Environmental Health Engineering and Management Journal. 2025;12:1370
http://ehemj.com
doi 10.34172/EHEM.1370
Environmental Health Engineering and Management Journal
Original Article
Open Access Publish Free

Air pollution forecasting using advanced machine learning techniques and ensemble stacking in Delhi

T. Sreenivasulu, G. Mokesh Rayalu*

School of Advanced Sciences, Vellore Institute of Technology, Vellore, India

Abstract

Background: Air pollution, primarily due to air particulates like PM10 and PM2.5, causes several respiratory health problems. Accurate forecasting of particulate matter concentration is crucial for managing complex and nonlinear data, allowing timely interventions for early warning systems and air pollution control. The study aimed to develop reliable machine learning models for forecasting PM2.5 and PM10 concentrations, providing actionable insights for air quality management and public health.

Methods: Despite challenges with unusual patterns and abrupt changes, the models achieved high accuracy, with R^2 values exceeding 0.96 and low RMSE values. MLP outperformed the RF and XGB models for both PM2.5 and PM10 predictions. MLP-based stacking models further enhanced prediction accuracy, achieving the lowest RMSE and highest R^2 values. For PM10, the weighted average approach provided better performance, striking an optimal balance between the different models' contributions.

Results: Despite unusual patterns and rapid jerks, our models had the highest R^2 (>0.96) and lowest RMSE values. MLP outperformed the RF and XGB models for both pollutants. It improves the PM2.5 concentration predictions of stacking models, notably those using MLP as the meta-learner. MLP-based stacking yielded the lowest error values. The weighted average strategy improved the PM10 performance more than the stacking models and provided a better balance between the model contributions.

Conclusion: Ensemble models achieved enhanced predictive accuracy by emphasizing the importance of selecting machine learning models and stacking methods based on air contaminants and data features, which is a crucial aspect of air quality management and public health.

Keywords: Machine learning, Random forest, Air pollution, Neural networks, India

Citation: Sreenivasulu T, Mokesh Rayalu G. Air pollution forecasting using advanced machine learning techniques and ensemble stacking in Delhi. Environmental Health Engineering and Management Journal. 2025;12:1370. doi: 10.34172/EHEM.1370.

Article History:
Received: 13 June 2024
Revised: 26 October 2024
Accepted: 16 October 2025
ePublished: 22 October 2025

***Correspondence to:**
G. Mokesh Rayalu,
Email: mokesh.g@vit.ac.in

Introduction

Air pollution is a global health issue, causing major health problems such as asthma, heart attacks, lung cancer, acute bronchitis, and respiratory and heart-related difficulties. Primary causes include carbon monoxide (CO), nitrogen dioxide (NO_2), sulfur dioxide (SO_2), ozone (O_3), and particulate matter (PM2.5 and PM10). These tiny particles, 30 to 35 times smaller than human hairs, can cause breathing and cardiac difficulties, leading to death in critical situations (1). The World Health Organization (WHO) reports that tiny particles in contaminated air cause around seven million deaths annually (2). Atmospheric particulate matter comprises a varied collection of particles derived from chemical and biological sources, which may profoundly affect the health of people and animals (3). Particulate matter is emitted from various sources, including vehicle exhaust fumes, power plants, fossil fuel combustion, agricultural pesticides, natural dust, and industrial facilities (4). Rapid urbanization and industries are causing an increase in air pollution, mainly in metropolitan cities (5). India, for example, is the third most polluted country with an annual average of 54.4 μg/m^3, more than 10 times the WHO PM2.5 yearly standards. According to the Lancet report, roughly 1.6 million people in India died in 2019 because of air pollution from both home and ambient sources (6). Delhi is the second-most polluted city in India with an annual PM2.5 mean of 54 μg/m^3 (7). Accurate prediction of particulate matter, including PM2.5 and PM10 levels, is essential for public health strategy and pollution management initiatives. This study aimed to predict daily PM2.5 and PM10 concentrations in Delhi using advanced models of machine learning and ensemble approaches to enhance prediction accuracy.

The literature review below includes a variety of approaches for forecasting various air pollutant concentrations. Some methodologies incorporated explanatory variables, whereas others used univariate models.

Traditional time-series models, such as autoregressive integrated moving average (ARIMA), seasonal autoregressive integrated moving average (SARIMA), exponential smoothing, and Holt-Winters models, produce inaccurate findings for environmental variables with nonlinearity, variability, and complicated seasonality. Liu and You (8) implemented ARIMA and three-layer neural network models to forecast Beijing's air quality index over long and short terms. However, the ARIMA model could not capture nonlinear relationships in the dataset. Bhatti et al. (9) utilized the SARIMA model to predict the monthly mean PM2.5 and PM10 levels in Lahore; however, the model failed to yield satisfying results due to the enormous amount of data with significant seasonality and non-stationarity.

Mani et al. (10) proposed that multilinear regression and ARIMA models worked well in forecasting the daily air quality index (AQI) because the variables' relationships are linear. Goyal et al. (11) used a multiple linear regression model to forecast PM2.5 and NOx levels in three separate Jaipur sites; nonetheless, the model produced adequate results after eliminating outliers from the data. Ventura et al. (12) employed the ANN and Holt-Winter models to forecast PM2.5 in three distinct locations in Brazil: rural, urban, and industrial. The ANN model produced consistent output in three regions and even enhanced performance after adding meteorological factors. Hasnain et al. (13) examined three distinct time series models, ARIMA, RF, and Prophet, for forecasting the daily mean of PM2.5 levels in multiple Chinese cities using five years of daily data. The RF model outperformed the other two models, whereas the ARIMA model failed to perform better due to the high unpredictability and complexity of the environmental variables.

Kumar et al. (14) proposed ET (extra trees)-AdaBoost, an ensemble technique for predicting PM2.5 levels based on two years of daily data from Delhi. The proposed model outperformed existing ML models because it combined the benefits of extra trees and AdaBoost to provide accurate and resilient results. Yin et al. (15) combined four different models, cluster linear regression, long-term MLP, Fourier series descriptor, and short-term MLP, to combine long-term and short-term capabilities for forecasting PM2.5 using 12 years of massive data from Puli township in Taiwan, which contains meteorological, periodic, and other regressive variables related to PM2.5. The combination of these models presents certain challenges, but it produces better results than the individual approaches.

Danesh Yazdi et al. (16) suggested an ensemble model for forecasting PM2.5 levels based on nine years of daily data from Greater London that included three machine learning models (ML): RF, gradient boosting machine, and k-nearest neighbors. The ensemble technique is limited when handling high-variability data, but it performs better than the individual models. Satish et al. (17) used an ANN-based ensemble model with machine learning models to estimate stream water quality in India's Godavari River basin. The stacking ANN model outperformed the individual ML models XGB, RF, and ET in terms of water quality parameter forecasting accuracy, reducing overfitting, and increasing resilience.

The ML and DL models have improved significantly in predicting air pollution, but for locations such as Delhi, which have a high variability, more robust models are required for long-term air quality forecasts. Traditional statistical models fail to capture nonlinear patterns, whereas some of the ML models fail to generalize well over long-period data. Existing research has not sufficiently explored hybrid stacking models, which combine multiple ML and DL techniques to solve the problem. This study aimed to bridge this gap by implementing innovative air quality forecasts using machine learning models like RF and XGB, as well as deep learning models like ANN. Moreover, this research introduces novel hybridization techniques to improve the long-term predictions of daily PM2.5 and PM10 concentrations. These techniques include stacking with linear regression, ANN, and the weighted average approach. We used advanced hybrid models in this study to push the limits of accuracy in forecasting such complex environmental data, which has significant implications for public health and policymaking in highly polluted cities like Delhi.

Materials and Methods

Data Description

This study uses air pollution measurements from several monitoring stations in Delhi, India. Figure 1 displays the study area map. In this study, we converted the hourly air pollution data variables to daily data by computing the daily average for each pollutant. This process involves adding the hourly values for each pollutant for each day and then dividing the total by the number of hourly measurements for that day. This process was performed for each pollutant variable in our dataset. The hourly air pollution data was gathered from Kaggle between November 25, 2020, and January 24, 2023. This dataset was freely obtained from the following website: https://www.kaggle.com/datasets/deepaksirohiwal/delhi-air-quality.

The conversion of hourly data to daily data is critical for long-term analysis and visualization. As a result, this change is critical for delivering reliable, precise, and actionable information about air pollution and its impact on human health and the environment. The revised data were divided into two sections: training and testing data. The training data are used to train the models, while the testing data are utilized to evaluate them. The training data were acquired from November 25, 2020, to November 5, 2022. On the other hand, the testing data ranges from November 6, 2022, to January 24, 2023.

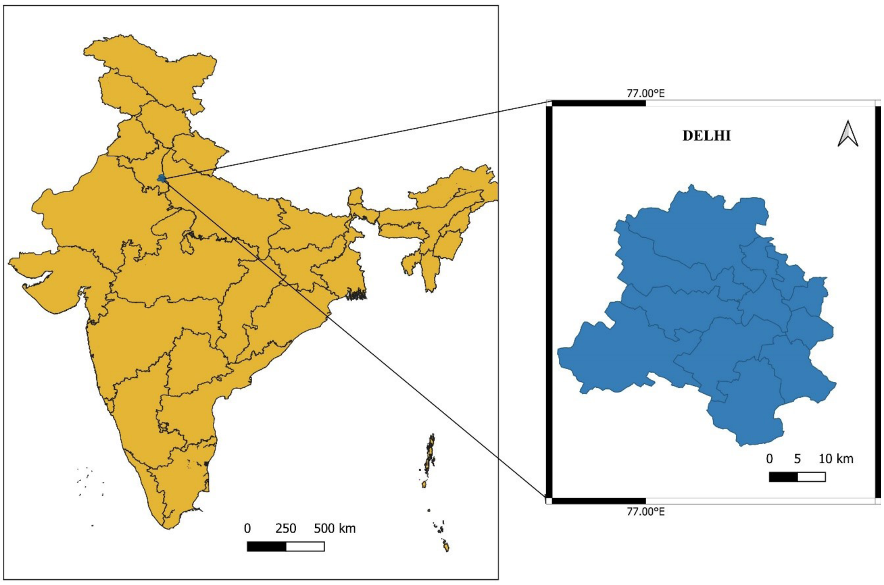

Figure 1. Study area map of Delhi, located in India

Methodology

This section presents a rigorous prediction methodology for air-quality evaluation, with an emphasis on specific matter concentrations, particularly PM2.5 and PM10. The methodology includes a multi-stage modeling strategy that uses various machine learning approaches to improve forecast accuracy and dependability.

Initially, we suggested three machine learning prediction models for PM2.5 and PM10 variables: Multilayer perceptron neural network (MLP), random forest (RF), and extreme gradient boosting (XGB). In the second step, a successful method known as stacking is used for the following initial predictions: This includes developing the hybrid model with different strategies that combine the strengths of the three base models. The proposed stacking model is MLP-RF-XGB. In the last stage, we extracted meta-features from both base models used in the hybrid model. In this case, the meta-model is linear regression and MLP, which successfully integrates predictions from the basic models. Finally, we used the ensemble model weighted average to forecast PM2.5 and PM10 concentrations. The flow of the work is shown in Figure 2.

Multilayer Perceptron Neural Networks (MLPS)

Artificial neural networks (ANNs) are sophisticated forms of artificial intelligence (AI). These ANNs are critical in time series forecasting because they produce reliable results. The ANN model is a popular intelligence model used in several domains, including time series modeling and forecasting. ANN models are popular because of their ability to extrapolate data without assuming a certain model structure (18). When standard linear models struggle to grasp the intricacies and nonlinear relationships in the data at hand, ANN models are employed to overcome the problems and produce more accurate findings than traditional time series models. ANN contains several models, but the single-hidden-layer multilayer perceptron is most commonly employed in time series modeling. A layer is created by joining two or more neurons, and a network can be made up of one or more layers of neurons linked together by a connection strength known as weight (19).

In general, MLPS has three layers: input, output, and hidden. In the input layer, the number of nodes represents the number of features or lagged observations used for prediction. The input layer sends data to the next layer without completing any actions. The hidden layer is located between the input and output layers. The proposed method collects input data from all nodes in the input layer and calculates the weighted total of these inputs. The obtained total is then processed through the activation function, introducing nonlinearity into the model. MLPS has one or more hidden layers that convert the input data into a more acceptable format for prediction by the output layer. Finally, the output layer has a single neuron that produces the regression results. In the proposed model, the hidden layer employs the rectified linear unit (ReLU) activation function, and regularization techniques, such as L2 regularization, are utilized to prevent overfitting. The model is trained using backpropagation with an optimizer such as Adam, and its performance is assessed using regression-appropriate metrics. Figure 3 shows the architecture of an MLPS.

Random Forest (RF)

RF is a type of ML model. It is an embedded learning-based prediction approach that incorporates several decision trees. It can be used for data mining in machine learning frameworks. RFs are typically employed for classification,

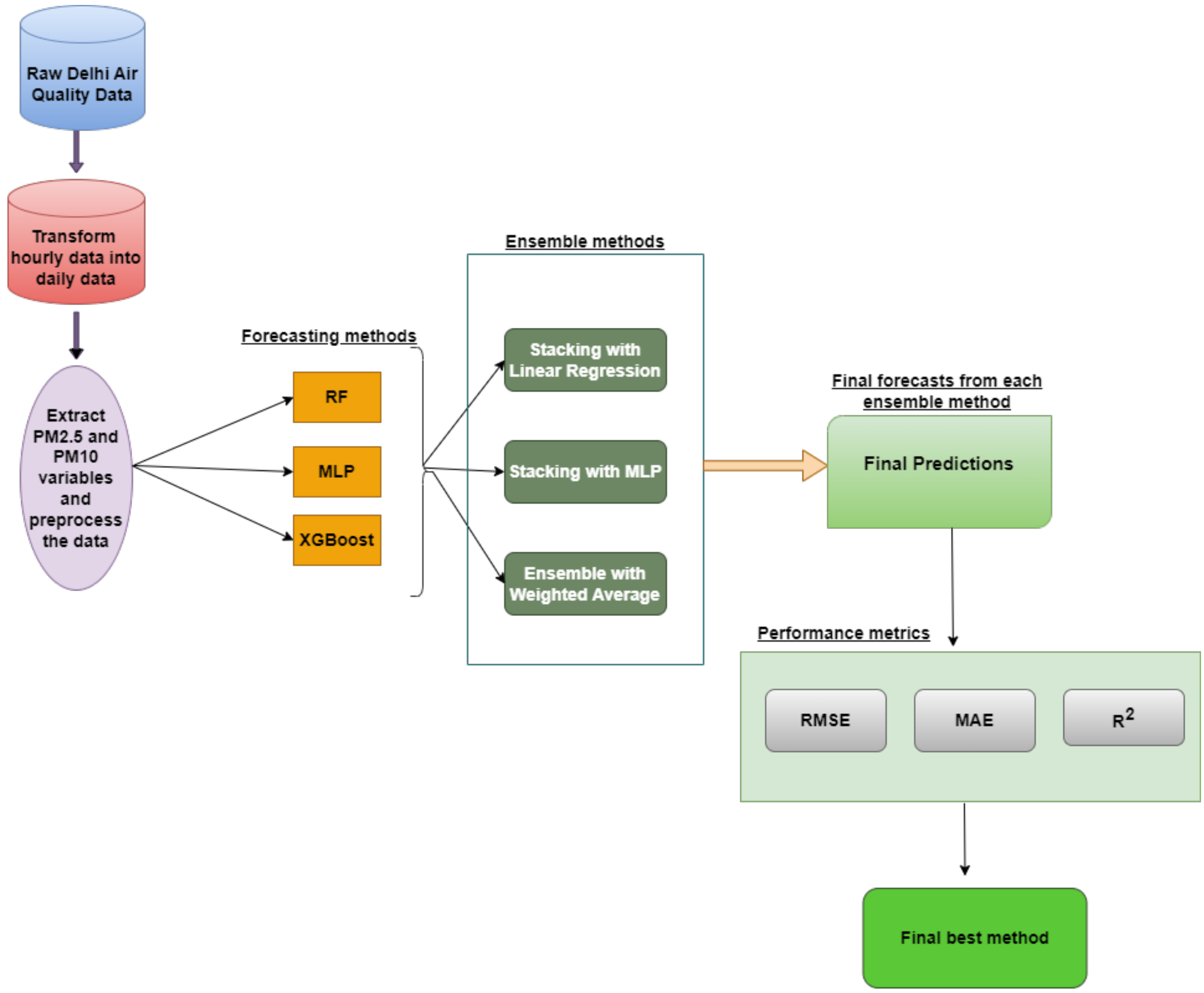

Figure 2. Comprehensive workflow for PM2.5 and PM10 forecasting

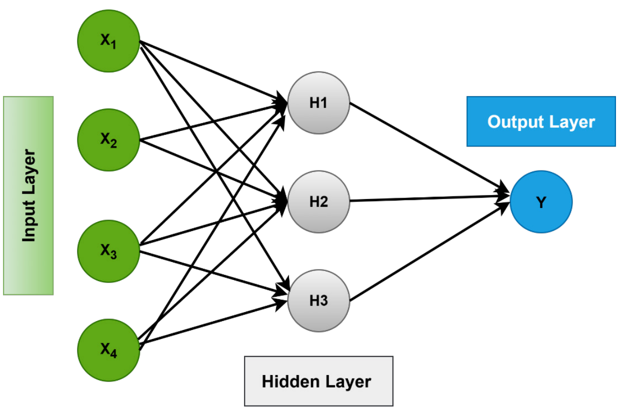

Figure 3. Design of a multi-layer perceptron model

regression, and forecasting. RFs perform well for time series forecasting, even when the data contains non-linear connections (20). For time series forecasting, the RF first generates a lag variable, which represents prior time steps, utilizes it as input, and then learns the time series pattern from it. To improve the model's performance, feedforward or other standard approaches, such as k-fold cross-validation, will be applied. After training the model, the next step is to make predictions based on the average forecasts of the various trees in the ensemble.

Extreme Gradient Boosting (XGB)

XGB is one of the most powerful gradient-boosting machine learning techniques, noted for its efficiency and high performance when applied to structured data, like time series data. It can be used for both classification and regression (21). XGB works well with temporal data, and when we use this technique with time-series data, we must first convert it into a format suited for supervised learning. This involves analyzing a dataset in which prior time steps are predictors of future values. The primary purpose of XGB is to assess prediction accuracy by building on the information gained from prior weak learners and adding new weak learners that are specifically designed to address and correct residual errors (22). In addition, XGB employs both regularizations, such as L1 and L2, to avoid overfitting and allows parallel processing to considerably accelerate estimations. The proposed method provides several hyperparameters that may be fine-tuned to improve performance, including the number of trees, tree depth, and learning rate.

Proposed Stacking Models

The ensemble technique is an ML strategy that combines many model predictions to provide superior outcomes when compared to individual strategies (18). Ensemble approaches include stacking. This is achieved by combining the many model predictions from the same dataset using a meta-model that incorporates meta-features from base models. In this study, three distinct metamodels, such as linear regression, MLP stacking, and weighted average, were utilized to aggregate predictions from base models, such as MLP, RF, and XGB.

Stacking with linear regression: Stacking using linear regression is an effective strategy that integrates meta-features from base models, such as MLP, RF, and XGB models, and learns to weight these features to minimize total prediction error on training data. It is posited that the best combination of base model predictions may be conveyed via a linear relationship.

Stacking with MLP: Stacking using MLP is a strong approach that can produce more robust predictions because it captures non-linear interactions and allows for hyperparameter adjustment. MLP consists of an input layer that sends nodes or previously lagged values to the next layer, as well as one or more hidden layers that alter the input data. The hidden layer calculates the weighted total of these inputs and runs it through the activation function, introducing nonlinearity into the model. The final prediction results are provided by the output layer, which has a single neuron for the regression work. The MLP can grasp complicated patterns in meta-features, thereby minimizing the total prediction error.

Weighted average ensemble: To improve the performance of a weighted average ensemble, compute a weighted average of all base model predictions and score them according to their performance, such as RMSE on testing data. The weights in the final ensemble are inversely proportional to the RMSE of each model. Models with lower RMSE values provide more weight, whereas models with larger RMSE values give less weight. By employing this stacking method, we aim to capitalize on the benefits of several techniques, such as ANN, RF, and XGB, to increase the resilience and accuracy of final predictions.

Evaluation Metrics

In this study, we employed three metrics to evaluate and compare the performance of various models. Root mean square error (RMSE), mean absolute error (MAE) (23), and R^2 coefficient of determination. The formulae for these measures are defined as follows:

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(\hat{y}_i - y_i\right)^2}$$

$$MAE = \frac{}{n}\sum_{i=}^{n}\left|\hat{y}_i - y_i\right|$$

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{n}\left(y_i - \overline{y}_i\right)^2}$$

Results

Summary statistics for PM2.5 and PM10 data reveal a concerning trend in air quality, with a daily mean of 237.72 µg/m^3 for PM2.5 and 299.39 µg/m^3 for PM10 levels significantly exceeding the WHO guidelines standard. We found that PM2.5 concentrations were 15 times higher, and PM10 levels exceeded the standard by 6.5 times the WHO daily guidelines limits. These elevated pollution levels pose serious health risks to the population in the city. Furthermore, the notable standard deviations of 180.30 µg/m^3 for PM2.5 and 206.69 µg/m^3 for PM10 indicate considerable variability and potential seasonal fluctuations in air quality, emphasizing the need for targeted interventions. Continuous monitoring and research will be essential to track progress and adapt advanced strategies to improve air quality effectively.

This section highlights the results of predicted values for PM2.5 and PM10 from individual and ensemble models using MLP, RF, and XGB.

Results of Individual Models

In this study, 90% of the data is used for training, while the remaining 10% (80 days) is used for testing to anticipate PM2.5 and PM10 air particle concentrations. First, the three individual models, such as MLP, RF, and XGB, were trained. To evaluate the performance of these three individual ML approaches, performance metrics such as RMSE, MAE, and R^2 were utilized, as shown in Table 1.

The ANN model was initially trained using the MLPRegressor from the Scikit-Learn Python module. GridSearchCV was used to determine the best parameters and models based on cross-validation performance. After tuning the hyperparameters, a higher R^2 value was obtained for both the PM2.5 and PM10 parameters, with an alpha of 0.001, two hidden layers of sizes 100 and 50, and a maximum of 1000 iterations. The MLP model fared better on PM10 than on PM2.5. The RMSE for PM10 was 0.9906, whereas that for PM2.5 was 0.9886.

After training the MLP model, the RF model, which is an ensemble approach that generates many decision

Table 1. Comparison of the prediction performances of three individual models for both Particulate Matter with a diameter of 2.5 microns or less (PM2.5) and Particulate Matter with a diameter of 10 microns or less (PM10)

Variable	Model	RMSE	MAE	R-square
PM2.5 (µg/m^3)	MLP	17.0643	13.1255	0.9886
	RF	26.4041	18.9136	0.9728
	XGB	26.6033	16.3819	0.9724
PM10 (µg/m^3)	MLP	18.0392	13.1632	0.9906
	RF	17.4015	13.7427	0.9912
	XGB	36.2245	17.4294	0.9621

trees during training and produces mean predictions from individual trees, was trained. The GridSearchCV technique was used to determine ideal parameters and models. Each combination was evaluated using 5-fold cross-validation. After tuning the hyperparameters, the RF model functioned admirably in terms of parameters, with the maximum depth of each tree in the forest set to 7, the minimum number of samples necessary to divide an internal node set to 2, and the number of trees in the forest set to 200. These tuning parameters produced low RMSE and high R^2 values for PM10 and PM2.5. Compared with the model's performance for PM2.5, the RF model fitted well for PM10, with R^2 values of 0.9912 and 0.9728, respectively.

To anticipate the daily average PM2.5 and PM10 levels, we used the XGB model, an accurate gradient approach known for its versatility and excellent performance with time-series data. GridSearchCV was utilized to tune the XGB model's hyperparameters using 5-fold cross-validation. The ideal parameters for PM2.5 are as follows: a 0.1 learning rate, 200 trees in the forest, and a maximum depth of 5 for each tree. The optimal PM10 parameters were a learning rate of 0.05, 200 trees in the forest, and a maximum depth of 5. The XGB model was trained on training data using optimal parameters and then tested on test data. The XGB model correctly predicted the daily average PM2.5 and PM10 levels with R^2 values of 0.9728 and 0.9621, respectively. A comparison of the actual and predicted values for both pollutants is shown in Figures 4 and 5.

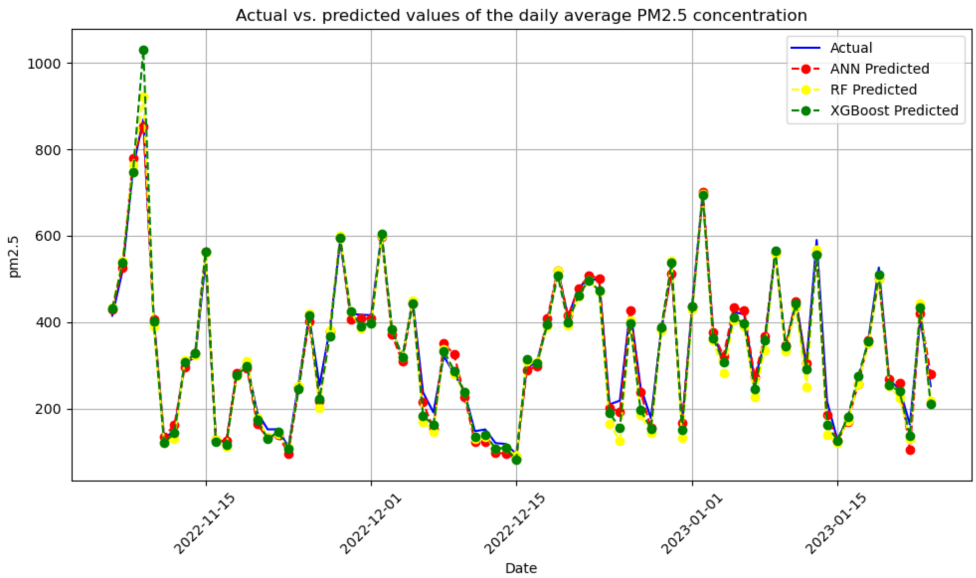

Figure 4. Comparison of PM2.5 Forecasting Results: ANN(MLP), RF, and XGB vs. Actual

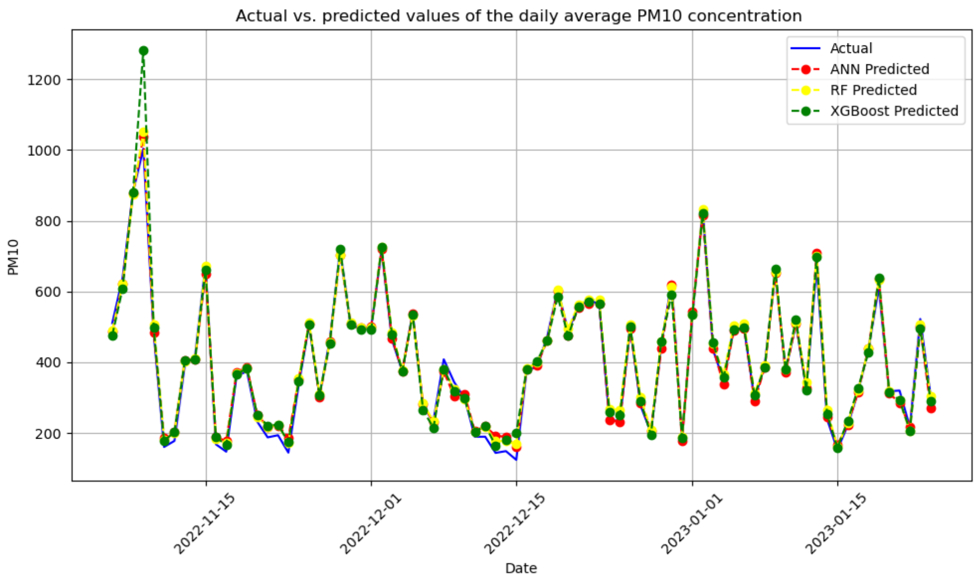

Figure 5. Comparison of PM10 Forecasting Results: ANN(MLP), RF, and XGB vs. Actual

Results of Stacking Techniques

Three stacking approaches were used to accurately forecast daily average PM2.5 and PM10 levels and enhance model performance. These methods include stacking with linear regression, the MLP regressor, and the weighted average method. To assess the performance of these three approaches, performance metrics such as RMSE, MAE, and R^2 were utilized, as shown in Table 2. The PM2.5 findings showed that all three stacking approaches worked remarkably well, with R^2 values near 0.99, indicating a high level of accuracy when projecting daily average PM2.5 values. Among the three approaches for PM2.5, stacking with an MLP regressor demonstrated the lowest RMSE (16.6885) and MAE (12.975). Figure 6 compares actual and predicted PM2.5 levels for three stacking approaches. The anticipated values are fairly similar to the real values.

Similarly, the same stacking methods were employed to forecast daily average PM10 levels. The evaluated metrics are summarized in Table 2. The results revealed that all three techniques were effective in predicting PM10 levels, with R^2 values greater than 0.985. The weighted average strategy produced the best results for PM10 prediction, with the lowest RMSE (18.9096) and greatest R^2 (0.9896). Figure 7 presents the actual and expected outcomes for the three stacking approaches, indicating that the predicted values are similar to the real values. When comparing the three stacking techniques used in this study for PM2.5 and PM10 pollutant concentrations, it is evident that stacking with MLP and the weighted average approach performed marginally better than stacking using regression. The least error values across the three models demonstrate that the stacked models successfully capture data variability.

Table 2. Comparison of the prediction performances of three ensemble models for both Particulate Matter with a diameter of 2.5 microns or less (PM2.5) and Particulate Matter with a diameter of 10 microns or less (PM10)

Variable	Stacking Model	RMSE	MAE	R-square
PM2.5 (μg/m^3)	Stacking with linear regression	16.7634	13.088	0.9890
	Stacking with MLP	16.6885	12.975	0.9891
	Weighted average	18.4221	13.2298	0.9867
PM10 (μg/m^3)	Stacking with linear regression	22.2821	13.9379	0.9856
	Stacking with MLP	21.1890	13.9118	0.9870
	Weighted average	18.9096	13.3324	0.9896

Discussion

We compared the accuracy of the three machine learning models, MLP, RF, and XGB, in forecasting daily average PM2.5 and PM10 pollutant concentrations. Each model was finetuned using optimal hyperparameters to enhance performance and evaluated using metrics such as RMSE, MAE, and R^2 values. Table 1 provides the performance metrics for the three models across both pollutants.

For PM2.5, the MLP model outperformed the RF and XGB models by achieving the lowest RMSE (17.0643) and MAE (13.1255), alongside the highest R^2 (0.9886). The superior performance of the MLP model shows that it effectively captures nonlinear patterns in air pollution data, as shown in Figure 4, which compares actual and predicted values for test data for three different models. For PM10, the RF model performed marginally better with the lowest RMSE (17.4015) and high R^2 (0.9912), since the RF model effectively captured important variance, although the MLP model showed similar strong results. Conversely, the XGB model displayed the highest error values for both pollutants, indicating its lower effectiveness compared to the other two models. Figure 5 shows a comparison of the three models for PM10 validation data.

To improve performance, we employ the stacking models, which combine the predictions of MLP, RF, and XGB using linear regression, MLP, and a weighted

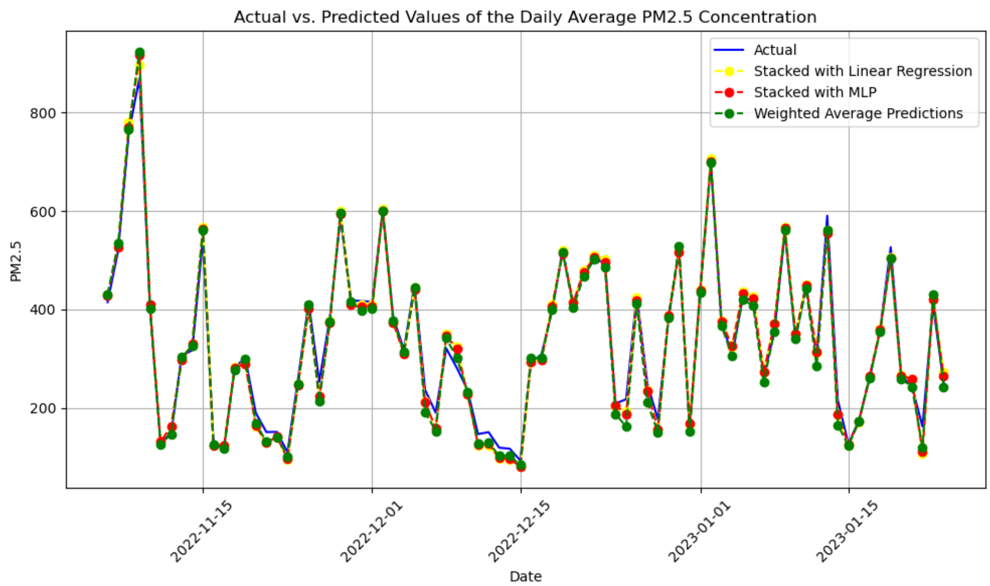

Figure 6. Comparison of ensemble approaches predicted vs actual values for PM2.5

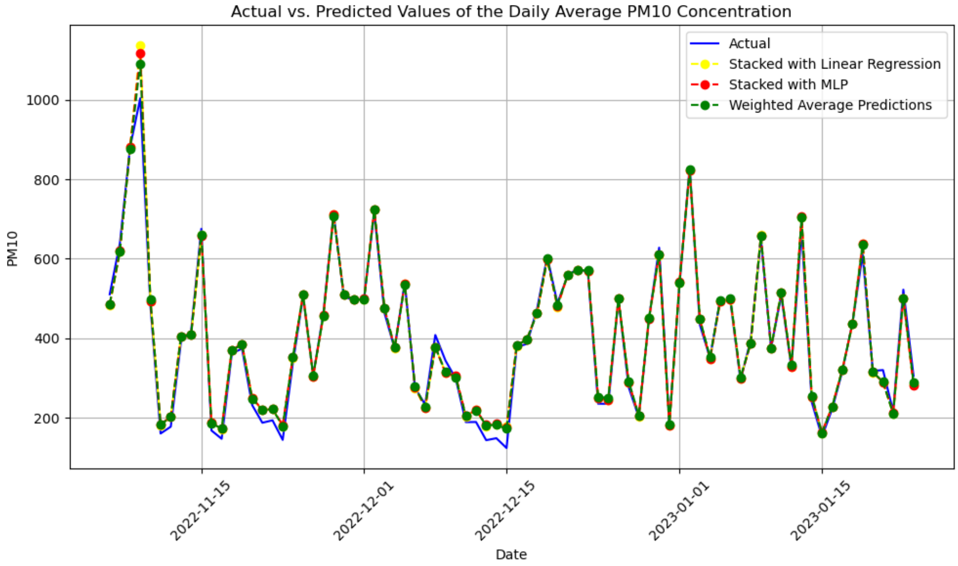

Figure 7. Comparison of ensemble approaches predicted vs actual values for PM10

average approach. Figures 6-7 illustrates the variation between actual and expected values for validation data for two pollutants. The stacking methods worked well to combine the best parts of the base models for PM2.5. This led to better performance accuracy, especially when MLP was used as a meta-learner, and an RMSE of 16.6885 and an R^2 of 0.9891 were reached. This demonstrates that leveraging MLP in a stacking framework can enhance the model's ability to predict complex patterns in pollutant concentrations. For PM10, stacking using linear regression and MLP resulted in marginally better results. Interestingly, the weighted average method outperformed both stacking methods with an RMSE of 18.9096 and an R^2 of 0.9896. This indicates that effective balancing of contributions from each individual model, rather than relying heavily on a single model, can yield better predictions in certain scenarios, especially when predicting PM10 concentrations.

The authors retrieve the latest literature on PM2.5 and PM10 forecasting, selecting the first seven articles based on similar research topics, geographical contexts, and methodological methods. The standard statistical measure RMSE is used to evaluate the overall performance of the developed models and comparative results are shown in Table 3. The average RMSE values are 52.19 μg/m^3 according to (24), 21.51 μg/m^3 according to (25), 30.32 μg/m^3 according to (26), 0.06 μg/m^3 according to (27), 47.36 μg/m^3 according to (28), 25.75 μg/m^3 according to (29), and 36.58 μg/m^3 according to (30) with respect to PM2.5. Upon examination, the suggested model attains an average RMSE of 16.68 μg/m^3. The proposed model ranks second among eight studies in terms of predictive accuracy. The suggested model demonstrates specific improvements over previous approaches.

The proposed models had the highest R^2 (>0.96) values and the lowest RMSE and MAE values compared to earlier studies. This was true even though the data showed strange patterns and sudden jumps. The results highlight that the MLP model is highly effective for predicting PM2.5, while the RF model excels for PM10, and stacking approaches, especially those using MLP, can significantly improve overall forecast accuracy. Overall, this study advances the performance of ensemble models for predicting air pollutant concentrations, highlighting the significance of selecting ML models and an appropriate stacking approach based on air pollutants and data behavior. The enhanced performance of stacking approaches implies that they can make more reliable predictions, which is critical for air quality management and public health measures. Accurate air pollution forecasts are essential for effective air quality management, enabling timely interventions and minimizing public health risks.

Table 3. Comparative exhibitions of daily average RMSE (μg/m^3) using the proposed model with recent models applied to PM2.5 data in Delhi

Year	Location	Target variables	Model	RMSE (μg/m^3)
2024	Delhi	PM2.5	Linear regression	52.19
2022	Delhi	PM2.5	LSTM	21.51
2023	Delhi	PM2.5	ELM-SO hybrid model	30.32
2024	Delhi	PM2.5 and AQI	Bi-LSTM	**0.06 and 0.10**
2024	Delhi	PM2.5	1D-CNN	47.36
2024	Delhi	PM2.5	Bi-RNN	25.75
2024	Delhi	PM2.5 and PM10	Five-layered ANN	36.58 and 56.4
2024	Delhi	PM2.5 and PM10	Proposed model	**16.68 and 18.9**

Conclusion

This study describes three separate machine learning algorithms, MLP, RF, and XGB, as well as ensemble models that use stacking techniques to forecast daily average PM2.5 and PM10 pollution levels without requiring exogenous factors. Despite the data's unpredictable oscillations, no obvious trend, and unexpected peaks, all of the models fitted well and produced the greatest long-term predictions with the highest R^2 values. Individually, the MLP model outperformed the other two models for both PM2.5, but RF outperformed both MLP and XGB for PM10. Stacking with MLP and a weighted average method outscored stacking with linear regression by a small margin. The ensemble models improved performance and provided more accurate results compared to individual models. Ultimately, the neural networks achieved good results in finding the underlying patterns in the data, and as a metamodel, they captured nonlinear interactions among the base models. In future studies, incorporating additional factors like meteorological and anthropogenic variables may further enhance the predictive power of these models, providing even greater value for policymakers and public health authorities.

Acknowledgments
The authors express their heartfelt gratitude to the Vellore Institute of Technology for providing the necessary facilities and assistance to conduct and complete this work.

Authors' contributions
Conceptualization: Mokesh Rayalu G.
Data curation: Sreenivasulu T.
Formal analysis: Sreenivasulu T.
Funding acquisition: Mokesh Rayalu G.
Investigation: Mokesh Rayalu G.
Methodology: Sreenivasulu T.
Project administration: Mokesh Rayalu G.
Resources: Mokesh Rayalu G.
Software: Sreenivasulu T.
Supervision: Mokesh Rayalu G.
Validation: Mokesh Rayalu G.
Visualization: Sreenivasulu T.
Writing–original draft: Sreenivasulu T.
Writing–review & editing: Mokesh Rayalu G.

Competing interests
The authors confirm that this paper is their work and have no conflicts of interest to disclose.

Ethical issues
There were no ethical issues in the writing of this article.

Funding
This research received no external funding.

References
1. Polezer G, Tadano YS, Siqueira HV, Godoi AFL, Yamamoto CI, de André PA, et al. Assessing the impact of PM2.5 on respiratory disease using artificial neural networks. Environ Pollut. 2018;235:394-403. doi: 10.1016/j.envpol.2017.12.111.
2. World Health Organization (WHO). Ambient (Outdoor) Air Pollution [Internet]. WHO; 2024. Available from: https://www.who.int/news-room/fact-sheets/detail/ambient-(outdoor)-air-quality-and-health. Accessed May 30, 2024.
3. Aboosaedi Z, Naddafi K, Nabizadeh Nodehi R, Hassanvand MS, Faridi S, Aliannejad R. Investigating the performance of urban air quality monitoring station in measuring PM2.5 and PM10: a case study in Tehran, Iran. Environ Health Eng Manag. 2023;10(4):451-8. doi: 10.34172/ehem.2023.48.
4. Dyah Prasetyawati N, Tjiptowibisono S, Pranoto P, Sunarto S. Swot analysis of factors causing air pollution and recommended control efforts in the city of Yogyakarta, Indonesia. Environ Health Eng Manag. 2024;11(1):1-7. doi: 10.34172/ehem.2024.01.
5. Fouladvand S, Nozari M, Nadafi K, Alimohammadi M, Khalui M, Hassanvand MS, et al. Assessment of bioaerosols, PM2.5, and PM10 in liver transplantation operating rooms in Tehran, Iran: implications for air quality. Environ Health Eng Manag. 2024;11(3):301-13. doi: 10.34172/ehem.2024.30.
6. Fuller R, Landrigan PJ, Balakrishnan K, Bathan G, Bose-O'Reilly S, Brauer M, et al. Pollution and health: a progress update. Lancet Planet Health. 2022;6(6):e535-47. doi: 10.1016/s2542-5196(22)00090-0.
7. IQAir. World Live Air Quality Map. Available from: https://www.iqair.com/world-air-quality. Accessed May 30, 2024.
8. Liu T, You S. Analysis and forecast of Beijing's air quality index based on ARIMA model and neural network model. Atmosphere. 2022;13(4):512. doi: 10.3390/atmos13040512.
9. Bhatti UA, Yan Y, Zhou M, Ali S, Hussain A, Qingsong H, et al. Time series analysis and forecasting of air pollution particulate matter (PM2.5): an SARIMA and factor analysis approach. IEEE Access. 2021;9:41019-31. doi: 10.1109/access.2021.3060744.
10. Mani G, Viswanadhapalli JK, Stonier AA. Prediction and forecasting of air quality index in Chennai using regression and ARIMA time series models. J Eng Res. 2022;10(2 Pt A):179-94. doi: 10.36909/jer.10253.
11. Goyal S, Sharma R. Prediction of the concentrations of PM2.5 and NOx using machine learning-based models. Mater Today Proc. 2023. doi: 10.1016/j.matpr.2023.07.121.
12. Ventura LM, de Oliveira Pinto F, Soares LM, Luna AS, Gioda A. Forecast of daily PM2.5 concentrations applying artificial neural networks and Holt-Winters models. Air Qual Atmos Health. 2019;12(3):317-25. doi: 10.1007/s11869-018-00660-x.
13. Hasnain A, Hashmi MZ, Khan S, Bhatti UA, Min X, Yue Y, et al. Predicting ambient PM2.5 concentrations via time series models in Anhui province, China. Environ Monit Assess. 2024;196(5):487. doi: 10.1007/s10661-024-12644-9.
14. Kumar S, Mishra S, Singh SK. A machine learning-based model to estimate PM2.5 concentration levels in Delhi's atmosphere. Heliyon. 2020;6(11):e05618. doi: 10.1016/j.heliyon.2020.e05618.
15. Yin PY, Yen AY, Chao SE, Day RF, Bhanu B. A machine learning-based ensemble framework for forecasting PM2.5 concentrations in Puli, Taiwan. Appl Sci. 2022;12(5):2484. doi: 10.3390/app12052484.

16. Danesh Yazdi M, Kuang Z, Dimakopoulou K, Barratt B, Suel E, Amini H, et al. Predicting fine particulate matter (PM2.5) in the greater London area: an ensemble approach using machine learning methods. Remote Sens (Basel). 2020;12(6):914. doi: 10.3390/rs12060914.
17. Satish N, Anmala J, Rajitha K, Varma MRR. A stacking ANN ensemble model of ML models for stream water quality prediction of Godavari River Basin, India. Ecol Inform. 2024;80:102500. doi: 10.1016/j.ecoinf.2024.102500.
18. Khashei M, Hajirahimi Z. A comparative study of series ARIMA/MLP hybrid models for stock price forecasting. Commun Stat Simul Comput. 2019;48(9):2625-40. doi: 10.1080/03610918.2018.1458138.
19. Singh ND, Sarada C, Ramasubramanian V, Sivaramane N, Krishnan M, Ravichandran S, et al. Forecasting of export prices of Indian shrimp: an ANN approach. Vision. 2023:09722629231186466. doi: 10.1177/09722629231186466.
20. Babu S, Thomas B. A survey on air pollutant PM2.5 prediction using random forest model. Environ Health Eng Manag. 2023;10(2):157-63. doi: 10.34172/ehem.2023.18.
21. Lin L, Liang Y, Liu L, Zhang Y, Xie D, Yin F, et al. Estimating PM2.5 concentrations using the machine learning RF-XGBoost model in Guanzhong urban agglomeration, China. Remote Sens (Basel). 2022;14(20):5239. doi: 10.3390/rs14205239.
22. Zhang L, Jánošík D. Enhanced short-term load forecasting with hybrid machine learning models: CatBoost and XGBoost approaches. Expert Syst Appl. 2024;241:122686. doi: 10.1016/j.eswa.2023.122686.
23. Xiao F, Yang M, Fan H, Fan G, Al-Qaness MA. An improved deep learning model for predicting daily PM2.5 concentration. Sci Rep. 2020;10(1):20988. doi: 10.1038/s41598-020-77757-w.
24. Kumar RP, Prakash A, Singh R, Kumar P. Machine learning-based prediction of hazards fine PM2.5 concentrations: a case study of Delhi, India. Discov Geosci. 2024;2(1):34. doi: 10.1007/s44288-024-00043-z.
25. Masood A, Ahmad K. Data-driven predictive modeling of PM2.5 concentrations using machine learning and deep learning techniques: a case study of Delhi, India. Environ Monit Assess. 2022;195(1):60. doi: 10.1007/s10661-022-10603-w.
26. Masood A, Hameed MM, Srivastava A, Pham QB, Ahmad K, Razali SF, et al. Improving PM2.5 prediction in New Delhi using a hybrid extreme learning machine coupled with snake optimization algorithm. Sci Rep. 2023;13(1):21057. doi: 10.1038/s41598-023-47492-z.
27. Shah V, Patel K, Gupta RK, Yagnik H. Advancements in air quality forecasting using artificial intelligence: a comprehensive analysis. In: 2024 Asia Pacific Conference on Innovation in Technology (APCIT). Mysore, India: IEEE; 2024. p. 1-6. doi: 10.1109/apcit62007.2024.10673466.
28. Mohan AS, Abraham L. An ensemble deep learning approach for air quality estimation in Delhi, India. Earth Sci Inform. 2024;17(3):1923-48. doi: 10.1007/s12145-023-01210-5.
29. Pande CB, Kushwaha NL, Alawi OA, Sammen SS, Sidek LM, Yaseen ZM, et al. Daily scale air quality index forecasting using bidirectional recurrent neural networks: case study of Delhi, India. Environ Pollut. 2024;351:124040. doi: 10.1016/j.envpol.2024.124040.
30. Gupta V, Gharekhan D, Samal DR. Machine learning based PM2.5 and 10 concentration modeling for Delhi city. J Indian Soc Remote Sens. 2025;53(1):81-99. doi: 10.1007/s12524-024-01962-7.