Journal of Information Hiding and Multimedia Signal Processing ©2025 ISSN 2073-4212
Ubiquitous International Volume **16**, Number **3**, September **2025**

PM2.5 Pollution Prediction Method Based on Spatiotemporal Fusion Deep Learning

Qiang Wang*

College of Computing and Information Technologies, National University
Manila 1008, Philippines
College of Big Data and Artificial Intelligence, Anhui Xinhua University
Hefei 230088, P. R. China
wangqiang@axhu.edu.cn

Vladimir Y. Mariano

College of Computing and Information Technologies, National University
Manila 1008, Philippines
vymariano@national-u.edu.ph

*Corresponding author: Qiang Wang
Received May 3, 2025, revised June 3, 2025, accepted June 4, 2025.

Abstract. *PM2.5 prediction is a typical time-series problem. Compared with the early application of single models, current research trends have increasingly focused on multi-model fusion strategies. Although complex network models have demonstrated promising performance in air pollution concentration prediction, prediction accuracy still has room for improvement due to limitations such as insufficient feature dimensions and inadequate mining of spatiotemporal correlations. This paper proposes a CNN-ECA-LSTM hybrid model that integrates spatiotemporal features and attention mechanisms. Conv1D is used to extract local spatiotemporal correlations from multi-source data, LSTM captures long-term dependencies of time series, ECA strengthens key feature weights, and achieves effective integration of spatiotemporal features and suppression of redundant information. Experimental research shows that the proposed model is significantly better than LSTM single model and CNN-LSTM hybrid model in predicting PM2.5 concentration, validating the synergistic enhancement of prediction accuracy by spatiotemporal fusion and attention mechanisms. This research not only provides an innovative solution for small-region pollution prediction but also offers valuable references for applying deep learning hybrid models to explore time-series big data in fields such as meteorological data, stock markets, economic analysis, and transportation.*

Keywords: PM2.5 Prediction; Spatiotemporal Fusion; LSTM; CNN-LSTM; CNN-ECA-LSTM

1. **Introduction.** The rapid development of modern technology, industry, and transportation has accelerated the global urbanization process. Concurrently, the sustained emission of various air pollutants has exacerbated environmental degradation, triggering global environmental issues such as the greenhouse effect, ozone layer depletion, and photochemical smog. Among numerous air pollutants, fine particulate matter PM2.5—particles with a diameter ≤ 2.5 micrometers—has become a critical risk factor for human health due to its ability to penetrate the human respiratory system and even enter the bloodstream [1], attracting widespread international attention. The Global Burden of Disease (GBD) study further confirms that PM2.5 has become the primary environmental risk

factor, with each 10 $\mu g/m^3$ increase in concentration associated with a roughly 7% higher all-cause mortality rate (with regional variations in this proportion) [2].

PM2.5 pollution not only threatens public health but also exerts far-reaching impacts on the global economy and policy-making. Economically, smoggy weather—with PM2.5 as the primary pollutant—significantly reduces transportation efficiency, leading to issues such as flight delays, railway suspensions, and highway closures [3]. It also compels industrial enterprises to undergo structural adjustments, increasing costs for pollution control and operational transformation. Additionally, chain reactions such as rising residential medical expenses, declining urban commercial attractiveness, and reduced foreign investment further exacerbate economic losses. In terms of policy, countries have implemented targeted measures: the United States established a strict pollution control system through the Clean Air Act; France promulgated the Building Energy Efficiency Regulations to promote low-carbon transitions.

PM2.5 comes from a wide range of sources, such as sandstorms, volcanic ash, particulate matter from forest fires, and sea salt crystals in nature; however, anthropogenic emissions play a dominant role, encompassing fossil fuel (coal, petroleum products) and biomass (crop straw, wood) combustion, industrial dust emissions, road dust, vehicle exhaust, and kitchen fumes [4]. According to WHO 2023 data, the global annual average PM2.5 concentration stands at 32.8 $\mu g/m^3$, with over 90% of days exceeding recommended limits in East and South Asia. China's current Ambient Air Quality Standards set an annual limit of 35 $\mu g/m^3$ for PM2.5, which aligns only with the WHO's first-stage transitional target; although revisions are moving toward the second-stage target of 25 $\mu g/m^3$, there remains a substantial gap from the WHO's ultimate standard of 5 $\mu g/m^3$. The 2023 China Ecological and Environmental Status Bulletin reveals that among 339 monitored cities nationwide, 136 (40.1%) failed to meet air quality standards, of which 105 cities (77.2% of Excess Cities) had excessive PM2.5 concentrations. These findings highlight that PM2.5 pollution remains a critical obstacle to improving urban air quality in China. Accurately predicting the trend of PM2.5 concentration changes has irreplaceable strategic significance for formulating scientific prevention and control strategies, safeguarding public health, and promoting sustainable urban development. By establishing precise prediction models, policymakers can obtain data-driven support for pollution early warning, emission reduction planning, and policy optimization, thereby facilitating the coordinated development of air quality and socioeconomic goals.

2. **Literature Review.** Current research on PM2.5 concentration prediction methods primarily focuses on three directions: time series analysis, machine learning, and deep learning. The application of time series analysis models for PM2.5 prediction can be traced back to the last century. Given that PM2.5 concentration data are fundamentally time-series data, traditional time series models often yield favorable results in predicting PM2.5 concentrations. For example, Suresh S et al. [5] developed three single models—ARIMA, LSTM, and Prophet—for predicting PM2.5 concentrations in Kodungaiyur, India. The results showed that ARIMA model has the best performance. A similar trend was observed in the study by Wongrin W et al. [6], which compared traditional statistical methods (ETS, ARIMA, DLM) with deep learning models (RNN, LSTM) for predicting PM2.5 in northern Thailand. Traditional statistical methods, particularly the ARIMA model, outperformed deep learning models with lower RMSE values. However, single time series prediction models often appear insufficient in comprehensive research, prompting scholars to explore integrations with other modeling approaches.

Machine learning algorithms exhibit significant advantages in solving general classification and regression problems, ensemble learning algorithms (such as RF, XGBOOST,

etc.) demonstrate superior performance in handling multidimensional data compared to traditional time-series models. Minh V T T et al. [7] used meteorological data from the WRF model to conduct short-term predictions of PM2.5 concentrations within 48 hours, concluding that among six single-algorithm models developed, the Extremely Randomized Trees Regressor (ETR) performed best with 74% accuracy. Ma X et al. [8] compared four machine learning methods (Linear SVR, K-Nearest Neighbors, Lasso, Gradient Boosting) for prediction, showing that Gradient Boosting significantly outperformed the other three, while the Linear SVR model yielded suboptimal results. Ji Haiyang [9] established four machine learning models—XGBOOST, RF, GBDT, and Adaboost—and found that GBDT achieved the best performance. Murugan et al. [10] used the Malaysian air pollution dataset to predict PM2.5 concentrations using two algorithms—Multilayer Perceptron (MLP) and Random Forest (RF)—and conducted a comprehensive accuracy comparison, the results indicate that random forests perform better. Van et al. [11] employed Pearson correlation coefficients for feature engineering on an AQI dataset to screen variables, then compared Decision Tree, Random Forest, and XGBoost algorithms. Their findings confirmed that XGBoost demonstrated superior prediction performance for AQI values.

With the continuous development of deep learning algorithms, an increasing number of scholars have turned to these powerful training frameworks. Unlike shallow learning, deep learning models feature multiple hidden layers, enabling diverse functionalities through different network architectures. Compared to traditional machine learning, deep learning's recurrent network models (e.g., LSTM, GRU) can explicitly model temporal dependencies in sequential data, making them particularly effective for time-series tasks like PM2.5 prediction. While single neural network models once outperformed traditional statistical and machine learning methods, their prediction accuracy is no longer sufficient for practical needs. Conversely, integrating multiple deep learning models has shown promise in improving performance. Hu J et al. [12] integrated BiLSTM (Bidirectional LSTM), CNN, and GRU into a hybrid model, which exhibited the lowest RMSE compared to individual models, indicating strong potential for pollutant concentration forecasting. Similar to machine learning, optimizing deep learning models with appropriate algorithms can significantly enhance performance. Notable examples include Zhang Y et al. [13] proposed SSA-LSTM model to tune traditional LSTM parameters, outperforming LSTM, BPNN, and BiLSTM in empirical tests. Wang Qianying and Yang Kexin [14] used a grid search-optimized SVR model to correct LSTM prediction errors, successfully improving forecasting accuracy. Xu Yixin [15] and Zhang Yanan et al. [16] applied wavelet analysis for data preprocessing before building BPNN models, both achieving satisfactory results. Peng Yuqing [17] introduced attention mechanisms into LSTM and GRU models, demonstrating substantial performance gains. This trend of developing PM2.5 prediction models that integrate spatiotemporal features through deep learning frameworks represents a cutting-edge research direction.

3. **Methodology.**

PM2.5 concentration variations are influenced by complex interactions among pollutant source emissions, geographical environment, and meteorological conditions, exhibiting significant nonlinear and dynamic characteristics. These features make it difficult for traditional time-series models to effectively capture changing patterns, limiting their prediction accuracy. With the advancement of deep learning techniques, related algorithms have gradually become the mainstream approach in PM2.5 concentration prediction due to their strong feature extraction and nonlinear fitting capabilities. Contrasting with early single-model applications, current research trends increasingly focus on multi-model fusion strategies or enhancing single-model performance through methods such as parameter optimization and feature engineering.

Although complex network models have demonstrated promising application results in the field of air pollution concentration prediction, there is still significant room for improvement in the prediction accuracy of PM2.5 concentration due to limitations such as insufficient feature dimensions and inadequate exploration of spatiotemporal correlations. In response to this situation, this study fully considers the coupling effects of air pollutants and meteorological factors within the region, deeply integrates the spatiotemporal evolution laws of pollutant diffusion, and constructs a PM2.5 concentration prediction model that fuses spatiotemporal correlations, as illustrated in Figure 1.

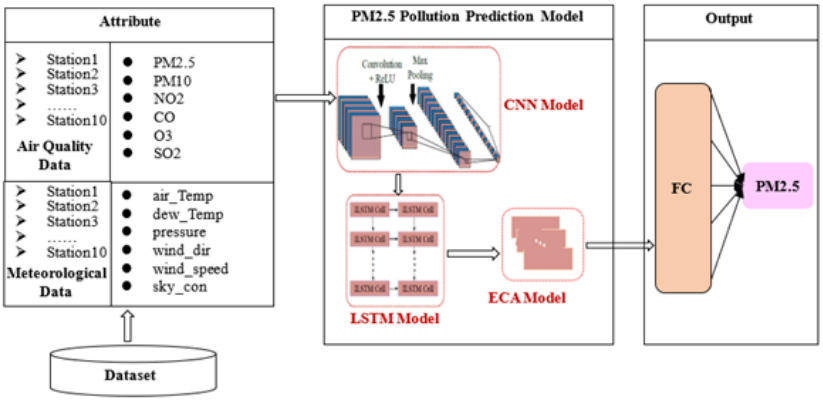

FIGURE 1. PM2.5 Concentration Prediction Model Integrating Spatiotemporal Correlations

3.1. **CNN Feature Processing.** CNN has been widely applied in areas such as computer vision and speech recognition, thanks to its outstanding performance in processing and analyzing data from images, audio, and video [18]. Essentially, CNN utilizes two core mechanisms—local connectivity and weight sharing—it significantly reduces the number of model parameters, effectively enhancing training efficiency and generalization ability. Its network structure is based on convolutional layers, pooling layers, and fully connected layers [19], and achieves classification, recognition, and segmentation tasks of complex data through multi-level feature extraction and transformation.

Conv1D is a variant of CNN. Especially suitable for processing one-dimensional temporal data, such as time series signals, text sequences, etc. Its workflow is shown in Figure 2.

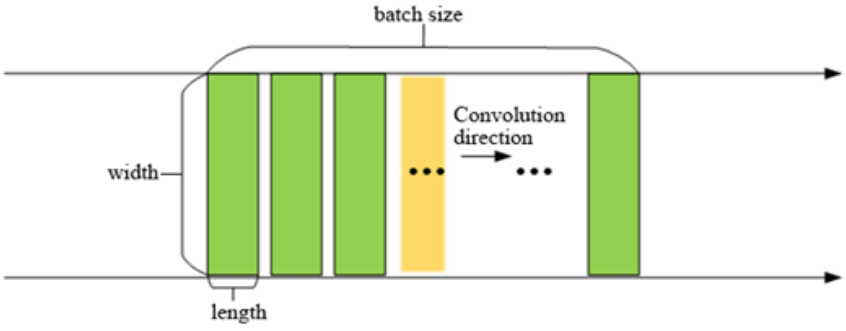

FIGURE 2. Schematic Diagram of One-Dimensional Convolution (Conv1D) Processing

This model employs a Conv1D structure to construct a CNN module for feature extraction. Considering the cross-influence of air pollutant characteristics among adjacent monitoring stations, during the model construction process, multi-station data are expanded in the single-station dimension to form a two-dimensional data structure that integrates information from multiple stations. The structural principle is illustrated in Figure 3.

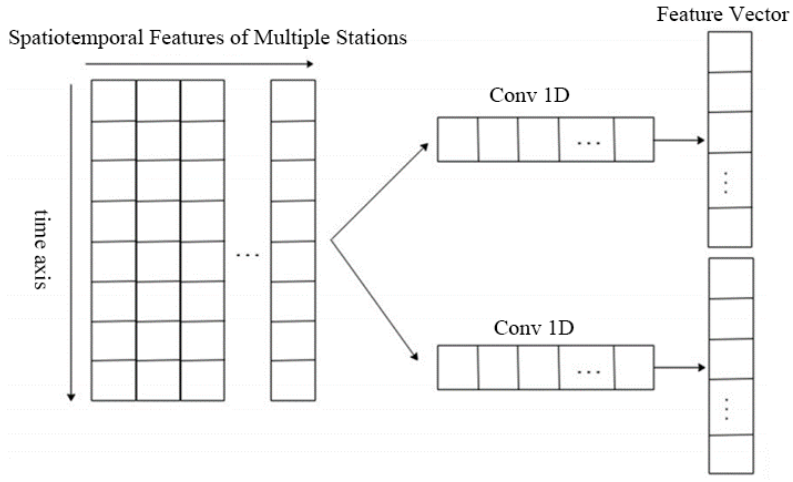

FIGURE 3. Structure Diagram of Conv1D

The calculation formulas for the Conv1D network layer are shown in Formula (1)-(2).

$$c = f(x \otimes W_c + b_c) \tag{1}$$

$$p = \max pooling(c) \tag{2}$$

Here, c represents the output of the convolutional layer, $f(x)$ is the activation function, and ReLU is chosen. $\otimes$ denotes the convolution operation, W_c is the weight matrix, b_c is the bias term, p is the output of the pooling layer, and $maxpooling(x)$ is the max-pooling function [20].

3.2. **LSTM for Capturing Long-Term Dependencies.** As an improved architecture of the RNN, LSTM network effectively overcomes the vanishing gradient problem of traditional RNNs when processing long sequential data, thanks to its unique gating mechanisms (input gate, forget gate, and output gate) and cell state design. The LSTM network unit consists of five core components: Cell State, Hidden State, Input Gate, Forget Gate, and Output Gate. Its neural network architecture is illustrated in Figure 4.

The calculation process of the LSTM network layer is shown in Formula (3)-(8).

$$f_t = \sigma(W_f \cdot [h_{t-1}, x_t] + b_f) \tag{3}$$

$$i_t = \sigma(W_i \cdot [h_{t-1}, x_t] + b_i) \tag{4}$$

$$\tilde{C}_t = \tanh(W_c \cdot [h_{t-1}, x_t] + b_c) \tag{5}$$

$$C_t = f_t \cdot C_{t-1} + i_t \cdot \tilde{C}_t \tag{6}$$

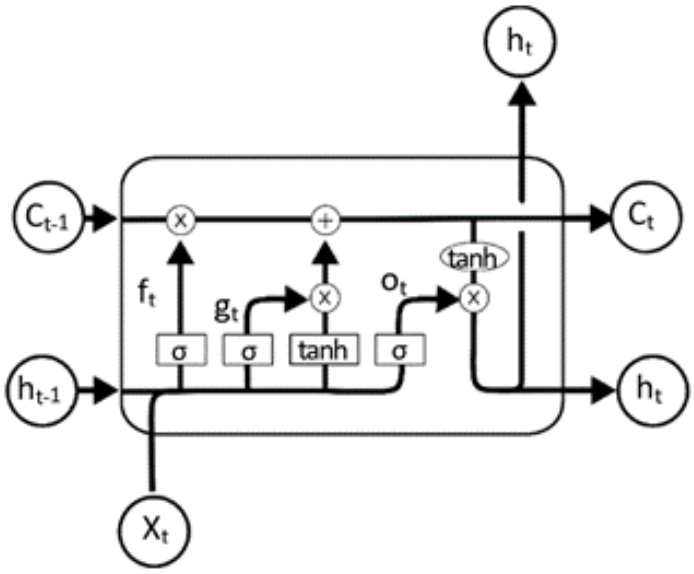

FIGURE 4. Structure Diagram of the LSTM Neural Network

$$o_t = \sigma(W_o \cdot [h_{t-1}, x_t] + b_o) \tag{7}$$

$$h_t = o_t \cdot \tanh(C_t) \tag{8}$$

Here, σ represents the Sigmoid activation function, and tanh represents the hyperbolic tangent activation function. W and b denote the weight matrix and bias vector, respectively. Through these iterative calculations, the LSTM network achieves deep feature extraction and effective memory of time-series data [21].

3.3. **ECA Focuses on Key Features.** Although the LSTM layer has a powerful ability to model long-term dependencies, in practical applications, its model performance is often significantly constrained by other components. During the PM2.5 prediction process, if there are information deficiencies or representational biases in the local features extracted by the Conv1D layer, even with the excellent temporal modeling capabilities of the LSTM layer, it is difficult to learn accurate long-term dependencies. In addition, the influencing factors of PM2.5 concentration are dynamically changing. For example, under different weather conditions, the influence weights of environmental features such as wind speed and humidity on PM2.5 diffusion vary significantly.

To address the above-mentioned issues, introducing an attention mechanism into the model architecture can effectively improve the prediction accuracy. By dynamically adjusting the feature weights, the attention mechanism enables the model to automatically focus on the key influencing factors of the current prediction task and enhance the representation ability of important features. Among them, the Efficient Channel Attention (ECA) mechanism demonstrates significant advantages with its unique design, and its structural schematic diagram is shown in Figure 5.

The ECA mechanism uses one-dimensional convolution to directly capture the dependencies between channels, innovatively eliminating the complex dimensionality reduction and increase processes in traditional attention mechanisms. This mechanism adaptively selects the size of the one-dimensional convolution kernel to achieve local cross-channel interactions without dimensionality reduction. It can accurately learn the importance

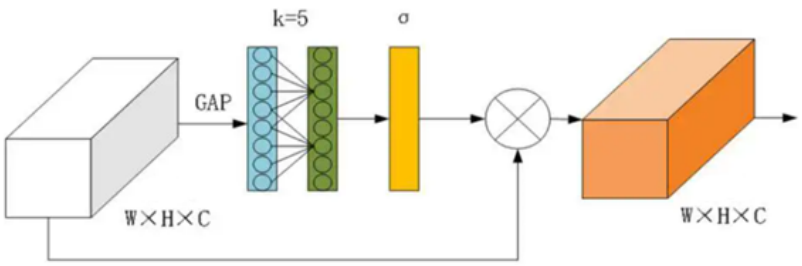

FIGURE 5. Structure Diagram of ECA (Efficient Channel Attention)

weights of each channel and dynamically adjust the feature response intensity accordingly. This design can reduce the number of parameters, lower computational complexity, and enhance the ability to focus on key features while maintaining model performance [22].

Specifically, the ECA module first adaptively calculates the kernel size k of the one-dimensional convolution based on the number of channels, as illustrated in Formula (9).

$$k = \lfloor C/\gamma + b \rfloor \tag{9}$$

Here, C denotes the number of channels in the input features, γ and b are hyperparameters. Taking the absolute value and rounding down to the nearest odd number ensures that the kernel size is an odd number. After obtaining the kernel size k, the ECA module applies one-dimensional convolution to the input features to learn the importance of each channel relative to the others. The calculation formula is shown in Formula (10).

$$y = Conv1D_k(x) \tag{10}$$

Here, x is the input features, and y is the output features.

4. Experiments and Results Analysis.

4.1. Data Sources and Correlation Analysis of the Experiments.

The air pollutant concentration data used in this experiment cover indicators such as PM2.5, PM10, SO_2, NO_2, O_3, and CO. All these data are sourced from the National Urban Air Quality Real-time Release Platform of the China National Environmental Monitoring Centre. The meteorological data (Air Temperature, Dew Point Temperature, Pressure, Wind Direction, Wind Speed, Sky Conditions) come from the public dataset of the National Centers for Environmental Information (NCDC). The study period is from 2019/1/1 to 2022/12/31, and the data are hourly records from 10 monitoring stations in Hefei urban area, Anhui Province, China [21]. Among them, the GaoXinQu monitoring station is selected as the target prediction station, and the other 9 adjacent stations are determined by the fixed adjacent radius method. The search radius for PM2.5 is set to 30 km, specifically including MingZhuGuangChang, SanLiJie, HuPoShanZhuan, DongPuShuiKu, ChangJiangZhongRoad, LuYangQu, YaoHaiQu, BaoHeQu, and BinHuXinQu. The detailed distribution of each monitoring station is shown in Table 1.

Previous studies have shown that PM2.5 concentration is not only closely related to the concentrations of air pollutants such as PM10 and CO but is also significantly influenced by meteorological factors such as wind speed and humidity, as well as geographical location. Considering that the formation of PM2.5 concentration is the result of the combined action of multiple factors, in order to apply experimental data to the prediction model

Table 1. Details of Monitoring Stations in Hefei City

STATION CODE	STATION NAME	CTRY	LONGITUDE	LATITUDE	SORT
1279	GaoXinQu	HeFei	117.1318	31.8403	Target station
1270	MingZhuGuangChang	HeFei	117.1959	31.7848	Adjacent station
1271	SanLiJie	HeFei	117.3072	31.8764	Adjacent station
1272	HuPoShanZhuan	HeFei	117.2588	31.8707	Adjacent station
1273	DongPuShuiKu	HeFei	117.1604	31.9052	Adjacent station
1274	ChangJiangZhongRoad	HeFei	117.2500	31.8571	Adjacent station
1275	LuYangQu	HeFei	117.2660	31.9436	Adjacent station
1276	YaoHaiQu	HeFei	117.3360	31.8586	Adjacent station
1277	BaoHeQu	HeFei	117.3027	31.7964	Adjacent station
1278	BinHuXinQu	HeFei	117.2776	31.7385	Adjacent station

more scientifically, need to explore the correlation between PM2.5 and various influencing factors.

This paper conducts a correlation study on six air pollutant data, namely PM2.5, PM10, SO_2, NO_2, O_3, and CO, as well as six meteorological data, including Air Temperature, Dew Point Temperature, Pressure, Wind Direction, Wind Speed, and Sky Conditions, using the Pearson correlation coefficient analysis method. The Pearson correlation coefficient is commonly used before regression analysis. It can quantify the linear relationship between two quantitative variables, including determining whether there is an association between variables, the direction of the association, and the degree of correlation [23]. The formula is (11).

$$\rho_{x,y} = \frac{cov(X,Y)}{\sigma_x \sigma_y} = \frac{E((X-\mu_x)(Y-\mu_y))}{\sigma_x \sigma_y} \tag{11}$$

Here, cov represents the covariance, σ_x and σ_y are the standard deviations, and E is the mathematical expectation.

Figure 6 analyzes the correlation between PM2.5 and various air pollutants. The coefficients between PM2.5 and PM10, NO_2, CO, and SO_2 are greater than 0, indicating a positive correlation. This means that the PM2.5 concentration increases with the increase of other air pollutants. Among them, the correlation coefficients of PM2.5 with CO, PM10, and NO_2, which rank in the top three, reach 0.67, 0.64, and 0.48 respectively, indicating a strong correlation. However, the correlation coefficient between PM2.5 and O_3 is -0.26, showing a negative correlation, which means that the PM2.5 concentration tends to decrease as the O_3 concentration increases.

This paper also includes meteorological data in the scope of research. The correlation analysis mainly focuses on the relationships between PM2.5 and Air Temperature, Dew Point Temperature, Pressure, Wind Direction, Wind Speed, and Sky Conditions, as illustrated in Figure 7. According to the data, there is also a certain degree of correlation between PM2.5 and meteorological data. However, the absolute values of the correlation coefficients do not exceed 0.4, indicating that the overall correlation between PM2.5 and meteorological data is relatively low. From the analysis results, the correlations between PM2.5 and Pressure, Sky Conditions are positive, which means that an increase in Pressure and Sky Conditions will lead to an increase in PM2.5 concentration. On the contrary, the correlations between PM2.5 and Air Temperature, Dew Point Temperature, Wind Speed are negative, indicating that as Air Temperature, Dew Point Temperature, and Wind Speed increase, the PM2.5 concentration will decrease. Therefore, this paper incorporates the meteorological data of the target station into the training of the prediction model.

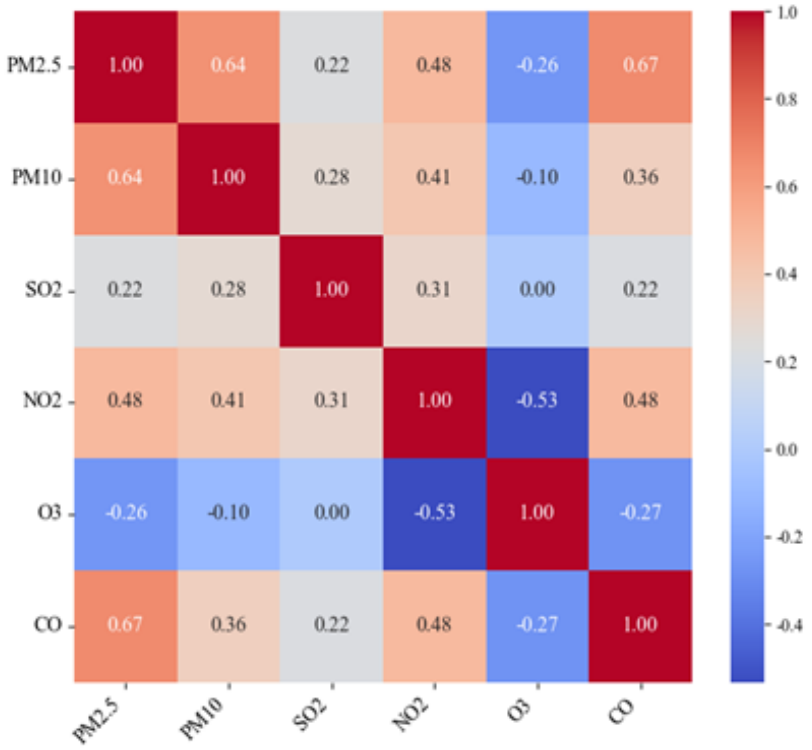

FIGURE 6. Correlation Coefficient Matrix Diagram of PM2.5 and Other Air Pollutants

4.2. **Experimental Data Preprocessing.** The air pollutant and meteorological data from the 10 monitoring stations in the experimental study are updated in real-time with an hourly time granularity. Based on the results of Pearson correlation coefficient analysis, to comprehensively capture the combined influence of multiple factors on the variation of PM2.5, the air pollutant data and meteorological data of the target station, the GaoXinQu monitoring station, are first integrated and concatenated. This operation fully considers the interactions among different air pollutants and meteorological factors to construct a multi-dimensional feature vector. Meanwhile, to depict spatiotemporal distribution characteristics of PM2.5 pollutants, the study further precisely matches and concatenates the PM2.5 concentration data from the other 9 adjacent stations (within a 30-km radius) according to timestamps. Eventually, an original dataset with an hourly unit is formed, as illustrated in Figure 8.

The descriptions of features for air pollutant data and meteorological data [24] are presented in Table 2.

In the integrated original dataset, due to the hourly time granularity of data collection, some missing values inevitably exist. To address this issue, a hierarchical processing strategy is implemented during the data preprocessing stage. First, data records with 10 consecutive missing time steps are deleted. For the remaining missing values, forward-filling and backward-filling methods are preferentially used for imputation. If the missing values at the front and back prevent the application of this strategy, the mean imputation method is employed for processing. After the above processing, a total of 35,064 time-series data records from January 1, 2019, to December 31, 2022, are finally obtained, and all data are arranged in strict chronological order. The temporal variation trends of key indicators such as PM2.5, PM10, SO2, Dew Point Temperature, and Air Temperature

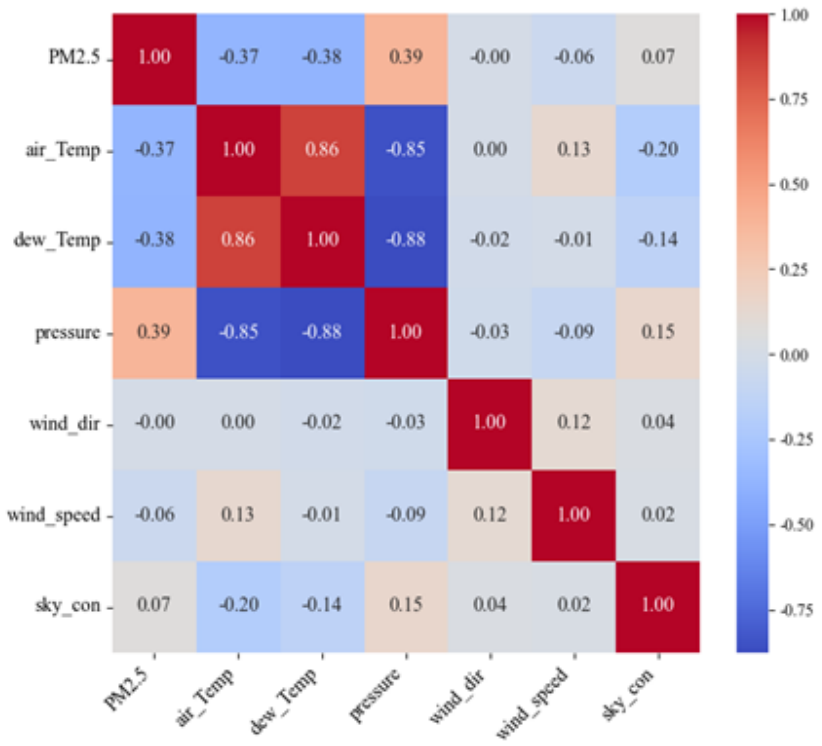

FIGURE 7. Correlation Coefficient Matrix Diagram of PM2.5 and Meteorological Data

TABLE 2. Feature Description of the Original Dataset

Feature	Description	Unit
PM2.5	PM2.5 Pollution	$\mu g/m^3$
PM10	PM10 Pollution	$\mu g/m^3$
SO2	Sulfur Dioxide	$\mu g/m^3$
NO2	Nitrogen Dioxide	$\mu g/m^3$
O3	Ozone	$\mu g/m^3$
CO	Carbon Monoxide	$\mu g/m^3$
air_Temp	Air Temperature	$°C$
dew_Temp	Dew Point Temperature	$°C$
pressure	Sea Level Pressure	kPa
wind_dir	Wind Direction	16 Direction
wind_speed	Wind Speed	m/s
sky_con	Sky Conditions	/

are shown in Figure 9. To ensure the effectiveness of model training and evaluation, the experimental dataset is divided into a training set, a validation set, and a test set at a ratio of 7:1:2. The specific division is shown in Table 3.

4.3. **Experimental Design.** This paper proposes a PM2.5 concentration prediction method based on a deep learning hybrid model for spatio-temporal fusion. This method deeply explores the spatio-temporal characteristics and interactive influences of air pollutants and meteorological factors. It organically integrates the advantages of CNN in local feature extraction, the long-term dependence modeling ability of LSTM, and introduces the ECA mechanism to enhance the response to key features, constructing a

date	hour	PM2.5	PM10	SO2	NO2	O3	CO	air_Temp	dew_Temp	pressure	wind_dir	wind_speed	sky_con
2020/1/1	0	38	75	7	78	5	0.5	2	-24	10321	50	6	8
2020/1/1	1	33	68	8	75	5	0.5	20	-30	10321	110	20	0
2020/1/1	2	40	67	7	65	9	0.4	30	-20	10321	130	40	0
2020/1/1	3	34	57	6	63	9	0.4	42	-16	10321	122	25	7
2020/1/1	4	38	56	6	61	10	0.4	50	-20	10321	130	30	7
2020/1/1	5	42	60	6	58	11	0.3	60	-20	10321	130	30	7
2020/1/1	6	39	68	7	61	8	0.4	57	-19	10282	85	29	8
2020/1/1	7	50	64	6	60	8	0.4	60	-30	10282	110	30	8
2020/1/1	8	51	70	7	60	7	0.4	60	-30	10282	120	30	8
2020/1/1	9	48	77	7	54	11	0.5	57	-5	10287	60	6	8
2020/1/1	10	49	81	8	50	15	0.5	60	-10	10287	70	20	8
2020/1/1	11	46	75	8	48	19	0.5	60	-10	10287	70	20	8
2020/1/1	12	49	70	8	49	20	0.5	56	10	10287	74	21	8
2020/1/1	13	50	69	10	50	25	0.5	60	0	10287	90	20	0
2020/1/1	14	55	61	10	52	26	0.5	50	10	10287	130	30	0
2020/1/1	15	44	65	9	57	24	0.5	50	20	10287	150	20	0
2020/1/1	16	46	71	8	55	27	0.4	50	20	10287	150	10	0
2020/1/1	17	52	78	8	64	15	0.4	50	20	10287	150	10	0
2020/1/1	18	52	74	7	74	8	0.5	45	33	10281	356	15	0
2020/1/1	19	48	81	9	79	5	0.5	50	20	10281	356	10	0
2020/1/1	20	55	83	9	78	5	0.7	50	30	10281	356	10	0
2020/1/1	21	60	84	9	79	5	0.8	45	32	10281	348	14	0
2020/1/1	22	60	86	6	78	5	0.7	40	30	10281	20	10	0
2020/1/1	23	57	66	5	75	5	0.7	40	40	10281	0	0	0
2020/1/2	0	56	57	5	74	5	0.7	45	31	10299	356	17	8
2020/1/2	1	55	56	5	73	5	0.8	50	30	10299	60	20	8
2020/1/2	2	50	58	5	71	5	0.8	60	30	10299	50	20	8
2020/1/2	3	55	59	5	69	5	0.9	66	34	10306	41	9	8
2020/1/2	4	59	62	5	66	5	0.8	70	30	10306	350	20	8

date	hour	1270PM2.5	1271PM2.5	1272PM2.5	1273PM2.5	1274PM2.5	1275PM2.5	1276PM2.5	1277PM2.5	1278PM2.5
2020/1/1	0	36	34	37	24	39	33	33	52	34
2020/1/1	1	36	37	35	32	36	37	39	50	29
2020/1/1	2	29	36	33	26	29	37	35	38	34
2020/1/1	3	28	38	35	29	34	35	35	42	29
2020/1/1	4	32	40	34	25	35	39	39	44	32
2020/1/1	5	37	44	39	32	43	35	40	45	39
2020/1/1	6	36	41	42	28	43	35	32	48	38
2020/1/1	7	48	37	42	32	38	44	36	41	48
2020/1/1	8	40	41	42	34	38	42	43	49	44
2020/1/1	9	38	45	45	39	40	46	48	53	45
2020/1/1	10	41	55	51	39	42	52	48	50	41
2020/1/1	11	44	67	51	39	43	52	48	52	47
2020/1/1	12	46	69	49	41	46	54	48	52	47
2020/1/1	13	44	57	50	42	47	58	51	53	47
2020/1/1	14	49	56	52	40	61	58	51	58	49
2020/1/1	15	49	59	48	45	49	47	56	56	50
2020/1/1	16	51	47	57	46	62	47	45	48	44
2020/1/1	17	46	50	53	45	50	45	45	46	48
2020/1/1	18	55	55	50	42	48	53	46	51	46
2020/1/1	19	45	55	52	36	57	54	50	52	42
2020/1/1	20	58	60	52	41	65	64	49	64	48
2020/1/1	21	48	60	56	50	71	65	50	66	52
2020/1/1	22	52	60	67	51	73	64	58	62	45
2020/1/1	23	50	65	65	50	58	64	51	63	51
2020/1/2	0	45	62	61	46	52	54	55	57	49
2020/1/2	1	53	70	56	41	55	52	54	56	47
2020/1/2	2	53	62	53	45	57	58	54	70	52
2020/1/2	3	54	65	64	45	57	51	55	62	52
2020/1/2	4	56	60	54	45	57	52	53	64	51

FIGURE 8. Partial View of the Original Dataset

TABLE 3. Dataset Partitioning Details

Data Category	Data Number	Percentage
Train Data	24304	69.31%
Validation Data	4000	11.41%
Test Data	6760	19.28%
Total	35064	100%

multi-dimensional feature representation system to solve the problem of accurate prediction of PM2.5 pollution concentration in small areas. At the data input level, the model is based on the air pollutant data and meteorological data of the target station, systematically analyzing the impact of the combined action of multiple factors on the change of PM2.5 concentration. Meanwhile, considering the spatio-temporal propagation characteristics of air pollutants, it dynamically couples the temporal and spatial dimensions, comprehensively taking into account the dual influencing factors of time and

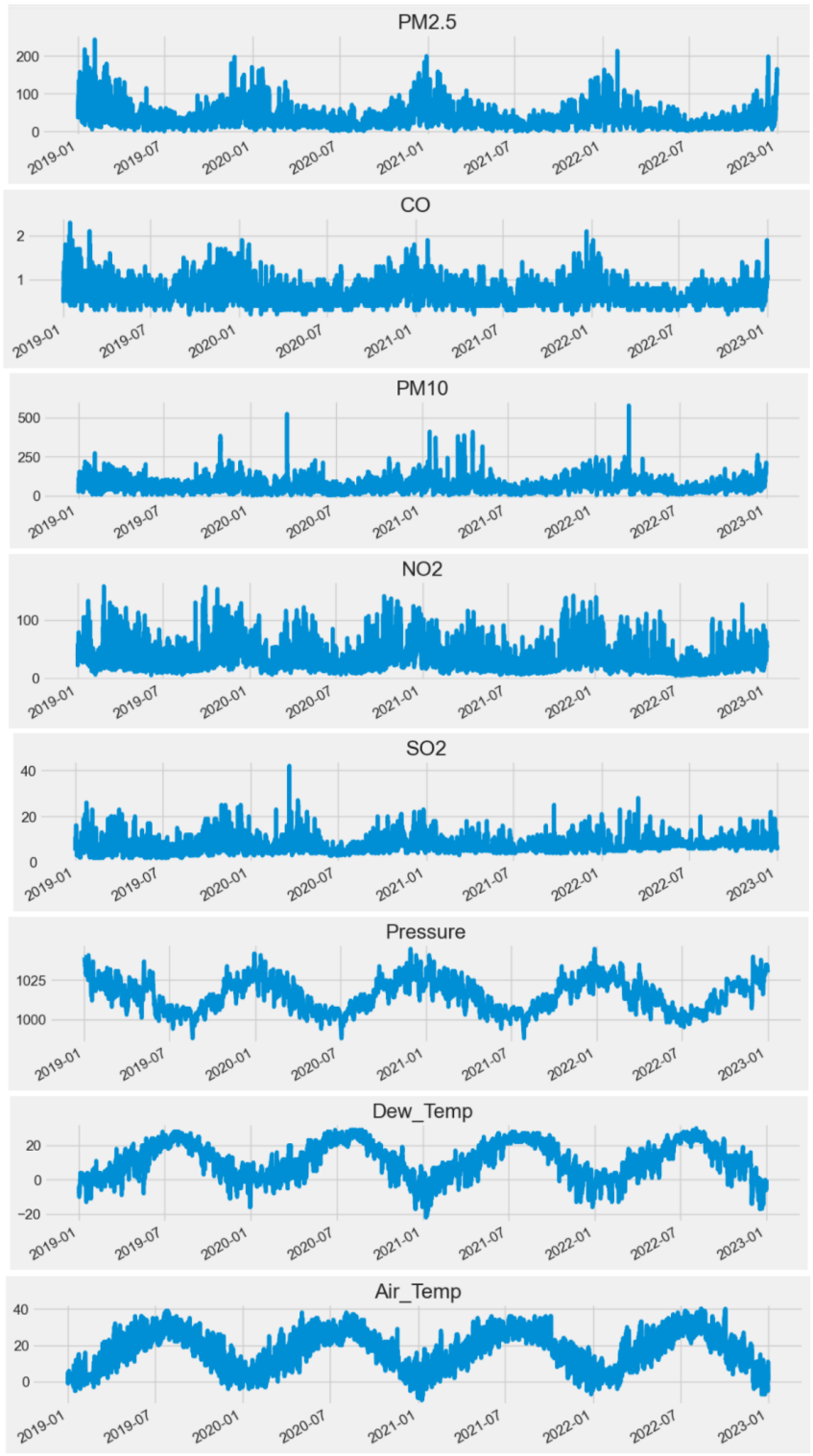

FIGURE 9. Temporal Trend Diagram of Experimental Data Features (Partial)

space in feature extraction. To reduce the computational complexity of the model, and prevent the risk of overfitting, the experiment only uses the PM2.5 data of the other 9 adjacent stations as input. By mining the spatial correlation between the target station and adjacent stations and the diffusion characteristics of pollutants, and integrating the

spatio-temporal information of multiple stations, the model's ability to capture the regional pollution propagation patterns is enhanced, thus achieving accurate prediction of PM2.5 in small areas. The model adopts a rolling prediction mechanism, using the PM2.5 data of the past 24 hours combined with other feature variables to predict the PM2.5 of the next hour. In subsequent iterations, the predicted values are combined with historical data to continuously generate future time-series prediction results. The network structure of the model and the optimal training parameters are shown in Table 4 and Table 5.

Table 4. Model Network Structure and Optimal Training Parameters

Layer Type	Kernel Size	Kernel Number	Nodes	Activation Function
Conv1	2	64	-	ReLU
Conv2	2	128	-	ReLU
LSTM1	-	-	256	-
Attention	-	-	256	Sigmoid
FC	-	-	1	-

Table 5. Model Training Parameter Settings

Parameter item	Parameter value
Num of CNN layers	2
Num of LSTM hidden layers	1
Num of hidden layer nodes	256
Num of Attention Layers	1
Learning rate	0.0001
Batch-size	128
Epochs	100
Loss function	MSE
Optimizer	Adam

In the CNN stage, model captures the local dependencies of adjacent time steps in the time series or spatially neighboring features, and extracts the underlying correlation patterns of multi-dimensional input data. Two layers of Conv1D are mainly constructed. The first layer of Conv1D uses 64 one-dimensional convolutional kernels with a size of 2 to perform preliminary feature abstraction on the input data, capturing basic features such as impact characteristics of other air pollution data on PM2.5 and the immediate correlation between meteorological factors and PM2.5. The second layer of Conv1D stacks 128 convolutional kernels with a size of 2. Based on the output of the first layer, it further extracts high-order composite features, such as the synergistic action patterns of multi-pollutant concentrations with Dew Point Temperature and Wind Speed, and the short-term coupling of pollutant concentrations between the target station and adjacent stations. By increasing the number of convolutional kernels, the feature representation space is expanded. Activation function of both layers of Conv1D is ReLU. In the LSTM stage, the model models the long term dependencies of time series and captures long - term change patterns of PM2.5 concentration, such as daily/weekly periodicity and seasonal trends. One layer of LSTM with 256 neurons is constructed to receive the feature sequence output by the two layers of Conv1D. Through the gating mechanism (forget gate, input gate, output gate), it selectively remembers key historical information and suppresses irrelevant noise. Next, the ECA is introduced to dynamically adjust the weights of spatio - temporal features and focus on the key information that contributes significantly to

PM2.5 prediction. A single - layer Attention is constructed, and the number of neurons is the same as the output dimension of LSTM (256). By calculating the attention weights of feature channels or time steps, it enhances the feature responses of meteorological factors (Pressure, Dew Point Temperature) during high - concentration pollution periods or strongly correlated pollutants (CO, PM10), suppresses redundant information, and the activation function is Sigmoid. Finally, the model enters the fully connected layer to output the prediction results. It maps the high-level abstract features to the predicted values of PM2.5 concentration, thus completing the regression task. The single-layer fully connected layer contains one neuron, which directly outputs the predicted value of PM2.5 concentration. This concise design of the output layer can integrate the features extracted and processed by the previous layers to obtain the final prediction results. Meanwhile, it avoids the risk of overfitting and the increase in computational complexity that may be caused by an excessive number of output neurons. The Adam optimizer is used in the experiment. The learning rate is initialized to 0.0001 and an exponential decay strategy is adopted. The Mean Square Error (MSE) is used as the loss function to measure the regression accuracy between the predicted values and the true values. The number of epochs is set to 100, and the batch size is set to 128.

4.4. **Analysis of Experimental Results.** In this paper, the Mean Absolute Error (MAE), Root Mean Square Error (RMSE), and Coefficient of Determination (R - Square, R^2) are used as evaluation indicators to measure the prediction accuracy of the model. Among them, the smaller the values of MAE and RMSE, the higher the prediction accuracy. The larger the R^2 value, the better the model fitting effect[21]. The calculation formulas are shown in Formula (12)-(14).

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|\hat{y}_l - y_i| \tag{12}$$

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(\hat{y}_l - y_i)^2} \tag{13}$$

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(y_i - \hat{y}_l)^2}{\sum_{i=1}^{n}(\bar{y}_l - \hat{y}_l)^2} \tag{14}$$

Where n is the total number of samples, y_i is the true value, $\hat{y}_l$ is the predicted value of the model, $\bar{y}_l$ is the average of the true values.

To verify the prediction performance of the model, a single LSTM model and a CNN-LSTM hybrid model were introduced as benchmarks in the experiment. In the comparative LSTM model, three layers of LSTM were utilized, with the number of neurons in the hidden layer of each LSTM set to 32. The remaining parameters were kept consistent with those of the CNN-ECA-LSTM model. The comparative CNN-LSTM model consisted of one layer of CNN and one layer of LSTM. The CNN in this model had a kernel size of 2 and 64 kernels, and the other parameters were the same as those of the CNN-ECA-LSTM model. The experiment employed the air pollutant data and meteorological data of the target station and adjacent stations over the past 24 hours to predict the PM2.5 concentration value for the next hour. That is, the time step (Timestep) was set to 24, the input feature dimension was 21, and the output feature dimension was 1. The experimental results are presented in Table 6. The results show that the constructed CNN-ECA-LSTM hybrid model performed best in PM2.5 concentration prediction, with a MAE of 4.24, a RMSE of 7.46, and a R^2 reaching 0.921, representing significant improvements compared to both the single LSTM model and the CNN-LSTM model without the attention

mechanism. Specifically, by integrating the local feature extraction ability of CNN and the long-term dependency capture ability of LSTM, the CNN-LSTM model reduced the MAE by 29.6% and increased the R^2 by 10.3% compared to the single LSTM model, verifying the positive impact of spatio-temporal feature fusion on prediction accuracy. After further introducing the ECA attention mechanism, the model's MAE decreased by 20.7% and the R^2 increased by 15.0% compared to the CNN-LSTM model. This indicates that the attention mechanism can effectively enhance the weights of key features, suppress redundant information, significantly improve the model's fitting ability to PM2.5 concentration changes and generalization performance, and provide an effective solution for accurate prediction in small areas.

TABLE 6. Comparison of Experimental Results

Model	MAE	RMSE	R2
LSTM	7.6	13.87	0.726
CNN-LSTM	5.35	11.81	0.801
CNN-ECA-LSTM	4.24	7.46	0.921

To observe the long-term prediction performance of the models, this paper selects the data of the next 500 hours from the same test set to compare the long-term prediction results, as illustrated in Figures 10-12. It can be found that CNN-ECA-LSTM model significantly outperforms LSTM model and CNN-LSTM model in the 500-hour long-term prediction. This indicates that through the dual mechanisms of spatio-temporal feature fusion and key information enhancement, the CNN-ECA-LSTM model effectively suppresses the problem of error accumulation in long term prediction and significantly improves the fitting accuracy for the long term evolution patterns of PM2.5. It also verifies the synergistic optimization effect of spatio-temporal modeling and the attention mechanism in long-sequence prediction.

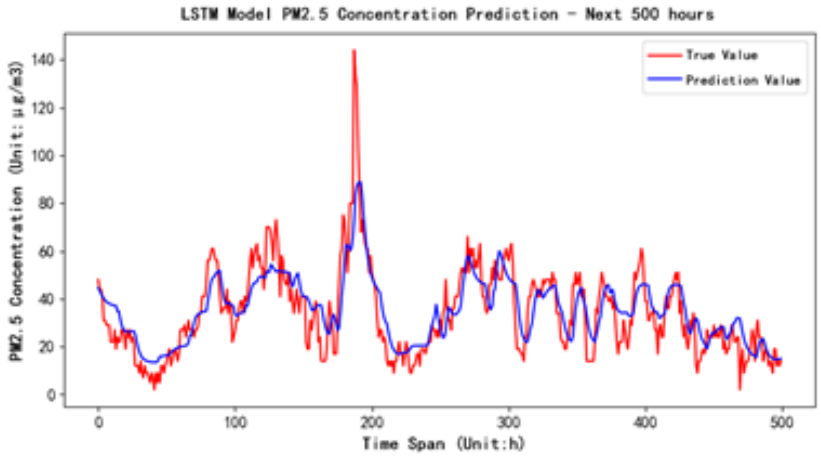

FIGURE 10. LSTM Model PM2.5 Concentration Prediction - Next 500 hours

5. **Conclusion.** This paper presents a CNN-ECA-LSTM deep learning hybrid model integrating spatiotemporal features and attention mechanisms, aiming to enhance the accuracy of PM2.5 pollution prediction. The model employs Conv1D to extract local spatiotemporal correlations from multi-source data, utilizes LSTM to capture long-term

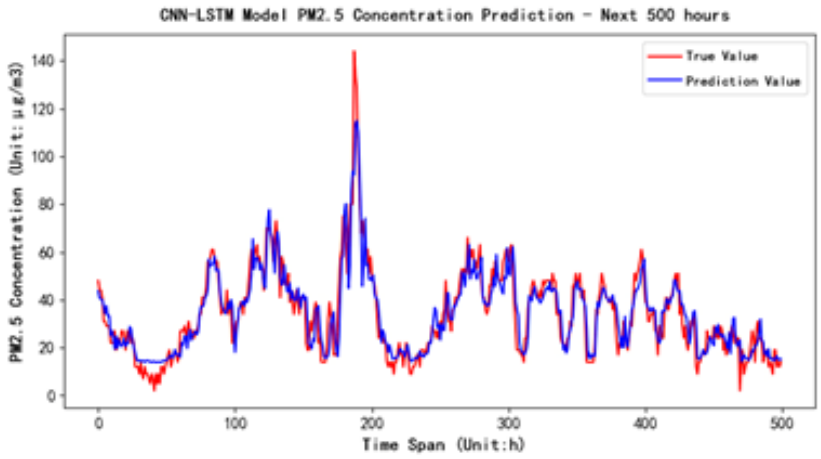

FIGURE 11. CNN-LSTM Model PM2.5 Concentration Prediction - Next 500 hours

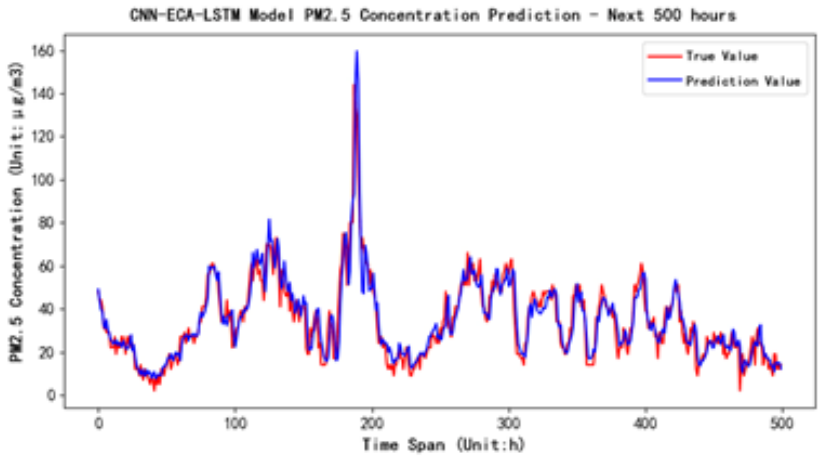

FIGURE 12. CNN-ECA-LSTM Model PM2.5 Concentration Prediction - Next 500 hours

dependencies in time series, and introduces ECA mechanism to reinforce the weights of key features. This framework enables effective integration of spatiotemporal features and suppression of redundant information, thereby significantly improving PM2.5 pollution prediction performance. The experimental dataset comprises hourly air pollutant and meteorological data from the GaoXinQu monitoring station in Hefei City from 2019 to 2022, with PM2.5 data from nine adjacent stations used as inputs to exploit spatial correlations and pollutant diffusion characteristics between the target station and its neighbors. The fusion of multi-station spatiotemporal information enhances the model's capability to capture regional pollution propagation patterns, enabling precise prediction of PM2.5 concentrations in small-scale areas. Comparative experiments demonstrate that the proposed model significantly outperforms the LSTM single model and the CNN-LSTM hybrid model in PM2.5 concentration prediction, fully validating the synergistic improvement of prediction accuracy through spatiotemporal fusion and attention mechanisms. This achievement not only provides an innovative solution for small-area pollution prediction but also offers a reusable technical framework and methodological reference for applying

deep learning hybrid models in big data domains with prominent temporal characteristics, such as meteorological data, stock markets, economic analysis, and transportation.

However, this study has several limitations. For instance, the model's hyperparameters (e.g., convolution kernel size, number of LSTM layers) currently rely primarily on empirical settings rather than data-driven automatic tuning, which may partially restrict the full realization of model performance. Additionally, due to the complexity of air pollutant diffusion mechanisms, this research cannot fully and accurately characterize the PM2.5 diffusion process. Future studies may consider incorporating PM2.5 diffusion models to further enhance the model's prediction accuracy.

Acknowledgment. This work was supported by the Key Scientific Research Project of the Anhui Provincial Department of Education (2022AH051867), the Provincial Quality Engineering Project of Anhui Higher Education Institutions (2023sysx029), and the College Student Innovation Training Program of Anhui Province (S202412216183).

REFERENCES

[1] Southerland V A, Brauer M, Mohegh A, "Global urban temporal trends in fine particulate matter (PM2.5) and attributable health burdens: estimates from global datasets," *The Lancet Planetary Health*, vol.6, no.2, pp.139-146, 2022.

[2] Zhu R X, Chen J, "Long-term effects of air pollution on hospital admissions and mortality for chronic obstructive pulmonary disease in Beijing, China," *The Clinical Respiratory Journal*, vol.17, no.7, pp.672-683, 2023.

[3] Ju Yang, "Prediction and Analysis of PM2.5 in Nanjing City Using Multiple Machine Learning Models," *Environmental Science Guide*, vol.44, no.2, pp.46-52, 2025.

[4] Yang Linxue, Yang Jun, Wu Xiaoying, "Study on the toxic mechanism of atmospheric PM2.5 induced apoptosis in human lung adenocarcinoma cells in Taiyuan City," *Journal of Environment and Health*, vol.42, no.4, pp.283-287, 2025.

[5] Suresh S, Sindhumol M R, Ramadurai M, "Forecasting particulate matter emissions using time series models," *Nature Environment & Pollution Technology*, vol.22, no.1, pp.221-228, 2023.

[6] Wongrin W, Chaisee K, Suphawan K, "Comparison of statistical and deep learning methods for forecasting PM2.5 concentration in northern Thailand," *Polish Journal of Environmental Studies*, vol.32, no.2, pp.1419-1431, 2023.

[7] Minh V T T, Tin T T, Hien T T, "PM2.5 forecast system by using machine learning and WRF model, a case study: Ho Chi Minh City, Vietnam," *Aerosol and Air Quality Research*, vol.21, no.12, pp.101-108, 2021.

[8] Ma X, Chen T, Ge R, "Time series-based PM2.5 concentration prediction in Jing-Jin-Ji area using machine learning algorithm models," *Heliyon*, vol.8, no.9, pp.10691-10702, 2022.

[9] Cai Haiyang, Li Yu, Fu Junyuan, "Research on the Application of Ensemble Learning in PM2.5 Prediction," *Henan Science and Technology*, vol.42, no.3, pp.26-30, 2023.

[10] Murugan R, Palanchamy N, "Smart city air quality prediction using machine learning," *2021 5th International Conference on Intelligent Computing and Control Systems (ICICCS)*, IEEE, pp.1048-1054, 2021.

[11] Van N H, Van Thanh P, Tran D N, "A new model of air quality prediction using lightweight machine learning," *International Journal of Environmental Science and Technology*, vol.20, no.3, pp.2983-2994, 2023.

[12] Hu J, Chen Y, Wang W, "An optimized hybrid deep learning model for PM2.5 and O3 concentration prediction," *Air Quality, Atmosphere & Health*, vol.16, no.4, pp.857-871, 2023.

[13] Zhang Y, He D, Wu Q, "Forecasting of PM2.5 concentration time series based on SSA-LSTM model," *International Conference on Statistics, Data Science, and Computational Intelligence (CSDSCI 2022)*, SPIE, pp.373-380, 2023.

[14] Wang Qianying, Yang Kexin, "PM2.5 concentration prediction based on LSTM-SVR hybrid model," *Information Technology and Informatization*, vol.258, no.9, pp.33-36, 2021.

[15] Xu Yixin, Ren Jie, Feng Lei, "Optimization of PM2.5 concentration prediction model based on wavelet analysis," *Environmental Monitoring Management and Technology*, vol.33, no.2, pp.24-28+34, 2021.

[16] Zhang Yanan, Sun Jianping, "A PM2.5 concentration prediction model based on wavelet neural network," *Electric Power Science and Engineering*, vol.36, no.1, pp.55-61, 2020.
[17] Peng Yuqing, Qiao Ying, Tao Huifang, "PM2.5 prediction model incorporating attention mechanism," *Sensors and Microsystems*, vol.39, no.7, pp.44-47, 2020.
[18] Rehman A, Chin-Feng L, "CNN-Based Intelligent Disease Detection and Identification Technique through Chest X-rays," *Journal of Information Hiding and Multimedia Signal Processing*, vol.15, no.4, pp.319-333, 2024.
[19] Feng Xu, Zhen Sui, Jiangang Ye, "Ternary Precursor Centrifuge Rolling Bearing Fault Diagnosis Based on Adaptive Sample Length Adjustment of 1DCNN-SeNet," *Processes*, vol.12, no.4, 2024.
[20] Weihu Liu, Shuo Liu, Yang Wang, "Research on Grain Pile Temperature Prediction Based on CNN-GRU Neural Network," *Springer Science and Business Media LLC*, vol.121, no.3, pp.214-226, 2022.
[21] Qiang Wang, Yiwen Zhang, "Research on PM2.5 Pollution Prediction Method in Hefei City Based on CNN-LSTM Hybrid Model," *Journal of Physics: Conference Series*, vol.24, no.1, pp.12006-12016, 2022.
[22] Zhu Yongbing, Cai Yuqin, Jiang Liyao, "Research on Quantitative Analysis Method of Intravenous Drug Concentration Using TDLAS Based on ECA-1D-CNN," *Spectroscopy and Spectral Analysis*, vol.45, no.5, pp.1341-1347, 2025.
[23] Li Dan, Fang Rong, Chen Jingyuan, "Research on the Vitality Characteristics, Influencing Factors, and Enhancement Strategies of Tourist Attractions Based on Multi-source Data: A Case Study of Hefei City," *Journal of Anhui Agricultural University*, vol.51, no.6, pp.987-995, 2024.
[24] Jiming Li, Guangyuan Xu, Xuezhen Cheng, "Combining spatial pyramid pooling and long short-term memory network to predict PM2.5 concentration," *Atmospheric Pollution Research*, vol.13, no.3, pp.309-317, 2021.