Towards Smarter Cities: Multivariate Spatiotemporal Forecasting of Urban Air Pollution Using Hybrid Deep Graph Frameworks

Homayoon Zahmatkesh[1*], Rahim Ali Abbaspour[1], Abbas Abedini[1], Saeid Homayouni[2]

[1]School of Surveying and Geospatial Engineering, College of Engineering, University of Tehran, Tehran, Iran
[2]Institut National de la Recherche Scientifique (INRS), Centre Eau Terre Environnement, Quebec, Canada

Abstract

Urban air pollution forecasting is a critical component in the design of sustainable and smart cities, as it directly influences public health policy, transportation regulation, and environmental resilience. This paper presents a hybrid deep learning (DL) framework that integrates graph convolutional networks (GCNs) with long short-term memory (LSTM) units to perform multivariate spatiotemporal predictions of six major air pollutants: PM2.5, PM10, SO_2, NO_2, CO, and O_3. Unlike conventional univariate or grid-based time series models, our approach leverages the underlying topological structure of sensor networks and the temporal dynamics across multiple pollutant variables. A spatial graph is constructed based on the geographic locations of twelve monitoring stations in Beijing, enhanced with pollutant-specific correlations to encode both proximity and functional similarity among the sensors. The proposed GCN-LSTM architecture was trained on over 35,000 hourly observations per station. It demonstrates robust forecasting capabilities, achieving a root mean square error (RMSE) as low as 0.0316 and an R^2 value of up to 0.87 across various pollutants. Comparative experiments confirm the superiority of the hybrid model over baseline architectures, such as standalone LSTM and GRU models. This emphasizes the effectiveness of spatiotemporal graph representation in capturing urban air pollution dynamics. This framework provides a scalable and real-time solution for air quality management, offering a valuable tool for policymakers and urban planners engaged in smart environmental governance.

Keywords: Air Pollution Forecasting, Multivariate Time Series, Spatiotemporal Deep Learning, Smart City

1. Introduction

Air pollution remains a critical environmental and public health concern, particularly within densely populated urban centers.

In 2021, the European Environment Agency (EEA) reported that an estimated 253,000 premature deaths across Europe were attributable to prolonged exposure to fine particulate matter (PM2.5). Moreover, nitrogen dioxide (NO_2) and ground-level ozone (O_3) were linked to approximately 52,000 and 22,000 deaths, respectively (EEA 2023). These figures, which far exceed the exposure thresholds recommended by the World Health Organization (WHO), highlight the pressing need for precise and timely air quality forecasting systems to support risk reduction and inform decision-making in urban contexts. Beyond mortality, a substantial body of epidemiological literature has established strong associations between long-term exposure to ambient air pollutants and numerous serious health outcomes, including asthma, bronchitis, pneumonia, emphysema, and lung cancer (Sunyer, Jarvis et al. 2006, Sarnat and Holguin 2007, Kravchenko, Akushevich et al. 2014, Buonanno, Stabile et al. 2017). These risks are especially pronounced among vulnerable groups such as children, the elderly, and individuals with pre-existing medical conditions. This emphasizes the importance of localized and proactive forecasting frameworks.

In the rapidly evolving landscape of smart cities, the ability to forecast air pollutant concentrations with high spatial and temporal resolution is essential for strengthening environmental resilience, enabling real-time alert systems, and supporting evidence-based urban policymaking. Nevertheless, accurately forecasting air pollution presents significant technical challenges due to the inherently nonlinear, nonstationary, and dynamic characteristics of pollutant dispersion. Urban air quality is shaped by a complex interaction of diverse factors, including traffic density, meteorological conditions, industrial emissions, and urban morphology, all of which exhibit variability across both space and time.

Traditional models, such as ARIMA (Kumar and Jain, 2010) and SVR (Sánchez et al., 2011), were among the earliest approaches for air quality forecasting, but they often struggle with rigid assumptions and a limited ability to capture nonlinear or multiscale patterns. These limitations have led to a growing interest in DL approaches, which offer greater flexibility for modeling complex environmental dynamics. GCNs have emerged as powerful tools for capturing spatial dependencies in non-Euclidean domains such as sensor networks. GCNs represent monitoring stations as graph nodes and define edges using spatial distances or pollutant correlations (Chen et al., 2023; Han et al., 2025; Yang et al., 2025). In parallel, RNNs, especially LSTM (Li et al., 2025) and GRU (Saikumar et al., 2025) architectures, are widely used for learning temporal patterns in time series, leveraging memory gates for effective sequence modeling. Recent studies have demonstrated the effectiveness of hybrid spatiotemporal architectures in predicting urban air quality. For instance, Qi, Li et al. in 2019) introduced a GC-LSTM model that achieved strong performance (R^2 = 0.72) for 72-hour PM2.5 forecasting by integrating spatial graph structures with temporal LSTM networks.

The integration of GCNs for spatial representation and LSTM/GRU networks for temporal modeling has recently emerged as a powerful paradigm for capturing the spatiotemporal structure of environmental data (Abbasi, Alesheikh et al. 2025, Kumar, Kour et al. 2025, Rad, Nematollahi et al. 2025, Ranjan and Singh 2025). Despite these advances, most existing frameworks prioritize single-pollutant

* Corresponding author

forecasting or treat pollutants independently, limiting their ability to model the complex chemical and atmospheric interactions that occur simultaneously in urban environments (Le, Bui et al. 2020, Jana, Middya et al. 2024).

While several previous studies have employed such hybrid architectures for predicting individual pollutants, many fail to consider the interdependencies among multiple coexisting pollutants in real-world urban settings.

To bridge this gap, the present study proposes a multivariate spatiotemporal DL framework—GCN-LSTM— designed for simultaneous forecasting of six major air pollutants: PM2.5, PM10, SO_2, NO_2, CO, and O_3. The model utilizes hourly data collected from twelve fixed monitoring stations in Beijing spanning the years 2013 to 2017. In contrast to conventional approaches that typically focus on single-target forecasting (de Souza, Soares et al. 2025, Lee, Lee et al. 2025, Yang, Kai et al. 2025). This framework formulates the task as a multivariate prediction problem, thereby capturing cross-pollutant interactions and improving predictive performance.

This study presents a unified DL framework combining GCNs for spatial modeling and LSTM networks for temporal modeling to support real-time air quality monitoring and early warnings. Unlike univariate models, it jointly predicts multiple pollutants, capturing their spatiotemporal interdependence. Applied to a five-year Beijing dataset, the framework achieves state-of-the-art performance and offers a scalable, interpretable solution for smart urban governance and environmental policy.

2. Methodology

This section outlines the overall modeling pipeline, including the structure of the input data, the construction of the spatial graph, and the architecture of the proposed GCN-LSTM framework. The objective is to capture both the spatial dependencies among air quality monitoring stations and the temporal dynamics across multiple pollutants to achieve accurate multivariate forecasting.

2.1 Data Description

The experimental analysis utilizes hourly air quality data from 12 monitoring stations in Beijing, spanning from March 1, 2013, to February 28, 2018, with 35,064 hourly records per station. Each record includes simultaneous readings of PM2.5, PM10, SO_2, NO_2, CO, and O_3 (Chen, 2017). To capture spatial dependencies, a sensor graph is constructed where nodes represent stations and edges reflect Pearson correlation strength (Figure 1) (Lingye et al., 2024). Unlike fully connected graphs (Chen et al., 2024), only significant correlations are preserved. Some nearby stations remain disconnected due to terrain barriers, highlighting the importance of geospatial and environmental context.

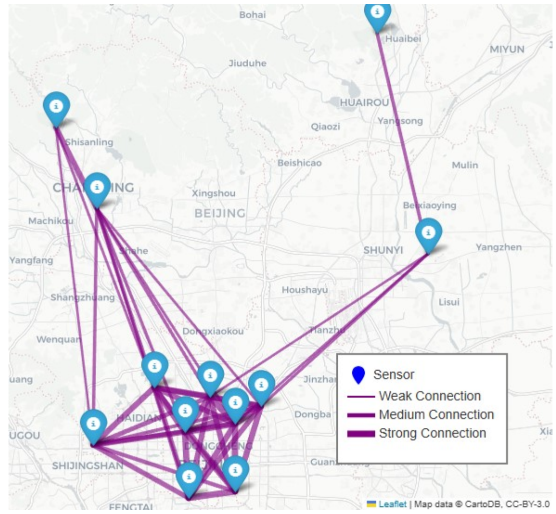

Figure 1. Location and the constructed correlation-based graph of air pollution sensors in Beijing

2.2 Exploratory Spatiotemporal Correlation Analysis

To gain a deeper understanding of the spatial and inter-variable dependencies present in the dataset, we conducted a comprehensive exploratory correlation analysis across both the sensor network and the recorded pollutant types. Two key visualizations were utilized to support this investigation.

2.2.1 Inter-Station Correlation of PM2.5: Using hourly PM2.5 concentrations recorded at 12 monitoring stations in Beijing from March 2013 to February 2018, we calculated the pairwise Pearson correlation coefficients to quantify spatial relationships. As illustrated in Figure 2, most stations exhibited high intercorrelation values (greater than 0.85), indicating strong spatial dependencies in PM2.5 levels across the urban sensor network. The highest correlations were observed between geographically proximate or environmentally similar stations, such as Guanyuan and Wanliu (r=0.96). This finding reinforces the appropriateness of graph-based models for capturing spatial dynamics in air pollution forecasting.

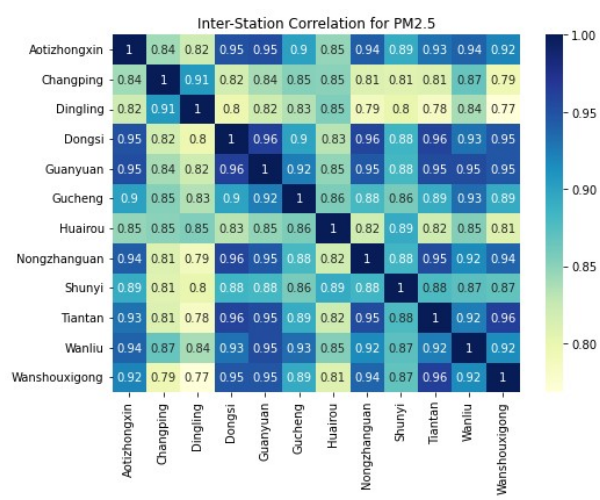

Figure 2. Inter-station Correlation of PM2.5

https://doi.org/10.5194/isprs-annals-X-4-W8-2025-835-2026

2.2.2 Intra-Station Pollutant Correlation (Aotizhongxin): In addition to analyzing spatial patterns, we examined the interdependence of pollutants at a representative station (Aotizhongxin). As illustrated in Figure 3, PM2.5 and PM10 demonstrated a strong positive correlation (r = 0.88), indicating their shared sources and chemical similarities. Notably, carbon monoxide (CO) also exhibited a significant correlation with PM2.5 and nitrogen dioxide (NO_2), suggesting common emission sources, such as vehicular traffic. In contrast, ozone (O_3) displays a negative correlation with other pollutants, particularly NO_2 and CO, which is consistent with established atmospheric chemical interactions and photochemical behavior.
These insights informed the design of our spatiotemporal DL architecture, particularly in defining graph edge weights and integrating pollutant co-dependencies in the multivariate modeling approach.

2.3 Spatiotemporal Forecasting

To effectively model the complex dynamics of urban air pollution, we explored and compared several DL architectures: GCN, LSTM, GRU, and a hybrid GCN-LSTM model. Each architecture offers distinct strengths in capturing spatial and temporal patterns.

2.3.1 GCN for Spatial Learning: GCNs were well-suited for capturing spatial dependencies between monitoring stations by modeling them as nodes within a graph structure. The connections between stations were derived using a combination of physical proximity and pollutant correlation. The GCN layers allowed each station to learn from its neighbors, providing a more comprehensive understanding of the spatial distribution of pollutants across the city.

2.3.2 LSTM and GRU for Temporal Modeling: To capture temporal patterns, LSTM and GRU networks were employed. These are specifically designed to handle sequential data and can effectively learn long-term dependencies, making them suitable for forecasting hourly pollution levels. While LSTM offers intricate control through multiple gates, GRU features a more streamlined structure that allows for faster training.

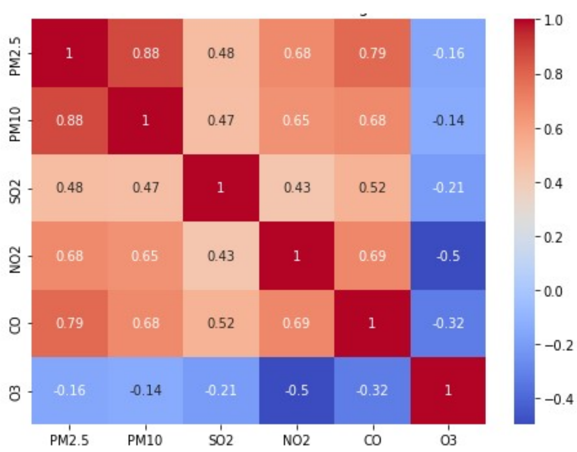

Figure 3. Intra-Station Pollutant Correlation

2.3.3 GCN-LSTM Hybrid for spatiotemporal modeling: Our primary model, the GCN-LSTM, synergistically combines the strengths of both methodologies. At each time step, the GCN extracts spatial features from all monitoring stations. These features are subsequently fed into an LSTM network, which captures their temporal evolution. This integration allows the model to effectively learn how pollution disperses across both spatial and temporal dimensions, thereby enhancing prediction accuracy for multiple pollutants concurrently.

3. Proposed Architecture

This study presents a multivariate spatiotemporal forecasting framework that utilizes GCNs in combination with recurrent architectures—LSTM and Gated Recurrent Unit (GRU)—to predict air pollutant concentrations across a network of monitoring stations. The architecture is specifically designed to simultaneously capture the spatial dependencies among sensors and the temporal evolution of pollutant levels.
The input data consist of hourly records for key air pollutants (PM2.5, PM10, SO_2, NO_2, CO, and O_3) collected from 12 stations in Beijing over the period from 2013 to 2018. These records are organized into a spatiotemporal tensor, where each node corresponds to a station, and each feature vector represents the multivariate pollutant readings at a specific time step.
To model spatial relationships, we construct a graph that incorporates both physical proximity and inter-station pollutant correlations, as determined by the Pearson correlation coefficient. This process results in an adjacency matrix that informs the GCN layer. The GCN learns spatial representations by aggregating feature information from neighboring stations at each time step.
These spatial features are subsequently fed into a temporal modeling layer, which can be implemented using either an LSTM or GRU architecture. Both variants are assessed to compare their effectiveness in capturing temporal dependencies and managing multivariate pollutant sequences. Following this, dropout regularization is applied, and two fully connected layers transform the temporal embeddings into the output space, producing simultaneous forecasts for all pollutants across all monitoring stations.
The model is optimized using the mean squared error (MSE) loss calculated across all output dimensions. A comparative analysis of the LSTM and GRU variants highlights the trade-offs between them.

https://doi.org/10.5194/isprs-annals-X-4-W8-2025-835-2026 | © Author(s) 2026. CC BY 4.0 License.

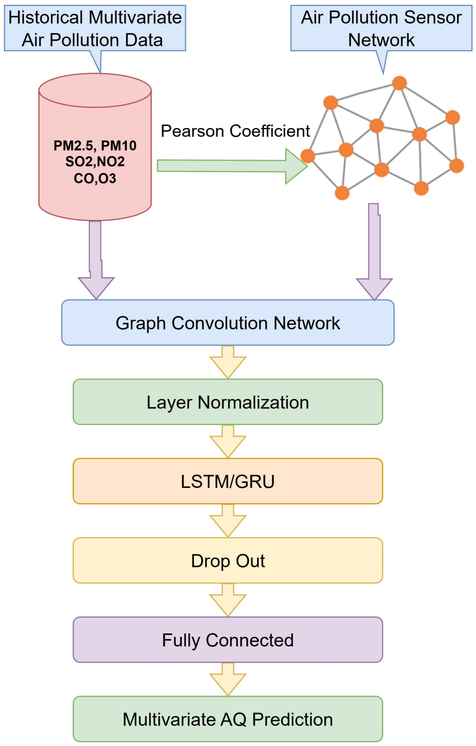

Figure 4. Proposed Architecture

4. Experiment and results

To evaluate the performance of the proposed GCN-LSTM framework, we conducted extensive experiments using a dataset collected from twelve air quality monitoring stations in Beijing between March 2013 and February 2018. The target variables included six key pollutants: PM2.5, PM10, SO2, NO2, CO, and O3. The model was trained on hourly records, comprising 35,064 samples per pollutant, and the results were compared with those of a GCN-GRU baseline to assess effectiveness.

4.1 Convergence Analysis

Figure 4 illustrates the training and validation loss curves over 50 epochs for both the GCN+GRU and GCN+LSTM configurations. While both models eventually converge, the GCN-LSTM demonstrates a smoother loss trajectory and consistently lower validation error, indicating superior generalization performance. Importantly, the close alignment between training and validation loss in the GCN-LSTM model suggests minimal overfitting. In contrast, models that experience overfitting typically exhibit an increasing gap between training and validation curves as training progresses. The GCN-LSTM's ability to maintain a narrow loss gap throughout training confirms its robustness and stability, making it well-suited for real-world forecasting tasks where unseen data may vary in distribution.

4.2 Accuracy of Multi-Pollutant Forecasting

To further evaluate the model, we plotted the predicted versus actual normalized values for each pollutant at the Aotizhongxin station, which is central and representative of urban pollution patterns. As illustrated in Figure 7, the GCN-LSTM model accurately captures both the magnitude and variability of pollutant trends, demonstrating strong alignment across PM2.5, PM10, and NO2. Although slight deviations are observed in highly dynamic pollutants such as O3 and SO2, the overall temporal patterns are well reproduced.

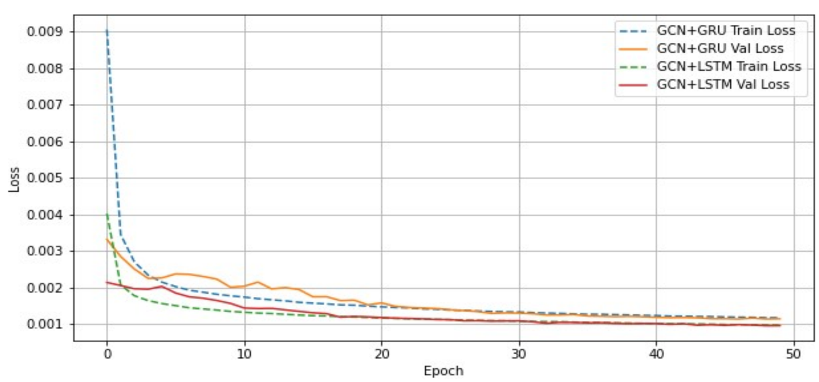

Figure 5. Convergence Comparison — GCN+GRU vs GCN+LSTM (Train and Validation Loss)

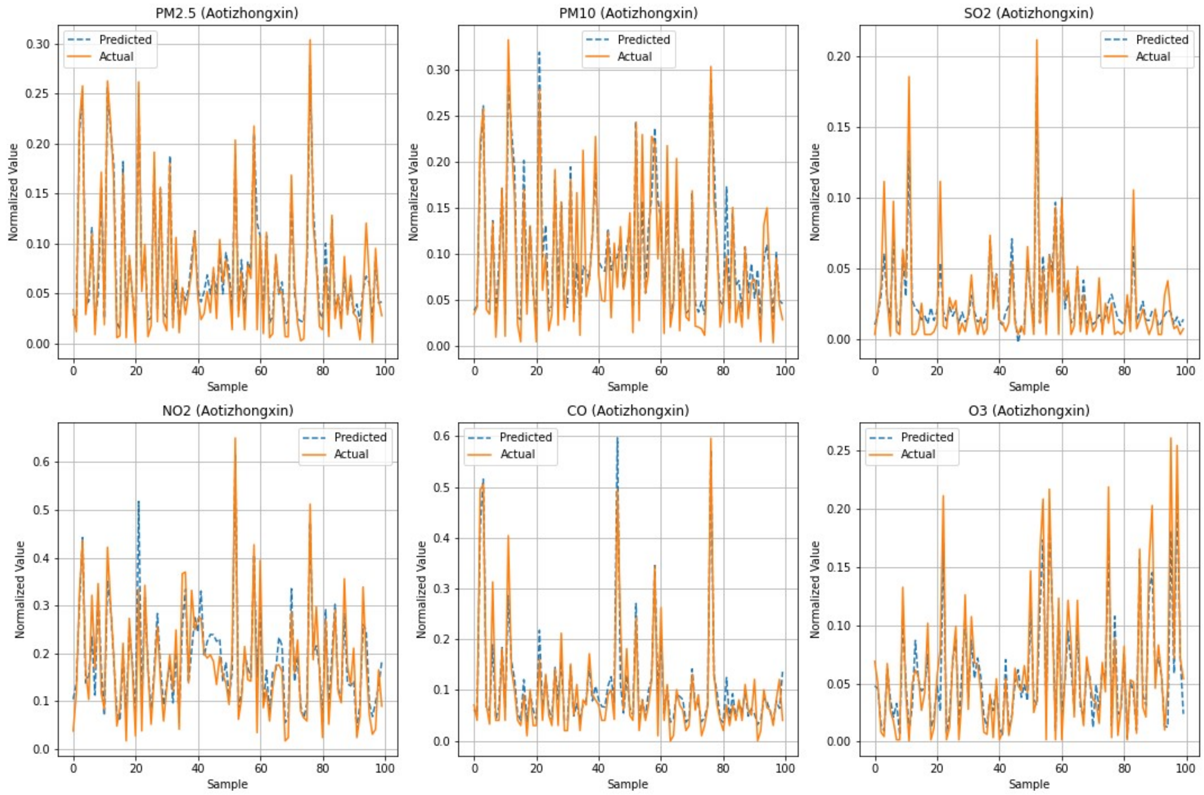

Figure 6. Predicted vs Actual Concentration (First 100 Samples at Aotizhongxin Station)

4.3 Quantitative Evaluation

The final performance was evaluated using the RMSE, MAE, and R^2 for models. The results confirm that the proposed GCN-LSTM framework consistently outperforms the GRU-based baseline across all six pollutants. In particular, the model achieves lower RMSE and MAE, indicating improved accuracy in predicting both short-term fluctuations and overall trends. Notably, the R^2 values exceed 0.87 for $PM_{2.5}$, demonstrating that the GCN-LSTM can capture a substantial proportion of the variance in actual pollutant concentrations. This underscores the model's robust generalization capability, particularly for pollutants that display complex and non-linear temporal behaviors, such as NO_2 and CO. The minimal difference between RMSE and MAE further indicates low error variance and a stable prediction profile. In contrast, the GCN-GRU demonstrates lower R^2 values and higher error metrics across pollutants, while still effective, reflecting its comparatively diminished capacity to learn long-range temporal dependencies.

MODEL	RMSE	MAE	R^2
GCN + GRU	96.56%	0.0202	0.85
GCN + LSTM	96.84%	0.0199	0.87

Table 1- Quantitative Evaluation between Configurations

5. Discussion

The experimental results confirm the effectiveness of the GCN-LSTM model in accurately forecasting multiple urban air pollutants. It consistently outperforms the GCN-GRU baseline, demonstrating lower RMSE and MAE values, as well as higher R^2 scores across all six pollutants. This enhanced performance stems from the ability of GCNs to capture spatial dependencies among monitoring stations, while LSTM networks effectively model temporal dynamics. Unlike previous univariate or less integrative models, this framework accounts for both pollutant co-occurrence and spatial propagation. Strong correlations, such as those between $PM_{2.5}$ and PM10, contribute to higher accuracy, whereas the model's performance on volatile pollutants like O_3 suggests that incorporating meteorological variables could enhance forecasting capabilities. The model shows promise for integration into smart city initiatives, particularly in light of increasing urbanization and the deployment of Internet of Things (IoT) sensors. However, the reliance on a static graph may limit its adaptability to changing emission patterns. Additionally, its generalizability beyond Beijing has yet to be evaluated. Future research should explore the use of dynamic graphs, multimodal data (e.g., weather and traffic information), and explainable AI techniques to improve performance, interpretability, and relevance to policymaking.

6. Conclusion

This study introduces a novel multivariate spatiotemporal DL framework that combines GCNs and LSTM networks for forecasting urban air quality. By jointly modeling spatial correlations across sensor stations and temporal dependencies within and between multiple pollutants, the proposed GCN-LSTM architecture outperforms traditional and baseline models across several evaluation metrics.

Applied to a comprehensive five-year dataset from Beijing, the framework delivers highly accurate predictions for six major pollutants and demonstrates robust generalization. These results validate the utility of integrating graph-based spatial modeling with sequence learning for environmental monitoring in complex urban environments.
Beyond its technical merits, the framework supports smart city initiatives through real-time alerts and data-informed planning. Future improvements, such as dynamic graphs, multimodal inputs, and broader urban deployment, could enhance performance and enable more adaptive air quality management.

References

Abbasi, M., A. Alesheikh, A. Lotfata, Z. Azizi, 2025: "Hybrid graph convolutional networks for air quality prediction: a systematic review of foundations, challenges, and opportunities." International Journal of Environmental Science and Technology: 1-36.

Buonanno, G., L. Stabile, L. Morawska, G. Giovinco, X. Querol, 2017: "Do air quality targets really represent safe limits for lung cancer risk?" Science of the Total Environment **580**: 74-82.

de Souza, A., D. G. Soares, E. Silva de Medeiros, J. F. de Oliveira-Júnior, D. Ozonur, M. C. Abreu, 2025: "Advancing Ozone Forecasting with LSTM Neural Networks: A Study on Time Window Variations and Performance Metrics." Ozone: Science & Engineering: 1-8.

EEA, 2023: Harm to human health from air pollution in Europe: burden of disease 2023, European Environmental Agency.

Jana, S., A. I. Middya, S. Roy, 2024: "Short-term air pollution prediction using graph convolutional neural networks." Technological Forecasting and Social Change **208**: 123684.

Kravchenko, J., I. Akushevich, A. P. Abernethy, S. Holman, W. G. Ross Jr, H. K. Lyerly, 2014: "Long-term dynamics of death rates of emphysema, asthma, and pneumonia and improving air quality." International Journal of Chronic Obstructive Pulmonary Disease: 613-627.

Kumar, S., V. Kour, A. Raj, T. Tapung, S. Mishra, R. Misra, T. Singh, 2025: "Optimizing Air Pollution Forecasting Models Through Knowledge Distillation: A Novel GCN and TRANS_GRU Methodology for Indian Cities." IEEE Access.

Le, V.-D., T.-C. Bui, S.-K. Cha, 2020: Spatiotemporal deep learning model for citywide air pollution interpolation and prediction. 2020 IEEE international conference on big data and smart computing (BigComp), IEEE.

Lee, C.-Y., J.-Y. Lee, S.-H. Han, J.-G. Kang, J.-B. Lee, D.-R. Choi, 2025: "Performance Evaluation of PM2. 5 Forecasting Using SARIMAX and LSTM in the Korean Peninsula." Atmosphere **16**(5): 524.

Qi, Y., Q. Li, H. Karimian, D. Liu, 2019: "A hybrid model for spatiotemporal forecasting of PM2.5 based on graph convolutional neural network and long short-term memory." Science of the Total Environment **664**: 1-10.

Rad, A. K., M. J. Nematollahi, A. Pak, M. Mahmoudi, 2025: "Predictive modeling of air quality in the Tehran megacity via deep learning techniques." Scientific Reports **15**(1): 1367.

Ranjan, S., S. K. Singh, 2025: "A deep dive into delhi's air pollution: forecasting PM 2.5 levels using a Bi-LSTM-GRU hybrid model." Earth Science Informatics **18**(2): 201.

Sarnat, J. A., F. Holguin, 2007: "Asthma and air quality." Current opinion in pulmonary medicine **13**(1): 63-66.

Sunyer, J., D. Jarvis, T. Gotschi, R. Garcia-Esteban, B. Jacquemin, I. Aguilera, U. Ackerman, R. de Marco, B. Forsberg, T. Gislason, 2006: "Chronic bronchitis and urban air pollution in an international study." Occupational and environmental medicine **63**(12): 836-843.

Yang, C., Z. Kai, X. Wang, T. Hu, H. Liu, 2025: "An optimized TCN–LSTM model for predicting PM 2.5 in metro systems." Environmental Science: Processes & Impacts.

This contribution has been peer-reviewed. The double-blind peer-review was conducted on the basis of the full paper.
https://doi.org/10.5194/isprs-annals-X-4-W8-2025-835-2026