Highlights

A spatial correlation prediction model of urban $PM_{2.5}$ concentration based on deconvolution and LSTM

Bo Zhang,Yuan Liu,RuiHan Yong,Guojian Zou,Ru Yang,Jianguo Pan,Maozhen Li

- The use of deconvolution neural network resolves the problem of excessive information loss incurred in traditional CNNs.
- The input data of the model takes into account the concentration of atmospheric pollutants, meteorological factors and environmental factors in the adjacent areas in predicting future $PM_{2.5}$.
- The presented Dev-LSTM model outperforms the classic models in mining the spatial temporal correlation of pollutants and subsequent predictions.

A spatial correlation prediction model of urban $PM_{2.5}$ concentration based on deconvolution and LSTM

Bo Zhang[a,b,c], Yuan Liu[a,*], RuiHan Yong[a], Guojian Zou[d], Ru Yang[a,*], Jianguo Pan[a,b,c,*] and Maozhen Li[e]

[a] *College of Information, Mechanical, and Electrical Engineering, Shanghai Normal University, Shanghai, 200234, China*
[b] *Institute of Artificial Intelligence on Education, Shanghai Normal University, Shanghai, 200234, China*
[c] *Shanghai Engineering Research Center of Intelligent Education and Bigdata, Shanghai Normal University, Shanghai, 200234, China*
[d] *Key Laboratory of Embedded Systems and Service Computing, Ministry of Education., Tongji University, Shanghai, 200234, China*
[e] *Department of Electronic and Electrical Engineering., Brunel University London, Kingston Lane, Uxbridge, Middlesex UB8 3PH, UK*

ARTICLE INFO

Keywords:
Air pollutant concentration prediction
Deconvolution
Dev-LSTM
Deep learnin

ABSTRACT

Precise prediction of air pollutants can effectively reducre the occurrence of heavy pollution incidents. With the current surge of massive data, deep learning appears to be a promising technique to achieve dynamic prediction of air pollutant concentration from both the spatial and temporal dimensions. This paper presents Dev-LSTM, a prediction model building on deconvolution and LSTM. The novelty of Dev-LSTM lies in its capability to fully extract the spatial feature correlation of air pollutant concentration data, preventing the excessive loss of information caused by traditional convolution. At the same time, the feature associations in the time dimension are mined to produce accurate prediction results. Experimental results show that Dev-LSTM outperforms traditional prediction models on a variety of indicators.

1. Introduction

Under the current circumstances, air pollution is becoming increasingly prominent, and the frequency of severely polluted weather has been increasing. When the concentration of air pollutants exceeds a certain standard, it can cause heavily polluted weather, and even have a significant impact on people's daily lives [1]. For example, under heavy smoggy weather, the visibility of the surrounding environment is reduced, which brings inconvenience to people's travel posing potential risks to traffic accidents. Air pollutants, i.e., substances that are present in the air can cause great harm to people's health and have been receiving a lot of attention [2–4]. It has become an imperative to predict the concentration of air pollutants to reduce the occurrence of heavy pollution incidents and carry out effective air quality control.

Air pollutant concentration prediction has been a research hotspot in academia [5, 6]. However, the change of air pollutant concentration is dynamic, and the production of air pollutants is complicated. The analysis and prediction work involves multiple departments, regions, and fields [7]. It needs to process a large amount of air pollutant concentration data and related meteorological information. At the same time, it is necessary to dig out the laws of dynamic changes of air pollutant concentration. So far, academia has done a lot of research work in the field of air pollutant concentration prediction, most of which are based on traditional air pollutant concentration and prediction methods with non-deep learning techniques [8, 9]. In face of the increasing sensor monitoring data, and the increasingly complex air pollutants' causes and their diffusion methods, researchers have encountered different forms of bottlenecks that need to be broken through [10]. For example, it is difficult to probe complex high-dimensional relationships from massive datasets.

The prediction of $PM_{2.5}$ concentration can be regarded as a time series processing problem, which can be predicted according to historical data, such as meteorological factors such as humidity and temperature, and other pollutant factors such as SO_2 and CO [11]. It has been proved by many works that there are complex interactions among these factors [11–14].

On the one hand, traditional air pollutant concentration prediction just performs shallow learning on the data, and thus cannot effectively utilize and integrate massive historical monitoring data, extract the deep connections between data features and conduct in-depth mining of the hidden features of historical data [15]. On the other hand, traditional air pollutant concentration prediction methods rely too much on the laws summarized from historical data, and fail to take into account the complexity and variability of air pollutant concentration changes [10, 16]. Air pollutant concentration varies in time and space. In terms of dimensions, they are all dependent. Specifically, air pollutant concentration that happened in the past affects current and future air pollutant concentration [17]. In addition, air pollution is a regional diffusion problem, which needs to consider the spatial dimension. Air pollution concentration in the vicinity of a target city also affects the air pollutant concentration in a target city [18]. This means that there is a spatial correlation between the air pollution effects of neighboring cities.

In recent years, a new generation of data analytical approaches represented by deep learning [19] is being adopted

*Corresponding author
1000513397@smaill.shnu.edu.cn (Y. Liu); yangru@shnu.edu.cn (R. Yang); panjg@shnu.edu.cn (J. Pan)

in many areas [20–30]. At the same time, some models in deep learning, because of their unique structure, can conduct linkage analysis on the temporal and spatial correlation of air pollutant distribution and diffusion [31], and then dynamically predict urban air pollutant concentrations. However, research in deep learning is far from sufficient on spatial-temporal correlation analysis. Convolutional Neural Networks (CNN) [14, 32–34] have been shown to be powerful in spatial data processing. However, the data dimension is compressed after the convolution operation, which would lose some useful information degrading the performance on accuracy.

This paper presents Dev-LSTM, a regional air pollutant concentration prediction model based on deconvolution neural network and LSTM network. Dev-LSTM is trained on air pollutant data and meteorological data of a target city and the neighboring cities. Dev-LSTM fully integrates air pollutant data and meteorological data, uses deconvolution network to extract the spatial feature correlation of the air pollutant concentration, and then uses LSTM to extract the time dimension feature correlation. Dev-LSTM dynamically analyzes the spatial and temporal dimension dependence of atmospheric pollutant concentrations.

The main contributions of the paper are as follows:

1) The use of deconvolutional neural networks retains the advantage that CNN extracts the correlation between features from the spatial dimension. More importantly, it resolves the problem that a traditional CNN over-compresses the data dimension resulting in excessive loss of information.
2) The input data set of the model includes air pollution and meteorological data, and also considers the influence of air pollutants and meteorological factors on future $PM_{2.5}$ concentrations. Considering that a air pollutants in a target city are affected by neighboring areas, the data from the surrounding cities are also utilized.
3) Experiments show that Dev-LSTM performs better on multi-dimensional indicators in terms of mining the temporal and spatial correlations of pollutants and achieving accurate prediction compared with traditional models.

The remainder of this paper is organized as follows. Section 2 reviews the related work. Section 3 details the design of Dev-LSTM. Section 4 evaluates the performance of Dev-LSTM and analyzes the experimental results, and Section 5 concludes the paper.

2. Related Work

In the past, academic research on air pollutant concentration prediction can generally be divided into two types: air pollutant concentration prediction based on non-deep learning and air pollutant concentration prediction based on deep learning. Prediction methods based on non-deep learning can also be called traditional prediction methods, which can be divided into predictions based on empirical models, predictions based on probability models, predictions based on comprehensive methods, and predictions based on traditional machine learning.

2.1. Empirical Model

The atmospheric pollutant concentration prediction methods based on empirical models focus on summarizing the law based on historical meteorological and atmospheric pollutant concentration data, establishing a linear regression equation, then solving the coefficients of the regression equation, and finally predicting a certain time in the future according to the solved regression equation. air pollutant concentrations. Moisan et al. [35] proposed a method based on dynamic multi-linear equations, which combines hourly, daily and annual air pollutant concentration seasonal changes to predict hourly $PM_{2.5}$ concentration in Santiago, Chile. Experimental results showed that the proposed model has the potential to surpass other linear prediction models in terms of the accuracy of air pollutant concentration prediction; Abdullah et al. [36] established a linear regression model (linear mixed model multiple linear regression), which used meteorological parameters and historical concentrations of air pollutants as input data to predict future air quality.

2.2. Probabilistic Model

Probabilistic model-based predictions are based on probabilistic methods in mathematics, combined with historical data to produce predictions of future atmospheric pollutant concentrations. Wu et al. [37] proposed a variational Bayesian approach for the adaptive air pollution prediction problem. They formulated the adaptive prediction task based on recent support data as a conditional inference problem on a parametric graphical model, and further proposed an end-to-end learning objective with two-step variational approximation to estimate intractable conditional likelihoods; Alyousifi et al.[38] used the maximum a posteriori method to determine the transition probability matrix of the discrete-time Markov chain model. based on the API hourly data collected from Peninsular Malaysia, a maximum a posteriori (MAP) method was proposed to estimate the Markov chain transition probability matrix for chain model under three different priors (Dirichlet, Jeffreys and Uniform).

Since the prediction of pollutants is only a summary of changes in pollutants based on historical experience, it is insufficient to represent the complex effects of volatile atmospheric environments.

2.3. Comprehensive Model

The air pollutant concentration prediction based on the integrated method is a method that integrates at least two technologies to produce the final deterministic prediction result. First, it analyzes the massive air pollutant data and meteorological data monitored by the sensors, eliminates redundant data or fills in missing data, and then obtains the change law of the data in the time dimension. And according to this law, the final prediction of the air pollutant concentration in the future is made. For example, the Land

Use Regression method [39] is combined with other methods to predict the air pollutant concentration. Specifically, regression kriging extended linear LUR models in [40] to improve the explanatory power and predictive performance of the models. Shi T et al. used the LUR model to predict the spatial concentration distribution of NO_2 and PM_{10} in the central urban agglomeration of Liaoning, and combined with the multiple linear regression method to establish an equation to obtain the spatial concentration distribution map of the target pollutants based on the significant variables in the heating season and non-heating season.

However, the influence conditions of air pollutants are too complex to be judged as a simple continuation of a certain law. Moreover, the large amount of background data and the daily accumulation of air pollution-related data are not independent, they are time-dependent and spatially correlated which shall be considered in prediction models.

2.4. Traditional Machine Learning Models

Air pollutant concentration prediction methods based on traditional machine learning mainly use machine learning to imitate and realize human behavior, learn new capabilities through known input from the outside world, and iterate the characteristics of their own capabilities based on existing experience [41]. For example, Li et al. [42] used Random Forest to establish a real-time prediction model of $PM_{2.5}$ concentration based on monitoring data and meteorological data. Zhao et al. [43] developed an enhanced geographically weighted regression (GWR) model to analyze the spatial distribution of $PM_{2.5}$ concentrations by combining Geodetector analysis and principal component analysis (PCA). Alternatively, Park et al. [44] used ANN (Artificial Neural Network) with a hidden layer to study the PM_{10} concentration inside and outside the subway station in Seoul.

Traditional prediction methods are mainly targeting at small sample data sets, and the learning of pollutant characteristics is still at a shallow level, and it is impossible to perform dynamic analysis of pollutants in time and space.

2.5. Deep Learning Models

As one of the hot topics in academic research in recent years, deep learning has been used to predict the concentration of air pollutants and can achieve better results than existing methods. Ch et al. [45] developed a dynamic statistical mixture model for the next two-day forecast of $PM_{2.5}$ concentrations in the Seoul Metropolitan Area in South Korea by combining the CMAQ- based prediction with the RNN algorithm [12]; Zhang B et al. utilized a combination of autoencoder and bidirectional LSTM model (AE-Bi-LSTM) [11] to extract feature concentrations between meteorological variables and $PM_{2.5}$, and then predicted $PM_{2.5}$ values over time series; Rui et al. [13] used a multilayer perceptron (MLP) to analyze and predict ambient $PM_{2.5}$ in eight regional core cities in China; While Ragab et al. [14] used 1D Deep Convolutional Neural Network (1D-CNN) and Exponential Adaptive Gradient (EAG) optimization to predict the API of selected locations in Klang, Malaysia. Their proposed method performed an average absolute better prediction accuracy values than the baseline model in terms of error (MAE), root mean square error (RMSE), mean absolute percentage error (MAPE), and correlation coefficient (R-Squared); Combining the advantages of convolutional neural network (CNN) and long short-term memory (LSTM) models, Ding et al. [32] proposed a hybrid CNN-LSTM model based on spatiotemporal correlation to predict daily $PM_{2.5}$ concentration in Beijing. CNN model is used to learn spatial features, while LSTM model is used to extract temporal information, which outperforms the same model without spatiotemporal correlation.

However, CNN-LSTM has two key problems. First of all, CNN loses information in extracting the hidden spatial features of pollutants and meteorological data, which can easily lead to the loss of feature information and the degradation of the model.

Recently, methods based on attention mechanism have been widely used in many research methods[46, 47]. In the research of pollutant concentration prediction in recent years, the temporal and spatial features of multi-site pollutant data are extracted by attention-based method. In the task of pollutant concentration prediction, the prediction model based on Att-ConvLSTM uses spatio-temporal attention and ConvLSTM method to weight the input data and extract spatio-temporal features, and the accuracy of the final prediction results has been greatly improved [48]. However, the current pollutant concentration prediction model based on Att-ConvLSTM has encountered the following challenges: First, it is difficult to deeply extract the hidden distribution features of pollutants and meteorological data. Second, the influence of various pollutants and meteorological factors on the prediction results is not considered. Thirdly, ConvLSTM method combines the advantages of CNN and LSTM model, and mainly extracts the spatio-temporal correlation characteristics of long-time series data. However, the single ConvLSTM network model has a major disadvantage [48]. On the one hand, the extraction of temporal and spatial features of pollutants and meteorological data lacks consideration of the influence of urban location on prediction. On the other hand, with the increasing number of ConvLSTM layers, more network degradation problems will appear in the model, and the training cost will increase rapidly. Therefore, it is difficult for Att-ConvLSTM to overcome the above two problems.

The current air pollutant prediction based on deep learning can solve the bottleneck problem of traditional prediction. However, there is not enough research work on the use of deep learning to predict the concentration of air pollutants, and to analyze time-space correlation problems. Research on the integration of different neural networks is still not mature enough to mine the correlation of spatial dimensions. Meanwhile, although the traditional CNN network can be used to mine spatial information, the data dimension is compressed after the convolution operation, which will lose useful information. Aiming at the spatial correlation and temporal correlation of air pollutant concentrations, the Dev-LSTM model presented in this paper

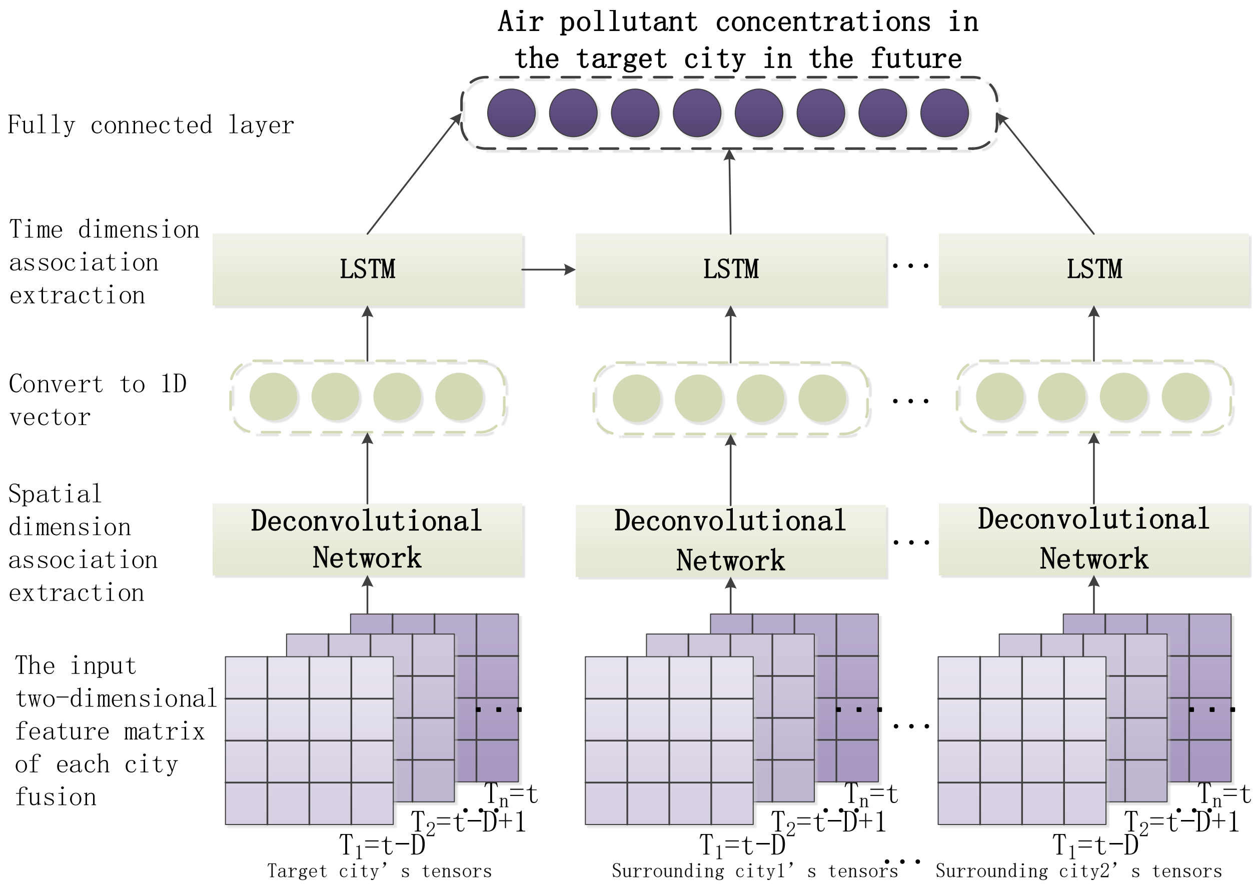

Fig. 1: Dev-LSTM prediction model architecture.

integrates deconvolutional neural network with LSTM network to provide spatiotemporal dynamic predictions of air pollutant concentrations.

3. Dev-LSTM

This section presents the design of the Dev-LSTM model. The architecture of Dev-LSTM is shown in Fig. 1.

3.1. Deconvolutional Neural Network

A traditional CNN is able to extract the correlation between features from the spatial dimension, but suffers from an excessive loss of information due to over compression of data dimensions. The deconvolutional neural network does not perform the learning and extraction of features from the perspective of continuously compressing the feature dimensions. Instead, by expanding the feature dimensions before passing the data features through the convolutional layer, it can still maintain a rich amount of information and fully learn the data characteristics after the convolution operation of lossy compression, which is not possible with traditional CNN.

Fig. 2 shows the network structure of the deconvolutional neural network (MP matrix represents the characteristic matrix of meteorological factors and air pollutant concentration).

In Fig. 2, the deconvolutional neural network first performs a reconstruction operation for the air pollutant concentration and meteorological feature matrix input in time series $\left(T_1, T_2, \dots T_n\right)$ (represented as the MP matrix in the figure), i.e., all low-dimensional feature matrices are reconstructed into high-dimensionality feature matrix. In order to keep the original information of the matrix, it is appropriate to use zero padding. The reconstructed MP matrix shown in Fig. 2 centered on the original MP matrix, around which expansion operations are performed. The feature matrix after the convolution operation is in the principle of the same dimension as the original input MP matrix, which is equivalent to another feature expression method of the input MP matrix. The corresponding input and output matrices are used to calculate the reconstruction error, and update the parameters of the deconvolution network according to the reconstruction error.

The traditional convolution network is subsampling, and the corresponding output size will be reduced. The deconvolution network is upsampling. The significance of upsampling is to restore the small-sized high-dimensional feature map, so as to do pixelwise prediction and obtain the classification information of data points.

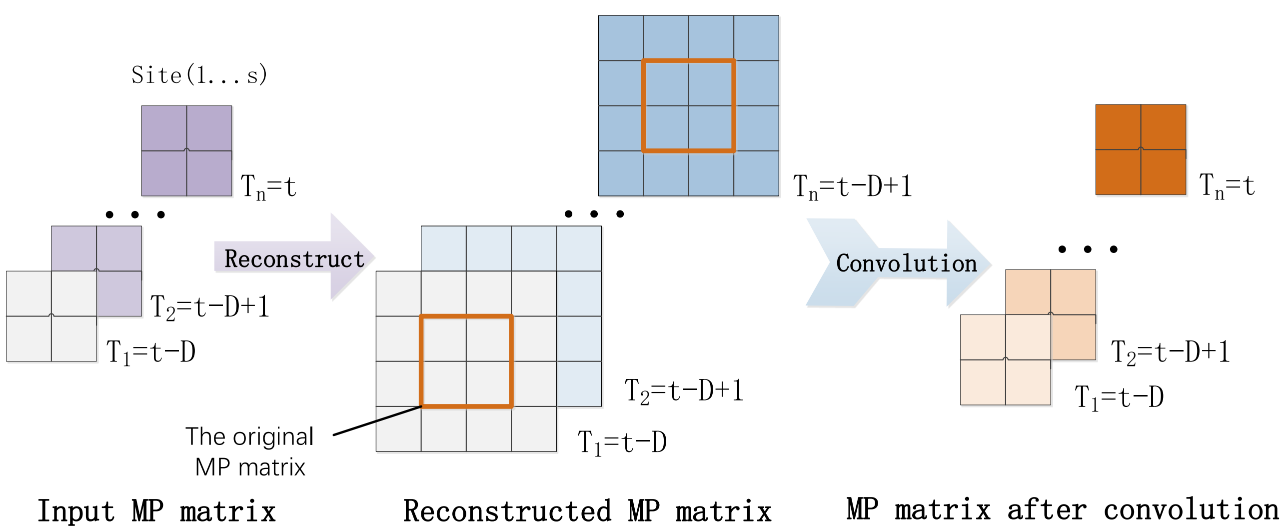

Fig. 2: The network structure of the deconvolutional neural network.

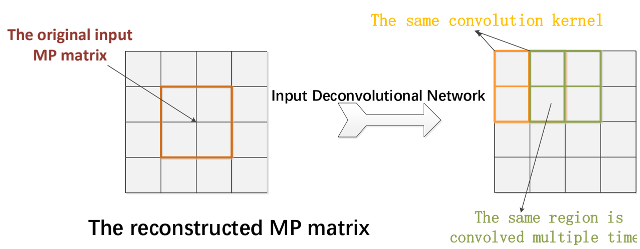

Fig. 3: Reconstruction and multiple convolution of the input matrix.

As shown in Fig. 3, compared with traditional CNN, each region of the input matrix is only be convolved once, and the deconvolution network performs multiple convolution operations on the same region, thereby realizing the full extraction of spatial dimension features.

3.2. Dev-LSTM Model Training

Fig. 4 is the architecture diagram of the entire Dev-LSTM air pollutant prediction model. The multi-city and multi-site historical data is used as the input of the Dev-LSTM model. The bottom layer of the model is a deconvolution neural network, which is used to extract the spatial correlation of the input data. The high layer of the model is an LSTM neural network, which is used for deconvolution. The output after the network extracts spatial features is extracted again, and the hidden connections of air pollutants are mined from the time dimension. This Dev- LSTM air pollutant prediction model is repeatedly trained and tuned until the model's performance reaches the desired result.

The training weight of the previous stage of the deconvolution network is the initial weight of the entire model during training. The output result of the spatial correlation extracted by the deconvolution network is the input of LSTM. The prediction of LSTM is to reduce the air pollutants in D hours $\left(T_1, T_2, \ldots, T_n\right)$ before Time t. The concentration value is used as the input of LSTM, and the predicted target is the concentration value of air pollutants in the target city hour after the time (D and N are both set time windows). Let x denote input, which is a dynamic time series, W denote weight matrix, h denote hidden layer information, and b denote bias. The training process of LSTM is shown below:

(1) LSTM selectively forgets the air pollutant concentration data and some meteorological factor data in the historical time with the Eq. (1),

$$f_t = \sigma\left(W_f\left[h_{t-1}, x_t\right] + b_f\right) \tag{1}$$

where f_t represents the information that needs to be forgotten at the current moment.

(2) After selectively forgetting some information, it is needed to determine the updated storage information in the cell, which comes from two parts, one is the input information determined by the Sigmoid function at the current moment, and the other is absolute and hide status information by the tanh function at the previous moment. These two parts are represented by the Eq. (2),

$$\begin{aligned} i_t &= \sigma\left(W_i\left[h_{t-1}, x_t\right] + b_i\right) \\ C_t' &= \tanh\left(W_c \cdot \left[h_{t-1}, x_t\right] + b_t\right) \end{aligned} \tag{2}$$

where C_t' is the updated storage information in the determined cell.

(3) Based on the hidden state information at the last moment and the updated storage information in the determined cell, the old state is updated using the Eq. (3),

$$C_t = f_t * C_{t-1} + i_t * C_t' \tag{3}$$

where C_t is the final storage information in the determined cell at the current time.

(4) Finally, output o_t, which is the predicted concentration of air pollutants in the target city is decided, and the

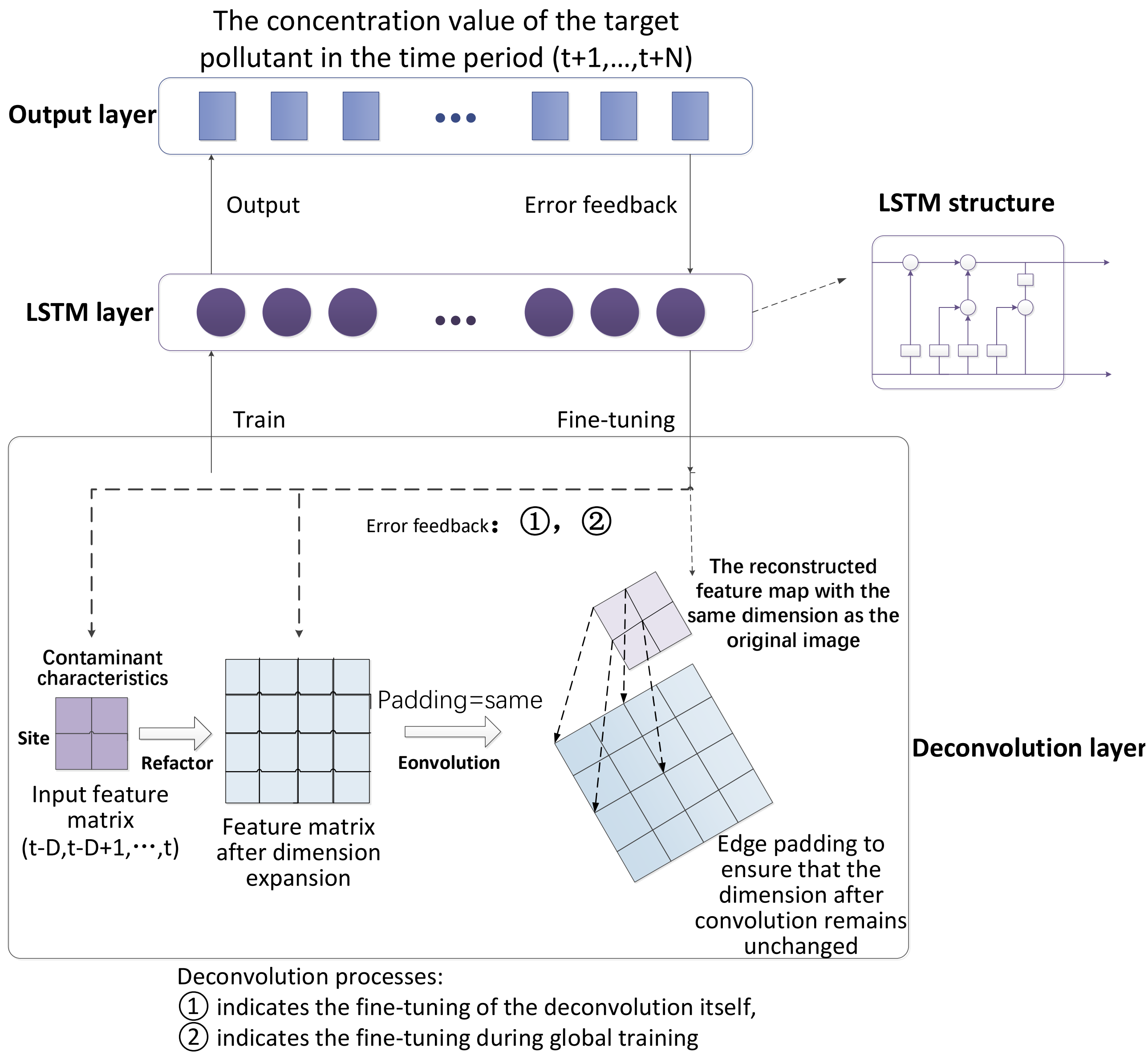

Fig. 4: The training process of the Dev-LSTM model.

hidden layer status information at the current moment is updated by using the Eq. (4),

$$
\begin{aligned}
o_t &= \sigma' \left(W_o \left[h_{t-1}, x_t \right] + b_o \right) \\
h_t &= o_t * \tanh \left(C_t \right)
\end{aligned} \tag{4}
$$

For the prediction result output by LSTM, this paper defines a loss function to measure the error between the predicted value and the true value, and adjusts the network repeatedly according to this error:

$$
E(\varphi) = \frac{1}{N} \sum_{i=1}^{N} \left(X_i - Y_i \right)^2 + \frac{\lambda}{2} \left((1-\zeta)|\varphi| + \zeta \varphi^T \varphi \right) \tag{5}
$$

In Eq. (5), N is the predicted length of time in the future, X_i represents the predicted value of the air pollutant concentration in the target city generated by Dev-LSTM, Y_i represents the true value of the corresponding air pollutant concentration; λ is a non-negative hyperparameter φ represents the set of weights of u, w, d, ξ in section 3.2 of this article, namely $\varphi = \{u, w, d, \xi\}$, and ζ is the parameter that controls the proportion of penalty terms used in L1 and L2, $\zeta \in (0, 1)$.

The process of training and fine-tuning is repeated, and the performance of the Dev-LSTM prediction model is continuously optimized, so that it can accurately organize the input of historical data from multiple cities and multiple sites, extract the spatial correlation between features during the training process and realize the full use of data and produce more accurate prediction results.

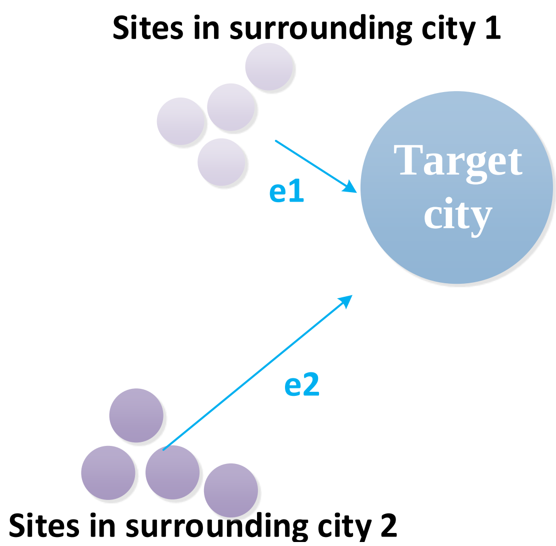

Fig. 5: Spatial effect on target city.

3.3. Spatial correlation between cities

In the experiment carried out in this study, because the concentration of pollutants is affected by the spatial relationship, we chose the neighboring cities close to the target city. The e1,e2 is the spatial influence of surrounding cities on the target city Fig. 5. Tensors of multiple stations in surrounding cities are input into the model in time series.

4. Performance Evaluation

4.1. Data description

In this work we selected Shanghai, China as the target city and Suzhou and Hangzhou as the surrounding cities of Shanghai to measure the impact on the air pollutant concentration of the target city, and incorporates the correlation of the spatial dimensions of air pollutant concentration changes into the prediction system. The training set spans 2 years, and the test set spans 1 year. The input hourly air pollutants and meteorological factors are shown in Table 1.

Table 1
Air pollutants and meteorological factors.

Air pollutants	Meteorological factors
AQI	Temperature
$^*PM_{2.5}$	Dew point
PM_{10}	Wind speed
SO2	Sky condition
NO_2	Precipitation
O_3	
CO	

The air quality data selected in this paper come from the national urban air quality real-time publishing platform of China National Environmental Monitoring Center, and the meteorological data come from the National Climate Data Center (NCDC).

Historical pollutant concentrations and meteorological data was collected from 7 cities from January 1, 2018 to January 1, 2020. This work selected 7 cities near the Yangtze River Delta (Shanghai, Hangzhou, Suzhou, Nantong, Wuxi, Shaoxing, Jiaxing). We selected 12 pollutants and meteorological factors: Air Quality Index (AQI), $PM_{2.5}$, PM_{10}, SO2, NO_2, O_3, CO, temperature, dew point, wind speed, sky condition(the code that denotes the fraction of the total celestial dome covered by clouds or other obscuring phenomena.), precipitation.

Among air pollutants in Table 1, $PM_{2.5}$ was selected as the target air pollutant, i.e., the forecast target of this work was the $PM_{2.5}$ concentration of Shanghai.

For the missing values of air pollutant concentration and meteorological data sets, if the data of a single day was missing, it was filled with the data of the previous day; if the data of multiple days was missing continuously, it was assumed that the data changes evenly during this period, that is, filled in according to the arithmetic sequence data.

Fig. 6 shows the annual value change of each pollutant concentration, including AQI. By observing the changes in the concentrations of pollutants such as $PM_{2.5}$, it can be found that the trend of changes in pollutant concentrations is generally consistent, which also reflects that there may be hidden relationships between pollutants.

Fig. 7 shows the annual change of meteorological factors. From Fig. 7, we can observe that the temperature and dew point have the same change. In addition, the numerical types and intervals of meteorological elements are quite different, but the changing trends are highly similar, which means that there may be mutual influences among meteorological elements. Thirdly, the meteorological factors are consistent with the change of $PM_{2.5}$ concentration, which implies the implicit relationship between air pollutants and meteorological factors. For example, between 4000-5000 hours, $PM_{2.5}$ reaches its maximum value, and the temperature and precipitation are also the maximum at this time. Therefore, combined with the existing research results [31], in the study of $PM_{2.5}$ concentration prediction, we take meteorological factors as part of the model input, and extract the hidden features between pollutants and meteorological factors.

4.2. Experimental setup

Based on a number of experiments, an optimal hyperparameter set was selected in this study. The validation set used in this study was closely related to the training stage. After each epoch, RMSE, MAE and Corr of the prediction model on the validation set were calculated. Therefore, we can choose the optimal model according to the model error calculated on the validation set. Specific parameters are shown in Table 2.

Fig. 8 depicts the prediction effect of Dev-LSTM model on test set. The x-axis represents the observed value of $PM_{2.5}$, and the y-axis represents the predicted value of

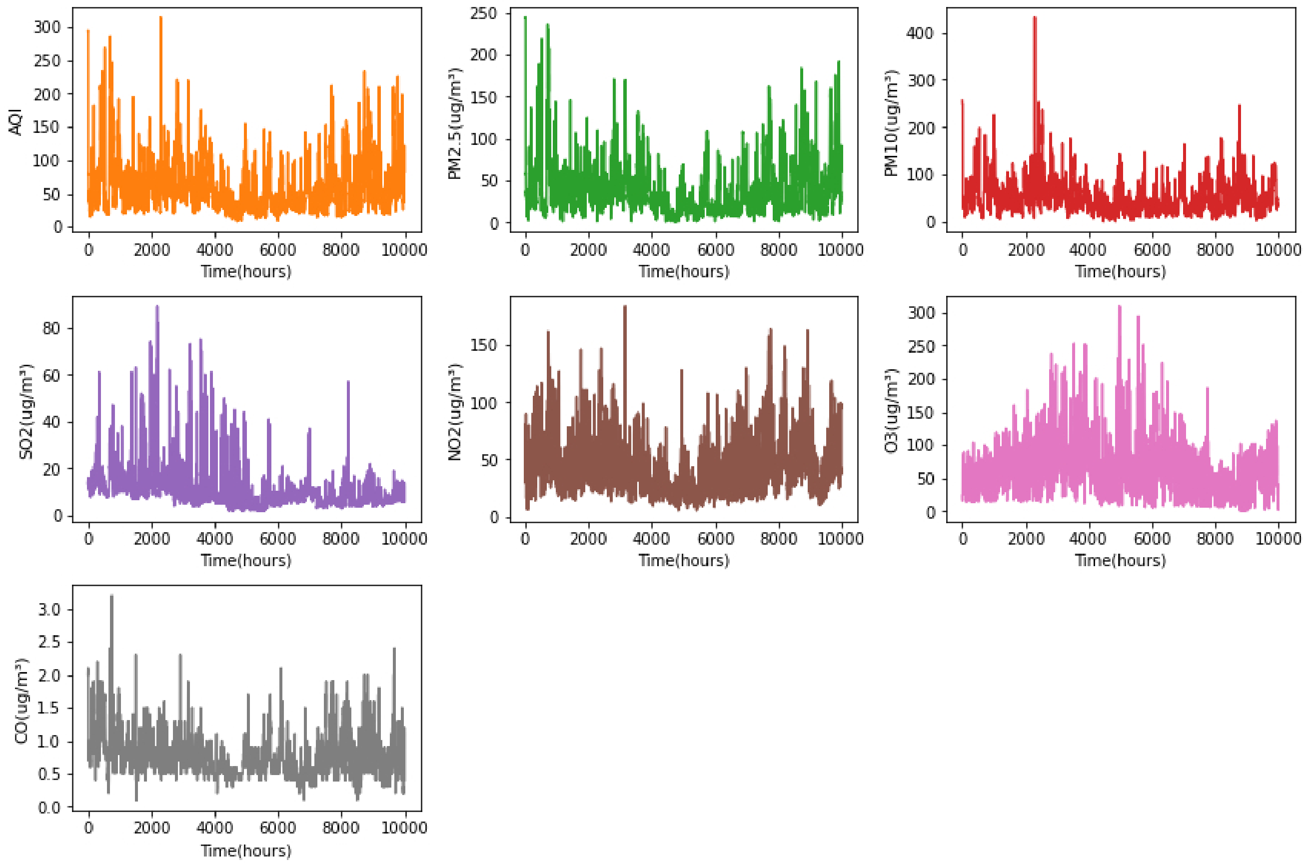

Fig. 6: Time series plot of air pollutant concentration data.

Table 2
Dev-LSTM model parameter settings.

Symbol	Description	Value
Kernel_size	Convolution kernel size	5*5
Batch_size	Batch size for each training	32
Stride	Convolution slidingś window step size	1
LNode_num	Number of LSTM nodes	128
Learning_rate	Learning rate	0.0005
Epochs	Number of iterations	150
Input_shape0	Input matrix dimensions	9*12(SH) 11*12(HZ) 7*12(SZ)
Dropout	Rate of dropout	0.2
D-N	Time in the past to predict the concentration of the future time	1-1

$PM_{2.5}$. The black line indicates the function, and the red dot indicates the deviation between the observed value and the predicted value. Figure 8 shows that the predicted data are basically consistent with the observed data. By calculating the correlation, the correlation coefficient between the predicted value and the observed value is 0.988, MAE is 2.982, and RMSE is 3.789.

In order to more intuitively show the performance difference of several prediction models on the test set, this paper compares multiple prediction models for RMSE, MAE and Corr values. As can be seen from Table 3, the Dev-LSTM prediction model has the lowest RMSE and MAE values and the highest correlation (Corr) between the true and the predicted value on the test set. Therefore, the performance of Dev-LSTM is better than these classic models.

4.3. Experimental comparison

As can be seen from Table 3, compared with the traditional methods, the LSTM and Bi-LSTM have better prediction results compared to the traditional methods because they can better extract the temporal dependence of pollutants.

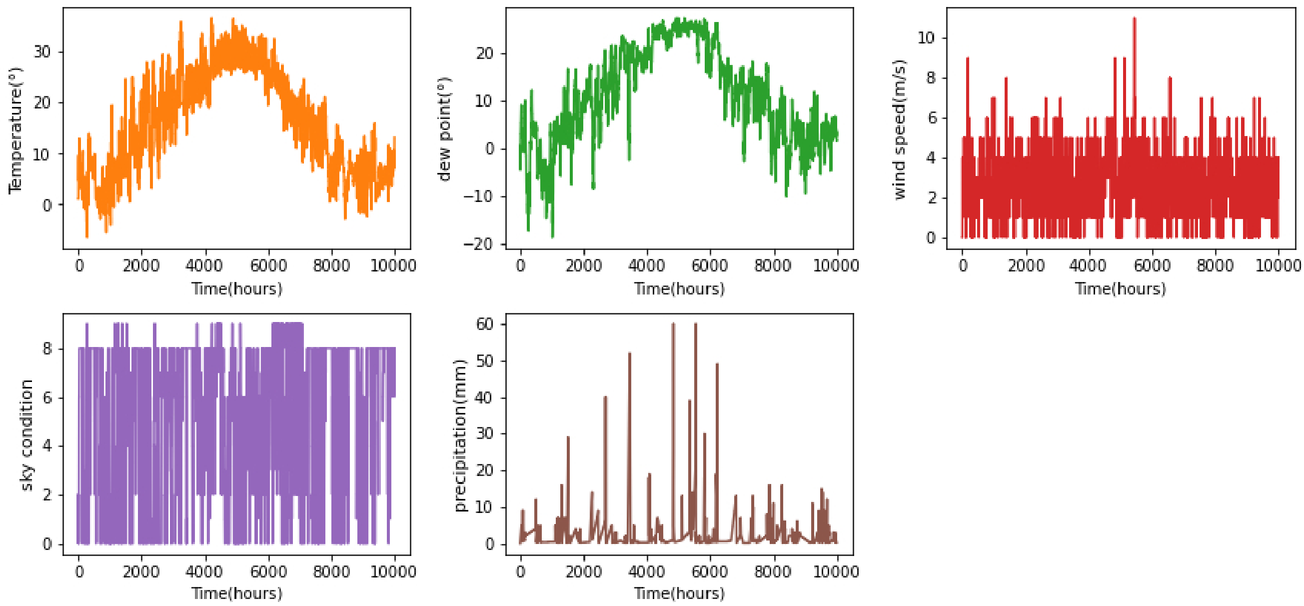

Fig. 7: Time series plots of meteorological data.

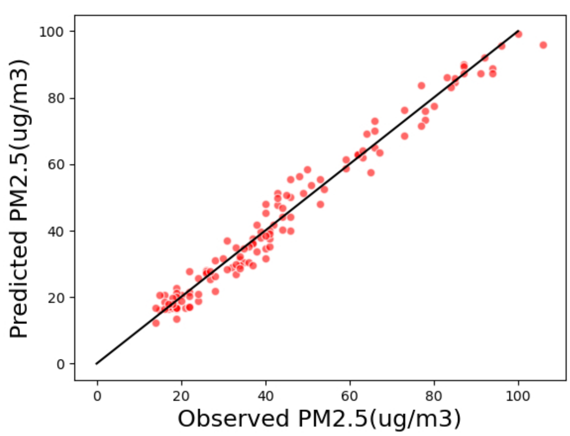

Fig. 8: The fitting degree between the observed value and the predicted value.

Table 3
Model performance comparison.

model	RMSE	MAE	Corr
DME [35]	7.88	4.97	-
CAMx [49]	34.454	-	0.712
CMAQ [50]	34.087	-	0.708
Variational Bayesian [37]	-	10.275	-
XGBoost [51]	6.07	3.94	-
ANN [44]	13.26	-	-
LSTM [52]	10.93	5.05	-
Bi-LSTM [53]	13.245	9.654	0.969
Att-ConvLSTM [48]	11.476	8.321	0.974
3D-CNN-GRU [54]	7.15	4.64	-
Dev-LSTM	**3.789**	**2.982**	**0.988**

Table 4
The impacts of each module.

model	RMSE	MAE	Corr
LSTM	4.634	3.673	0.985
CNN	4.140	3.331	0.985
CNN-LSTM	4.127	3.394	0.985
BP	4.001	3.193	0.986
Deconvolution	3.970	3.233	0.986
Dev-LSTM	**3.789**	**2.982**	**0.988**

Att-ConvLSTM, 3D-CNN-GRU and Dev-LSTM have better prediction results. Because these five methods can better deal with the long-term sequence dependence problem with spatial characteristics. The RMSE, MAE and Corr reached the optimal values of 3.789, 2.982 and 0.988 respectively. Next, comparing the prediction results of 3D-CNN-GRU and Dev-LSTM in Table 3, the prediction accuracy of Dev-LSTM is higher than that of 3D-CNN-GRU, which proves that deep FDN-Learning has better spatio-temporal feature ability than integrating CNN and LSTM methods in extracting pollutants and meteorological data. In order to verify the effect of each module, we designed corresponding experiments. As shown in Table 4, CNN model is superior to LSTM model, and better results can be obtained if the two models are combined. If the deconvolution network is taken into account, the deconvolution network is superior to CNN and CNN-LSTM in this task. Therefore, we propose Dev-LSTM, a deep learning model that integrates deconvolution network and LSTM. This experiment is also an ablation experiment of our proposed model, which proves that every module in the model structure is indispensable.

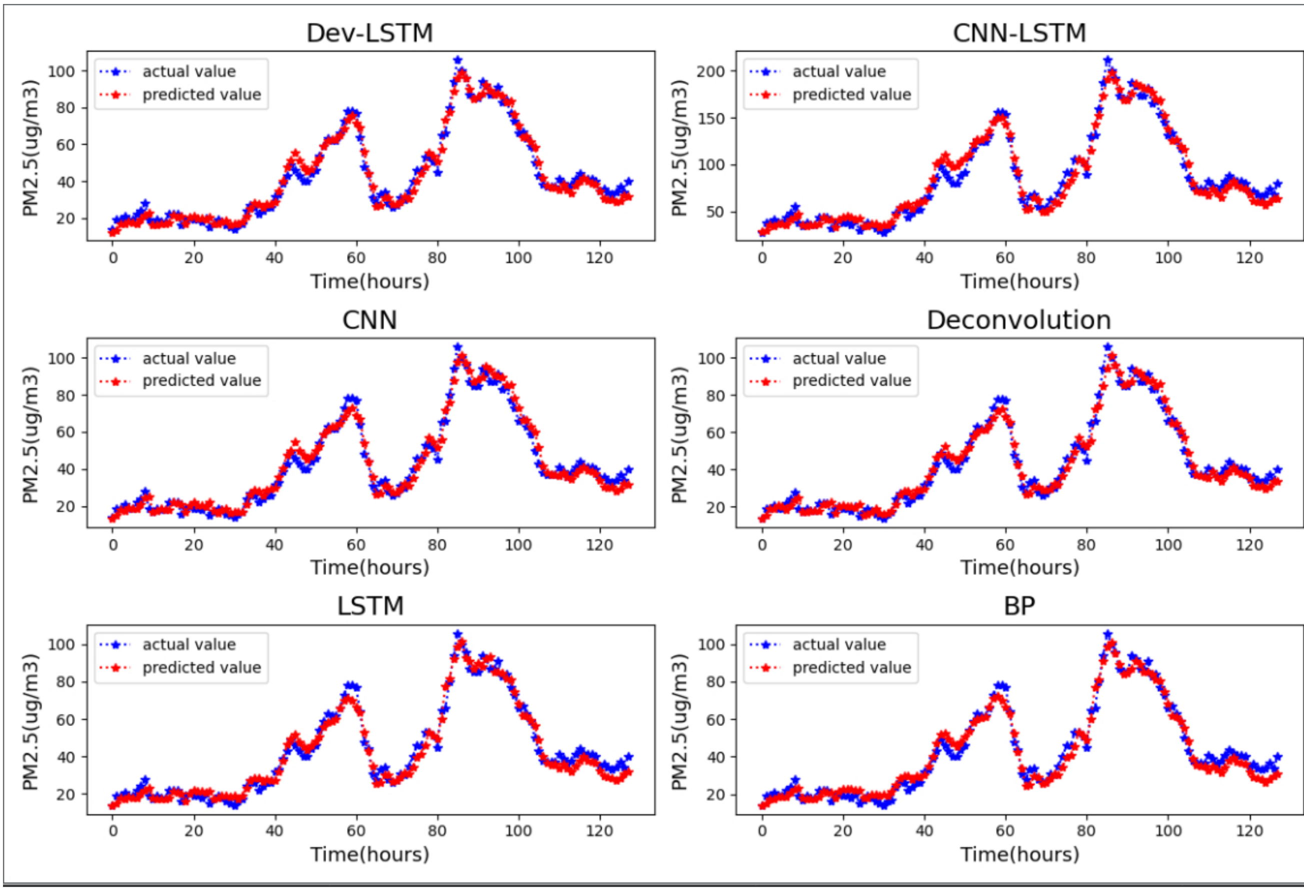

Fig. 9: The fitting degree between the observed value and the predicted value.

Fig. 9 shows the fitting trends of Dev-LSTM, CNN-LSTM, CNN, Deconvolution, LSTM and BP models respectively. Compared with CNN-LSTM, CNN, Deconvolution, LSTM and BP models, they are trained in the same training set and tested in the same test set. Table 4 is almost consistent with the predicted and observed results of the Dev-LSTM model described in Fig. 9, and has a good fitting effect on the sudden change of $PM_{2.5}$ concentration (e.g. 40-60 hours, 80-100 hours). The results show that Dev-LSTM is able to extract the temporal and spatial correlation characteristics of complex pollutant concentrations and meteorological data in many cities in the region, solve the long-term dependence problem in pollutant prediction, and effectively deal with the sudden change of pollutant concentrations.

In this work we also varied some of the hyperparameters of the Dev-LSTM prediction model to show the performance of the prediction model. Table 5 lists the results of Dev-LSTM on the test set when the convolution kernel of the deconvolution network and the LSTM network take different numbers of nodes. It can be concluded that in the Dev-LSTM prediction model, by setting the convolution kernel of the deconvolution network to 3*3, and the number of nodes of the LSTM network to 256, the integration of two neural networks, results in smaller values on RMSE and MAE, and a larger value on Corr. The model performance in this configuration is better than other convolution kernel-node number combinations.

In order to verify the prediction effect of the Dev-LSTM prediction model under different time periods, we also considered different time slots. That is, different time periods in the past were used to predict the concentration of air pollutants in different time periods of the future. Then various index values of the Dev-LSTM prediction model under different are listed in Table 6. As can be seen from this table, the optimal prediction duration of the Dev-LSTM prediction model is 1-1. As the prediction duration increases, various performance indicators gradually deteriorate. The increase in time series leads to more redundant information, which is a problem faced by all forecasting models. Therefore, under the configuration of 1-1, the Dev-LSTM prediction model has been able to surpass the performance of other models.

4.3.1. *Spatial weighted prediction performance*

In order to measure the spatial influence of several surrounding cities selected in this work on the concentration of air pollutants in the target city and the spatial correlation of air pollutant concentrations, this paper conducted comparative experiments for analysis. The experimental results are shown in the Table 7, where NC1 and NC2 represent Suzhou and Hangzhou, respectively, and TC represents Shanghai.

Table 5
A performance change of Dev-LSTM with varied parameters.

convolution kernel - number of LSTM nodes	RMSE	MAE	Corr
2*2-128	5.549	4.369	0.985
2*2-256	3.981	3.125	0.987
3*3-128	4.032	3.195	0.986
3*3-256	4.038	3.243	0.986
4*4-128	4.776	3.884	0.986
4*4-256	3.896	3.106	0.987
5*5-128	**3.789**	**2.982**	**0.988**
5*5-256	3.916	3.013	0.986
6*6-128	3.914	3.106	0.986

Table 6
Performance changes of Dev-LSTM under different time lengths.

D	N	RMSE	MAE	Corr
1h	**1h**	**3.789**	**2.982**	0.988
3h	1h	6.746	4.627	**0.989**
6h	1h	6.768	4.676	0.988
6h	2h	9.878	6.764	0.967
12h	1h	7.227	5.029	0.977
18h	1h	6.161	4.797	0.979
24h	12h	21.277	15.227	0.686
24h	24h	26.788	19.569	0.413

Table 7
Influence comparison in spatial dimension.

City	RMSE	MAE	Corr
TC	3.853	2.913	0.987
NC1+TC	3.906	3.087	0.987
NC2+TC	4.058	3.345	0.985
NC1+NC2+TC	**3.789**	**2.982**	**0.988**

As shown in Table 7, the model has the worst performance with only the target city TC. The reason is that the prediction model does not extract the correlation of multi-city and multi-site in the spatial dimension, and thus there are relatively large errors. NC1 represents Suzhou City, which is a little closer to Shanghai. Compared with NC2, i.e. Hangzhou, NC1 is a little farther away from Shanghai, and the prediction error of the model is smaller. This shows that the distance in the spatial dimension can have different degrees of impact on the concentration of pollutants in the target city. When the monitoring data of surrounding cities are incorporated into the prediction system as much as possible, the prediction error generated by the model is the smallest. Therefore, the comprehensive consideration of the impact of the surrounding cities on the pollutant concentration of the target city plays a very important role in improving the accuracy of the model prediction.

4.3.2. *Predictive model generalization ability*

To verify the generalization ability and effectiveness of the Dev-LSTM model, we applied Dev-LSTM to the prediction of pollutant concentrations in other cities in China. We use monitoring data from multiple cities to train the predictive model and test the generalization ability of the model in seven cities in Yangtze River Delta region and Beijing-tianjin-hebei region. The experimental results are shown in Table 8.

Table 8
Dev-LSTM's generalization ability.

City	RMSE	MAE	Corr
Shanghai	3.789	2.982	0.988
Suzhou	4.513	3.436	0.988
Hangzhou	5.818	4.277	0.951
Wuxi	4.262	3.397	0.988
Nantong	6.233	4.460	0.981
Jiaxing	6.116	5.036	0.974
Shaoxing	4.867	3.456	0.969

(a) The Yangtze River Delta region

City	RMSE	MAE	Corr
Beijing	6.870	5.125	0.977
Tianjin	11.286	7.992	0.985
Langfang	9.211	6.367	0.972
Zhangjiakou	5.930	4.758	0.930
Tangshan	17.122	13.240	0.892
Chengde	6.222	4.863	0.955
Baoding	16.850	12.411	0.959

(b) The Beijing-tianjin-hebei region

4.3.3. *Other experiments*

In addition, we conducted some experiments on the task of predicting $PM_{2.5}$ in the air in several cities in the future. The relevant experimental results are shown in Fig. 10 and Fig. 11.

In this paper, the surrounding cities of the target city are sorted according to the distance, and the influence of the number of surrounding cities introduced into the Dev-LSTM on the prediction results is verified. The abscissa of Fig. 10

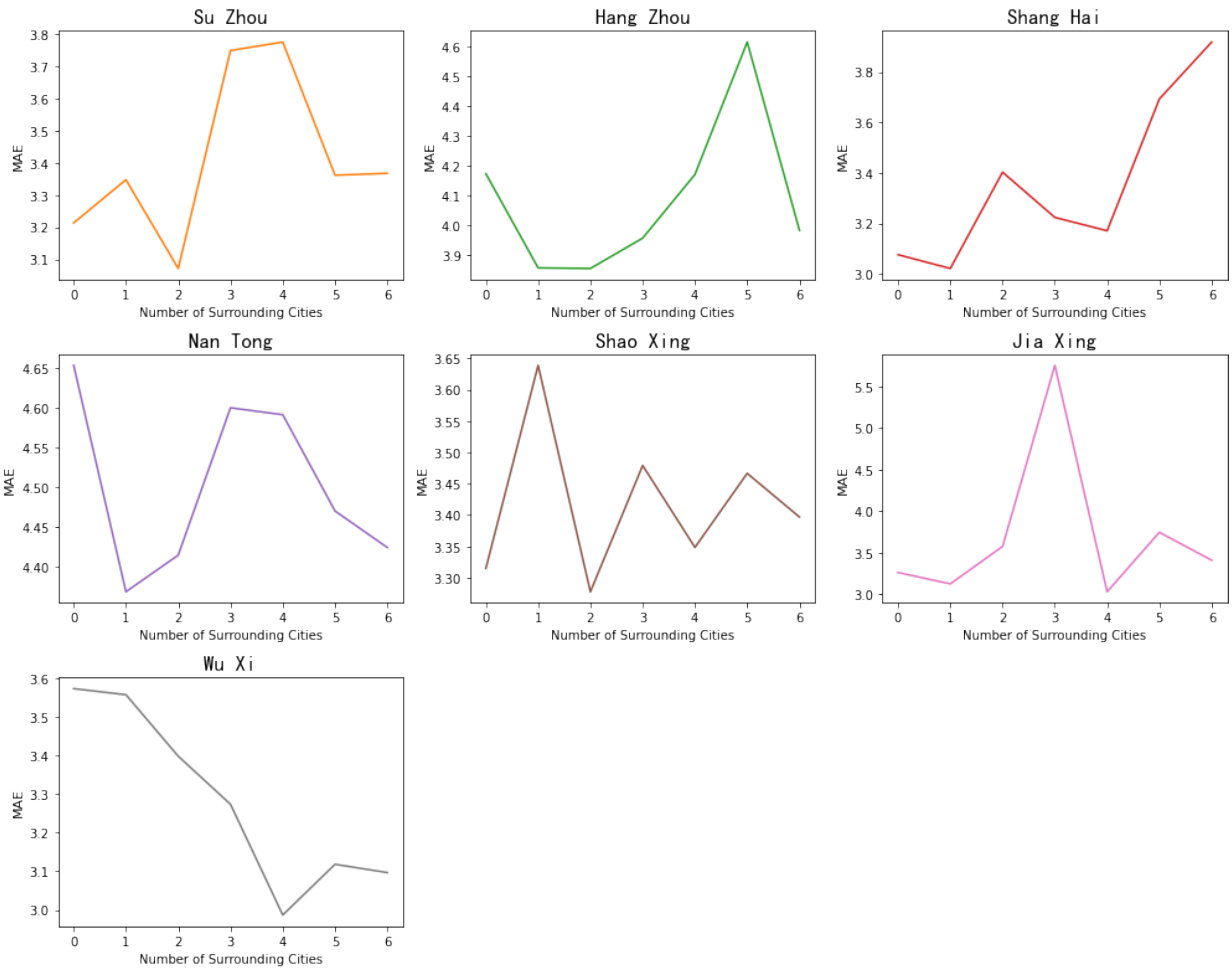

Fig. 10: The MAE of Dev-LSTM With a varied numbers of surrounding cities.

shows the number of data from surrounding cities, and the ordinate shows the MAE of the Dev-LSTM on the test set. As can be seen from Fig. 10, in the task of predicting the $PM_{2.5}$ content in the air in the next hour, the Dev-LSTM can achieve the best effect in almost all cities (except Wuxi) after introducing the data of 1-2 surrounding cities.

We also conducted some experiments on the impact of time lengths of target cities on the prediction results. The experiments on Dev-LSTM prediction in the next 1- 24 hours were conducted, and the performance of the Dev-LSTM is reflected by the similarity between the observed value and the predicted value. From Fig. 11, we can see that the model obtains the best results in predicting the $PM_{2.5}$ content in the air in the next 1-12 hours through the meteorological and pollutant data in the past 1-3 hours; However, if the forecast span is more than 12 hours, it should be considered to increase the length of the input past time with input data of 3-5 hours being a good choice.

5. Conclusion

In this paper we have presented Dev-LSTM, a prediction model that combines deconvolution neural network with LSTM to extract the spatial dimensional characteristics of pollutants, resolving the problem of excessive loss of information caused by traditional CNN. Dev-LSTM achieves an accurate prediction of air pollutant concentration in both spatial and temporal dimensions. Compared with traditional predictive models, Dev-LSTM shows superior predictive capabilities on multi-dimensional indicators.

Currently Dev-LSTM does not consider the geographic and geomorphic conditions contained in the spatial information. A future research will be to introduce more types of data to the Dev-LSTM prediction model to further enhance its performance in association extraction in the spatial dimension.

Acknowledgement

This work is funded by National Natural Science Foundation of China (61802258, 615723263), Natural Science

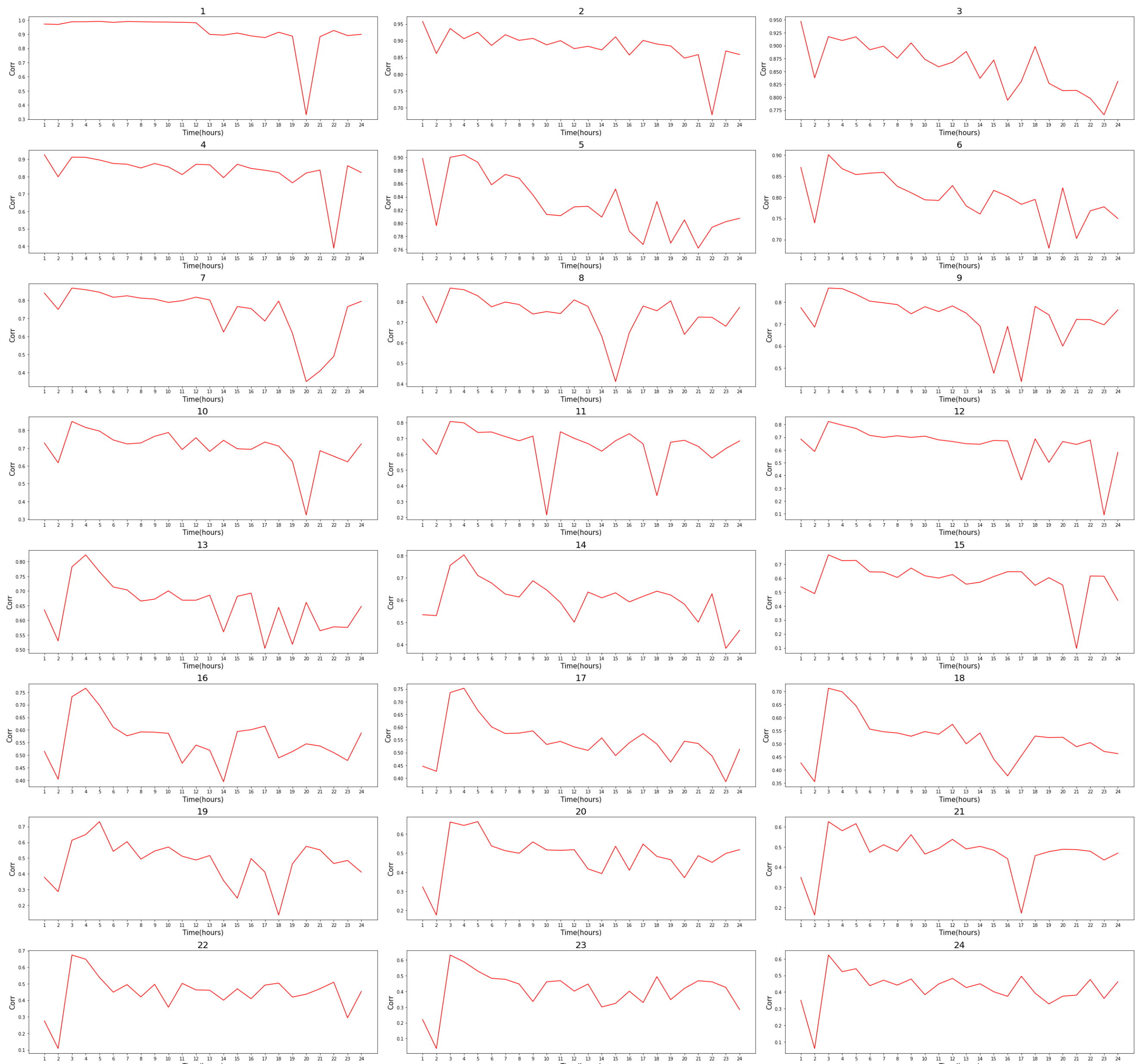

Fig. 11: The Corr of Dev-LSTM with a varied input time length.

Foundation of Shanghai (18ZR1428300, 20ZR1455600) and National Key Research and Development Program of China under Grant No.2022YFB4501704.

References

[1] Stephan Gabet, Clémentine Lemarchand, Pascal Guénel, and Rémy Slama. Breast cancer risk in association with atmospheric pollution exposure: a meta-analysis of effect estimates followed by a health impact assessment. *Environmental health perspectives*, 129(5):057012, 2021.

[2] Fengping Hu and Yongming Guo. Health impacts of air pollution in china. *Frontiers of Environmental Science & Engineering*, 15(4):1–18, 2021.

[3] Marco Travaglio, Yizhou Yu, Rebeka Popovic, Liza Selley, Nuno Santos Leal, and Luis Miguel Martins. Links between air pollution and covid-19 in england. *Environmental pollution*, 268:115859, 2021.

[4] Nicolas Huneeus, Hugo Denier van der Gon, Paula Castesana, Camilo Menares, Claire Granier, Louise Granier, Marcelo Alonso, Maria de Fatima Andrade, Laura Dawidowski, Laura Gallardo, et al. Evaluation of anthropogenic air pollutant emission inventories for south america at national and city scale. *Atmospheric Environment*, 235:117606, 2020.

[5] Cole Brokamp, Eric B Brandt, and Patrick H Ryan. Assessing exposure to outdoor air pollution for epidemiological studies: Model-based and personal sampling strategies. *Journal of Allergy and Clinical Immunology*, 143(6):2002–2006, 2019.

[6] Ling Han, Zhaobin Sun, Juan He, Yu Hao, Qiaoling Tang, Xiaoling Zhang, Canjun Zheng, and Shiguang Miao. Seasonal variation in health impacts associated with visibility in beijing, china. *Science of the Total Environment*, 730:139149, 2020.

[7] Akhilesh Kumar Yadav, Saba Shirin, Christina Emmanouil, and Aarif Jamal. Effect of seasonal and meteorological variability of air pollution in singrauli coalfield. *Aerosol Science and Engineering*,

6(1):61–70, 2022.
[8] WC Leong, RO Kelani, and Z Ahmad. Prediction of air pollution index (api) using support vector machine (svm). *Journal of Environmental Chemical Engineering*, 8(3):103208, 2020.
[9] KS Harishkumar, KM Yogesh, Ibrahim Gad, et al. Forecasting air pollution particulate matter (pm2. 5) using machine learning regression models. *Procedia Computer Science*, 171:2057–2066, 2020.
[10] Qi Liao, Mingming Zhu, Lin Wu, Xiaole Pan, Xiao Tang, and Zifa Wang. Deep learning for air quality forecasts: a review. *Current Pollution Reports*, 6(4):399–409, 2020.
[11] Bo Zhang, Hanwen Zhang, Gengming Zhao, and Jie Lian. Constructing a pm2. 5 concentration prediction model by combining auto-encoder with bi-lstm neural networks. *Environmental Modelling & Software*, 124:104600, 2020.
[12] Chen Zhang, Liping Di, Ziheng Sun, Li Lin, G Yu Eugene, and Juozas Gaigalas. Exploring cloud-based web processing service: A case study on the implementation of cmaq as a service. *Environmental Modelling & Software*, 113:29–41, 2019.
[13] Rui Feng, Han Gao, Kun Luo, and Jian-ren Fan. Analysis and accurate prediction of ambient pm2. 5 in china using multi-layer perceptron. *Atmospheric environment*, 232:117534, 2020.
[14] Mohammed G Ragab, Said J Abdulkadir, Norshakirah Aziz, Qasem Al-Tashi, Yousif Alyousifi, Hitham Alhussian, and Alawi Alqushaibi. A novel one-dimensional cnn with exponential adaptive gradients for air pollution index prediction. *Sustainability*, 12(23):10090, 2020.
[15] Zhongang Qi, Tianchun Wang, Guojie Song, Weisong Hu, Xi Li, and Zhongfei Zhang. Deep air learning: Interpolation, prediction, and feature analysis of fine-grained air quality. *IEEE Transactions on Knowledge and Data Engineering*, 30(12):2285–2297, 2018.
[16] Tongwen Li, Huanfeng Shen, Qiangqiang Yuan, Xuechen Zhang, and Liangpei Zhang. Estimating ground-level pm2. 5 by fusing satellite and station observations: a geo-intelligent deep learning approach. *Geophysical Research Letters*, 44(23):11–985, 2017.
[17] Xiang Li, Ling Peng, Xiaojing Yao, Shaolong Cui, Yuan Hu, Chengzeng You, and Tianhe Chi. Long short-term memory neural network for air pollutant concentration predictions: Method development and evaluation. *Environmental pollution*, 231:997–1004, 2017.
[18] Chiou-Jye Huang and Ping-Huan Kuo. A deep cnn-lstm model for particulate matter (pm2. 5) forecasting in smart cities. *Sensors*, 18(7):2220, 2018.
[19] Yann LeCun, Yoshua Bengio, and Geoffrey Hinton. Deep learning. *nature*, 521(7553):436–444, 2015.
[20] Jian Wang, Hengde Zhu, Shui-Hua Wang, and Yu-Dong Zhang. A review of deep learning on medical image analysis. *Mobile Networks and Applications*, 26(1):351–380, 2021.
[21] John Jumper, Richard Evans, Alexander Pritzel, Tim Green, Michael Figurnov, Olaf Ronneberger, Kathryn Tunyasuvunakool, Russ Bates, Augustin Žídek, Anna Potapenko, et al. Highly accurate protein structure prediction with alphafold. *Nature*, 596(7873):583–589, 2021.
[22] Fatemeh Mohammadi Shakiba, Milad Shojaee, S Mohsen Azizi, and Mengchu Zhou. Real-time sensing and fault diagnosis for transmission lines. *International Journal of Network Dynamics and Intelligence*, pages 36–47, 2022.
[23] Maciej Szankin and Alicja Kwasniewska. Can ai see bias in x-ray images? *International Journal of Network Dynamics and Intelligence*, pages 48–64, 2022.
[24] From emotion ai to cognitive ai. *International Journal of Network Dynamics and Intelligence*, 1:65–72, 2022.
[25] Deep common spatial pattern based motor imagery classification with improved objective function. *International Journal of Network Dynamics and Intelligence*, pages 73–84, 2022.
[26] Chang-Mei Liang, Yan-Wen Li, Yan-Hong Liu, Peng-Fei Wen, and Hua Yang. Segmentation and weight prediction of grape ear based on sfnet-resnet18. *Systems Science & Control Engineering*, 10(1):722–732, 2022.
[27] Shuqiao Geng, Changan Zhu, Yi Jin, Lichao Wang, and Hailong Tan. Gaze control system for tracking quasi-1d high-speed moving object in complex background. *Systems Science & Control Engineering*, 10(1):367–376, 2022.
[28] Peng Lu, Baoye Song, and Lin Xu. Human face recognition based on convolutional neural network and augmented dataset. *Systems Science & Control Engineering*, 9(sup2):29–37, 2021.
[29] Han Li, Peishu Wu, Nianyin Zeng, Yurong Liu, and Fuad E Alsaadi. A survey on parameter identification, state estimation and data analytics for lateral flow immunoassay: from systems science perspective. *International Journal of Systems Science*, 53(16):3556–3576, 2022.
[30] Ye Zhao, Xiao He, Lifeng Ma, and Hongjian Liu. Unbiasedness-constrained least squares state estimation for time-varying systems with missing measurements under round-robin protocol. *International Journal of Systems Science*, 53(9):1925–1941, 2022.
[31] Wojciech Samek, Grégoire Montavon, Sebastian Lapuschkin, Christopher J Anders, and Klaus-Robert Müller. Explaining deep neural networks and beyond: A review of methods and applications. *Proceedings of the IEEE*, 109(3):247–278, 2021.
[32] Chen Ding, Guizhi Wang, Xinyue Zhang, Qi Liu, and Xiaodong Liu. A hybrid cnn-lstm model for predicting pm2. 5 in beijing based on spatiotemporal correlation. *Environmental and Ecological Statistics*, 28(3):503–522, 2021.
[33] Nianyin Zeng, Peishu Wu, Zidong Wang, Han Li, Weibo Liu, and Xiaohui Liu. A small-sized object detection oriented multi-scale feature fusion approach with application to defect detection. *IEEE Transactions on Instrumentation and Measurement*, 71:1–14, 2022.
[34] Han Li, Peishu Wu, Zidong Wang, Jingfeng Mao, Fuad E Alsaadi, and Nianyin Zeng. A generalized framework of feature learning enhanced convolutional neural network for pathology-image-oriented cancer diagnosis. *Computers in Biology and Medicine*, 151:106265, 2022.
[35] Stella Moisan, Rodrigo Herrera, and Adam Clements. A dynamic multiple equation approach for forecasting pm2. 5 pollution in santiago, chile. *International Journal of Forecasting*, 34(4):566–581, 2018.
[36] Samsuri Abdullah, Marzuki Ismail, Ali Najah Ahmed, and Ahmad Makmom Abdullah. Forecasting particulate matter concentration using linear and non-linear approaches for air quality decision support. *Atmosphere*, 10(11):667, 2019.
[37] Zhiyuan Wu, Ning Liu, Guodong Li, Xinyu Liu, Yue Wang, and Lin Zhang. A variational bayesian approach for fast adaptive air pollution prediction. In *2021 IEEE International Conference on Big Data (Big Data)*, pages 1748–1756. IEEE, 2021.
[38] Yousif Alyousifi, Kamarulzaman Ibrahim, Wei Kang, and Wan Zawiah Wan Zin. Markov chain modeling for air pollution index based on maximum a posteriori method. *Air Quality, Atmosphere & Health*, 12(12):1521–1531, 2019.
[39] Hehua Zhang and Yuhong Zhao. Land use regression for spatial distribution of urban particulate matter (pm10) and sulfur dioxide (so2) in a heavily polluted city in northeast china. *Environmental monitoring and assessment*, 191(12):1–14, 2019.
[40] Tuo Shi, Yuanman Hu, Miao Liu, Chunlin Li, Chuyi Zhang, and Chong Liu. Land use regression modelling of pm2. 5 spatial variations in different seasons in urban areas. *Science of the Total Environment*, 743:140744, 2020.
[41] Kevin P Murphy. *Machine learning: a probabilistic perspective*. MIT press, 2012.
[42] Xintong Li and Xiaodong Zhang. Predicting ground-level pm2. 5 concentrations in the beijing-tianjin-hebei region: a hybrid remote sensing and machine learning approach. *Environmental pollution*, 249:735–749, 2019.
[43] Rui Zhao, Liping Zhan, Mingxing Yao, and Linchuan Yang. A geographically weighted regression model augmented by geodetector analysis and principal component analysis for the spatial distribution of pm2. 5. *Sustainable Cities and Society*, 56:102106, 2020.
[44] Sechan Park, Minjeong Kim, Minhae Kim, Hyeong-Gyu Namgung, Ki-Tae Kim, Kyung Hwa Cho, and Soon-Bark Kwon. Predicting

pm10 concentration in seoul metropolitan subway stations using artificial neural network (ann). *Journal of hazardous materials*, 341:75–82, 2018.

[45] Ho Chang-Hoi, Ingyu Park, Hye-Ryun Oh, Hyeon-Ju Gim, Sun-Kyong Hur, Jinwon Kim, and Dae-Ryun Choi. Development of a pm2. 5 prediction model using a recurrent neural network algorithm for the seoul metropolitan area, republic of korea. *Atmospheric Environment*, 245:118021, 2021.

[46] Peishu Wu, Zidong Wang, Baixun Zheng, Han Li, Fuad E Alsaadi, and Nianyin Zeng. Aggn: Attention-based glioma grading network with multi-scale feature extraction and multi-modal information fusion. *Computers in Biology and Medicine*, 152:106457, 2023.

[47] Tom Brown, Benjamin Mann, Nick Ryder, Melanie Subbiah, Jared D Kaplan, Prafulla Dhariwal, Arvind Neelakantan, Pranav Shyam, Girish Sastry, Amanda Askell, et al. Language models are few-shot learners. *Advances in neural information processing systems*, 33:1877–1901, 2020.

[48] Zhe Xu and Yi Lv. Att-convlstm: Pm 2.5 prediction model and application. In *Advances in Natural Computation, Fuzzy Systems and Knowledge Discovery: Volume 1*, pages 30–40. Springer, 2020.

[49] Yuan-Yuan Zhu, Yu-Xiao Gao, Bing Liu, Xiao-Yan Wang, Li-Li Zhu, Rong Xu, Wei Wang, Jun-Nan Ding, Jian-Jun Li, and Xiao-Li Duan. Concentration characteristics and assessment of model-predicted results of pm 2.5 in the beijing-tianjin-hebei region in autumn and winter. *Huan Jing ke Xue= Huanjing Kexue*, 40(12):5191–5201, 2019.

[50] Jianjun Chen, Jin Lu, Jeremy C Avise, John A DaMassa, Michael J Kleeman, and Ajith P Kaduwela. Seasonal modeling of pm2. 5 in california's san joaquin valley. *Atmospheric environment*, 92:182–190, 2014.

[51] Hongbin Dai, Guangqiu Huang, Huibin Zeng, and Fangyu Zhou. Pm2. 5 volatility prediction by xgboost-mlp based on garch models. *Journal of Cleaner Production*, 356:131898, 2022.

[52] Billy Peralta, Tomás Sepúlveda, Orietta Nicolis, and Luis Caro. Space-time prediction of pm2. 5 concentrations in santiago de chile using lstm networks. *Applied Sciences*, 12(22):11317, 2022.

[53] Zhe Xu and Yi Lv. Att-convlstm: Pm 2.5 prediction model and application. In *Advances in Natural Computation, Fuzzy Systems and Knowledge Discovery: Volume 1*, pages 30–40. Springer, 2020.

[54] Marjan Faraji, Saeed Nadi, Omid Ghaffarpasand, Saeid Homayoni, and Kay Downey. An integrated 3d cnn-gru deep learning method for short-term prediction of pm2. 5 concentration in urban environment. *Science of The Total Environment*, 834:155324, 2022.