

## Research Article

# An Air Quality Prediction Model Based on a Noise Reduction Self-Coding Deep Network

Jianxian Cai , Xun Dai , Li Hong, Zhitao Gao, and Zhongchao Qiu

*College of Electronic Science and Control Engineering, Institute of Disaster Prevention, Sanhe 065201, China*

Correspondence should be addressed to Jianxian Cai; [jianxian.cai@hotmail.com](mailto:jianxian.cai@hotmail.com)

Received 12 February 2020; Accepted 3 April 2020; Published 15 May 2020

Academic Editor: Peng-Yeng Yin

Copyright © 2020 Jianxian Cai et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Aiming at remedying the problem of low prediction accuracy of existing air pollutant prediction models, a denoising autoencoder deep network (DAEDN) model that is based on long short-term memory (LSTM) networks was designed. This model created a noise reduction autoencoder with an LSTM network to extract the inherent air quality characteristics of original monitoring data and to implement noise reduction processing on monitoring data to improve the accuracy of air quality predictions. The LSTM network structure in the DAEDN model was designed as bidirectional LSTM (Bi-LSTM) to solve the problem of a lag in the unidirectional LSTM prediction results and thereby to further improve the prediction accuracy of the prediction model. Using air pollutant time series data, the DAEDN model was trained using hourly  $PM_{2.5}$  concentration data collected in Beijing over 5 years. The experimental results show that the DAEDN model can extract more stable features from the noisy input after training was completed. The models were evaluated using RMSE and MAE, and the results show that the indexes are 15.504 and 6.789; compared with unidirectional LSTM, it is reduced by 7.33% and 5.87%, respectively. In addition, the new prediction model essentially considered the time series properties of the prediction of the concentration of spatial pollutants and the fully integrated environmental big data, such as air quality monitoring, meteorological monitoring, and forecasting.

## 1. Introduction

Air quality prediction is highly significant to any government's emergency management of severely polluted weather. Predictions not only warn the public to reasonably avoid highly polluted weather but also provide time for the government to implement appropriate emergency measures to mitigate atmospheric pollution, such as limiting the production and emissions of heavily polluting enterprises and restricting motor vehicles [1]. At the same time, air quality forecasting is an effective technical means to implement scientific decision-making and comprehensively manage the environment in an effort to strengthen air pollution prevention and control, and it provides an important way to quickly convert relevant environmental monitoring information into a basis for air pollution prevention and decision-making. For those reasons, air pollution prediction is highly valued by the state. In accord with the requirements of China's State Council's Notice on

Printing and Distributing Action Plan for Air Pollution Prevention and Control (Guo Fa [2013] No. 37), the Beijing-Tianjin-Hebei, Yangtze River Delta, and Pearl River Delta regions were established in 2014. Construction of regional, provincial, and municipal levels of heavy-pollution weather monitoring and early warning systems were to be completed by the end of that year. Other provinces (autonomous regions and municipalities), subprovincial cities, and capital cities were to be completed by the end of 2015. As a core function of heavy-pollution weather monitoring and early warning systems, air quality prediction has an important influence on the function of the entire system. However, air quality prediction is a complex, systematic undertaking, and improving the accuracy of predictions is an urgent and difficult problem in the field of air pollution prevention.

A goal of air quality prediction is to predict the degree of air pollution in an area for the next day, basing that prediction on past air pollutant emissions and meteorological conditions, atmospheric diffusion, and geographical features

[2]. There are many pollutants in the atmosphere. Among them,  $\text{SO}_2$ ,  $\text{NO}_2$ ,  $\text{CO}_2$ ,  $\text{NO}$ ,  $\text{CO}$ , and fine particulate matter ( $\text{PM}_{2.5}$  and  $\text{PM}_{10}$ ) are very important pollutants, so the urgency of predicting air pollution is mainly to predict these six pollutants. Research on air quality prediction began in the 1960s, and in the beginning, there was no way to achieve quantitative prediction of atmospheric pollutants. In the 1980s, people began to quantify pollutants by using mathematical statistical prediction methods and numerical analyses. The statistical prediction method [3], which uses a mathematical technique combining factor analysis and regression analysis to replace the physical, chemical, and biological process prediction methods, has three significant shortcomings: (1) the accuracy level of its prediction is based on whether there is a sufficient number of detailed and true historical weather data; (2) analysis of long-term historical monitoring data takes significant time, energy, and financial resources if sufficient data are available; and (3) extreme weather conditions often cannot be accurately predicted because of sudden events such as sandstorms, tornadoes, and thunderstorms. The numerical prediction method is a scientific, effective method that mathematically models the change law of atmospheric pollutants and uses mathematical models to approximate the changes in the trend of pollutants. However, the method's implementation process is relatively complicated and the efficiency is not high.

In the 21st century, the era of information storms and big data has arrived. Data collection is no longer a technical problem, so an air quality prediction method based on machine learning is proposed. Machine learning uses a type of algorithm that does not provide accurate results but does produce approximate solutions [4]. Common machine learning algorithms are (1) classification and regression, (2) clustering, (3) recommended algorithms, (4) association rules, and (5) artificial neural networks. Air quality prediction methods based on machine learning algorithms have overcome some of the shortcomings of the older statistical prediction methods and numerical predictions mentioned above and have become the mainstream of air quality prediction research. So far, air quality prediction methods based on machine learning have achieved some good results. Sahafzadeh and Ahmadi [5] took historical data on air quality in Bushehr City from 1951 to 2013, and using the k-mean algorithm, they constructed the city's air quality prediction model. Athanasiadis et al. [6] focused on the application of classification algorithms to air quality prediction. Although people have continuously tried to apply classification algorithms, clustering algorithms, and logistic regression algorithms in machine learning algorithms for air quality prediction, air quality is affected by a variety of geographical conditions, human activities, and the atmospheric environment, making it a complex, multidimensional, large-scale system that is driven by multiple feature factors, and the relationships between those feature factors are intricate and complex. Even more important is the fact that air quality has extremely significant nonlinear characteristics. Therefore, an increasing number of experts and scholars have begun to focus their research on air quality prediction by using artificial neural networks that are

inherently good at dealing with nonlinear problems in machine learning algorithms and that have strong noise tolerance.

With the continuous deepening of relevant research, various prediction models based on neural network technology have been constructed, and artificial neural network technology has achieved an irreplaceable position in the field of atmospheric pollutant prediction and has become a hot research topic [7–10]. Azid and others [11] combined principal component analysis and neural networks to establish a prediction model for the Malaysian Air Quality Index (AQI). Mishra and others [12] used multiple linear regression analysis and artificial neural networks to predict  $\text{PM}_{2.5}$  concentrations in New Delhi, India. That experiment proved that the prediction results of neural networks are superior. Neural networks have strong nonlinear fitting capabilities and can map complex nonlinear relationships. However, as the number of layers of neural network increases, the gradient descent algorithm may converge to a local minimum, with a resulting error that leads to a result ratio. Shallow networks are even worse, and neural networks have additional shortcomings, such as overfitting, poor generalization ability, slow convergence speed, and low prediction accuracy [13].

In 2015, three leading figures in the field of machine learning, Yann LeCun, Yoshua Bengio, and Geoffrey Hinton, published a landmark article titled "Deep Learning" in the journal *Nature* [14]. Deep learning technology has since evolved into the current field of artificial intelligence. One very active research focus in China has shown huge advantages in the fields of image recognition and speech recognition, and it continues to develop and change. Deep learning can be an effective method for big data processing by training big data, mining it, and capturing deep connections within big data, thus improving classification and prediction accuracy. In addition, the deep learning model includes faster training, and with an increase of training samples, it can show better performance growth than the general method does. Practice has proved that air pollution prediction models based on deep learning can better overcome the shortcomings of existing prediction methods for three reasons. (1) In recent years, with China's increased attention and investment in environmental monitoring, a large number of air pollutants have been monitored in real time. Data have been accumulated over a long period of time, including air pollutant concentrations and meteorological conditions. In the context of environmental big data, deep learning technology can integrate massive, multisource environmental protection data and can use sufficient observational data as training samples to ensure that the deep-learning-based air pollution prediction model has high accuracy [15]. (2) The deep learning model can deeply explore the inherent data relationships among the factors that affect pollutant concentrations and can establish a more accurate proxy model of a complex mechanism model between air pollutant concentrations and impact factors. Deep mining extracts advanced and semantic patterns and rules of air quality changes and organically integrates multiple models and expertise to achieve effective air quality analyses

[16]. (3) The deep learning model has strong scalability. By properly setting the input factors, other methods can be integrated into the model, which then can avoid the defects and uncertainty of the single air pollution prediction model to a certain extent and can improve the accuracy of predictions. In short, with the establishment of a large number of air quality detection systems, the relevant data have gradually become richer, and that advantage makes it possible to use deep learning technology to predict air quality [17]. Deep learning that is based on deep neural network models has a high order of magnitude and complexity. Larger data show a greater advantage than traditional machine learning methods and overcome the shortcomings of traditional machine learning methods in model building and feature extraction. Many scholars have begun to invest in research and have achieved good results. Zheng Yi and Zhu Chengzhang [18] applied a deep belief network (DBN) to the prediction of regional  $PM_{2.5}$  daily average data. By optimizing the DBN network parameter settings, their experimental results were compared with a backpropagation (BP) neural network and radial basis function (radial basis function or RBF) model prediction results. The DBN-based prediction method could better predict the daily average change trend of  $PM_{2.5}$  in the region, and the prediction accuracy significantly improved. Dong Ting et al. [19] proposed an AQI prediction method based on space-time optimized input stacked denoising autoencoders (SDAEs), and the SDAE model had better prediction performance when compared with other models. Yin Wenjun et al. [20] aimed at remedying the shortcomings of traditional statistical methods and artificial neural network models in the prediction of urban AQI indexes in the context of big data, by proposing a DBN-based prediction method that obtained hierarchical data feature representations from model training. Such predictions have been more instructive than traditional methods. For the case of air quality prediction using a shallow neural network model when the prediction result is not good, Xiang Li et al. [21] adopted a spatiotemporal stacked autoencoder (SAE) to extract air quality data features, using greedy training and working with spatiotemporal neural network (STNN) and autoregressive and moving average (ARMA) models. A comparison between their model and the support vector regression (SVR) model showed that the SAE had superior performance. Because it considers the correlation between space and time, it can simultaneously predict the air quality of different monitoring points, and it reflects the temporal stability of air quality in each season. Bun Theang Ong, Komei Sugiura, and Koji Zettsu [22] proposed an automatic encoder-based deep recurrent neural network (AE-DRNN) model based on environmental data collected by sensors and used it to predict  $PM_{2.5}$  concentrations in Japan. With the help of a sparsity of autoencoders (AEs), the model was pretrained and the data features were extracted, and then the DRNN completed the prediction. The results showed that the prediction for time series was better than that of the conventional AE model. Zhang and Ding [23] proposed a method based on an extreme learning machine (ELM) to predict air pollutant concentrations, and it overcame the

feedforward artificial neural networks (FFANNs), had convergent convergence, and was easily trapped in local extremes, so the ELM further improved prediction accuracy, robustness, and generalization. Fan Junxiang and others [24] constructed a deep neural network composed of long short-term memory (LSTM) layers and fully connected layers. Their model was trained using air quality data and meteorological data from the Beijing-Tianjin-Hebei region, and the results performed better than traditional deep recurrent neural networks (DRNNs) did, thereby confirming the effectiveness of the deep learning framework in spatiotemporal predictions. Liu Bingchun et al. [25] first decomposed the historical time series of daily air pollutant concentrations into different frequencies by wavelet decomposition and recombined them into a high-dimensional training data set; subsequently, the high-dimensional data set was used to train the LSTM prediction model, and repeated experiments adjusted the parameters to obtain the optimal prediction model. Those research results show that the combined model had a higher prediction accuracy and stability for predicting pollutant concentration than the traditional LSTM model did.

Although deep neural networks have been applied successfully in air quality modelling, it has some shortcomings. First, there is noise in air quality data and meteorological monitoring data, and existing air quality prediction methods are very sensitive to noise, which affects to a certain extent the accuracy of their predictions. Second, the theory and learning algorithms still contain many intractable problems. The biggest challenge is the problem of the time-consuming training phase [26], and the current solutions include improving the learning parameters through reasonable selection. The convergence speed of a deep network [27], using a hardware accelerator based on the graphic processing unit (GPU), has been applied to the algorithm operation and has achieved significant acceleration convergence effects [28–31], but the hardware equipment cost and maintenance costs are too high, making it uneconomical, and it does not improve the convergence speed from the perspective of the algorithm. In tandem with the era of big data, the amount of information for processing data will increase exponentially, and traditional deep networks cannot quickly converge or even complete learning tasks. Therefore, one direction for the future development of deep networks lies in quickly and economically completing the full learning of large amounts of data [32].

The LSTM networks have a positive effect on the prediction of time series signals, so they are suitable for air quality prediction. For this study, based on an LSTM network, a denoising autoencoder deep network (DAEDN) model was designed to solve the low prediction accuracy of existing air pollutant prediction models. This model designed a noise reduction autoencoder with an LSTM structure to extract the inherent air quality characteristics of the original monitoring data and to implement noise reduction processing on the monitoring data to improve the accuracy of air quality prediction. The LSTM networks in the DAEDN model were all designed as a two-way structure to solve the problem of lagging in one-way LSTM prediction

results and further improve the prediction accuracy of air quality prediction models. Using the hourly PM<sub>2.5</sub> concentration data collected by Beijing's 12 air quality-monitoring stations over 5 years, the prediction model of the study was analyzed and verified.

## 2. Methods

The DAEDN model designed in this study was based on the use of time series data of air pollutants as experimental data. The network structure was extended based on the encoder-decoder framework. The structure of the study's DAEDN model is shown in Figure 1. In the figure, the green solid line frame contains the main input of the model, including historical air quality data, primarily the air quality index AQI and PM<sub>2.5</sub>, PM<sub>10</sub>, SO<sub>2</sub>, NO<sub>2</sub>, O<sub>3</sub>, CO, and other pollutant concentrations. The yellow dotted frame contains the auxiliary input time data, along with the time data taken from 1-hour intervals, which corresponded to the historical air quality data one by one. The input layer used a denoising autoencoder (DAE) to extract the inherent air quality characteristics of the original monitoring data to achieve noise reduction processing on the monitoring data and thus to improve the model's prediction accuracy. At the same time, the internal structure of the DAE was designed as a Bi-LSTM (bidirectional long short-term memory or Bi-LSTM) network that was used to solve the problem of the lag in the prediction results of the unidirectional LSTM structure and thus to improve the prediction accuracy of the model. The middle layer was a fully connected layer that combined the air quality features extracted by the input layer. The output layer still used the Bi-LSTM structure to generate a predicted output of air quality.

The input layer used a DAE structure that was based on an extension of the autoencoder and added noise to the input data, based on that AE. The AE was a three-layer unsupervised neural network. By extracting the most important features that could represent the input data, the output reproduced the input signal as much as possible. The AE consisted of an encoder, a hidden layer, and a decoder. The encoder converted the input data from being high dimensional to low dimensional, in order to extract the data's features, and the decoder converted the data back from low-dimensional to high-dimensional outputs, thus verifying whether the extracted features could represent the input data well. The ultimate goal of the AE training process was to minimize reconstruction errors, which meant essentially reducing the difference between the input data and the expression of its features. The network structure is shown in Figure 2.

The functional relationship between the input layer of the encoder and the hidden layer can be expressed as

$$y = s(W_1x + b_1), \quad (1)$$

where  $s(x)$  is the encoder activation function,  $W_1$  is the adjacent node weight,  $b_1$  is the adjacent node offset,  $x$  is the input layer data, and  $y$  is the hidden layer data. The above

formula is the encoding process, and the decoding process can be expressed as

$$z = s(W_2y + b_2), \quad (2)$$

where  $s(y)$  is the decoder activation function,  $W_2$  is the adjacent node weight,  $b_2$  is the adjacent node offset, and  $z$  is the output layer data.

The structure of the DAE is shown in Figure 3. The DAE added noise to the input data on the basis of the AE, which was to randomly erase certain nodes of the input layer with a certain probability distribution. At that point, the encoder automatically learned to remove the noise, thereby obtaining an input signal that was not contaminated by noise. The trained encoder with a noise reduction function could extract more robust features from the noisy input; that is, the input data  $x$  became  $x'$ , which improved the pan of the input data by the self-encoding neural network model and increased the model's ability to improve data processing accuracy.

The DAE internal structure of the input layer was designed as a bidirectional LSTM. Commonly, DAEs are based on fully connected neural networks, and it is relatively rare to use a bidirectional LSTM network.

The structure of the LSTM unit is shown in Figure 4. The LSTM is a variant of recurrent neural networks (RNNs). Although in theory a RNN can handle any long-distance dependence problem, in reality, that is difficult to achieve due to problems such as gradient disappearance and explosion. The LSTM provided a solution by introducing a gate mechanism and a memory unit by replacing the hidden layer neural unit in the RNN with an LSTM unit.

The historical information stored in the LSTM was controlled by the input gate, the forget gate, and the output gate, which were calculated as follows:

$$\begin{aligned} \tilde{C}_t &= \tan h(W_C \cdot [h_{t-1}, x_t] + b_C), \\ i_t &= \sigma(W_i \cdot [h_{t-1}, x_t] + b_i), \\ f_t &= \sigma(W_f \cdot [h_{t-1}, x_t] + b_f), \\ C_t &= f_t \otimes \tilde{C}_t + i_t \otimes \tilde{C}_t, \\ h_t &= O_t \cdot \tan h(C_t). \end{aligned} \quad (3)$$

Here,  $x_t$  is the input data at time  $t$ ,  $h_t$  is the output state value of the LSTM unit at time  $t$ ,  $\tilde{C}_t$  is the candidate value of the memory unit at time  $t$ ,  $i_t$  is the state value at time  $t$  of the input gate, and  $f_t$  is the state at time  $t$  of the forget gate. Of the values,  $W$  is the corresponding weight,  $b$  is the corresponding paranoid parameter, and  $\otimes$  represents the dot product between the elements and is multiplied point by point. The state value of the memory unit was adjusted by the input gate and the forget gate.

However, for an input sequence, at a time node  $t$ , an LSTM network only contains information before  $t$  and does not contain information after  $t$ . That situation will cause the problem of the lag in the prediction result and lose the function of real-time prediction in practical applications. To solve that problem, the model in this study used a Bi-LSTM network. The Bi-LSTM network structure is shown in Figure 5.

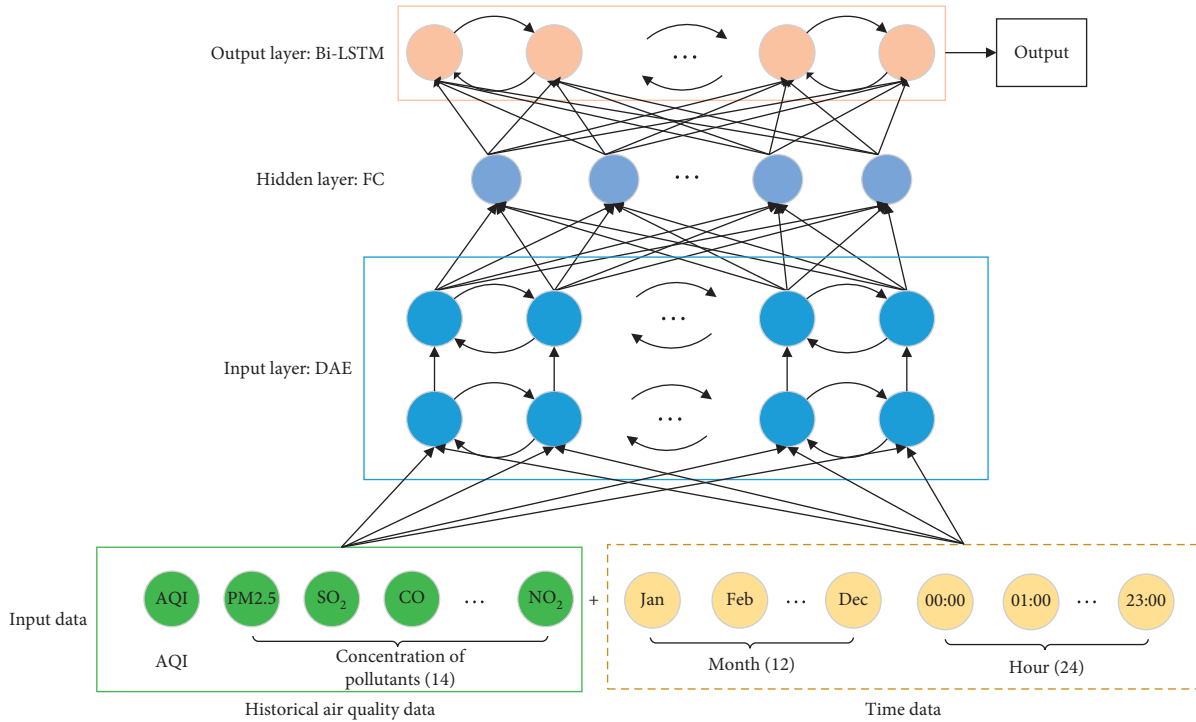


FIGURE 1: The DAEDN model.

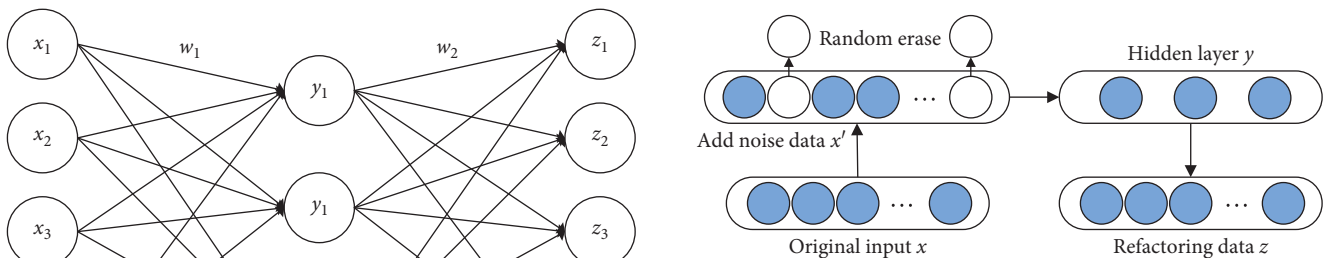


FIGURE 2: The AE network.

The network structure included a forward LSTM and a reverse LSTM. The forward LSTM obtained a sequence  $h_a$  according to the normal input. The reverse LSTM reversed the input and then passed through a network with the same structure as the forward LSTM but with different weight parameters. Finally, it obtained a sequence and then reversed that sequence to obtain  $h_b$ . Ultimately, the two sequences were added to obtain  $H$ , which was the final result through the Bi-LSTM network:

$$H = h_a \oplus h_b. \quad (4)$$

A Bi-LSTM can simultaneously use the historical and future information in the sequence, divide the sequence

information into two directions for input into the model, use two hidden layers to save the input information in both directions, and connect the corresponding outputs of the hidden layers to the same output layer. The two structures are the same and independent of each other, but they only accept different sequence inputs. Therefore, the final hidden layer vector contains the data of the positive and negative time series of the data set, which solves the problem of the lag of the prediction result in the unidirectional LSTM and improves the accuracy of the air quality predictions.

### 3. Evaluation Index and Data Preprocessing

**3.1. Evaluation Index.** The research background of this study was the air pollution index, and the type of problem was a regression (a real value prediction of air pollutant concentration). The data sets used in the experiments were real data sets. For such a data set, it is possible to make a certain accuracy prediction based on life experience, even without training the model, so it obviously was inappropriate to use

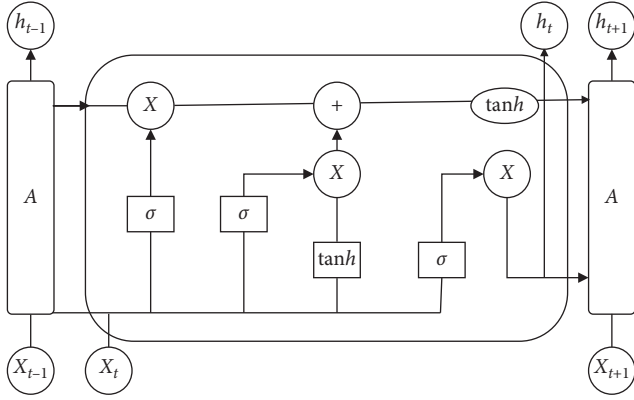


FIGURE 4: The LSTM cell.

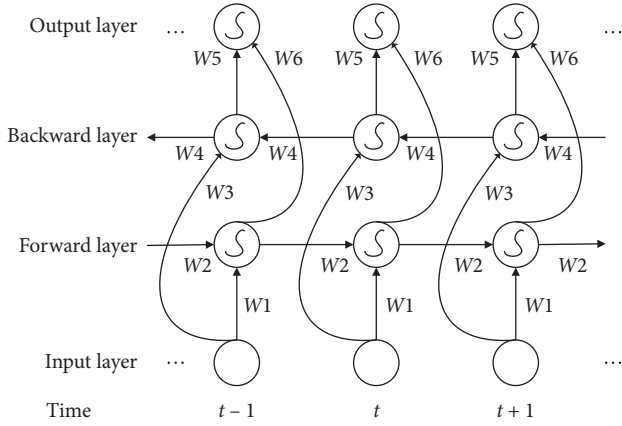


FIGURE 5: The Bi-LSTM network.

the accuracy rate to judge the performance of the prediction model. In this study, when training the model, the root-mean-square error (RMSE) and the mean absolute error (MAE) were used as the evaluation criteria for prediction accuracy. The RMSE and MAE calculations are shown in the following equations:

$$\text{RMSE} = \sqrt{\frac{1}{n} \sum_{i=1}^n (x_i - x'_i)^2}, \quad (5)$$

$$\text{MAE} = \frac{1}{n} \sum_{i=1}^n |x_i - x'_i|. \quad (6)$$

In the above two formulas,  $n$  is the data length, that is, the number of hours in the test set,  $x_i$  is the true value of the air pollution index at the  $i$ th hour, and  $x'_i$  is the predicted value of the air pollution index at the  $i$ th hour.

**3.2. Data and Modelling.** The data set used in this paper comes from Beijing Municipal Environmental Monitoring Center (<http://www.bjmemc.com.cn/>). The area targeted in this study was Beijing, and the data sources were divided into three categories: (1) air quality-monitoring data, (2) pollutant concentration-monitoring data were  $\text{PM}_{2.5}$ ,  $\text{PM}_{2.5\ 24\text{h}}$ ,  $\text{PM}_{10}$ ,

$\text{PM}_{10\ 24\text{h}}$ ,  $\text{SO}_2$ ,  $\text{SO}_{2\ 24\text{h}}$ ,  $\text{NO}_2$ ,  $\text{NO}_{2\ 24\text{h}}$ ,  $\text{O}_3$ ,  $\text{O}_{3\ 24\text{h}}$ ,  $\text{O}_{3\ 8\text{h}}$ ,  $\text{O}_{3\ 8\text{h}\ 24\text{h}}$ ,  $\text{CO}$ , and  $\text{CO}_{24\text{h}}$ , and (3) time data were month and hour.

The source data were updated every hour. The air quality-monitoring data were the air quality index AQI, and the gas concentration-monitoring data were  $\text{PM}_{2.5}$  concentration,  $\text{PM}_{10}$  concentration,  $\text{SO}_2$  concentration,  $\text{NO}_2$  concentration,  $\text{O}_3$  concentration, and  $\text{CO}$  concentration. The time span of the data was from May 13, 2014, to December 31, 2018, a total of 4 years and 7 months. The data from May 13, 2014, to December 31, 2017, were selected as the training data set, the data from January 1, 2018, to June 30, 2018, were used as the validation data set, and the data from July 1, 2018, to December 31, 2018, were used as the test data set. The data sets are not intersected with each other, which can effectively achieve the goal of continuous optimization of model training.

**3.3. Data Preprocessing.** Data loss may occur in the data set. Without changing the structure of the neural network, the average value of the data in the same period was used to replace the missing value, as shown in the following equations:

$$x_t^d \leftarrow m_t^d x_t^d + (1 - m_t^d) \bar{x}^d, \quad (7)$$

$$m_t^d = \begin{cases} 1, & x_t^d \text{ valid} \\ 0, & x_t^d \text{ loss} \end{cases}, \quad (8)$$

$$\bar{x}^d = \frac{\sum_{t'=1}^T m_{t'}^d x_{t'}^d I_{\text{hour}}(s_t, s_{t'})}{\sum_{t'=1}^T m_{t'}^d I_{\text{hour}}(s_t, s_{t'})}, \quad (9)$$

$$I_{\text{hour}}(s_t, s_{t'}) = \begin{cases} 1, & s_t = s_{t'} \\ 0, & s_t \neq s_{t'} \end{cases}. \quad (10)$$

In the above four formulas,  $t' (< t)$  represents the time step in which the  $d$ -dimensional component was recently observed,  $\bar{x}^d$  represents the mean value of the  $d$ -dimensional component of the observation at the current time in the same month,  $x_t^d$  represents the current valid observation of the component, and  $s_t$  represents the time corresponding to the  $t$ -th time step ( $s_1 = 0$ ).

Many factors affect air quality, and each factor has its own physical properties and dimensions. Direct analysis of those factors will affect the accuracy of the results, so to facilitate network training and prevent problems such as “overfitting” in the calculation process, the original data needed to be normalized first to put the different impact factors in the same order of magnitude for more accurate data analysis. This study used the min-max normalization method, which was to perform a linear transformation on each attribute of the original data. After normalization, the data were between (0, 1). The normalized function was as in the following equation:

$$x' = \frac{x - \min}{\max - \min}, \quad (11)$$

where  $x$  is the data before normalization,  $x'$  is the data after normalization,  $\min$  is the minimum value of all data in the influence factor to which  $x$  belongs, and  $\max$  is the maximum value of all data in the influence factor to which  $x$  belongs.

After normalizing the training set, the test data should also be standardized in the same way, so as to ensure that the test data and the training set are scaled at the same proportion. However, most air quality and meteorological data values do not have exact boundaries. For individual test data that are smaller than the minimum or larger than the training set, in order to normalize the data to fall within the interval (0, 1), one must add the following restrictions on the basis of the equation

$$x' = \begin{cases} 0, & x < \min \\ 1, & x > \max \end{cases} \quad (12)$$

#### 4. Experimental Results and Analysis

By using the DAEDN model in Figure 1, we built an experimental process framework, as shown in Figure 6. The process included three primary steps: (1) data preprocessing, to process missing values and other issues in the original data set to complete the original data; (2) data fusion, based on the time series distribution of the data, formatting the data, adding time steps, and then generating time series data for training and testing the model; and (3) training and evaluating the model, using the generated input data to train the DAEDN model, and using the RMSE and MAE to evaluate the prediction effectiveness of the network.

This study used the programming language Python and the deep learning libraries TensorFlow and Keras to build and train the deep network models. The Python packages used are NumPy, Matplotlib, Pandas, SciKit-Learn, etc. In comparison experiments, different network structure parameters were unified.

First, we set the number of neurons in the input and output layers of the model to 17 and 1, respectively. The input is a time series data with 17-dimensional features that can be summarized in three categories: (1) air quality-monitoring data were AQI, (2) pollutant concentration-monitoring data were  $PM_{2.5}$ ,  $PM_{2.5\ 24h}$ ,  $PM_{10}$ ,  $PM_{10\ 24h}$ ,  $SO_2$ ,  $SO_2\ 24h$ ,  $NO_2$ ,  $NO_2\ 24h$ ,  $O_3$ ,  $O_3\ 24h$ ,  $O_3\ 8h$ ,  $O_3\ 8h\ 24h$ ,  $CO$ , and  $CO\ 24h$ , and (3) time data were month and hour. The output is a 1-dimensional scalar, an AQI value for the Beijing area in the next hour. We took the first 30633 data as the training data, the next 4325 data as the validation data, and the last 4053 data as the test data, and we standardized those data separately. We set the number of training samples (batch size) in each batch to 72 and the time step to 50, and then we defined the neural network variables. After the training of the model, the validation set was used to verify the prediction performance of the model, and the hyperparameters were fine-tuned to obtain the values. The number of hidden layer nodes is set to 128, and the learning rate is set to 0.1. Adam is an optimization extension of the gradient descent

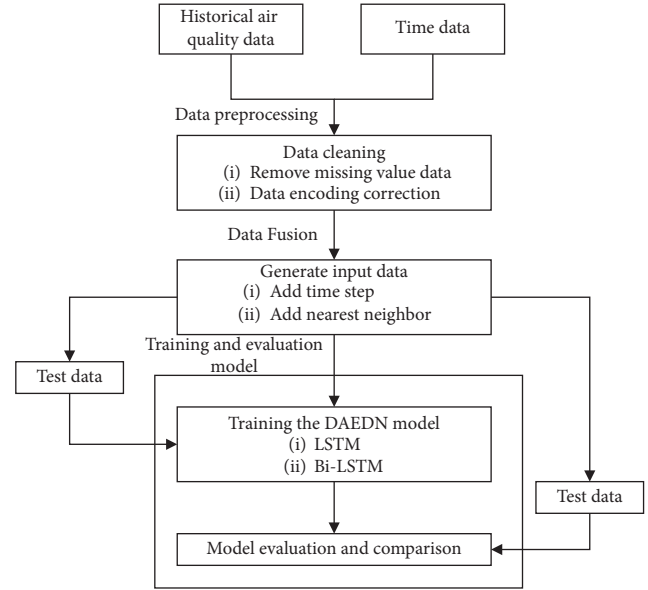


FIGURE 6: Experimental flowchart.

algorithm, which maintains an independent and adaptive learning rate for each network weight and has a good convergence effect. The activation function is set to ReLU, which is simple and prevents the gradient from disappearing.

When inputting features, one needs to convert the tensor into two dimensions for calculation. The input of the layer, and finally the tensor, are converted into three-dimension input for the LSTM cell. In the training model, the number of iterations can be represented, with the more iterations there are, the more accurate the prediction result is, but the longer the processing takes. The trained model can predict the AQI of the next hour.

**4.1. Performance Analysis of the DAEDN Model's Noise Reduction.** Two types of deep network models for the structure were established in the experiment. The input layer had both a DAE structure and non-DAE structure, and the noise reduction effect of the DAE structure was verified by comparison. The experimental results are shown in Table 1. It can be seen from the results that the RMSE value with the DAE structure was 15.459 and the MAE value was 7.000, which were 7.61% and 2.94% lower than the values without the DAE structure. It can be seen that the prediction accuracy of the DAE structure model was significantly higher than that of the non-DAE structure model, which indicates that the DAE structure had a good noise reduction effect on the processing of air quality data, removing the noise contained in the data and reducing the overall noise. The impact on network model training effectively improved the prediction accuracy of the model.

**4.2. Experimental Analysis of Lag Suppression.** When using an LSTM network for prediction, there is a certain resulting lag due to the accumulation of errors. In order to verify the

TABLE 1: Evaluation criteria for model prediction accuracy.

Model	RMSE	MAE
Non-DAE structure	16.732	7.212
DAE structure	15.459	7.000

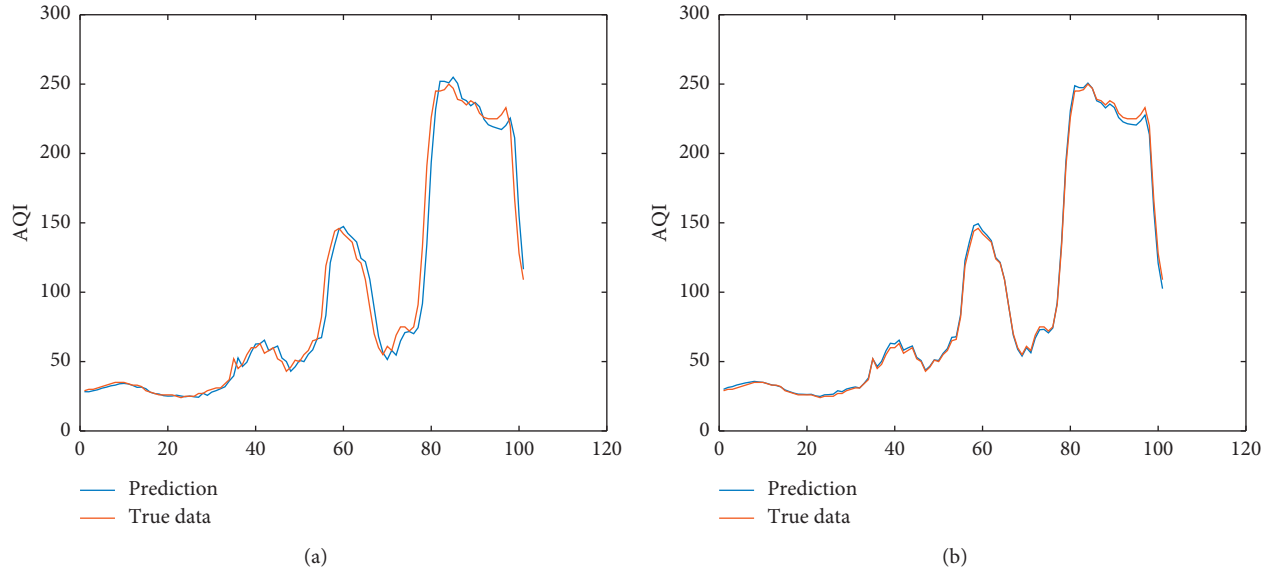


FIGURE 7: Deep network predictions and real values based on the DAEDN model. (a) LSTM. (b) Bi-LSTM.

suppression effect of the Bi-LSTM network on the hysteresis phenomenon, an experimental comparison with a unidirectional LSTM network was performed, and the results are shown in Figure 7. It can be seen that the use of the Bi-LSTM network had a certain suppression effect on the lag phenomenon, and the lag phenomenon greatly affected the real-time performance of the model's prediction. That effect is very important in practical applications, because real-time performance is a key factor in air quality prediction. Here, the inhibition effect of the Bi-LSTM on the lag effectively guaranteed the real-time performance of the model prediction.

From the perspective of accuracy, the accuracy evaluation criteria of the LSTM and the Bi-LSTM are shown in Table 2. The RMSE of the Bi-LSTM was 15.504 and the MAE was 6.789. Compared with the values from the LSTM, the reduction values for the Bi-LSTM were 5.12% and 4.54%, respectively, and the prediction accuracy was improved. In an actual application of air quality prediction, the problem of the lag in the prediction of the LSTM greatly affects one's judgment of the results. Aiming at remedying that problem, utilizing the characteristics of the Bi-LSTM allowed us to simultaneously use the historical information and future information in the sequence, thus suppressing the problem of the lag in the prediction results of the LSTM to some extent.

*4.3. Comparative Analysis of Different Network Models.* At present, the commonly used air quality prediction methods are BP neural networks, deep recurrent neural networks (DRNNs), DBNs, and others. In order to verify the prediction effect of our DAEDN model and compare the

TABLE 2: Evaluation criteria for the LSTM and Bi-LSTM structure prediction accuracy.

Model	RMSE	MAE
LSTM	16.340	7.112
Bi-LSTM	15.504	6.789

DAEDN model with the test results of those commonly used models, we conducted comparison experiments, the input features of each model are 17-dimensional time series data and the output is a 1-dimensional data. The prediction effects of the models on the test data set are shown in Figure 8.

It can be seen from Figure 8 that among the different model structures, the accuracy of the BP model was relatively low, the performance of the DRNN was slightly better than the performance of the BP model, the DBN model had the best prediction performance among the first three, and the DAEDN model in our study and using the test data set had the greatest improvement of all in accuracy. Compared with the BP model, the DAEDN model's RMSE and MAE values were reduced by 7.29% and 21.85%, respectively, to achieve by far the best prediction accuracy. This shows that in the time series regression prediction problem, a DAEDN can extract the characteristics of the data better and has certain advantages in air quality prediction.

Figure 9 shows the study's training and verification loss values for three of the deep network models: the BP, DBN, and DAEDN models. Combining Figures 8 and 9, from the perspective of prediction accuracy and convergence speed, the performance of the BP network model was far lower than



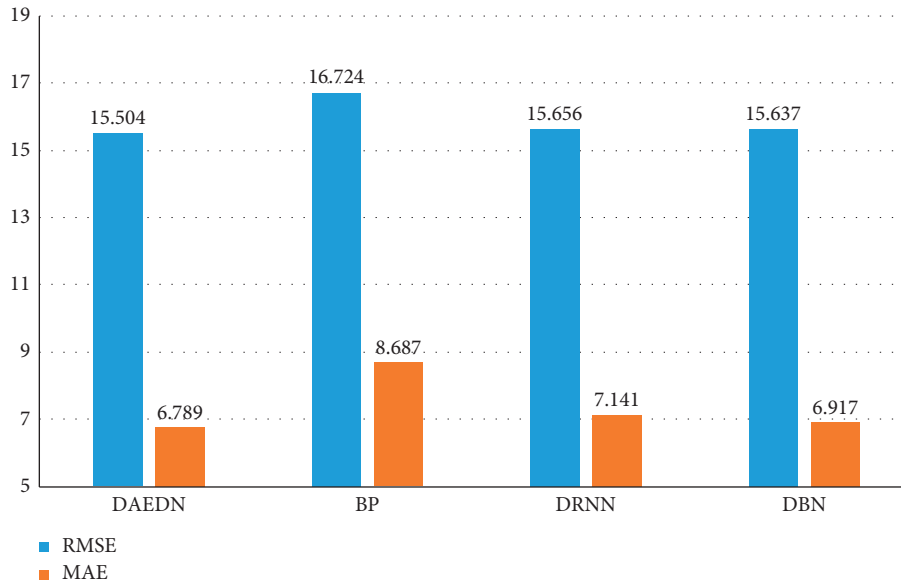


FIGURE 8: Evaluation criteria for prediction accuracy of each model.

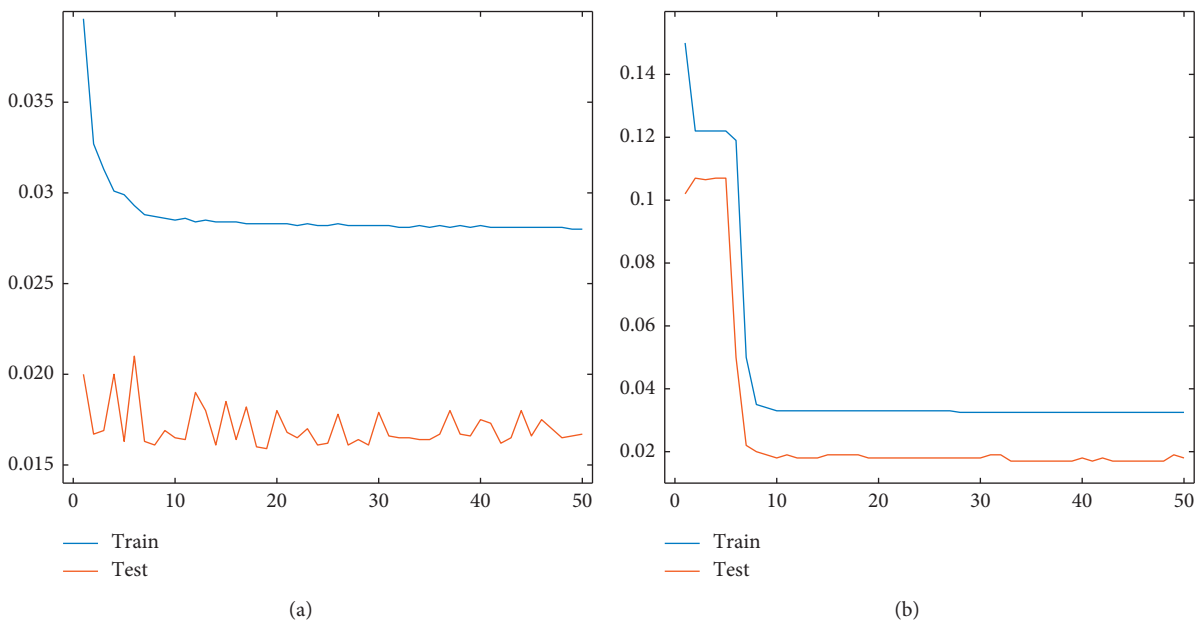
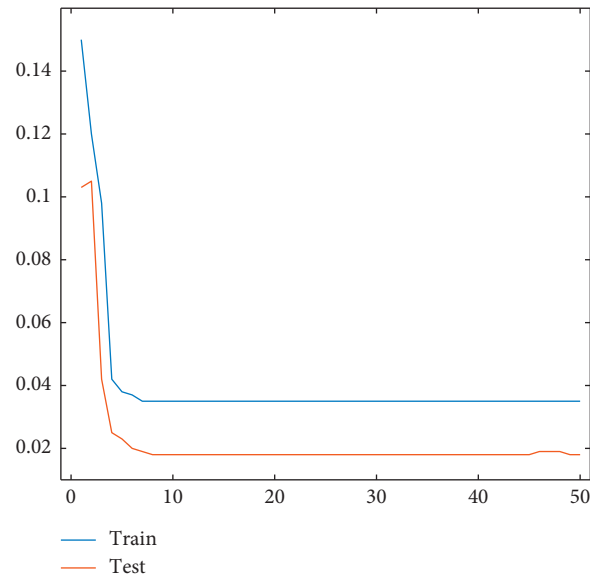


FIGURE 9: Continued.



(c)

FIGURE 9: BP, DRNN, and DAEDN models' deep network training loss and verification loss values. (a) BP. (b) DRNN. (c) DAEDN.

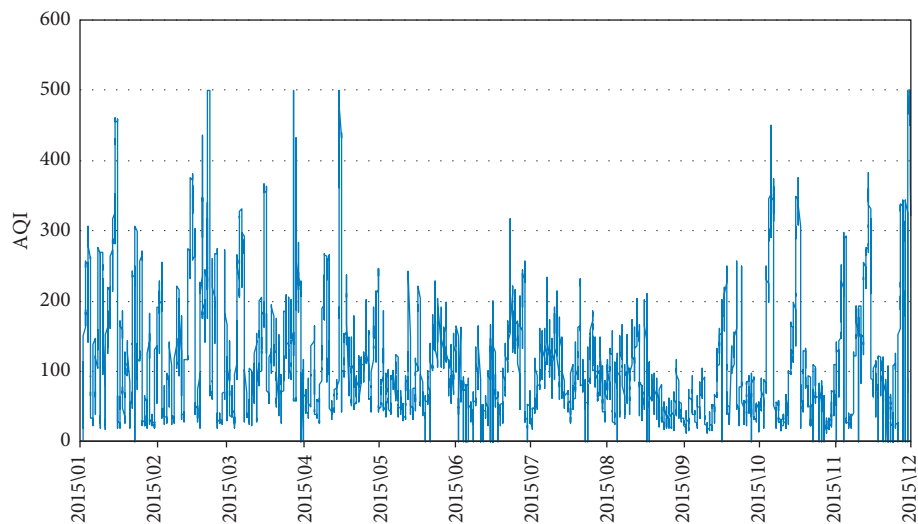


FIGURE 10: Beijing air quality index data from 2015.

that of the DBN and DAEDN network models. The DBN and DAEDN network models were similar in accuracy, but the DAEDN model was significantly better than the DBN in terms of convergence speed. In the early stage, the DBN model experienced a long period of iterative stagnation, which led to a decrease in the model's convergence speed. The DAEDN model did not show such a stagnation, which improved the training speed of the model to a certain extent.

**4.4. Seasonal Forecast Performance Analysis.** Taking the data from 2015 as an example, air quality in different quarters of the year were affected by objective factors, such as human activities and production and construction, and by seasonal changes that varied greatly, as can be seen in Figure 10. The

April-October, October-December air quality indexes were significantly higher, with more days exceeding 300, and the May-September 2015 air quality index was almost always below 300, with only one day exceeding 300. From those changes in the AQI, it can be seen that the severity of pollution varied from quarter to quarter, and whether the model could make good predictions based on the characteristics of each quarter became a focus of our testing and analyses.

April-May is the peak period of sandstorm weather in Beijing, which causes serious air pollution. At the same time, as the weather gets warmer, human travel and industrial production begin to increase, as the weather changes sharply and unstable, the fluctuation of pollution index has a greater impact. In the fourth quarter (October-December), the

TABLE 3: Evaluation criteria for forecast accuracy by quarters.

Model	RMSE	MAE
First quarter (Jan–Mar)	13.176	6.994
Second quarter (Apr–Jun)	20.729	9.019
Third quarter (Jul–Sep)	12.179	6.189
Fourth quarter (Oct–Dec)	16.345	8.014
All year	15.504	6.789

winter in Beijing comes. It can be seen that the air pollution index increases significantly after the start of central heating. Coal and other burning have some influence on the air quality.

For this situation, the dataset is divided into four datasets by quarter and trained and tested using the DAEDN model, respectively. The prediction accuracy of the model is shown in Table 3. It can be seen that among the models that are trained according to the quarter, the predictions in the first and third quarters are better, while the predictions in the second and fourth quarters are slightly worse. Among them, the second quarter is most affected, but the accuracy is still within the acceptable range, which indicates that the DAEDN model has a strong adaptability and still has a good predictive effect under the interference.

## 5. Conclusions

Using as its basis the structural framework of an LSTM network, this study designed a DAEDN air quality prediction model. Through training and learning, the relationships among air quality levels, pollutant factor concentrations, and meteorological data were used to make real-time predictions. Taking the air quality and meteorological data of the Beijing area from 2014 to 2018 as our sample for experimental analysis, the following conclusions were obtained: (1) the DAE structure of the DAEDN model input layer could effectively reduce noise and improve prediction accuracy; (2) the model's Bi-LSTM structure made good use of historical and future information to eliminate the lag and improve prediction accuracy; (3) the prediction performance of the study's surface DAEDN prediction model was superior to the prediction results of BP, DRNN, and DBN network models; and (4) the test results divided by quarters indicated that the prediction accuracy of the model was different in different quarters, but the quarterly accuracy levels basically remained near the accuracy level of the annual average prediction.

## Data Availability

All experimental data and calculated data that support the findings of this study are available from the corresponding author upon reasonable request.

## Conflicts of Interest

The authors declare no conflicts of interest.

## Acknowledgments

This work was supported by Scientific Research Plan Projects for Higher Schools in Hebei Province (no. ZD2018304), the Special Fund of Fundamental Scientific Research Business Expense for Higher School of Central Government (no. ZY20180111), and the China Scholarship Council for one-year study at the University of Alberta.

## References

- [1] International Agency for Research on Cancer, Carcinogen List, <http://samr.cfda.gov.cn/WS01/CL1991/215896.html>, 2017.
- [2] W. H. Ren, Z. Z. Su, and H. D. Zhao, "Advances in research on forecast of urban ambient air pollution," *Environmental Protection Science*, vol. 36, no. 3, pp. 9–11, 2010.
- [3] M. Liu, S. Wang, and H. Lin, "Establishment and application of winter ambient air quality statistical forecast model in Shenyang," *China Environmental Monitoring*, vol. 30, no. 4, pp. 10–15, 2014.
- [4] Q. Wang, "Numerical prediction and analysis of typical PM<sub>2.5</sub> pollution process in Shanghai in autumn," *China Environmental Monitoring*, vol. 30, no. 2, pp. 7–13, 2014.
- [5] E. Sahafzadeh and E. Ahmadi, "Prediction of air pollution of boushehr city using data mining," in *Proceedings of the 2009 Second International Conference on Environmental and Computer Science*, pp. 33–36, Dubai, UAE, December 2009.
- [6] I. N. Athanasiadis, K. D. Karatzas, and P. A. Mitkas, "Classification techniques for air quality forecasting," in *Proceedings of the Fifth ECAI Workshop on Binding Environmental Sciences and Artificial Intelligence*, Riva del Garda, Italy, August 2006.
- [7] C.-D. Neagu, N. Avouris, E. Kalapanidas, and V. Palade, "Neural and neuro-fuzzy integration in a knowledge-based system for air quality prediction," *Applied Intelligence*, vol. 17, no. 2, pp. 141–169, 2002.
- [8] B. G. Ibarra, A. Elias, and A. Barona, "Prediction of air pollution levels using neural networks: influence of spatial variability," *WIT Transactions on Ecology & the Environment*, vol. 116, pp. 409–417, 2008.
- [9] A. Azid, H. Juahir, and M. T. Latif, "Feed-forward artificial neural network model for air pollutant index prediction in the Southern Region of Peninsular Malaysia," *Journal of Environmental Protection*, vol. 4, no. 3, pp. 1–10, 2013.
- [10] H. Abderrahim, M. R. Chellali, and A. Hamou, "Forecasting PM 10, in Algiers: efficacy of multilayer perceptron networks," *Environmental Science and Pollution Research*, vol. 23, no. 2, pp. 1–8, 2016.
- [11] A. Azid, H. Juahir, and M. E. Toriman, "Prediction of the level of air pollution using principal component analysis and artificial neural network techniques: a case study in Malaysia," *Water, Air, & Soil Pollution*, vol. 225, no. 8, pp. 20–63, 2014.
- [12] D. Mishra, P. Goyal, and A. Upadhyay, "Artificial intelligence based approach to forecast PM 2.5 during haze episodes: a case study of Delhi, India," *Atmospheric Environment*, vol. 102, pp. 239–248, 2015.
- [13] H. B. Lu, D. D. Li, and H. Y. Sun, "Chaotic time series prediction of BP neural network optimized by PSO," *Computer Engineering and Applications*, vol. 51, no. 2, pp. 224–229, 2015.
- [14] L. C. Yann, B. Yoshua, and H. Geoffrey, "Deep learning," *Nature*, vol. 521, no. 7553, pp. 436–444, 2015.
- [15] F. Wang, S. Y. Cheng, and M. J. Li, "Genetic algorithm optimized neural network for prediction of atmospheric

- pollution,” *Journal of Beijing University of Technology*, vol. 35, no. 9, pp. 84–88, 2009.
- [16] X. H. Wang, Y. P. Zhang, and Y. Zhao, “Air quality prediction based on dynamic granular wavelet neural network,” *Computer Engineering and Applications*, vol. 49, no. 6, pp. 221–224, 2013.
- [17] Y. Yang, “Forecasting analysis of air quality index during heating period based on GA-BP neural network,” *Journal of Shaanxi University of Science and Technology*, vol. 34, no. 4, pp. 171–176, 2016.
- [18] Yi. Zheng and C. Y. Zhu, “Prediction of PM<sub>2.5</sub> based on deep belief network,” *Journal of Shandong University (Engineering Science Edition)*, vol. 44, no. 6, pp. 19–25, 2014.
- [19] T. Dong, J. H. Zhao, and Y. Hu, “AQI level prediction based on spatiotemporal optimized deep neural networks,” *Computer Engineering and Applications*, vol. 53, no. 21, pp. 17–23+41, 2017.
- [20] W. J. Yin, D. W. Zhang, and J. H. Yan, “Big data air pollution prediction based on deep learning,” *China Environmental Management*, vol. 7, no. 6, pp. 46–52, 2015.
- [21] X. Li, L. Peng, Y. Hu, J. Shao, and T. Chi, “Deep learning architecture for air quality predictions,” *Environmental Science and Pollution Research*, vol. 23, no. 22, pp. 22408–22417, 2016.
- [22] B. T. Ong, K. Sugiura, and K. Zettsu, “Dynamically pre-trained deep recurrent neural networks using environmental monitoring data for predicting PM<sub>2.5</sub>,” *Neural Computing and Applications*, vol. 27, no. 6, pp. 1553–1566, 2016.
- [23] J. Zhang and W. Ding, “Prediction of air pollutants concentration based on an extreme learning machine: the case of Hong Kong,” *International Journal of Environmental Research and Public Health*, vol. 14, no. 2, p. 114, 2017.
- [24] J. X. Fan, Q. Li, and Y. J. Zhu, “Research on air pollution spatio-temporal prediction model based on RNN,” *Science of Surveying and Mapping*, vol. 42, no. 7, pp. 76–83, 2017.
- [25] B. C. Liu, M. Z. Lai, X. Qi, and H. Wang, “Prediction of air pollutant concentration in beijing based on wavelet-LSTM model,” *Environmental Science and Technology*, vol. 42, no. 08, pp. 142–149, 2019.
- [26] Y. Bengio, “Learning deep architectures for AI,” *Foundations and Trends in Machine Learning*, vol. 2, no. 1, pp. 1–127, 2009.
- [27] N. Lopes and B. Ribeiro, “Improving convergence of restricted Boltzmann machines via a learning adaptive step size,” *Progress in Pattern Recognition, Image Analysis, Computer Vision, and Applications* Springer, Berlin, Germany, vol. 7441 of Lecture Notes in Computer Science, pp. 511–518, 2012.
- [28] R. Raina, A. Madhavan, and A. Y. Ng, “Large-scale deep unsupervised learning using graphics processors,” in *Proceedings of the 26th Annual International Conference on Machine Learning*, pp. 873–880, New York, USA, June 2009.
- [29] D. L. Ly, V. Paprotski, and Y. Danny, “Neural networks on GPUs: restricted Boltzmann machines,” Technical Report, University of Toronto, Toronto, Canada, 2009.
- [30] N. Lopes and B. Ribeiro, “Towards adaptive learning with improved convergence of deep belief networks on graphics processing units,” *Pattern Recognition*, vol. 47, no. 1, pp. 114–127, 2014.
- [31] R. N. Le and Y. Bengio, “Representational power of restricted Boltzmann machines and deep belief networks,” *Neural Computation*, vol. 20, no. 6, pp. 1631–1649, 2008.
- [32] G. E. Hinton, “Training products of experts by minimizing contrastive divergence,” *Neural Computation*, vol. 14, no. 8, pp. 1771–1800, 2002.