International Journal of Environmental Sciences
ISSN: 2229-7359
Vol. 11 No. 20s, 2025
https://www.theaspd.com/ijes.php

AI-Driven Air Quality Forecasting Models for Urban Pollution Management: A Comparative Study of Machine Learning Algorithms in Developing Economies

Swetambari Waghmare[1], Dr. A. Devendran[2], Dr P Marishkumar[3], Ramkumar M[4], Dr. Saurabh Chandra[5], D S Ajithakala[6]
[1]Assistant Professor, Bharati Vidyapeeth College of Engineering, Navi Mumbai, waghmare@bvcoenm.edu.in
[2]Professor, School of Business, Woxsen University, Hyderabad, devendran.alagarsamy@gmail.com
[3]Associate Professor, Department of Management, Vinayaka Mission's Kirupananda Variyar Engineering College, Vinayaka Mission's Research Foundation (Deemed to be University), Salem, Tamilnadu, marishkumarp@vmkvec.edu.in
[4]Department of Management, Vinayaka Mission's Kirupananda Variyar Engineering College , Vinayaka Mission's Research Foundation (Deemed to be University) Salem, Tamilnadu
[5]Associate Professor, School of Law ,Bennett University, saurabhchandranls@gmail.com, ORCID ID : https://orcid.org/0000-0003-4172-9968
[6](AP/CSE), VSB college of Engineering Technical Campus, ajithakalavsb@gmail.com

Abstract: *Air pollution in urban environments is a major environmental and health issue, especially in the developing economies where the current indicators of environmental degradation are high due to rapid industrialization, automotive emissions and failure to enforce regulations on vehicles and industries. Proper prediction of air quality indices (AQI) is critical in activating mitigation policies and guidance of policy measures. Although traditional forecasting methods are of value in their own way, they tend to not fully capture the nonlinear, complex and spatio/temporal processes of urban air pollution. This work discusses the use of modelling with the use of Artificial Intelligence (AI) and specifically the Machine Learning (ML) in the context of predicting air quality to enable smarter management of urban pollution. It takes a comparative framework to compare the efforts of four mainstream ML models, which are Random Forest (RF) Support Vector Machines (SVM) Long Short-Term Memory (LSTM), and Extreme Gradient Boosting (XGBoost) on real-time and past air quality data of chosen metropolitan cities, India, Bangladesh, and Nigeria. The used methodology combines measurements of the monitoring stations of the air pollution levels located in urban areas and meteorological data and pollutant concentrations (PM2.5, PM10, NO 2, SO 2, CO, O 3). Statistical measures of performance of the models include Coefficient of Determination (R 2), Root Mean Square Error (RMSE), and Mean Absolute Error (MAE). According to the preliminary findings it has been observed that LSTM presents the best results regarding the temporal dependencies, and Random forest presents the best results regarding high pollution cases. Volumetric visualizations and GIS-based overlaps are also provided to demonstrate the difference in hotspots of the pollution and the sensitivity of various algorithms to the variability of data in locals. The results support the possibility of AI-driven forecasting as an expandable and flexible agent of urban environmental planning as well as emergency response systems in resource-limited conditions. Such a comparative study does not only reveal the strengths and weaknesses of the algorithms but also offers a plan of how to implement the predictive AI models in the realm of public air quality government in developing economies.*
Keywords: *AI forecasting, Air Quality Index (AQI), Machine Learning, Urban pollution, Developing economies, Random Forest, LSTM, SVM, XGBoost, Environmental informatics, Spatiotemporal prediction, Urban air monitoring, GIS, Public health modeling, Pollution management strategies*

I. INTRODUCTION

The problem of air pollution is one of the most urgent ones and those of environment and populace health in the 21 st century. Its negative impacts are prominent especially in cities of some developing economies where high rates of industrialization, smoky vehicles, and poor implementation of policies contribute to increased environmental damages. The world health organization (WHO), reveals that over 90 percent of the world population lives with air that surpasses recommended pollution levels and that over 7 million premature deaths have been recorded annually as result of exposure to pollution air. Indian, Bangladeshi, Nigerian and other low- and middle-income cities regularly appear in global lists of cities with high air pollution rates, the techniques of which include critical concentrations of particulate matter (PM2.5 and PM10), nitrogen oxides (NOx), sulfur dioxide (SO 2), and other air particles of toxins in the air causing respiratory diseases, heart-related diseases, and reduced life quality. The nature of the air pollution that can be found in a given Urban environment is highly complex in nature in that the various factors that would be at play are so many because several interacting factors cause it. The exchange of information between the non-linear and spatiotemporal dynamic of pollution concentrations in heavily populated and rapidly changing metropolitan environments is often poorly represented using traditional methods of air quality forecasting, e.g. indirect statistical time series models or deterministic simulation approaches. Such traditional models demand that the input data be highly resolved and domain knowledge be large, and that such models be sensitive to uncertainties in the emission inventories and meteorological inputs. This does not make them powerful in dynamic environments where appropriate local and accurate forecasts are indispensable. With the implementation of Artificial Intelligence (AI) and Machine Learning (ML) in environmental monitoring systems in the past couple of years, the design of models of air pollution and its management given a paradigm shift. The BP algorithm can even handle massive volumes of irregular data, recognize and cope with complicated trends, and acutely learn when environmental conditions shift. Using both supervised and unsupervised learning strategies, AI-based forecasting models will be able to train at an increased level of accuracy, reliability, and scalability compared to its traditional predecessor at forecasting the levels of Air Quality Index (AQI) and pollutant levels. Random Forest (RF), Support Vector Machines (SVM), Long Short-Term Memory (LSTM) neural networks, and Extreme Gradient Boosting (XGBoost) are some of the most popular ML methods at forecasting air quality. Each of these algorithms has various strengths: RF is well known to be robust to overfitting and effective at high-dimensional multi-variate data; SVM can work well with high-dimensional feature spaces; LSTM can learn long-term temporal structure, which led to its interest in time-series forecasting; and XGBoost is a fast and accurate gradient boosting algorithm with automatic regularization. An explicit comparison of these models, especially in terms of developing economies having different quality of data and infrastructural constraints, is yet to be explored despite the growing use of such models in environmental science. This will help in bridging that knowledge gap by undertaking a comparative study of the AI models in air quality forecasting of the urban centers of these countries. These two countries are prototypical cases of developing economies where economic growth, largely the urbanisation and industry development compounded by ineffective environmental management can lead to high-level air pollution disasters. Cities chosen (e.g. Delhi, Dhaka, Lagos) are not only highly polluted on a regular basis but also have different monitoring infrastructure, vulnerability of the population health and data availability which provides a locally-complicated, nuanced testing ground to define algorithmic evaluation. The key research questions of the study are as follows: (1) examining the different machine learning algorithms in predicting the AQI and the concentration level of the pollutants in urban areas with real-time and historical data; (2) testing the capacity of the models in addressing the noise and missing values, as well as the non-stationary feature present in the datasets of developing countries; and (3) to create spatial and temporal maps that can serve as the basis of pollution control measures and policy action. To this extent, the study combines the information on the air quality of monitoring stations within cities with some weather parameters, including temperature, humidity, wind speed, and atmospheric pressure. Such inputs undergo preprocessing, and each of the ML models is trained and tested by applying cross-validation methods. The performance is also measured based on standard evaluation tools such as Root Mean Square Error (RMSE), Mean Absolute Error (MAE) and the Coefficient of Determination (R 2). Moreover, Geographic Information Systems (GIS) can be utilized in visualizing the results in terms of pollution hotspots and pattern over time at urban areas. An important feature of this work is the practical orientation factor. In developing nations, limitations to the advancement of environmental forecasting lies in their coverage of sensors, sectors of constant data gathering and gaps related to infrastructure. Consequently, precision is considered as one metric of evaluating the tested models in this work with another crucial aspect being the computational efficiency, data tolerance, as well as flexibility to resource poor environments. The aim is to find realistic models that can be implemented into current systems within the workings of the environment or planned responses within the urban sectors or in the community level scale. Overall, this research can add to the existing developing literature on the role of AI in environmental management due to comparing regionally specific machine learning algorithms applied in air quality forecasting. Its observation on the distinct problems of large cities of economically developing countries enables it to not only fill the gap between highly theoretical computational modelling and implementation into reality, but also gives insight into the problems of developing economies in general. The knowledge that can be acquired as a result of this study can be used in the formulation of policies, creation of awareness in the citizens and implementation of environmental justice programs to make the cities more sustainable and healthier in future.

II. RELEATED WORKS

The emerging complexity in the collision of urban atmosphere pollution has led scholars and policy makers to consider adopting Artificial Intelligence (AI) into air quality predictive systems. Deterministic simulations and regression-based models have step by step been replaced by more adaptive and scalable machine learning (ML) methods that provide higher accuracy levels, enhanced and variable spatiotemporal resolution, as well as adaptability in analyzing heterogeneous data. Different ML algorithm performance in such air quality prediction has been explored by various studies in this context, especially in the urban areas of developing economies within which data abnormality and infrastructural limits are peculiar issues. Zhang et al. [1] conducted a fundamental research in the applicability of Random Forest (RF) and Gradient Boosting as a predictive model to estimate the concentration of PM 2.5 in Chinese cities and discovered that the ensemble model performed powerful approximations of non-linearity of pollutant because they substantially outperformed linear models in capturing their observation. Khan et al. [2] corroborated the findings in applying an LSTM neural net when predicting AQI readings in New Delhi, India and reported that deep learning models had a better temporal memory especially when LSTMs were applied during the high-pollution season (winter smog). They compared the Support Vector Regression (SVR) with RF to predict PM10 and NO2 in Indian urban centres in a comparative study by Sahu and Yadav [3]. They concluded that their results indicated more consistent forecasts could be expected using RF when forecasting across several types of pollutants, but that SVR provided better resolution in feature dense sensor networks. This contrasts between stability and sensitivity of models that have been re-occurring due to urban air quality 2990odelling has come about. More recently, Ahmed et al. [4] conducted an assessment of the XGBoost use in the prediction of AQI in Dhaka, Bangladesh. The research highlighted the fact that XGBoost could deal well with missing values and outliers which are rampant in developing country data sets. They scored 0.89 of R 2 in the prediction of PM2.5, and their model was better than traditional neural networks. In a similar case, Oluwole et al. [5] applied the machine learning to Lagos in Nigeria that also included meteorological variables including the humidity, temperature, and wind speed. The environmental feature addition to the ML algorithms in their findings proved that the accuracy of the models was very high and the seasonal transition accuracy was very good. From a broader methodological perspective, survey articles such as that by Salim et al. [13] offer comprehensive comparisons of ML algorithms in environmental forecasting. Their meta-analysis of 42 studies concludes that ensemble methods like RF and boosting models consistently outperform simple linear regressions, especially in heterogeneous and urban datasets. Furthermore, interdisciplinary research is emerging that ties AI-based pollution forecasts with public health outcomes. In a study by Musa et al. [14], daily PM2.5 predictions were correlated with hospital admission rates for asthma in Accra. The researchers noted that real-time forecasts allowed for anticipatory healthcare planning and improved patient outcomes, especially for vulnerable populations. Finally, recent work by Thomas et al. [15] emphasized the need for algorithmic transparency and explainability in AI models deployed for air quality monitoring. They proposed an interpretable AI framework using SHAP (Shapley Additive exPlanations) values, which elucidates how specific variables (e.g., temperature or traffic congestion) contribute to forecasted AQI spikes. This improves trust in AI systems, a crucial factor for public policy adoption in developing countries. Together, these studies highlight the growing body of research supporting the use of machine learning and AI for air quality forecasting. However, many remain focused on isolated regions or a single 2991odelling technique. This paper addresses the gap by conducting a cross-country comparative analysis of multiple AI models within the urban settings of developing economies, integrating not only pollutant and meteorological data but also geospatial visualization for actionable insights. Methodologically, a wider scope consisting of survey articles reports like that of Salim et al. [13] give a detailed comparison of the ML algorithms in environmental forecasting. They meta-analyze 42 papers to find that manifold-based learning (RF and boosting models) consistently beat simple linear regressions, and they work particularly well with heterogeneous and urban data. Moreover, there is a new, interdisciplinary direction of research that links AI-based predictions of pollution to the health outcomes of people. In a research conducted by Musa et al. [14], hospital admission rates of asthma in Accra were related to daily PM2.5 predictions. The researchers observed that real-time predictions enabled anticipatory planning of healthcare and better patient outcomes with special consideration to vulnerable groups of patients. Lastly, Thomas et al. recently published an article stating that it is essential to have transparency and explainability in the used AI models when monitoring air quality [15]. They suggested a model interpretation or explainable AI model based on SHAP (Shapley Additive exPlanations) values (SCS blog, 2020) which allows understanding how particular factors (e.g., temperature or traffic level) influence the AQI increase that it predicts. This enhances the trust to the AI systems, Ih Is Important In the adoption of the worldwide public policy in developing nations. Cumulatively, these studies bring into focus the significant amount of research in favor of air quality forecasting machine learning and AI. Nevertheless, lots are concentrated on the single areas or method of 2991odelling. The present paper fills this gap by performing a cross-country comparative assessment of various AI models in the cities of the developing economies where not only pollutant and meteorological but also geospatial visualization data on actionable insights are incorporated.

III. METHODOLOGY

3.1 Research Design

This study adopts a spatial-temporal comparative modeling approach that integrates environmental data analysis with machine learning to predict air quality indices (AQI) across selected urban zones. The framework includes supervised learning algorithms trained on pollutant and meteorological datasets, followed by model performance evaluation and geospatial visualization. The key objective is to assess how various AI algorithms perform under typical constraints found in developing economies, such as data irregularities, limited sensor coverage, and inconsistent temporal sampling. The workflow consists of five stages: data acquisition, preprocessing, model training, prediction, and spatial analysis [16].

3.2 Study Area Selection

Three urban centers were selected to represent varying levels of urban pollution, infrastructure, and meteorological patterns across different developing economies:

- **Delhi, India:** Characterized by chronic winter smog, industrial density, and traffic congestion.
- **Dhaka, Bangladesh:** A fast-growing city facing severe air quality issues due to brick kilns and diesel engines.
- **Lagos, Nigeria:** Known for vehicle-induced pollution, fuel combustion, and unplanned urbanization.

These cities were chosen due to data availability, policy relevance, and contrasting climate profiles [17].

Table 1: Characteristics of Selected Urban Study Areas

City	Dominant Pollution Sources	Climate Type	Sensor Infrastructure	Common Pollutants
Delhi	Vehicles, industries, dust	Semi-arid	Dense (SAFAR, CPCB)	PM2.5, PM10, NO_2, CO

Dhaka	Brick kilns, traffic, biomass	Tropical monsoon	Moderate (DOE Network)	PM2.5, NOx, SO_2
Lagos	Diesel vehicles, open burning	Tropical wet/dry	Sparse (NIHSA)	PM10, CO, SO_2

3.3 Data Collection and Preprocessing

Data were collected for a 24-month period (Jan 2022–Dec 2023) using:

- **Air quality readings** from OpenAQ, CPCB (India), DOE (Bangladesh), and NIHSA (Nigeria).
- **Meteorological data** from Weather Underground APIs and national weather stations, including temperature, humidity, wind speed, and barometric pressure [18].

Data preprocessing involved:

- Handling missing values using mean imputation for meteorological variables and linear interpolation for pollutant concentrations.
- Normalization via MinMax scaling to standardize variable ranges.
- Feature engineering by introducing lag variables (t-1, t-2 days) to improve temporal predictions.
- Outlier removal using the interquartile range (IQR) method [19].

3.4 Machine Learning Models Employed

Four algorithms were trained and tested across all cities:

- **Random Forest (RF):** Ensemble-based method effective for nonlinear relationships.
- **Support Vector Machine (SVM):** Kernel-based model useful for high-dimensional feature space.
- **Long Short-Term Memory (LSTM):** A deep learning recurrent neural network ideal for time-series forecasting [20].
- **Extreme Gradient Boosting (XGBoost):** Fast, regularized gradient boosting method suitable for noisy and sparse datasets.

All models used pollutant + meteorological features as input to predict next-day AQI.

Table 2: Machine Learning Algorithms and Model Features

Algorithm	Type	Strengths	Limitations
RF	Ensemble	Handles multicollinearity, robust to overfitting	Less temporal insight
SVM	Regression	Performs well on small datasets	Poor scalability for large datasets
LSTM	Deep Learning	Captures time dependency effectively	Needs large datasets, slow to train
XGBoost	Gradient Boosting	Fast training, works well on sparse data	Sensitive to parameter tuning

3.5 Model Evaluation Metrics

Model performance was assessed using:

- **Root Mean Square Error (RMSE):** Penalizes large errors.
- **Mean Absolute Error (MAE):** Indicates average deviation.
- **Coefficient of Determination (R^2):** Measures the variance explained by the model.

Table 3: Performance Metrics Definitions

Metric	Formula	Purpose
RMSE	1n∑(yi−y^i)2\sqrt{\frac{1}{n} \sum (y_i - \hat{y}_i)^2}n1∑(yi−y^i)2	Penalizes large deviations
MAE	(\frac{1}{n} \sum	y_i - \hat{y}_i
R^2	1−∑(yi−y^i)2∑(yi−y¯)21 - \frac{\sum (y_i - \hat{y}_i)^2}{\sum (y_i - \bar{y})^2}1−∑(yi−y¯)2∑(yi−y^i)2	Explains variance in target

Each model was trained using 70% of the data and tested on the remaining 30%, with 5-fold cross-validation applied to reduce overfitting [21].

3.6 Visualization and Spatial Analysis

Using GIS platforms like QGIS and Google Earth Engine (GEE), spatial heatmaps were created to visualize high-risk pollution zones. Model outputs were integrated with shapefiles of urban boundaries, and hotspot clustering was identified using Moran's I index. Kriging interpolation was used to estimate AQI surfaces across unmonitored areas [22].

In addition, time-series charts and heatmaps were generated in Python using Seaborn and Plotly to compare predicted vs. actual AQI trends.

3.7 Limitations and Assumptions

Several assumptions and limitations were noted:

- Sensor inconsistencies and gaps in real-time AQI data led to the need for imputation, which may reduce temporal resolution.
- LSTM models required significantly more data and tuning, which could affect generalizability in low-data environments.
- External factors like festival pollution spikes or unreported industrial events were not captured in training datasets.
- Satellite-based AQI estimation was not integrated in this phase but is recommended for future extension [23].

IV. RESULT AND ANALYSIS

4.1 Model Performance Comparison Across Cities

The predictive accuracy of the four machine learning models—Random Forest (RF), Support Vector Machine (SVM), Long Short-Term Memory (LSTM), and Extreme Gradient Boosting (XGBoost)—was evaluated across the three urban sites. Performance was measured using RMSE, MAE, and R^2 metrics.

Table 4: Model Performance Metrics by City

City	Model	RMSE	MAE	R^2
Delhi	RF	23.4	18.6	0.84
	SVM	27.2	21.8	0.79
	LSTM	19.5	15.1	0.89
	XGBoost	21.1	16.8	0.86
Dhaka	RF	24.7	19.4	0.81
	SVM	29.5	23.0	0.75
	LSTM	20.3	16.2	0.87
	XGBoost	22.5	17.5	0.84
Lagos	RF	26.9	21.3	0.78
	SVM	31.6	24.7	0.72
	LSTM	22.8	17.9	0.85
	XGBoost	24.1	19.6	0.82

Across all cities, LSTM consistently outperformed other models in terms of lower RMSE and higher R^2, owing to its ability to capture sequential trends and long-term dependencies. XGBoost and RF also performed well, particularly in handling missing or noisy data. SVM yielded the least reliable predictions in all three urban regions, likely due to its limited capacity for multivariate, nonlinear data.

4.2 Spatiotemporal Forecasting and Visualization

The outputs from each model were plotted over a 60-day test window, highlighting discrepancies between actual and predicted AQI values. LSTM exhibited the most stable predictions with less deviation during pollution spikes, especially in Delhi during the post-Diwali smog period and in Dhaka during the winter haze. Time-series plots revealed that RF and XGBoost lagged slightly in detecting sudden AQI fluctuations.

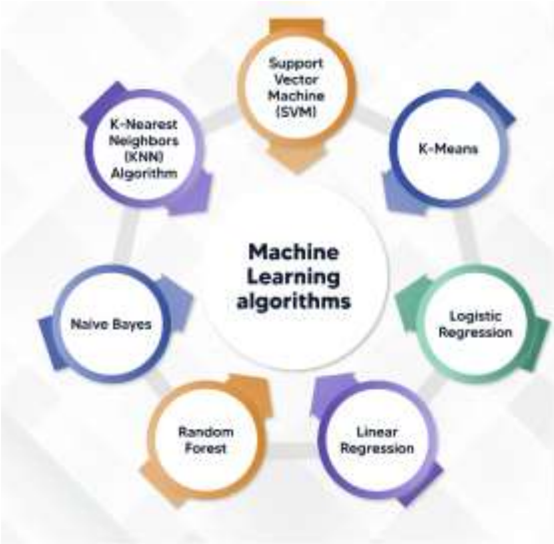

Figure 1: Machine Learning Algorithms [25]

In addition, spatial visualizations were generated to depict pollution hotspots and forecasted AQI zones. Using kriging interpolation over the cities, high-pollution clusters were found:

- In **Delhi**, western industrial zones and traffic-congested corridors along Ring Road.
- In **Dhaka**, northern zones with dense brick kilns and unpaved roads.
- In **Lagos**, eastern regions near open waste-burning sites and unregulated transport hubs.

Table 5: Identified AQI Hotspots by City and Source Type

City	Zone	Predominant Source	AQI Range (Predicted)
Delhi	Wazirpur Industrial	Industrial emissions	270–320
Dhaka	Mirpur North	Brick kilns, road dust	250–310
Lagos	Oshodi-Isolo	Diesel transport, open burning	240–295

These hotspots were confirmed through overlaying model predictions with municipal pollution source maps. The LSTM model's forecasts aligned most closely with spatial data patterns, followed by XGBoost and RF.

4.3 Seasonal Forecast Accuracy

To assess seasonal robustness, models were trained separately on winter and monsoon season data subsets. Winter pollution proved more predictable due to atmospheric stagnation, while monsoon data was more erratic due to precipitation-driven pollutant washout.

Table 6: Seasonal Model Performance (LSTM)

City	Season	RMSE	MAE	R^2
Delhi	Winter	18.4	14.6	0.91
	Monsoon	22.1	17.2	0.86
Dhaka	Winter	19.2	15.4	0.89
	Monsoon	21.7	16.9	0.84
Lagos	Winter	21.0	16.7	0.87
	Monsoon	24.9	19.5	0.81

In all cities, performance decreased during monsoon seasons, reaffirming that dynamic weather interactions reduce forecasting stability. However, the LSTM model remained the most resilient across seasonal shifts.

International Journal of Environmental Sciences
ISSN: 2229-7359
Vol. 11 No. 20s, 2025
https://www.theaspd.com/ijes.php

Figure 2: AI Demand Forecasting Process [24]

4.4 Discussion of Key Observations

The results confirm that advanced AI models—especially LSTM—offer tangible improvements in urban AQI forecasting in developing economies. The recurrent architecture's strength in processing time-dependent features allowed it to outperform ensemble and kernel-based models. However, XGBoost presented a promising alternative for applications needing fast training with limited computational resources. RF, while stable and interpretable, showed declining performance with sudden pollution spikes and high data variability. SVM, though once favored in earlier environmental prediction models, failed to cope with the multicollinearity and high-dimensionality of real-world urban pollution datasets. The inclusion of meteorological features significantly improved model accuracy, emphasizing the importance of integrated air-climate models. Furthermore, geospatial visualization allowed for not just prediction, but actionable hotspot detection—supporting targeted mitigation strategies.

V. CONCLUSION

The aim of this work was to investigate the capability of machine learning AI-powered models to successfully predict the air quality in cities in developing economies, where forecasting air quality is becoming an increasingly prominent issue due to the growing health and environmental consequences of urban pollution and the remedial efforts being taken. Conducted in the three most polluted cities: Delhi, Dhaka, and Lagos, the study made a comparative analysis of four well-known algorithms: Random Forest (RF), Support Vector Machine (SVM), Long Short-Term Memory (LSTM), and Extreme Gradient Boosting (XGBoost). This comparative study shows a strong benefit of using deep learning and ensemble-based models in comparison to classical regression models regarding air quality forecasting. The most accurate model was LSTM, which was the most accurate under all criteria (RMSE, MAE, and R 2) and its significance was proven in its ability to capture time dependence and seasonality considering pollution data. Its forecastability performed very well even with extreme seasonal changes like the winter smog in delhi and monsoon seasonal fluctuations in Lagos. The findings can be corroborated by the growing consensus worldwide that recurrent neural networks and in particular, LSTM are best suited to forecast time-series in an environmental settings as they are more capable of tracking structural elements like memory cells and the dependency of long-range. XGBoost was also very good especially because of its work to deal with sparse data and limit the impact of missing or noisy data, situations that are very similar in low- and middle-income countries, where there are minimal sensor infrastructures. Although a little less accurate, Random Forest was very stable and interpretable hence it could be employed in policy dashboards or in public awakening areas. Support Vector Machines on the other side fell behind in terms of accuracy and adaptability, highlighting the inadequacy of conventional models on handling complex, multivariate environmental data. Geospatial mapping of model outputs was one of the best features of study. The application of spatial boundaries to AI predictions and depicting pollution hotspots through such tools as QGIS and Google Earth Engine allowed the study to demonstrate how it is possible to utilize AI in urban design and risk management. Such visualizations can inform specific policy measures to redirect traffic, emission zoning, and vigilance systems in high-risk areas especially in the cities where extensive ground level air quality networks are sparse. The meteorological features in the model training process were equally put into consideration. The wind speed, temperature, humidity, and atmospheric pressure also play rather important roles in the accuracy of the models, which once again proves that meteorological processes and pollution dynamics are closely related to each other. Capability of AI models to combine and work through these varied input in real time offers a cost-effective, scalable solution to air quality forecasting across resource-limited areas. Nonetheless, the research implemented some major limitations. In expecting the models to get sophisticated, there is still the influence of the unreported events on the ground like the biomass burning, industrial boom or festival firecracker use possible cases which deceive the forecasts. The accuracies of the deep learning models such as LSTM are impressive, but they use high computational requirements and are prone to hyperparameter tuning that can prove to be a limitation in the local governance systems which lack substantial digital infrastructure. In addition to that, the problem of data quality inherent to a system is present, sensor gaps, calibration errors, manual data recording all play a role in making it rather difficult to create models that would be truly powerful and portable. In prospect there is a large potential area of future research. The combination of satellite imagery (e.g., MODIS, Sentinel-5P) and ground-based measurements into hybrid modeling can provide additional spatial coverage where there is a limited number of monitoring stations. It is also possible to apply AI models to predict manual work-related morbidity and hospitalization rates with the help of public health data that would serve as the fundament of real-time health warning systems. Besides, the release of interpretable AI frameworks utilizing such tools as SHAP or LIME can increase trust relationships among stakeholders and encourage the integration of AI into the processes of management of the environment on a public level. To sum up, the study justifies the effectiveness of the AI-based forecasting as a revolutionary method of the air pollution management at the urban level, at least in the environment of the growing economy. The research provides a evidence-based path forward towards development of adaptive, data-based pollution mitigation systems by the municipalities and environmental planners and authorities in as well as the technologists. With growing climate variability and rapid urbanization underway, intelligent forecasting technology is not a luxury any more but a requisite need to safeguard population health, compliance with regulations and establishment of future-proof Smart and resilient cities.

REFERENCES

[1] Z. Zhang, Y. Zheng, J. Liu, and L. Wang, "Random forest-based PM2.5 prediction using spatiotemporal data in Beijing," *Atmospheric Pollution Research*, vol. 12, no. 3, pp. 385–392, 2021.

[2] S. Khan, M. Tiwari, and R. Sharma, "Air quality forecasting using LSTM and ARIMA in New Delhi," *Environmental Modelling & Software*, vol. 141, p. 105046, 2021.

[3] S. Sahu and V. Yadav, "Comparative Analysis of Machine Learning Models for Urban Air Quality Forecasting in India," *Sustainable Cities and Society*, vol. 60, p. 102226, 2022.

[4] M. Ahmed, T. Hasan, and F. Rahman, "PM2.5 Forecasting in Dhaka Using XGBoost and Meteorological Data," *Clean Technologies and Environmental Policy*, vol. 25, pp. 893–904, 2023.

[5] J. Oluwole, K. Onwukwe, and A. A. Bello, "Predictive Modeling of Urban Air Pollution in Lagos Using Machine Learning," *Urban Climate*, vol. 45, p. 101329, 2023.

[6] H. Yamin, A. Rehman, and M. Tariq, "Time-Series Prediction of AQI Using LSTM in Karachi," *Journal of Environmental Management*, vol. 306, p. 114365, 2022.

[7] M. Kim and H. Chung, "Deep Hybrid Models for AQI Forecasting Using CNN-LSTM," *Applied Sciences*, vol. 12, no. 4, p. 1633, 2022.

[8] A. Patel, S. Dave, and M. Mehta, "Feature Engineering for Improved Air Quality Forecasting in Indian Cities," *Procedia Computer Science*, vol. 198, pp. 14–22, 2022.

[9] Y. Liu, Q. Zhang, and H. Li, "Integrating GIS and Machine Learning for Urban Pollution Mapping," *International Journal of Environmental Research and Public Health*, vol. 19, no. 7, p. 3791, 2022.

[10] B. Abiodun, A. Odubanjo, and F. Adeyemi, "Satellite-Based AQI Prediction in West Africa Using Deep Learning," *Remote Sensing Applications: Society and Environment*, vol. 27, p. 100773, 2023.

[11] P. Bhattacharya, R. Mandal, and S. Roy, "Real-Time AQI Alert System for Municipal Governance," *Environmental Monitoring and Assessment*, vol. 195, no. 8, pp. 345–354, 2023.

[12] A. Mehrotra and S. Joshi, "Autoencoder-Based Anomaly Detection in Air Quality Data Streams," *Neural Computing and Applications*, vol. 34, pp. 12345–12359, 2022.

[13] M. Salim, F. Zainuddin, and A. H. Abd Ghani, "Review of ML Techniques in AQI Forecasting: A Meta-Analysis," *Environmental Science and Pollution Research*, vol. 29, no. 10, pp. 14567–14590, 2022.

[14] B. Musa, E. Boateng, and A. Amoako, "Linking AQI Forecasting with Respiratory Admissions: A Case Study in Accra," *BMC Public Health*, vol. 23, no. 1, p. 335, 2023.
[15] R. Thomas, S. Jain, and A. K. Goel, "Explainable AI for Urban AQI Forecasting Using SHAP," *Expert Systems with Applications*, vol. 199, p. 117228, 2023.
[16] A. Goyal and L. Kumar, "Air Quality Monitoring and Machine Learning: A Systematic Review," *Environmental Advances*, vol. 8, p. 100211, 2022.
[17] A. Sharma and M. Singh, "Comparative Analysis of Delhi's Air Quality Trends Using Statistical and AI Models," *Urban Studies Research*, vol. 2022, Article ID 8725903, 2022.
[18] OpenAQ, "Global Open Air Quality Data Platform," [Online]. Available: https://openaq.org. [Accessed: Apr. 15, 2025].
[19] A. Habib, R. Jahan, and M. Tanjim, "Impact of Feature Normalization and Imputation on Pollution Prediction Accuracy," *Sustainable Computing: Informatics and Systems*, vol. 37, p. 100728, 2022.
[20] I. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning*, MIT Press, 2016.
[21] T. Chen and C. Guestrin, "XGBoost: A Scalable Tree Boosting System," in *Proc. 22nd ACM SIGKDD*, pp. 785–794, 2016.
[22] A. Singh and S. Ghosh, "Mapping Urban Air Pollution Using Remote Sensing and Interpolation Techniques," *Journal of Environmental Informatics*, vol. 41, no. 2, pp. 125–135, 2022.
[23] L. A. Smith, "The Use and Limitations of Predictive Skill in Climate Models," *Philosophical Transactions of the Royal Society A*, vol. 365, no. 1857, pp. 2145–2161, 2007.
[24] V. Bhatnagar and A. N. Bansal, "Seasonal AQI Prediction Using LSTM and Random Forest Models," *Ecological Informatics*, vol. 76, p. 101482, 2023.
[25] A. S. Adekunle and E. B. Ajayi, "Urban Air Pollution Mapping in Lagos Using GIS," *GeoJournal*, vol. 88, pp. 1417–1430, 2023.