Forecasting Indoor Nitrogen-Dioxide Concentrations in a Student Dormitory Using RNN–LSTM Models

by

Jia XIE[a,b], Banglong WU,[a,b] Shuting YANG [a,b], and Yaowen XIA [a,b,*]

AFFILIATIONS

[a,*]*School of Information Science and Technology, Yunnan Normal University, Kunming, 650500, China*

[b]*Southwest United Graduate School, Kunming, Yunnan 650092, China*

a) Authors to whom correspondence should be addressed: yaowenxia@ynnu.edu.cn

Air quality has a profound impact on the urban environment, with both positive and negative effects. Developing effective strategies for improving and predicting air quality is crucial for urban environmental management. This study was conducted from June 2021 to March 2022, with a sample size of 584. The study collected air quality data from dormitory buildings, including indoor temperature, wind speed, humidity, and nitrogen dioxide (NO_2) concentration. Additionally, we conducted a qualitative analysis to explore the relationship between atmospheric parameters and average NO_2 concentration. We compared MLP with recurrent neural networks (RNN) and long short-term memory (LSTM) models for optimization in predicting NO_2 concentrations. Experimental results showed that the combination of RNN and LSTM significantly improved the accuracy of NO_2 concentration predictions. This study provides important references for monitoring NO_2 concentrations in university dormitories and improving air quality.

Keywords: *NO2; LSTM; RNN; Air Quality Index*

1 Introduction

1.1 Background and significance of the study

Following the 20th century, the country embarked on a vigorous path of industrialization, which consequently led to an increase in air pollution levels. Each year, approximately 5.5 million individuals die prematurely as a result of anthropogenic air pollution [1]. Air pollutants have historically been, and continue to be, significant contributors to chronic diseases and mortality, thereby adversely affecting public health [2].

[*] [*] Corresponding author.

E-mail:yaowenxia@ynnu.edu.cn (Y.Xia)

The effects of air pollutants on the environment and human health are influenced by factors such as time, space, duration of effects, and concentration. Numerous pollutants can be found in the air, including carbon monoxide (CO), nitrogen dioxide (NO_2), and particulate matter (PM2.5) [3-4]. Among these, nitrogen dioxide (NO_2) is particularly concerning, as it can be life-threatening and may cause breathing difficulties, headaches, and other uncomfortable conditions when its concentration exceeds manageable levels. A healthy environment is crucial for higher education students, who spend, on average, approximately 90% of their time indoors [5]. Therefore, effectively controlling air pollution, especially in residential areas of schools, is a critical issue that must be addressed. Accurate and effective prediction and analysis of nitrogen dioxide (NO_2) concentrations are essential for ensuring indoor comfort.

Artificial intelligence (AI), a powerful tool, has been widely utilized in image understanding, speech recognition, natural language processing [6-7], and air quality prediction [8]. Numerous researchers have conducted extensive studies on air quality prediction using AI technologies, including Recurrent Neural Networks (RNN) and Long Short-Term Memory (LSTM) networks. Chang et al. proposed an aggregated LSTM model method to predict PM2.5 concentrations over 1 to 8 hours [9]. Gao and Li introduced Graph-based Long Short-Term Memory (GLSTM) to forecast PM2.5 concentrations in Gansu Province, China [10]. Seng et al. utilized LSTM to predict air quality pollution indicators, including PM2.5, carbon monoxide (CO), nitrogen dioxide (NO_2), ozone (O_3), and sulfur dioxide (SO_2) [11]. Yan et al. developed a multi-site forecasting model for Air Quality Index (AQI) predictions over the next 1-6 hours [12].

Recently, LSTM has gained popularity for predicting air pollutant concentrations due to its ability to solve complex tasks quickly and accurately [13-15][16]. As an emerging alternative technique, LSTM demonstrates strong performance in air quality prediction for individual sites, surpassing traditional persistent statistical models [17]. It has been developed, improved, and applied to time series forecasting because of its capacity to learn from long time series without encountering the vanishing gradient problem [18]. Cheng liang Xu et al. employed the LSTM model to accurately predict indoor temperature trends, revealing a slight advantage in the performance of horizontal indoor temperature predictions [19].

Recurrent Neural Networks (RNN) are among the most prevalent architectures in the field. Hochreiter et al. proposed a method that combines spatio-temporal convolutional neural networks (CNN) with long short-term memory (LSTM) networks to predict the concentration of the next day's daily average PM2.5 in Beijing City [20]. Rao et al. utilized pollutant concentration data from Visakhapatnam, India, to construct a model for predicting air quality using RNNs, with validation results indicating improved prediction accuracy [21]. Krishan et al. employed the LSTM method to forecast concentrations of O3, PM2.5, NO_2, and CO at a location in NCT-Delhi, demonstrating that it was more effective than other deep learning methods [22]. Zheng Yangyang et al. established an LSTM model to predict the Air Quality Index (AQI) of Taiyuan City, confirming its advantages in terms of high prediction accuracy and broad applicability [23]. Jie Huang et al. utilized an RNN with strong memory capacity to leverage the correlation information along the time axis for predicting future PM2.5 concentrations. [24] Kristiani et al. employed two types of time series data with trained models to forecast air pollution over an eight-hour period, concluding that LSTM is more suitable than other neural networks for constructing deep learning prediction models for air pollution data [25]. Bihter Das et al. applied LSTM, RNN, and multilayer perceptron (MLP) models to predict PM10 and SO_2 atmospheric pollutants in 2022, with results indicating that the LSTM outperformed both the MLP and RNN models [26]. Kim et al. developed a water usage prediction model for individual customers using LSTM, comparing it with traditional time series prediction models and finding that the LSTM model demonstrated superior performance [27].

Apeksha Aggarwal et al. propose a methodological model for air quality prediction, which is based on air quality data from 15 locations in India. This model demonstrates superiority over existing prediction models when compared to other learning models [28]. Freeman et al.predicted 8-hour average surface ozone (O_3) concentrations using a deep learning approach that incorporates recurrent neural networks (RNN) with long short-term memory (LSTM) [29]. The accuracy of the mean absolute error (MAE) measurement improved significantly when the number of features used to train the LSTM model was reduced from 25 to 5. The results indicate that the predictions are computable. Patricio Perez et al. developed a neural network and a linear model utilizing input variables including current hour-by-hour PM2.5 values for 18 h and 19 h, forecasted temperatures, wind speeds, precipitation, and measured concentrations of NO_2, CO, and O_3, to predict the maximum value of the 24-hour moving average one day in advance. The experiments reveal that the neural network model outperforms the linear model in terms of accuracy [30]. R. Janarthanan et al. combined deep learning with support vector regression (SVR) and LSTM to classify AQI values [31]. Their proposed deep learning model provides accurate and specific AQI values for designated locations within the city, demonstrating the capability to accurately predict urban AQI. Al-Janabi et al. proposed a method for predicting air pollutant concentrations, including PM2.5, PM10, NO_2, CO, O_3, and SO_2 for the next two days using deep learning with RNN [32]. Their design employs a Particle Swarm Optimization algorithm to determine the optimal structure for its operation, along with a novel predictor based on unsupervised learning, specifically LSTM combined with PSO, which yields favorable prediction results.

Based on the current state of research on air quality, it can be preliminarily concluded that using recurrent neural networks (RNNs) to predict nitrogen dioxide (NO_2) concentrations in university dormitories is feasible. Long Short-Term Memory (LSTM) networks are particularly noteworthy due to their ability to address gradient explosion and vanishing backpropagation error issues, as well as their effectiveness in capturing long-term and short-term dependencies. Therefore, enhancing RNN models with LSTM significantly improves the accuracy of data predictions. Air quality data is inherently closely related to temporal factors and exhibits time series characteristics with distinct periodicity. Given the time-sensitive nature of the data, time prediction has undoubtedly become a key research area. RNNs and LSTMs perform exceptionally well in time series analysis and prediction, with their feature extraction and nonlinear representation capabilities enabling the automatic extraction of deep features from the data, thereby improving prediction accuracy [33]. However, research on NO_2 air quality prediction in student dormitories is relatively scarce, both domestically and internationally. Therefore, this paper aims to use LSTM for optimized prediction, comparing RNN and LSTM optimization to predict NO_2 concentration, which has important research value.

This study focuses on predicting and analyzing the concentration of NO_2 in student dormitories using RNN. Specifically, the air quality within the student dormitory is monitored and recorded daily. The collected data is then processed and analyzed using an RNN model, which is optimized through LSTM. RNN and LSTM were selected for their proven effectiveness in time series prediction tasks, which are crucial for accurate air quality forecasting. In the realm of air pollution research and forecasting, traditional methods typically employ statistical regression models and numerical forecasting techniques. While these approaches are convenient and quick, they often yield high error rates and low forecast accuracy, failing to meet contemporary forecasting demands. This paper builds upon RNN methodologies to study and analyze NO_2 predictions. It processes data through sampling and prediction, comparing training values to target values to assess the satisfaction of prediction results. LSTM is utilized for enhancement, and the predicted values are contrasted with actual values to determine whether the anticipated outcomes are achieved.

2 Materials and Methods

2.1 Air Quality Index (AQI)

The Air Quality Index (AQI) is a dimensionless indicator used to measure air quality conditions. According to the Technical Specifications for the Air Quality Index (AQI) (Trial Version) issued by China, the index is divided into six levels, each representing a different air quality scenario. As shown in Table 1, this table illustrates the classification of air quality.

Tab. 1 Air Quality Classification

Air index	Air rating index	Air Index	Classification and color
≤50	Grade 1	Excellent	Green
≤100	Grade 2	Good	Yellow
≤150	Grade 3	Slightly polluted	Orange
≤200	Grade 4	Moderately polluted	Red
≤300	Grade 5	Heavily polluted	Purple
＞300	Grade 6	Severely polluted	Maroon

2.2 Back Propagation Neural Networks (BPNN）

The BPNN is defined as a 'generic model integrated with an error correction mechanism.' By analyzing the training outcomes and the anticipated results, the model adjusts its weights and thresholds iteratively to ensure that the output matches the expected outcomes. This method is among the most commonly employed models in neural network applications [34]. Figure 1 depicts the architecture of the BP neural network model. This network is composed of three distinct layers: the input layer, the hidden layer, and the output layer. The input layer functions to gather input data, the hidden layer houses the neural network’s parameters, and the output layer consists of a single output parameter.

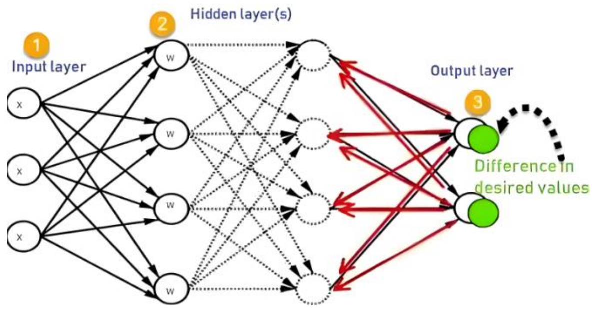

Fig. 1 BP Neural Network Model

2.3 Recurrent Neural Network (RNN)

Backpropagation (BP) Neural Networks identify linked data without utilizing previous data as input for subsequent data, which limits the algorithm's ability to mine data for pattern generation [35]. Given the necessity for internal patterns in time series data, RNN is well-suited for processing and predicting sequential data as shown in Figure 2. These networks capture the state of change over time, with data collected at various time points. RNN is inherently memory-based, as the input data are interconnected. They are recognized for their effective predictive modeling capabilities when dealing with time series data, employing a sequence of input data with cyclic connections among segments. In RNN, neurons within the same hidden layer are interconnected, allowing for the repeated application of a training function to the hidden state [36].

As illustrated in Figure 2, the right graph depicts the unfolded view of the model. The left graph represents the initial time step in the sequence where the input X is introduced. The matrix U serves as the weight matrix that facilitates the connection between the input and the hidden layer. The variable w denotes a state within the hidden layer. The output S from the hidden layer is generated alongside the subsequent input X and is stored in the memory unit to produce the final result O. In RNN, the nodes are interconnected between the hidden layers, and the inputs to the hidden layer encompass not only the outputs from the input layer but also the outputs from the hidden layer at the preceding time step [37]. The RNN generates an output for each input in conjunction with the model's state.

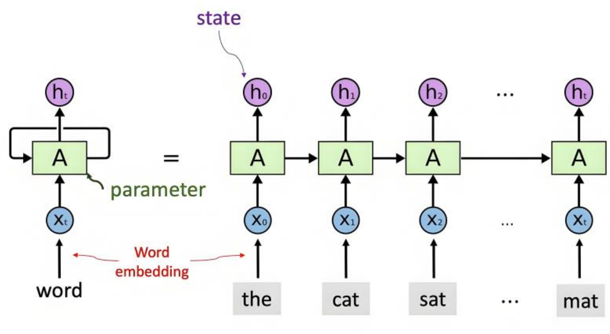

Fig. 2 Cyclic Neural Network Model

2.4 Long Short-Term Memory (LSTM)

Long Short-Term Memory (LSTM) networks are a specific type of Recurrent Neural Network (RNN) that have been developed to overcome the long-term dependency issues commonly associated with standard RNNs. All RNNs display a chaining mechanism through the repetitive use of neural network components. LSTM networks have proven to be exceptionally capable of capturing long-term dependencies, which is a vital element in analyzing sequential data. Furthermore, they adeptly address the complexities introduced by nonlinear, periodic, seasonal, and sequential dependencies within data sequences. This is particularly illustrated in the context of air quality prediction, which serves as the primary focus of this research, as depicted in Figure 3.

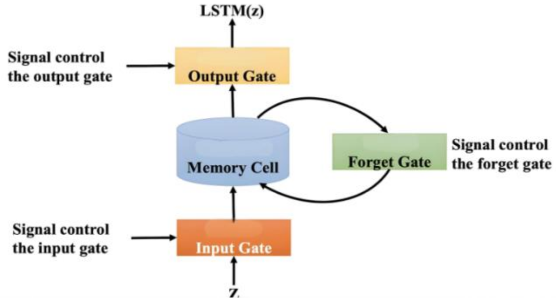

Fig. 3 The structure of LSTM

The figure 3 illustrates that the input gate regulates the amount of new information entering the network, the forgetting gate governs the retention of information from previous moments, and the output gate controls the amount of information exiting the network [38]. In this paper, we employ a multilayer LSTM to optimize the RNN neural network. The optimization of LSTM necessitates the establishment of a controlled learning rate, which is activated when the results remain unchanged for more than one consecutive iteration during training [39].

3 NO_2 Predictive modelling

3.1 Data collection

This study collected air quality data from student dormitories at Yunnan Normal University between June 2021 and March 2022. Due to significant temperature differences between morning and evening, measurements were taken once in the morning and once in the evening each day to ensure that the data used for model optimization and prediction was more rigorous during the analysis process. A handheld portable nitrogen dioxide tester was placed in the activity area of the student dormitories, and this study chose to collect data next to the desks. The collected data encompass four parameters: temperature, humidity, air quality, and wind speed. The key air quality data are illustrated in Figure 4.

time	temp	hum	wind	air	no2
02.06.2021 08:00	19.8	73	2	24	23
02.06.2021 14:00	26.7	47	3	36	5
03.06.2021 08:00	20.2	68	2	35	14
03.06.2021 14:00	27.4	44	2	46	8
04.06.2021 08:00	13.9	61	2		14
04.06.2021 14:00	20.8	43	1	37	18
05.06.2021 08:00	19.6	66	1	49	32
05.06.2021 14:00	24.8	48	2	52	10
06.06.2021 08:00	18.6	82	1	28	11
06.06.2021 14:00	24.2	61	2	44	5
07.06.2021 08:00	17.1	94	1	24	14
07.06.2021 14:00	23.0	65	2	31	6
08.06.2021 08:00	15.6	94	2	28	15
08.06.2021 14:00	16.4	92	1	30	8
09.06.2021 08:00	17.5	90	2	39	25
09.06.2021 14:00	22.5	71	3	30	6
10.06.2021 08:00	17.7	96	3	27	36

Fig. 4 Key Data of Air Quality

The NO_2 concentration, illustrated in Figure 5, reveals a period from November 2021 to December 2021 during which no data were collected, resulting in a linear representation. This diagram facilitates a clear observation of both the maximum and minimum NO2 concentration values, which can be further analyzed to assess compliance with standards.

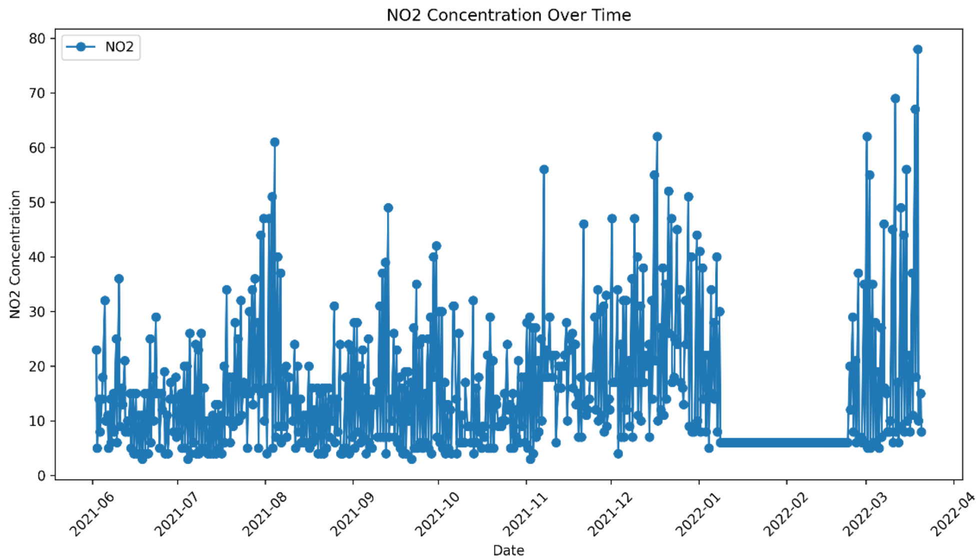

Fig. 5 NO_2 Concentration

3.2 Data processing

The data collected in this study needs to be converted in the model to ensure recognizability, such as converting strings to date types. Therefore, before data preprocessing, strings must be converted to date types recognizable by the model, and factors such as missing data and data standardization must also be considered during model construction. This study encountered missing data issues during data collection, which may be attributed to the twice-daily data collection interval. Methods for handling missing data include directly deleting all data for missing dates and various data imputation techniques. While the direct deletion method is simple and convenient, it may negatively impact the accuracy of subsequent data analysis, leading to poor prediction results. Data imputation methods include backward imputation and forward imputation. This study compared forward imputation and backward imputation, as shown in Figure 6, where forward imputation yields significantly lower MAE and RMSE values than backward imputation. Forward imputation performs better in missing value imputation tasks with smaller errors, making it a suitable method for handling missing data in this study. After data processing, standardization will be performed. The purpose of data standardization is to ensure that all parameters of different features (i.e., input layers) are within the same range. This facilitates the training of the neural network and produces more reliable prediction results.

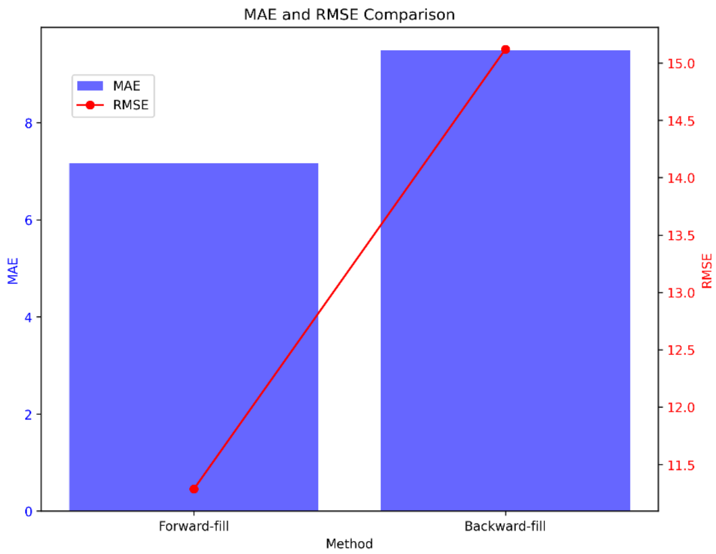

Fig. 6 MAE and RMS Comparison

3.3 Data training

In the training set, insufficient data can lead to incomplete training of the network, thereby hindering the achievement of desired results. Conversely, an excess of data may result in overfitting, which adversely impacts the accuracy of subsequent predictions. This paper utilizes data from five parameters—temperature, humidity, wind speed, air quality, and NO_2—collected from June 2021 to March 2022 at the student dormitory of Yunnan Normal University, located in Chenggong District, Yunnan Province. Specifically, data from the first five days will be used to predict outcomes for the following day, while data from the penultimate day will be collected for processing. After conducting data analysis, the first 80th percentile of the dataset will be allocated for training, with the remaining 20th percentile reserved for testing.

3.4 Data forecasting

Relevant research indicates that wind speed, humidity, temperature, and air quality significantly influence NO_2 concentration. Consequently, incorporating these parameters as input variables into the model may enhance prediction accuracy. In this paper, we utilize wind speed, humidity, temperature, and air quality as input parameters within a recurrent neural network (RNN) framework. The analysis is performed using jupyter notebook, with the necessary Python libraries installed for training. We employ the keras API as the front end and TensorFlow as the back end to construct the neural network prediction model. The dataset is divided into training samples, test data, and target data, followed by predictions made after the training phase.

3.5 Design of neural network structure

In this study, data from the first five days was used to predict conditions for the following days, specifically analyzing six relevant factors: temperature, humidity, wind, air quality index, and nitrogen dioxide over a six-day period. This indicates that the model consists of five input layers, while the prediction focuses solely on nitrogen dioxide concentration values, resulting in only one output layer. This paper employs an MLP model to compare the RNN model and the optimized LSTM model that need to be validated in this experiment, further examining whether LSTM is suitable for the predictions in this experiment.

4. Results

4.1 NO_2 concentration prediction and analysis

As shown in Figures 7 and 8, this study used a classic MLP model for simulation training. The training and validation loss curves were visualized, and the actual values of NO_2 were compared with the predicted values. The results were not satisfactory. There was a significant difference between the predicted values and the actual values.

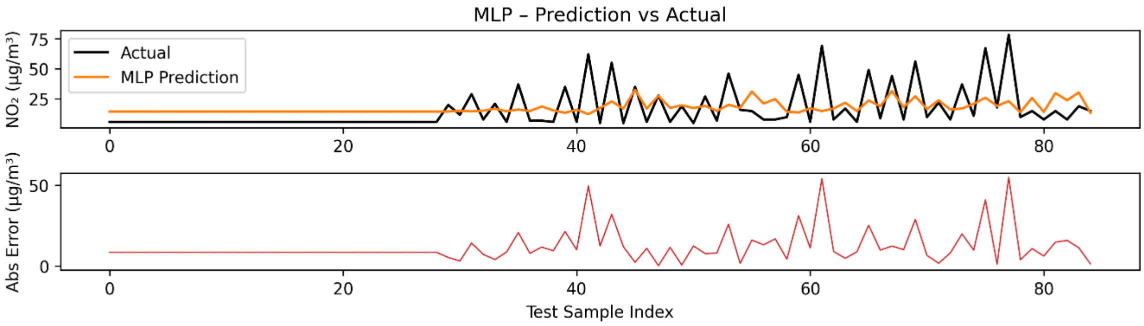

Fig. 7 MLP Prediction and Actual Comparison

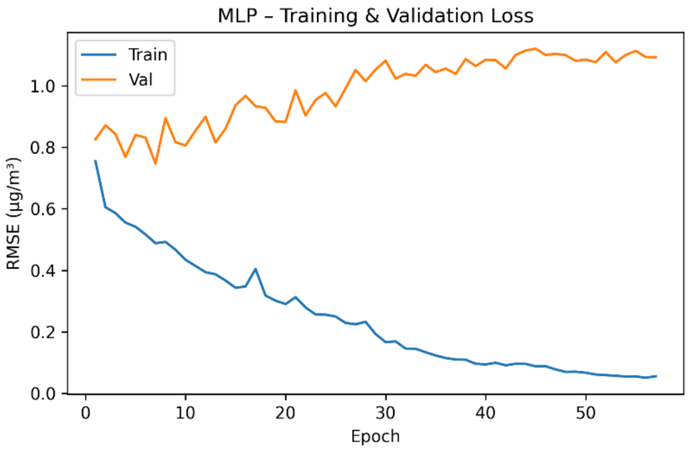

Fig. 8 MLP Prediction and Validation Loss

The results of the MLP prediction model do not represent the ideal results we hope to achieve. As shown in Figures 9 and 10, the same training operation was performed on the RNN.

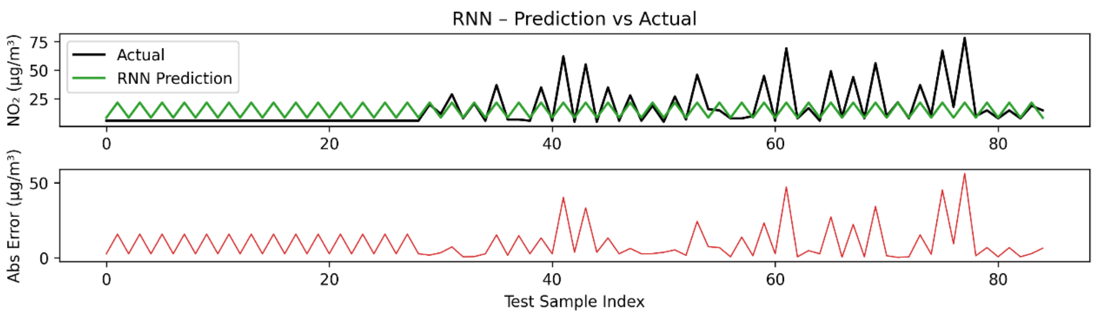

Fig.9 MLP Prediction and Actual Comparison

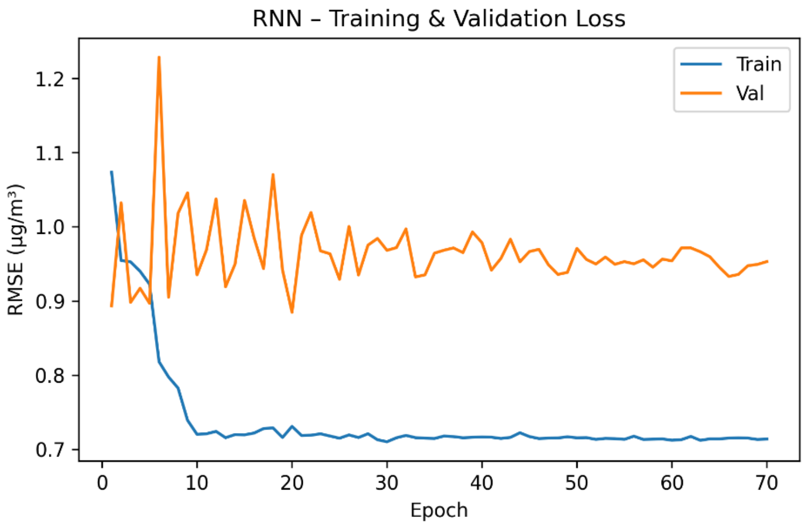

Fig.10 MLP Prediction and Validation Loss

We compared MLP and RNN, as shown in Figure 11, and it is clear that RNN is an improvement over MLP. Next, we will use the trained RNN model to predict the NO2 concentration value for March 21, 2022.

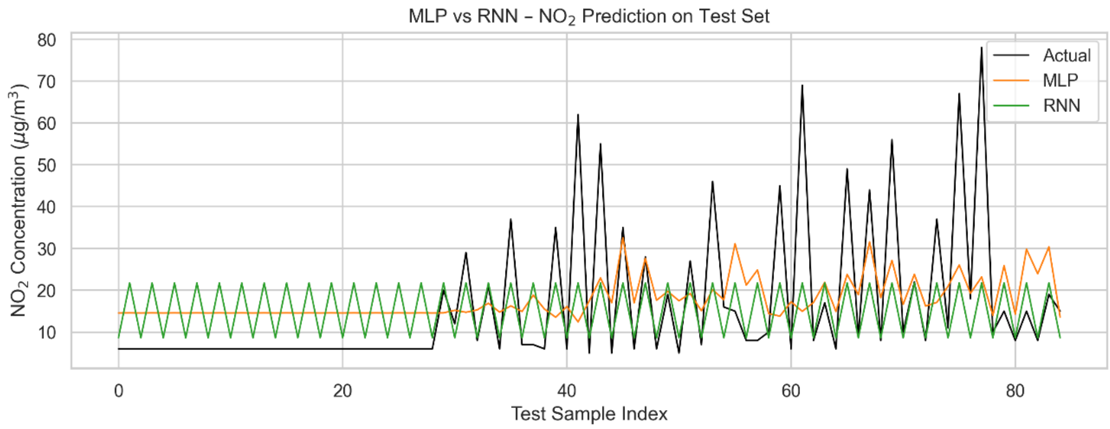

Fig.11 RNN vs MLP

The trained RNN model predicted a NO_2 concentration value of 21.8 for March 21, 2022, while the actual value for that day was 15, as shown in Figure 12. As can be seen, there is a significant difference between the predicted and actual values. For this experimental study, while the RNN model shows some improvement over the MLP model, it still falls far short of the desired results. Therefore, we will optimize the model by training it using an LSTM model to assess whether the performance can be improved.

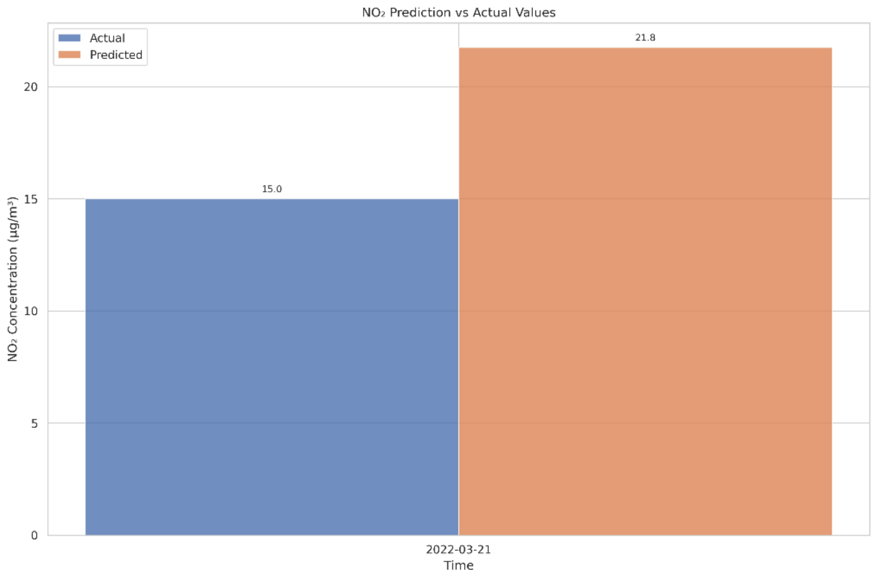

Fig.12 RNN Prediction Values

As shown in Figures 13 and 14, using the LSTM model optimization, we first examine the comparison between actual values and predicted values in the LSTM model, as well as the comparison between training and validation losses. It can be seen that the LSTM model performs better than the MLP model and RNN model in terms of training effectiveness.

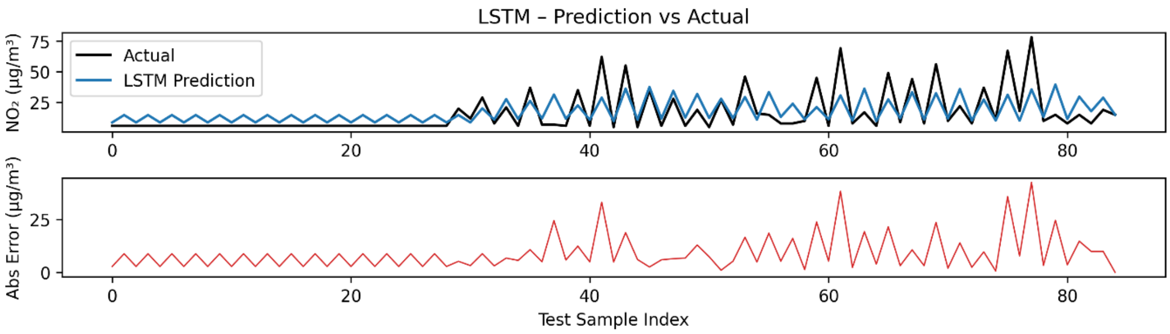

Fig. 13 LSTM Prediction and Actual Comparison

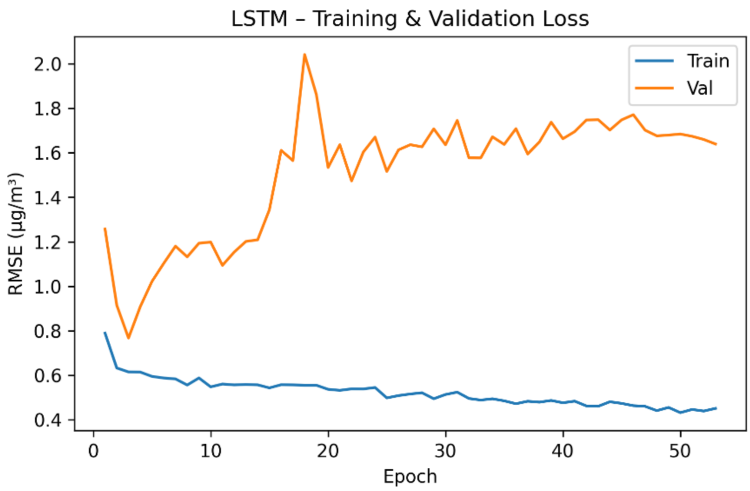

Fig.14 LSTM Prediction and Validation Loss

The RNN model and LSTM model trained in this experiment were compared, as shown in Figure 15. It can be seen that the LSTM model performs better than the RNN model in terms of accuracy.

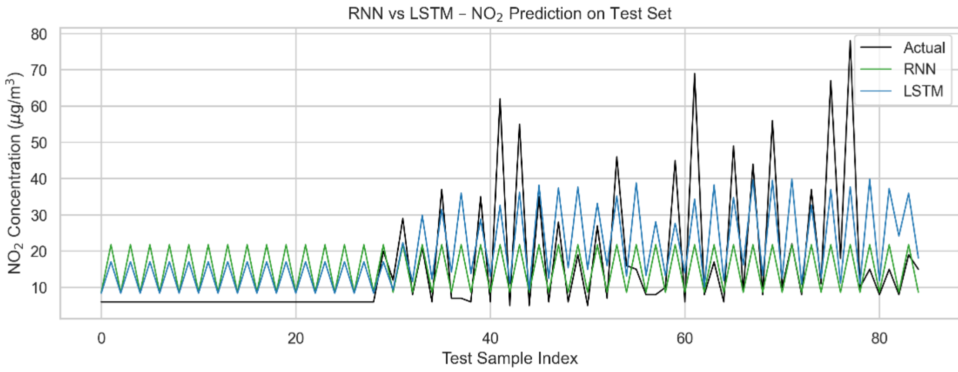

Fig.15 RNN vs LSTM

By comparing these three models, as shown in Figure 16, the LSTM model clearly outperforms the other two models and is closer to the actual values. As can be seen from the labeled RMSE in the figure, the LSTM model is significantly superior to the RNN and MLP models. To further demonstrate the feasibility of the LSTM model, we will use 5-fold cross-validation to verify that the LSTM model is suitable for this experiment.

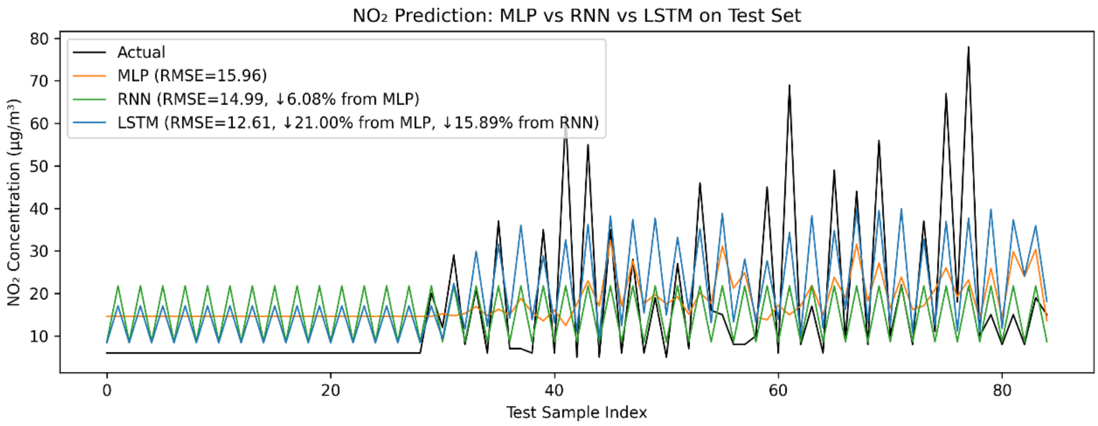

Fig.16 RNN vs LSTM vs MLP

Five-fold cross-validation was performed on the three models, as shown in Figure 17. LSTM achieved the lowest error in both RMSE and MAPE metrics and was statistically significantly better than the Naive baseline (paired t-test $p < 0.001$). Therefore, it can be clearly seen that LSTM is the best choice in this study.

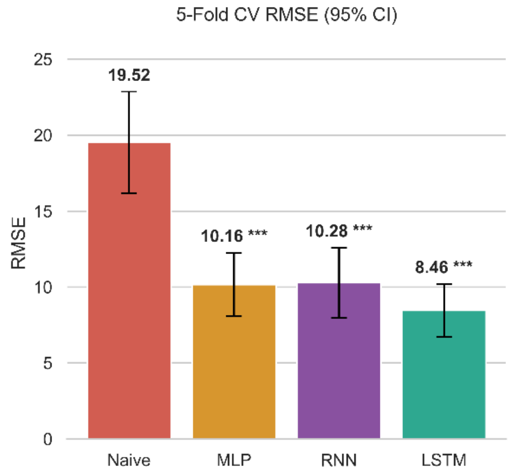

Fig.16 5-Fold CV RMSE

4.2 NO_2 Projected results

The LSTM model has clearly reached a satisfactory state. At this point, we use the model to predict the nitrogen dioxide concentration value for March 21, 2022.

On March 21, 2022, the actual measured value of nitrogen dioxide (NO_2) concentration at Yunnan Normal University was 15. As shown in Figure 17, the predicted measured value of nitrogen dioxide (NO_2) concentration was 14.999865, demonstrating extremely high accuracy, indicating that the prediction and analysis successfully achieved the expected objectives. Additionally, these results provide valuable data for ongoing nitrogen dioxide concentration prediction research, thereby contributing to the protection of faculty and student health.

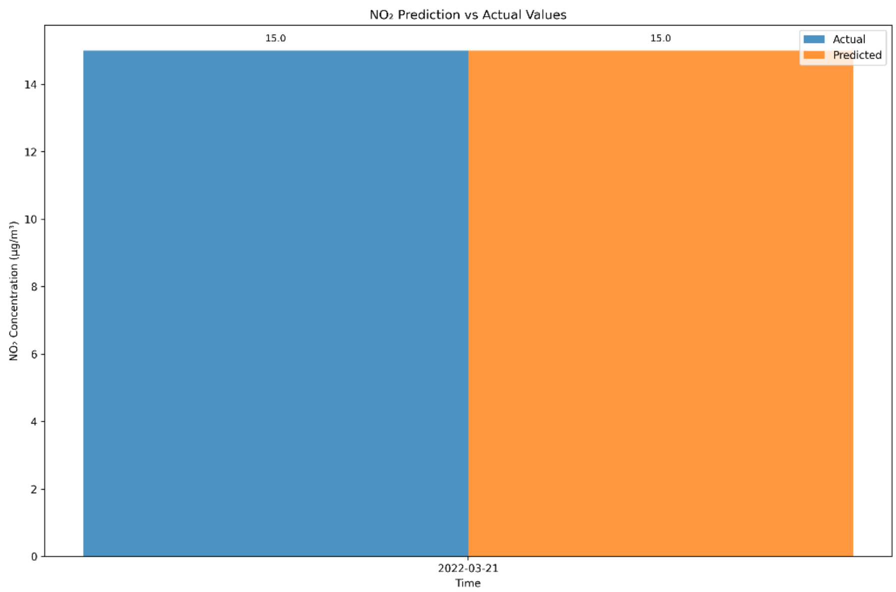

Fig. 17 LSTM Prediction Results

4.3 Relationship between NO_2 concentration and air parameters

A qualitative analysis was further conducted on the relationship between commonly used air parameters and the average concentration of nitrogen dioxide (NO_2), as demonstrated below. NO_2 concentration is negatively correlated with the indoor temperature of the dormitory as shown in Table 2; specifically, as the temperature increases, the NO_2 concentration decreases.

Tab. 2 Relationship between Air Temperature and Average NO_2 Levels

Temp (^{0}C)	$No_2(\mu g/m^{-3})$
$5 \leq T \leq 10$	30.778
$10 < T \leq 15$	23.878
$15 < T \leq 20$	23.832
$20 < T \leq 27$	21.782

The concentration of NO_2 is negatively correlated with wind speed as shown in Table 3. As wind speed increases, the concentration of NO_2 gradually decreases. This indicates that a suitable increase in wind speed facilitates the diffusion of indoor NO_2 concentrations.

Tab. 3 Relationship between Wind Speed and Average NO_2 Content

Air velocity (m/s)	$No_2(\mu g/m^{-3})$
$\mathrm{W} \leq 1.0$	23.898
$1.0 < W \leq 3.0$	23.832
$3.0 < W \leq 5.0$	23.797
$W > 5.0$	20.864

The concentration of NO_2 is negatively correlated with humidity as shown in Table 4; as indoor humidity increases, the concentration of NO_2 decreases. Consequently, regulating indoor humidity levels can reduce nitrogen dioxide concentration to a certain extent and improve comfort.

Tab. 4 Relationship between Humidity and Average NO_2 Content

Humidity level (%)	$No_2(\mu g/m^{-3})$
$H \leq 20$	48.000
$20 < H \leq 50$	30.607
$50 < H \leq 80$	23.832
$H > 80$	23.816

5 Conclusion

This study first reviews research reports on air quality over the past three years and summarizes the current state of development both domestically and internationally. The paper introduces the concepts of BP neural networks, RNNs (recurrent neural networks), and LSTMs (long short-term memory networks), and discusses the corresponding indicators for assessing air quality. It then outlines the basic methods for designing predictive models. During the data collection phase, issues such as converting strings into date types recognizable by the model and handling missing data using forward chaining were addressed. Parameters such as wind speed, humidity, and temperature are used as inputs for predicting nitrogen dioxide (NO_2) concentrations. The final prediction results indicate that the predictions from the LSTM-optimized model align with actual data from the same day, demonstrating positive outcomes. Clearly, when predicting NO_2 concentrations, comparing the LSTM-optimized model with MLP and RNN models validates the suitability of LSTM for this study.

Although there have been improvements in air quality in recent years, the issue remains complex and challenging. Traditional data monitoring methods have proven inadequate for effectively processing the vast amounts of data required for accurate air quality predictions. As a result, the application of deep learning techniques for air quality forecasting is gradually gaining traction. Educational institutions are crucial environments for students, with dormitories serving as their homes away from home. Therefore, it is essential to enhance the prediction and analysis of NO_2 concentration levels in student dormitories to ensure that students can study and rest in a healthy environment.

The predictive models and optimization methods selected in this study have a significant impact on nitrogen dioxide concentration predictions. By actively implementing effective measures to optimize dormitory environments, a healthier and more comfortable learning and living environment can be created for students, thereby promoting their physical and mental health, enhancing learning efficiency, and laying a solid foundation for their comprehensive development. Additionally, this study focuses on using the LSTM optimization model to compare predictions with actual values on the same day to assess accuracy, which has certain limitations. Future research could build on existing findings or explore more precise algorithms and prediction models to expand the scope of predictions and conduct a more comprehensive evaluation of the model's effectiveness under different environmental conditions.

Acknowledgments

The support of the National Natural Science Foundation of China (Grant Nos. 12362023) is gratefully acknowledged.

Reference

[1] Roser, Max. "Data review: how many people die from air pollution? " Our world in data (2024).

[2] Ayturan, et al. (2018). Air Pollution Modelling with Deep Learning: A Review. 1. 58-62.

[3] Kai-Fa Lu, et al, Characterizing temporal and vertical distribution patterns of traffic-emitted pollutants near an elevated expressway in urban residential areas, Building and Environment, Volume 172,2020,106678, ISSN 0360-1323.

[4] Tang Zhixiang. Research and implementation of air quality prediction based on BP neural network [D]. Xi'an Electronic Science and Technology University,2018.

[5] N.E. Klepeis, et al, The national human activity pattern survey (NHAPS): a resource for assessing exposure to environmental pollution, Expo, Anal. Environ. Epidemiol. 11 (2001) 231–252.

[6] T. -H. Chan, et al, "PCANet: A Simple Deep Learning Baseline for Image Classification?" inIEEE Transactions on Image Processing, vol. 24, no. 12, pp. 5017-5032, Dec. 2015, doi: 10.1109/TIP.2015.2475625.

[7] A. -r. Mohamed, et al, "Deep Belief Networks using discriminative features for phone recognition,"2011 IEEE International Conference on Acoustics, Speech and Signal

Processing (ICASSP), Prague, Czech Republic, 2011, pp. 5060-5063, doi: 10.1109/ICASSP.2011.5947494.
[8] Hao, et al. (2016). Deep learning. International Journal of Semantic Computing, 10(03), 417–439.
[9] Chang, et al, An LSTM based aggregated model for air pollution forecasting, Atmos. Pollu. Res. 11 (2020) 1451–1463.
[10] X. Gao, W.D. Li, A graph-based LSTM model for PM2.5 forecasting, Atmos. Pollu. Res. 12 (2021) 101150.
[11] D. Seng, et al, Spatiotemporal prediction of air quality based on LSTM neural network, Alex. Eng. J. 60 (2021) 2021–2032.
[12] R. Yan, et al, Multi-hour and multi-site air quality index forecasting in Beijing using CNN, LSTM, CNN-LSTM, and spatiotemporal clustering, Expert Syst. Appl. 169 (2021) 114513.
[13] B. Zhang, et al, A novel Encoder-Decoder model based on read-first LSTM for air pollutant prediction, Sci. Total Environ. 765 (2021) 144507.
[14] Rui Zhang, et al Predicting the concentrations of VOCs in a controlled chamber and an occupied classroom via a deep learning approach, Building and Environment, Volume 207, Part B,2022,108525, ISSN 0360-1323.
[15] Li, et al, 2017b. Long short-term memory neural network for air pollutant concentration predictions: method development and evaluation.
[16] Xue, et al. District Heating Load Prediction Algorithm Based on Feature Fusion LSTM Model.Energies2019,12, 2122.
[17] Xing cheng Lu, et al, Development and application of a hybrid long-short term memory – three dimensional variational technique for the improvement of PM2.5 forecasting, Science of The Total Environment, Volume 770,2021,144221, ISSN 0048-9697.
[18] HAN Wei, WU Yan lan, REN Fu. Air pollutant prediction based on fully connected and LSTM neural networks[J]. Geographic Information World,2018,2503:34-40.
[19] Cheng liang Xu, et al, Improving prediction performance for indoor temperature in public buildings based on a novel deep learning method, Building and Environment, Volume 148,2019, Pages 128-135, ISSN 0360-1323.
[20] Hochreiter, S., Schmidhuber, J., 1997. Long short-term memory. Neural Comput. 9,1735e1780.
[21] F.BIANCOFIORE, M.BUSILACCHIO, M.VERDECCHIA, et al. Recursive neural network model for analysis and forecast of PM10 and PM2.5[J]. Atmospheric Pollution Research,2017,8(4):652~659.
[22] Krishan, M., Jha, S., Das, J.et al. Air quality modelling using long short-term memory (LSTM) over NCT-Delhi, India. Air Qual Atmos Health12, 899–908 (2019).
[23] Zheng Yangyang, et al. Application of Keras-based LSTM model for air quality index prediction[J]. Practice and Understanding of Mathematics, 2019(7):6. DOI:CNKI:SUN:SSJS.0.2019-07-017.
[24] Jie HUANG, et al. Hourly concentration prediction of PM2.5based on RNN-CNN ensemble deep learning model. Journal of Zhe Jiang University (Science Edition), 2019,

46(3): 370-379.
[25] Kristiani, E.; et al. Short-Term Prediction of PM2.5Using LSTM Deep Learning Methods.Sustainability2022,14, 2068.
[26] Bihter Das, Ömer Osman Dursun, Suat Toraman, Prediction of air pollutants for air quality using deep learning methods in a metropolitan city, Urban Climate, Volume 46,2022,101291, ISSN 2212-0955.
[27] Kim, J.; et al. Development of a Deep Learning-Based Prediction Model for Water Consumption at the Household Level.Water2022,14, 1512.
[28] Apeksha Aggarwal, Durga Toshniwal, A hybrid deep learning framework for urban air quality forecasting, Journal of Cleaner Production, Volume 329,2021,129660, ISSN 0959-6526.

[29] Freeman, et al, J. (2018). Forecasting air quality time series using deep learning. Journal of the Air & Waste Management Association,68(8), 866–886.
[30] Patricio Perez, et al, PM2.5 forecasting in Coyhaique, the most polluted city in the Americas, Urban Climate, Volume 32,2020,100608, ISSN 2212-0955.
[31] R. Janarthanan, et al, A deep learning approach for prediction of air quality index in a metropolitan city, Sustainable Cities and Society, Volume 67,2021,102720, ISSN 2210-6707.
[32] Al-Janabi, S., Mohammad, M. & Al-Sultan, A. A new method for prediction of air pollution based on intelligent computation. Soft Comput24, 661–680 (2020). https://doi.org/10.1007/s00500-019-04495-1.
[33] Bekkar, A., Hssina, B., Douzi, S.et al. Air-pollution prediction in smart city, deep learning approach. J Big Data8, 161 (2021). https://doi.org/10.1186/s40537-021-00548-1.
[34] Suh. Research on air quality prediction based on improved BP neural network[D]. Nanchang University,2020.
[35] FAN Junxiang, et al. Research on air pollution spatio-temporal forecasting model based on RNN[J]. Surveying and Mapping Science,2017,4207:76-83+120.
[36] Tello-Leal, et al. Evaluation of Deep Learning Models for Predicting the Concentration of Air Pollutants in Urban Environments.Sustainability2024,16, 7062. https://doi.org/10.3390/su16167062.
[37] YANG Li, et al. A review of recurrent neural network research[J]. Computer Applications,2018,38S2:1-6+26.
[38] Xie Chongbo. Research on urban air quality prediction based on recurrent neural network[D]. Southwest University of Science and Technology,2019.
[39] HAN Wei, et al. Air pollutant prediction based on fully connected and LSTM neural networks[J]. Geographic Information World,2018,2503:34-40.

Paper submitted: 02.06.2025
Paper revised: 25.07.2025
Paper accepted: 01.08.2025