Theoretical and Applied Climatology (2025) 156:464
https://doi.org/10.1007/s00704-025-05715-5

RESEARCH

Effective air pollution prediction using wavenet deep learning with Xgboost (1DCNN-BiLSTM-XgRC) for urban US embassies

Naushad Ahmad[1] · **Vipin Kumar**[1]

Received: 28 April 2025 / Accepted: 5 August 2025

Abstract

The key contributors to climate change include air pollution and atmospheric radiation. $PM_{2.5}$ is especially harmful among various pollutants, posing significant risks to human health and the environment. The reality highlights the pressing need for accurate forecasting models to address this challenge effectively. In this study, we proposed a hybrid Pro-1: Proposed 1DCNN-BiLSTM model, where a wavenet architecture combined with two 1DCNN branches feeds into a BiLSTM layer for $PM_{2.5}$ prediction. Furthermore, introduce an enhanced version, the Proposed 1DCNN-BiLSTM-XGBoost residual correction model, which incorporates XGBoost-based residual correction to optimize prediction accuracy. The proposed models are compared to traditional deep learning models like BiLSTM, CNN, GRU, LSTM, and RNN. The Pro-1 model demonstrates a noteworthy improvement in RMSE compared to the CNN model, recording a value of 79.909 ± 19.760 across the dataset. It achieves a minimum RMSE of 3.463 in the Colombo dataset, while the Pro-2 model has an RMSE of 0.2658 in the same dataset. The Pro-1 model's maximum RMSE reaches 23.778 in the Ulaanbaatar dataset, whereas Pro-2 peaks at 2.554. Pro-2 consistently outperforms all models, achieving the lowest RMSE, MAE, and MSE across all cities, as demonstrated by Friedman's post hoc test, which ranked 1st with statistically significant improvements. The model ranking was $1DCNN - BiLSTM - XGBoost > 1DCNN - BiLSTM >$ BiLSTM $>$ GRU $>$ LSTM $>$ RNN $>$ CNN to the RMSE. Further, the Diebold-Mariano test, AIC-BIC test, and Taylor diagrams will be used to validate the proposed models. These results establish the proposed models as the most effective predictive model for $PM_{2.5}$ forecasting, offering improved accuracy and reliability for air pollution monitoring.

1 Introduction

Air pollution is hazardous to human health. The quality of the air decreases day by day. It is one of the most common causes of environmental issues, leading to droughts, heat waves, floods, and forest fires. Climate change looks like a natural phenomenon, but humans cause the truth behind it Suthar and Singh (2025). Climate change has long-term and short-term effects on humans. High levels of pollutant gases are one of the causes of modifying the natural characteristics of the atmosphere. Air pollution is a critical issue that demands immediate attention, as it significantly contributes to climate change and adversely affects human health (Saxena 2025). Both pollutant and meteorological parameters alter the climate, making it essential to monitor and manage these factors effectively (Singh and Suthar 2025). Exceeding pollutant concentrations like particulate matter, carbon dioxide, nitrogen oxide, sulfur dioxide, and carbon monoxide affect humans and the atmosphere (Arooj et al. 2025). Furthermore, exceeding meteorological parameters like temperature, humidity, solar radiation, and wind speed are the most common causes of the effects of weather conditions, drought, and human health (Duarte et al. 2025). Understanding the dynamics of air quality changes requires considering both meteorological and socio-economic variables, as demonstrated in recent studies conducted in Hohhot City, China (Guo et al. 2024c). Particulate matter consists of a mixture of gas molecules, dust particles, and water droplets, influenced by solar radiation, complicating our efforts to combat climate change. Air pollution $PM_{2.5}$ is particulate matter 2.5 micrometers in size. Among the various pollutants, particulate matter $PM_{2.5}$ and PM_{10}

✉ Vipin Kumar
rt.vipink@gmail.com

Naushad Ahmad
naushad13bhu@gmail.com

[1] Computer Science and Information Technology, Mahatma Gandhi Central University, Zila School, 845401 Motihari, Bihar, India

Published online: 15 August 2025

are particularly harmful, causing short-term and long-term health issues such as respiratory diseases, cardiovascular problems, and premature death (Zhang et al. 2025; Amin et al. 2025). Sustainable development goals and agreements like those established in the parties' conferences emphasize reducing air pollution levels to meet World Health Organization or national standards (Rana et al. 2024). The escalating levels of air pollution demand a multifaceted approach that integrates technological advancements. AI-driven predictive models and real-time monitoring systems can enhance early warning mechanisms, enabling communities to reduce exposure to hazardous air pollutants proactively. Figure 1 has divided air pollution and climate change into the causes of industrialization and urbanization in the increased pollution (Saxena 2025). Exceeded pollutant concentrations can directly and indirectly affect human health and climate change (Xu et al. 2025).

Before the advancement of artificial intelligence, air quality forecasting was predominantly conducted using traditional statistical models such as ARIMA (Belmahdi et al. 2020), SARIMA (Gocheva-Ilieva and Ivanov 2019), GARCH (Chiroma et al. 2016), AR, and MA. These models were adequate but limited in capturing complex nonlinear patterns and long-term dependencies in air pollution data (Bekhor et al. 2013). With the emergence of machine learning techniques, models such as SVM, LR, DT, XGBoost, and RF have been increasingly employed to predict air quality levels and pollutant concentrations (Pande et al. 2024), including particulate matter $PM_{2.5}$ and gaseous pollutants such as NO2, SO2, O3, and CO (Essamlali et al. 2024). Machine learning models are applied by Salehie et al. (2024) to intelligent air quality monitoring systems that provide real-time pollution alerts and personalized recommendations, encouraging individuals to adopt eco-friendly habits like reducing vehicle dependency, optimizing energy consumption, and supporting green initiatives (Satpathy et al. 2025).

Utilizes deep learning models (Ansari and Quaff 2025), including LSTM, GRU, BiLSTM, CNN, and RNN, has further enhanced the predictive capabilities of air pollution models by capturing spatiotemporal dependencies and complex interactions within atmospheric data (Samal

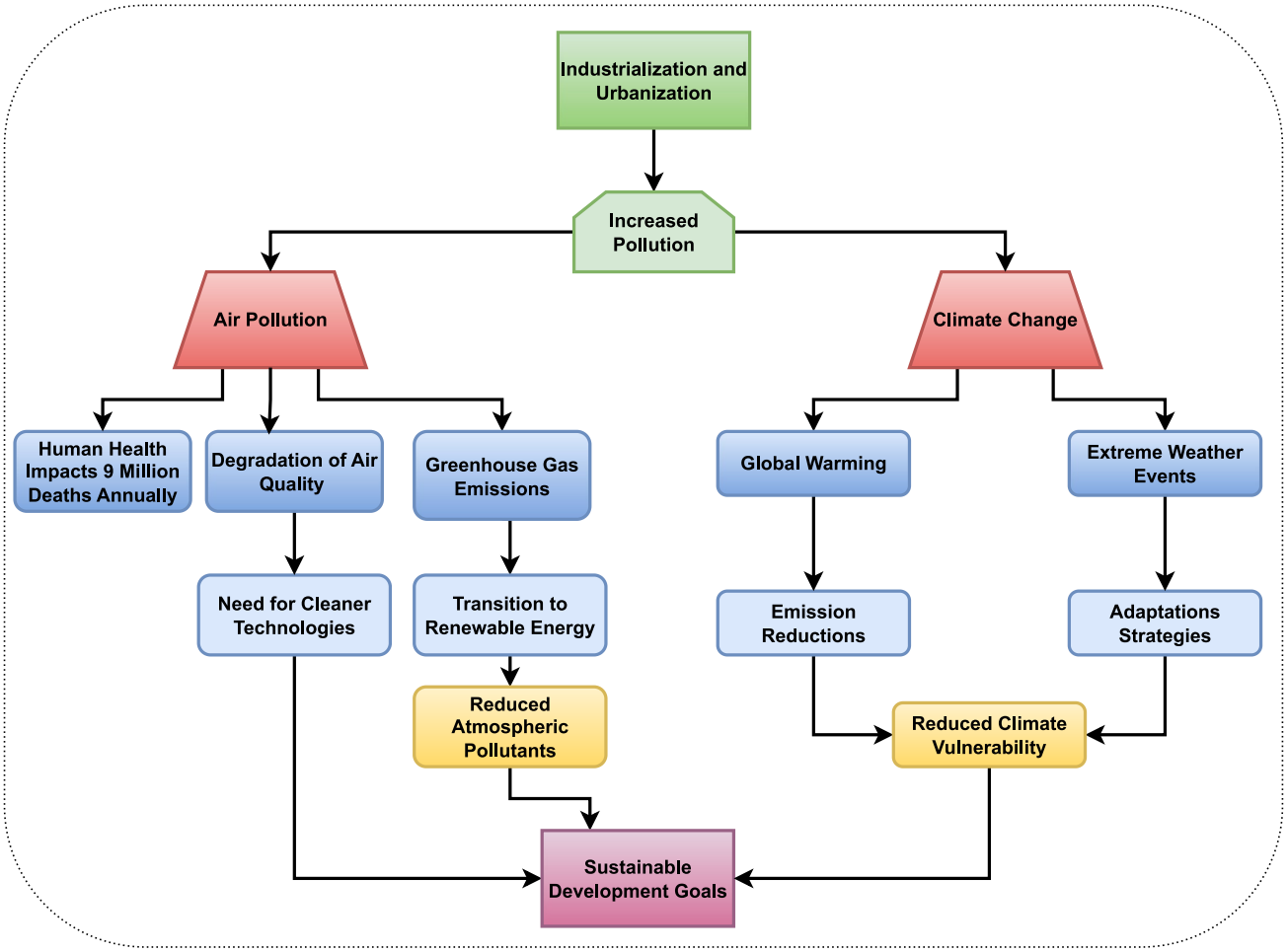


Fig. 1 Impact of Industrialization and Pollution on Air Quality, Climate Change, and Sustainable Developments

et al. 2022). Air pollution prediction using traditional deep learning models employs powerful techniques such as feedforward neural networks, recurrent neural networks, and long-term memory networks (Zaini et al. 2022). These advanced models analyze historical data on pollutant concentrations, meteorological conditions, and other influential factors (Zhang et al. 2022). By capturing complex nonlinear relationships and temporal dependencies within the data, they enable highly accurate forecasts of pollution levels. Training these models requires a carefully curated dataset to achieve optimal performance, often involving crucial preprocessing steps like normalization, feature selection, and data augmentation (Jairi et al. 2024). While these models hold significant potential, their effectiveness can be limited by the availability of comprehensive data and the need for diligent hyperparameter tuning. These practical methods can significantly enhance our capacity to predict and address air pollution.

Hybrid models, which integrate traditional statistical models with machine learning and deep learning approaches, have been developed to leverage the strengths of each methodology. These models aim to improve prediction accuracy by combining statistical rigor with the flexibility and adaptability of AI-driven techniques (Zong and Guan 2024). Hybrid and deep learning models, including ANN, have also been successfully applied to predict monthly average and extreme atmospheric parameters, such as temperature, highlighting their robustness and generalizability in environmental modeling tasks (Guo et al. 2023a). Hybrid deep learning models for air pollution prediction represent a notable advancement in forecasting accuracy and reliability (Yang et al. 2020). By integrating multiple algorithms, these models exploit CNNs for spatial feature extraction, combined with RNNs or LSTMs to capture essential temporal dependencies. The potent combination facilitates the analysis of complex data from various sources, including meteorological factors, traffic patterns, and historical pollution records (Ulpiani et al. 2021). The fusion of convolutional and recurrent neural networks, such as CNN-LSTM hybrids, is effective for monthly climate prediction tasks by capturing spatial and temporal dependencies simultaneously (Guo et al. 2024a). Furthermore, researchers successfully address overfitting issues and enhance model generalization through innovative ensemble techniques and the strategic stacking of diverse model architectures.

In Fig. 2, air pollution modelings are designed by the physical process-based method and data-driven time series models. The physical process-based methods are very old in predicting air quality using numerical techniques. Traditional machine learning models like MLP, DT, RF, SVR, and Xgboost are used to predict air pollution concentrations, whereas deep learning models like LSTM, GRU, CNN, BiLSTM, and RNN are also used. The combination of traditional machine learning and deep learning models performed well in most cases. The wavenet architecture uses the model so that the model's performance should increase, similar to ensemble techniques.

1.1 Research problem and objective of research

The research problem focuses on the growing concern of air pollution and its detrimental impact on public health and the environment, compounded by the challenges of accurate and timely prediction using traditional models.

- **Challenges in accurate air quality prediction :** Traditional models struggle to capture complex spatial-temporal dependencies, leading to unreliable forecasts.
- **Limitations in error correction and computational efficiency :**Existing approaches lack effective residual learning mechanisms and face trade-offs between accuracy and computational cost.
- **Develop a hybrid model :** Integrate WaveNet, CNN, BiLSTM, and XGBoost to enhance predictive accuracy and robustness while improving the model's ability to extract meaningful patterns from air pollution data.
- **Optimize computational efficiency :** Balance model complexity with performance to ensure real-world applicability.

The research aims to develop and evaluate hybrid deep-learning models that can effectively predict air pollution levels by integrating various data sources and methodologies. Specifically, the study seeks to enhance prediction accuracy, account for temporal and spatial dynamics, and identify key influencing factors. The proposed approach is a sophisticated and effective way to use the strengths of WaveNet (CNN, BiLSTM). See Table 1 for the definition of the abbreviation and XGBoost for predictive modeling. By integrating these components, you create a hybrid model that excels at capturing both temporal and spatial patterns while correcting errors through residual learning. The model is highly suitable for complex tasks like air quality prediction, where accuracy and robustness are critical.

1.2 Contributions of this research

The study comprehensively evaluates traditional deep learning models, including BiLSTM, CNN, GRU, LSTM, and RNN, to address the pressing issue of pollutant concentration prediction at U.S. embassies. Introduced the innovative Pro-1 model, combining two 1DCNN branches through element-wise addition with BiLSTM, resulting in highly accurate predictions. Additionally, the Pro-2 model builds

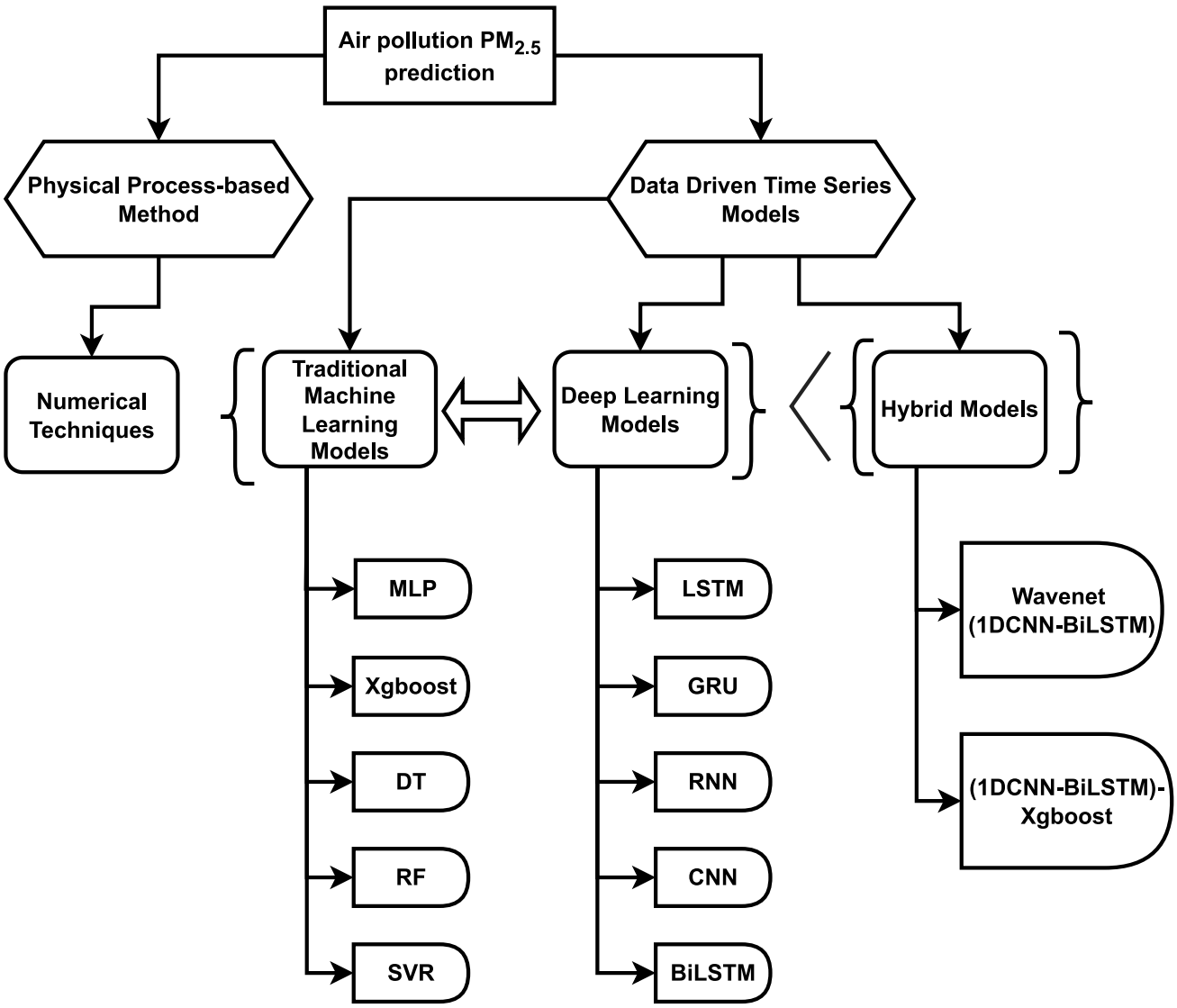


Fig. 2 Categorization of literature review based on air pollution forecasting using different techniques

Table 1 List of the abbreviations

Abbreviation	Definition	Abbreviation	Definition
AI	Artificial Intelligence	PM10	Particulate Matter 10 μm
MAE	Mean Absolute Error	MSE	Mean Absolute Error
ARIMA	Autoregressive Integrated Moving Average	NO	Nitrogen Dioxide
SARIMA	Seasonal Autoregressive Integrated Moving Average	SO_2	Sulfur Dioxide
GARCH	Generalized Autoregressive Conditional Heteroskedasticity	O	Ozone
AR	Autoregressive	CO	Carbon Monoxide
MA	Moving Average	LSTM	Long Short-Term Memory
SVM	Support Vector Machine	GRU	Gated Recurrent Unit
LR	Linear Regression	BiLSTM	Bidirectional LSTM
DT	Decision Tree	CNN	Convolutional Neural Network
XGBoost	eXtreme Gradient Boosting	RNN	Recurrent Neural Network
RF	Random Forest	PSO	Particle Swarm Optimization
$PM_{2.5}$	Particulate Matter 2.5 μm	RMSE	Root Mean Squared Error

on this foundation by integrating XGBoost for advanced residual prediction, significantly enhancing performance. Our proposed models excel at capturing both short-term and long-term dependencies, clearly outperforming other existing approaches. Below, summarize the pivotal contributions of this study:

- $PM_{2.5}$ datasets were collected from the AirNow website, covering major urban cities based on population and pollution levels. The dataset spans from 2017 to 2024 and includes cities such as Abu Dhabi, Beijing, Colombo, Delhi, Dhaka, Jakarta, Kampala, Kuwait City, Manama, and Ulaanbaatar.
- A new hybrid deep learning model, **Pro-1: Proposed 1DCNN-BiLSTM model**, was developed to enhance $PM_{2.5}$ forecasting accuracy. An advanced version, **Pro-2: Proposed 1DCNN-BiLSTM-XGBoost model**, further improved predictions by optimizing residual errors using XGBoost.
- The proposed models **Pro-1** and **Pro-2** were compared against traditional deep learning models such as BiLSTM, CNN, GRU, LSTM, and RNN.
- The predictive performance of the proposed models was statistically validated using the Friedman ranking and the Diebold-Mariano test.

The paper begins with the Introduction Section 1, and the other sections are structured as follows: Section 2 provides a comprehensive review of the literature, while Section 3, divided into several subsections, covers the Materials and EDA Section 4 details the Models Methodology, like 1DCNN-BiLSTM and 1DCNN-BiLSTM-XGBoost Section 5 presents the Results, and finally, Section 6 offers the Conclusion. Declarations are mentioned at the end of the paper.

2 Literature review

The He et al. (2022) utilized an artificial neural network to predict monthly $PM_{2.5}$ concentrations in Liaocheng, China, highlighting the importance of regional data characteristics and model adaptability in air quality forecasting. Comparative analyses of deep learning models, such as those applied in Weifang City for climate forecasting, offer valuable insights into the relative strengths and weaknesses of each architecture (Guo et al. 2024b). A recent comparative analysis in Dezhou City demonstrated that the performance of deep learning models for monthly $PM_{2.5}$ forecasting can vary significantly, underscoring the importance of hybrid architectures and appropriate optimization strategies (He and Guo 2024).

In these review papers, Agbehadji and Obagbuwa worked on a systematic review of the machine and deep learning techniques for spatiotemporal air quality prediction using various ML and DL models, including random forest and decision tree classifiers. Findings of the research combining ML and DL techniques can address data limitations and process the nonlinear characteristics of air pollutants. The research gap needs more studies on the impact of military activities and fires on O_3 concentration (Agbehadji and Obagbuwa 2024). Gangwar et al. review air pollution monitoring and forecasting systems using IoT, big data, and machine learning. Research outcomes integration of IoT and big data with ML models enhances air pollution monitoring and forecasting. The research gap challenges in data quality, standardization, and real-time processing need to be addressed (Gangwar et al. 2023). Blanco et al. Forecast urban air pollution by integrating satellite observations and meteorological forecasts. Machine learning models using satellite and meteorological data. Models demonstrated efficacy in predicting pollutant levels for the forthcoming day, achieving a percentage error of around 30%. The research gap needs validation in diverse urban environments and exploration of additional data sources (Blanco et al. 2024). Table 2 has the recent literature review on air pollution prediction using deep learning and combines machine learning and deep learning models.

Before the advancement of AI, only traditional statistical models, such as ARIMA, SARIMA, GARCH, AR, and MA, were used to forecast air levels in a limited time frame. Machine learning models, such as SVM, LR, DT, Xgboost, and RF, predict air quality levels or pollutant concentrations like $PM_{2.5}$, PM_{10}, NO_2, SO_2, O_3, and CO. Several studies have employed machine learning techniques for air pollution prediction. Bosco and Kowsalya (2024) proposed an ensemble learning approach that combines multiple machine learning algorithms to improve prediction accuracy for air pollution in Rome, Italy, achieving an RMSE of 6.7, MAE of 4.5, and MSE of 44.89. Blanco et al. (2024) integrated satellite observations and meteorological forecasts with machine learning models to forecast urban air pollution in Milan, Italy, reporting an RMSE of 9.546 and an MAE of 6.492. Their study highlighted the necessity of validating models in diverse urban environments and incorporating additional data sources to enhance prediction reliability. These studies emphasize the role of machine learning models in air pollution forecasting and the potential benefits of integrating multiple data sources for improved accuracy.

Deep learning models, such as LSTM, GRU, BiLSTM, CNN, and RNN, are also used to predict air pollution concentration or combined air pollution levels. Deep learning models have been extensively used to develop more accurate and efficient air pollution prediction systems. Pranolo et al. (2023) optimized LSTM, CNN, and MLP models using PSO to improve $PM_{2.5}$ prediction in Beijing, China. The author evaluated the RMSE and MAPE on a Monthly,

Table 2 Literature review on air pollution prediction using deep learning models

Reference	Purpose	Models used	Evaluation parameters (RMSE, MAE, MSE)	Dataset area	Research gap
Pranolo et al. (2023)	Optimize deep learning architectures for $PM_{2.5}$ prediction using Particle Swarm Optimization (PSO).	LSTM, CNN, and MLP optimized with PSO.	RMSE and MAPE with the Monthly, Weekly, and Daily basis	Beijing, China	Further research is needed on the applicability of PSO in optimizing other deep learning models for air pollution prediction.
Tao et al. (2019)	Develop a methodology for air pollution forecasting using deep learning models.	1D ConvNets and Bidirectional GRU.	RMSE: 0.3330, and MAE: 0.2063	Beijing, China	Exploration of other deep learning architectures and external factors influencing air pollution.
Rad et al. (2025)	Forecast concentrations of CO, O_3, NO_2, SO_2, PM_{10}, and $PM_{2.5}$ in Tehran.	GRU, FCNN, CNN	RMSE: 7.03, MAE: 5.01, MSE: 58.12	Tehran, Iran	Further research is needed to explore the impact of additional environmental variables on model performance.
Khan et al. (2024)	Predict Air Quality Index (AQI) in megacities using attention-enhanced deep multitask spatiotemporal learning.	Attention-enhanced LSTM	RMSE, MAE, MSE are applied over SO2, NO2, O3, and CO	Los Angeles, California	Exploration of model performance in different urban settings and integration with real-time data.
Blanco et al. (2024)	Forecast urban air pollution by integrating satellite observations and meteorological forecasts.	Machine Learning models leveraging satellite and meteorological data	RMSE: 9.546, and MAE: 6.492	Milan, Italy	Need for validation in diverse urban environments and exploration of additional data sources.
Chen et al. (2025)	Develop a hybrid deep learning approach for air pollution prediction considering spatio-temporal correlations.	CNN-LSTM integrated with a spatio-temporal attention mechanism.	RMSE: 10.513, and MAE: 6.919	Beijing, China	Further exploration of the impact of different attention mechanisms on model performance is needed.
Bosco and Kowsalya (2024)	Propose a novel approach for air pollution prediction using machine learning techniques.	Ensemble learning combining multiple machine learning algorithms.	RMSE: 6.7, MAE: 4.5, MSE: 44.89	Rome, Italy	Investigation into integrating deep learning techniques with ensemble methods.
Che et al. (2024)	Develop a short-term air quality prediction model using deep learning systems coupled with multi-factor decomposition.	Deep learning models integrated with multi-factor decomposition techniques.	RMSE and MAE are utilized over spring, summer, autumn, and winter seasons	Beijing, China	needs research on long-term prediction capabilities and model generalization.
Srivastava and Kumar Das (2023)	Propose an air pollution prediction system using the XRSTH-LSTM algorithm.	XRSTH-LSTM	RMSE: 0.167845, and MSE: 0.051248	Mumbai, India	Further studies on the scalability and real-time application of the model.

Weekly, and Daily basis. Tao et al. (2019) developed a forecasting methodology using 1D ConvNets and Bidirectional GRU, reporting an RMSE: 0.3330, and MAE: 0.2063, emphasizing the need for exploring additional deep learning architectures and external factors affecting air pollution. Rad et al. (2025) applied GRU, FCNN, and CNN models to forecast multiple air pollutants in Tehran, Iran, achieving an RMSE: 7.03, MAE: 5.01, MSE: 58.12, suggesting further research into the impact of environmental variables on model performance. Khan et al. (2024) proposed an attention-enhanced LSTM for predicting AQI in Los Angeles, California. The author used RMSE, MAE, and MSE over SO_2, NO_2, O_3, and CO, demonstrating the potential of attention mechanisms in deep learning-based air pollution models. Che et al. (2024) introduced deep learning models integrated with multi-factor decomposition techniques to develop a short-term air quality prediction model in Beijing, China. RMSE and MAE are utilized over the spring, summer, autumn, and winter seasons, stressing the need for research on long-term prediction capabilities and model generalization. Srivastava and Kumar Das (2023) presented an air pollution prediction system using the XRSTH-LSTM algorithm in Mumbai, India, reporting an RMSE of 0.167845 and an MSE of 0.051248, with a focus on scalability and real-time application. Recent work has shown that ANN-based models can effectively predict hourly $PM_{2.5}$ and PM_{10} concentrations, achieving high accuracy when tailored to city-specific meteorological and emission data (Guo et al. 2023b).

Hybrid models use a combination of traditional statistical models, machine learning models, and deep learning models. Hybrid deep learning models have enhanced air pollution prediction accuracy by leveraging multiple model architectures (Ahmad and Kumar 2025a, b, 2023). Chen

et al. (2025) developed a hybrid CNN-LSTM model integrated with a spatio-temporal attention mechanism for air pollution prediction in Beijing, China, achieving an RMSE: 10.513, and MAE: 6.919. Emphasizing the need for further exploration of different attention mechanisms. Li et al. (2024) introduced a composite deep learning model, BWO-BiLSTM, to predict atmospheric particulate matter mass concentration in Shanghai, China. Advocating for its application to different pollutants and environmental conditions. These hybrid models demonstrate the advantages of combining deep learning techniques to improve air pollution forecasting systems' predictive accuracy and robustness.

3 Materials (Data)

The study area map illustrates the geographical distribution of U.S. embassies where $PM_{2.5}$ air quality data was collected from the AirNow website. These embassies are strategically located in major urban centers with significant air pollution concerns, including Abu Dhabi, Beijing, Colombo, Delhi, Dhaka, Jakarta, Kampala, Kuwait City, Manama, and Ulaanbaatar. The dataset spans 2017 to 2024, ensuring a comprehensive temporal analysis of air quality trends. The inclusion of all U.S. embassies in the study area provides a globally representative dataset, allowing for robust model evaluation across diverse climatic and pollution conditions. The spatial distribution ensures that the proposed forecasting models are tested under varied environmental settings, enhancing their generalizability and reliability for real-world air quality monitoring.

3.1 Exploratory data analysis

The authors compiled 17 comprehensive datasets from all US embassies and consulates, representing a population of 1.5 million, with historical data available since 2017. Among these datasets, only 10 contain fewer than 10,000 missing values, and several do not include data for the year 2023. The high proportion of missing values can significantly hinder effective model building. To address this issue, the author imputed all missing values by using the mean of the corresponding $PM_{2.5}$ features. The approach calculates the average while excluding zeros, effectively minimizing the impact of outliers and ensuring our analysis remains accurate and reliable. The data originates from US embassies and official air monitoring stations, utilizing $PM_{2.5}$ datasets from multiple cities, including Abu Dhabi, Beijing, Colombo, Delhi, Dhaka, Jakarta, Kampala, Kuwait City, Manama, and Ulaanbaatar in Fig. 3.

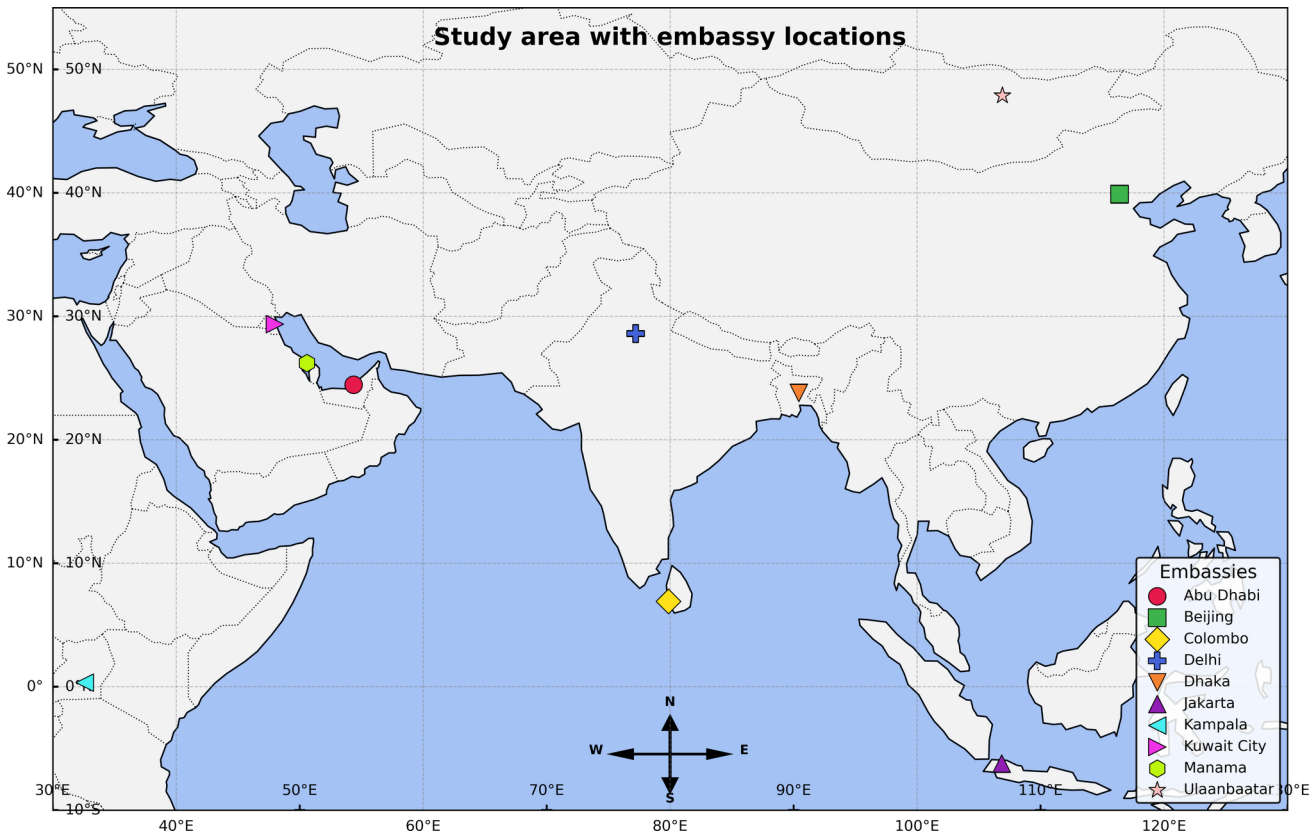


Fig. 3 Worldwide location of US embassies that is utilized for research

The statistical summary of the exploratory data analysis reveals significant variations in $PM_{2.5}$ concentrations across different cities. Delhi and Dhaka have the highest mean values, at 97.91 and 91.70 μg/m^3, respectively, indicating severe air pollution. Both cities demonstrate notably high standard deviations, suggesting considerable fluctuations in $PM_{2.5}$ levels. In contrast, Colombo and Jakarta show relatively lower mean concentrations of 22.20 and 33.79 μg/m^3, respectively, with smaller variations. Additionally, Abu Dhabi, Beijing, and Kuwait City report extreme maximum values of 688.7, 670.8, and 808.6 μg/m^3, respectively, indicating occasional severe pollution events (see Table 3). The wide interquartile ranges observed in many cities, particularly Ulaanbaatar and Manama, further highlight the variability in pollution levels. These insights are vital for predictive modeling, as they suggest that $PM_{2.5}$ levels are highly city-specific, necessitating tailored deep-learning approaches to enhance prediction accuracy while accounting for extreme pollution events and seasonal variations. Guo et al. (2020) introduced hybrid approaches, such as the integration of wavelet decomposition with artificial neural networks and meteorological variables, which have demonstrated enhanced forecasting capabilities by effectively managing data non-stationarity and seasonality.

The time series decomposition of $PM_{2.5}$ data reveals notable disparities in pollution levels among cities, with alarming peaks observed in Delhi and Ulaanbaatar. The trend analysis indicates that certain cities are experiencing concerning increases in pollution over time, while residual analysis uncovers substantial noise, particularly in heavily polluted areas. ACF plots demonstrate strong autocorrelation, suggesting that past pollution levels significantly influence future values. Meanwhile, PACF assists in identifying optimal lag structures for time series modeling. The analysis underscores the pressing need for robust predictive models that effectively capture trends, seasonal patterns, and interdependencies for accurate forecasting of $PM_{2.5}$ levels, as illustrated in Figs. 4 and 5.

The similarity in the $PM_{2.5}$ distribution patterns observed in Kampala and Manama, as well as in Delhi and Dhaka, can be attributed to underlying environmental and meteorological factors. Kampala and Manama, despite being in different geographical regions, share similarities in pollution sources such as urban emissions, vehicular pollution, and localized industrial activities, which contribute to comparable air quality distributions. Similarly, Delhi and Dhaka are densely populated South Asian megacities with high levels of industrial emissions, construction activities, and biomass burning, leading to similar $PM_{2.5}$ concentration patterns.

Table 3 US embassies location, geographical details, population, and statistical summary of exploratory data analysis

City	Country	Latitude	Longitude	Altitude (m)	Population (M)	Mean	Std	Min	25%	50%	75%	Max
Abu Dhabi	UAE	24.4667	54.3667	27	3.8	37.62	23.68	0.1	24.0	37.6	44.3	688.7
Beijing	China	39.9042	116.4074	44	21.5	43.23	44.24	0.1	14.6	30.8	55.3	670.8
Colombo	Sri Lanka	6.9271	79.8612	1	5.6	22.20	14.58	0.1	11.8	21.8	26.6	178.7
Delhi	India	28.6139	77.2090	216	32.2	97.91	89.55	0.1	37.1	68.3	128.15	1546.9
Dhaka	Bangladesh	23.8103	90.4125	4	22	91.70	73.96	0.2	37.5	70.1	122.9	844.9
Jakarta	Indonesia	-6.2088	106.8456	8	30	33.79	18.89	0.1	19.3	32.3	44.7	168.5
Kampala	Uganda	0.3476	32.5825	1190	8.5	53.47	26.77	0.1	36.0	52.0	61.9	319.9
Kuwait City	Kuwait	29.3759	47.9774	55	4.82	46.19	41.98	0.6	26.6	38.1	52.5	808.6
Manama	Bahrain	26.2279	50.5856	5	1.7	50.26	32.14	0.1	33.4	47.3	57.1	802.3
Ulaanbaatar	Mongolia	47.8864	106.9057	1300	1.7	58.64	77.43	0.1	12.4	35.0	58.6	891.0

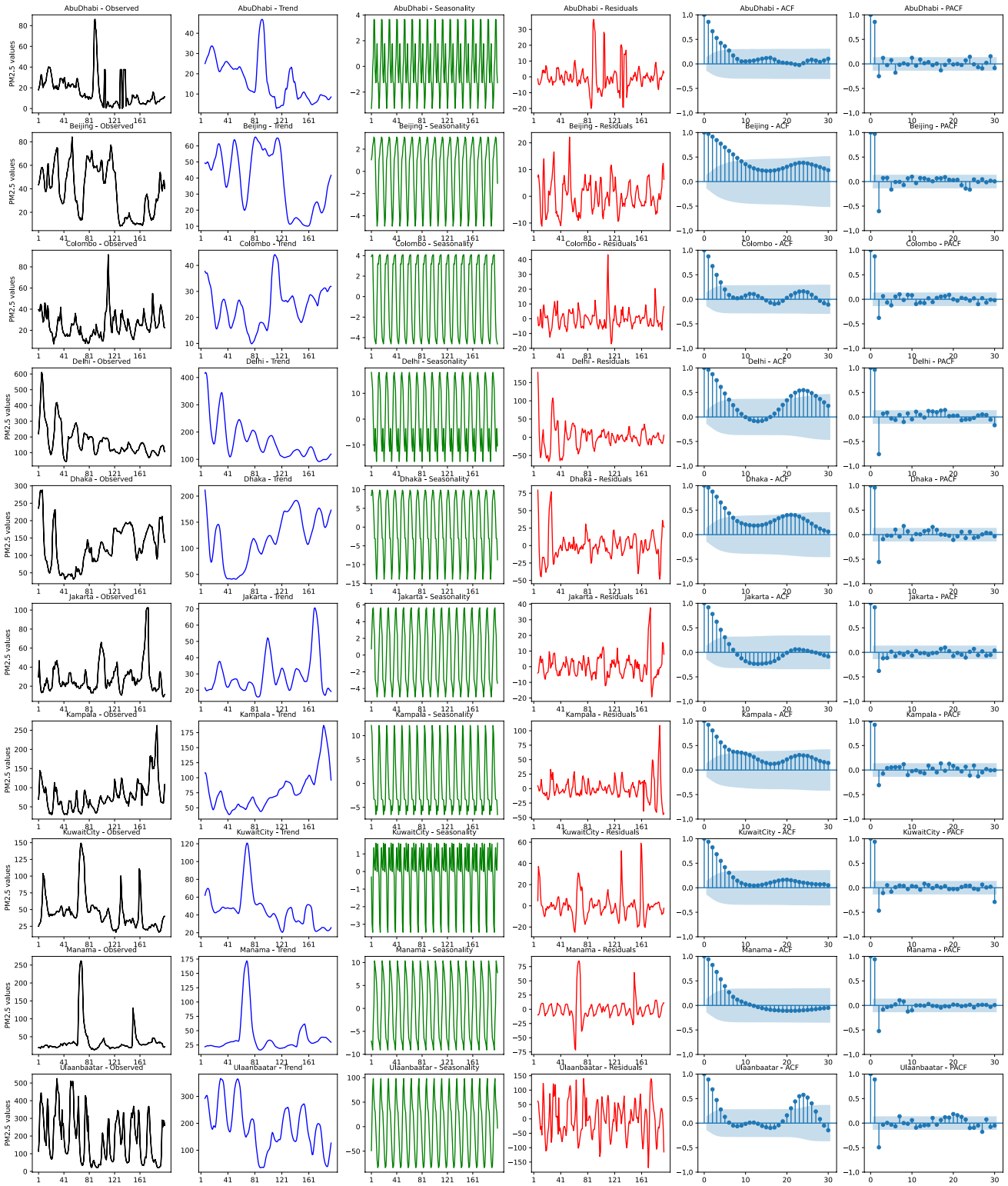


Fig. 4 The data description of all datasets with observed, trend, residual, seasonal, ACF, and PACF

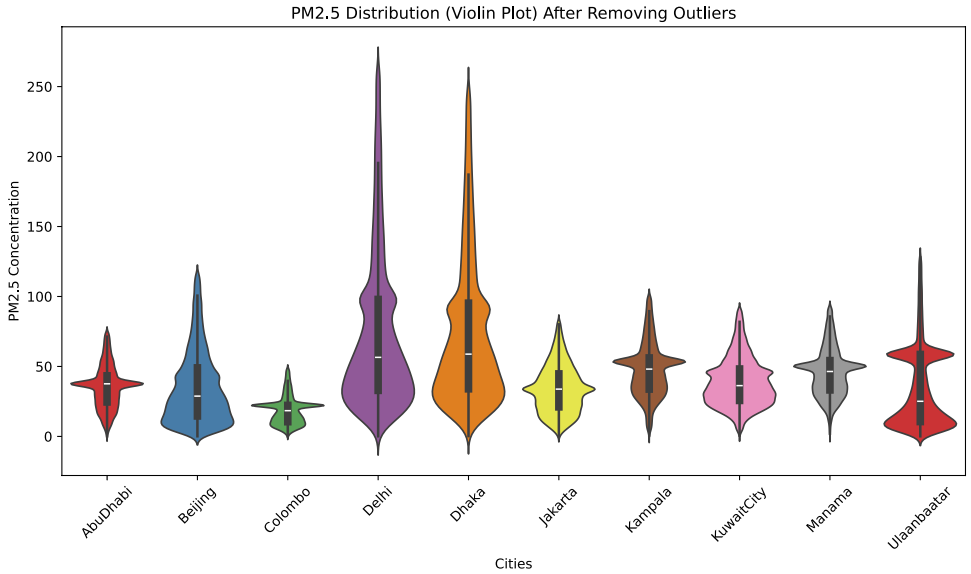


Fig. 5 The $PM_{2.5}$ concentration distribution of the all 10 dataset of US embassies

4 Methods and methodology

Traditional statistical and machine learning models often fail to capture air pollution data's highly nonlinear and complex temporal patterns. This research addresses the challenge of accurately predicting air pollution concentrations at urban U.S. embassies, crucial monitoring points in international cities. The comparison of LSTM and ANN models for daily ozone forecasting in Liaocheng City illustrates the practical advantages of recurrent structures in capturing temporal dependencies in atmospheric pollutant levels (Guo et al. 2025). Hybrid models combining wavelet transformation with deep learning architectures have demonstrated enhanced accuracy in daily $PM_{2.5}$ forecasting, as observed in recent work focused on Guangzhou City (He et al. 2025). The motivation behind this work is to build a robust, efficient, and generalizable model that can support proactive air quality management and health advisories in urban regions where air quality data quality and prediction precision are critical. Additionally, recent studies have explored wavelet-based neural architectures for daily $PM_{2.5}$ prediction, where the integration of wavelet decomposition and artificial neural networks showed improved performance by capturing both high- and low-frequency variations in air pollutant data (Guo et al. 2023c).

To gather the $PM_{2.5}$ air pollution data from the most polluted urban U.S. embassies, the authors focus on cities with populations exceeding one million, ensuring data availability from 2017.

$$X = \{X_1, X_2, \ldots, X_n\} \tag{1}$$

Here n is the total number of US embassies containing $PM_{2.5}$ concentration data. The D is the total dataset that combines each US embassy, like Abu Dhabi, Beijing, Colombo, Delhi, Dhaka, Jakarta, Kampala, Kuwait City, Manama, and Ulaanbaatar. Separately denoted by X_1, X_2, X_3 ,..... soo on see Eq. 1. A min-max scaler was applied to normalize the dataset, all dataset X between 0 and 1:

$$X' = \frac{X - X_{\min}}{X_{\max} - X_{\min}} \tag{2}$$

where X' represents the scaled values, and $X_{\min}$ and $X_{\max}$ are the minimum and maximum values of the feature across the dataset of the US embassies of $PM_{2.5}$. The dataset was split into training, validation, and testing sets using a 70:15:15 ratio. Here, give enough data to train the model so that there is no overfitting of the proposed and traditional deep learning models. 15 percent of the whole data is given for the validation to analyze the performance of the models, and 15 percent is given for the test in new scenarios for the model to be applied to unseen data.

$$X \rightarrow X_{\text{train}} : X_{\text{val}} : X_{\text{test}} = 70\% : 15\% : 15\% \tag{3}$$

4.1 Pro-1: Proposed 1DCNN-BiLSTM model

The Pro-1 model integrates convolutional and recurrent neural network capabilities to effectively capture both spatial and temporal patterns in $PM_{2.5}$ data. The algorithm of Pro-1 is shown in the Algorithm 1 where details mathematical description are as follow:

- **Input series:** Given an input time series $X \in \mathbb{R}^{T\times d}$ where T is the sequence length and d is the feature dimension, process this through the dual-branch architecture.
- **1DCNN:** Each 1DCNN branch employs multiple convolutional filters to extract distinct feature representations:

$$H^{(l)}_{\text{CNN}_1} = \text{ReLU}(W^{(l)}_{\text{CNN}_1} * H^{(l-1)}_{\text{CNN}_1} + b^{(l)}_{\text{CNN}_1}) \tag{4}$$

$$H^{(l)}_{\text{CNN}_2} = \text{ReLU}(W^{(l)}_{\text{CNN}_2} * H^{(l-1)}_{\text{CNN}_2} + b^{(l)}_{\text{CNN}_2}) \tag{5}$$

where $l = 1, ..., L$ denotes the layer index, with $H^{(0)}_{\text{CNN}_1} = H^{(0)}_{\text{CNN}_2} = X$. The convolution operation $*$ with filter $W_{\text{CNN}} \in \mathbb{R}^{k\times d_{\text{in}}\times d_{\text{out}}}$ (where k is the kernel size, d_{in} and d_{out} are input and output dimensions) is defined as:

$$(W * X)_t = \sum_{i=0}^{k-1} W_i X_{t+i} \tag{6}$$

The two 1DCNNs utilize different kernel sizes and filter numbers to capture complementary features at varying temporal scales. Here, the first 1DCNN focuses on short-term patterns while the second 1DCNN captures medium-range dependencies. This estimation is consistent with prior analyses in convolutional sequence models such as those presented in Shao et al. (2024); Nguyen et al. (2024), which emphasize the trade-off between expressivity and efficiency in temporal convolutional architectures. **Feature Fusion Process:** The element-wise addition allows informative features from both branches to be preserved while streamlining the model's parameter space:

$$H_{\text{combined}} = H^{(L)}_{\text{CNN}_1} + H^{(L)}_{\text{CNN}_2} \tag{7}$$

The fusion approach performs implicit feature selection, with the strongest activations from either 1DCNN dominating the combined representation.

- **BiLSTM model:** The BiLSTM processes the fused features $H_{\text{combined}} \in \mathbb{R}^{T\times d'}$ by analyzing the sequence in both forward and backward directions:

$$\overrightarrow{h_t} = \text{LSTM}_{\text{forward}}(H_{\text{combined},t}, \overrightarrow{h_{t-1}}, \overrightarrow{C_{t-1}}) \tag{8}$$

$$\overleftarrow{h_t} = \text{LSTM}_{\text{backward}}(H_{\text{combined},t}, \overleftarrow{h_{t+1}}, \overleftarrow{C_{t+1}}) \tag{9}$$

Each LSTM cell computes the following for time step t:

$$\begin{aligned} f_t &= \sigma(W_f[h_{t-1}, H_{\text{combined},t}] + b_f) \\ i_t &= \sigma(W_i[h_{t-1}, H_{\text{combined},t}] + b_i) \\ \tilde{C}_t &= \tanh(W_C[h_{t-1}, H_{\text{combined},t}] + b_C) \\ C_t &= f_t \odot C_{t-1} + i_t \odot \tilde{C}_t \\ o_t &= \sigma(W_o[h_{t-1}, H_{\text{combined},t}] + b_o) \\ h_t &= o_t \odot \tanh(C_t) \end{aligned} \tag{10}$$

The hidden states from both directions are concatenated for each time step, see Eq. 10 for the more aligned equation:

$$h_t = [\overrightarrow{h_t}, \overleftarrow{h_t}] \in \mathbb{R}^{2\times d_h} \tag{11}$$

where d_h is the hidden dimension of each LSTM direction.

- **Multi-timestep processing:** For $PM_{2.5}$ prediction, the final hidden state can be further processed to predict multiple future time steps:

$$\hat{Y}_{\text{Pro-1}} = W_y h_T + b_y \tag{12}$$

where h_T is the concatenated hidden state at the final time step, and $\hat{Y}_{\text{Pro-1}} \in \mathbb{R}^F$ with F being the forecast horizon.

- **Learning process:** The model parameters $\Theta = \{W_{\text{CNN}_1}, b_{\text{CNN}_1}, W_{\text{CNN}_2}, b_{\text{CNN}_2}, W_f, b_f, W_i, b_i, W_C, b_C, W_o, b_o, W_y, b_y\}$ are optimized by minimizing the mean squared error loss:

$$\mathcal{L}(\Theta) = \frac{1}{N}\sum_{j=1}^{N} \|Y_j - \hat{Y}_{\text{Pro-1},j}\|_2^2 \tag{13}$$

where N is the number of training samples.

4.2 Pro-2: Proposed 1DCNN-BiLSTM-XGBoost residual correction model

Pro-2 extends Pro-1 by implementing a residual correction mechanism using XGBoost regression. The ensemble approach systematically addresses prediction biases in the Pro-1 model. The Algorithm 2 shows the structure of Pro-2 processing.

- **Residual computation:** For the training dataset, compute residuals as the error between actual values and Pro-1 predictions:

$$RD_{\text{train},i} = Y_{\text{actual,train},i} - \hat{Y}_{\text{Pro-1,train},i} \tag{14}$$

These residuals represent structured prediction errors that may follow certain patterns based on input data characteristics.

- **XGBoost regression model deployment:** The XGBoost model f_{XGB} is trained to predict residuals based on Pro-1 predictions:

$$f_{\text{XGB}} : \hat{Y}_{\text{Pro-1,train}} \rightarrow RD_{\text{train}} \tag{15}$$

Mathematically, the XGBoost model approximates the residual function using an ensemble of *K* decision trees:

$$f_{\text{XGB}}(\hat{Y}) = \sum_{k=1}^{K} f_k(\hat{Y}) \tag{16}$$

where each f_k is a decision tree structure defined as:

$$f_k(\hat{Y}) = w_{q(\hat{Y})} \tag{17}$$

where q represents the tree structure that maps each input to the corresponding leaf index, and w is the vector of leaf weights. The objective function optimized during training includes both loss and regularization terms:

$$\mathcal{L}_{\text{XGB}} = \sum_i l(RD_{\text{train},i}, f_{\text{XGB}}(\hat{Y}_{\text{Pro-1,train},i})) + \sum_k \Omega(f_k) \tag{18}$$

where l is the loss function (typically squared error for regression) and $\Omega(f) = \gamma T + \frac{1}{2}\lambda\|w\|^2$ is the regularization term with T representing the number of leaves in the tree.

- **Residual prediction for test data:** For test data, first generate predictions using Pro-1:

$$\hat{Y}_{\text{Pro-1,test}} = \mathcal{F}_{\text{Pro-1}}(X_{\text{test}}) \tag{19}$$

Then, the trained XGBoost model predicts the corresponding residuals:

$$\hat{RD}_{\text{test}} = f_{\text{XGB}}(\hat{Y}_{\text{Pro-1,test}}) \tag{20}$$

- **Final ensemble prediction:** The final prediction combines the Pro-1 prediction and the predicted residual:

$$\hat{Y}_{\text{Final}} = \hat{Y}_{\text{Pro-1,test}} + \hat{RD}_{\text{test}} \tag{21}$$

The correction mechanism systematically addresses specific biases or limitations in the Pro-1 model's forecasting capabilities. Pro-2 offers theoretical advantages through bias-variance decomposition. The Pro-1 model effectively captures the main temporal patterns in the data, while the XGBoost residual model specifically targets systematic biases that the Pro-1 model consistently exhibits. This decomposition allows each component to specialize in capturing different aspects of the underlying data generation process.

4.3 Performance measures

To ensure the author identifies the most effective model for our regression problem, the author rigorously evaluates its performance using a range of error measurement tools. By applying RMSE, MAE, and MSE, the authors aim to decisively pinpoint the best model, allowing us to achieve the highest level of accuracy and reliability in our predictions.

$$\text{MSE} = \frac{1}{n}\sum_{i=1}^{n}(Y_i - \hat{Y}_{\text{Final},i})^2 \tag{22}$$

$$\text{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(Y_i - \hat{Y}_{\text{Final},i})^2} \tag{23}$$

$$\text{MAE} = \frac{1}{n}\sum_{i=1}^{n}|Y_i - \hat{Y}_{\text{Final},i}| \tag{24}$$

Here $\hat{Y}_{\text{Final}}$ is the adjusted prediction for Pro-2, $\hat{Y}_{\text{test}}$ is the initial prediction from Pro-1, and RD_{test} is the predicted residual adjustment. Also, Y_i represents the actual observed value for $PM_{2.5}$, n is the number of samples, and MSE, RMSE, and MAE Eqs. 22 to 24 are used to evaluate the accuracy of $\hat{Y}_{\text{Final}}$ by measuring the error between the actual values and the final predictions.

The combination of CNN and BiLSTM allows the model to capture temporal and spatial features in the data, making it highly versatile for complex datasets. By modeling the residuals separately, XGBoost acts as a corrective mechanism, improving the overall accuracy of the predictions. The WaveNet structure integrates convolutional and bidirectional LSTM layers, creating a unified deep-learning framework for sequential data processing. It uses dilated causal convolutions to capture long-range dependencies in sequential data, making it highly effective for time-series forecasting or sequential data modeling. WaveNet serves as the backbone, providing the ability to model temporal patterns and dependencies in the input data. On top of the WaveNet architecture, you incorporate stacked convolutional layers presented in Fig. 6. These layers help extract hierarchical spatial features from the data, which can be particularly useful if

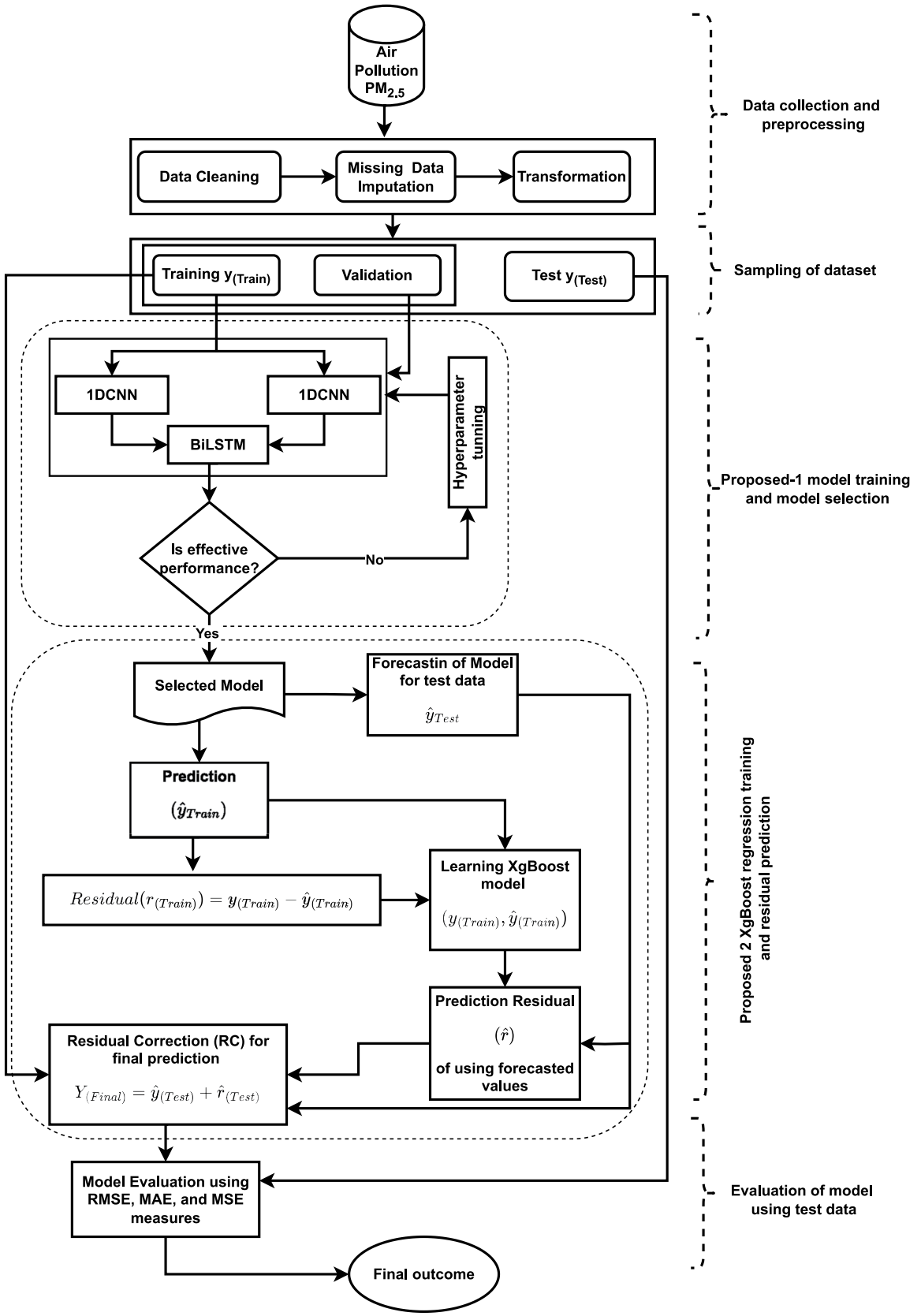


Fig. 6 Hybrid deep learning framework for $PM_{2.5}$ air pollution prediction using 1DCNN-BiLSTM, and XGBoost

your input data has a spatial structure, multi-dimensional time series, or grid-based data. The stacked CNN layers enhance the model's ability to capture local patterns and relationships in the data, complementing the global temporal patterns learned by WaveNet. After the CNN layers, add Bidirectional LSTM layers. BiLSTM networks are designed to process sequential data in both forward and backward directions, capturing dependencies from past and future time steps. This bidirectional context is beneficial for tasks where future context can improve predictions of air quality forecasting and where future meteorological conditions might influence current pollutant levels. After training the proposed model, WaveNet will calculate the residuals and the differences between the actual values and the model's predictions.

The output of the first Pro-1 model prediction results is stored in $\hat{y}$, and the residual is computed concerning the same $\hat{y}$, with the difference of the actual training values. Then the equation should be look like $(R = y_{Train} - \hat{y}_{Train})$. These residuals represent the errors or unexplained patterns that the initial model could not capture. Residuals are a powerful way to identify and address the limitations of the initial model, ensuring that no significant patterns in the data are overlooked. Then, an XGBoost model was trained to predict these residuals. XGBoost is a gradient-boosting algorithm known for handling complex, non-linear relationships and providing high accuracy in regression tasks. By focusing on the residuals, XGBoost learns to correct the errors made by the initial model, effectively refining the overall prediction.

The diagram illustrates the two-stage prediction framework. First, an initial prediction $\hat{y}_{Test}$ is generated. The residuals $(R = y_{Train} - \hat{y}_{Train})$ are modeled using XGBoost to predict $\hat{r}_{Test}$. The final prediction is obtained by summing the initial prediction and the modeled residual: $Y_{Final} = \hat{y}_{Test} + \hat{r}_{Test}$. The final prediction is obtained by adding the XGBoost residual prediction to the proposed model's prediction. The ensemble approach ensures that the strengths of the deep learning model WaveNet and the gradient boosting model XGBoost are combined, leading to a more accurate and robust prediction.

Algorithm 1 Pro-1: 1DCNN-BiLSTM model for $PM_{2.5}$ prediction.

Require: $PM_{2.5}$ time series data X, split into X_{train}, X_{val}, and X_{test}
Ensure: Predicted $PM_{2.5}$ values $\hat{Y}_{\text{Pro-1}}$
Initialization:
Initialize weights for CNN branches: $W_{\text{CNN}_1}, W_{\text{CNN}_2}, b_{\text{CNN}_1}, b_{\text{CNN}_2}$
Initialize BiLSTM parameters: $W_f, W_i, W_o, W_C, b_f, b_i, b_o, b_C$
Initialize output layer parameters: W_y, b_y
Set maximum epochs E and learning rate η
Training of the model:
for *epoch* $= 1$ to E **do**
  for each batch b in X_{train} **do** // See Eqs. 4 to 5
    $H_{\text{CNN}_1} = \text{ReLU}(W_{\text{CNN}_1} * b + b_{\text{CNN}_1})$
    $H_{\text{CNN}_2} = \text{ReLU}(W_{\text{CNN}_2} * b + b_{\text{CNN}_2})$
    `// Feature Fusion`
    $H_{\text{combined}} = H_{\text{CNN}_1} + H_{\text{CNN}_2}$
    `// BiLSTM`
    Initialize forward and backward states: $\overrightarrow{h_0}, \overrightarrow{C_0}, \overleftarrow{h_T}, \overleftarrow{C_T}$
    for $t = 1$ to T `(Forward)` **do** // See Eq. 8
      Update cell state: $\overrightarrow{C_t} = f_t \odot \overrightarrow{C_{t-1}} + i_t \odot \tilde{C}_t$
      Update hidden state: $\overrightarrow{h_t} = o_t \odot \tanh(\overrightarrow{C_t})$
    end for
    for $t = T$ to 1 `(Backward)` **do** // See Eq. 9
      Compute LSTM gates f_t, i_t, o_t using current input and future state
      Update cell state: $\overleftarrow{C_t} = f_t \odot \overleftarrow{C_{t+1}} + i_t \odot \tilde{C}_t$
      Update hidden state: $\overleftarrow{h_t} = o_t \odot \tanh(\overleftarrow{C_t})$
    end for
    for $t = 1$ to T **do** // See Eq. 11
      $h_t = [\overrightarrow{h_t}, \overleftarrow{h_t}]$
    end for
    `// Output layer`
    $\hat{Y}_{\text{batch}} = W_y h_T + b_y$
    Backward pass:
    Compute loss: $\mathcal{L} = \frac{1}{|b|} \sum_{i \in b} \|Y_i - \hat{Y}_i\|^2$
    Update all parameters via backpropagation using learning rate η
  end for
  `// Validation`
  Compute validation loss on X_{val}
  if validation loss stops improving for p epochs **then**
    Early stopping
  end if
end for
Testing:
Generate final predictions $\hat{Y}_{\text{Pro-1}}$ on X_{test}
return $\hat{Y}_{\text{Pro-1}}$

Algorithm 2 Pro-2: 1DCNN-BiLSTM-XGBoost model for $PM_{2.5}$ prediction.

Require: $PM_{2.5}$ time series data X, split into X_{train}, X_{val}, and X_{test}
Ensure: Final predicted $PM_{2.5}$ values $\hat{Y}_{\text{Final}}$
Train Pro-1 model
Train the 1DCNN-BiLSTM model as per Algorithm 1
Obtain predictions $\hat{Y}_{\text{Pro-1,train}}$ on training data
Obtain predictions $\hat{Y}_{\text{Pro-1,test}}$ on test data
Residual computation // See Eq. 14
$RD_{\text{train}} = Y_{\text{actual,train}} - \hat{Y}_{\text{Pro-1,train}}$
XGBoost residual modeling
Set number of trees K, max depth D, learning rate α, regularization parameters λ, γ
// Train XGBoost model // See Eqs. 15 to 18
$f_{\text{XGB}} = \text{TrainXGBoost}(\hat{Y}_{\text{Pro-1,train}}, RD_{\text{train}})$, where:
$f_{\text{XGB}}(\hat{Y}) = \sum_{k=1}^{K} f_k(\hat{Y})$
f_k is the k-th regression tree
Objective: Minimize $\sum_i (RD_{\text{train},i} - f_{\text{XGB}}(\hat{Y}_{\text{Pro-1,train},i}))^2 + \Omega(f)$
$\Omega(f) = \gamma \cdot (\text{number of leaves}) + \frac{1}{2}\lambda \sum (\text{leaf weights})^2$
Prediction on test data // See Eq. 19
$\hat{RD}_{\text{test}} = f_{\text{XGB}}(\hat{Y}_{\text{Pro-1,test}})$
$\hat{Y}_{\text{Final}} = \hat{Y}_{\text{Pro-1,test}} + \hat{RD}_{\text{test}}$ // See Eq. 21
return $\hat{Y}_{\text{Final}}$

4.4 Computational complexity analysis

The computational complexity of each 1DCNN layer is $\mathcal{O}(T \cdot k \cdot d_{\text{in}} \cdot d_{\text{out}})$, where T is the sequence length, k is the kernel size, d_{in} is the number of input channels, and d_{out} is the number of output channels. This derives from performing a convolution operation across the time dimension at every timestep for each filter. For multi-layer convolutional networks, this cost scales linearly with the number of layers L.

4.4.1 Complexity of Pro-1: Proposed 1DCNN-BiLSTM model

The proposed-1 model integrates CNN with biLSTM to capture both spatial and temporal patterns in $PM_{2.5}$ data. Its complexity can be described as follows in terms of Time and space:

- **Time complexity:** Each 1DCNN branch processes input sequences of length T with kernel size k, input dimension d_{in}, and output dimension d_{out}, resulting in a complexity of $\mathcal{O}(T \times k \times d_{in} \times d_{out})$ per branch. The biLSTM operates on the fused features with a complexity of $\mathcal{O}(T \times d' \times d_h)$, where d' is the combined CNN feature dimension and d_h is the hidden state dimension. The total time complexity is with E epochs and N training samples. The overall time complexity is shown in Eq. 25:

$$\mathcal{O}(E \times N \times (T \times k \times d_{in} \times d_{out} + T \times d' \times d_h)) \quad (25)$$

- **Space complexity:** The active storage for storig intermediat representation is $\mathcal{O}(T \times d_{out} + T \times d_h)$. And parameter storage is $\mathcal{O}(k \times d_{in} \times d_{out} + d' \times d_h + d_h \times F)$ where F represents the forecast horizon.

4.4.2 Complexity of Pro-2: Proposed 1DCNN-BiLSTM-XGBoost model

The Pro-2 model extends the first model by incorporating an XGBoost regression mechanism for residual correction, adding the following complexity:

- **Time complexity:** The base Pro-1 time complexity will be remain with Pro-2 with computing residuals required operation i.e., $\mathcal{O}(N)$. XgBoost includes the tree-based ensemble learning with K trees, which introduces the time complexity, i.e., $\mathcal{O}(K \times N \times \log(N))$. Therefore, total time complexity is a combination of Pro-1 and XgBoost time complexity i.e., shown in Eq. 26:

$$\mathcal{O}(E \times N \times (T \times k \times d_{in} \times d_{out} + T \times d' \times d_h) + K \times N \times \log(N)) \quad (26)$$

- **Space complexity:** The maximum space is utilized by the base nodel Pro-1 with XgBoost model additional space i.e., $\mathcal{O}(K \times D)$ where D represents the maximum tree depth. Then total requirement of space is $\mathcal{O}(k \times d_{in} \times d_{out} + d' \times d_h + d_h \times F + K \times D)$.

5 Experiments and results analysis

In this section, the authors thoroughly analyze the experimental results, employing various effective graphical and numerical techniques to validate the proposed models. The author utilized robust statistical tools, including the AIC-BIC test, the Friedman ranking non-parametric test, and the Diebold-Mariano test, to ensure a comprehensive evaluation. The author implemented key metrics such as RMSE, MAE, and MSE to assess the model's performance across all datasets accurately. Additionally, the author created detailed visualizations comparing our proposed models with traditional deep-learning approaches, incorporating various evaluation parameters to showcase their effectiveness. We also included a Taylor diagram and line plots that illustrate the performance of all models across each dataset, further reinforcing the strength and reliability of our findings.

5.1 Experimental setup

Table 4 experimental setup for both proposed models, which uses the parameters and corresponding values. In

Table 4 Experimental setup for the Pro-1 and Pro-2 $PM_{2.5}$ prediction

Parameter	Pro-1 value	Pro-2 value
Time Series Window	10 time steps	10 time steps
Batch Size	10	10
Model Architecture	WaveNet (1DCNN + Bidirectional LSTM)	WaveNet, XGBoost Residual
Dilation Rate	2	2
LSTM Units	75	75
Dropout Rate	0.1	0.1
Optimizer	SGD (learning rate = 0.01)	SGD (learning rate = 0.01)
Loss Function	Mean Squared Logarithmic Error (MSLE)	Mean Squared Logarithmic Error (MSLE)
XGBoost Parameters	–	n_estimators=500, lr=0.05, max_depth=5
Regularization	–	alpha=1 (L1), lambda=1 (L2)
Training Epochs	200	200
Callbacks	ReduceLROnPlateau, EarlyStopping	ReduceLROnPlateau, EarlyStopping
Evaluation Metrics	MAE, MSE, and RMSE	MAE, MSE, and RMSE

this experiment, a time series window of 10 time steps and 10 steps was utilized for a batch size of 10. The model architecture in both models is two branches: 1DCNN, then BiLSTM. The Xgboost was used in Pro-2 to improve the model's accuracy. Uses of the regularization in the Xgboost regression model, in which alpha and lambda are one. The authors trained all traditional and proposed models for epoch 200 with a callback reduceLROnPlateau, and EarlyStopping when the model fully qualified for all the datasets.

5.2 Analysis of results

Table 5 performance comparison of traditional deep learning models BiLSTM, CNN, GRU, LSTM, and RNN with the proposed Pro-1 model using RMSE, MAE, and MSE for the US Embassies dataset reveals key insights. The Pro-1 model consistently achieves lower error values, indicating improved predictive accuracy. Regarding RMSE, Pro-1 outperforms traditional models across most cities, with significant reductions for Delhi at 13.90, Dhaka at 13.63, and Manama at 4.28, highlighting its robustness in highly polluted regions. Similarly, in MAE, Pro-1 achieves the lowest errors for Delhi at 7.30 and Manama at 2.23, demonstrating enhanced short-term prediction precision. The MSE values further confirm the superior performance of Pro-1, particularly in Kuwait City 62.42 and Manama 18.42, where it significantly reduces error margins. These results emphasize that the hybrid Pro-1 model is more effective in capturing the complex temporal dependencies of $PM_{2.5}$ levels, making it the best choice for accurate air pollution forecasting across diverse environments.

5.2.1 Quantitative analysis

In our analysis of RMSE, the author found a significant positive average percentage improvement across most models, except for the BiLSTM. The Pro-1 model showcases an impressive percentage improvement in RMSE compared to the CNN model, boasting a value of 79.909 ± 19.760 across the entire dataset. In contrast, the BiLSTM demonstrated much less variability, with a standard deviation of only ±9.206.

Comparing traditional models with our proposed model yields significant insights in the MAE analysis. The BiLSTM model recorded a minimal percentage improvement of -0.849, highlighting its close performance concerning our proposed model across various datasets. In contrast, the CNN model achieved an impressive average improvement of 98.039, demonstrating that our proposed model significantly outperforms the CNN. Furthermore, the CNN exhibited the highest standard deviation at ± 28.290, while the BiLSTM maintained the lowest standard deviation at ± 13.060 across all datasets, further reinforcing the robustness of our proposed model.

In the evaluation of MSE, the proposed model achieved an impressive minimum positive percentage improvement average of 3.855 with the GRU model, particularly excelling on the dataset labeled CNN 208. The standard deviations across all models evaluated over the 10 datasets were as follows: ±27.826 for BiLSTM, ±97.621 for CNN, ±37.264 for GRU, ±38.964 for LSTM, and ±28.379 for RNN. These results underscore the proposed model's strong performance and consistency.

Table 6 comparison between the proposed models demonstrates significant performance improvements, with Pro-2 consistently outperforming Pro-1 across all cities in RMSE, MAE, and MSE metrics. The most substantial improvements are observed in Kampala, where Pro-2 reduces RMSE by 1597.74%, MAE by 1470.80%, and MSE by 53925.87%, showcasing its superior predictive capability. Colombo also exhibits remarkable gains, with MSE improving by 16868.03%, indicating Pro-2's

Table 5 Performance comparison of Pro-1 model over traditional deep learning models using RMSE, MAE, and MSE of US embassies dataset

Error	Station	BiLSTM	%imp	CNN	%imp	GRU	%imp	LSTM	%imp	RNN	%imp	Pro-1
RMSE	AbuDhabi	4.08	**2.40%**	7.39	**85.59%**	3.91	-1.87%	5.46	**37.23%**	4.21	**5.79%**	**3.98**
	Beijing	**6.07**	-2.00%	11.05	**78.30%**	6.28	**1.32%**	6.87	**10.89%**	6.26	**1.00%**	6.20
	Colombo	3.47	**0.18%**	5.80	**67.42%**	3.56	**2.72%**	3.52	**1.63%**	3.57	**3.09%**	**3.46**
	Delhi	14.28	**2.72%**	28.11	**102.21%**	14.06	**1.16%**	16.24	**16.80%**	15.58	**12.04%**	**13.90**
	Dhaka	13.76	**0.91%**	27.39	**100.86%**	14.41	**5.69%**	14.68	**7.66%**	14.58	**6.94%**	**13.64**
	Jakarta	4.53	**0.46%**	8.22	**82.36%**	4.59	**1.73%**	4.54	**0.56%**	4.63	**2.56%**	**4.51**
	Kampala	9.32	-18.13%	17.04	**49.66%**	9.61	-15.65%	**9.05**	-20.50%	9.62	-15.54%	11.39
	KuwaitCity	8.32	**5.32%**	15.30	**93.66%**	8.23	**4.16%**	8.66	**9.58%**	8.51	**7.73%**	**7.90**
	Manama	4.73	**10.37%**	8.05	**87.92%**	5.68	**32.60%**	5.22	**21.77%**	5.31	**23.91%**	**4.29**
	Ulaanbaatar	20.26	-14.80%	34.74	**46.11%**	**20.19**	-15.09%	20.52	-13.70%	20.28	-14.73%	23.78
		Average	-1.27		**79.91**		**1.41**		**7.79**		**3.68**	
		Std Dev.	**± 9.21**		**±19.76**		**±12.98**		**±16.32**		**±11.48**	
MAE	AbuDhabi	2.33	**2.11%**	4.48	**96.10%**	**2.23**	-2.57%	3.42	**49.83%**	2.50	**9.25%**	2.29
	Beijing	**3.13**	-4.02%	6.77	**107.86%**	3.45	**5.96%**	3.86	**18.47%**	3.35	**2.72%**	3.26
	Colombo	2.05	**0.83%**	3.90	**91.39%**	2.17	**6.46%**	2.14	**5.08%**	2.19	**7.37%**	**2.04**
	Delhi	7.74	**6.02%**	17.87	**144.77%**	7.72	**5.70%**	9.67	**32.35%**	8.86	**21.33%**	**7.30**
	Dhaka	8.08	**2.55%**	17.56	**123.00%**	8.59	**9.12%**	8.90	**13.02%**	8.76	**11.18%**	**7.88**
	Jakarta	3.06	**0.15%**	6.17	**101.74%**	3.14	**2.74%**	**3.06**	-0.03%	3.15	**3.21%**	3.06
	Kampala	5.37	-23.03%	10.87	**55.77%**	5.97	-14.46%	**4.90**	-29.83%	5.99	-14.19%	6.98
	KuwaitCity	3.83	**10.91%**	7.29	**111.33%**	3.78	**9.54%**	4.10	**18.85%**	4.05	**17.30%**	**3.45**
	Manama	2.58	**15.44%**	3.99	**78.53%**	3.33	**49.10%**	2.92	**30.58%**	3.02	**34.90%**	**2.24**
	Ulaanbaatar	**9.38**	-20.45%	18.03	**52.90%**	9.84	-16.57%	9.86	-16.36%	9.97	-15.46%	11.79
		Average	-0.85		**96.04**		**4.10**		**7.89**		**7.14**	
		Std Dev.	**±13.06**		**±28.29**		**±17.35**		**±22.36**		**±17.25**	
MSE	AbuDhabi	16.67	**5.11%**	54.57	**244.04%**	**15.27**	-3.72%	33.49	**111.15%**	18.13	**14.28%**	15.86
	Beijing	**36.92**	-5.26%	122.11	**213.36%**	39.77	**2.07%**	47.76	**22.58%**	39.81	**2.16%**	38.97
	Colombo	12.04	**0.36%**	33.62	**180.32%**	12.68	**5.73%**	12.39	**3.32%**	12.75	**6.34%**	**11.99**
	Delhi	206.85	**6.24%**	790.27	**305.91%**	198.02	**1.71%**	270.57	**38.97%**	253.15	**30.03%**	**194.69**
	Dhaka	189.39	**1.82%**	750.34	**303.39%**	207.94	**11.79%**	219.05	**17.76%**	212.78	**14.39%**	**186.01**
	Jakarta	20.53	**0.93%**	67.67	**232.69%**	21.05	**3.49%**	20.57	**1.13%**	21.42	**5.29%**	**20.34**
	Kampala	87.28	-64.10%	290.67	**19.56%**	92.43	-61.98%	**82.06**	-66.25%	92.65	-61.89%	243.13
	KuwaitCity	69.46	**11.28%**	234.07	**274.99%**	67.72	**8.49%**	75.35	**20.72%**	72.59	**16.28%**	**62.42**
	Manama	22.81	**23.83%**	64.88	**252.17%**	39.85	**116.30%**	28.46	**54.50%**	29.52	**60.22%**	**18.42**
	Ulaanbaatar	410.48	-44.95%	1207.53	**61.95%**	**407.71**	-45.32%	422.30	-43.36%	411.17	-44.86%	745.64
		Average	-6.98		**208.14**		**3.86**		**16.05**		**4.57**	
		Std Dev.	**±27.83**		**±97.62**		**±37.26**		**±38.96**		**±28.38**	

Table 6 The comparison of the Pro-1 and Pro-2 with the percentage improvement

Models	RMSE			MAE			MSE		
	Pro-1	Pro-2	%imp	Pro-1	Pro-2	%imp	Pro-1	Pro-2	%imp
AbuDhabi	3.98	**0.483464**	**723.23%**	2.285	**0.33492**	**582.25%**	15.861	**0.233737**	**6685.89%**
Beijing	6.197	**0.69986**	**785.46%**	3.257	**0.392671**	**729.45%**	38.966	**0.489805**	**7855.46%**
Colombo	3.463	**0.265859**	**1202.57%**	2.037	**0.152063**	**1239.58%**	11.993	**0.070681**	**16868.03%**
Delhi	13.902	**1.309116**	**961.94%**	7.303	**0.982865**	**643.03%**	194.692	**1.713786**	**11260.39%**
Dhaka	13.637	**1.286225**	**960.23%**	7.876	**0.844181**	**832.98%**	186.009	**1.654375**	**11143.48%**
Jakarta	4.51	**0.453247**	**895.04%**	3.056	**0.276669**	**1004.57%**	20.338	**0.205432**	**9800.54%**
Kampala	11.389	**0.670834**	**1597.74%**	6.98	**0.444359**	**1470.80%**	243.126	**0.450018**	**53925.87%**
KuwaitCity	7.899	**3.1996**	**146.87%**	3.451	**1.122783**	**207.36%**	62.421	**10.23744**	**509.74%**
Manama	4.286	**0.679409**	**530.84%**	2.236	**0.342127**	**553.56%**	18.423	**0.461597**	**3891.26%**
Ulaanbaatar	23.778	**2.554386**	**830.87%**	11.792	**1.365496**	**763.57%**	745.637	**6.524888**	**11327.60%**

effectiveness in reducing large error magnitudes. Similarly, Delhi and Dhaka achieved over 960% RMSE improvement, highlighting Pro-2's robustness in high-pollution regions. Ulaanbaatar records 830.87% RMSE and 11327.60% MSE improvement, emphasizing its ability to handle extreme pollution variations. Even in cities with moderate pollution, such as Jakarta and Manama, Pro-2 shows 895% and 530% RMSE enhancements, respectively. These results confirm that Pro-2 significantly enhances air pollution forecasting accuracy, making it a highly efficient model for $PM_{2.5}$ prediction across diverse environments.

Furthermore, the Pro-1 model achieves a remarkable minimum RMSE of 3.463 in the Colombo dataset, while the Pro-2 model follows with an outstanding RMSE of 0.2658 in the same dataset. Furthermore, the Pro-1 model's highest RMSE value reaches 23.778, whereas the Pro-2 model maintains a significantly lower maximum of 2.554 in the Ulaanbaatar dataset. These results highlight the superior performance and reliability of the Pro-1 model, underscoring its effectiveness in various contexts.

5.2.2 Graphical analysis

The boxplot analysis of RMSE across different models confirms the superior performance of Pro-2, which consistently achieves the lowest RMSE across all cities, demonstrating enhanced prediction accuracy and stability. In high-pollution cities like Delhi, Dhaka, and Ulaanbaatar, Pro-2 significantly reduces error dispersion, outperforming traditional deep learning models and Pro-1. The Pro-2 model achieves impressive RMSE values ranging from 0.4 to 0.6, as clearly depicted in the red diagram of the Abu Dhabi dataset. In comparison, the overall RMSE variation for all traditional deep learning models and the proposed models across the same dataset spans a significant range from 0 to 12 in Fig. 7. This demonstrates the superior performance and consistency of the Pro-2 model in this analysis. The error spread in Pro-1 is narrower than in traditional models but still higher than in Pro-2, indicating that Pro-2 provides the most stable and reliable predictions. These results validate Pro-2 as the most effective model for minimizing RMSE and improving $PM_{2.5}$ forecasting.

The boxplot analysis of MAE clearly illustrates that Pro-2 surpasses both traditional deep learning models and Pro-1. Achieving the lowest MAE across all cities, Pro-2 showcases reduced error dispersion and significantly enhanced prediction accuracy. In high-pollution cities such as Delhi, Dhaka, and Ulaanbaatar, Pro-2 consistently outperforms its competitors. While Pro-1 shows some improvement over traditional models, it still demonstrates more significant variability. The MAE values for Pro-2, ranging from 0.08 to 0.12 as illustrated in the red diagram of the Colombo dataset, starkly contrast with the overall MAE variation of 0 to 4 observed in traditional and proposed models. The compelling evidence firmly establishes Pro-2 as the preferred choice for precise predictions. The spread of MAE values in Pro-2 is significantly narrower, indicating more excellent prediction stability and reliability. These findings confirm that Pro-2 is the most effective model for minimizing prediction errors and improving $PM_{2.5}$ forecasting accuracy across diverse urban environments Fig. 8.

The boxplot analysis of MSE across different models highlights the significant performance advantage of Pro-2, which consistently achieves the lowest MSE across all cities, indicating superior predictive accuracy and error reduction. Pro-2 significantly reduces error dispersion compared to traditional deep learning models and Pro-1 in high-pollution cities like Delhi, Dhaka, and Ulaanbaatar. In the Abu Dhabi dataset, the tail values approach and exceed zero. The values of interest are significantly closer to zero, revealing a clear pattern. The MSE ranges from 0 to 140. An analysis of the Pro-2 model indicates an impressive performance range of 0.1 to 0.4, surpassing all traditional deep learning models as well as the Pro-1 model, thereby demonstrating its potential for improved $PM_{2.5}$ predictive accuracy in Fig. 9. The variance in Pro-1 is lower than in traditional models but still higher than in Pro-2, reinforcing Pro-2's stability and reliability. Pro-2's lower MSE suggests better adaptability to

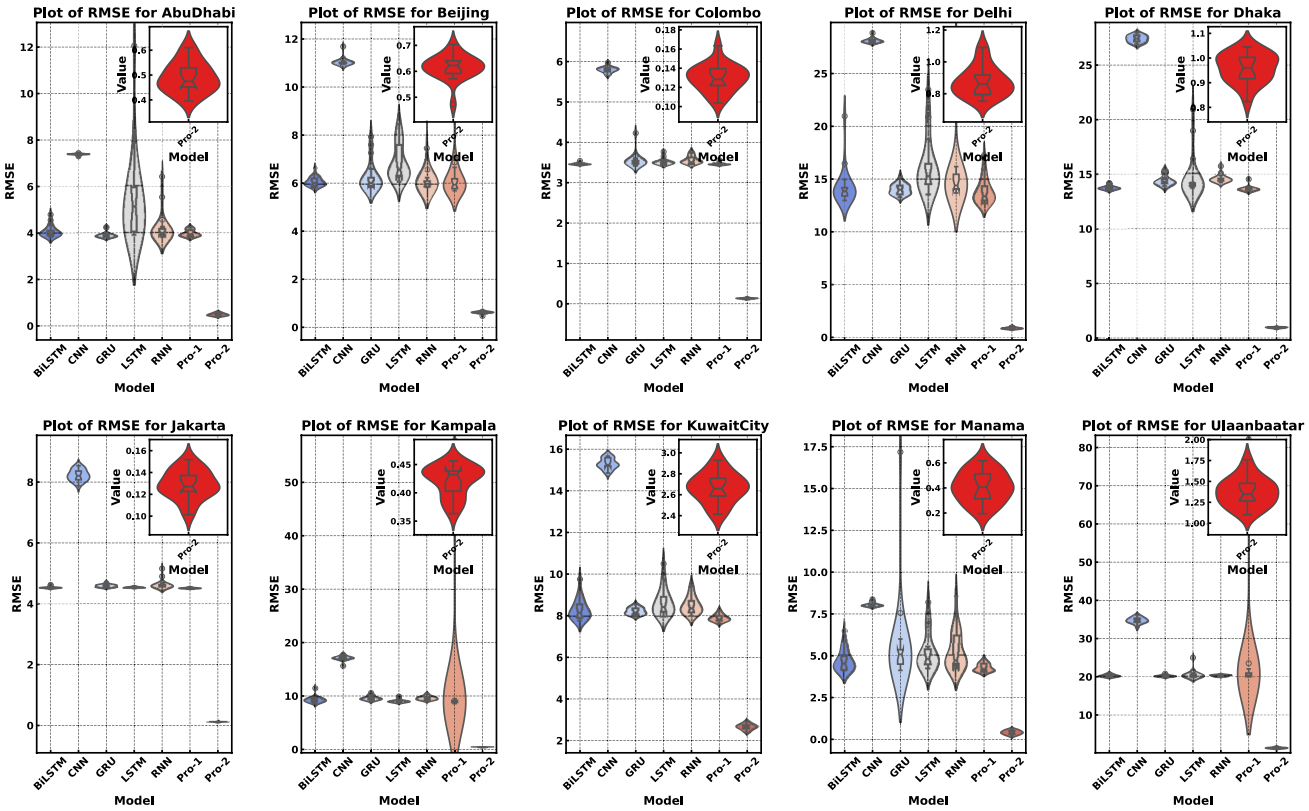


Fig. 7 Performance of Pro-1 and Pro-2 along with other traditional models based on RMSE measures

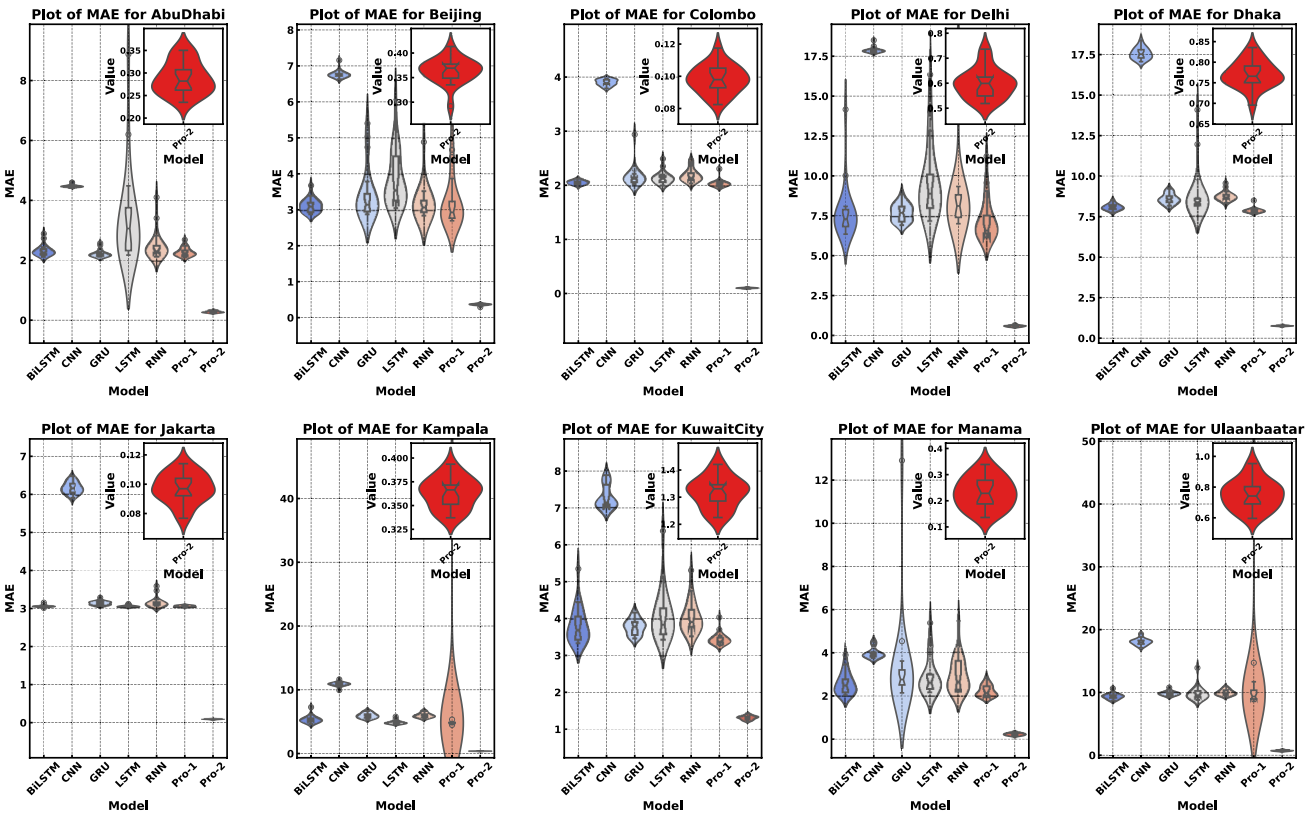


Fig. 8 Performance of Pro-1 and Pro-2 and other traditional models based on MAE measures

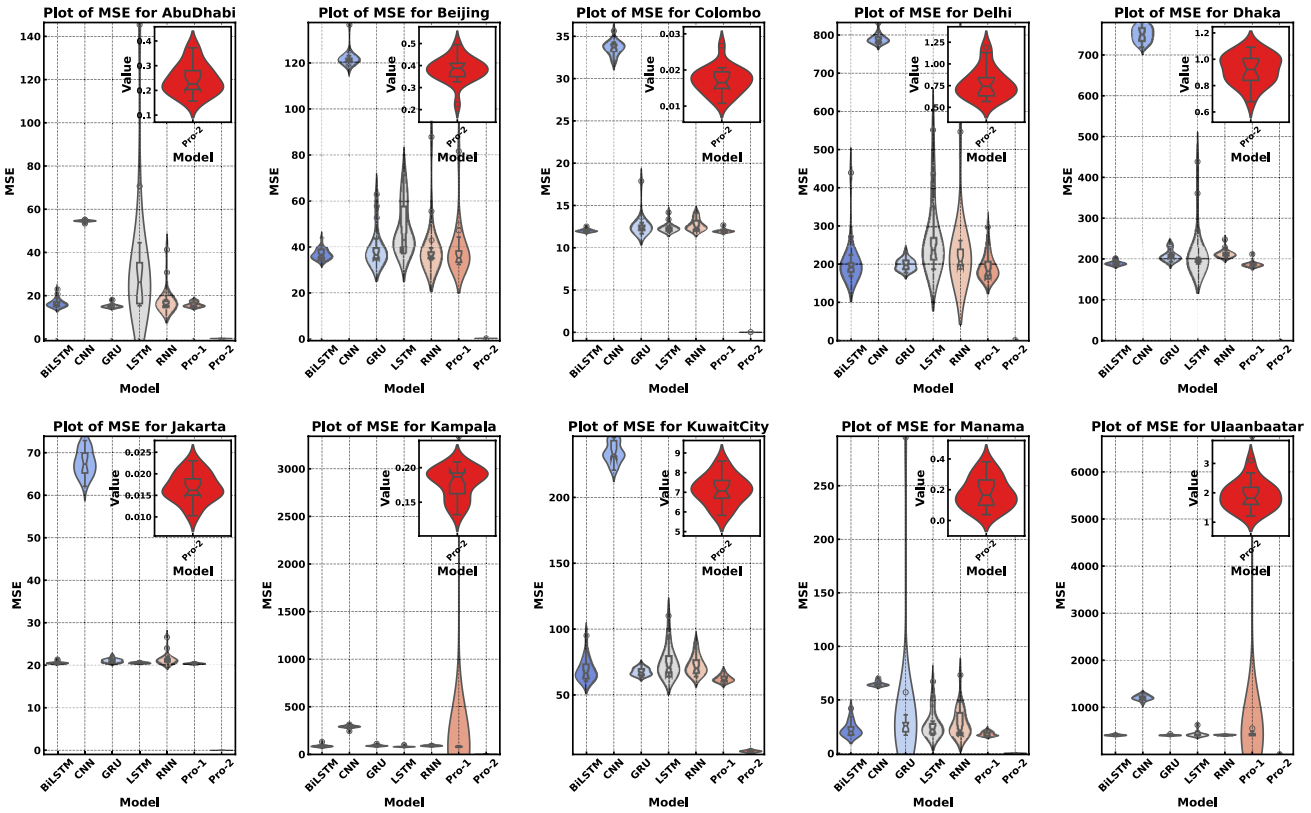


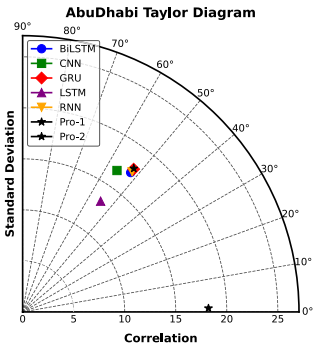
Fig. 9 Performance of Pro-1 and Pro-2 along with other traditional models based on MSE measures

extreme pollution variations, particularly in highly fluctuating regions such as Kampala and Kuwait City.

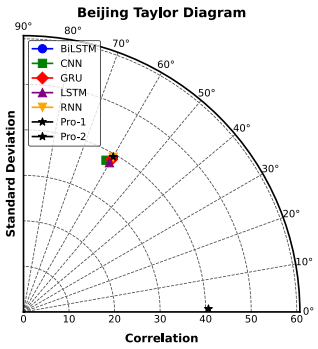
Figure 10 Taylor diagram is a graphical tool designed to visually evaluate and compare the performance of multiple models by integrating three key statistical metrics standard deviation, which measures the variability of the model relative to observations correlation coefficient, which quantifies how well the model captures the pattern of the observed data ranging from -1 to 1 and RMSE, which assesses the total difference between model outputs and observations. By combining these metrics into a single plot, the Taylor diagram provides a concise and intuitive way to identify which model performs closest to the actual observations, making it a valuable tool for model comparison and selection in fields like climate science, hydrology, and other data-driven disciplines.

The Taylor diagrams across multiple cities comprehensively visualize model performance regarding correlation, standard deviation, and RMSE. Pro-2 consistently achieves the highest correlation values and lowest standard deviation in all cities, indicating its superior predictive capability. Particularly in high-pollution regions like Delhi, Dhaka, and Ulaanbaatar, Pro-2 significantly improves over traditional models and Pro-1, with better alignment to observed values. In cities with moderate pollution variability, such as Beijing, Jakarta, and Kuwait City, Pro-2 maintains stable predictions with reduced deviation and improved correlation. Traditional models exhibit higher variance and lower correlation, reinforcing their limitations in capturing complex pollution dynamics. The Taylor diagrams confirm that Pro-2 provides the most accurate and reliable $PM_{2.5}$ forecasts, making it the most effective model for air pollution prediction across diverse urban environments.

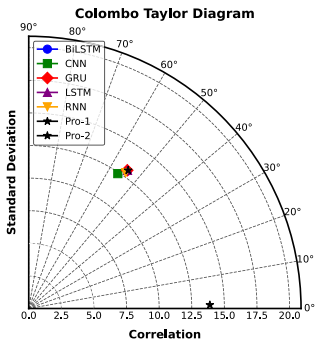
Figure 11 line plot comparison of Pro-1, Pro-2, and traditional models against actual $PM_{2.5}$ values across multiple cities highlights Pro-2's superior predictive accuracy. Pro-2 closely follows the actual $PM_{2.5}$ trends in all cities, showing minimal deviation, while traditional models exhibit higher discrepancies. Traditional models underestimate peak pollution levels in high-pollution cities like Delhi, Dhaka, and Ulaanbaatar, whereas Pro-2 captures extreme variations more effectively. Similarly, in cities with moderate pollution levels like Abu Dhabi, Beijing, and Kuwait City, Pro-2 demonstrates smoother trend alignment, while Pro-1 shows slight deviations. The CNN and RNN models show the highest errors, consistently diverging from the actual values. These results confirm that Pro-2 is the most reliable model for accurately capturing $PM_{2.5}$ fluctuations, making it the best choice for air pollution forecasting.



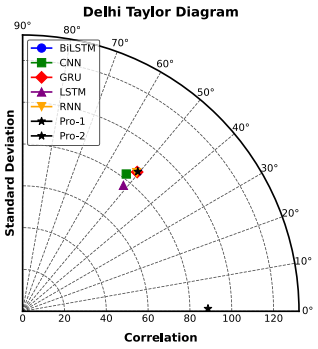
(a) Abu Dhabi



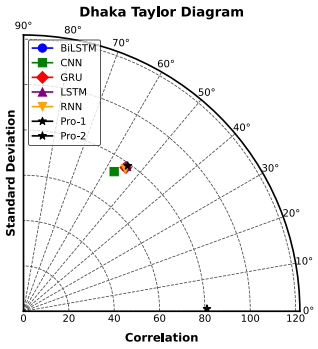
(b) Beijing



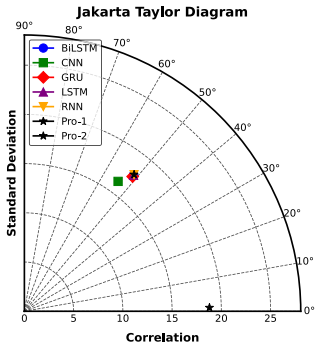
(c) Colombo



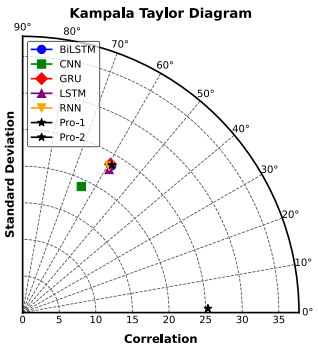
(d) Delhi



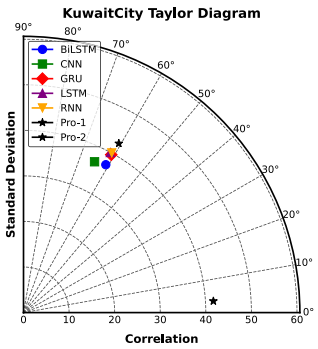
(e) Dhaka



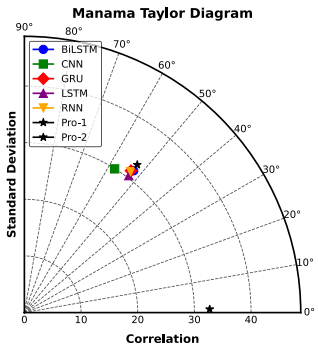
(f) Jakarta



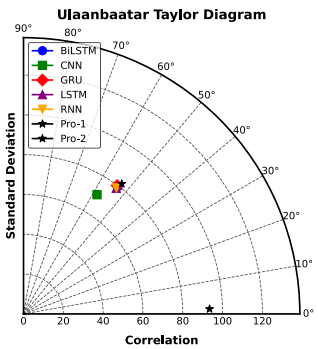
(g) Kampala



(h) Kuwait City



(i) Manama



(j) Ulaanbaatar

Fig. 10 The model's predictive correlation representation using Taylor diagrams of Pro-1, Pro-2, and traditional models across multiple cities

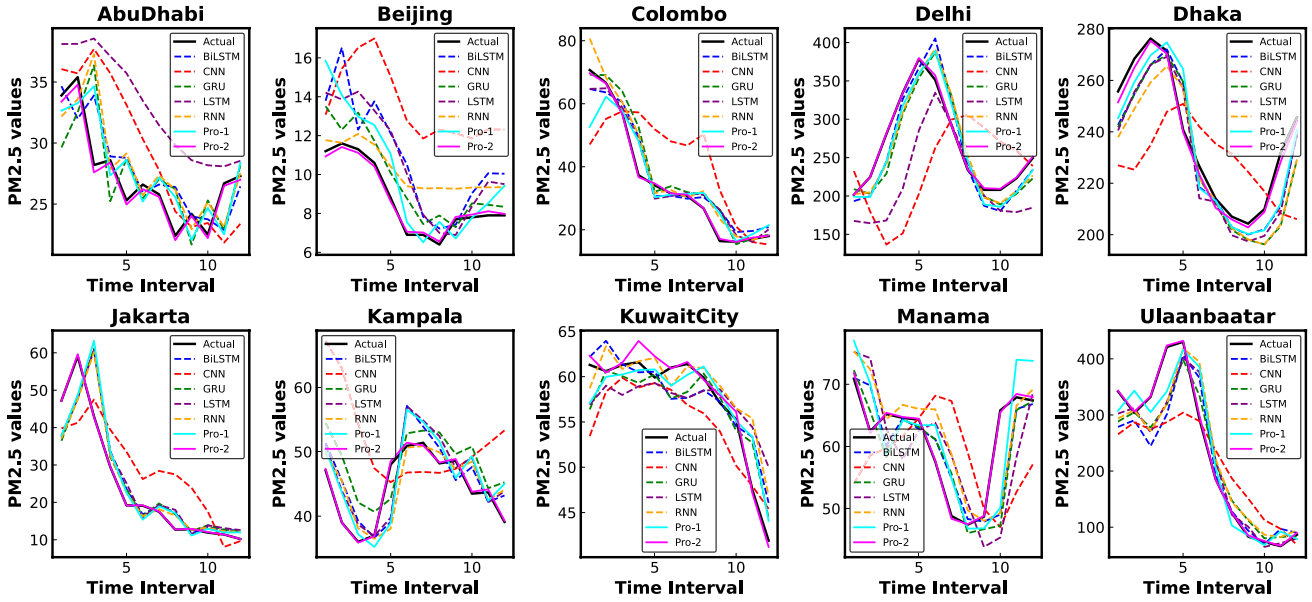


Fig. 11 Line plot of the Pro-1 and Pro-2 model vs. all traditional models

5.2.3 Statistical test analysis

The Diebold-Mariano test is a statistical tool used to compare the predictive accuracy of two competing models, determining whether the difference in their forecast performance is statistically significant. The null hypothesis (H_0) states no difference in predictive accuracy between the two models, while the alternative hypothesis (H_1) suggests that one model is significantly better. The test produces a DM statistic, and for statistical significance at the 5% level, this value must fall outside the range of -1.960 to +1.960. A negative DM statistic below -1.960 indicates that the first model, often referred to as the proposed model, is significantly better, whereas a positive DM statistic above +1.960 suggests that the first model is significantly worse than the second model. The test is particularly useful for evaluating the performance of forecasting models in practical applications.

The Diebold-Mariano test results in Table 7 indicate that Pro-2 significantly outperforms all traditional deep learning models and Pro-1 in predictive accuracy across all cities. The large negative test statistics suggest that the forecasting errors of Pro-2 are consistently lower than those of the compared models. The most significant performance improvements are observed in Colombo -43.69 vs. CNN, Abu Dhabi -44.58 vs. LSTM, and Dhaka -40.90 vs. CNN, highlighting Pro-2's ability to capture complex pollution patterns. Even in cities with relatively lower pollution levels, such as Kuwait City -12.35 vs. CNN and Manama -12.62 vs. CNN, Pro-2 maintains a clear advantage. These results confirm that Pro-2 provides statistically superior predictions, making it the most reliable model for forecasting across diverse environments for $PM_{2.5}$.

The AIC and BIC are essential tools for model selection, widely used in statistical modeling, time series forecasting, and machine learning to balance goodness of fit and model complexity. Both criteria focus on relative differences between models rather than their absolute values, meaning negative values are not problematic as long as comparisons are made within the same dataset. The negative values show that the MSE is much closer to zero for the same dataset of

Table 7 Diebold-Mariano test over the predictive values of Pro-1 and Pro-2 corresponding to traditional model and DL-DR models

	Cities	Pro-1	BiLSTM	CNN	GRU	LSTM	RNN
Pro-2 vs All Models	AbuDhabi	-36.2688	-37.3197	-41.277	-36.6913	-44.5782	-36.7837
	Beijing	-27.9301	-27.6176	-28.6948	-27.4595	-28.2126	-27.3008
	Colombo	-39.3649	-39.4862	-43.6853	-39.1349	-39.1185	-37.8586
	Delhi	-30.4897	-30.8523	-33.963	-30.7506	-31.9573	-30.2029
	Dhaka	-37.6685	-37.6204	-40.9041	-37.5694	-37.4656	-37.4757
	Jakarta	-35.437	-35.4159	-40.5382	-34.9687	-35.2201	-34.988
	Kampala	-24.254	-24.3078	-27.8496	-23.6846	-24.352	-23.849
	KuwaitCity	-11.5105	-11.8066	-12.3516	-11.433	-11.3878	-11.3433
	Manama	-11.6888	-11.7456	-12.62	-10.471	-12.5721	-10.7063
	Ulaanbaatar	-28.9818	-28.787	-32.5653	-28.7902	-28.8229	-28.7571

the proposed model. The key principle is that a lower AIC or BIC indicates a better model, regardless of whether the values are positive or negative. Differences more significant than 10 between models provide strong evidence favoring the model with the lower value, helping to identify the most suitable model for the given dataset. These criteria are invaluable for making informed decisions in model selection processes.

The AIC and BIC results significantly improve predictive performance for the Pro-2 model compared to traditional deep learning models and Pro-1. Across all cities, Pro-2 consistently achieves substantially lower AIC and BIC values, indicating superior model efficiency and better generalization. The most notable improvements are in Colombo AIC reduction of 75,206.6, BIC reduction of 75,206.6, and Abu Dhabi AIC and BIC reduction of over 72,986, highlighting pro-2's strong performance in diverse environments. Similarly, high-pollution cities like Delhi and Dhaka exhibit reductions of over 78,000 in AIC and BIC, reinforcing Pro-2's ability to handle complex pollution dynamics, see Table 8. These findings confirm that Pro-2 is the most effective model, significantly outperforming all traditional approaches by minimizing information loss and improving predictive accuracy.

The Friedman test is a non-parametric statistical test used to identify significant differences in performance across multiple models when evaluated over multiple datasets. If the Friedman test reveals a significant difference (p-value < 0.05), post hoc tests such as Holm correction are applied to determine which specific pairs of models differ significantly. To construct a post hoc comparison table for $\alpha = 0.05$, the table typically includes pairwise comparisons between models, the test ranking statistic, the p-value, and an indication of whether the difference is statistically significant. A p-value < 0.05 signifies a statistically significant difference. After applying the Friedman test, average ranks for each model across all datasets are calculated, with the model achieving the highest rank closest to 1 being identified as the best-performing model. The structured approach helps in systematically comparing and selecting the most effective model.

Table 9 Friedman post hoc comparison test results confirm the statistical superiority of Pro-2 over all other models BiLSTM, CNN, GRU, LSTM, RNN, and Pro-1 in RMSE, MAE, and MSE metrics. Pro-2 ranks 1st in all three error measures, reinforcing its robustness and accuracy. The CNN model ranks the lowest, 7th with a p-value of 0.00, indicating the weakest predictive performance. The RNN and LSTM models also show significantly higher errors with p-values close to zero, confirming their lower reliability. While Pro-1 performs better than traditional models, its p-value of 0.038434 for RMSE and 0.029727 for MAE suggests it is statistically inferior to Pro-2. These results validate that Pro-2 is the most statistically significant model, consistently delivering superior prediction accuracy across all key metrics.

The comparison Table 10 presents the performance of the proposed Pro-2 model against five state-of-the-art models across ten cities using RMSE, MAE, and MSE metrics. Numerically, the Pro-2 model outperforms all competing models with significantly lower error values across almost all cities. For instance, in highly polluted cities like Delhi, Dhaka, and Ulaanbaatar, Pro-2 achieves drastically reduced RMSEs (1.309, 1.286, and 2.554 respectively) compared to others (e.g., up to 115.107 RMSE in Delhi by Hu et al. 2023). The same trend holds for MAE and MSE, where Pro-2 consistently maintains the lowest errors, indicating superior predictive accuracy and robustness. Especially in cities like Abu Dhabi and Colombo, Pro-2 shows exceptional performance with RMSEs of 0.483 and 0.266, respectively, and corresponding MSEs of just 0.234 and 0.071, demonstrating its effectiveness in both moderately and highly polluted regions. These numerical results underscore the model's capability to generalize well across diverse geographic and pollution profiles.

5.2.4 Overall conclusion of analysis

Pro-2 consistently outperforms all models, achieving the lowest RMSE, MAE, and MSE across all cities, as demonstrated by Friedman's post hoc test, which ranked 1st with statistically significant improvements. The model ranking was $Pro\text{-}2 > Pro\text{-}1 > \text{BiLSTM} > \text{GRU} > \text{LSTM} > \text{RNN} > \text{CNN}$ CNN to the RMSE. The Diebold-Mariano test further confirms that Pro-2 significantly reduces forecasting errors, with the most critical enhancements observed in highly polluted cities such as Delhi, Dhaka, and Ulaanbaatar. AIC and BIC tests indicate Pro-2's superior model efficiency, achieving reductions of over 75,000 in Colombo and 72,000 in Abu Dhabi, demonstrating better generalization. The Taylor diagrams validate Pro-2's higher correlation and reduced standard deviation, proving its stability and accuracy across diverse pollution patterns. Line plot comparisons show that Pro-2 closely follows actual $PM_{2.5}$ variations, whereas traditional models exhibit significant discrepancies. These results establish the proposed models as the most effective predictive model for $PM_{2.5}$ forecasting, offering improved accuracy and reliability for air pollution monitoring. The model can be utilized for urban air pollution forecasting.

The 3D bar plot in Fig. 12 presents the Friedman ranking analysis of five $PM_{2.5}$ prediction state-of-the-art models (Vo et al. 2023; Sun et al. 2021; Hu et al. 2023; Govande et al. 2025; Rad et al. 2025), to the proposed Pro-2 (Final Model) evaluated across three performance metrics: RMSE,

Table 8 AIC and BIC test over the predictive values of the traditional vs. Pro-2 model

City	Model	Trad. AIC	AIC Pro-2	diff	Trad. BIC	BIC Pro-2	diff
AbuDhabi	Pro-1	54754.5	-18231.9261	**72986.43**	54790.51	-18195.915	**72986.43**
	CNN	54502.75	-18231.9261	**72734.68**	54538.76	-18195.915	**72734.68**
	BiLSTM	55559.3	-18231.9261	**73791.22**	55595.31	-18195.915	**73791.22**
	GRU	54699.76	-18231.9261	**72931.68**	54735.77	-18195.915	**72931.68**
	LSTM	54861.68	-18231.9261	**73093.61**	54897.69	-18195.915	**73093.61**
	RNN	54568.13	-18231.9261	**72800.06**	54604.14	-18195.915	**72800.06**
Beijing	Pro-1	73149.22	-5838.97939	**78988.2**	73185.23	-5802.9678	**78988.2**
	CNN	73083.34	-5838.97939	**78922.32**	73119.36	-5802.9678	**78922.32**
	BiLSTM	73260.28	-5838.97939	**79099.26**	73296.29	-5802.9678	**79099.26**
	GRU	72997.39	-5838.97939	**78836.36**	73033.4	-5802.9678	**78836.36**
	LSTM	72829.73	-5838.97939	**78668.71**	72865.74	-5802.9678	**78668.71**
	RNN	73108.73	-5838.97939	**78947.71**	73144.74	-5802.9678	**78947.71**
Colombo	Pro-1	49848.4	-25358.1938	**75206.6**	49884.41	-25322.182	**75206.6**
	CNN	49821.06	-25358.1938	**75179.26**	49857.07	-25322.182	**75179.26**
	BiLSTM	50183.95	-25358.1938	**75542.15**	50219.96	-25322.182	**75542.15**
	GRU	49901.44	-25358.1938	**75259.63**	49937.45	-25322.182	**75259.63**
	LSTM	49877.04	-25358.1938	**75235.23**	49913.05	-25322.182	**75235.23**
	RNN	49748.79	-25358.1938	**75106.99**	49784.8	-25322.182	**75106.99**
Delhi	Pro-1	85545.29	5908.974783	**79636.31**	85581.3	5944.9863	**79636.31**
	CNN	85533.91	5908.974783	**79624.93**	85569.92	5944.9863	**79624.93**
	BiLSTM	85974.12	5908.974783	**80065.15**	86010.13	5944.9863	**80065.15**
	GRU	85533.61	5908.974783	**79624.64**	85569.63	5944.9863	**79624.64**
	LSTM	84955.18	5908.974783	**79046.21**	84991.19	5944.9863	**79046.21**
	RNN	85499.35	5908.974783	**79590.37**	85535.36	5944.9863	**79590.37**
Dhaka	Pro-1	85204.74	6569.46142	**78635.28**	85240.75	6605.473	**78635.28**
	CNN	85177.68	6569.46142	**78608.22**	85213.69	6605.473	**78608.22**
	BiLSTM	85442.84	6569.46142	**78873.38**	85478.85	6605.473	**78873.38**
	GRU	85129.56	6569.46142	**78560.1**	85165.57	6605.473	**78560.1**
	LSTM	85228.87	6569.46142	**78659.4**	85264.88	6605.473	**78659.4**
	RNN	85029.3	6569.46142	**78459.84**	85065.31	6605.473	**78459.84**
Jakarta	Pro-1	54797.61	-18999.0604	**73796.67**	54833.62	-18963.049	**73796.67**
	CNN	54795.81	-18999.0604	**73794.87**	54831.82	-18963.049	**73794.87**
	BiLSTM	55121.65	-18999.0604	**74120.71**	55157.67	-18963.049	**74120.71**
	GRU	54657.34	-18999.0604	**73656.4**	54693.35	-18963.049	**73656.4**
	LSTM	54753.85	-18999.0604	**73752.91**	54789.86	-18963.049	**73752.91**
	RNN	54779.84	-18999.0604	**73778.9**	54815.86	-18963.049	**73778.9**
Kampala	Pro-1	63064.96	-7690.97325	**70755.94**	63100.98	-7654.9617	**70755.94**
	CNN	63020	-7690.97325	**70710.97**	63056.01	-7654.9617	**70710.97**
	BiLSTM	63474.1	-7690.97325	**71165.07**	63510.11	-7654.9617	**71165.07**
	GRU	63233.34	-7690.97325	**70924.31**	63269.35	-7654.9617	**70924.31**
	LSTM	62824.13	-7690.97325	**70515.1**	62860.14	-7654.9617	**70515.1**
	RNN	63164.39	-7690.97325	**70855.37**	63200.4	-7654.9617	**70855.37**
KuwaitCity	Pro-1	74158.77	21146.04746	**53012.73**	74194.78	21182.059	**53012.73**
	CNN	72960.83	21146.04746	**51814.78**	72996.84	21182.059	**51814.78**
	BiLSTM	73874.98	21146.04746	**52728.93**	73910.99	21182.059	**52728.93**
	GRU	73517.8	21146.04746	**52371.75**	73553.81	21182.059	**52371.75**
	LSTM	73771.4	21146.04746	**52625.35**	73807.41	21182.059	**52625.35**
	RNN	73660.5	21146.04746	**52514.45**	73696.51	21182.059	**52514.45**
Manama	Pro-1	66832.96	-8875.4598	**75708.42**	66868.97	-8839.4483	**75708.42**
	CNN	66341.44	-8875.4598	**75216.9**	66377.45	-8839.4483	**75216.9**
	BiLSTM	67698.43	-8875.4598	**76573.89**	67734.44	-8839.4483	**76573.89**
	GRU	66510.41	-8875.4598	**75385.87**	66546.42	-8839.4483	**75385.87**
	LSTM	66116.96	-8875.4598	**74992.42**	66152.97	-8839.4483	**74992.42**
	RNN	66488.28	-8875.4598	**75363.74**	66524.3	-8839.4483	**75363.74**
Ulaanbaatar	Pro-1	86584.95	20280.49455	**66304.46**	86620.96	20316.506	**66304.46**

Table 8 (continued)

City	Model	Trad. AIC	AIC Pro-2	diff	Trad. BIC	BIC Pro-2	diff
	CNN	86569	20280.49455	**66288.5**	86605.01	20316.506	**66288.5**
	BiLSTM	87494.31	20280.49455	**67213.81**	87530.32	20316.506	**67213.81**
	GRU	86769.13	20280.49455	**66488.64**	86805.14	20316.506	**66488.64**
	LSTM	86543.7	20280.49455	**66263.21**	86579.72	20316.506	**66263.21**
	RNN	86559.92	20280.49455	**66279.43**	86595.93	20316.506	**66279.43**

Table 9 Post Hoc Comparison Table for $\alpha = 0.05$ (FRIEDMAN) RMSE, MAE, and MSE

Models	RMSE			MAE			MSE		
	Rank	p value	p_{Holm}	Rank	p value	p_{Holm}	Rank	p value	p_{Holm}
BiLSTM	3.2	0.022773	0.025	3.2	0.022773	0.025	3.2	0.022773	0.025
CNN	7	0	0.008333	7	0	0.008333	7	0	0.008333
GRU	3.89	0.002684	0.016667	4	0.001901	0.016667	3.8	0.003752	0.016667
LSTM	4.9	0.000054	0.0125	4.6	0.000194	0.0125	4.9	0.000054	0.0125
RNN	5	0.000035	0.01	5.1	0.000022	0.01	5.1	0.000022	0.01
Pro-1	3	0.038434	0.05	3.1	0.029727	0.05	3	0.038434	0.05
Pro-2	1	———	———	1	———	——-	1	———	———

MAE, and MSE. The Friedman test, a non-parametric statistical method for comparing multiple models over multiple datasets, assigns lower ranks to better-performing models. The Pro-2 model consistently achieves the best rank of 1.0 across all three metrics, demonstrating statistically superior and robust performance. In contrast, Vo et al. (2023) rank the worst, with scores near 4.9 for all parameters, while the remaining models exhibit moderate performance with ranks ranging from approximately 3.2 to 4.2. This analysis confirms the dominance of the final model in predictive accuracy and generalization ability for air pollution forecasting tasks.

6 Limitations and potential applications

The limitations outlined below are now explicitly addressed:

- **Data dependency:** The model's performance is dependent on the quality and availability of high-resolution air pollution and meteorological datasets. Inaccurate or missing data may lead to reduced predictive accuracy and misinformed decisions.
- **Computational cost:** The hybrid architecture, especially with residual correction using XGBoost, can increase training and inference time, which may limit real-time deployment on resource-constrained systems. Optimization strategies or model pruning may be necessary for edge deployment.
- **Generalization to rural areas:** The model was trained and validated on data from urban U.S. embassy stations, so its generalizability to rural or less-monitored regions may require further validation. Transfer learning or domain adaptation techniques could help improve performance in these regions.
- **Temporal drift and changing pollution patterns:** The model may not adapt well to sudden policy changes, industrial shifts, or climate anomalies that significantly alter pollution dynamics. Periodic retraining or the use of adaptive learning algorithms may be necessary. The inclusion or exclusion of certain meteorological and emission-related features can significantly impact model accuracy. A systematic feature importance or selection framework is required to ensure robust performance.
- **Limited interpretability:** Deep learning models often act as black boxes, making it difficult to interpret or explain predictions to policymakers or non-technical stakeholders. The applications we explored are not only promising but also crucial for our future endeavors and must be prioritized:
- **Urban air quality monitoring systems:** Deployment in real-time forecasting systems to aid city-level environmental decision-making. This can assist municipal authorities in issuing timely pollution warnings and traffic control advisories. Embedding the forecasting pipeline into IoT-based urban systems for dynamic environmental response. This integration allows for adaptive control of ventilation, traffic, or industrial operations based on predicted pollution levels.
- **Health risk assessment tools:** Integrating model predictions with public health platforms to inform exposure risk and emergency alerts. It can support the development of targeted interventions for vulnerable populations such as children and the elderly.

Table 10 Comparison of the state-of-the-art models with Pro-2 (final model) using RMSE, MAE, and MSE metrics across 10 cities

Dataset	RMSE						MAE						MSE					
	Rad et al. (2025)	Vo et al. (2023)	Gov-ande et al. (2025)	Hu et al. (2023)	Sun et al. (2021)	Pro-2	Rad et al. (2025)	Vo et al. (2023)	Gov-ande et al. (2025)	Hu et al. (2023)	Sun et al. (2021)	Pro-2	Rad et al. (2025)	Vo et al. (2023)	Gov-ande et al. (2025)	Hu et al. (2023)	Sun et al. (2021)	Pro-2
AbuDhabi	5.102	10.844	6.350	4.606	4.465	0.483	3.041	6.665	4.670	2.690	2.547	0.335	26.029	117.587	40.324	21.219	19.937	0.234
Beijing	8.139	17.463	10.502	6.968	106.181	0.700	4.531	10.659	7.435	3.377	98.098	0.393	66.248	304.959	110.296	48.553	11274.362	0.490
Colombo	21.943	7.523	3.792	13.148	3.689	0.266	18.027	4.783	2.317	10.332	2.169	0.152	481.482	56.599	14.379	172.871	13.609	0.071
Delhi	18.518	63.269	26.750	115.107	100.888	1.309	10.543	37.712	17.614	75.503	63.797	0.983	342.931	4002.954	715.557	13249.683	10178.466	1.714
Dhaka	19.264	48.519	19.951	76.322	78.669	1.286	11.459	30.158	13.059	51.052	51.495	0.844	371.116	2354.046	398.027	5825.079	6188.752	1.654
Jakarta	4.840	12.301	5.486	4.934	5.617	0.453	3.305	8.872	3.963	3.365	3.964	0.277	23.422	151.302	30.101	24.347	31.550	0.205
Kampala	10.521	23.196	11.087	10.050	31.822	0.671	6.221	13.244	7.671	5.313	21.830	0.444	110.684	538.075	122.919	101.009	1012.625	0.450
KuwaitCity	9.695	19.619	8.062	24.205	52.124	3.200	4.441	9.567	3.623	15.597	46.288	1.123	93.998	384.924	64.995	585.876	2716.899	10.237
Manama	5.341	10.777	5.829	27.281	4.293	0.679	2.707	5.587	4.170	23.200	2.017	0.342	28.523	116.135	33.980	744.262	18.429	0.462
Ulaanbaatar	21.690	49.997	24.921	21.677	21.315	2.554	9.171	22.876	11.141	8.878	8.722	1.365	470.474	2499.661	621.064	469.872	454.345	6.525

7 Policy suggestions

A multi-faceted approach is required to effectively mitigate $PM_{2.5}$ pollution, incorporating real-time monitoring, predictive analytics, regulatory enforcement, and sustainable urban development. The following policy recommendations aim to enhance air quality management and public health protection:

- **Strengthening air quality monitoring and data integration** Governments should expand real-time air quality monitoring networks by integrating satellite data, meteorological models, and urban sensor networks. Advanced machine learning techniques should be employed to improve forecasting accuracy, enabling proactive pollution control measures.
- **AI-Driven early warning systems and public health alerts** Implement AI-based early warning systems that provide real-time pollution forecasts and health alerts, particularly for vulnerable populations such as children, the elderly, and individuals with respiratory conditions. Mobile applications and digital platforms should disseminate actionable air quality advisories.
- **Strict emission regulations and industrial control measures** Enforce stringent industrial emission standards, upgrade vehicle emission norms to Euro VI-equivalent or higher, and incentivize the transition to cleaner fuels and electric mobility. Regular audits and real-time emission tracking should be mandated for industries and transport sectors.
- **Sustainable urban planning and green infrastructure** Promote environmentally conscious urban planning by increasing green spaces, developing low-emission zones, and encouraging non-motorized transport infrastructure such as pedestrian-friendly streets and cycling lanes. Adopting green roofs and urban forests should be incentivized to act as natural air purifiers.
- **Public awareness and community engagement** Launch large-scale public awareness campaigns on the health risks of $PM_{2.5}$ exposure and promote behavioral changes such as reducing biomass burning and adopting energy-efficient appliances. Citizen science initiatives should be encouraged, allowing communities to participate in air quality monitoring and advocacy efforts.

8 Conclusion

Air pollution is a crucial problem for humans and the atmosphere. Air pollution $PM_{2.5}$ has the most detrimental health effects on humans and the environment. $PM_{2.5}$ can have long-term and short-term health effects on individuals of

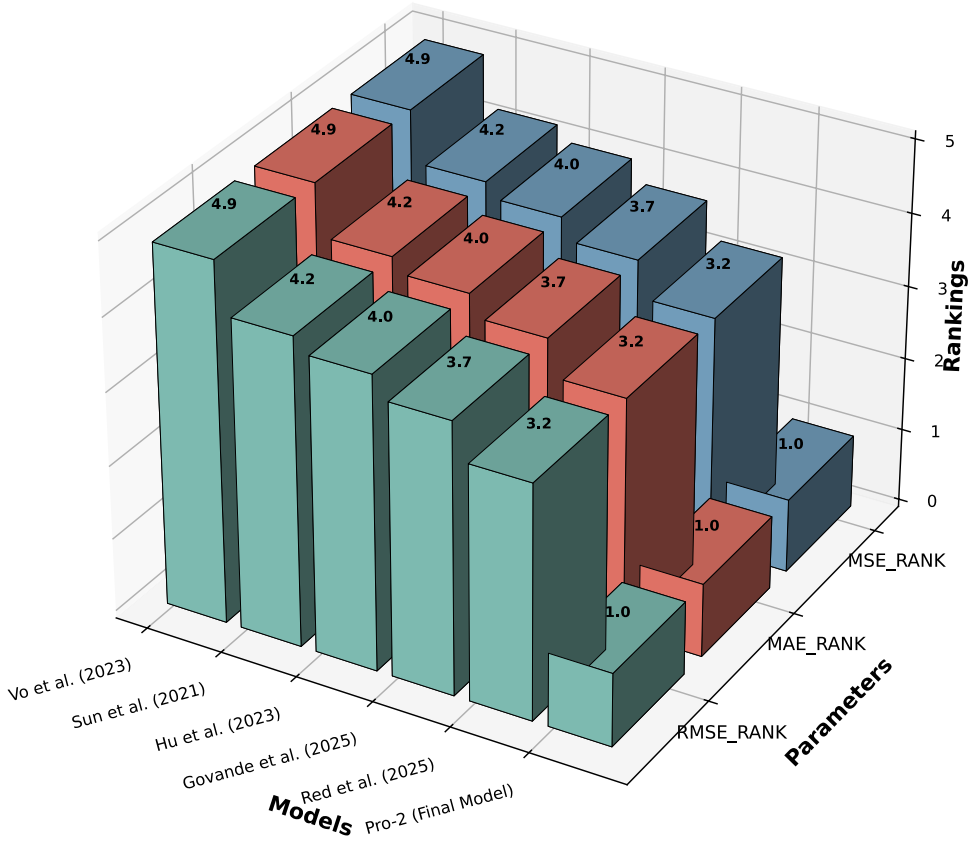


Fig. 12 Friedman ranking comparison of state-of-the-art $PM_{2.5}$ prediction models over the final proposed model using RMSE, MAE, and MSE metrics

all ages. The work is crucial for accurately forecasting air pollution, which is $PM_{2.5}$. The study uses a comprehensive dataset from US Embassies and official monitoring stations to evaluate $PM_{2.5}$ prediction across multiple cities. In this innovative study, we introduced models integrating a WaveNet architecture with two 1DCNN branches, connecting them to a BiLSTM layer for precise predictions. The Pro-1 model significantly enhances accuracy by optimizing errors through XGBoost for residual forecasting, referred to as Pro-2. In a comprehensive comparison, our models outperformed traditional deep learning approaches, including BiLSTM, CNN, GRU, LSTM, and RNN. Pro-1 achieved an impressive RMSE of 79.909 ± 19.760, with a remarkable minimum RMSE of 3.463 in the Colombo dataset. Pro-2 further excelled, achieving an RMSE of just 0.2658. Pro-1 peaked at 23.778 in the Ulaanbaatar dataset, while Pro-2 reached a superior 2.554. Pro-2 consistently set new standards, demonstrating the lowest RMSE, MAE, and MSE across all cities. The performance was validated by Friedman's post hoc test, confirming significant advantages. The RMSE ranking is as follows: $Pro\text{-}2 > Pro\text{-}1 > \text{BiLSTM} > \text{GRU} > \text{LSTM} > \text{RNN} > \text{CNN}$. The authors employed the Diebold-Mariano test, the AIC-BIC test, and Taylor diagrams to further substantiate our models. Friedman ranking comparison of state-of-the-art $PM_{2.5}$ prediction models shows that the final proposed model ranked first across all cases when evaluated using RMSE, MAE, and MSE metrics. These findings establish our models as leading predictive tools for $PM_{2.5}$ forecasting, providing exceptional accuracy and reliability for monitoring urban air pollution. Future work can focus on integrating real-time sensor data and advanced feature selection techniques to enhance prediction accuracy further.

Acknowledgements We acknowledge the AirNow platform for providing access to air quality data from U.S. embassies and consulates.

Author Contributions Naushad Ahmad: Conceptualization, Methodology, Software, Formal analysis, Data Curation, Writing - Original Draft, Visualization. Vipin Kumar: Methodology, Validation, Formal analysis, Investigation, Resources, Writing - Review & Editing, Supervision.

Funding No funding was received for conducting this study.

Data Availability No datasets were generated or analysed during the current study.

Code Availability All code was implemented in Python. Interested researchers can request access by contacting the corresponding author.

Declarations

Competing Interests The authors declare no competing interests.

Ethics Approval and Consent to Participate Not applicable

Consent for Publication Not applicable

References

Agbehadji IE, Obagbuwa IC (2024) Systematic review of machine learning and deep learning techniques for spatiotemporal air quality prediction. Atmosphere 15(11):1352. https://doi.org/10.3390/atmos15111352

Ahmad N, Kumar V (2023) A hybrid time series model for the spatio-temporal analysis of air pollution prediction based on pm 2.5. In: International conference on advanced network technologies and intelligent computing, pp 62–81. https://doi.org/10.1007/978-3-031-64067-4_5. Springer

Ahmad N, Kumar V (2025a) Enhancing pm 2.5 air pollution forecasting with novel random imputation based on hybrid rnn-bidirectional gru (nri rnn-bigru) model. SN Comput Sci 6(6):637. https://doi.org/10.1007/s42979-025-04167-y

Ahmad N, Kumar V (2025b) Spatio-temporal forecasting using a hybrid bigru-1dcnn model for pm 2.5 concentrations in Delhi, India (2018-2023) across multiple monitoring stations. Water Air Soil Pollut 236(7):459. https://doi.org/10.1007/s11270-025-08103-x

Amin M, Ramadhani AAT, Putri RM, Auliani R, Torabi SE, Hanami ZA, Suryati I, Bachtiar VS (2025) A review of particulate matter (pm) in Indonesia: trends, health impact, challenges, and options. Environ Monit Assess 197(1):1–19. https://doi.org/10.1007/s10661-024-13426-z

Ansari A, Quaff AR (2025) Data-driven analysis and predictive modelling of hourly air quality index (aqi) using deep learning techniques: a case study of Azamgarh, India. Theor Appl Climatol 156(1):74. https://doi.org/10.1007/s00704-024-05304-y

Arooj F, Mushtaq K, Iftikhar A, Luqman M, Rehman SA, Naseer R, Safdar S (2025) Comparative assessment of ambient air quality in large and small ruminant farms: insights from Lahore and Pattoki. J Wildl Biodivers 9(1):314–331. https://doi.org/10.5281/zenodo.14635192

Bekhor S, Broday DM et al (2013) Data-driven nonlinear optimisation of a simple air pollution dispersion model generating high resolution spatiotemporal exposure. Atmos Environ 79:261–270. https://doi.org/10.1016/j.atmosenv.2013.06.005

Belmahdi B, Louzazni M, Bouardi AE (2020) A hybrid Arima-Ann method to forecast daily global solar radiation in three different cities in Morocco. Eur Phys J Plus 135(11):925. https://doi.org/10.1140/epjp/s13360-020-00920-9

Blanco G, Barco L, Innocenti L, Rossi C (2024) Urban air pollution forecasting: a machine learning approach leveraging satellite observations and meteorological forecasts. https://doi.org/10.48550/arXiv.2405.19901

Bosco JJ, Kowsalya V (2024) A novel approach for air pollution prediction using machine learning techniques. In: 2024 Third International Conference on Electrical, Electronics, Information and Communication Technologies (ICEEICT), pp 1–6. https://doi.org/10.1109/ICEEICT61591.2024.10718439 . IEEE

Che J, Hu K, Xia W, Xu Y, Li Y (2024) Short-term air quality prediction using point and interval deep learning systems coupled with multi-factor decomposition and data-driven tree compression. Appl Soft Comput 166:112191. https://doi.org/10.1016/j.asoc.2024.112191

Chen G, Chen S, Li D, Chen C (2025) A hybrid deep learning air pollution prediction approach based on neighborhood selection and spatio-temporal attention. Sci Rep 15(1):3685. https://doi.org/10.1038/s41598-025-88086-1

Chiroma H, Abdul-kareem S, Shukri Mohd Noor A, Abubakar AI, Sohrabi Safa N, Shuib L, Fatihu Hamza M, Ya'u Gital A, Herawan T (2016) A review on artificial intelligence methodologies for the forecasting of crude oil price. Intell Autom Soft Comput 22(3):449–462. https://doi.org/10.1080/10798587.2015.1092338

Duarte EdSF, Lucio PS, Henriques-Rodrigues L, Costa MJ (2025) Impact of monthly air pollution and weather conditions on cardiorespiratory mortality in Portuguese metropolitan areas. Sci Rep 15(1):4147. https://doi.org/10.1038/s41598-025-88473-8

Essamlali I, Nhaila H, El Khaili M (2024) Supervised machine learning approaches for predicting key pollutants and for the sustainable enhancement of urban air quality: a systematic review. Sustainability 16(3):976. https://doi.org/10.3390/su16030976

Gangwar A, Singh S, Mishra R, Prakash S (2023) The state-of-the-art in air pollution monitoring and forecasting systems using iot, big data, and machine learning. Wirel Pers Commun 130(3):1699–1729. https://doi.org/10.1007/s11277-023-10351-1

Gocheva-Ilieva S, Ivanov A (2019) Assaying sarima and generalised regularised regression for particulate matter pm10 modelling and forecasting. Int J Environ Pollut 66(1–3):41–62. https://doi.org/10.1504/IJEP.2019.104520

Govande A, Attada R, Shukla KK (2025) Predicting pm2.5 levels over Indian metropolitan cities using recurrent neural networks. Earth Sci Inf 18(1):1–16. https://doi.org/10.1007/s12145-024-01491-4

Guo Q, He Z, Li S, Li X, Meng J, Hou Z, Liu J, Chen Y (2020) Air pollution forecasting using artificial and wavelet neural networks with meteorological conditions. Aerosol Air Qual Res 20(6):1429–1439. https://doi.org/10.4209/aaqr.2020.03.0097

Guo Q, He Z, Wang Z (2023a) Prediction of monthly average and extreme atmospheric temperatures in Zhengzhou based on artificial neural network and deep learning models. Front For Glob Change 6:1249300. https://doi.org/10.3389/ffgc.2023.1249300

Guo Q, He Z, Wang Z (2023b) Prediction of hourly pm2. 5 and pm10 concentrations in Chongqing City in China based on artificial neural network. Aerosol Air Qual Res 23(6):220448. https://doi.org/10.4209/aaqr.220448

Guo Q, He Z, Wang Z (2023c) Predicting of daily pm2. 5 concentration employing wavelet artificial neural networks based on meteorological elements in Shanghai, China. Toxics 11(1):51. https://doi.org/10.3390/toxics11010051

Guo Q, He Z, Wang Z (2024a) Monthly climate prediction using deep convolutional neural network and long short-term memory. Sci Rep 14(1):17748. https://doi.org/10.1038/s41598-024-68906-6

Guo Q, He Z, Wang Z, Qiao S, Zhu J, Chen J (2024b) A performance comparison study on climate prediction in Weifang City using different deep learning models. Water 16(19):2870. https://doi.org/10.3390/w16192870

Guo Q, He Z, Wang Z (2024c) The characteristics of air quality changes in Hohhot city in China and their relationship with meteorological and socio-economic factors. Aerosol Air Qual Res 24(5):230274. https://doi.org/10.4209/aaqr.230274

Guo Q, He Z, Wang Z (2025) Assessing the effectiveness of long short-term memory and artificial neural network in predicting daily ozone concentrations in liaocheng city. Sci Rep 15(1):6798. https://doi.org/10.1038/s41598-025-91329-w

He Z, Guo Q (2024) Comparative analysis of multiple deep learning models for forecasting monthly ambient pm2. 5 concentrations: a case study in Dezhou City, China. Atmosphere 15(12):1432. https://doi.org/10.3390/atmos15121432

He Z, Guo Q, Wang Z, Li X (2022) Prediction of monthly pm2. 5 concentration in Liaocheng in China employing artificial neural network. Atmosphere 13(8):1221. https://doi.org/10.3390/atmos13081221

He Z, Guo Q, Wang Z, Li X (2025) A hybrid wavelet-based deep learning model for accurate prediction of daily surface pm2. 5 concentrations in Guangzhou City. Toxics 13(4):254. https://doi.org/10.3390/toxics13040254

Hu J, Chen Y, Wang W, Zhang S, Cui C, Ding W, Fang Y (2023) An optimized hybrid deep learning model for pm2. 5 and o3 concentration prediction. Air Qual Atmos Health 16(4):857–871. https://doi.org/10.1007/s11869-023-01317-0

Jairi I, Ben-Othman S, Canivet L, Zgaya-Biau H (2024) Enhancing air pollution prediction: a neural transfer learning approach across different air pollutants. Environ Technol Innov 36:103793. https://doi.org/10.1016/j.eti.2024.103793

Khan H, Tso J, Nguyen N, Kaushal N, Malhotra A, Rehman N (2024) Novel approach for predicting the air quality index of megacities through attention-enhanced deep multitask spatiotemporal learning. https://doi.org/10.48550/arXiv.2407.11283

Li X, Zhao H, Cheng J, He Q, Gao S, Mao J, Zhou C, Gong X, Rao Z (2024) Bwo-bilstm & cnn composite model for prediction of atmospheric particulate matter mass concentration. Atmos Pollut Res 15(11):102273. https://doi.org/10.1016/j.apr.2024.102273

Nguyen E, Poli M, Durrant MG, Kang B, Katrekar D, Li DB, Bartie LJ, Thomas AW, King SH, Brixi G et al (2024) Sequence modeling and design from molecular to genome scale with evo. Science 386(6723):9336. https://doi.org/10.1126/science.ado9336

Pande CB, Radhadevi L, Satyanarayana MB (2024) Evaluation of machine learning and deep learning models for daily air quality index prediction in Delhi City, India. Environ Monit Assess 196(12):1215. https://doi.org/10.1007/s10661-024-13351-1

Pranolo A, Mao Y, Wibawa AP, Utama ABP, Dwiyanto FA (2023) Optimized three deep learning models based-pso hyperparameters for Beijing pm2. 5 prediction. https://doi.org/10.48550/arXiv.2306.07296. arXiv:2306.07296

Rad AK, Nematollahi MJ, Pak A, Mahmoudi M (2025) Predictive modeling of air quality in the Tehran megacity via deep learning techniques. Sci Rep 15(1):1367. https://doi.org/10.1038/s41598-024-84550-6

Rana B, Bandyopadhyay J, Halder B (2024) Investigating the relationship between urban sprawl and urban heat island using remote sensing and machine learning approaches. Theor Appl Climatol 155(5):4161–4188. https://doi.org/10.1007/s00704-024-04874-1

Salehie O, Jamal MHB, Shahid S (2024) Characterization and prediction of pm2. 5 levels in Afghanistan using machine learning techniques. Theor Appl Climatol 155(9):9081–9097. https://doi.org/10.1007/s00704-024-05172-6

Samal KKR, Babu KS, Das SK (2022) Multi-output spatio-temporal air pollution forecasting using neural network approach. Appl Soft Comput 126:109316. https://doi.org/10.1016/j.asoc.2022.109316

Satpathy I, Nayak A, Jain V (2025) The green city: sustainable and smart urban living through artificial intelligence. In: Utilizing technology to manage territories, pp 273–304. IGI Global. https://doi.org/10.4018/979-8-3693-6854-1.ch009

Saxena V (2025) Water quality, air pollution, and climate change: investigating the environmental impacts of industrialization and urbanization. Water Air Soil Pollut 236(2):1–40. https://doi.org/10.1007/s11270-024-07702-4

Shao Z, Wang F, Xu Y, Wei W, Yu C, Zhang Z, Yao D, Sun T, Jin G, Cao X et al (2024) Exploring progress in multivariate time series forecasting: comprehensive benchmarking and heterogeneity analysis. IEEE Trans Knowl Data Eng. https://doi.org/10.1109/TKDE.2024.3484454

Singh S, Suthar G (2025) Machine learning and deep learning approaches for pm2. 5 prediction: a study on urban air quality in Jaipur, India. Earth Sci Inf 18(1):97. https://doi.org/10.1007/s12145-024-01648-1

Srivastava H, Kumar Das S (2023) Air pollution prediction system using xrsth-lstm algorithm. Environ Sci Pollut Res 30(60):125313–125327. https://doi.org/10.1007/s11356-023-28393-0

Sun Q, Zhu Y, Chen X, Xu A, Peng X (2021) A hybrid deep learning model with multi-source data for pm 2.5 concentration forecast. Air Qual Atmos Health 14:503–513. https://doi.org/10.1007/s11869-020-00954-z

Suthar G, Singh S (2025) Application of machine learning models for pm2. 5 prediction in Bengaluru using precursor air pollutants and meteorological data. Theor Appl Climatol 156(3):1–21. https://doi.org/10.1007/s00704-025-05413-2

Tao Q, Liu F, Li Y, Sidorov D (2019) Air pollution forecasting using a deep learning model based on 1d convnets and bidirectional gru. IEEE Access 7:76690–76698. https://doi.org/10.1109/ACCESS.2019.2921578

Ulpiani G, Ranzi G, Santamouris M (2021) Local synergies and antagonisms between meteorological factors and air pollution: a 15-year comprehensive study in the Sydney region. Sci Total Environ 788:147783. https://doi.org/10.1016/j.scitotenv.2021.147783

Vo MT, Vo A, Bui H, Le T (2023) A hybrid deep learning approach for pm2. 5 concentration prediction in smart environmental monitoring. Intell Autom Soft Comput 36(3). https://doi.org/10.32604/iasc.2023.034636

Xu J, Su Z, Liu C, Nie Y, Cui L (2025) Climate change, air pollution and chronic respiratory diseases: understanding risk factors and the need for adaptive strategies. Environ Health Prev Med 30:7–7. https://doi.org/10.1007/s40726-025-00347-9

Yang G, Lee H, Lee G (2020) A hybrid deep learning model to forecast particulate matter concentration levels in Seoul, South Korea. Atmosphere 11(4):348. https://doi.org/10.3390/atmos11040348

Zaini N, Ean LW, Ahmed AN, Malek MA (2022) A systematic literature review of deep learning neural network for time series air quality forecasting. Environ Sci Pollut Res 1–33. https://doi.org/10.1007/s11356-021-17442-1

Zhang B, Rong Y, Yong R, Qin D, Li M, Zou G, Pan J (2022) Deep learning for air pollutant concentration prediction: a review. Atmos Environ 290:119347. https://doi.org/10.1016/j.atmosenv.2022.119347

Zhang X, Xi Z, Yang M, Zhang X, Wu R, Li S, Pan L, Fang Y, Lv P, Ma Y et al (2025) Short-term effects of combined environmental factors on respiratory disease mortality in Gingdao city: a time-series investigation. PloS One 20(1):0318250. https://doi.org/10.1371/journal.pone.0318250

Zong Z, Guan Y (2024) Ai-driven intelligent data analytics and predictive analysis in industry 4.0: transforming knowledge, innovation, and efficiency. J Knowl Econ 1–40. https://doi.org/10.1007/s13132-024-02001-z