Atmospheric Science and Air Quality, 2026, 1(1), 5-9
https://www.hillpublisher.com/Journals/asaq/
ISSN Online:

Research on Air Pollution Prediction Model Based on Convolutional Neural Networks

Jingyi Shen*, **Ye Zhu**, **Yuxin Yang**

North China University of Science and Technology, School of Science, Tangshan 063210, Hebei, China.

How to cite this paper: Jingyi Shen, Ye Zhu, Yuxin Yang. (2026). Research on Air Pollution Prediction Model Based on Convolutional Neural Networks. *Atmospheric Science and Air Quality*, 1(1), 5-9.
DOI: 10.26855/asaq.2026.06.002

Received: November 22, 2025
Accepted: December 16, 2025
Published: January 19, 2026

***Corresponding author:** Jingyi Shen, North China University of Science and Technology, School of Science, Tangshan 063210, Hebei, China.

Abstract

With the acceleration of industrialization and urbanization, air pollution has become one of the environmental problems to be solved in the world. Fine particulate matter (PM2.5), as a major air pollutant, has had a profound impact on human health, ecosystems and climate change. The purpose of this study is to explore the application potential of convolutional neural network (CNN), a deep learning technology, in PM2.5 concentration prediction, and to provide scientific basis and technical support for the prevention and control of air pollution by building an efficient and accurate prediction model. In the model construction stage, this study innovatively introduced convolutional neural network (CNN) into the field of air pollution prediction, and proposed a CNN prediction model combining time series and spatial data. The model uses the convolutional layer to automatically extract the local features and spatial dependence of the data, reduces the data dimension, reduces the computational complexity, and completes the prediction task of PM2.5 concentration through the fully connected layer.

Keywords

Air pollution prediction; Convolutional neural network (CNN); PM2.5 concentration; Deep learning

1. Introduction

1.1 Research Background

In the context of globalization, air pollution has become a transnational and far-reaching global environmental problem. With the acceleration of industrialization, energy consumption, and urbanization, pollutant emissions have increased significantly, leading to deteriorating air quality and seriously threatening human health, ecosystem balance, and the global climate. According to a World Health Organization (WHO) report, millions of people die each year from air pollution, most of whom are directly related to excessively high concentrations of pollutants such as PM2.5. Sources of pollution include not only anthropogenic activities such as industrial emissions and vehicle exhaust, but also natural phenomena such as dust storms and volcanic eruptions. Although natural factors exist, anthropogenic emissions have become the primary source.

Faced with this challenge, developed countries have effectively reduced pollutant emissions through stringent environmental regulations, advanced technologies, and comprehensive monitoring systems. For example, European countries have implemented strict vehicle emission standards, promoted clean energy, and established regional air quality monitoring networks. Developing countries, however, face greater difficulties in pollution control due to limitations in their economies, technologies, and experience. But with increasing global environmental awareness and strengthened international cooperation, many developing countries have begun to take measures such as formulating environmental regulations, strengthening monitoring and enforcement, and promoting clean energy. The

Note: The article was originally published in *Scientific Development Research*, and this is an authorized translation of it.

international community has also increased its technical support and financial assistance to these countries.

In conclusion, air pollution has become a pressing global issue. Different countries have adopted different governance strategies, but the common goal is to reduce pollution and improve air quality. Strengthening international cooperation and technology sharing is the future trend. This study aims to explore the application potential of convolutional neural networks (CNNs) in PM2.5 concentration prediction, providing a more accurate and effective solution for air pollution control.

1.2 Current Research Status

Significant progress has been made in air pollution prediction technology in recent years, but traditional methods still face many challenges. Statistical methods, such as time series analysis and multiple linear regression, while able to capture long-term trends in data, fall short when dealing with complex nonlinear relationships and non-stationary time series. Emerging machine learning algorithms, such as random forests and support vector machines, have improved prediction accuracy to some extent by automatically learning data features. However, these algorithms are performance-limited when handling large-scale, high-dimensional datasets with complex spatial relationships, making it difficult to fully capture the potential spatial and temporal dependencies in the data.

With the rapid development of deep learning technology, especially the widespread application of Convolutional Neural Networks (CNNs), new ideas have been provided for air pollution prediction. CNNs, with their unique convolution and pooling operations, have achieved great success in image processing, and their powerful feature extraction and pattern recognition capabilities have made it possible to process complex air pollution data. In recent years, some scholars have attempted to apply CNNs to air quality prediction and have achieved initial success, demonstrating their significant advantages in capturing the spatiotemporal variation characteristics of air pollutants. However, existing research mainly focuses on validation in single regions or datasets, and the generalization ability and real-time prediction capabilities of the models still need further improvement. Therefore, this study focuses on optimizing the CNN model structure to improve its prediction accuracy and robustness under multi-source heterogeneous data, aiming to provide a more scientific and efficient solution for air pollution prevention and control.

1.3 Research Objectives

This study aims to explore the potential and advantages of convolutional neural networks (CNNs) in the field of air pollution prediction, especially the accurate prediction of fine particulate matter (PM2.5) concentration. With the acceleration of urbanization and the improvement of industrialization, air pollution has become increasingly serious, posing a huge threat to public health, ecological environment and climate system (Meng, 2016). Therefore, developing a method that can efficiently and accurately predict air pollution concentration is of great significance for formulating scientific pollution prevention and control strategies, protecting public health and promoting sustainable development.

Specifically, this study aims to achieve the above objectives through the following efforts: First, to construct a CNN prediction model based on multi-source heterogeneous data to fully utilize various information such as meteorological, traffic, and industrial emissions, thereby improving the comprehensiveness and accuracy of predictions; second, to optimize the model structure and parameter settings, introducing advanced optimization algorithms and regularization techniques to enhance the model's generalization ability and robustness; third, to verify the model's prediction performance through extensive experiments and to compare and analyze it with traditional machine learning algorithms and other deep learning models to demonstrate the superiority of the proposed CNN model in terms of prediction accuracy and efficiency; finally, to apply the research results to practical air pollution prevention and control work, providing strong support for government decision-making, environmental monitoring, and public health.

2. Method

2.1 Data Preprocessing

2.1.1 Data source

This study used PM2.5 concentration data from a certain region over many years, as well as synchronous meteorological data (such as temperature, humidity, wind speed, etc.), which came from multiple monitoring stations and meteorological departments.

2.2.2 Data cleaning
To ensure data quality, missing values were imputed and outliers were handled. Missing values were imputed using interpolation or the average value method, while outliers were identified and removed using statistical methods.

2.2.3 Feature selection
Based on literature analysis and correlation analysis, features highly correlated with PM2.5 concentration were selected as input variables, including but not limited to temperature, humidity, wind speed, wind direction, and air pressure.

2.2 Model Construction

2.2.1 Selection of Convolutional Neural Network
The reason for choosing CNNs as the modeling tool is their ability to effectively handle complex relationships in spatial and temporal data, especially demonstrating outstanding performance in image recognition. Although air pollution data is not typical image data, CNNs can still be competent for such prediction tasks by constructing appropriate feature matrices and reasonable network structures.

2.2.2 Network structure
The CNN model designed in this study includes an input layer, several convolutional layers, pooling layers, fully connected layers, and an output layer. The convolutional layers use convolutional kernels of different sizes to extract multi-scale features; the pooling layers are used to reduce the feature dimensionality and computational cost; and the fully connected layers map the feature vectors output by the convolutional layers to the final prediction result.

2.2.3 Loss function and optimizer
Mean squared error (MSE) is used as the loss function to measure the difference between the predicted and actual values. The optimizer chosen is Adam, whose adaptive learning rate helps accelerate convergence and improve the model's training efficiency.

3. Experiment

3.1 Experimental Setup

3.1.1 Division of training set and test set
The dataset is divided into training and test sets in chronological order to ensure that the test set does not contain information leakage from the training set. The training set is used for model training and parameter tuning, and the test set is used to evaluate the generalization ability of the model (Jiao, 2024).

3.1.2 Hyperparameter settings
After multiple experiments and optimizations, suitable hyperparameters such as learning rate, batch size, and number of training epochs were determined. These parameters control the magnitude of weight adjustments during each model update and the number of samples used to calculate gradient updates during each training iteration, thus better combining momentum and adaptive learning rate to facilitate faster convergence.

3.2 Experimental Results

3.2.1 Model performance
The model performance was evaluated using metrics such as root mean square error (RMSE) and mean absolute error (MAE). Experimental results show that the CNN model constructed in this study outperforms traditional machine learning algorithms and some deep learning models in terms of prediction accuracy.

3.2.2 Modeling results
As shown in Figure 1.

3.2.3 Error analysis
As shown in Figure 2.

Figure 1. CNN model predictions for PM2.5.

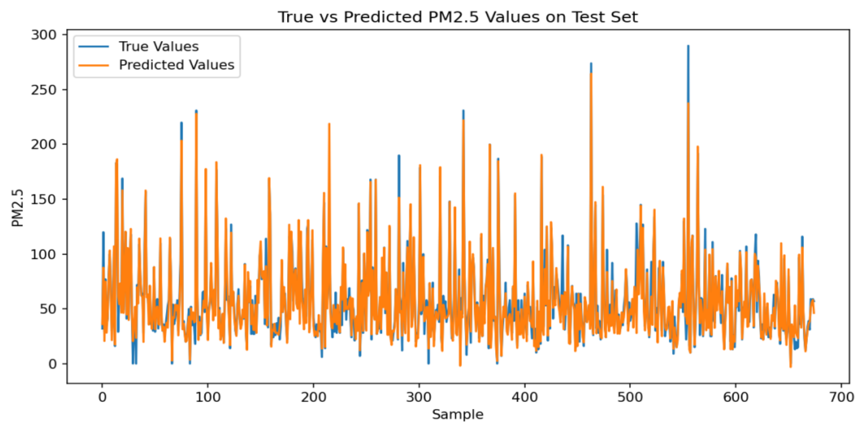

Figure 2. CNN model predictions for PM2.5.

The experimental results of this study demonstrate that CNN exhibits excellent performance in air pollution prediction. This is mainly due to the powerful feature extraction and complex pattern recognition capabilities of CNN, which enables it to extract effective features highly correlated with PM2.5 concentration from multidimensional meteorological data and make accurate predictions.

Although the CNN model constructed in this study has achieved a certain breakthrough in prediction accuracy, it still has some limitations. For example, the model's ability to predict extreme weather conditions needs further improvement; model training requires a large dataset and a long computation time.

Future research can be improved in the following aspects: first, optimize the network structure and introduce new technologies such as attention mechanisms; second, expand data sources and improve the diversity and completeness of data; and third, develop more efficient model training algorithms and computing platforms.

4. Conclusion

This study constructs an atmospheric pollution prediction model based on convolutional neural networks, which enables efficient prediction of air quality. Experimental results show that the model performs well in predicting PM2.5 concentration and other related pollutants, significantly outperforming traditional prediction methods (Chen & Xie, 2015). The following is a summary of the main contributions of this study and an outlook on future research directions.

(1) Excellent model performance: The CNN model constructed in this study can automatically extract key features from meteorological data and effectively capture the nonlinear relationship of atmospheric pollutant concentration changes, thus achieving high-precision prediction. The low RMSE and MAE values in the experimental results verify the effectiveness of the model.

(2) Data processing innovation: In the data preprocessing stage, this study adopted advanced data cleaning and feature selection methods, which effectively improved data quality and reduced the impact of redundant information on model performance. This laid a solid foundation for subsequent model training.

(3) Huge application potential: This study not only provides a new technical approach for air pollution prediction, but also demonstrates the broad application prospects of deep learning in the field of environmental science. In the future, this model can be further extended to the prediction of other air quality parameters and regions.

(4) Promoting environmental decision-making: High-precision air pollution forecast results can provide strong support for the government to formulate environmental protection policies, enterprises to optimize emission control and public health early warning, which helps to promote the continuous improvement of environmental quality.

5. Future Research Directions

(1) Multi-source data integration: Future research can explore how to combine more types of data sources (such as satellite remote sensing data, traffic flow information, etc.) to improve the prediction accuracy and generalization ability of the model.

(2) Model optimization and extension: Further optimize the architecture and hyperparameters of the CNN model, and combine it with more advanced deep learning techniques (such as residual networks, graph neural networks, etc.) to improve model performance. In addition, the model can also be considered for application to other air pollutant prediction tasks.

(3) Real-time prediction platform: Apply the model to the actual environment to build a real-time air pollution prediction system. This requires further optimization and compression of the model to ensure that it meets the requirements of real-time performance and computing resources.

(4) Model Interpretation and Interpretability: Although deep learning models perform well in prediction, their internal workings and decision-making processes are often difficult to understand. Future research could focus on improving the interpretability of deep learning models to better understand how they work and their prediction results.

(5) Interdisciplinary collaboration: Air pollution prediction involves knowledge and technology from multiple disciplines. Future research can strengthen interdisciplinary collaboration with disciplines such as environmental science, computer science, and statistics to jointly promote the development and application of air pollution prediction technology.

References

Chen, Q., & Xie, L. (2015). Discussion on the application of neural networks in air pollution prediction. *Information Technology and Informatization*, (6), 84-86.

Guo, Q. (2015). Air pollution prediction based on neural network. *Electronic Testing*, (18), 75-76. https://doi.org/10.16520/j.cnki.1000-8519.20151022.001

Huang, Q., Zhang, H., & Sun, J. (2024). Research on the Environmental Economic Policy System for Air Pollution Control. *Leather Manufacturing & Environmental Protection Technology*, *5*(15), 161-163.

Jiao, S. (2024). Research on Computer Information Management System Based on Artificial Intelligence Technology. *Communications World*, *31*(4), 190-192.

Meng, J. (2016). Research progress on the health effects of PM2.5 pollution. *Occupational Health*, *32*(20), 2873-2880.

Tang, X. (2018). An atmospheric pollution prediction model based on an improved BP neural network algorithm. *Journal of Henan University of Science and Technology (Natural Science Edition)*, *46*(1), 74-78.

Xiong, R. (2024). Construction of Highway Traffic Flow Prediction Model. *China Transportation Informatization*, (7), 106-108.

Xu, Q. (2020). Air Quality Assessment and Prediction Based on Humanoid Intelligent Optimization Algorithm (Master's thesis, North University of China).