Received December 22, 2020, accepted December 29, 2020, date of publication January 11, 2021, date of current version January 21, 2021.

Digital Object Identifier 10.1109/ACCESS.2021.3050836

A Hybrid Deep Learning Framework for Long-Term Traffic Flow Prediction

YIQUN LI[1], SONGJIAN CHAI[1,2], ZHENGWEI MA[1,3], AND GUIBIN WANG[1], (Member, IEEE)
[1]College of Mechatronics and Control Engineering, Shenzhen University, Shenzhen 518060, China
[2]College of Physics and Optoelectronic Engineering, Shenzhen University, Shenzhen 518060, China
[3]College of Urban Transportation and Logistics, Shenzhen Technology University, Shenzhen 518118, China

Corresponding author: Guibin Wang (wanggb@szu.edu.cn)

This work was supported in part by the National Natural Science Foundation of China under Grant 51507103 and Grant 52002251, in part by the Foundations of Shenzhen Science and Technology Committee under Grant JCYJ20170817100412438 and Grant JCYJ20190808141019317, in part by the Natural Science Foundation of Guangdong Province under Grant 2018A030310523, in part by the Special Innovation Project in Higher Education of Guangdong Province under Grant 2018KTSCX351, and in part by the Teaching Reform Project in Higher Education of Guangdong Province (Study on the Practical Teaching System of Multi-integration of School-enterprise and Industry-education for Automotive Service Engineering Major Under the Background of New Engineering).

ABSTRACT An accurate and reliable traffic flow prediction is of great significance, especially the long-term traffic flow prediction e.g., 24 hours, which can help the traffic decision-makers formulate the future traffic management strategy. However, the long-term traffic flow prediction imposes great challenges for decision-makers due to the nonlinear and chaotic feature of traffic flow. Therefore, in this paper, we proposed a hybrid deep learning model based on wavelet decomposition, convolutional neural network-long and short-term memory neural network (CNN-LSTM), called W-CNN-LSTM, to prediction next-day traffic flow. The wavelet decomposition technology is used to decompose the original traffic flow data into high-frequency data and low-frequency data for the improvement of predictive accuracy. The decomposed sequences are fed into a CNN-LSTM deep learning model, where the long-term temporal features of traffic flow can be well captured and learned. The numerical experiment is carried out against five benchmarks based on England traffic flow dataset; the results show that the proposed hybrid approach can achieve superior forecasting skill over the benchmarks.

INDEX TERMS Traffic flow prediction, long-term prediction, wavelet decomposition, CNN-LSTM model.

I. INTRODUCTION

A. MOTIVATION

The rapid development of urbanization brings great benefits to people, but also brings about some inconvenience, which urges the researchers to solve these challenges in their industries. For instance, in order to deal with the economic issues and the technical issues in energy, [1] propose a novel transactive energy trading framework. In order to implement voltage control, a distributed online voltage control algorithm is proposed in [2]. Reference [3] propose two distributed voltage control algorithms to overcome these challenges in multiphase unbalanced distribution networks. For transportation, traffic problems caused by the rapid increase of the number of motor vehicles, such as traffic congestion, traffic accidents and traffic delays, impose huge challenges and pressure to urban transportation management system. Making a reliable traffic management plan based on forecasting traffic flow is the effective way to deal with traffic problems [4]. The traffic flow represents the number of vehicles passing through the road on each time interval. However, the number of vehicles on the road increases with the increase of people's travel demands, and the road network becomes more complex with the rapid development of the city, which makes the traffic flow more prominent in complexity and randomness. In addition, now, transportation filed has entered the era of big data. All these make it more difficult for the traditional traffic flow prediction model to fit the traffic data and to make a best forecast [5]. Moreover, most of the traditional traffic flow prediction models only focus on single time-step prediction, although of scientific significance, then cannot satisfy the practical application of multi-time step traffic flow prediction. An accurate long-term traffic flow prediction is of great practical significance to the transportation management

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

system and congestion analysis early warning system of intelligent transportation system [6].

B. LITERATURE REVIEW

In the past decade, with the development of intelligent transportation, traffic flow prediction has become a research hotspot in the intelligent transportation. Many experts and scholars have committed to the traffic flow prediction research and proposed a large number of prediction methods.

Accurate traffic condition prediction is the premise to realize active traffic control and dynamic traffic distribution effectively. Reference [7] propose localized space-time autoregressive for traffic flow forecasting on urban road network and formulated a new parameters estimation method to reduce computational complexity. A hybrid short-term traffic flow prediction model based on the multifractal characteristics of traffic flow time series is proposed in [8]. Reference [9] establish a simple and effective mixed traffic flow prediction model, which combines Auto-regressive Integrated Moving Average Model (ARIMA) and genetic programming (GP) model to capture the different aspects of the underlying traffic flow patterns. The grey correlation prediction of traffic data with panel data characteristics is studied in [10], which adopt the ARMIA model for prediction. For this, [11] simulate the response of the expressway system to the change of traffic state and proposed a spatiotemporal traffic flow prediction model without off-line parameter calibration. Reference [12] in order to improve the accuracy of traffic flow prediction, a combination prediction model based on GM, ARIMA and GRNN is proposed, and then used to establish a combination model of road traffic flow based on fixed weight. Reference [13] develops a novel Bayesian combination method (BCM) to improve the performance of traditional BCM for short-term traffic flow forecasting. Three single predictors, ARMIA, Kalman Filter (KF) and back propagation neural network (BPNN), were designed and incorporated into the BCM to make full use of the advantages of each method.

However, in recent years, traffic data has experienced significant growth and entered the era of traffic big data. Due to the nonlinearity and randomness of traffic flow, researchers turned their attentions to using the deep learning technologies for traffic flow prediction. By taking into account the spatial and temporal correlations, a novel deep-learning-based traffic flow prediction method is proposed in [14]. Reference [15] studies the correlation between weather parameters and traffic flow and proposed a novel overall framework to improve traffic flow prediction. Furthermore, [16] proposes a traffic prediction method based on the deep belief networks model structure and multitask regression is applied to predict the traffic flow of single output and multitask output. A deep code learning technique is proposed in [17], and applied to the Macao intelligent system. Recently, the recurrent neural network (RNN) and the variant of RNN (like LSTM [18]) are widely regarded as an appropriate method to capture the temporal and spatial information in a variety of fields, such as text classification [19], energy [20], or transportation [21].

However, most of the existing forecasting studies in traffic field focus on the short-term traffic flow prediction, while the long-term traffic flow prediction is particularly meaningful since it involves more information about the future for traffic system planning an effective plan. Therefore, we propose a hybrid deep learning algorithm, in which the CNN-LSTM is used to predict the traffic flow for next 24 hours, and the wavelet decomposition method is used to decompose the information of the original traffic volume data. The experimental results show that our proposed model can achieve accurate prediction performance, which is better than the traditional model (e.g., ARIMA), Neural networks (e.g., NLP), and deep learning model (e.g., LSTM, CNN and CNN-LSTM).

The main contributions of our work are briefly summarized as follows:

1). We proposed a hybrid deep learning model to long-term multi-variable traffic flow modeling tasks, and applied wavelet decomposition to improve the prediction performance.

2). Based on the publicly available dataset, we conduct extensive experiments to evaluate the performance of W-CNN-LSTM, experimental results show the superiority of the proposed model against the benchmark.

The rest of the paper is organized as follows: Section II introduces the proposed hybrid forecasting framework, W-CNN-LSTM, and describes the assessment indicators of performance for our proposed model and the benchmarks; Section III describes the experimental procession of W-CNN-LSTM; Section IV is case studies; Section V presents the concluding remarks.

II. THE PROPOSED HYBRID DEEP LEARNING FRAMEWORK

In this section, a hybrid day-ahead traffic flow forecasting deep learning framework, which comprises the wavelet decomposition and CNN-LSTM model, was formulated and proposed.

A. WAVELET TRANSFORM

In signal processing, Fourier transform can reveal the internal relationship between time-domain data and frequency-domain data [22]. It is a popular classical signal decomposition method in the traditional stationary signal analysis and processing. The wavelet transform [23] is developed based on the short-time Fourier transform, which overcomes the shortcoming of the Fourier transform, such as the window problem, to decompose the original time-domain data through the filter. The wavelet transform filter can be divided into low pass filter and high pass filter, and through these filters, the original time-domain data can be decomposed into a set of low-frequency data CAn and several sets of high-frequency data CD1 to CDn. The common wavelet transform has two forms, continuous wavelet transform (CWT) and discrete wavelet transform (DWT).

CWT is a complex signal transform process used to decompose a continuous time function into several wavelet, as follow as:

$$X_w(a,b) = \int_{-\infty}^{+\infty} f(t)\psi_{ab}^{\wedge}(t)dt \tag{1}$$

where $a > 0$ is scalability factor, b is translation factor. Both factors control the scale of the wavelet transform. $f(t)$ is the original function. $\psi_{ab}^{\wedge}(t)$ is conjugate function of $\psi_{ab}(t)$. $\psi_{ab}(t)$ is Mother Wavelet, express as:

$$\psi_{ab}(t) = \frac{1}{\sqrt{a}}\psi\left(\frac{t-a}{b}\right) \tag{2}$$

In practical applications, DWT is more common than continuous wavelet transform, is a discrete input and discrete output, and is no simple and clear formula to express the relationship between input and output, can only be expressed by hierarchical architecture. The expression is shown in Fig.1. $X[n]$ denote the input signal of length n, $L[n]$ and $H[n]$ are the loss pass filter and high pass filter, respectively, $\downarrow Q$ is the downsampling filter.

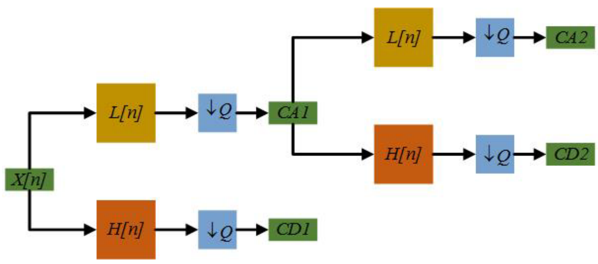


FIGURE 1. Algorithm structure of DWT.

After decomposition, the reconstructed signal is by adding up all the low-frequency and high-frequency signals, express as:

$$f(t) = CA_n l(\psi(t)) + \sum_{i=1}^{n} CD_i h(\psi(t)) \tag{3}$$

where $l(\psi_{ab}(t))$ denote low pass filter, $h(\psi_{ab}(t))$ denote high pass filter.

B. CNN-LSTM MODULE

The CNN-LSTM framework for forecasting traffic flow consists of a series connection of CNN and LSTM. The CNN network [24] applied in this method only comprises the convolutional layer and ReLU activation layer. CNN pass convolution operation to learning these complex traffic flow features such as temporal information and the traffic flow eigenvalue of last days. Convolution operation can reduce the number of neuron parameters and make the hybrid model deeper. The CNN structure of proposed model is shown in Fig.2. If $X = x_1, x_2, \ldots, x_n$ is the traffic information input vector, where n denotes the 24 hours unit per window. The operation of convolution layer and activation layer as follows,

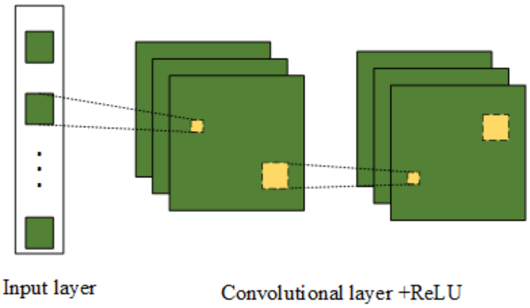


FIGURE 2. Algorithm structure of CNN.

$$x_j^l = f(g(\sum_{i \in \mathrm{I}_m} x_i^{l-1} * w_{ij}^l + b_j^l)) \tag{4}$$

where I_m denotes the number of feature map, $w^{l_{ij}}$ and b_j^l are weights of the kernel and bias for i-th input feature map and j-th output feature map corresponding to l-th convolutional layer, respectively. $g(\cdot)$ is a user-defined activation function, $f(\cdot)$ is the ReLU activation as follows as expression (5), * represent a convolutional operation.

$$f(x) = \max(0, x) \tag{5}$$

RNN model is widely applied in time series learning. However, the traditional RNN is not very suitable for the long time series prediction due to the gradient problem. Thus, a variant of the RNN, i.e., LSTM network is proposed to mitigate gradient explosion or disappearance through gating mechanisms and cell memory. As such, the long temporal dependence can be well learned using this configuration. The structure of an LSTM for information flows is sketched in Fig.3. The update process of LSTM model at timestep t can be described as follows:

$$f_t = \delta(W_f x_t + U_f h_{t-1} + b_f) \tag{6}$$
$$i_t = \delta(W_i x_t + U_i h_{t-1} + b_i) \tag{7}$$
$$o_t = \delta(W_o x_t + U_o h_{t-1} + b_o) \tag{8}$$
$$c_t^* = \tanh(W_{c*} x_t + U_{c*} h_{t-1} + b_{c*}) \tag{9}$$
$$c_t = f_t \circ c_{t-1} + i_t \circ c_t^* \tag{10}$$
$$h_t = o_t \tanh(c_t) \tag{11}$$

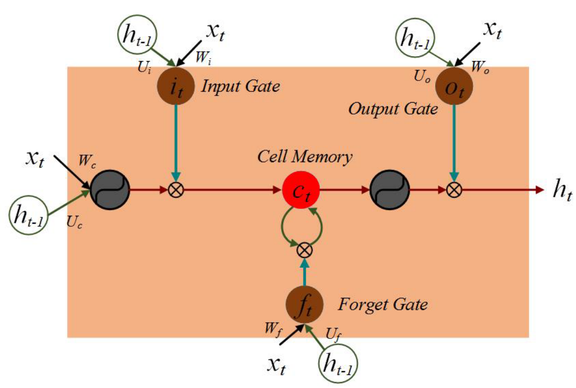


FIGURE 3. Algorithm structure of LSTM.

where $\circ$ denotes the Hadamard product, i_t, f_t and o_t are the output of input gate, forget gate and output gate. c_t^* is the new state of t time-step cell memory, c_t is the final state of t time-step cell memory that is participate in next time-step cell memory operation and h_t is the final output of the memory unit. W_i, W_f, W_o, W_{c^*}, U_i, U_f, U_o and U_{c^*} are coefficient matrixes of these gates; b is bias, $\delta(\cdot)$ is sigmoid function as formulated in (12), $\tanh(\cdot)$ is tanh function as formulated in (13).

$$\delta(x) = \frac{1}{1+e^{-x}} \tag{12}$$

$$\tanh(x) = \frac{1-e^{-2x}}{1+e^{-2x}} \tag{13}$$

Then, the full connection (FC) layer activates the input information h_t by (14) and yield the final traffic flow prediction. The entire model is depicted in Fig.4.

$$y_t = g(h_t) \tag{14}$$

where y_t is the final traffic flow prediction output of t-th time, $g(\cdot)$ denotes the FC layer activation function.

This work aims to predict the hourly traffic flow for the next day by using a set of explanatory variables in the previous days, including the traffic flow information and calendar information. The mapping relationship between the point estimates and the inputs can be formulated under the deep learning framework as,

$$f_{t+1:t+p} = cnnlstm(H_{:t}, s_t \left| F_{t+1:t+p} \right.) \tag{15}$$

where $f_{t+1:t+p}$ denotes the multi-step traffic volume output, p is the timestep, $H_{:t}$ denotes the variable of historical data, including the traffic volume and temporal information. $F_{t+1:t+p}$ is the known characteristic variable of traffic flow in the future p time stamps.

C. EVALUATION CRITERIA

An excellent point forecasting model is required to accurately capture the future traffic flow trends. To verify the performance of the proposed traffic flow prediction model, we applied three evaluation indexes, including root mean square error (RMSE), mean absolute error (MAE), and goodness of fit (R-Square). The expression of these evaluation indexes are as follows:

$$RMSE = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(y_i - y_i{}^*)^2} \tag{16}$$

$$MAE = \frac{1}{N}\sum_{i=1}^{N}\left|y_i - y_i{}^*\right| \tag{17}$$

$$R^2 = 1 - \frac{\sum_{i=1}^{N}(y_i - y_i{}^*)^2}{\sum_{i=1}^{N}(y_i - \hat{y}_i)^2} \tag{18}$$

where N represents the number of traffic flow, y_i is the real traffic flow data, and $y^*{}_i$ is the predicted traffic flow after wavelet reconstruction by (3). $\hat{y}_i$ is the mean value of the real traffic flow data. Naturally, the smaller the RMSE and MAE index values, the more accurate the model prediction. R^2 infinitely close to 1 denotes that the predictions are as close as the real values. In addition, we applied mean square error (MSE) as the loss function for model training.

$$MSE = \frac{1}{N}\sum_{i=1}^{N}(y_i - y_i{}^*)^2 \tag{19}$$

III. EXPERIMENTAL SETUP

A. DATA DESRCRIPTION AND PREPROCESSING

The traffic flow data used in our work is collected from highways England organization [25], which records the traffic volume data with 15 minutes resolution from 1 Jul 2018 to 28 Jan 2020 on the M6 main northbound lanes between the J4 and J4A in England. However, a large number of missing data is observed between 1 Oct 2018 to 27 Oct 2018, thus we only retained the samples measured from 1 Jul 2018 to 30 Sep 2018 and that from 28 Oct 2018 to 28 Jan 2020. The former period will be used for testing and the latter is for training.

In our work, we predict the day-ahead hourly traffic volume based on the traffic information in the prior days. In order to fulfill our prediction requirements, the 15-min traffic flow data is converted into the hourly data beforehand. Apart from the traffic volume data, we also include the calendar variables, such as year, month, day, hour and holiday. The traffic data information for this long-term traffic flow prediction study is listed in TABLE 1.

TABLE 1. Traffic flow and feature data information.

Variables	Abbr.	Description
Traffic data	RTV	Raw traffic volume (veh/h)
Temporal data	MOY	Month of the year (1-12)
	DOY	Day of the year (1-365)
	DOM	Day of the month (1-31)
	DOW	Day of the week (0-6)
	HOD	Hour of the day (0-23)
	HD	Holiday:1; holiday eve or the day after the holiday:0.5; other:0

Then, we normalize the original traffic data and the temporal data according to formula (20). Next, decompose the normalized raw traffic volume by the DWT, which yield a low-frequently data and several group high-frequently data, same as shown in Fig.5. Finally, we divided last datasets into training set, verification set according to the condition of 8:2. The training set is applied to train different hyper-parameters model and update the weights and bias of neuron cell. And then verification set verify the skill of these hyper-parameters models, which is through the formula (3) reconstruct the prediction values and inverse normalization, and the prediction values is calculate the critical indexes with the real observed traffic flow. Finally, the reconstructed prediction values of test set are inverse normalization to calculate the evaluation indexes by (16)-(18) as the model predictive performance

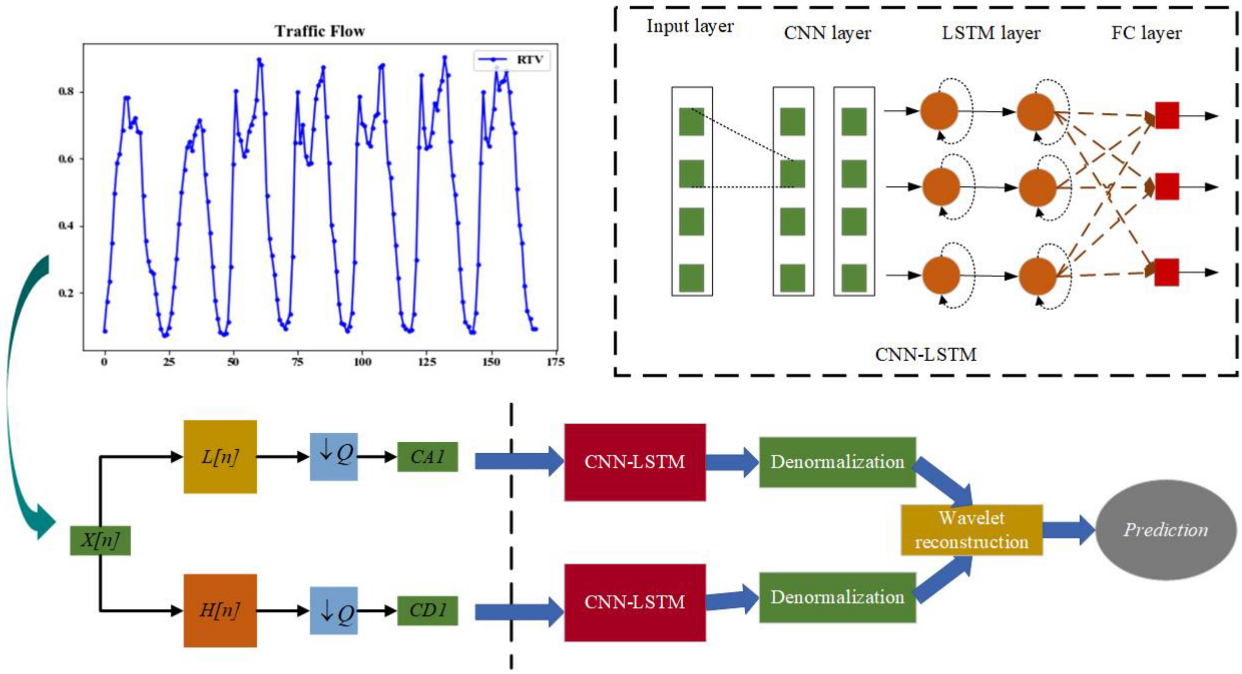


FIGURE 4. The overall proposed W-CNN-LSTM structure. The normalized raw traffic flow is put into the DWT and processed by the filters. Then, these wavelet datasets are fed to the CNN-LSTM, respectively. The high and low frequency data predicted by CNN-LSTM are denormalization and reconstruction to output the next-day traffic flow prediction values.

TABLE 2. The determination results of four groups of wavelet parameters.

Model	RMSE	MAE	R^2
W-CNN-LSTM (2- order)	466.268	303.979	0.894
W-CNN-LSTM (3- order)	436.717	295.569	0.908
W-CNN-LSTM (4- order)	475.001	299.408	0.891
W-CNN-LSTM (5- order)	479.871	339.727	0.888

TABLE 3. The experiment results of different historical input scale.

Scale	RMSE	MAE	R^2
Last 24hour	436.717	295.569	0.908
Last 48hour	449.699	288.145	0.903
Last 72hour	420.117	268.359	0.914
Last 96hour	465.900	298.997	0.901

evaluation.

$$x^{norm} = \frac{x - x^{\min}}{x^{\max} - x^{\min}} \tag{20}$$

B. PARAMETERS DETERMINATION OF WAVELET DECOMPOSITION

In order to make full advantages of DWT, and improve the performance of hybrid model, it is necessary to select the order of wavelet decomposition beforehand. To this end, we tested four different decomposition orders to find the optimal order of wavelet decomposition, based on the last day as the historical input. The training dataset was applied to feed the four groups of models, and then the experimental

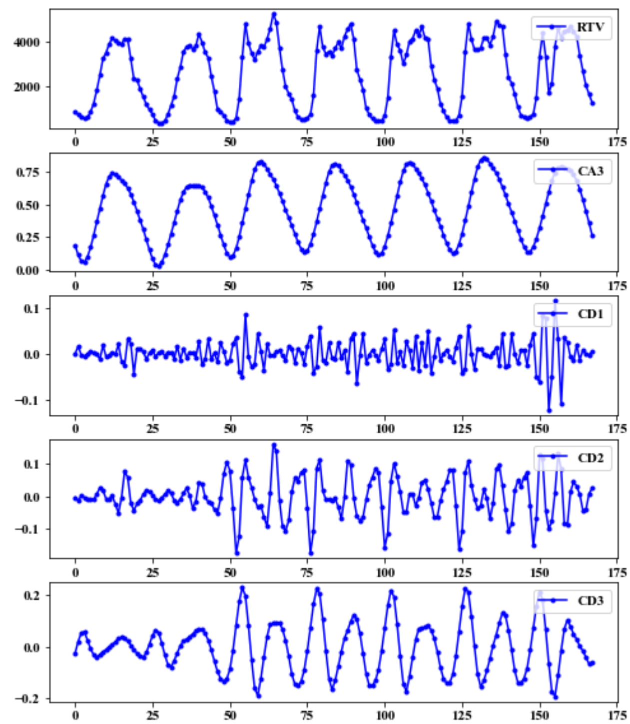


FIGURE 5. 168 hours of traffic flow raw data and 3 order wavelet decomposition (the low frequency and high frequency data).

results were produced by the test set, the final model results are shown in TABLE 2. It can be seen that the prediction performance of the 3-order W-CNN-LSTM is more excellent,

TABLE 4. The W-CNN-LSTM and benchmarks setup parameters.

Model	W-CNN-LSTM				LSTM	CNN	MLP	CNN-LSTM	ARIMA
	CA3	CD1	CD2	CD3	RTV	RTV	RTV	RTV	RTV
Learning rate	Lr=0.002	Lr=0.002	Lr=0.002	Lr=0.002	Lr=.082	Lr=0.002	Lr=.082	Lr=0.002	(2,1,0)
Optimizer	Nadam	Nadam	Nadam	Nadam	Adadelta	Nadam	Adadelta	Nadam	
Layer number	2 CNN 2 LSTM	2 CNN 2 LSTM	2 CNN 2 LSTM	2 CNN 2 LSTM	3	3	-	2 CNN 2 LSTM	
CNN activation	Relu	Sigmoid	Simoid	Relu	-	Tanh	-	Relu	
CNN neuron	30	40	40	30	-	80	-	60	
LSTM activation	Tanh	Tanh	Tanh	Tanh	Tanh	-	-	Tanh	
LSTM neuron	60	40	40	60	80	-	-	40	
Dense layer	3	3	3	3	3	1	4	3	
Dense activation	Tanh	Tanh	Tanh	Tanh	Relu	tamh	Tanh	Tanh	
Dense neuron	50	20	50	50	90	50	80	50	
Batch-size	50	50	50	50	50	50	50	60	
Epoch	180	180	180	180	180	180	180	180	
L2	0.001	0.001	0.001	0.001	0.002	0.002	0.002	0.001	

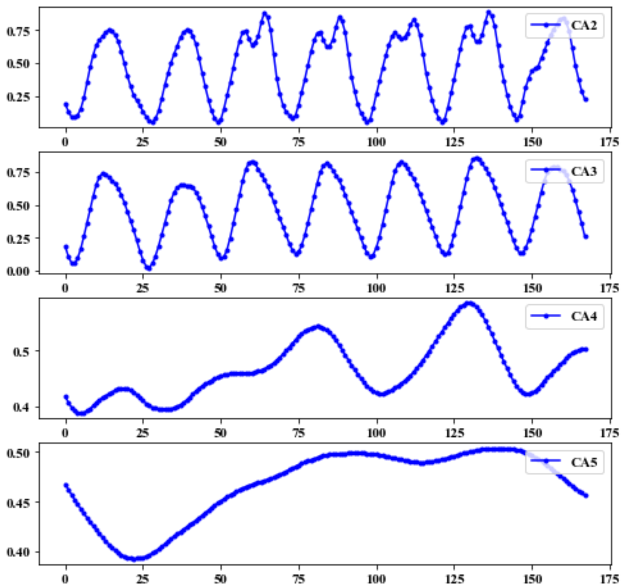


FIGURE 6. The low pass data of different order wavelet decomposition.

and the forecasting performance of the 2, 4 and 5 order decomposition is unsatisfactory, especially the 5-order W-CNN-LTSM.

This conclusion can be also obtained from the vibration trend of the low-frequency arrays of the four groups of wavelet decomposition, show in Fig.6. It is obviously that the fluctuation trends of CA3 is relatively stable, which can explain the reason that the accuracy of order 3 are better than that of order 2, 4 and 5. This indicates that the stationarity of the low-frequency filter data can determine whether the wavelet decomposition plays an auxiliary role well.

LSTM is an important part framework of CNN-LSTM and yield the traffic flow vector features according to the traffic historical information fed. The scale of historical data has an impact on LSTM exploring the traffic flow eigenvalue. Therefore, according to the results of TABLE 2, 3-order W-CNN-LSTM is satisfactory in long-term traffic forecasting, we apply the 3-order model framework to set up another three experiments to find the scale of historical data that can enhance model accuracy in the 24 hours ahead traffic prediction. The final experimental results are shown in TABLE 3. It can be seen again that 3-order hybrid model can achieve superior forecasting skill. Where R^2 indexes are all larger than 0.90, and MAE values are less than 300. It can be seen taking the traffic information in last 3 days as the historical scale of LSTM can activate the LSTM skill well. Therefore, we adopted the 3-order hybrid model with last 3 days as the historical input as our prediction framework and compared with the benchmark model.

IV. CASE STUDIES

In this section, we verified the effectiveness of the proposed W-CNN-LSTM model against the benchmarks, including traditional statistical model, ARMIA, and four advanced models: LSTM, Multilayer Perceptron (MLP), CNN, as well as CNN-LSTM. ARIMA is the most commonly used benchmark for single point forecasts of traffic prediction. the LSTM forecast method is a widely used deep learning model in traffic prediction and is known to be easy to outperform for short look-ahead time. In the experiment, these deep learning/machine learning models needs to find the best hyper-parameters, including batch size, number of neurons, layer number of neural networks, and activation function of neural network. For ARIMA, auto_arima is used to search the optimal parameters automatically. After comprehensive experiment, we obtained the final configuration results of these models through the evaluation of the verification set, as shown in TABLE 4.

To be fair, the traffic volumes in the last three days are taken as the historical information of the next for MLP, CNN and LSTM. The final experiment results are shown in TABLE 5.

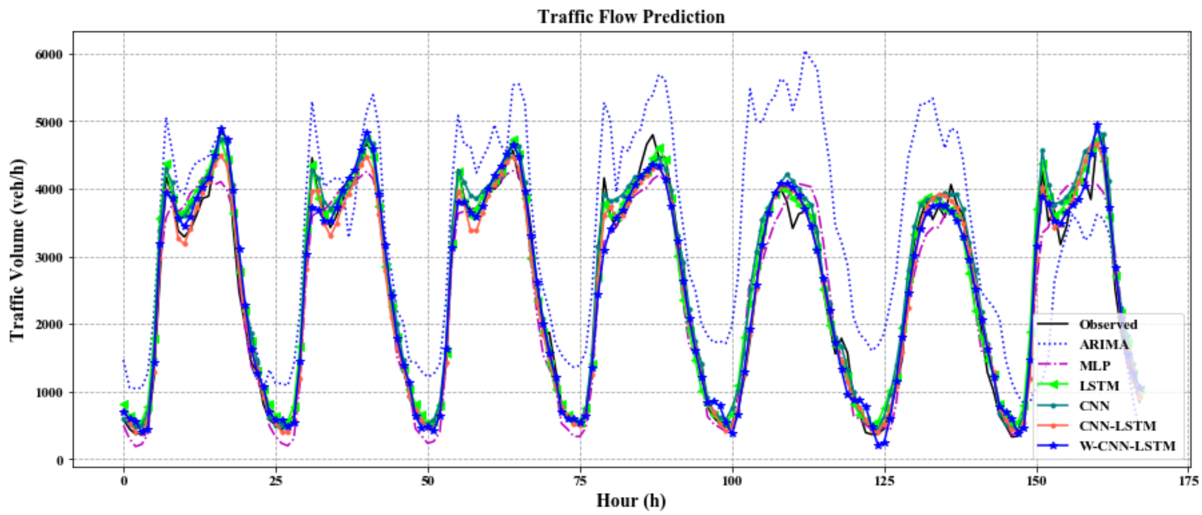


FIGURE 7. 24hour-ahead traffic flow forecasting results for various prediction models in one week.

TABLE 5. Experiment results of W-CNN-LSTM and benchmarks.

Model	RMSE	MAE	R^2
W-CNN-LSTM	420.117	268.359	0.914
ARIMA	1372.429	1084.228	0.099
LSTM	495.040	314.862	0.881
MLP	460.265	312.976	0.898
CNN-LSTM	448.780	279.991	0.902
CNN	492.338	303.318	0.882

As we can see, the traffic flow prediction performance of the traditional statistical model does not satisfy the long-term traffic prediction performance, the R^2 is less than 0.1 and the RMSE or MAE is greater than 1000. Then, MLP, CNN and LSTM have roughly the same performance with relatively high R^2 index. Compared with experimental results of ARIMA, machine learning model can output satisfactory prediction values in the long-term traffic flow prediction. Further, from the RMSE and MAE, it is obviously that CNN-LSTM is more accurate than LSTM and CNN since combining the advantages of both, which CNN layers can explore the features between several variables affecting traffic flow and LSTM can explores the long-term dependency. This result indicates that the CNN-LSTM model is more suitable to model the long-term traffic flow patterns than the original RNN variants model. Furthermore, the wavelet transform can enhance the predictive performance of CNN-LSTM model. The hybrid deep learning mode on the basis of wavelet decomposition and W-CNN-LSTM is more accurate than the CNN-LSTM. Fig.7 shows the traffic flow predictions for various models. It can be seen that deep learning model can well predict the trend of traffic flow.

V. CONCLUSION

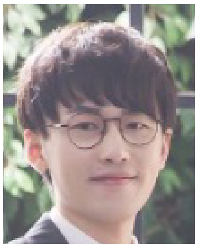
Long-term traffic flow is a new milestone for traffic flow prediction and a new field worth exploring. In order to maximize the performance of CNN-LSTM in the day-ahead traffic flow prediction, the original traffic flow data were firstly decomposed through wavelet transform, and each group of decomposed data was used to train an independent CNN-LSTM model. The predicted traffic flow data from decomposed data and independent CNN-LSTMs were reconstructed as the final predictions. Through verifying on the real-life traffic flow data measured in the England highway, the proposed W-CNN-LSTM model shows superior predictive performance than ARMIA, LSTM, CNN, MLP as well as its counterpart without wavelet decomposition process.

REFERENCES

[1] J. Li, C. Zhang, Z. Xu, J. Wang, J. Zhao, and Y.-J.-A. Zhang, "Distributed transactive energy trading framework in distribution networks," *IEEE Trans. Power Syst.*, vol. 33, no. 6, pp. 7215–7227, Nov. 2018.
[2] J. Li, Z. Xu, J. Zhao, and C. Zhang, "Distributed online voltage control in active distribution networks considering PV curtailment," *IEEE Trans. Ind. Informat.*, vol. 15, no. 10, pp. 5519–5530, Oct. 2019.
[3] J. Li, C. Liu, M. E. Khodayar, M.-H. Wang, Z. Xu, B. Zhou, and C. Li, "Distributed online VAR control for unbalanced distribution networks with photovoltaic generation," *IEEE Trans. Smart Grid*, vol. 11, no. 6, pp. 4760–4772, Nov. 2020.
[4] E. I. Vlahogianni, M. G. Karlaftis, and J. C. Golias, "Short-term traffic forecasting: Where we are and where we're going," *Transp. Res. C, Emerg. Technol.*, vol. 43, pp. 3–19, Jun. 2014.
[5] M. Längkvist, L. Karlsson, and A. Loutfi, "A review of unsupervised feature learning and deep learning for time-series modeling," *Pattern Recognit. Lett.*, vol. 42, pp. 11–24, Jun. 2014.
[6] B. Yu, X. Song, F. Guan, Z. Yang, and B. Yao, "K-nearest neighbor model for multiple-time-step prediction of short-term traffic condition," *J. Transp. Eng.*, vol. 142, no. 6, Jun. 2016, Art. no. 04016018.
[7] M. Lippi, M. Bertini, and P. Frasconi, "Short-term traffic flow forecasting: An experimental comparison of time-series analysis and supervised learning," *IEEE Trans. Intell. Transp. Syst.*, vol. 14, no. 2, pp. 871–882, Jun. 2013.
[8] H. Zhang, X. Wang, J. Cao, M. Tang, and Y. Guo, "A hybrid short-term traffic flow forecasting model based on time series multifractal characteristics," *Appl. Intell.*, vol. 48, no. 8, pp. 2429–2440, 2018.
[9] C. Xu, Z. Li, and W. Wang, "Short-term traffic flow prediction using a methodology based on autoregressive integrated moving average and genetic programming," *Transport*, vol. 31, no. 3, pp. 343–358, Sep. 2016.

[10] J. Yang, X. Xiao, S. Mao, C. Rao, and J. Wen, "Grey coupled prediction model for traffic flow with panel data characteristics," *Entropy*, vol. 18, no. 12, p. 454, Dec. 2016.
[11] C. Dong, Z. Xiong, C. Shao, and H. Zhang, "A spatial–temporal-based state space approach for freeway network traffic flow modelling and prediction," *Transportmetrica A, Transp. Sci.*, vol. 11, no. 7, pp. 547–560, Aug. 2015.
[12] Z. Yu, T. Sun, H. Sun, and F. Yang, "Research on combinational forecast models for the traffic flow," *Math. Problems Eng.*, vol. 2015, pp. 1–10, Oct. 2015.
[13] J. Wang, W. Deng, and Y. Guo, "New Bayesian combination method for short-term traffic flow forecasting," *Transp. Res. C, Emerg. Technol.*, vol. 43, pp. 79–94, Jun. 2014.
[14] Y. Lv, Y. Duan, W. Kang, Z. Li, and F.-Y. Wang, "Traffic flow prediction with big data: A deep learning approach," *IEEE Trans. Intell. Transp. Syst.*, vol. 16, no. 2, pp. 865–873, Apr. 2015.
[15] A. Koesdwiady, R. Soua, and F. Karray, "Improving traffic flow prediction with weather information in connected cars: A deep learning approach," *IEEE Trans. Veh. Technol.*, vol. 65, no. 12, pp. 9508–9517, Dec. 2016.
[16] W. Huang, G. Song, H. Hong, and K. Xie, "Deep architecture for traffic flow prediction: Deep belief networks with multitask learning," *IEEE Trans. Intell. Transp. Syst.*, vol. 15, no. 5, pp. 2191–2201, Oct. 2014.
[17] D. Li, L. Deng, Z. Cai, B. Franks, and X. Yao, "Intelligent transportation system in Macao based on deep self-coding learning," *IEEE Trans. Ind. Informat.*, vol. 14, no. 7, pp. 3253–3260, Jul. 2018.
[18] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural Comput.*, vol. 9, no. 8, pp. 1735–1780, 1997.
[19] C. Zhou, C. Sun, Z. Liu, and F. C. M. Lau, "A C-LSTM neural network for text classification," 2015, *arXiv:1511.08630*. [Online]. Available: http://arxiv.org/abs/1511.08630
[20] S. Chai, Z. Xu, Y. Jia, and W. K. Wong, "A robust spatiotemporal forecasting framework for photovoltaic generation," *IEEE Trans. Smart Grid*, vol. 11, no. 6, pp. 5370–5382, Nov. 2020.
[21] R. Fu, Z. Zhang, and L. Li, "Using LSTM and GRU neural network methods for traffic flow prediction," in *Proc. 31st Youth Academic Annu. Conf. Chin. Assoc. Autom. (YAC)*, Nov. 2016, pp. 324–328.
[22] R. N. Bracewell and R. N. Bracewell, *The Fourier Transform and Its Applications*. New York, NY, USA: McGraw-Hill, 1986.
[23] D. Zhang, "Wavelet transform," in *Fundamentals of Image Data Mining*. New York, NY, USA: Springer, 2019, pp. 35–44.
[24] S. Albawi, T. A. Mohammed, and S. Al-Zawi, "Understanding of a convolutional neural network," in *Proc. Int. Conf. Eng. Technol. (ICET)*, Aug. 2017, pp. 1–6.
[25] *Highways Agency Network Traffic Flow Data*. Accessed: 2020. [Online]. Available: http://tris.highwaysengland.co.uk/

YIQUN LI received the B.Eng. degree in traffic engineering from Guangzhou University, China, in 2018. He is currently pursuing the M.E. degree in transportation engineering from Shenzhen University, Shenzhen, China. His main research interests include traffic information forecasting and artificial intelligence application in transportation and renewable energy.



SONGJIAN CHAI received the M.S. and Ph.D. degrees in electrical engineering from The Hong Kong Polytechnic University, Hong Kong, in 2012 and 2018, respectively. He is currently a Postdoctoral Research Fellow with Shenzhen University, Shenzhen, China. His research interests include variable renewable generation forecasting, electricity price forecasting, power system uncertainty analysis, and artificial intelligence application in power engineering.

ZHENGWEI MA received the B.E. degree in transportation from the North China University of Water Resources and Electric Power, China, in 2010, and the Ph.D. degree in vehicle engineering from the South China University of Technology, China, in 2015. Since 2015, he has been an Assistant Professor with the College of Mechatronics and Control Engineering, Shenzhen University. Since 2017, he has been an Assistant Professor and an Associate Professor with the College of Urban Transportation and Logistics, Shenzhen Technology University. His research interests include the design, control and optimization in energy efficiency, safety, and man–machine of smart and sustainable mobility, especially for electric vehicles.

GUIBIN WANG (Member, IEEE) received the B.E. degree in electrical engineering from Shandong University, Jinan, China, in 2009, and the Ph.D. degree in electrical engineering from Zhejiang University, Hangzhou, China, in 2014. From 2011 to 2014, he was also a Research Assistant with the Department of Electrical Engineering, The Hong Kong Polytechnic University, Hong Kong. He is currently an Associate Research Professor with Shenzhen University, Shenzhen, China. His main research interests include electric vehicles and renewable energy.