Published by :
http://www.ijert.org

Comparative Analysis of Machine Learning Models for AQI Prediction in Indian Metro Cities

Asif Iqbal
Department of Computer Science and Engineering
Jadavpur University, Kolkata, India

Priyankar Sarkar
Department of Computer Science and Engineering
Jadavpur University, Kolkata, India

Nandini Mukherjee
Department of Computer Science and Engineering
Jadavpur University, Kolkata, India

***Abstract*—This paper presents a comparative study of machine learning algorithms for predicting the Air Quality Index (AQI) in three major Indian metro cities: Delhi, Bengaluru, and Kolkata. The analysis was carried out using four machine learning models, including Linear Regression, Random Forest, XGBoost, and LSTM, on a dataset spanning from 2019 to 2023, which contained daily AQI values along with concentrations of key pollutants such as PM2.5 and PM10. The performance of each machine learning model was compared using established evaluation metrics such as mean absolute error (MAE), root mean squared error (RMSE), and R-squared (R2). The main findings of this study highlight how the inclusion of both particulate matter (PM2.5) and PM10 significantly influences model performance.**

***Keywords*—aqi; pm2.5; pm10; air quality prediction; machine learning**

I. INTRODUCTION (HEADING 1)

Air pollution is becoming a significant global issue. It has severe effects on human health, ecosystems, and overall quality of life. The World Health Organization noted that it is becoming the main thread in the environment. It is also responsible for approximately 4.2 million deaths each year due to outdoor pollution alone [1]. Rapid urbanization, especially in Africa and Asia, is exacerbating the problem of rising concentrations of pollutants, especially PM2.5 [2]. This situation requires urgent global action to mitigate its effects.

Particulate matter (PM2.5 and PM10) significantly affects air quality, leading to numerous major health impacts and environmental difficulties. PM particles (PM2.5 and PM10) are so small that they can fly throughout the air and enter the bloodstream and lungs deeply during breathing, resulting in cardiovascular and respiratory disorders.

Urbanization and industrialization in India, especially in metropolises like Delhi, Kolkata, and Bengaluru, have drastically deteriorated air quality. Manufacturing factories and industrial activities emit pollutants like PM2.5, PM10, nitrogen dioxide, etc., which are affecting the urban environment [3].

Rapid population migration has led to increased construction activities and vehicular emissions, particularly in cities like Delhi, Bengaluru, and Kolkata [4].

Industrial and transportation activities are expanding rapidly. Due to this, the complexity of the air quality monitoring process has increased. But on the other hand, it is essential to monitor and predict the air quality index (AQI). Traditional methods for air quality monitoring, such as monitoring stations, are costly and inadequate for fine-grained monitoring across extensive areas, such as metropolitan cities. Recent studies focus on the application of IoT-based frameworks and low-cost sensors for the monitoring of air quality in some areas. However, such a framework frequently yields inaccurate estimates of various atmospheric pollutants, resulting in misguided predictions of environmental risks. Recently, numerous researchers have been investigating various machine learning models for air quality index (AQI) prediction and utilizing supervised machine learning algorithms that are effective for measuring environmental pollution [5].

This paper presents a comparative study of machine learning algorithms for predicting the Air Quality Index (AQI) in three major Indian metropolitan cities: Delhi, Bengaluru, and Kolkata. Machine learning models commonly used for AQI prediction include Random Forest Regression (RF), XGBoost, Long Short-Term Memory (LSTM), and neural networks. In this study, we utilized a dataset spanning from 2019 to 2023, which contained daily AQI values along with concentrations of key pollutants such as PM2.5 and PM10. We included monthly temporal features, represented by sine and cosine transformations, to capture seasonal trends, given the daily availability of the data. We evaluated the models in two phases: first, we used PM2.5 and temporal features as input, and then we added PM10 as an additional input feature to evaluate its impact on AQI prediction. The performance of each machine learning model was compared using metrics such as Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), and R-squared (R^2). One of the key findings of this study highlights how the inclusion of both PM2.5 and PM10 significantly influences AQI prediction accuracy. The results provide insights into the strengths of different machine learning techniques and their effectiveness in predicting AQI for urban areas like Delhi, Bengaluru, and Kolkata.

This paper is organized as follows. It presents a state of art of our work in Section II, followed by the detailed methodology of our work in Section III, Rsults are shown in the Section III. Conclusion of our study is described in Section IV.

II. STATE OF ART

There are various approaches and areas that are already being investigated for air quality monitoring and prediction. These studies aim to understand pollution sources, assess health impacts, and develop mitigation strategies. These studies aim to understand pollution sources, assess health impacts, and develop mitigation strategies. Recent advancements in computational techniques have introduced machine learning as a powerful tool for AQI forecasting.

In a study, authors [6] have shown that various machine learning and deep learning algorithms are applied on a dataste containing various pollutants concentration for AQI prediction. They recorded random forest classifier predicts better than other.

Authors [7] has evaluated various data forecasting methods for predicting air quality index (AQI) values for PM 2.5. They propose a hybrid approach combining two deep learning models, long-short-term memory (LSTM) and gated recurrent unit (GRU), to predict AQI values. The model takes into account various pollutant contents and meteorological characteristics like temperature, humidity, and wind speed. The efficiency of the models is evaluated using RMSE, MAE, and R-squared values, and the proposed model shows the best performance.

The correlation between various air pollutants, including PM 2.5 and PM 10 concentrations, NO2, CO2, CO, etc are examined. The authors [8], have proposed a probabilistic voting ensemble based model to predict AQI. Different ML models, including simple linear regression (SLR), support vector regression (SVR), random forest (RF), and probabilistic voting ensemble, have been compared. RF outperformed PVE in terms of MAE and RMSE, while PVE outperformed RF in terms of R-squared.

A study has conducted by [9], using various machine-learning models to predict AQI levels across three regions. They found stacking ensemble, AdaBoost, and random forest as the best predictive models. The stacking ensemble outperformed AdaBoost in R-squared and RMSE.

Authors [10] have found that a hybrid deep learning model combining CNN and LSTM outperformed traditional algorithms and models in predicting air quality in a real dataset from Beijing.

In a study, authors have developed a multiple linear regression model that predicts air quality index (AQI) based on air pollutants and meteorological parameters, outperforming existing algorithms [12].

In a study, authors [12] employed multiple preprocessing techniques on a dataset, subsequently utilizing a support vector machine with a radial basis function kernel model. They say that their suggested model works better than other kernel support vector machine methods, like the long short-term memory (LSTM) and seasonal autoregressive integrated moving average (SARIMA) techniques.

The literature survey reveals that the majority of researchers have concentrated solely on applying and comparing machine learning algorithms for AQI prediction. Maximum author does not show what type of meteorological features and what type of temporal features they have used in their study. Either they have used the entire dataset or are doing some preliminary data processing.

To address these gaps, our study evaluates four machine learning models—linear regression, random forest, XGBoost, and LSTM—on datasets spanning five years (2019–2023) from three major Indian cities: Bengaluru, Delhi, and Kolkata. This comparative approach is unique in that it not only applies a diverse set of models but also emphasizes the roles of PM2.5 and PM10 in AQI prediction. By incorporating both pollutants and examining their combined effect on predictive accuracy, this work provides deeper insights into their significance. Our findings demonstrate the necessity of pollutant-specific modelling for AQI prediction, highlighting the enhanced performance of models when including PM10 alongside PM2.5. This broader scope of analysis contributes to a more comprehensive understanding.

This understanding of model effectiveness and pollutant influence serves as a valuable resource for future research. The study ultimately aims to inform and improve AQI management systems by integrating diverse machine learning techniques and pollutant-specific insights.

III. METHODOLOGY

This study follows a structured approach to investigate the performance of various machine learning models for predicting the air quality index (AQI). This study also analyses the impact of particulate matter (PM2.5 and PM10) on AQI prediction. Methodology consists of various steps, starting with the data collection process, followed by data preprocessing, feature selection, and model implementation. The next subsection provides a detailed description of these steps.

A. Data Collection

Daily air quality data for Delhi, Kolkata, and Bengaluru was collected from the Central Pollution Control Board’s (CPCB) website. The span of the data we have used is from the year 2019 to 2023. The dataset contains the concentration of various air pollutants, including PM2.5, PM10, carbon monoxide, ozone, etc., along with temporal data like month and year.

B. Data Preprocessing

A dataset can provide some meaningful insight; on the other hand, it also contains some noise, outliers, missing values, and inconsistencies. This unnecessary noise may affect the study negatively. Therefore, data preprocessing is crucial to address these unnecessary errors and noises. This stage involves the removal of unnecessary data from the dataset. As we already mentioned, our main focus is to analyze how particulate matter affects the AQI prediction for various cities, so we have removed all other pollution information. The imputation method handles null values. To handle the missing values in our case, we first used the mean imputation technique and then the KNN imputation technique. Mean imputation is a simple process and can handle continuous values like PM2.5 and PM10. On the other hand, KNN imputation is able to solve critical relationships between the data.

After addressing the missing data, our focus shifts to data normalization. Before feature engineering and model implementation, data normalization is a crucial step that ensures all the features in the model are on the same scale.

Some features may have larger magnitude datasets may dominate the modelling process due to their size. This type of dataset may hamper the overall performance of a model. Data normalization is a process that handles it. In our proposed work, the standardized scaling technique is used to normalize the input features to ensure that they all have the same scale or range. The standardization process modifies the data to achieve a mean of 0 and a standard deviation of 1. This guarantees that all features contribute uniformly to the model and inhibits any individual feature from prevailing due to variations in scale.

C. Feature Selection

Our primary goal in this work is to investigate the impact of particulate matter (PM2.5 and PM10) on the AQI prediction process and their correlation. To achieve this, we must conduct a feature selection process. We include only PM2.5 and PM10, along with the month, as input features, and our main target variable is the AQI.

D. Data Analysis

Generally, a dataset can provide insight into the location. Here, we have used five years of air quality data from three major metro cities in India. This motivates us to further explore the dataset. How are the pollutants correlated? What is the daily, monthly, and yearly pollution trend? We analyzed the data to answer this type of question and found some interesting stories behind them.

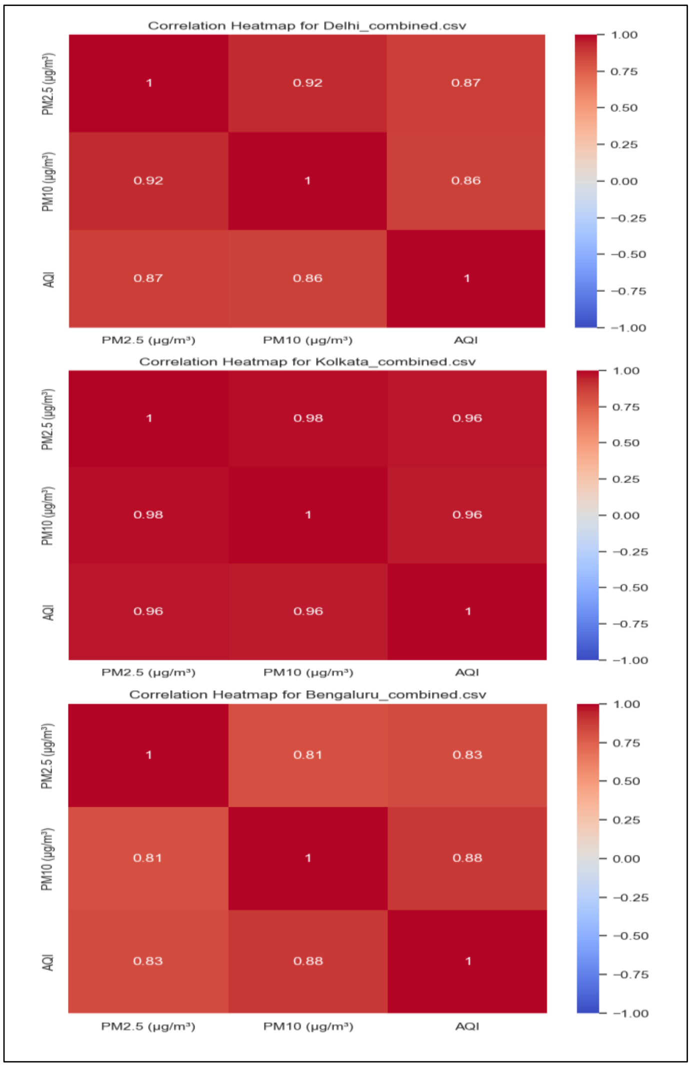

Fig. 1. Correlation Heatmap of 3 cities.

a. Correlation Matrix and Heat Map

We've started our data analysis process by creating a correlation matrix, followed by its visualization as a correlation heatmap. A correlation matrix can reveal the linear relationship between different features of a dataset. Generally, if two features have a strong positive correlation, then the value lies close to 1, and an increase in one feature is associated with an increase in the other. Conversely, if there is a weak to negligible significant relationship between the two features, the value will be close to 0, and a value close to -1 indicates a strong negative correlation. In our case, as shown in Figure 1, the analysis revealed a strong positive correlation between PM2.5, PM10, and AQI across all three cities: Delhi, Bengaluru, and Kolkata. This finding reinforces the significant role these pollutants play in determining air quality. The analysis motivated us to focus on these pollutants in our study, first evaluating models using PM2.5 alone and subsequently incorporating PM10.Additionally, Figure 3 demonstrates that among the three cities, Kolkata exhibited the strongest correlations between features, followed by Delhi and Bengaluru. This indicates that the relationship between particulate matter and AQI is more pronounced in Kolkata, which could reflect regional differences in pollutant sources and environmental conditions.

b. Pollution Trends: Yearly, Monthly and Seasonal Analysis

To better understand how the air quality changes over time, we have examined the yearly, monthly, and seasonal trends in AQI, PM2.5, and PM10 for each city: Delhi, Bengaluru, and Kolkata. We use line graphs to visualise these trends and demonstrate how pollution levels fluctuate over the years and seasons.

- Yearly Trend: In our study we have used a data span of 5 years, beginning from the year 2019 to the year 2023. One of the reasons we chose these years was the COVID-19 pandemic. We all know that India, along with several countries of the world, faced a pandemic that began from 2020 to 2022. The Government of India enforced a lockdown that started in March 2020 and extended up to April 2022 in various phases. The lockdown process involved the complete or partial closure of various industrial areas, plants, and factories. Traffic numbers were lower. Therefore, we eagerly anticipate the impact of these events on air pollution. The yearly analysis of data exhibits that PM2.5, PM10, and AQI have shown a general fluctuation over the five-year period, though the magnitude varies across the cities. For instance, Delhi consistently displays a higher pollution level over the entire period. The line graph shown in Figure 2, confirms that there is a little impact of the lockdown imposed by the government on the air quality of Delhi. Only in 2020, Delhi's air quality exhibits a lower AQI compared to all other years. Among these three cities, the air quality of Bengaluru is quite good compared to others. Figure 2 shows During the lockdown period, the air quality of Kolkata was healthier than earlier.
- Monthly Trend: Upon monthly analysis of the dataset, we observe a cyclical pattern in the pollutant levels and pollution levels of these metro areas. The graph for all three cities consistently shows an increase in pollution, starting in October and reaching a peak in either December or January. We also observe a decrease in both the pollution level and the pollutant level in all these cities from July to September. It is also observed from the figure 3, that Delhi shows a higher pollution level throughout the year, highest in the months of October to December. Kolkata, having a moderate pollution level, exhibits the lowest pollution level from May to September. Bengaluru city maintains the lowest pollution level throughout the entire year.

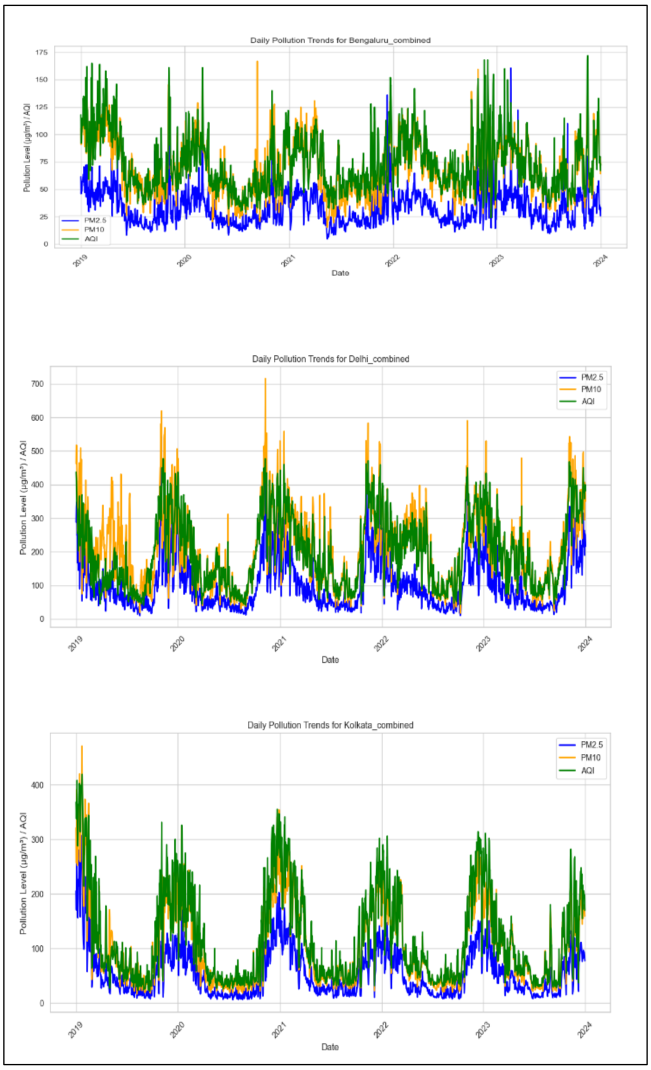

Fig. 2. Yearly Trend of Pollutants and Pollution of 3 cities.

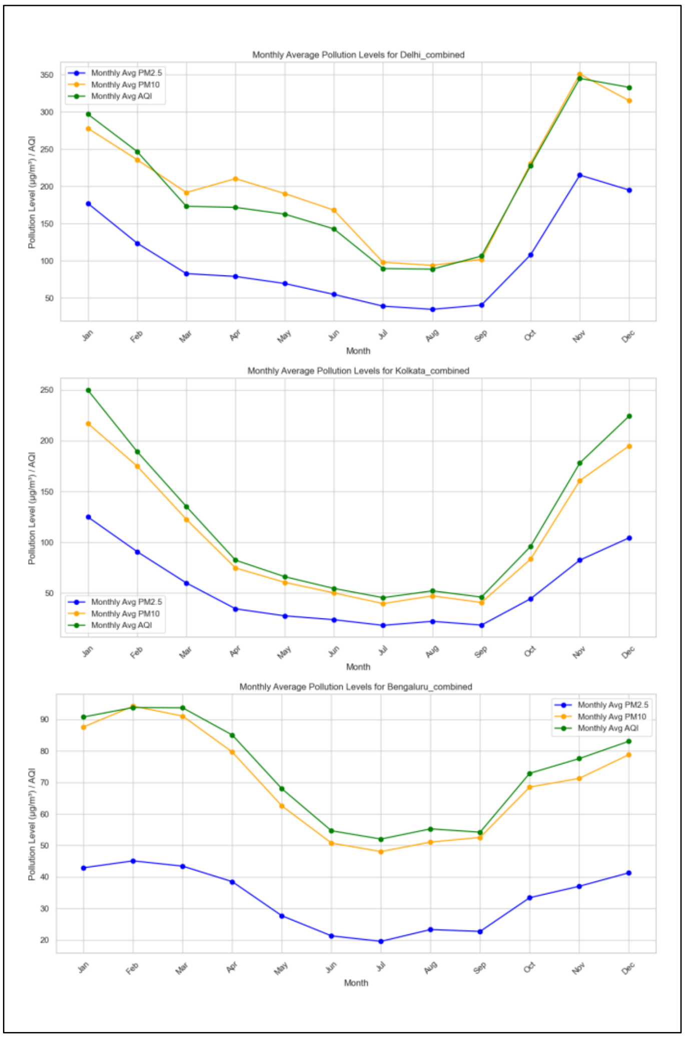

Fig. 3. Monthly Trend of Pollutants and Pollution of 3 Cities.

- Seasonal Trend: The seasonal trend is analysed and recorded in Figures 10, 8, and 9. We observe a common trend where the pollution level increases during the winter and decreases during the monsoon season. Winter consistently showed the highest pollution levels, with AQI often crossing the severe threshold in Delhi. Summer and monsoon seasons displayed a significant reduction in particulate matter, especially in Bengaluru and Kolkata, highlighting the impact of natural cleansing mechanisms such as wind and rain. Post-monsoon (October) showed a sharp increase in pollution levels, particularly in Delhi, due to post-harvest stubble burning and festival-related activities like firecrackers.

E. Model Implementation

In this study, one of our main objectives is to make a comparison between the performance of various different machine learning models for AQI prediction. In order to complete this, we have used four different machine learning models on preprocessed data from three metropolitan cities in India, including Delhi, Kolkata, and Bengaluru. We evaluated and compared the models using established evaluation metrics such as mean squared error (MSE), root mean squared error (RMSE), mean absolute error (MAE), and R-squared. We will delve deeper into the process of splitting the dataset and the models employed in the next subsection.

1. Dataset Splitting Strategy

In our case, we have divided the entire dataset into three parts for each city. 60% of the data was used to train the models. In this learning process, the models learn about the underlying pattern of the data and the hidden relationship between the features. 20% of the data are used for the validation set. We use it to adjust the model's hyperparameters and evaluate its performance during the training phase. Validation sets generally prevent overfitting issues by evaluating the model on unseen data. The Remaining 20% of the data is used to test the model. These data are kept separate from the training process to provide an unbiased evaluation of the final model's performance. This splitting ratio ensures that the models are trained by using sufficient data while keeping enough data for validation and testing.

2. Feature Selection

Selecting a feature from the entire dataset is crucial for training the model effectively. In our case, we not only focused on AQI prediction, but also sought to understand the crucial role of particulate matter (PM2.5 and PM10) in AQI prediction. To achieve this, we have selected PM2.5 and PM10 as features in our training dataset, and we have also added cosine and sine transformations of months as temporal features.

3. Machine Learning Models

We have chosen four machine learning models for our study. We chose linear regression, random forest, XGBoost, and LSTM for this study to provide a diverse range of approaches and leverage their strengths in addressing the AQI prediction.

Linear regression is a fundamental machine learning model and can establish a baseline performance for the task. It assumes a linear relationship between the input features and target variables. As we have found a strong correlation (mentioned in Section III.D.a) between the input features and the target variable of our case, we decided to start with linear regression.

Random Forest is a robust ensemble learning method that is based on decision trees. This model effectively captures the non-linear relationships between features without requiring extensive preprocessing. This reason helps us decide whether to include this model in our study.

XGBoost is known for its powerful and efficient gradient-boosting framework. It combines the strengths of boosting to improve accuracy while preventing overfitting through regularization. This model is helpful for AQI prediction as it has the ability to handle non-linearities.

Long Short-Term Memory is a type of recurrent neural network (RNN) designed to capture sequential patterns and dependencies over time. In our particular case, we want to examine how the neural network reacts with our dataset, so we have included this in our study.

The next section of this paper discusses the detailed results and analysis of this study.

IV. RESULT AND DISCUSSION

This section primarily focuses on our empirical investigation into AQI prediction using particulate matter concentration analysis. The dataset spanning from the year 2019 to 2023 for three Indian metro cities (Bengaluru, Delhi, and Kolkata) is divided into training (60%), validation (20%), and testing (20%) sets to make a robust performance evaluation of the machine learning model.

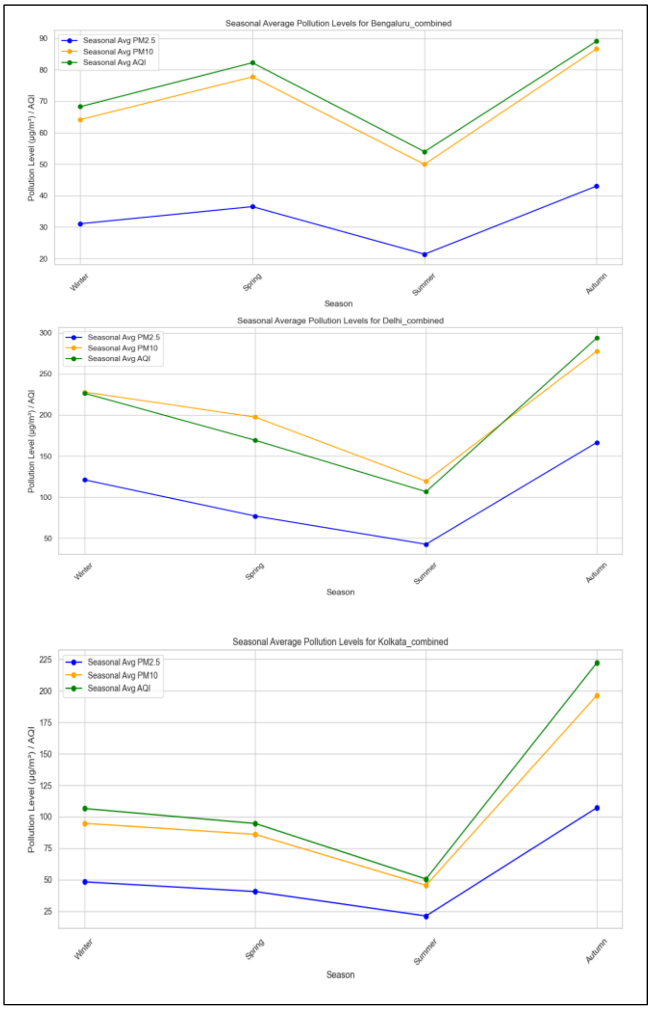

Fig. 4. Seasonal Trend of Pollutants and Pollution of 3 Cities.

We have implemented four machine learning models—linear regression, random forest, XGBoost, and LSTM—for AQI prediction. We evaluated the models in two phases: first, we used PM2.5 as a meteorological feature and the sine and cosine transformation of the month as a temporal feature, and in the second phase, we incorporated PM10 as a meteorological feature alongside PM2.5. This approach aimed to understand the individual and combined effect of particulate matter on AQI prediction.

We evaluated the performance of these models using some recognized regression metrics. The following metrics are used to assess the performance of the model:

1. Mean Squared Error (MSE): It measures the average squared difference between the predicted and actual values, where lower values indicate better performance.
2. Root Mean Squared Error (RMSE): The square root of MSE, providing interpretability in the same units as AQI.
3. Mean Absolute Error (MAE): It expresses the average magnitude of errors, providing a direct measure of prediction accuracy.
4. R-Squared: It indicates the proportion of variance explained by the model.

1. Comparative Performance Analysis

Now in this section we make a discussion about the performance of the machine learning models one by one as follows:

- Linear regression: We can clearly observe from Table 1 that this model showed a strong performance in terms of Q-squared for all three cities, particularly Kolkata. The inclusion of PM10 has also improved the accuracy across all the cities. For instance, from Tables 1 and 2, it is clear that in the case of Bengaluru, the inclusion of PM10 increases the R-squared value from 0.74 to 0.82. It clearly indicates that the addition of PM10 has a strong influence on AQI prediction.
- Random Forest: This model has shown consistent performance across the cities. In maximum cases, it outperforms the linear regression, especially for Delhi. Table 1 describes the R-squared value for Delhi in the case of the random forest model, which has increased from 0.78 to 0.83, and for Kolkata, the same metric got a value of 0.94 instead of 0.93. Inclusion of PM10 drops the MSE value to 0.7 (shown in Table 2), which is very low and proof of how PM10 has a strong influence in AQI prediction.
- XGBoost: The performance is competitive, especially for Delhi and Kolkata. The model yielded a lower R-squared value compared to the random forest method. From the Tables 1 and 2, it is clear that for this model, the inclusion of PM10 has not had any significant role in AQI prediction.
- LSTM: LSTM struggled compared to the tree-based models, especially for Delhi and Bengaluru. However, for Kolkata, it performed relatively better, with an R^2 of approximately 0.85 for both the cases (Tables 1 and 2). The sequential nature of LSTM made it more sensitive to the data characteristics.

All the models in the three cities significantly performed better while PM10 was added as a meteorological feature with PM2.5 and temporal feature. For instance, in Delhi, the R-squared value for Linear Regression increased from 0.78 to 0.83, indicating a more robust predictive capability with the added feature. Similarly, Random Forest showed a significant improvement in Bengaluru, with the R-squared value rising from 0.71 to 0.78. In Kolkata, the RMSE for XGBoost decreased from 0.41 to 0.40, demonstrating enhanced accuracy in error minimization. While LSTM models exhibited only minor improvements with the inclusion of PM10, the overall performance remained less optimal compared to the other techniques. These results underscore the value of incorporating PM10 in predicting AQI, as it captures additional variability that PM2.5 alone could not explain.

At the end of this discussion, we have to confess some limitation which can be addressed later. LSTM model underperformed in this study, which need a proper intervention regarding this. Further exploration, hyperparameter tuning and feature engineering could improve the result for deep learning model like LSTM. We can consider all other major meteorological and temporal features with these datasets for better accuracy.

Tbale. 1. Validation and Test Result for only PM2.5 as Feature

Model	**City**	**Validation Result**				**Test Result**			
		MSE	**RMSE**	**MAE**	**R2**	**MSE**	**RMSE**	**MAE**	**R2**
Linear Regression	Bengaluru	28%	53%	38%	74%	26%	51%	38%	74%
	Delhi	23%	48%	35%	79%	24%	49%	35%	78%
	Kolkata	8%	29%	20%	92%	6%	24%	17%	93%
Random Forest	Bengaluru	33%	58%	41%	69%	29%	54%	37%	71%
	Delhi	22%	46%	33%	80%	18%	43%	31%	83%
	Kolkata	7%	27%	17%	93%	6%	24%	16%	94%
XGBoost	Bengaluru	38%	61%	45%	65%	35%	59%	45%	65%
	Delhi	29%	54%	45%	73%	30%	54%	45%	73%
	Kolkata	21%	46%	37%	81%	17%	41%	34%	81%
LSTM	Bengaluru	33%	58%	42%	65%	36%	60%	41%	57%
	Delhi	26%	51%	40%	69%	19%	43%	34%	81%
	Kolkata	8%	29%	20%	90%	9%	31%	22%	85%

Model	**City**	**Validation Result**				**Test Result**			
		MSE	**RMSE**	**MAE**	**R2**	**MSE**	**RMSE**	**MAE**	**R2**
Linear Regression	Bengaluru	22%	47%	32%	80%	18%	43%	30%	82%
	Delhi	19%	43%	30%	82%	19%	43%	30%	83%
	Kolkata	8%	28%	19%	93%	5%	23%	23%	94%
Random Forest	Bengaluru	22%	47%	33%	80%	21%	46%	33%	78%
	Delhi	17%	42%	28%	84%	15%	38%	26%	86%
	Kolkata	7%	26%	16%	94%	5%	23%	15%	94%
XGBoost	Bengaluru	34%	58%	42%	69%	28%	53%	41%	72%
	Delhi	28%	53%	45%	73%	28%	53%	43%	74%
	Kolkata	21%	46%	37%	80%	17%	41%	34%	81%
LSTM	Bengaluru	33%	58%	42%	65%	33%	58%	41%	61%
	Delhi	25%	50%	38%	70%	17%	41%	33%	83%
	Kolkata	8%	29%	20%	90%	9%	30%	22%	85%

Tbale. 2. Validation and Test Result for both PM2.5 and PM10 as Features

V. CONCLUSION

This study presents a comparative analysis of machine learning models for predicting the Air Quality Index (AQI) in three Indian metro cities: Bengaluru, Delhi, and Kolkata. The analysis was carried out using four models, including Linear

Regression, Random Forest, XGBoost, and LSTM—on a dataset spanning five years (2019–2023). The primary objective was not only to predict AQI but also to understand the impact of particulate matter (PM2.5 and PM10) on model performance. The results demonstrate that the inclusion of PM10 with PM2.5 significantly improves model performance across all cities. Among the models, Random Forest and XGBoost consistently outperformed others in terms of accuracy, robustness, and interpretability, particularly in Delhi and Kolkata. On the other hand, LSTM, though capable of capturing temporal patterns, underperformed compared to tree-based models due to the limited temporal complexity in the dataset.

City-wise, Kolkata exhibited the strongest correlations among features, resulting in the highest predictive performance across all models. Delhi's data proved well-suited for tree-based models, while Bengaluru showed moderate predictive accuracy due to its relatively complex pollution patterns. This work emphasizes the utility of machine learning techniques in environmental monitoring and decision-making processes. The findings support for incorporating both PM2.5 and PM10 in AQI prediction models and underscore the need to tailor models to city-specific characteristics for improved accuracy. Future research could explore more advanced deep learning architectures or incorporate additional meteorological and temporal features to further enhance AQI predictions.

ACKNOWLEDGMENT (HEADING 5)

First author of this article, Asif Iqbal is working as a JRF under AICTE Doctoral Fellowship (ADF) provided by AICTE, Government of India.

REFERENCES

[1] Shaddick, G., Thomas, M.L., Mudu, P. et al. Half the world's population are exposed to increasing air pollution. npj Clim Atmos Sci 3, 23 (2020). https://doi.org/10.1038/s41612-020-0124-2.

[2] Qin Zhou, Mir Muhammad Nizamani, Hai-Yang Zhang et al. The Air We Breathe: An In-Depth Analysis of PM2.5 Pollution in 1312 Cities from 2000 to 2020, 27 April 2023, PREPRINT (Version 1) available at Research Square [https://doi.org/10.21203/rs.3.rs-2740958/v1].

[3] Nawhath Thanvisitthpon, Kraiwuth Kallawicha, H. Jasmine Chao, Chapter 10 - Effects of urbanization and industrialization on air quality, Editor(s): Mohammad Hadi Dehghani, Rama Rao Karri, Teresa Vera, Salwa Kamal Mohamed Hassan, Health and Environmental Effects of Ambient Air Pollution, Academic Press, 2024, Pages 231-255, https://doi.org/10.1016/B978-0-443-16088-2.00003-X.

[4] Ena Jain, & Debopam Acharaya. (2023). A Mobile Sensing Based Stochastic Model to Forecast AQI Variation of Pollution Hotspots on Urban Neighborhoods. International Journal of Next-Generation Computing, 14(2). https://doi.org/10.47164/ijngc.v14i2.1195

[5] Hamdi A. Al-Jamimi, Sadam Al-Azani, Tawfik A. Saleh, Supervised machine learning techniques in the desulfurization of oil products for environmental protection: A review, Process Safety and Environmental Protection, Volume 120, 2018, Pages 57-71, https://doi.org/10.1016/j.psep.2018.08.021.

[6] Sanjeev, D. (2021). Implementation of machine learning algorithms for analysis and prediction of air quality. International Journal of Engineering Research & Technology (IJERT), 10.

[7] Sarkar, N., Gupta, R., Keserwani, P. K., and Govil, M. C. (2022). Air quality index prediction using an effective hybrid deep learning model. Environmental Pollution, 315.

[8] Xiang, X., Fahad, S., Han, M. S., Naeem, M. R., and Room, S. (2023). Air quality index prediction via multi-task machine learning technique: spatial analysis for human capital and intensive air quality monitoring stations. Air Quality, Atmosphere and Health, 16:85–97.

[9]] Liang, Y. C., Maimury, Y., Chen, A. H. L., and Juarez, J. R. C. (2020). Machine learning-based prediction of air quality. AppliedvSciences (Switzerland), 10:1–17.

[10] Hu, K., Guo, X., Gong, X., Wang, X., Liang, J., and Li, D. (2022). Air quality prediction using spatio-temporal deep learning. Atmospheric Pollution Research, 13.

[11] Sigamani, S. and Venkatesan, R. (2022). Air quality index prediction with influence of meteorological parameters using machine learning model for iot application. Arabian Journal of Geosciences, 15.

[12] Maltare, N. N. and Vahora, S. (2023). Air quality index prediction using machine learning for ahmedabad city. Digital Chemical Engineering, 7.