

Received October 7, 2021, accepted November 29, 2021, date of publication December 6, 2021, date of current version January 6, 2022.

Digital Object Identifier 10.1109/ACCESS.2021.3133497

# FA-LSTM: A Novel Toxic Gas Concentration Prediction Model in Pollutant Environment

YU CONG<sup>1</sup>, XIMENG ZHAO<sup>2</sup>, KE TANG<sup>2</sup>, GE WANG<sup>3</sup>, YANFEI HU<sup>4</sup>, AND YINGKUI JIAO<sup>5</sup>

<sup>1</sup>Tianjin Key Laboratory for Control Theory and Applications in Complicated Systems, School of Electrical and Electronic Engineering, Tianjin University of Technology, Tianjin 300384, China

<sup>2</sup>China Institute for WTO Studies, University of International Business and Economics, Beijing 100029, China

<sup>3</sup>College of Management and Economics, Tianjin University, Tianjin 300072, China

<sup>4</sup>School of Cyber Security, University of Chinese Academy of Science, Beijing 100049, China

<sup>5</sup>State Key Laboratory of Precision Measuring Technology and Instruments, Tianjin University, Tianjin 300072, China

Corresponding authors: Yu Cong (congyu.tju@gmail.com) and Ke Tang (kxt131@126.com)

This work was supported in part by the Major Demonstration Project of Tianjin Internet Cross-Border Integration Innovation and Technology under Grant 18ZXRH00240.

**ABSTRACT** Real-time monitoring and accurate prediction of toxic gas concentration in the future are of great significance for emergency capability assessment and rescue work. At present, the method of gas concentration prediction based on artificial intelligence still has problems of low accuracy, slow convergence speed and equal feature importance. This paper proposes a feature-aware LSTM model to predict pollutant gas concentration. First of all, we design a set of multi-component toxic gas monitoring equipment that applies in pollution environment, which can at the same time monitor CO, NO<sub>2</sub>, NH<sub>3</sub>, HCN, H<sub>2</sub>S and SO<sub>2</sub>, six common pollutants; To accurately estimate the toxic gas concentration, we combine the collected gas data and the environmental parameters and regard them as the input features, and then we obtain toxic gas data based on the sampling policy and the environmental data as our data-set. Finally, we train a FA-LSTM gas concentration prediction model on these data-set. We test the proposed model and compared with ARIMA, ETS and BP network on the same test set. Experimental results show that the proposed model outperforms traditional concentration prediction model. Also, it is better than other state-of-the-art models in predicting accuracy.

**INDEX TERMS** Pollution emergency decision, toxic gas, air pollution prediction, time series, LSTM.

## I. INTRODUCTION

A contaminated site, also known as a “brownfield,” refers to the space environment that carries harmful substances due to accumulation, storage, treatment, disposal or other means (such as migration), causing harm or potential risk to human health and the region [1]. In heavily polluted sites, there are often sudden, hidden and highly lethal chemical asphyxiating gases, such as volatile gases such as carbon monoxide, hydrogen cyanide and hydrogen sulfide. These toxic gases pose a great threat to the personal safety of personnel during the repair operation of polluted sites.

The accuracy of toxic gas concentration prediction directly determines the effectiveness of on-site solution implementation. Therefore, it is of great significance for the life and health of repair workers to monitor the toxic gas in the

The associate editor coordinating the review of this manuscript and approving it for publication was Sun-Yuan Hsieh<sup>1</sup>.

polluted environment in real time and to predict the amount of toxic gas or its development trend in real time, accurately, scientifically and reasonably. Among polluting gas predictions, it is challenging to accurately predict the concentration of polluting gas. First of all, the prediction of gas concentration is closely related to the terrain environment and surrounding temperature or humidity, and different terrain also has different effects on the surrounding wind speed. The coupling of each influencing factor and its strong nonlinear correlation bring great challenges to the analysis for technical personnel. Secondly, the high demand on modelling expert is another challenge in the prediction of pollutant gas composition. Professional technicians are usually better at determining gas content and analyzing which factors are affecting its concentration; The data are then handed over to modelling expert, and the level of these experts further constrains the accuracy or the efficiency on the concentration prediction of polluted gases. The accuracy of the model directly affects the

subsequent decision execution, and the low prediction will bring great damage to the safety of personnel or property. Therefore, how to quickly, accurately and conveniently predict the composition of polluting gas is a very urgent need.

At present, there are two kinds of gas concentration prediction methods, one is based on mechanism process, and the other is based on historical data. In concentration prediction, mechanism method obtains the results of target variables by analyzing the relationship between various influencing factors and target variables [2]. For scenes with strong non-linear and coupling, this method has low accuracy and an extremely complex analysis process. The calculation time of the model can reach up to several hours. For emergencies, the emergency response ability of this model cannot effectively offer solutions for field staff [3], [4], and its practicability is low. However, the method based on historical data can extract the hidden relationship between variables by analyzing data, which has high accuracy and practicability. Currently, gas concentration prediction methods based on historical data mainly include traditional modeling methods and artificial intelligence methods. Traditional modeling methods based on historical samples include multiple linear regression method [2], Kalman filter technology [5], time-space series model methods, etc. (existing time-space series models such as autoregressive comprehensive moving average model (ARIMA) [6], seasonal ARIMA [7], Gaussian plume diffusion model [8]). The prediction method based on multiple linear regression approximates the gas diffusion model to a linear model, which reduces the accuracy of prediction [6]. The establishment of a nonlinear model based on space-time series requires experienced personnel to determine parameters and the orders of the equation, which hinders the deployment of the model. In order to simplify the complexity of model building, a series of pollutant prediction models based on machine learning were proposed [9]. However, such methods have the limits of low accuracy, slow training speed [10], [11], or the model is difficult to converge and optimize [12]. The latest concentration predictions avoid the problem of the above methods, however, there are still exists defects of predicting a single gas concentration or regarding multiple influencing factors as the equal. In fact, the influence of various factors on the target object varies frequently, for example, the maximum effect of on concentration in the morning is humidity, and at noon wind speed has bigger influence on the concentration.

Based on the above analysis, we propose a FA-LSTM model to predict gas concentration in the process of contaminated site treatment. FA-LSTM is short for feature aware LSTM for multi-objective prediction model. Specifically, we use multi-component toxic gas monitoring equipment to monitor six high-risk gas pollutants in the polluted site, which are NH<sub>3</sub>, CO, NO<sub>2</sub>, HCN, H<sub>2</sub>S and SO<sub>2</sub>. We used the time series of the monitoring concentration data as input feature. In addition, the environmental parameters in this area predicted by the weather forecast are used as the auxiliary feature input to establish the model. In the process of model building,

we add attention mechanism to distinguish the importance of each feature. Finally, the model is used to predict the pollutant concentration in the following moments. We evaluate our proposed model and four other state-of-the-art time series prediction models based on the same test dataset. The evaluation results show that the performance of our proposed model on the pollutant gas concentration prediction is obviously better than other methods.

In summary, the contributions of this paper are as follows:

(1) We novelly combined the monitoring concentration data collected in the field with environmental parameters to predict multi-gas pollutant concentrations; the effective combination of these features helps improve the accuracy of the model.

(2) We novelly treat the weights of multiple features differently; the weight difference makes the results of the model more sensitive to the changes of features, and solves the problem that feature weights cannot be distinguished.

(3) We propose a combined model called FA-LSTM, we evaluate the results of our proposed model and compare it with the state-of-the-art model.

The rest of this paper is as follows: the second section introduces the basic theory of the model and designs the LSTM model in this paper; The third section focuses on the implementation process and methods, including the construction of monitoring equipment and a brief description of data acquisition. Firstly, the data is preprocessed to obtain the sampling data set, then the data is loaded in a specific way, trained with the model in this paper, and finally evaluated the performance. In the fourth section, the experimental results are analyzed and discussed. Finally, this paper draws a conclusion.

## II. BASIC OF LSTM ALGORITHM

The long-term and short-term memory network is a special recurrent neural network (RNN). Cyclic neural network has the ability to remember and analyze historical data, that is, the output of cyclic neural network does not only depend on the current input layer, but also contains historical data information, so it is very suitable for processing the data of predicting time-series related information. The structure of RNN is shown in Figure 1.

Bengio *et al.* described that gradient disappearance and gradient explosion would occur in the training process of RNN in practical application [13], which could not process a long input sequence. As a variant of RNN, LSTM model has a unique design structure: controllable self-circulation is skillfully introduced to generate a path that enables the gradient to flow sustainably for a long time, and effectively overcome the gradient disappearance problem caused by the passage of time and the increase of network layers in machine learning of RNN. On the basis of RNN, LSTM adds a memory cell state and three control gates: input gate, forgetting gate and output gate. The input gate determines how much network input is saved to the unit state at the current moment; The forgetfulness gate determines how much of the unit state

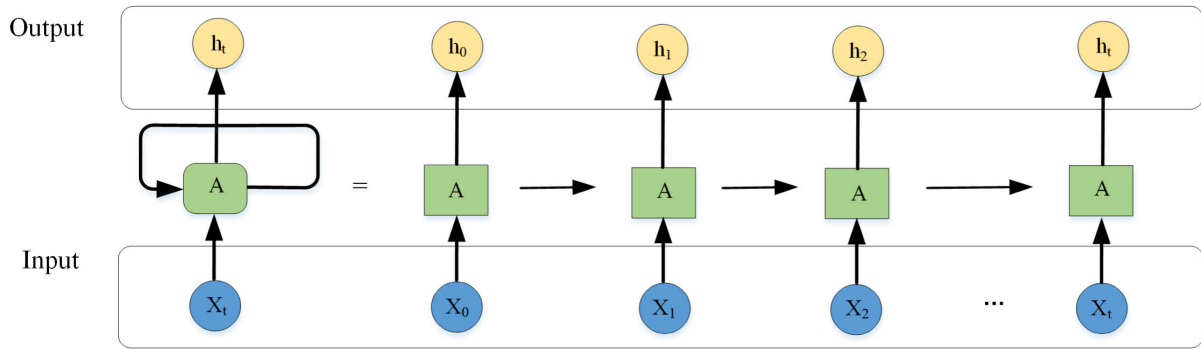


FIGURE 1. Example RNN structure expansion.

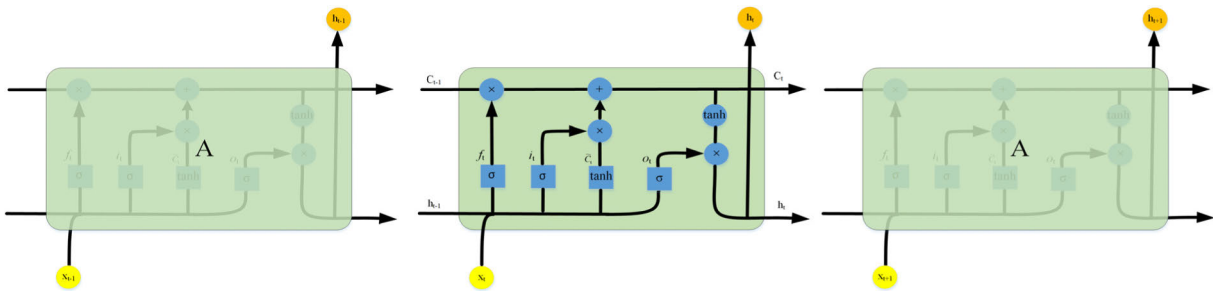


FIGURE 2. The memory unit structure of LSTM.

from the previous moment is retained to the current moment; The output gate determines how much output the unit state has to the current output value. These three special gate structures can effectively solve the gradient disappearance problem and have memory function, and are suitable for dealing with long-term dependence problem. Figure 2 shows the structure of duplicate modules and hidden layers of the memory unit LSTM.

The general working principle of LSTM can be expressed by Equation (1):

$$\begin{cases} i_t = \sigma(W_i \cdot [h_{t-1}, x_t] + b_i) \\ f_t = \sigma(W_f \cdot [h_{t-1}, x_t] + b_f) \\ C_t = f_t * C_{t-1} + i_t * \tilde{C}_t \\ \tilde{C}_t = \tanh(W_c \cdot [h_{t-1}, x_t] + b_c) \\ o_t = \sigma(W_o \cdot [h_{t-1}, x_t] + b_o) \\ h_t = o_t * \tanh(C_t) \end{cases} \quad (1)$$

where  $i_t$ ,  $f_t$ , and  $o_t$  represent input gate, forgetting gate and output gate respectively, and  $W$  ( $W_i, W_f, W_o, W_c$ ) and  $H$  ( $H_i, H_f, H_o, H_c$ ) are weight matrices;  $b$  ( $b_i, b_f, b_o, b_c$ ) represents the deviation vector;  $\sigma$  indicates the sigmoid activation function, and the value range of the function is  $0 \sim 1$ ;  $\tanh(\cdot)$  represents hyperbolic tangent activation function, and the output range is  $-1$  to  $1$ ;  $\tilde{C}_t$  is the candidate value of the state of the memory unit at time  $t$ , and  $C_t$  represents the state of the current memory unit at time  $t$ ;  $h_t$  is the output value at time  $t$ .

### III. MODEL DESIGN

The goal of the model is to predict the pollution gas content at a future moment based on input historical samples and current environmental characteristics. We need to learn a model that can output the pollutant at the next time according to the input at any time. Therefore, during training, the model output should meet the following relationship: the loss of square error (MSE) is selected as the loss function during predicted results and actual results is the smallest. The mean model training, and the formula is as follows:

$$MSE = \frac{1}{n} \sum_{i=1}^n (y_i - Y_i)^2 \quad (2)$$

where  $y_i$  represents the predicted value and  $Y_i$  represents the real value. The square of the difference between the real value and the predicted value is then summed and averaged to represent the training error. The above formula is the objective function optimized during training.

When constructing the network, we choose LSTM network, because it effectively solves the problems of long-term dependence on information and gradient disappearance during deep network training. We designed the following network structure: three-layer hidden layer structure, which enables us to achieve higher model accuracy. The network model is shown in Figure 3. The specific parameters corresponding to each layer are shown in Table 1. Among the hidden layers, the correction linear unit (Relu) function is selected as the activation function. Compared with the

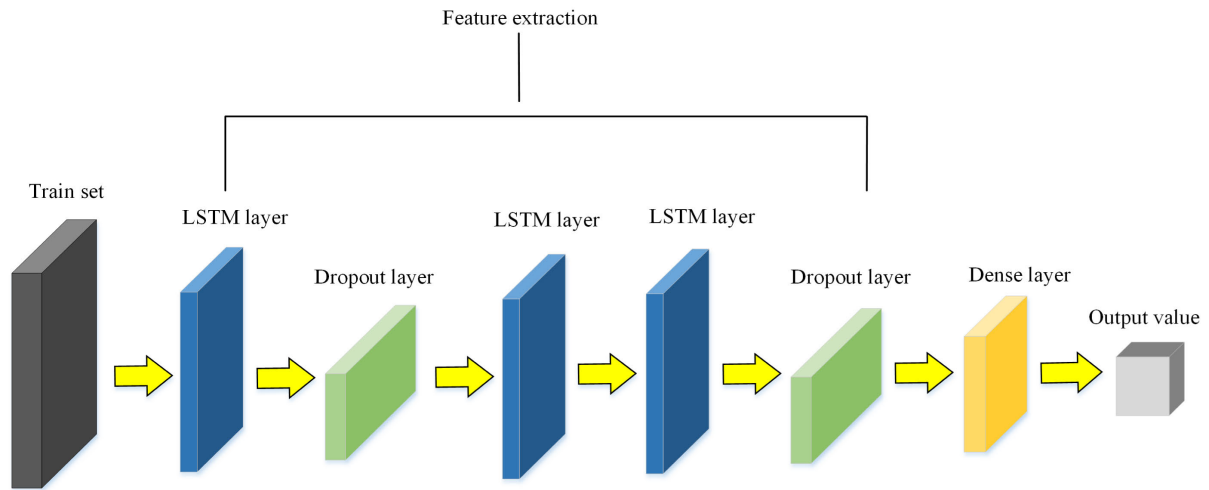


FIGURE 3. Multi-layer LSTM network model architecture.

TABLE 1. Concrete parameters of each layer of the model.

LSTM-layer1 neurons	Activation	LSTM-layer2 neurons	Activation	LSTM-layer3 neurons	Activation	Dense neurons	Activation
100	Relu	100	Relu	10	Relu	1	linear

classical Sigmoid activation function, Relu can avoid the occurrence of negative predicted concentration value, and is better than other activation functions in terms of statistical performance and computational cost [14]. The activation function of the output layer is Linear. In addition, in order to improve the generalization ability of deep learning model and avoid overfitting, we use the most commonly used method in deep learning to prevent neural network overfitting, the dropout method. The key idea for Dropout is to randomly delete neurons (and their connections) from the neural network during training to reduce co-adaptive relationships [15]. In this paper, the Dropout was added after the first LSTM layer and the last LSTM layer respectively, and the Dropout rate was set to 0.2. The better weight of the optimization model of Adam optimizer is selected [16]. The learning rate is set too large, and the number of training iterations of the model is too small, which leads to continuous oscillation at the optimal point and failure of normal convergence. The learning rate is too small, and the model is under-fitted. The loss value between 0.01 and 0.0001 of the learning rate declines smoothly and converges. Too few iterations cannot make the model obtain accurate prediction results. Therefore, the trained data sets need to be transmitted in the same neural network for many times. Combined with the complexity of the experiment in this paper, 300 epochs can meet the training accuracy requirements.

#### IV. EXPERIMENT EVALUATION

##### A. EXPERIMENTS SETUP

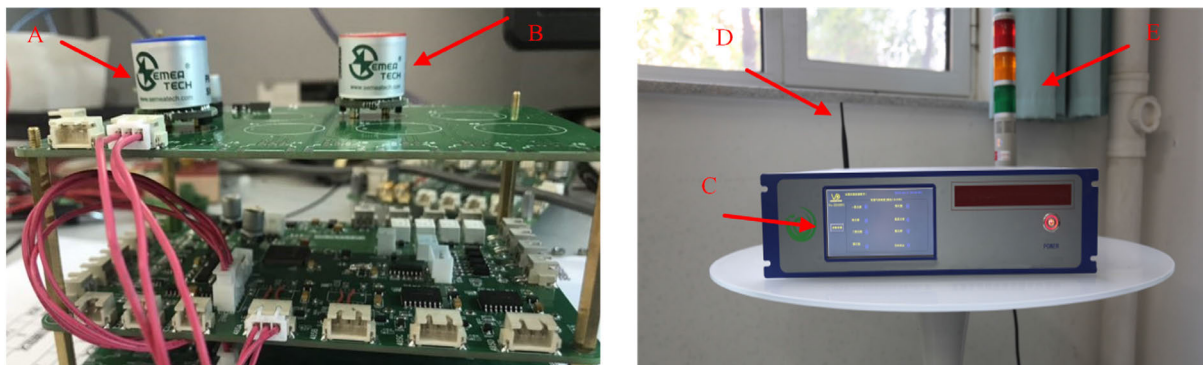
The test environment is as follows: the OS is Ubuntu16.04, CPU is 2 × Intel Xeon Gold5120 CPU, and GPU is 8 × 32GB

V100 SXM2 NVLINK GPU. The results are presented in the locale based on Python 3.7 and verified in the toxic gas sampling data set. Based on Python 3.7 language environment, Keras is used as the deep learning framework for training and prediction. Keras encapsulates many components of TensorFlow and Theano, two of the best open source deep learning frameworks [17]. Users can design various kinds of networks only by calling APIs, which is very convenient when optimizing networks.

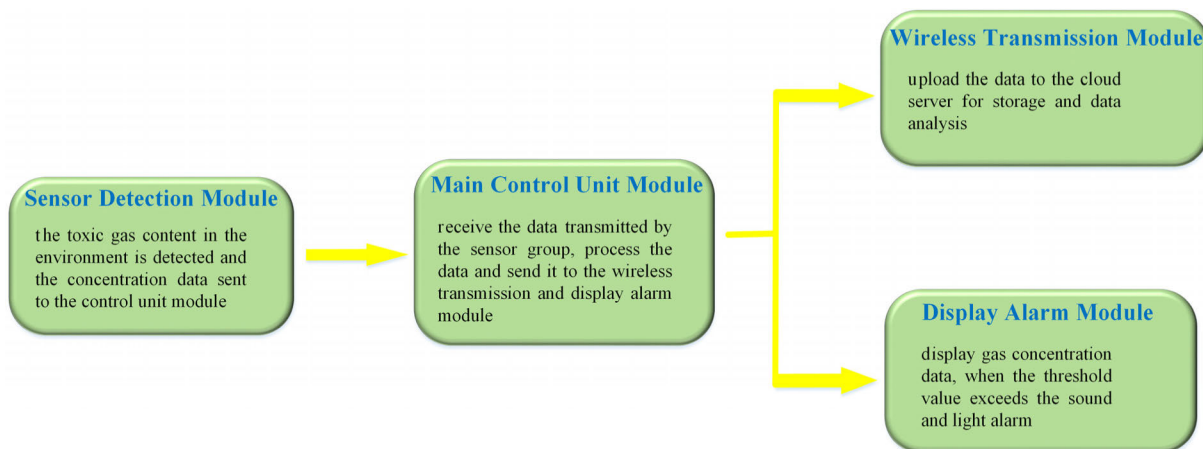
The multi-component toxic gas monitoring equipment based on embedded system technology designed in this paper is shown in Figure 4. The device uses STM32F407VGT6 micro-controller as the controller and six ECM-SMART electrochemical gas sensors to detect the corresponding gas concentration value. The 4G wireless transmission module and MCGS man-machine interactive display are responsible for data uploading and display.

##### B. DATA COLLECTION

The experimental data in this paper are mainly from the multi-component toxic gas monitoring equipment at each collection node in the contaminated site. Each acquisition node is a set of embedded system. Using the modular design idea, the toxic gas acquisition node is divided into four modules according to its functions, namely sensor detection module, control unit module, display alarm module and wireless transmission module. The data acquisition is mainly completed by the sensor detection module, which is composed of sensor group. The sensor group is composed of six ECM-SMART electrochemical gas sensors. The parameters are shown in Table 2, namely hydrogen cyanide, carbon monoxide,



**FIGURE 4.** Internal structure (Left) and the overall structure (Right) of Toxic gas monitoring. A and B: ECM-SMART electrochemical gas sensor; C: MCGS touch panel; D: Wireless transmission antenna; E: Audible and visual alarm.



**FIGURE 5.** Flow chart of data acquisition.

hydrogen sulfide, nitrogen dioxide, ammonia and sulfur dioxide. The sensor has the trait of small volume, low power consumption, high resolution, linearity and good repeatability, and meets the measurement requirements of toxic gas to be measured [18]. The data is detected by the sensor group, and the signal is transmitted to the single chip microcomputer through the UART bus. After calculation and processing, the data are respectively transmitted to 4g DTU and uploaded to the cloud server for storage and display. The alarm module can display the gas concentration on the HMI screen, set the gas concentration alarm threshold, and give an audible and visual alarm according to the set threshold to remind relevant personnel to evacuate as soon as possible. The data acquisition process is shown in Figure 5.

The toxic gas sampling data set includes the concentration of six pollutant gases including SO<sub>2</sub> NO<sub>2</sub> CO H<sub>2</sub>S HCN NH<sub>3</sub>. The data set can truly and comprehensively reflect the pollution situation of each toxic gas in the polluted site. The experimental site is a polluted site in Tianjin Dong jiang Port Free Trade Zone (located at 39.02N and 117.44 E), where a warehouse preserving dangerous goods fired and exploded in 2015, and all kinds of toxic gases are still trapped even after relevant treatment. During the experiment, 10 monitoring

devices are deployed in different locations of the site for all-round and multi-angle sampling. In order to display the prediction results more clearly, this paper uses 5 months data, and the sampling time is selected from January 15 to June 15, 2021. The sampling interval of concentration data of each group is 5 minutes, and there are 43200 sample for each type of gas. Since the data set is made by sampling in real polluted site, it has certain practical value. Finally, we get the time series of concentration data of six pollution factors through data preprocessing.

At present, weather forecasts have reached a level that people can rely on. If the characteristics of weather forecast can be well used for pollution prediction, the accuracy of air pollution forecast can be improved to a certain extent. The environmental features selected in this paper are shown in the following Table 3:

Finally, we divide the data set into training set and test set. The first 85% of the original data set is used as a training set for the training model, and the last 15% is used as a test set to evaluate the model performance. When the model is trained using training set and saved, its parameters can be read directly to predict the concentration value in the testing process.

TABLE 2. Performance paramter of each sensor.

Sensor Name	Range (ppm)	Resolution (ppm)	Largest Load (ppm)	Sensitivity (μA/ppm)	Baseline Drift
Carbon Monoxide	0-50	0.1	100	0.50±0.10	<1 ppm
Sulfur Dioxide	0-20	0.05	20	0.55±0.15	<0.5 ppm
Hydrogen Cyanide	0-20	0.2	100	0.10±0.02	<1 ppm
Hydrogen Sulfide	0-10	0.1	100	1.30±0.15	<0.2 ppm
Nitrogen Dioxide	0-20	0.05	20	0.65±0.20	<0.5 ppm
Ammonia	0-100	0.3	200	0.135±0.03	<2 ppm

TABLE 3. Selection of environment parameter.

Parameters	Symbol	Unit
Wind direction	$\theta$	°
Average wind speed	$\bar{v}$	m/s
Temperature	T	°C
Heat flux	$H_f$	W/m <sup>2</sup>
Atmosphere stability length	L	m

C. DATA PREPROCESSING

The original sampling data set is usually chaotic, and abnormal data cannot be avoided (due to packet loss and other reasons). Therefore, it is necessary to preprocess the data on the premise of data analysis. The abnormal data obtained by monitoring mainly include outliers, missing values and repeated values. For individual outliers, direct deletion of records has a slight impact on the whole. For a large number of missing values, air quality data is a time series. If all records of missing values are deleted, the characteristics in the time dimension are also deleted accordingly, and an appropriate filling method needs to be selected to fill the missing values. Therefore, this paper proposes to clean the missing data based on z-Score algorithm [19]. Firstly, the abnormal value of the given data is detected based on the Z-score algorithm for deletion, and then the lost value data is filled by the mean filling method. Finally, duplicate data in the data samples were deleted and the duplicate values were removed.

$$Z_i = \frac{X_i - \mu}{\sigma}, \quad |Z_i| > Z_{thr} \quad (3)$$

Z-score is a parameter anomaly detection method in one-dimensional or low-dimensional feature space [20]. This technique assumes that the data is Gaussian distribution, and the outliers are the data points at the tail of the distribution, so they are far away from the average value of the data. The distance depends on the set threshold  $Z_{thr}$  of the normalized data point  $Z_i$  calculated using the formula.  $Z_{thr}$  values are generally set to 2.5, 3.0 and 3.5. Where  $X_i$  is a data point,  $\mu$  is the average of all points  $X_i$ ,  $\sigma$  is the standard deviation of all points  $X_i$ . Then, after standardization, the abnormal value is also standardized, and its absolute value is greater than  $Z_{thr}$ .

Data normalization linearly reduces the data from a value range to a new value range. In this paper, the min max scale method is used to normalize the data [21], so that an original value  $x$  is mapped to the value  $x^i$  of interval

[0, 1]. The Min-Max Scale normalization can be expressed as Equation (4), where  $max\{x_i\}$  and  $min\{x_i\}$  represent the maximum and minimum values of the current concentration data respectively

$$x'_i = \frac{x_i - min\{x_i\}}{max\{x_i\} - min\{x_i\}} \quad (4)$$

Due to normalization, the final result must be inverse normalized to restore the original interval. The inverse normalization expression is Equation (5).

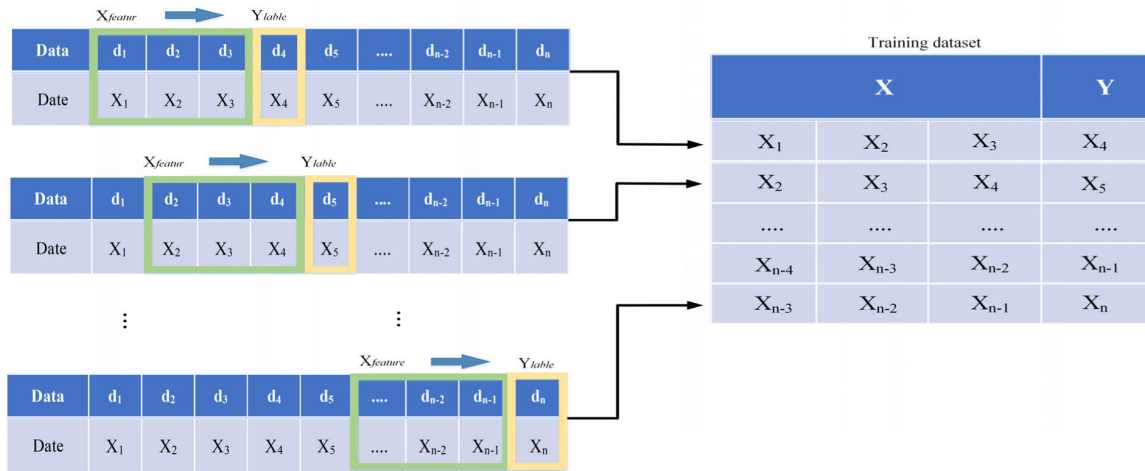
$$x_i = x'_i(max\{x_i\} - min\{x_i\}) + min\{x_i\} \quad (5)$$

D. MODEL DATA LOADING

The objective of the experiment is to use the multi-layer LSTM network to build a regression prediction model based on the time series data of toxic gas concentration characteristics at the pollution site. The LSTM model consists of three parts: the input layer, the hidden layer and the output layer. The input layer is the concentration characteristics and environmental characteristics of the training set in the toxic gas sampling data set. The training data loaded into the model is shown in Figure 6. The data intercepted by the time window each time will be used as the input feature  $X_{feature}$  of the model, and the next item will be used as the concentration label  $Y_{lable}$  of 10 s in the future that needs to be predicted. The number of moves of the time window each time is 1. In this way, the model input sample data required for training will be obtained from the training data set in turn, and these window data will be randomly scrambled into the model training, so as to make the training more consistent with the real distribution of data, so as to improve the generalization ability of the prediction model.

E. BASE MODEL

In order to evaluate the performance of the proposed method, two common benchmark models are established with the same training set. The first is the prediction method based on statistical time series, which is the auto-regressive comprehensive moving average model (ARIMA) and exponential smoothing model (ETS). Both of these methods are based on historical data and their accuracy depends on the estimated parameter order. The other is the prediction model based on machine learning algorithm, namely BP model and SVM model. The selection of these two types of benchmark models



**FIGURE 6.** Schematic diagram of loading training data in model.

can enable us to verify the effect of the LSTM model more comprehensively. For each prediction model, we selected the optimal parameters according to a specific method. Table 4 describes the optimal parameters of each model.

**F. EVALUATION METRIC**

In this paper, the mean absolute error (MAE), root mean square error (RMSE) and correlation coefficient (R) are selected to evaluate the prediction performance of the model. MAE represents the average value of the absolute error between the true value and the predicted value, avoiding the case that the positive and negative cancel each other [22].

RMSE is the square root of the square of the deviation between predicted value and true value and the ratio of observed number n, which can better reflect the prediction accuracy [23].

$$\left\{ \begin{aligned} MAE &= \frac{1}{n} \sum_{i=1}^n |C_i - C'_i| \\ RMSE &= \sqrt{\frac{1}{n} \sum_{i=1}^n (|C_i - C'_i|)^2} \\ r &= \frac{Cov(C_i, C'_i)}{\sqrt{Var(C_i) \cdot Var(C'_i)}} \end{aligned} \right. \quad (6)$$

**TABLE 4.** The most optimized parameters in each model.

Algorithm	The best Parameters
ARIMA	Length of one seasonal period (s):12 Order and seasonal Order: Auto-ARIMA function
ETS	Trend: Additive Seasonality: Additive Length of one seasonal period (s):12 Lag:20
SVM	Regularization parameter (C):50 Kernel: Radial Basis Function (RBF) Lag:15
BP	Number of hidden layers = 1 Number of units in the hidden layer (n):100 Learning rate (lr) :0.001 Number of epochs (E): 200

where  $C_i$  is the monitored concentration,  $C'_i$  is the predicted concentration,  $Var$  is the calculated variance, and  $Cov$  is the calculated covariance.

**V. EXPERIMENT RESULTS**

We used the trained LSTM network model, ARMLIA linear model, ETS model, SVM model and BP model to simultaneously predict the concentration of each component monitoring gas in the contaminated site on a certain day. In order to verify that the effect of the model has nothing to do with the specific data distribution (time dimension), we selected the data of five months for verification, and randomly selected one day in each month of the five months to verify the stability of the model. Finally, we give the mean and standard deviation of the five predictions.

**A. COMPARSION WITH STATISTICAL MODELS**

ETS and ARIMA are widely used statistical models in time series prediction. Here we compare the performance of our methodology with those two models, using our proposed methodology, actual monitoring data and forecast data from ARIMA and ETS, as shown in Figure 7. This model was designed and optimized by the Statsmodels1 package in Python, and the performance evaluation results are shown in Table 5.

It is clear from the chart that ETS is superior to ARIMA in the test set of toxic gas sampling data set consisting of six component monitoring gases. This result is due to the fact that ETS can handle both linear and nonlinear modes, while ARIMA can only handle linear modes. However, The LSTM model designed by us has achieved the best results on RMSE, MAE and r, and the prediction effect is far better than RMSE and MAE.

**B. COMPARSION WITH MACHINE LEARNING MODELS**

In this section, we compare the performance of the proposed method with two machine learning algorithms, support vector machine (SVM) and BP neural network (BP). SVM is a

TABLE 5. Average predictive performance of traditional models on test sets (5 months data).

Toxic Gas	Model	RMSE	MAE	r
SO <sub>2</sub>	ARIMA	39.3865±0.113	21.2154±0.116	0.4754±0.074
	ETS	32.4522±0.074	19.2145±0.114	0.5214±0.056
	LSTM	<b>4.1453±0.012</b>	<b>3.1458±0.017</b>	<b>0.9415±0.047</b>
NO <sub>2</sub>	ARIMA	32.5475±0.241	18.2125±0.257	0.3458±0.341
	ETS	28.3547±0.142	17.2145±0.245	0.5214±0.259
	LSTM	<b>6.2451±0.024</b>	<b>3.1245±0.475</b>	<b>0.8912±0.047</b>
CO	ARIMA	18.1475±0.496	9.4521±0.297	0.5121±0.247
	ETS	18.0145±0.248	8.1421±0.175	0.6574±0.169
	LSTM	<b>6.1285±0.121</b>	<b>2.3145±0.814</b>	<b>0.9522±0.027</b>
H <sub>2</sub> S	ARIMA	10.4829±0.269	5.1245±0.147	0.5414±0.149
	ETS	8.2364±0.158	5.0121±0.128	0.6218±0.128
	LSTM	<b>5.1479±0.054</b>	<b>3.1548±0.036</b>	<b>0.9457±0.022</b>
HCN	ARIMA	10.1456±0.147	8.2541±0.196	0.7654±0.142
	ETS	9.5247±0.128	7.1421±0.174	0.7985±0.117
	LSTM	<b>6.1542±0.119</b>	<b>4.1475±0.102</b>	<b>0.9124±0.002</b>
NH <sub>3</sub>	ARIMA	42.1235±0.285	20.1485±0.184	0.6125±0.174
	ETS	37.4578±0.165	17.4875±0.369	0.8745±0.158
	LSTM	<b>12.2547±0.020</b>	<b>7.1548±0.048</b>	<b>0.9354±0.085</b>

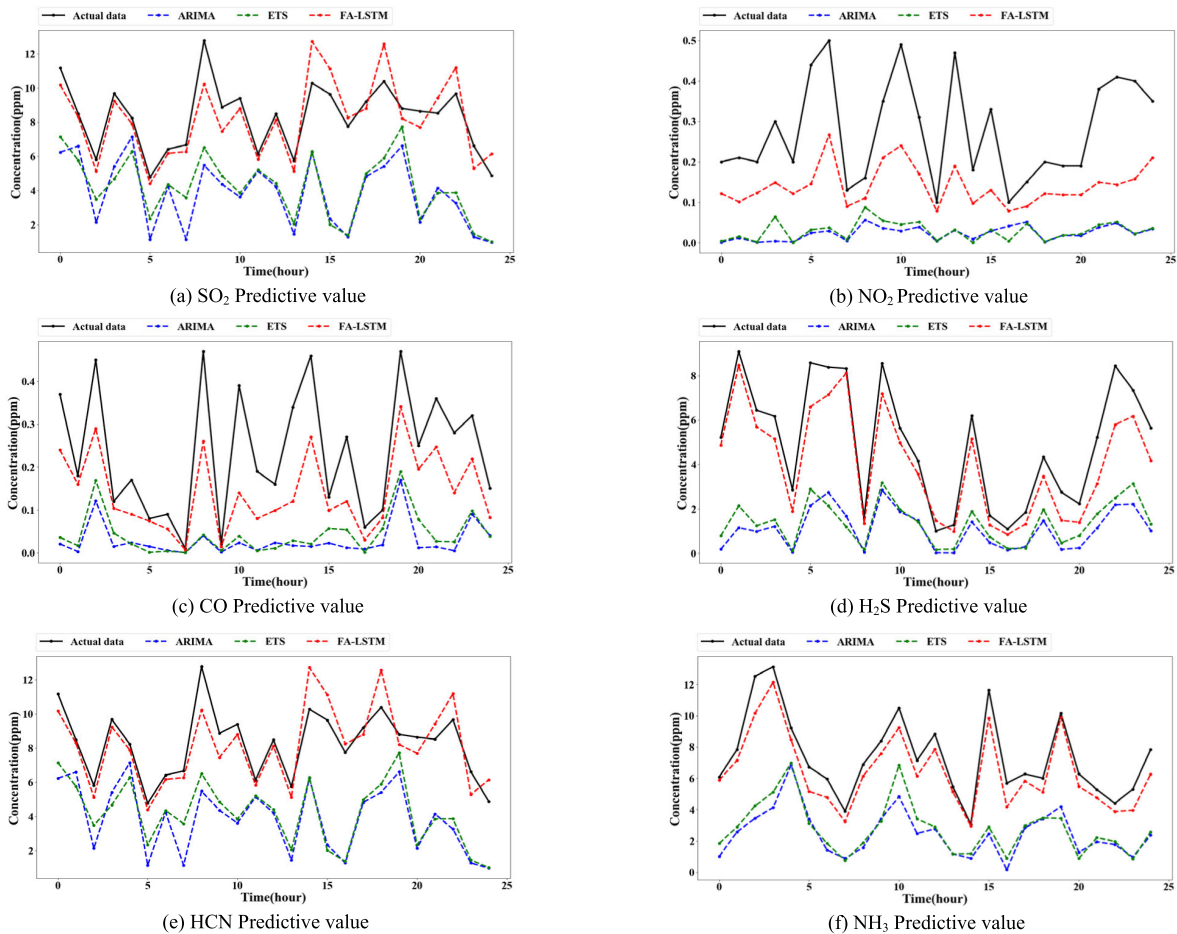


FIGURE 7. Predictions result of each traditional model on monitoring data.

machine learning method based on statistical learning theory used to solve classification and regression tasks [25].

This article is implemented and optimized using Scikit-Learn Package2 in Python; The neural network with hidden layer



**TABLE 6. Average predictive performance of ML models on test sets (5 months data).**

Toxic gas	Model	RMSE	MAE	r
SO <sub>2</sub>	SVM	20.1245±0.247	19.0145±0.158	0.6554±0.174
	BP	15.2154±0.258	14.3254±0.113	0.7014±0.142
	LSTM	<b>4.1453±0.012</b>	<b>3.1458±0.017</b>	<b>0.9415±0.047</b>
NO <sub>2</sub>	SVM	21.1514±0.176	17.2451±0.148	0.5216±0.136
	BP	16.2541±0.169	11.2548±0.249	0.6245±0.101
	LSTM	<b>6.2451±0.024</b>	<b>3.1245±0.475</b>	<b>0.8912±0.047</b>
CO	SVM	16.5484±0.132	7.1541±0.156	0.6648±0.201
	BP	12.0015±0.105	6.1024±0.123	0.8488±0.156
	LSTM	<b>6.1285±0.121</b>	<b>2.3145±0.814</b>	<b>0.9522±0.027</b>
H <sub>2</sub> S	SVM	8.0251±0.139	4.2151±0.110	0.7182±0.157
	BP	6.2410±0.117	3.5284±0.105	0.8145±0.126
	LSTM	<b>5.1479±0.054</b>	<b>3.1548±0.036</b>	<b>0.9457±0.022</b>
HCN	SVM	9.5414±0.254	8.5412±0.168	0.8541±0.162
	BP	8.8451±0.172	7.9841±0.095	0.8089±0.083
	LSTM	<b>6.1542±0.119</b>	<b>4.1475±0.102</b>	<b>0.9124±0.002</b>
NH <sub>3</sub>	SVM	28.2541±0.116	17.2541±0.184	0.7148±0.162
	BP	21.4786±0.055	12.2241±0.123	0.8124±0.131
	LSTM	<b>12.2547±0.020</b>	<b>7.1548±0.048</b>	<b>0.9354±0.085</b>

using back propagation algorithm is one of the widely used neural network types in time series prediction [26]. The design of BP neural network is realized by using Keras in Python. Figure 8 shows the comparison of actual data and predicted data in our proposed method and machine learning methods including support vector machine and neural network. The performance comparison results are shown in Table 6.

BP model and LSTM model designed in this paper are obviously superior to SVM model in RMSE, MAE and r, and the AVERAGE RMSE, MAE and r of BP model are 52.84%, 34.92% and 20.14% lower than that of this model. However, due to the generalization ability of BP neural network for complex nonlinear problems, it still has certain universality in time series prediction. Compared with the other two models, SVM model has no obvious advantages. The reason for this result may be that the SVM algorithm depends too much on the selection of parameters, and it is difficult to obtain the optimal parameters without external algorithm optimization. However, compared with Figure 7 and Figure 8, the accuracy of SVM is higher than that of traditional statistical prediction methods. Compared with artificial neural network, SVM with fewer parameters still has certain advantages in time series prediction.

From the perspective of data and conclusions, this paper introduces the LSTM deep learning model, which has good memory ability for historical information, into the prediction analysis of gas concentration in contaminated sites. The model shows good performance and minimum standard deviation in the prediction results of each month, indicating that the designed LSTM model is better than the traditional statistical theory and classical artificial intelligence algorithm in the prediction accuracy and stability, and has stronger adaptability and generalization ability. Although the gradient

problem of RNN has been solved in LSTM network to some extent, it is still very difficult for the sequence of high order. In addition, if the time span of the sequence is very large and the network is very deep, the calculation will be very heavy and time-consuming. In the subsequent work, we will further explore the research direction mentioned above. Future research can also consider using other deep learning methods, such as the prediction of relevant time series based on the attention mechanism network.

## VI. RELATED WORKS

Currently, there are a variety of methods for predicting the composition size of gaseous pollutants. The main methods can be divided into two categories: one is based on mathematical calculation and the other is based on historical sample modeling. Mathematical calculation models mainly include Gaussian diffusion model and computational fluid dynamics (CFD) model. Gaussian plume diffusion model predicts gas diffusion according to wind speed and direction and calculates according to diffusion expression.

The expression of calculation formula is relatively simple, and the calculation period is short, but this model is only applicable to gas flow on flat ground, the prediction of gas in complex geographical environment such as polluted sites is unreliable [27]. CFD model is suitable for predicting gas diffusion on complex urban terrain with dense buildings. Because the flexible representation of complex geometry determines the high calculation accuracy of the model, the model has good practicability for different types of urban terrain, but the calculation time of the model can be up to several hours. For emergencies, the emergency response capability of the model cannot effectively develop solutions for staff [3], [4].

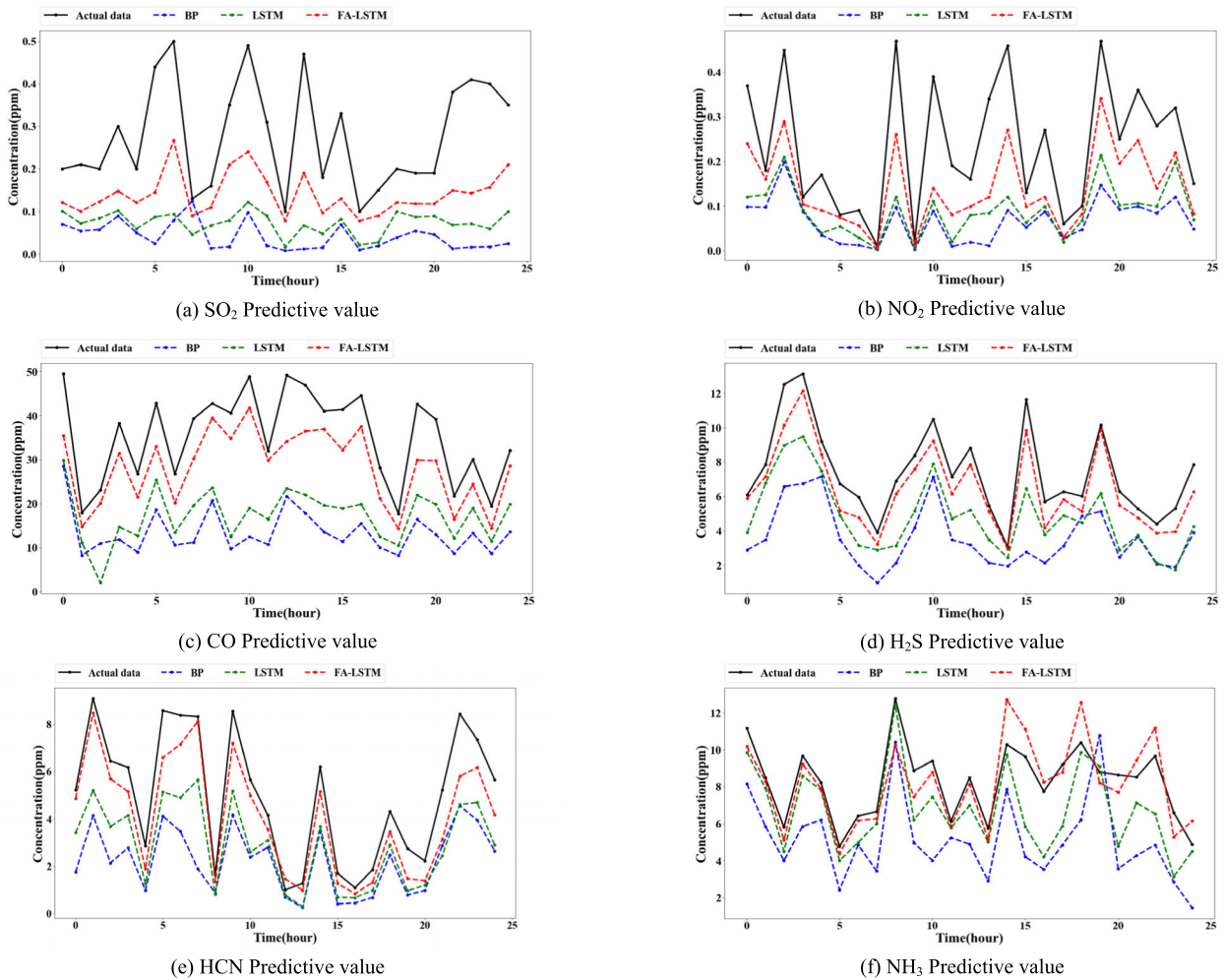


FIGURE 8. Predictions result of each machine learn model on monitoring data.

The other is based on historical sample modeling methods, including traditional modeling methods and emerging artificial intelligence methods. Traditional modeling methods based on historical samples include multiple linear regression method, Kalman filter technology, time-space series model method, etc. (existing time-space series models such as ARIMA [6], seasonal ARIMA [7], gaussian plume diffusion model [8]). The prediction method based on multiple linear regression approximates the gas diffusion model to linear model, which reduces the accuracy of prediction. The method based on spatio-temporal sequence mode requires experienced personnel to determine the parameters and order, which hinders the deployment of the model. To simplify the complexity of model building process, a series of pollutant prediction models based on machine learning are proposed [9]. For example, the model combining cellular automata and artificial neural network is used to predict the methane gas diffusion model in two-dimensional environment, which not only maintains a certain accuracy, but also is much better than the CFD model in calculation time [10]. In addition, the

gas pollution prediction model based on BP neural network has certain nonlinearity and generalization capability [11]. Although this kind of model can get rid of the dependence of the modeling process on experienced personnel, and the artificial neural network has some outstanding ability in gas prediction in the above research, the model has the problems of slow training speed and low prediction accuracy.

To solve the problems in machine learning model, Saratha Sathasivam proposed a recurrent neural network (RNN) [12]. RNN folds the hidden layer on the structure and iteratively calculates the data features of time series, so that it has certain memory ability. However, RNN has the problem of gradient disappearance, which makes it difficult for RNN to learn the state characteristics of long-term sequence data. Therefore, Sepp Hochreiter and Jürgen Schmidhuber proposed a variant of recurrent neural network, long-term and short-term memory neural network (LSTM) [13]. This method adds memory units to the neural units of RNN hidden layer, which effectively solves the problems of long-term dependence and gradient disappearance of information. The LSTM model has

been widely used in the field of environmental science. For example, a more effective and accurate wind speed prediction model was established based on LSTM, and the hourly solar radiation intensity was predicted and the carbon dioxide concentration in the forest environment was monitored according to the weather forecast data set [28]. Qian Fei, Chen Li et al directly applied the LSTM model designed to the gas diffusion problem, conducted experiments with the classical prairie grass project data set and compared the performance of other diffusion models, but this experiment only used the environmental characteristics in the data set as the input, ignoring the continuous influence of gas concentration [29]. Qin et al proposed a prediction scheme of urban PM<sub>2.5</sub> pollutant concentration based on convolutional neural network and LSTM. The model used CNN to extract the spatial representation between monitoring stations as the input feature. By learning the features contained in the air pollution concentration under the time series of historical data, LSTM was used to predict the future air pollution concentration. This method only pays attention to the time and space effect of monitoring data, and ignores the effect of wind speed, wind direction, temperature, atmospheric stability and other external factors on gas prediction [30]. For gas prediction in time series, gas concentration at each sampling point and current environmental parameters play a decisive role in the accuracy of prediction. Although work A also collects relevant weather data and distinguishes feature weights, its output is single, while our forecast output is multi-pollutant gas component and we consider the coupling relationship among each output.

## VII. CONCLUSION

In this paper, the LSTM deep learning model is applied to the real contaminated site treatment process, with the purpose of predicting the gas content of each component in the site, and helping the site personnel to make reasonable and scientific decisions. Combined with Dropout regularization technology, we constructed an optimized multi-layer LSTM network model and applied it to gas concentration prediction analysis. The toxic gas sampling data set is constructed based on the detection data of on-site multi-component toxic gas monitoring equipment. In the experiment, we compared the proposed method with some typical time series prediction techniques in the two dimensions of statistics and artificial intelligence. The experimental results show that the proposed method is superior to other methods in prediction accuracy and accuracy, and achieves the best performance in RMSE, MAE and r, followed by BP and SVM, which means that the means of artificial intelligence have strong ability in time series data analysis. In addition, the experimental results show that the performance of traditional statistical methods such as ARIMA and ETS is poor. In view of the strong self-learning ability, good generalization ability and high adjust ability of the model of LSTM network, its model has broad prospects in the prediction of toxic gas concentration with time series characteristics in the process of pollution control and prevention.

## ACKNOWLEDGMENT

The authors would like to thank the Engineering Research Center for Optoelectronic Devices and Communication Technology, Ministry of Education, Tianjin University of Technology, and the Tianjin Key Laboratory of Control Theory and Applications for Complex Systems for their assistance.

## REFERENCES

- [1] P. Guo, H. Li, and G. Zhang, "Contaminated site-induced health risk using Monte Carlo simulation: Evaluation from the brownfield in Beijing, China," *Environ. Sci. Pollut. Res.*, vol. 28, pp. 25166–25178, Jun. 2021.
- [2] J. Chen, K. de Hoogh, J. Gulliver, and B. Hoffmann, "A comparison of linear regression, regularization, and machine learning algorithms to develop Europe-wide spatial models of fine particles and nitrogen dioxide," *Environ. Int.*, vol. 130, Sep. 2019, Art. no. 104934.
- [3] M. Pontiggia, M. Derudi, M. Alba, M. Scaioni, and R. Rota, "Hazardous gas releases in urban areas: Assessment of consequences through CFD modelling," *J. Hazardous Mater.*, vol. 176, nos. 1–3, pp. 589–596, Apr. 2010.
- [4] S. R. Hanna, O. R. Hansen, M. Ichard, and D. Strimaitis, "CFD model simulation of dispersion from chlorine railcar releases in industrial and urban areas," *Atmos. Environ.*, vol. 43, no. 2, pp. 262–270, Jan. 2009.
- [5] S. Metia, S. Oduro, Q. Ha, and H. Duc, "Air pollution prediction using Matérn function based extended fractional Kalman filtering," in *Proc. 13th Int. Conf. Control Automat. Robot. Vis. (ICARCV)*, Dec. 2014, pp. 758–763.
- [6] R. Wongsathan and I. Seedadan, "A hybrid ARIMA and neural networks model for PM-10 pollution estimation: The case of Chiang Mai city moat area," *Proc. Comput. Sci.*, vol. 86, pp. 273–276, Mar. 2016.
- [7] S. Chattopadhyay and G. Chattopadhyay, "Modeling and prediction of monthly total ozone concentrations by use of an artificial neural network based on principal component analysis," *Pure Appl. Geophys.*, vol. 169, no. 10, pp. 1891–1908, Oct. 2012.
- [8] R. Wang, B. Chen, S. Qiu, Z. Zhu, Y. Wang, Y. Wang, and X. Qiu, "Comparison of machine learning models for hazardous gas dispersion prediction in field cases," *Int. J. Environ. Res. Public Health*, vol. 15, no. 7, p. 1450, Jul. 2018.
- [9] D. Ma and Z. Zhang, "Contaminant dispersion prediction and source estimation with integrated Gaussian-machine learning network model for point source emission in atmosphere," *J. Hazardous Mater.*, vol. 311, pp. 237–245, Jul. 2016.
- [10] P. Lauret, F. Heymes, L. Aprin, and A. Johannet, "Atmospheric dispersion modeling using artificial neural network based cellular automata," *Environ. Model. Softw.*, vol. 85, pp. 56–69, Nov. 2016.
- [11] J. Ouyang and A. Wang, "The application of concentration forecasting of air pollutant based on BP neural network in MATLAB," *Environ. Sci. Manage.*, to be published.
- [12] H. Hewamalage, C. Bergmeir, and K. Bandara, "Recurrent neural networks for time series forecasting: Current status and future directions," *Int. J. Forecasting*, vol. 37, no. 1, pp. 388–427, 2021.
- [13] C. Gulcehre, K. Cho, R. Pascanu, and Y. Bengio, "Learned-norm pooling for deep feedforward and recurrent neural networks," in *Proc. Joint Eur. Conf. Mach. Learn. Knowl. Discovery Databases*. Berlin, Germany: Springer, 2014, pp. 530–546.
- [14] L. Xu, C.-S. Choy, and Y.-W. Li, "Deep sparse rectifier neural networks for speech denoising," in *Proc. IEEE Int. Workshop Acoustic Signal Enhancement (IWAENC)*, Sep. 2016, pp. 1–5.
- [15] N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever, and R. Salakhutdinov, "Dropout: A simple way to prevent neural networks from overfitting," *J. Mach. Learn. Res.*, vol. 15, no. 1, pp. 1929–1958, 2014.
- [16] Y.-Q. Zhang and X.-Y. Wang, "A new image encryption algorithm based on non-adjacent coupled map lattices," *Appl. Soft. Comput.*, vol. 26, pp. 10–20, Jan. 2015.
- [17] D. J. Gunn, Z. Liu, R. Dave, X. Yuan, and K. Roy, "Touch-based active cloud authentication using traditional machine learning and LSTM on a distributed tensorflow framework," *Int. J. Comput. Intell. Appl.*, vol. 18, no. 4, Dec. 2019, Art. no. 1950022.
- [18] H. Li, X. Mu, Y. Yang, and A. J. Mason, "Low power multimode electrochemical gas sensor array system for wearable health and safety monitoring," *IEEE Sensors J.*, vol. 14, no. 10, pp. 3391–3399, Oct. 2014.

- [19] M. Akhoondzadeh, "A MLP neural network as an investigator of TEC time series to detect seismo-ionospheric anomalies," *Adv. Space Res.*, vol. 51, no. 11, pp. 2048–2057, Jun. 2013.
- [20] Y. Su, Y. Zhao, M. Sun, S. Zhang, and X. Wen, "Detecting outlier machine instances through Gaussian mixture variational autoencoder with one dimensional CNN," *IEEE Trans. Comput.*, early access, Mar. 9, 2021, doi: [10.1109/TC.2021.3065073](https://doi.org/10.1109/TC.2021.3065073).
- [21] X. H. Cao, I. Stojkovic, and Z. Obradovic, "A robust data scaling algorithm to improve classification accuracies in biomedical data," *BMC Bioinf.*, vol. 17, no. 1, p. 359, Sep. 2016.
- [22] R. Hemeimat, "Demand forecasting for motor vehicle spare parts," *Amer. J. Oper. Res.*, pp. 113–120, 2016.
- [23] D. Haboudane, J. R. Miller, E. Pattey, P. J. Zarco-Tejada, and I. B. Strachan, "Hyperspectral vegetation indices and novel algorithms for predicting green LAI of crop canopies: Modeling and validation in the context of precision agriculture," *Remote Sens. Environ.*, vol. 90, no. 3, pp. 337–352, 2004.
- [24] R. Taylor, "Interpretation of the correlation coefficient: A basic review," *J. Diagnostic Med. Sonogr.*, vol. 6, no. 1, pp. 35–39, 1990.
- [25] L. Wang, "Support vector machines: Theory and applications," in *Machine Learning and Its Applications* (Advanced Lectures). Springer-Verlag, Jan. 2001.
- [26] F. S. Wong, "Time series forecasting using backpropagation neural networks," *Neurocomputing*, vol. 2, no. 4, pp. 147–159, Jul. 1991.
- [27] A. Mazzoldi, T. Hill, and J. J. Colls, "CFD and Gaussian atmospheric dispersion models: A comparison for leak from carbon dioxide transportation and storage facilities," *Atmos. Environ.*, vol. 42, no. 34, pp. 8046–8054, Nov. 2008.
- [28] H. Liu, X. Mi, and Y. Li, "Smart multi-step deep learning model for wind speed forecasting based on variational mode decomposition, singular spectrum analysis, LSTM network and ELM," *Energy Convers. Manage.*, vol. 159, pp. 54–64, Mar. 2018.
- [29] F. Qian, L. Chen, J. Li, C. Ding, X. Chen, and J. Wang, "Direct prediction of the toxic gas diffusion rule in a real environment based on LSTM," *Int. J. Environ. Res. Public Health*, vol. 16, no. 12, p. 2133, Jun. 2019.
- [30] J. Ni, H. Yang, J. Yao, Z. Li, and P. Qin, "Toxic gas dispersion prediction for point source emission using deep learning method," *Hum. Ecol. Risk Assessment: Int. J.*, vol. 26, no. 2, pp. 557–570, Jan. 2019.



**KE TANG** was born in Henan, China, in 1990. He received the M.S. degree in international business, in 2013. He is currently pursuing the Ph.D. degree in international economics with the University of International Business and Economics.



**GE WANG** was born in Shaanxi, China, in 1995. She is currently pursuing the degree major in business administration with the College of Management and Economics, Tianjin University.



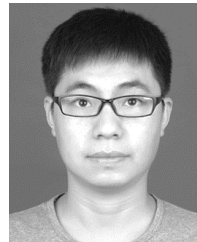
**YANFEI HU** was born in Henan, China, in 1989. He received the B.S. and M.S. degrees in automation, in 2012 and 2015, respectively. He is currently pursuing the Ph.D. degree major in computer science with the University of Chinese Academy of Science. He had worked as a Software Engineer in a company that develops automotive electronics. His research interest includes explanation in deep learning model.



**YU CONG** was born in Tianjin, China, in 1996. He is currently pursuing the degree major in control science and technology with the School of Electrical and Electronic Engineering, Tianjin University of Technology.



**XIMENG ZHAO** was born in Jilin, China, in 1987. She received the M.S. degree in national economics, in 2012. She is currently pursuing the Ph.D. degree in international economics with the University of International Business and Economics. After five years working for companies, such as Detecon consulting and Mercedes-Benz.



**YINGKUI JIAO** was born in Henan, China, in 1981. He received the B.S. and M.S. degrees in automation, in 2004 and 2015, respectively. He is currently pursuing the Ph.D. degree in precision instruments and machinery with Tianjin University. He had worked as a Software Engineer in a company that develops electronics device.

...