The Air Quality Prediction Based on a Convolutional LSTM Network

Canyang Guo[1], Wenzhong Guo[1], Chi-Hua Chen[1], Xin Wang[2], and Genggeng Liu[1(✉)]

[1] College of Mathematics and Computer Science,
Fuzhou University, Fuzhou, China
`liugenggeng@fzu.edu.cn`
[2] College of Intelligence and Computing, Tianjin University, Tianjin, China

Abstract. In recent years, the rapid development of industrial technology has been accompanied by serious environmental pollution. In the face of numerous environmental pollution problems, particulate matter (PM2.5) which has received special attention is rich in a large amount of toxic and harmful substances. Furthermore, PM2.5 has a long residence time in the atmosphere and a long transport distance, so analyzing PM2.5 distributions is an important issue for air quality prediction. Therefore, this paper proposes a method based on convolutional neural networks (CNN) and long short-term memory (LSTM) networks to analyze the spatial-temporal characteristics of PM2.5 distributions for predicting air quality in multiple cities. In experiments, the records of environmental factors in China were collected and analyzed, and three accuracy metrics (i.e., mean absolute error (MAE), root mean square error (RMSE), and mean absolute percentage error (MAPE)) were used to evaluate the performance of the proposed method in this paper. For the evaluation of the proposed method, the performance of the proposed method was compared with other machine learning methods. The practical experimental results show that the MAE, RMSE, and MAPE of the proposed method are lower than other machine learning methods. The main contribution of this paper is to propose a deep multilayer neural network that combines the advantages of CNN and LSTM for accurately predicting air quality in multiple cities.

Keywords: Air quality prediction · Convolutional neural network · Long short term memory · PM2.5

1 Introduction

In recent years, with the development of industrial development, the problem of environmental pollution has become more and more serious, which has attracted significant attention. There are many factors that cause environmental pollution, such as sulfur dioxide, nitrogen oxides, fine particles (PM2.5) and so on. PM2.5 is the main cause of smog among them [1]. PM2.5 can stay in the atmosphere for a long time, and it also can enter the body by breathing, accumulating in the trachea or lungs and affecting the health of the body [2]. PM2.5 is larger than viruses and smaller than bacteria. It is easy to carry toxic substances into the human body [3]. For the

environment and human health, the threat from PM2.5 is enormous. Therefore, the prediction and control of PM2.5 concentration are quite important issues.

This paper proposes a deep multilayer neural network model combining convolutional neural network (CNN) and long short-term memory (LSTM) to predict PM2.5 concentration. The model is able to predict the future PM2.5 concentration data based on the past PM2.5 concentration data. This study collects the environmental data from January 2015 to December 2017 in nine cities (i.e., Ningde City, Nanping City, Fuzhou City, Sanming City, Putian City, Quanzhou City, Longyan City, Zhangzhou City, and Xiamen City) in the Fujian Province of the People's Republic of China as a training set, and the environmental data from January to October 2018 as a testing set. For the evaluation of the proposed method, mean absolute error (MAE), root mean square error (RMSE) and mean absolute percentage error (MAPE) are used as accuracy metrics. Experimental results show that the proposed method is superior to other machine learning methods.

This paper is organized as follows. Section 2 presents a literature review on air quality prediction. Section 3 presents the prediction methods based on deep learning techniques. Section 4 describes the data processing and gives the practical experimental results and discussions. Section 5 summarizes the contributions of this study and discusses the future work.

2   Literature Reviews

In 2013, the World Health Organization's International Agency for Research on Cancer (IARC) published a report stating that PM2.5 is carcinogenic to humans and is considered a universal and a major environmental carcinogen [4]. Therefore, the prediction and control of PM2.5 are particularly important issues for air quality maintenance and urban development. At present, the air quality prediction methods are mainly classified into two categories: (1) the mechanism models based on the atmospheric chemical modes are called deterministic models [5, 6]; (2) the statistical models based on machine learning algorithms are called machine learning models [6, 7].

Gu et al. [8] designed a new picture-based predictor of PM2.5 concentration (PPPC) which employs the pictures acquired using mobile phones or cameras to make a real-time estimation of PM2.5 concentration. Although this method can estimate the PM2.5 concentration more accurately, it can just evaluate the current PM2.5 concentration and cannot predict the future PM2.5 concentration. Mahajan et al. designed a PM2.5 concentration prediction model that combined a neural network based hybrid model and clustering techniques like grid-based clustering and wavelet-based clustering [9]. The main focus is to achieve high accuracy of prediction with reduced computation time. A hybrid model was applied to do a grid-based prediction system for clustering the monitoring stations based on the geographical distance [9]. In 2014, Elangasinghedeng et al. [10] proposed the complex event-sequence analyses of PM10 and PM2.5 in coastal areas by using artificial neural network models and k-means clustering method. The study presented a new approach based on artificial neural network models, and the k-means clustering method was used to analyze the relationships between the bivariates of concentration–wind and speed–wind direction for

extracting source performance signals from the time series of ambient PM2.5 and PM10 concentrations [10]. Li et al. [11] proposed a LSTM network for air pollutant concentration prediction, and Tsai et al. [12] showed the way of air pollution prediction based on RNN and LSTM. The method collected PM2.5 data from 66 stations in Taiwan from 2012 to 2016 and establishes LSTM and RNN models for precise prediction of air quality. Yu et al. [13] predicted the concentration of PM2.5 through the Eta-Community Multiscale Air Quality (Eta-CMAQ) model. The method is based on the chemical composition of PM2.5 and has certain references, but the cost of chemical analysis is too expensive. Verma et al. [14] proposed the use of a bi-directional LSTM model to predict air pollutant severity levels ahead of time. The models are robust and have shown superiority over an artificial neural network model in predicting PM2.5 severity levels for multiple stations in New Delhi City [14].

Because of the small particle size of PM2.5, it stays in the atmosphere for a long time and the transport distance is long. Therefore, PM2.5 concentrations have very close relationships with time and space. The proposed method based on a CNN and a LSTM network to perfectly extract the spatial-temporal characteristics of PM2.5 distributions for air quality prediction.

3 Prediction Methods Based on Deep Learning Techniques

The concepts and processes of CNNs, LSTM networks, and convolutional long-short-term memory (ConvLSTM) networks are presented in the following subsections.

3.1 Convolutional Neural Networks

Deep neural networks have achieved remarkable performance at the cost of a large number of parameters and high computational complexity [15]. A convolutional neural network is a feedforward neural network that contains convolutional computation and has a deep structure. The difference between CNN and the fully connected neural network is the weight sharing. CNN [16, 17] has two advantages: (1) the number of weights is reduced, and the amount of training is greatly reduced; (2) spatio features can be effectively extracted. The network model can process multi-dimension data. In this paper, the input convolutional layer data is a 5×5 two-dimensional matrix. As shown in Fig. 1, the variables x_1 to x_{25} are inputs, and the variables w_1 to w_4 are convolution kernels which function to filter data and extract features. The variables h_1 to h_{16} are feature maps obtained after convolution.
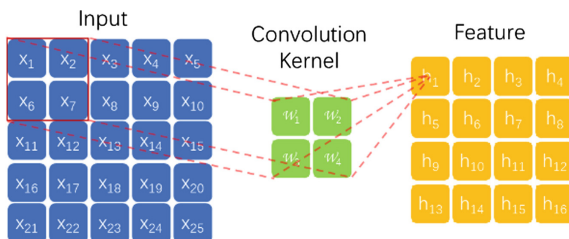

Fig. 1. A two-dimensional convolutional operation

3.2   Long Short-Term Memory Networks

A recurrent neural network (RNN) [18] is an artificial neural network that has a tree-like hierarchical structure, and the nodes of RNN recursively input information in the order in which they are connected. A LSTM [19, 20] network, a special RNN, differs from RNN in learning long-term dependencies. The repeating module in a conventional RNN contains only a single layer (shown in Fig. 2(a)), and the repeating module in a LSTM network contains four interacting neural network layers.
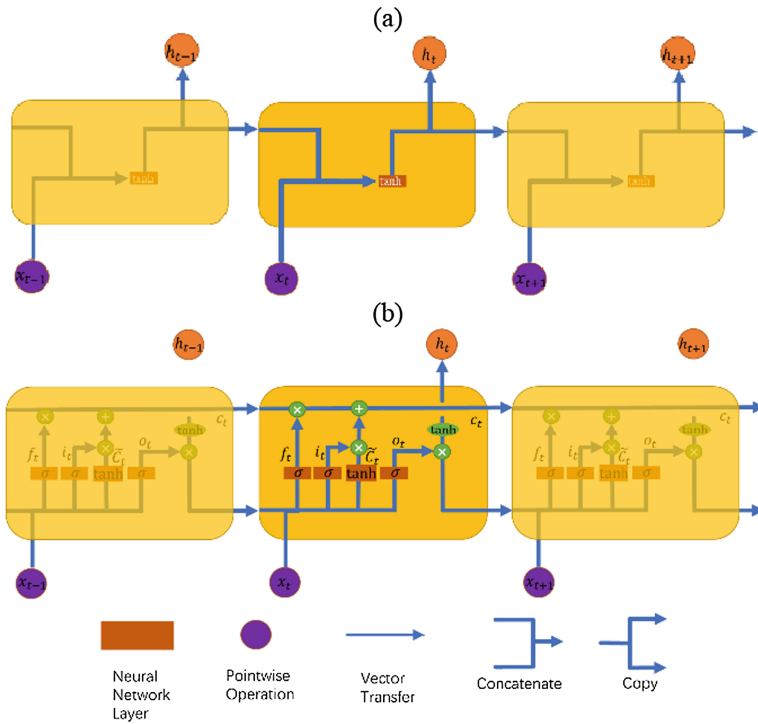

Fig. 2. RNN and LSTM networks

The LSTM network can remove or add information to the cell state and manage it by the gate structure. The LSTM network includes forgetting gates, input gates and output gates. The function σ in the module represents a sigmoid function, and the formula is as shown in Eq. (1). The sigmoid layer outputs a number between 0 and 1, which represents how much each component should pass the threshold. The value of "1" means that all ingredients pass, and the value of "0" means that no ingredients are allowed to pass.

$$S(t) = \frac{1}{1 + e^{-t}} \tag{1}$$

3.2.1 The Forgetting Gate

In Fig. 3, the forgetting gate decides which information to discard. The formula is shown in Eq. (2).

$$f_t = \sigma(W_{xf}x_t + W_{hf}h_{t-1} + W_{cf} \circ c_{t-1} + b_f) \tag{2}$$

Where W is the weight matrix and b is the deviation vector matrix. Both W and b need to learn during the training process. Where $\circ$ is the Hadamard product.
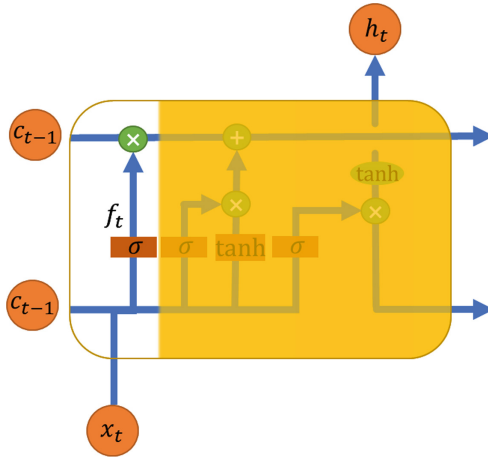

Fig. 3. The forgetting gate in a LSTM network

3.2.2 The Output Gate

In Fig. 4, the input gate determines which information to remember. The formulas are shown in Eqs. (3), (4) and (5).

$$i_t = \sigma(W_{xi}x_t + W_{hi}h_{t-1} + W_{ci} \circ c_{t-1} + b_i) \tag{3}$$

$$\tilde{C}_t = \tanh(W_{xc}x_t + W_{hc}h_{t-1} + b_c) \tag{4}$$

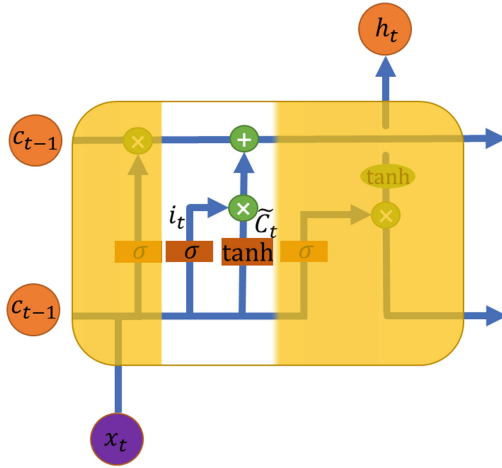$$C_t = f_t \circ C_{t-1} + i_t \circ \tilde{C}_t \tag{5}$$

Fig. 4. The output gate in a LSTM network

3.2.3 The Input Gate

In Fig. 5, the input gate decides which information to update. The formulas are shown in Eqs. (6) and (7).

$$o_t = \sigma(W_{xo}x_t + W_{ho}h_{t-1} + W_{co} \circ c_t + b_o) \tag{6}$$
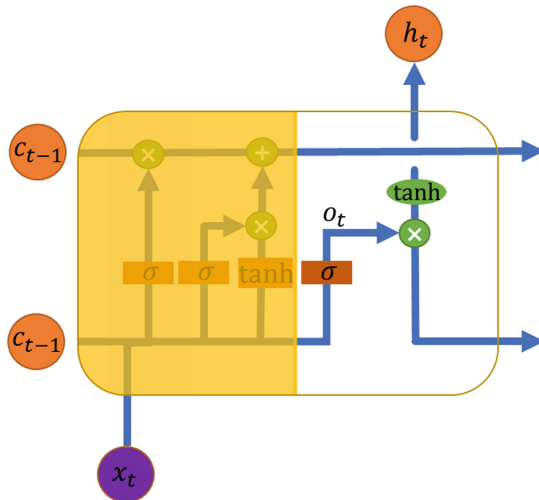
$$h_t = o_t \circ \tanh(c_t) \tag{7}$$

Fig. 5. The input gate in a LSTM network

3.3   Convolutional Long Short-Term Memory Networks

The ConvLSTM [21] network not only has the timing modeling capabilities of a LSTM network, but also extracts spatio features like a CNN. As shown in Fig. 6, the ConvLSTM network differs from the normal LSTM network in that the internal LSTM is internally calculated by a similar feedforward neural network and can be called FC-LSTM [21]. A ConvLSTM network uses convolutional calculations instead of fully connected calculations.

The derivation formulas have also changed, and the new derivations are shown in Eqs. (8), (9), (10), (11), (12) and (13).

$$f_t = \sigma(W_{xf} * X_t + W_{hf} * H_{t-1} + W_{cf} \circ C_{t-1} + b_f) \tag{8}$$

$$i_t = \sigma(W_{xi} * X_t + W_{hi} * H_{t-1} + W_{ci} \circ C_{t-1} + b_i) \tag{9}$$

$$\tilde{C}_t = \tanh(W_{xc} * X_t + W_{hc} * H_{t-1} + b_c) \tag{10}$$

$$C_t = f_t \circ C_{t-1} + i_t \circ \tilde{C}_t \tag{11}$$

$$o_t = \sigma(W_{xo} * X_t + W_{ho}H_{t-1} + W_{co} \circ C_t + b_o) \tag{12}$$
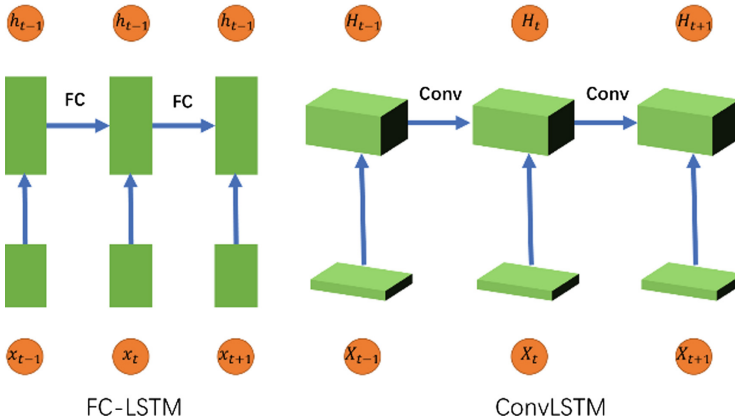
$$H_t = o_t \circ \tanh(c_t) \tag{13}$$

Fig. 6. The differences between a FC-LSTM network and a ConvLSTM network

4   Practical Experimental Environments and Results

This section illustrates the selected features for air quality prediction in Subsect. 4.1 and discusses the practical experimental results in Subsect. 4.2.

4.1 Practical Experimental Environments

For training and testing the air quality prediction methods, this study collected the environmental data in nine cities in the Fujian Province of the People's Republic of China from January 2015 to October 2018. The environmental factors include 7 dimensions which are air quality index (AQI), PM2.5, PM10, SO2, NO2, CO and O3; the AQI reflects the degree of air pollution. Seven environmental factors (i.e., AQI, PM2.5, PM10, SO2, NO2, CO and O3) at the t-th timestamp are elected as the inputs of neural networks, and the parameter of PM2.5 at the $(t + 1)$-th timestamp is elected as the output of neural networks. The mean squared error loss function is adopted for optimizing neural networks. In experiments, the environmental data from January 2015 to December 2017 is used as a training dataset, and the environmental data from January 2018 to October 2018 is used as a testing dataset.

For data pre-processing, if there are abnormal values or missing values in a record, the record will be deleted [22]. For data normalization, the data is processed by min-max normalization method [23] and represented by a number between 0 and 1. The number of records is N, and the value of the i-th record (x) can be normalized by Eq. (14).

$$z_i = \frac{x_i - \min_{1 \le n \le N}(x_n)}{\min_{1 \le n \le N}(x_n) - \min_{1 \le n \le N}(x_n)} \tag{14}$$

4.2 Practical Experimental Results and Discussions

For the evaluation of the proposed ConvLSTM method, multi-layer perception (MLP) neural networks [24–26], CNNs, LSTM networks are implemented and used to predict the air quality in the selected cities in Fujian Province. In order to compare the performance of each prediction method comprehensively and objectively, MAE, RMSE and MAPE were used as accuracy metrics. The value of the n-th actual data is defined as o_n, and the value of the n-th predicted data is defined as p_n. These three accuracy metrics can be estimated by Eqs. (15), (16) and (17), respectively. The practical experimental results based on these three accuracy metrics are shown in Tables 1, 2 and 3.

$$MAE = \frac{1}{N} \sum_{n=1}^{N} |o_n - p_n| \tag{15}$$

$$RMSE = \sqrt{\frac{\sum_{n=1}^{N} (o_n - p_n)^2}{N}} \tag{16}$$

$$MAPE = \frac{1}{N} \sum_{n=1}^{N} \frac{|o_n - p_n|}{o_n} \tag{17}$$

Table 1. The MAEs of each prediction method for each city

CITY	MLP	CNN	LSTM	CONVLSTM
Ningde	7.2652	7.2915	7.9911	**6.7214**
Nanping	6.0939	6.0738	6.4391	**5.3384**
Fuzhou	7.0248	7.1274	7.3394	**6.4475**
Sanming	6.2867	6.1960	6.3652	**5.4708**
Putian	8.4391	8.3612	8.4000	**8.1412**
Quanzhou	7.2165	7.4500	**6.8798**	6.9295
Longyan	7.0222	6.9505	6.7233	**6.3074**
Zhangzhou	7.4520	7.7403	7.6557	**6.7840**
Xiamen	6.3989	6.6250	6.2187	**5.9813**
Average	7.0221	7.0906	7.1125	**6.4579**

Table 2. The RMSEs of each prediction method for each city

City	MLP	CNN	LSTM	ConvLSTM
Ningde	9.7771	9.8717	10.7989	**9.5848**
Nanping	10.8600	10.9898	10.7856	**10.0890**
Fuzhou	9.3981	9.5205	9.6814	**8.8822**
Sanming	9.3716	9.2888	9.2222	**8.6397**
Putian	13.2202	12.9197	12.9752	**12.4293**
Quanzhou	10.7698	10.5632	10.3454	**10.0866**
Longyan	13.2247	12.9709	12.7514	**12.4693**
Zhangzhou	11.7122	11.5926	12.0828	**10.9533**
Xiamen	8.9357	8.9459	8.5964	**8.1712**
Average	10.8077	10.7404	10.8044	**10.1450**

Table 3. The MAPEs of each prediction method for each city

City	MLP	CNN	LSTM	ConvLSTM
Ningde	0.3952	0.3902	0.4271	**0.3322**
Nanping	0.3887	0.3906	0.4114	**0.3289**
Fuzhou	0.4118	0.4213	0.4238	**0.3442**
Sanming	0.3223	0.3186	0.3297	**0.2732**
Putian	0.4065	0.4268	0.4001	**0.3777**
Quanzhou	0.3455	0.3689	0.3313	**0.3272**
Longyan	0.3308	0.3305	0.3247	**0.2984**
Zhangzhou	0.2835	0.3047	0.2815	**0.2572**
Xiamen	0.3347	0.3614	0.3056	**0.2975**
Average	0.3577	0.3681	0.3595	**0.3152**

The MAEs from low to high are generated by ConvLSTM (6.4579), MLP (7.0221), CNN (7.0906) and LSTM (7.1125). Furthermore, the MAPEs from low to high are generated by ConvLSTM (0.3152), MLP (0.3577), LSTM (0.3595) and CNN (0.3681). Finally, the RMSEs from low to high are generated by ConvLSTM (10.1450), CNN (10.7404), LSTM (10.8044) and MLP (10.8077). From the comparison results, the performance of ConvLSTM is significantly better than the other methods, which proves that the superiority of the ConvLSTM network in predicting PM2.5 concentration.

A case study of air quality prediction by the proposed ConvLSTM network for each city is shown in Fig. 7. The actual records are illustrated as blue polylines, and the predicted records are expressed as orange polylines. In experiments, the predicted values of PM2.5 concentration by the proposed ConvLSTM network are roughly consistent with the actual values. Some large errors may be generally caused by human behaviors. For instance, a large number of fireworks and firecrackers are released during the Spring Festival and New Year's Eve, which causes the rising of PM2.5 concentration.
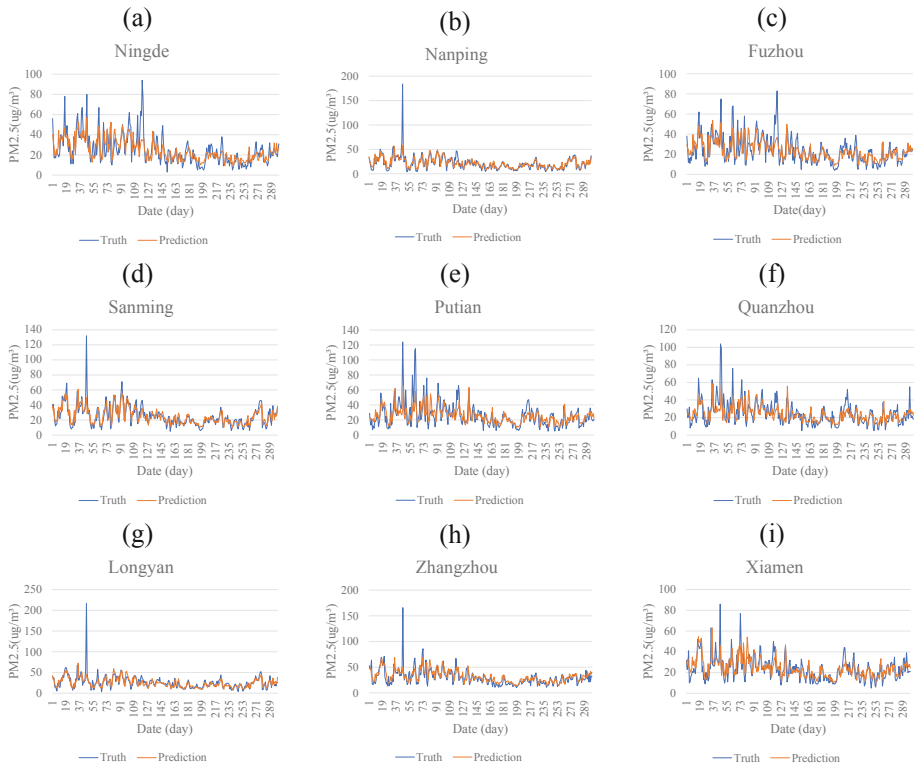
Fig. 7. The prediction results by the proposed ConvLSTM method

5 Conclusions and Future Work

A deep multi-layer neural network model based on CNN and LSTM (i.e., the ConvLSTM method) is proposed to analyze the spatio-temporal features for predicting air quality in multiple cities. A case study of the prediction of PM2.5 concentration in the Fujian Province of the People's Republic of China is given in this study, the proposed model estimate and predict the future concentration of PM2.5 in accordance with the past concentration of PM2.5. In experiments, the performances of each prediction method (e.g., MLP, CNN, LSTM, and ConvLSTM) were evaluated by MAE, MAPE and RMSE. The practical experimental results show that the proposed model combines the advantages of CNN and LSTM for analyzing the spatio-temporal features and improving the accuracy of PM2.5 concentration prediction.

In the future, this study can be applied to the prediction and control of air quality for other cities. Furthermore, the human behaviors can be detected and considered for the improvement of air quality prediction.

References

1. Querol, X., et al.: PM10 and PM2.5 source apportionment in the Barcelona metropolitan area, Catalonia, Spain. Atmos. Environ. **35**(36), 6407–6419 (2001)
2. Schwartz, J., Laden, F., Zanobetti, A.: The concentration-response relation between PM2.5 and daily deaths. Environ. Health Perspect. **110**(10), 1025–1029 (2002)
3. Bell, M.L., Francesca, D., Keita, E.: Spatial and temporal variation in PM2.5 chemical composition in the United States for health effects studies. Environ. Health Perspect. **115**(7), 989–995 (2007)
4. Badyda, A.J., Grellier, J., Dąbrowiecki, P.: Ambient PM2.5 exposure and mortality due to lung cancer and cardiopulmonary diseases in Polish cities. Adv. Exp. Med. Biol. **944**, 9–17 (2017)
5. Chan, C.K., Yao, X.: Air pollution in mega cities in China. Atmos. Environ. **42**(1), 1–42 (2008)
6. Kermanshahi, B.S., et al.: Artificial neural network for forecasting daily loads of a Canadian electric utility. In: Proceedings of the Second International Forum on Applications of Neural Networks to Power Systems, Yokohama, Japan (1993)
7. Fleming, S.W.: Artificial neural network forecasting of nonlinear Markov processes. Can. J. Phys. **85**(3), 279–294 (2007)
8. Gu, K., Qiao, J., Li, X.: Highly efficient picture-based prediction of PM2.5 concentration. IEEE Trans. Industr. Electron. **66**(4), 3176–3184 (2019)
9. Mahajan, S., Liu, H.M., Tsai, T.C., Chen, L.J.: Improving the accuracy and efficiency of PM2.5 forecast service using cluster-based hybrid neural network model. IEEE Access **6**, 19193–19204 (2018)

10. Elangasinghe, M.A., Singhal, N., Dirks, K.N., Salmond, J.A., Samarasinghe, S.: Complex time series analysis of PM10 and PM2.5 for a coastal site using artificial neural network modelling and k-means clustering. Atmos. Environ. **94**, 106–116 (2014)
11. Li, X., et al.: Long short-term memory neural network for air pollutant concentration predictions: method development and evaluation. Environ. Pollut. **231**, 997–1004 (2017)
12. Tsai, Y.T., Zeng, Y.R., Chang, Y.S.: Air pollution forecasting using RNN with LSTM. In: Proceedings of 2018 IEEE 16th International Conference on Dependable, Autonomic and Secure Computing, 16th International Conference on Pervasive Intelligence and Computing, 4th International Conference on Big Data Intelligence and Computing and Cyber Science and Technology Congress, Athens, Greece (2018)
13. Yu, S., et al.: Evaluation of real-time PM2.5 forecasts and process analysis for PM2.5 formation over the eastern United States using the Eta-CMAQ forecast model during the 2004 ICARTT study. J. Geophy. Res. Atmos. **113**, D06204 (2008)
14. Verma, I., Ahuja, R., Meisheri, H., Dey, L.: Air pollutant severity prediction using bi-directional LSTM network. In: Proceedings of 2018 IEEE/WIC/ACM International Conference on Web Intelligence, Santiago, Chile (2018)
15. Zhao, H., Xia, S., Zhao, J., Zhu, D., Yao, R., Niu, Q.: Pareto-based many-objective convolutional neural networks. In: Meng, X., Li, R., Wang, K., Niu, B., Wang, X., Zhao, G. (eds.) WISA 2018. LNCS, vol. 11242, pp. 3–14. Springer, Cham (2018). https://doi.org/10.1007/978-3-030-02934-0_1
16. Krizhevsky, A., Sutskever, I., Hinton, G.E.: ImageNet classification with deep convolutional neural networks. In: Proceedings of the 25th International Conference on Neural Information Processing Systems, Nevada, USA (2012)
17. Lawrence, S., Giles, C.L., Tsoi, A.C., Back, A.D.: Face recognition: a convolutional neural-network approach. IEEE Trans. Neural Netw. **8**(1), 98–113 (1997)
18. Mikolov, T., Karafiát, M., Khudanpur, S.: Recurrent neural network based language model. In: Proceedings of the 11th Annual Conference of the International Speech Communication Association, Chiba, Japan (2010)
19. Hochreiter, S., Schmidhuber, J.: Long short-term memory. Neural Comput. **9**(8), 1735–1780 (1997)
20. Baddeley, A.D., Warrington, E.K.: Amnesia and the distinction between long-and short-term memory. J. Verbal Learn. Verbal Behav. **9**(2), 176–189 (1970)
21. Shi, X., Chen, Z., Wang, H., Yeung, D.Y., Wong, W.K., Woo, W.C.: Convolutional LSTM network: a machine learning approach for precipitation nowcasting. In: Proceedings of the 28th International Conference on Neural Information Processing Systems, Montréal, Canada (2015)
22. Kotsiantis, S.B., Kanellopoulos, D., Pintelas, P.E.: Data preprocessing for supervised leaning. Int. J. Comput. Electr. Autom. Control Inf. Eng. **1**(12), 4091–4096 (2007)
23. Jain, Y.K., Bhandare, S.K.: Min max normalization based data perturbation method for privacy protection. Int. J. Comput. Commun. Technol. **3**(4), 45–50 (2014)
24. Haykin, S.: Neural Networks: A Comprehensive Foundation, 2nd edn. Prentice Hall, Upper Saddle River (1998)
25. Gardner, M.W., Dorling, S.R.: Artificial neural networks (the multilayer perceptron)—a review of applications in the atmospheric sciences. Atmos. Environ. **32**(14–15), 2627–2636 (1998)
26. Haykin, S.O.: Neural Networks and Learning Machines, 3rd edn. Prentice Hall, Upper Saddle River (2008)