis and LSTM_BILSTM method," *IEEE Transactions on Intelligent Transportation Systems*, vol. 22, pp. 1–10, 2021.

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

[27] P. Wu, Z. Huang, Y. Pian, L. Xu, J. Li, and K. Chen, "A combined deep learning method with attention-based LSTM model for short-term traffic speed forecasting," *Journal of Advanced Transportation*, vol. 2020, Article ID 8863724, 15 pages, 2020.

[28] L. Li, H. Zhou, H. Liu, C. Zhang, and J. Liu, "A hybrid method coupling empirical mode decomposition and a long short-term memory network to predict missing measured signal data of SHM systems," *Structural Health Monitoring*, vol. 20, no. 4, pp. 1778–1793, 2021.

[29] H. Huang, J. Chen, R. Sun, and S. Wang, "Short-term traffic prediction based on time series decomposition," *Physica A: Statistical Mechanics and Its Applications*, vol. 585, no. 1, Article ID 126441, 2022.

[30] L. Li, X. Qu, J. Zhang, H. Li, and B. Ran, "Travel time prediction for highway network based on the ensemble empirical mode decomposition and random vector functional link network," *Applied Soft Computing*, vol. 73, pp. 921–932, 2018.

[31] S. Lu, Q. Zhang, G. Chen, and D. Seng, "A combined method for short-term traffic flow prediction based on recurrent neural network," *Alexandria Engineering Journal*, vol. 60, no. 1, pp. 87–94, 2021.