IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received 7 October 2022, accepted 20 October 2022, date of publication 26 October 2022, date of current version 3 November 2022.

Digital Object Identifier 10.1109/ACCESS.2022.3217242

RESEARCH ARTICLE

A Hybrid Air Quality Index Prediction Model Based on CNN and Attention Gate Unit

JINGYANG WANG[1], LUKAI JIN[1], XIAOLEI LI[1], SIYUAN HE[2], MIN HUANG[1], AND HAIYAO WANG[3]

[1]School of Information Science and Engineering, Hebei University of Science and Technology, Shijiazhuang 050018, China
[2]FedUni Information Engineering Institute, Hebei University of Science and Technology, Shijiazhuang 050018, China
[3]School of Ocean Mechatronics, Xiamen Ocean Vocational College, Xiamen 361100, China

Corresponding author: Min Huang (huangmin@hebust.edu.cn)

This work was supported in part by the Innovation Foundation for Postgraduate of Hebei Province under Grant CXZZSS2022082, and in part by the Foundation of Hebei University of Science and Technology under Grant 2019-ZDB02.

ABSTRACT Air Quality Index (AQI) is the crucial foundation for measuring air quality, which reflects the influence of air quality on people's health and life to a certain extent. In this paper, a hybrid AQI prediction model based on Convolutional Neural Network (CNN) and Attention Gate Unit (AGU) is proposed to deal with the problems of the ''vanishing gradient'' and ''exploding gradient'' of Recurrent Neural Network (RNN). AGU is a new model proposed in this paper to embed the attention mechanism and Data Adjustment Module (DAM) into the gated unit. The attention mechanism enhances the learning ability of the gated unit, and the DAM makes the gated unit more sensitive to historical data learning. In this model, CNN plays a role in extracting features from time series data. AGU can make differentiated learning of historical data and finally produce prediction results. The model evaluation indexes used in the experiments are Mean Absolute Error (MAE), Mean Square Error (MSE), and R Squared (R^2). The experimental results show that the overall performance of the AQI prediction model based on CNN-AGU is superior to that of other models by comparing with the other nine models on the same data set.

INDEX TERMS CNN, AGU, AQI prediction, machine learning.

I. INTRODUCTION

Air pollution has caused great harm to the development of the national economy and people's physical and mental health, which has attracted significant attention from all walks of life. AQI is used to evaluate air quality, and the evaluation results can fully reflect the actual air quality [1], [2], [3]. The prediction of AQI has been one of the research hotspots in air pollution monitoring and treatment for a long time.

According to the regulations of the relevant environmental protection departments, if the AQI value is within the range of 0-50, the AQI level is one, and the air quality status is excellent. If the AQI value is within the range of 51-100, the AQI level is two, and the air quality status is good. If the AQI value is 101-150, the AQI level is three, and the air quality status belongs to mild pollution. If the AQI value is 151-200, the AQI level is four, and the air quality status belongs to medium pollution. If the AQI value is within the range of 201-300, the AQI level is five, and the air quality status belongs to heavy pollution. If it is above 300, the AQI level is six, and the air quality status belongs to very serious pollution [4].

The air pollution problem is becoming more and more serious, which has attracted significant attention from all walks of life. Various countries have established air quality supervision mechanisms to better monitor and improve air quality to ensure people's well-being and economic construction and prevent pollution incidents. These initiatives have accumulated many historical monitoring data on air quality. The vast amount of data brings new opportunities to study air quality and challenges in making good use of them. Therefore, it is of great social significance and value to construct an efficient AQI prediction model to treat air pollution effectively.

The continuous development of machine learning makes it possible to predict AQI accurately. The standard neural

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

network is not suitable for processing time series data, but the network structure of RNN connects each time state in a series. The latter state is affected by the former state, so RNN effectively the time series data processing is realized [5], [6]. However, when the amount of data is significant, RNN training often has the problems of ''vanishing gradient'' and ''exploding gradient.'' Long Short-Term Memory (LSTM) and Gated Recurrent Unit (GRU) are proposed, which can effectively alleviate the problems of ''vanishing gradient'' and ''exploding gradient'' [7], [8], [9]. LSTM and GRU have solved the gradient problem of RNN to some extent, but it is still not enough. They still cannot learn the relationship between data more deeply when dealing with longer sequences.

Through the study of LSTM and GRU, it is concluded that gated technology is used to decide the forgetting and saving data [10], [11], [12]. However, these gated units have some defects in processing time series data. In terms of the time dimension, the importance of meteorological data features is not consistent, and the features hidden in abnormal meteorological data are more important than those in normal meteorological data. This kind of difference in temporal features is difficult to be reflected by standard LSTM and GRU. Based on this problem, this paper proposes an AGU, which embeds the attention mechanism into the gated unit to enhance the learning ability of the gated unit. At the same time, AGU introduces DAM to avoid the error of prediction results caused by the saturation of function, which affects the prediction accuracy, so that the gated unit can learn the historical data more thoroughly. AGU combines the advantages of GRU and attention mechanism, GRU has fewer network parameters, and the attention mechanism has a strong learning ability. AGU uses the attention mechanism to make up for the ability of the gated unit to pay attention to specific features adaptively and deals with the problem of data redundancy or loss in the learning process. Usually, the attention mechanism is often used as an additional layer of RNN to form a combined model. Compared with the combined model, AGU makes the model more straightforward and easier to train. Although AGU makes up for the shortcomings of the existing gated technology that cannot pay attention adaptively to specific features and learn thoroughly, a single AGU cannot fully extract data features. This paper combines CNN efficiently extracting data features and AGU to propose a CNN-AGU model for AQI prediction. The accuracy and effectiveness of the proposed CNN-AGU model are verified by comparing the prediction results of nine different models through experiments.

To sum up, the major contributions of this paper are as follows:

(1) This paper proposes a new model, AGU, which embeds the attention mechanism and DAM into the gated unit. The attention mechanism enhances the learning ability of the gated unit. DAM makes the gated unit more sensitive to historical data learning.

(2) Compared with LSTM and GRU, the MAE, MSE, and R^2 of AGU are better than LSTM and GRU in predicting AQI with the same data set and experimental environment.

(3) This paper proposes a new hybrid AQI prediction model based on CNN and AGU, CNN-AGU. The introduction of CNN achieves the extraction of characteristic values well and improves the CNN-AGU model's prediction accuracy about AQI.

The rest of the paper is organized as follows: Section II introduces the literature review and analysis of related models. Section III presents the principle of CNN, AGU, and the architecture of the CNN-AGU. Section IV introduces the experimental tools, experimental data, data set sources, data preprocessing, data set segmentation, models adjustment and validation, experimental results. Section V includes the analysis and discussion of the experimental results. Section VI concludes the work of this paper and introduces further research.

II. RELATED WORK

As for the study of AQI and the concentration of various pollutants in the air, people used traditional statistical methods such as multiple regression to study the variation rule of pollutant concentration in the early stage. Abdul-Wahab et al. used pollutants in the air and various meteorological factors as variables and used multiple regression analysis methods to fit the model of O_3 concentration change. The experiment showed that this method was effective in predicting O_3 concentration [13]. The study combined Multiple Linear Regression (MLR) with Principal Components Analysis (PCA) to solve the problem of multicollinearity among independent variables. The MLR model was one of the most widely used methods to express the dependence of dependent variables on several independent variables. But the disadvantage of this model was that it was limited to solving the linear problem with a small data variation range or the problem with less severe nonlinearity. Most problems that needed to be solved were nonlinear problems, but the MLR model could not produce a high degree of fitting for nonlinear problems. The commonly used numerical prediction models included the WRF-Chem and CAMQ models in the United States. These methods explored a series of complex changes in pollution data and the atmospheric environment, and they could accurately predict the concentration of some factors in the air [14], [15]. The WRF-Chem model could simulate a more realistic atmospheric environment. It added many chemical modules into the WRF framework to realize the complete online coupling of meteorological and chemical models. This method made up for the loss of important information about atmospheric processes due to the separation of meteorology and chemistry caused by the offline transmission of the CAMQ model. Models-3 CMAQ promoted environmental modeling technology by integrating physical and chemical algorithms, realized the sharing of technologies and models, and promoted the collaborative

development of meteorological and air quality modules in the system. However, these numerical prediction methods involved a variety of fields, such as mathematics, physics, chemistry, and meteorology. And they required various more complex data and high-performance computer equipment for modeling and analysis.

In recent years, machine learning algorithms have also made some air quality prediction achievements. Some scholars used traditional machine learning algorithms to predict air quality and pollutant concentration. Sun et al. used Random Forest Regression (RFR) to estimate the solar radiation level based on the air quality data set. The experimental results showed that the predicted values were close to the observed values, and the RFR model performed well in predicting the solar radiation value [16]. Still, RFR risked over-fitting to predict nonlinear air quality data. Silva et al. used Support Vector Regression (SVR) to predict AQI in California. The experimental results showed that SVR could accurately predict pollutant concentration and AQI within hours [17]. However, the single traditional SVR model is often due to the lack of computing power, which is unsuitable for training on large-scale data sets.

RNN had advantages in processing time series data, and more and more scholars began to apply RNN to AQI research. Kim et al. used a data-driven air quality prediction model and applied RNN to it. Compared with other data-driven prediction modeling methods, the prediction results of RNN had better interpretability and modeling performance [18]. The study used Partial Least Squares (PLS) to select the input variables of the RNN model to optimize the prediction model. However, the problems of the ''vanishing gradient'' and ''exploding gradient'' caused by the long-term dependence of data on the model would occur in the training of RNN [19]. He and Luo used transfer entropy to select meteorological factors that affected significant changes in AQI and studied the prediction accuracy of AQI in different prediction times in the next 0-48 hours through LSTM. This study showed that LSTM had better robustness than traditional neural networks and had the advantage of short-term prediction [20]. Pu et al. improved the wavelet denoising method to process air quality data and constructed LSTM to predict AQI. The results showed that LSTM had good applicability in air quality prediction [21]. The above two studies showed that LSTM greatly affected processing time series tasks. Although LSTM could effectively alleviate the ''vanishing gradient'' and ''exploding gradient'' problems, training a single LSTM would produce instability. For more extended sequence data, capturing long-term interdependence and achieving high-precision prediction was challenging. The Paperswirhcode website showed that the optimal model for sunspot time series tasks was BD-LSTM (Bi-directional Long Short-Term Memory). The study verified LSTM on seven different time series problems, and experimental results showed that BD-LSTM performed well in processing time series tasks [22]. Although BD-LSTM could process the dependency relationship between data forward and backward, it still could not transmit the long-distance dependence information well for the excessively long sequence. Zhou et al. used GRU to train four models with PM2.5 concentration data in four seasons in Beijing. The results showed that the prediction accuracy of the GRU model was high, and the method was effective for the time series prediction of air pollutants [23]. Becerra-Rico et al. studied PM10 data in Mexico City through GRU. The experimental results showed that GRU was better than LSTM in predicting time length. LSTM and GRU could predict such highly nonlinear data without a significant difference in performance [24]. But the single GRU model could not extract the input data features efficiently. Xie et al. used CNN and GRU to predict PM2.5 concentration. They used CNN to extract data features automatically and input CNN results into the GRU network to learn the dependence between air quality data further. Moreover, the combined model could predict PM2.5 concentration with high accuracy [25]. There was still the problem of feature learning, such as the GRU could not constantly remember or forget data according to the importance of data. In the future, when the data and model volumes were gradually increasing, further enhancing the learning ability of models was the bottleneck of GRU development. Mei et al. proposed a CNN-GRU-Attention model to realize the prediction of time series data. The study found that the prediction accuracy of the combined model was better than that of the single model. And compared with other models, the prediction effect of the CNN-GRU-Attention was more accurate, and the stability of the model was enhanced [26]. However, integrating multiple models would lead to the complexity of the whole combined model, which increased the amount of computation and was inconvenient for model training.

III. MODELS

A. CNN

CNN can achieve the parameters sharing mechanism on the one hand. Different regions share the same filter, so they share the same set of parameters [27]. Through this mechanism, the number of parameters of the network is significantly reduced, the amount of calculation is reduced, and over-fitting is effectively avoided. On the other hand, the local connection structure of CNN makes any output unit only related to the part of the input units [28]. In addition, CNN allows data to be defective and distorted, so CNN has good compatibility with the data.

CNN is a good feature learner, which can automatically extract features according to the problem. 1D convolution is often used for series data, such as air pollutant concentration, stock index, and other time series data. 2D convolution is often used in the field of image processing, which can extract spatial dimension features of images [29]. 3D convolution enables the convolution kernel to perform operations in the spatial and temporal dimensions, which can well integrate the information of these dimensions. 3D convolution can be used to deal with problems such as videos that have not only spatial features of images but also temporal features [30].

Time series data can be directly used as the input of CNN to effectively decrease the complexity of feature extraction. The trained CNN can efficiently and accurately realize feature extraction of time series data [31], [32], [33].

B. AGU

As shown in Fig. 1, GRU generates the next predicted value, relying only on the output state of the previous cell and the input value of the current input. In the attention mechanism, the generation of each predicted value will be related to all historical data [34].

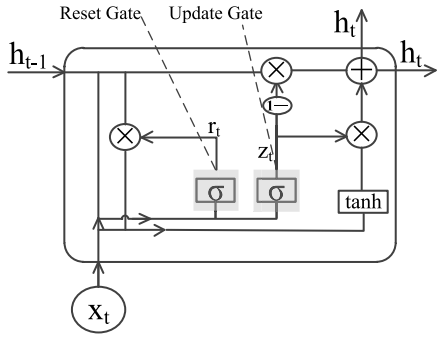


FIGURE 1. GRU structure.

The principle of Attention is to calculate the matching score between the current input sequence and the output vector. The higher the matching score, that is, the higher the relative score of the focus point, and then summarize the input information by weighted average.

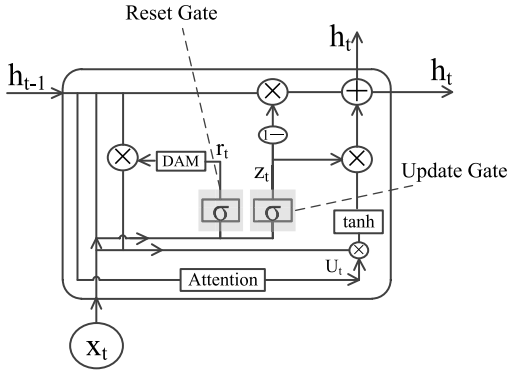


FIGURE 2. AGU structure.

As shown in Fig. 2, compared with GRU, AGU introduces the DAM and attention mechanism module inside the gated unit. The reset gate r_t controls how much past information is to be forgotten, and its calculation process is the same as the reset gate of GRU, as shown in formula (1).

$$r_t = \sigma\left(W_r \cdot [h_{t-1}, x_t] + b_r\right) \tag{1}$$

In the formula, σ represents the sigmoid activation function, x_t represents the input vector of the current cell, h_{t-1} represents the hidden state passed down from the previous cell, and W_r and b_r represent the weight and bias of the reset gate, respectively.

The update gate z_t helps the model solve how much past information should be transferred to the future, and its calculation process is the same as the update gate of GRU, as shown in formula (2).

$$z_t = \sigma\left(W_z \cdot [h_{t-1}, x_t] + b_z\right) \tag{2}$$

In the formula, W_z and b_z represent the weight and bias of the update gate, respectively.

When x is less than -6 or greater than 6, the sigmoid function value will approach 0 or 1, respectively. Therefore, when the input data are in $(-\infty, -6)$ or $(6, +\infty)$, the mapping function value will not change significantly. That is, it will approach a saturation state, which will lead to a decrease in learning sensitivity. Therefore, DAM is introduced into AGU in this paper. The module can map the sigmoid function values approaching 0 and 1 as x on the tanh activation function. According to the size of the input value, the mapping function value will also change significantly, which improves the model's sensitivity to the input data and increases the nonlinearity of the neural network model, as shown in formula (3).

$$\mathrm{DAM} = \tanh(r_t) \tag{3}$$

In the formula, r_t stands for the output of the reset gate, and tanh stands for the activation function.

$$\alpha_t = \sigma\left(W_\alpha \cdot [h_{t-1}, x_t] + b_\alpha\right) \tag{4}$$

$$U_t = \mathrm{softmax}(\alpha_t) \tag{5}$$

$$h'_t = \tanh(U_t \cdot [\mathrm{DAM} \cdot h_{t-1}, x_t] + b_t) \tag{6}$$

α_t represents the matching score between the input sequence x_t and h_{t-1}. And σ is used to calculate the matching score, as shown in formula (4). The attention probability distribution U_t is obtained by normalizing α_t with the softmax function, as shown in formula (5). Firstly, the matching score α_t is calculated by h_{t-1} and x_t, and then the adaptive weighted average value U_t is obtained so that it is more scientific to calculate and sum historical data. The attention distribution U_t is taken as the weight of the input sequence of the current time and the information of the previous time to be forgotten. That is, the weighted sum of x_t and $\mathrm{DAM} \cdot h_{t-1}$ is carried out. Finally, the candidate hidden state h'_t is computed by the tanh activation function, as shown in formula (6).

The phase of updating h_t uses the same z_t to forget and remember simultaneously. The operation of this step is to selectively forget some unimportant information in h_{t-1} passed down from the previous step and selectively remember h'_t containing the current step information, as shown in formula (7).

$$h_t = (1 - z_t) \cdot h_{t-1} + z_t \cdot h'_t \tag{7}$$

C. CNN-AGU

Fig. 3 shows the overall architecture of the CNN-AGU. The model consists of four parts: data preprocessing layer, CNN layer, AGU layer, and output layer, which is described in detail as follows.

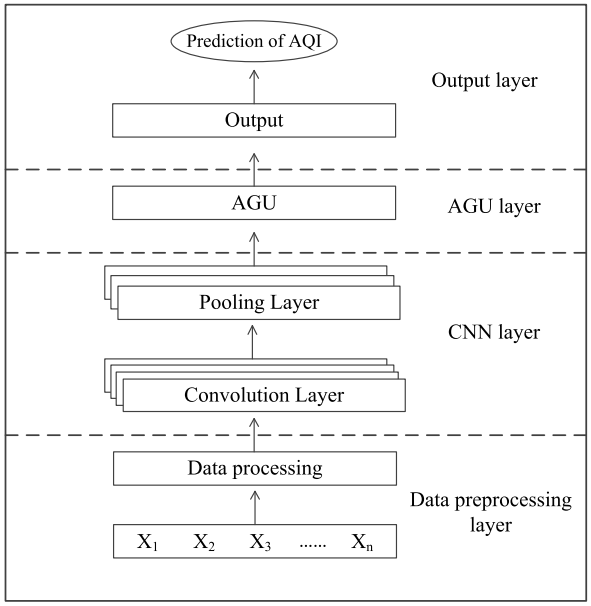


FIGURE 3. CNN-AGU model architecture.

(1) Data preprocessing layer: For different abnormal data, different methods are used to deal with them, such as forward recurrent complement, feature numerical replacement, and so on; for the significant difference between values of different variables, standardized processing is carried out. The data set is divided into three parts according to the partition method introduced in Section IV.C.

(2) CNN layer: The preprocessed data are concatenated into a two-dimensional structure, where the rows represent the time length, and the columns represent the dimension of the features. The batch-sized samples are used for batch training during modeling and then enter the CNN layer. This layer uses the convolution kernel to complete the convolution operation and extracts the features information of data in different dimensions. And then, the pooling layer reduces the dimension of the extracted information to alleviate the overfitting of the model. After CNN transformation, the output results are concatenated into a matrix, which can be regarded as the new expression of features.

(3) AGU layer: The filtered new features are passed as input to the AGU layer, and the layer is used to train the time series rules in the data. The AGU's update gate controls how much useful information needs to be passed down at the previous and current moment. The reset gate and the DAM of the AGU control how much past information needs to be forgotten. The DAM maps data changes on different activation functions to enhance the learning ability of the model. Then, the layer receives the sequence information and the hidden state at the previous moment in the attention gate and learns the weight matrix, which can express the importance of the data. The weight matrix is used to calculate the candidate hidden state, which helps the memory cells quickly discard the unimportant content and enhances the learning ability of the model. The AGU layer takes the hidden state of the last moment in the time series as its output.

(4) Output layer: Finally, the final output value, namely the AQI value, is obtained through the weighted sum and processing of the fully connected layer.

IV. EXPERIMENTS

A. EXPERIMENTAL TOOLS AND DATA SET

In the experiments, the CNN-AGU model is built using TensorFlow 1.14.0-GPU, Keras2.1.0 deep learning platform, and PyCharm 2018 3.5 × 64 development tool. The hourly air quality data and meteorological data from 00:00 on January 1, 2018, to 23:00 on June 30, 2021, in Shijiazhuang, Hebei Province, China, are selected as the data set. These data are obtained from the Shanghai Qingyue Data website and China Urban Meteorological Data interface (NowAPI), and a total of 39408 data pieces are collected.

According to China's air quality evaluation standards, the primary pollutants involved in AQI evaluation are SO_2, O_3, PM2.5, NO_2, PM10, and CO. By analyzing the factors affecting AQI, it is found that meteorological conditions have a certain influence on the diffusion, dilution, and accumulation of pollutants. Under situations of relatively stable air pollution emissions, AQI is closely related to typical meteorological factors such as precipitation, temperature, and wind [35], [36]. In the experiment, meteorological factors such as temperature, humidity, wind, and weather that affect AQI are added to the air quality data set. Therefore, the inputs involved in AQI prediction include these meteorological factors and six primary pollutants.

B. DATA PREPROCESSING

1) MISSING DATA HANDLING

This paper analyzes the missing data from the attributes and duration and takes corresponding processing methods to fill in the specific missing data. Therefore, this paper proposes a combined missing value processing method, as shown in Table 1.

TABLE 1. Combination of missing value processing methods.

duration of the missing data	Attributes of missing data	
	pollutant concentration and meteorological data	AQI
1h	forward recurrent complement	definition of AQI
t>1h	moving average method	definition of AQI

According to the combination of missing value processing methods proposed in this paper, specific data missing situations will be elaborated as follows:

(1) In view of the lack of AQI, according to the calculation formula of AQI, its value is determined by the maximum value of the Individual Air Quality Index (IAQI) of various pollutants, which is related to the corresponding pollutant

TABLE 2. Partial original data.

Date	AQI	SO_2 (μg/m3)	O_3 (μg/m3)	PM2.5 (μg/m3)	NO_2 (μg/m3)	PM10 (μg/m3)	CO (μg/m3)	Temperature (°C)	humidity (%)	wind	weather
2020/2/28 2:00	68	10	29	49	31	70	0.8	0°C	90%	3	Overcast
2020/2/28 3:00	68	10	31	49	28	73	0.8	-1°C	92%	3	Overcast
2020/2/28 3:00	68	10	31	49	28	73	0.8	-1°C	92%	3	Overcast
2020/2/28 4:00	75	10	35	55	26	80	0.8	-1°C	92%	3	Overcast

concentration and has nothing to do with the duration of the missing data. Therefore, no matter how long the duration of the missing data is, it is necessary to analyze the concentration values of various pollutants first to judge whether there is any missing situation. If there is any missing concentration value, the corresponding missing value processing should be carried out according to the duration of the missing data. The missing AQI will be assigned according to the maximum method when all the pollutant concentration values exist.

(2) For the lack of pollutant concentration and meteorological data, and considering the duration of the missing data, there will be the following three situations:

1) There is no value for the concentration of a particular pollutant, and the corresponding AQI exists but is not the maximum value of other IAQI. In this case, the IAQI of the missing pollutant is equal to the AQI value, from which the concentration value of the pollutant can be back-calculated;

2) Suppose one or more data is missing for one hour. In this case, if AQI exists and is the maximum value in the existing pollutants IAQI data, or AQI does not exist, it is necessary to fill each pollutant or meteorological data by forwarding recurrent complement [37], as shown in formula (8):

$$X_t^d = X_{t-1}^d \tag{8}$$

In the formula, X_{t-1}^d represents the effective observed value of d-dimensional component data at the previous time;

3) For the data missing for more than one hour, if the forward recurrent complement method is used, the filling value of certain data in consecutive missing periods will be constant. The data are so smooth that the network cannot find connections between the data well. In this case, the moving average method fills the missing data. As shown in formula (9):

$$X_t^d = (X_{t-1}^d + X_{t-2}^d + \ldots + X_{t-k}^d)/k \tag{9}$$

Firstly, a time window is set, which is denoted as k. The moving average method means this time window slides along the time axis to get the average value. When calculating the average value, the data of $t-1, t-2, \ldots, t-k$ time steps are used. X_{t-k}^d represents the d-dimensional component data of k time steps before the current missing time.

2) DATA NORMALIZATION

Some original data have abnormal situations, such as inconsistent format and repeated data, as shown in Table 2. Therefore, before using this data set for model training and evaluation, the experiment needs to process the original data, such as cleaning out abnormal data and standardizing the format.

The data in the table at 3:00 on February 28, 2020, is duplicated. In the experiment, the last item of repeated data is saved, and the repeated data before this item is deleted. To ensure the uniform format of the input data, delete the unit symbol of the data, such as temperature data with ° and humidity data with the % symbol. Features are sometimes not always continuous value but may be classified value. For example, weather situations can be divided into "Overcast" and "Sunny." The experiment usually needs to digitize them [38], as shown in Table 3.

TABLE 3. Weather data preprocessing.

Weather data	Quantification	Conversion data
Haze	0001	1
Fog	0010	2
Sunny	0011	3
Cloudy	0100	4
Overcast	0101	5
Rain, Sleet, Light snow, Snow, Light rain	0110	6
Moderate snow, Moderate rain	0111	7
Heavy torrential rain, Thunder shower, Torrential rain, Heavy rain, Blizzard, Shower, Heavy snow, Hail, Snow shower	1000	8
Extremely torrential downpours	1001	9
Freezing rain , Sandstorm , Floating dust , Dusty weather , Severe sandstorm	1010	10

The sample data of the experiment are multi-dimensional. That is, multiple features represent a sample. In the problem of AQI prediction, the factors affecting AQI include pollutants such as CO, NO_2, and some meteorological factors. The dimensions of features and magnitude of value are different. If the original data are used directly, their influence on AQI will be different. While different features can have the same scale through normalization, thus reducing the influence of different values on model training, speeding up the model convergence, and making comprehensive evaluation and analysis. Min-Max normalization is used in the experiment to carry out corresponding linear transformation processing on sample data to make the function values kept

in the interval [0, 1] [39]. The whole process does not destroy the existing data structure, as shown in formula (10).

$$y_i = \frac{x_i - \min\limits_{1\le j\le n}\{x_j\}}{\min\limits_{1\le j\le n}\{x_j\} - \min\limits_{1\le j\le n}\{x_j\}} \tag{10}$$

3) THREE-DIMENSIONAL TIME SERIES DATA CONSTRUCTION

Time series data refer to the data column information recorded in a time sequence, obtained by indexing the time dimension. The air quality data and meteorological data studied in this paper are also time series data. The obtained data set S can be represented by $\{X_1, X_2, \ldots, X_n\}$, where n represents the number of time. Each time step X_t is a d-dimensional vector representing the set of pollutant concentration observations and meteorological data at a certain time point. X_t can be expressed as $\{C_1, C_2, \ldots, C_{11}\}$, where C_1-C_7 are AQI value and CO, NO_2, O_3, PM10, PM2.5, and SO_2 concentration, respectively, and C_8-C_{11} are temperature, humidity, wind, and weather respectively. Therefore, the dimension of the data set is (n, d). Then the time series sample set generated by the data set is three-dimensional data with the size of (N, T, D), where N is the number of time slices, that is, the number of samples, T is the number of time steps in a sample, and D is the number of features of the input data. In the experiment, the data set is constructed according to the setting that the step size is one and the sequence length is 24. Finally, the sample set's dimensions are (Y-23, 24, 11), where Y represents the number of time, as shown in Fig. 4.

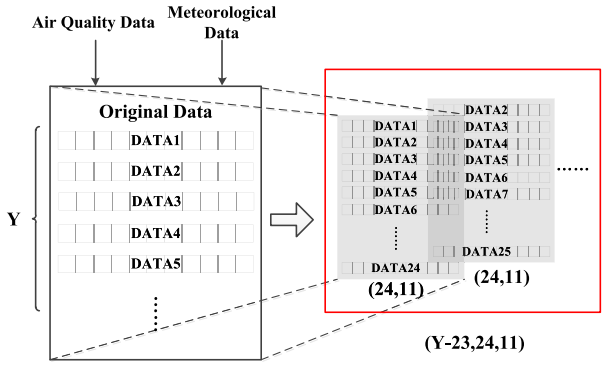


FIGURE 4. Time series construction process.

C. DATA SET SEGMENTATION

The sample data of the experiment are limited, so we divide the sample set into training set, verification set, and test set according to the ratio of 6: 2: 2 [40]. Firstly, 7877 samples are intercepted from the end of 39385 samples as the test set. Then 23,631 learning samples are intercepted as the training set and 7877 as the validation set.

D. MODELS ADJUSTMENT AND VALIDATION

The experiment is not only the comparison between different models but also the selection of models themselves. Usually, the generalization ability of the neural network is more robust than that of the linear model. However, many parameters in the neural network still need to be adjusted and selected. The experiment needs to adjust these hyperparameters to make the model's generalization ability reach its best level. The model needs to be initially evaluated on the validation set to avoid data leakage. Once the best parameters are found, the experiment tests the model last on the test set and takes the error as the approximation of the generalization error. The final experimental parameters are shown in Table 4.

TABLE 4. Models parameters.

Model	Layer	Parameters
MLR	MLR	fit_intercept=True, normalize=False, copy_X=True, n_jobs=1
RFR	RFR	max_depth=400, max_features='sqrt', min_samples_leaf=1, min_impurity_split=2, n_estimators=100
SVR	SVR	kernel='rbf',C=1.0,epsilon=0.1
LSTM	LSTM	kernel_initializer='he_normal', activation='sigmoid',units=128
GRU	GRU	kernel_initializer='he_normal', activation='tanh', units=128
BD-LSTM	LSTM	kernel_initializer='he_normal', activation='relu', units=16
AGU	AGU	kernel_initializer='he_normal', activation='tanh', units=128
CNN-LSTM	Conv1D	filters=16, padding='valid', kernel_size=1, activation='tanh', padding='valid'
	MaxPooling1D	padding='valid', pool_size=1
	LSTM	kernel_initializer='he_normal', activation='tanh', units=64
CNN-GRU	Conv1D	filters=16, padding='valid', kernel_size=1, activation='tanh', padding='valid'
	MaxPooling1D	padding='valid', pool_size=1
	GRU	kernel_initializer='he_normal', activation='tanh', units=128
CNN-AGU	Conv1D	filters=16, padding='valid', kernel_size=1, activation='tanh', padding='valid'
	MaxPooling1D	padding='valid', pool_size=1,
	AGU	kernel_initializer='he_normal', activation='tanh', units=128

Through the preliminary evaluation of the validation set on the ten models, different prediction models have been adjusted to an appropriate and effective state under the same experimental environment, as shown in Table 5. In terms of model training time, compared with LSTM and GRU, AGU needs to match the current input sequence with the corresponding hidden state at the previous moment in the training process because of the inclusion of the attention

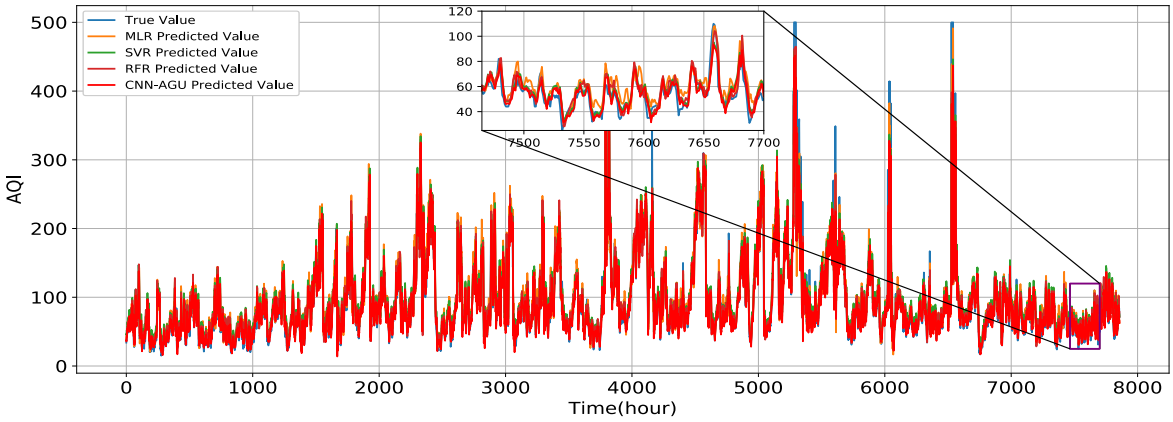


FIGURE 5. True Value and the Predicted Value of MLR, SVR, RFR, CNN-AGU.

TABLE 5. Validation set experimental results.

Model	MAE	R^2	MSE	Training time(s)
CNN-AGU	7.087	0.957	180.361	133.485
CNN-GRU	8.013	0.9453	185.4	111.366
AGU	8.242	0.949	194.231	127.135
CNN-LSTM	8.931	0.9466	195.826	121.791
GRU	9.003	0.9331	210.271	111.062
BD-LSTM	9.082	0.9327	214.542	167.577
LSTM	9.523	0.93	220.396	111.536
RFR	15.364	0.8905	402.629	48.632
MLR	18.753	0.8795	437.271	20.135
SVR	19.264	0.875	483.626	54.023

TABLE 6. Test set experimental results.

Model	MAE	R^2	MSE
CNN-AGU	8.107	0.949	183.322
CNN-GRU	8.429	0.9253	189.542
AGU	8.742	0.938	196.4631
GRU	9.199	0.9102	230.001
CNN-LSTM	9.9114	0.9211	199.3654
BD-LSTM	9.971	0.9204	231.869
LSTM	10.001	0.92	232.777
RFR	15.296	0.8727	400.4241
MLR	20.164	0.8631	440.166
SVR	21.1	0.855	523.136

mechanism in AGU. This matching process is used to calculate the weight of the matching degree, which is limited to the current sequence pair and is not the overall weight like the weight of the traditional neural network models. It increases slightly in training time, but AGU performs well in MAE, MSE, and R^2.

E. EXPERIMENTAL RESULTS

1) EXPERIMENTAL RESULTS ON AIR QUALITY DATA SET

To validate the AQI prediction model based on CNN-AGU, the prediction results of this model are compared with the other nine models using MAE, MSE, and R^2. The experimental results of each model are shown in Table 6. The results show that the evaluation indexes of the CNN-AGU prediction model are better than other models.

(1) The R^2 of traditional machine learning and statistical models' SVR, MLR and RFR are between 0.85 and 0.88, which are far lower than the fitting degree of neural network prediction models. Because SVR, MLR and RFR models have a poor ability to learn outliers, their MAE and MSE performance is poor. Compared with traditional statistical and machine learning methods, the structure of RNN is recursive according to the sequence direction, and the cyclic units are connected in a chain. This structure makes the output result become the result of the joint action of the input at that moment and all history and finally achieves the purpose of modeling time series. Therefore, this kind of model is more efficient in processing temporal tasks. Compared with the LSTM prediction model, the R^2 of SVR is 0.065 lower, MAE is 11.099 higher, and MSE is 290.359 higher. Predicted values of SVR, RFR, MLR and CNN-AGU versus true values are shown in Fig. 5.

(2) LSTM, GRU, and AGU are all variants of RNN in essence. GRU has the shortest training time in the experimental process due to its relatively simple internal structure, and the fitting effect is equivalent to LSTM. Compared with the GRU structure, the DAM is added to AGU to make the model more efficient in processing input data, and the addition of the attention mechanism makes the memory cells quickly discard unimportant content. Compared with LSTM, the R^2 of AGU increases by 0.018, and MAE and MSE decrease by 1.259 and 36.3139, respectively. Compared with GRU, R^2 of AGU increases by 0.0278, MAE and MSE decrease by 0.457 and 33.5379, respectively. Predicted values of LSTM, GRU, and AGU versus true values are shown in Fig. 6.

(3) CNN can process data efficiently and extract features automatically because of its shared convolution kernel. Combined with RNN, CNN further enhances the feature extraction ability of the combined model. When LSTM, GRU and AGU are combined with CNN, MAE decreases and R^2 increases. Compared with LSTM, the MAE of CNN-LSTM reduces by 0.9%, the MSE reduces by 14.35%, and R^2 rises by 0.12%.

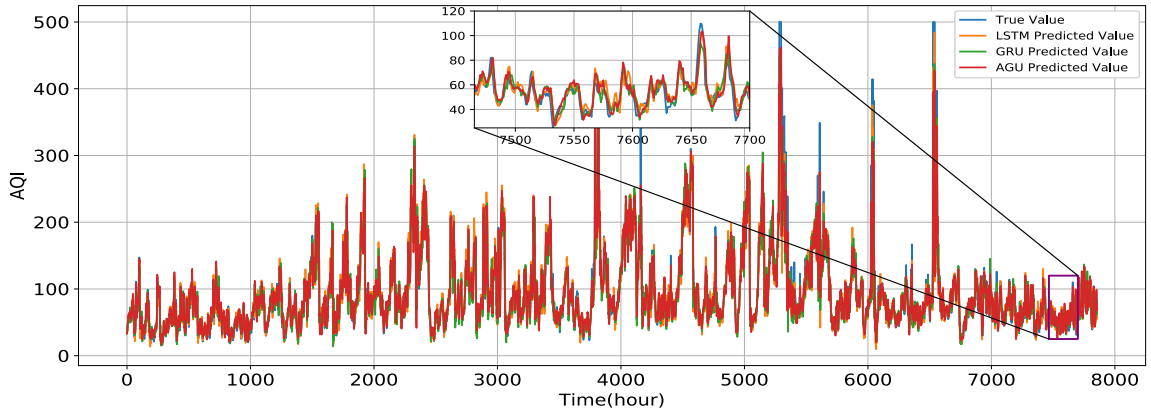


FIGURE 6. True Value and the Predicted Value of LSTM, GRU, AGU.

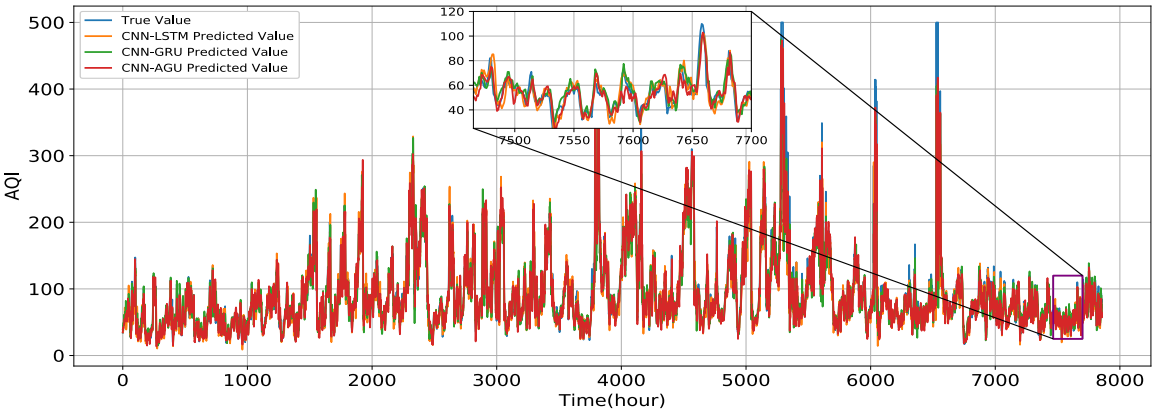


FIGURE 7. True Value and the Predicted Value of CNN-LSTM, CNN-GRU, CNN-AGU.

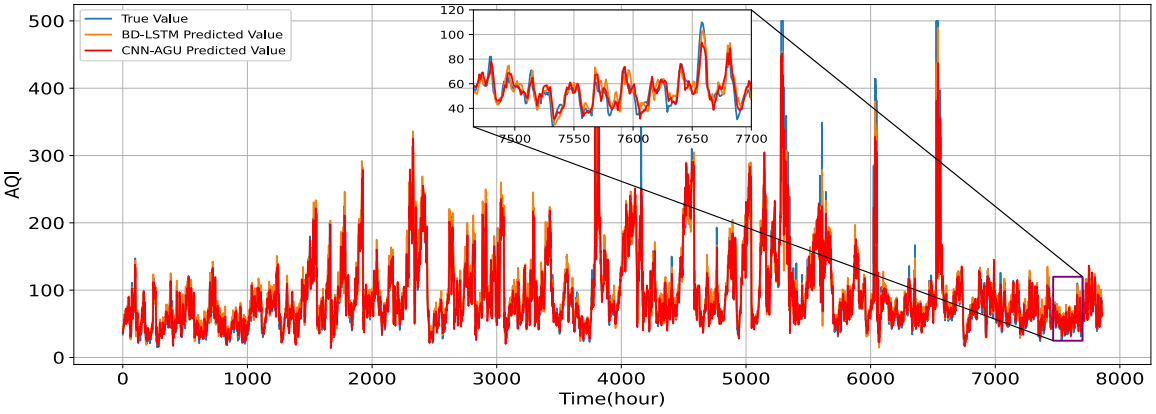


FIGURE 8. True Value and the Predicted Value of BD-LSTM, CNN-AGU.

Compared with GRU, the MAE of CNN-GRU reduces by 8.37%, the MSE reduces by 17.59%, and R^2 rises by 1.66%. Compared with AGU, the MAE of CNN-AGU reduces by 7.26%, the MSE reduces by 6.69%, and R^2 rises by 1.17%.

(4) Compared with CNN-LSTM, the MAE of CNN-AGU is 1.8044 lower than that of CNN-LSTM, the MSE of CNN-AGU decreases by 16.0434, and the R^2 of CNN-AGU increases by 0.0279. Compared with CNN-GRU, the MAE of CNN-AGU is 0.322 lower than that of CNN-GRU, the MSE of CNN-AGU decreases by 6.22, and the R^2 of CNN-AGU increases by 0.0237. The predicted values of the above three models versus the true values are shown in Fig. 7.

(5) BD-LSTM shows the best performance in processing sunspot time series tasks. Therefore, the experiment uses

BD-LSTM as a baseline model and adds this model to AQI time series tasks to verify the performance of the CNN-AGU model. BD-LSTM is a combination of forward LSTM and backward LSTM. It can capture the dependencies between long-distance data. However, for excessively long sequences, the initial information of the series is still not well transmitted. AGU adds the attention mechanism based on GRU structure, which does not rely on the order between data, but mines information by calculating the similarity between data. Hence, it reduces the loss of information. Moreover, the BD-LSTM has insufficient ability to extract features. Compared with the BD-LSTM prediction model, the R^2 of CNN-AGU is 0.0286 higher, MAE is 1.864 lower, and MSE is 48.547 lower. Predicted values of BD-LSTM and CNN-AGU versus true values are shown in Fig. 8.

2) EXPERIMENTAL RESULTS ON GOLD FUTURES DATA SET

To further verify the generalization ability of the CNN-AGU, the model is verified in predicting the gold futures price.

The gold futures data set used in this research comes from the API interface provided by Tushare, a third-party financial data website. There are 5827 pieces of data and nine input items in the gold futures price data set: futures settlement price, closing price, trading volume, futures holdings, AU99.99, USD_CNY, dollar index, Dow Jones Industrial Average, and NASDAQ index. The output of the model is the futures closing price.

The gold futures and air quality data sets are nonlinear time series data. The experimental results of the proposed model and baseline models on the gold futures data set are shown in Table 7.

TABLE 7. Experimental results on gold futures data set.

Model	MAE	R^2	MSE
CNN-AGU	56.4836	0.9766	5370.0341
CNN-GRU	58.1646	0.9648	5566.1358
CNN-LSTM	58.9654	0.9636	5762.1354
AGU	60.2155	0.9621	6394.2615
GRU	60.9957	0.9611	6813.1354
BD-LSTM	61.0516	0.9607	6674.2861
LSTM	63.1548	0.9615	7069.5465
SVR	73.3991	0.9544	8200.4646
RFR	76.1454	0.9569	8336.3542
MLR	89.5615	0.9231	9202.1342

In the prediction task of gold futures price with a small amount of data, CNN-AGU has the best MAE, MSE, and R^2 compared with the other nine models. It shows that the proposed model has strong adaptability in new samples.

V. DISCUSSION

Compared with nine other models for predicting the AQI task, the CNN-AGU model proposed in this paper is effective. In addition, this paper also verifies the CNN-AGU model on the gold futures data set, which is also time series data. These experimental results show that the model proposed in this paper is optimal compared with the baseline models on the air quality data set with a large amount of data and the gold futures data set with a small amount of data.

The fitting degree of LSTM, GRU, and AGU are higher than that of SVR, MLR, and RFR from the comparative experiments. The traditional machine learning and statistical models are not easy to learn the in-depth information in AQI data, so they cannot produce a high fitting degree.

Compared with LSTM and GRU, MAE, MSE and R^2 of the AGU model are the best. The introduction of CNN further enhances the model's prediction accuracy. For example, after the introduction of CNN, MAE and MSE of CNN-LSTM, CNN-AGU and CNN-GRU are reduced, and R^2 are correspondingly increased. Compared with other models, MAE, MSE and R^2 of CNN-AGU are the best. The reasons for CNN-AGU to achieve high-precision AQI prediction are as follows:

(1) AGU inherits the advantages of neural network and gated technology, and effectively learns historical data by using gated technology while reducing the probability of ''vanishing gradient'' and ''exploding gradient''. Therefore, this model has certain advantages in learning the nonlinear characteristics of time series data, such as air quality.

(2) AGU adds attention mechanism and DAM on the basis of GRU, which not only enhances the sensitivity of this kind of RNN to input data but also improves the rationality of parameter allocation learning greatly, which enables the model to pay attention to and remember important parts adaptively in the learning process.

(3) CNN can automatically mine deeply hidden features from the data. The introduction of CNN improves the model's data feature extraction and prediction accuracy.

VI. CONCLUSION

This paper proposes a new AQI prediction model based on CNN-AGU. In this model, the data features extracted by the CNN are the input data of the AGU. The AGU embeds the attention mechanism into the gated unit, which enhances the memory ability of the gated unit. At the same time, AGU also embeds the DAM, which makes the gated unit more sensitive to historical data learning. The comparative experiments show that the comprehensive evaluation of CNN-AGU is superior to the other nine models.

AQI prediction is an important research direction, and future research is mainly divided into the following two aspects:

(1) The experiment mainly uses the temporal attention mechanism to process time series data. However, in real life, factors affecting AQI include not only time series data (such as pollutants, meteorological data, and the traffic flow) but also spatial data (such as Point of Interest (POI) and road network). For example, the air quality near parks is much better than near factories in general, and the road network strongly correlates with transportation modes. For processing these spatial feature data, our future research work is to introduce the spatial attention mechanism into the prediction model.

(2) In the development of processing time series tasks, transformers have become popular in recent years. Compared with CNN, this model can pay more attention to local details [41]. The model will be more known and explored in future research work.

REFERENCES

[1] G. Kyrkilis, A. Chaloulakou, and P. A. Kassomenos, "Development of an aggregate air quality index for an urban Mediterranean agglomeration: Relation to potential health effects," *Environ. Int.*, vol. 33, no. 5, pp. 670–676, Jul. 2007, doi: 10.1016/j.envint.2007.01.010.

[2] P. Poursafa, M. Mansourian, M.-E. Motlagh, G. Ardalan, and R. Kelishadi, "Is air quality index associated with cardiometabolic risk factors in adolescents? The CASPIAN-III study," *Environ. Res.*, vol. 134, pp. 105–109, Oct. 2014, doi: 10.1016/j.envres.2014.07.010.

[3] X.-J. Wen, L. Balluz, and A. Mokdad, "Association between media alerts of air quality index and change of outdoor activity among adult asthma in six states, BRFSS, 2005," *J. Community Health*, vol. 34, no. 1, pp. 40–46, Sep. 2008, doi: 10.1007/s10900-008-9126-4.

[4] H. Y. Wang, J. Y. Wang, and X. H. Wang, "An AQI level forecasting model using chi-square test and BP neural network," in *Proc. 2nd Int. Conf. Intell. Inf. Process.*, Bangkok, Thailand, Jul. 2017, pp. 1–16, doi: 10.1145/3144789.3144817.

[5] U. Saini, R. Kumar, V. Jain, and M. U. Krishnajith, "Univariant time series forecasting of agriculture load by using LSTM and GRU RNNs," in *Proc. IEEE Students Conf. Eng. Syst. (SCES)*, Jul. 2020, pp. 1–6, doi: 10.1109/SCES50439.2020.9236695.

[6] A. Tokgöz and G. Ünal, "A RNN based time series approach for forecasting Turkish electricity load," in *Proc. 26th Signal Process. Commun. Appl. Conf. (SIU)*, Izmir, Turkey, May 2018, pp. 1–4, doi: 10.1109/SIU.2018.8404313.

[7] A. U. Muhammad, A. S. Yahaya, S. M. Kamal, J. M. Adam, W. I. Muhammad, and A. Elsafi, "A hybrid deep stacked LSTM and GRU for water price prediction," in *Proc. 2nd Int. Conf. Comput. Inf. Sci. (ICCIS)*, Sakaka, Saudi Arabia, Oct. 2020, pp. 1–6, doi: 10.1109/ICCIS49240.2020.9257651.

[8] F. Landi, L. Baraldi, M. Cornia, and R. Cucchiara, "Working memory connections for LSTM," *Neural Netw.*, vol. 144, pp. 334–341, Dec. 2021, doi: 10.1016/j.neunet.2021.08.030.

[9] Z. Yu, D. S. Moirangthem, and M. Lee, "Continuous timescale long-short term memory neural network for human intent understanding," *Frontiers Neurorobotics*, vol. 11, p. 42, Aug. 2017, doi: 10.3389/fnbot.2017.00042.

[10] K. Ohno and A. Kumagai, "Recurrent neural networks for learning long-term temporal dependencies with reanalysis of time scale representation," in *Proc. IEEE Int. Conf. Big Knowl. (ICBK)*, Auckland, New Zealand, Dec. 2021, pp. 182–189, doi: 10.1109/ICKG52313.2021.00033.

[11] Z. Cheng, Y. Xu, M. Cheng, Y. Qiao, S. Pu, Y. Niu, and F. Wu, "Refined gate: A simple and effective gating mechanism for recurrent units," 2020, *arXiv:2002.11338*.

[12] X. Y. Li and S. Q. Jiang, "Bundled object context for referring expressions," *IEEE Trans. Multimedia*, vol. 20, no. 10, pp. 2746–2760, Oct. 2018, doi: 10.1109/TMM.2018.2811621.

[13] S. A. Abdul-Wahab, C. S. Bakheit, and S. M. Al-Alawi, "Principal component and multiple regression analysis in modelling of ground-level ozone and factors affecting its concentrations," *Environ. Model. Softw.*, vol. 20, no. 10, pp. 1263–1271, Oct. 2005, doi: 10.1016/j.envsoft.2004.09.001.

[14] G. A. Grell, S. E. Peckham, R. Schmitz, S. A. McKeen, G. Frost, W. C. Skamarock, and B. Eder, "Fully coupled 'online' chemistry within the WRF model," *Atmos. Environ.*, vol. 39, no. 37, pp. 6957–6975, Dec. 2005, doi: 10.1016/j.atmosenv.2005.04.027.

[15] D. W. Byun and K. L. Schere, "Review of the governing equations, computational algorithms and other components of the models-3 community multiscale air quality (CMAQ) modeling system," *Appl. Mech. Rev.*, vol. 59, no. 2, pp. 51–78, Mar. 2006, doi: 10.1115/1.2128636.

[16] H. Sun, D. Gui, B. Yan, Y. Liu, W. Liao, Y. Zhu, C. Lu, and N. Zhao, "Assessing the potential of random forest method for estimating solar radiation using air pollution index," *Energy Convers. Manage.*, vol. 119, pp. 121–129, Jul. 2016, doi: 10.1016/j.enconman.2016.04.051.

[17] S. Silva, F. M. Clemente, M. Castelli, A. Popovi, and L. Vanneschi, "A machine learning approach to predict air quality in California," *Complexity*, vol. 2020, no. 332, pp. 1–32, Aug. 2020, doi: 10.1155/2020/8049504.

[18] M. H. Kim, Y. S. Kim, J. Lim, J. T. Kim, S. W. Sung, and C. Yoo, "Data-driven prediction model of indoor air quality in an underground space," *Korean J. Chem. Eng.*, vol. 27, no. 6, pp. 1675–1680, Nov. 2010, doi: 10.1007/s11814-010-0313-5.

[19] Q. Yin, R. Zhang, and X. Shao, "CNN and RNN mixed model for image classification," in *Proc. MATEC Web Conf.*, vol. 277, 2019, p. 02001, doi: 10.1051/matecconf/201927702001.

[20] H. He and F. Luo, "Study of LSTM air quality index prediction based on forecasting timeliness," *IOP Conf. Earth Environ. Sci.*, vol. 446, no. 3, 2020, Art. no. 032113, doi: 10.1088/1755-1315/446/3/032113.

[21] T. Pu, F. Lin, Y. Zhao, and Z. Fu, "Air quality prediction method based on improved wavelet denoising and LSTM," *J. Phys., Conf.*, vol. 1748, no. 3, Jan. 2021, Art. no. 032033, doi: 10.1088/1742-6596/1748/3/032033.

[22] R. Chandra, S. Goyal, and R. Gupta, "Evaluation of deep learning models for multi-step ahead time series prediction," *IEEE Access*, vol. 9, pp. 83105–83123, 2021, doi: 10.1109/ACCESS.2021.3085085.

[23] X. Zhou, J. Xu, P. Zeng, and X. Meng, "Air pollutant concentration prediction based on GRU method," *J. Phys., Conf.*, vol. 1168, Feb. 2019, Art. no. 032058, doi: 10.1088/1742-6596/1168/3/032058.

[24] J. Becerra-Rico, M. A. Aceves-Fernández, K. Esquivel-Escalante, and J. C. Pedraza-Ortega, "Airborne particle pollution predictive model using gated recurrent unit (GRU) deep neural networks," *Earth Sci. Informat.*, vol. 13, no. 3, pp. 821–834, May 2020, doi: 10.1007/s12145-020-00462-9.

[25] H. Xie, L. Ji, Q. Wang, and Z. Jia, "Research of PM2.5 prediction system based on CNNs-GRU in Wuxi urban area," *IOP Conf. Earth Environ. Sci.*, vol. 300, no. 3, Jul. 2019, Art. no. 032073, doi: 10.1088/1755-1315/300/3/032073.

[26] P. Mei, M. Li, Q. Zhang, G. Li, and L. Song, "Prediction model of drinking water source quality with potential industrial-agricultural pollution based on CNN-GRU-attention," *J. Hydrol.*, vol. 610, Jul. 2022, Art. no. 127934, doi: 10.1016/j.jhydrol.2022.127934.

[27] Y. Sun and Q. Liu, "Attribute recognition from clothing using a faster R-CNN based multitask network," *Int. J. Wavelets, Multiresolution Inf. Process.*, vol. 16, no. 2, Mar. 2018, Art. no. 1840009, doi: 10.1142/S021969131840009X.

[28] C. Guo, H. Wang, T. Jian, Y. He, and X. Zhang, "Radar target recognition based on feature Pyramid fusion lightweight CNN," *IEEE Access*, vol. 7, pp. 51140–51149, 2019, doi: 10.1109/ACCESS.2019.2909348.

[29] Z. Wang, W. Zheng, C. Song, Z. Zhang, J. Lian, S. Yue, and S. Ji, "Air quality measurement based on double-channel convolutional neural network ensemble learning," *IEEE Access*, vol. 7, pp. 145067–145081, 2019, doi: 10.1109/ACCESS.2019.2945805.

[30] Z. Wang, S. Yue, and C. Song, "Video-based air quality measurement with dual-channel 3-D convolutional network," *IEEE Internet Things J.*, vol. 8, no. 18, pp. 14372–14384, Sep. 2021, doi: 10.1109/JIOT.2021.3068375.

[31] C. X. Liu, K. L. Li, J. Liu, and C. Chen, "LHCnn: A novel efficient multivariate time series prediction framework utilizing convolutional neural networks," in *Proc. 21st Int. Conf. High Perform. Comput. Commun., IEEE 17th Int. Conf. Smart City, IEEE 5th Int. Conf. Data Sci. Syst. (HPCC/SmartCity/DSS)*, Zhangjiajie, China, Aug. 2019, pp. 2324–2332, doi: 10.1109/HPCC/SmartCity/DSS.2019.00323.

[32] B. Xin and W. Peng, "Prediction for chaotic time series-based AE-CNN and transfer learning," *Complexity*, vol. 2020, pp. 1–9, Sep. 2020, doi: 10.1155/2020/2680480.

[33] Y. Qiao, Y. Wang, C. Ma, and J. Yang, "Short-term traffic flow prediction based on 1DCNN-LSTM neural network structure," *Modern Phys. Lett. B*, vol. 35, no. 2, Oct. 2020, Art. no. 2150042, doi: 10.1142/S0217984921500421.

[34] D. T. Tran, A. Iosifidis, J. Kanniainen, and M. Gabbouj, "Temporal attention-augmented bilinear network for financial time-series data analysis," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 30, no. 5, pp. 1407–1418, May 2019, doi: 10.1109/TNNLS.2018.2869225.

[35] C. H. Zhang, F. S. Li, L. M. Chao, and J. S. Guo, "Correlations between air quality status and meteorological factors in Hohhot city," *J. Arid Land Resour. Environ.*, vol. 32, no. 2, pp. 87–93, Feb. 2018, doi: 10.13448/j.cnki.jalre.2018.053.

[36] B. V. Bhaskar and V. M. Mehta, "Atmospheric particulate pollutants and their relationship with meteorology in Ahmedabad," *Aerosol Air Quality Res.*, vol. 10, no. 4, pp. 301–315, 2010, doi: 10.4209/aaqr.2009.10.0069.

[37] S. Lim, S. J. Kim, Y. Park, and N. Kwon, "A deep learning-based time series model with missing value handling techniques to predict various types of liquid cargo traffic," *Exp. Syst. Appl.*, vol. 184, Dec. 2021, Art. no. 115532, doi: 10.1016/j.eswa.2021.115532.

[38] J. Wang, J. Li, X. Wang, J. Wang, and M. Huang, "Air quality prediction using CT-LSTM," *Neural Comput. Appl.*, vol. 33, no. 10, pp. 4779–4792, Nov. 2020, doi: 10.1007/s00521-020-05535-w.
[39] L. Munkhdalai, T. Munkhdalai, K. H. Park, H. G. Lee, M. Li, and K. H. Ryu, "Mixture of activation functions with extended min-max normalization for forex market prediction," *IEEE Access*, vol. 7, pp. 183680–183691, 2019, doi: 10.1109/ACCESS.2019.2959789.
[40] I. Goodfellow, Y. Bengio, and A. Courville, "Machine learning basics," in *Deep Learning*. Cambridge, MA, USA: MIT Press, 2016, pp. 118–120. [Online]. Available: https://www.deeplearningbook.org/
[41] Z. Wang, Y. Yang, and S. Yue, "Air quality classification and measurement based on double output vision transformer," *IEEE Internet Things J.*, vol. 9, no. 21, pp. 20975–20984, Nov. 2022, doi: 10.1109/JIOT.2022.3176126.

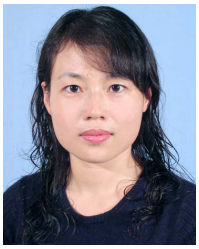
JINGYANG WANG received the B.Eng. degree in computer software from Lanzhou University, China, in 1995, and the M.Sc. degree in software engineering from the Beijing University of Technology, China, in 2007. He is currently a Professor with the School of Information Science and Engineering, Hebei University of Science and Technology, Shijiazhuang, China. His research interests include machine learning, natural language processing, big data processing, and distributed computing.

LUKAI JIN is currently pursuing the master's degree with the Hebei University of Science and Technology. Her research interests include machine learning and deep learning.

XIAOLEI LI is currently pursuing the master's degree with the Hebei University of Science and Technology. His research interests include machine learning and deep learning.

SIYUAN HE is currently pursuing the bachelor's degree with the Hebei University of Science and Technology. Her research interests include machine learning and deep learning.

MIN HUANG is currently an Associate Professor with the School of Information Science and Engineering, Hebei University of Science and Technology. His research interests include machine learning, data processing and security, and artificial intelligence.



HAIYAO WANG received the B.Eng. degree in mechanical design and manufacturing, and the M.Sc. degree in industry engineering from the Hefei University of Technology, China, in 1998 and 2009, respectively. She is currently an Associate Professor with the School of Ocean Mechatronics, Xiamen Ocean Vocational College, Xiamen, China. Her research interests include machine learning, process optimization, and efficiency improvement.