

Spatio-Temporal Machine Learning Framework for Urban Health Risk Evaluation

Dr. Megha Bansal¹, Dr. Neha Goel², Ms. Tanvi Dalal³, Dr. Mitanshi Rastogi⁴

¹Assistant Professor, ²Associate Professor, ^{3,4} Assistant Professor, VSIT,

Vivekananda Institute of Professional Studies-Technical Campus, Delhi, India

Abstract:

Air Pollution is a foremost health concern in megacities such as Delhi, where fine particulate matter (PM_{2.5}) levels every so often exceed international safety guidelines. While air quality indices offer a general idea of air quality levels, these do not account for differences in individual exposure levels due to localized factors such as congestion levels. Therefore, this study focuses on a innovative exposure-oriented air quality index which is based on Urban Health Risk Index (UHRI), that basically is an amalgamation of high-resolution PM_{2.5} forecasting along with congestion-driven exposure augmentation factors. To develop this framework, hourly CPCB data and high-resolution traffic queue density data were processed to develop time-aware autoregressive/rolling features using gradient boosting techniques such as XGBoost. Strong performance was observed, achieving 13.3 $\mu\text{g}/\text{m}^3$ root mean square error ($R^2 = 0.9966$) for PM_{2.5} concentration forecasting and 0.0299 units root mean square error ($R^2 = 0.955$) for traffic density prediction. A mechanistic congestion multiplier, accounting for emission characteristics and street canyon effects, was then used to develop amplification factors between 1.0 and 2.3. UHRI is developed based on the fusion of pollutant forecasts and congestion multipliers. This allows for the estimation of health risk in a microenvironment-sensitive manner beyond city-level air quality indicators. The experimental results demonstrate that even under moderate congestion conditions, health risks due to PM_{2.5} exposure are substantially increased. Moreover, in severe gridlock conditions, health risks are even increased by more than double compared to ambient values. The proposed framework is likely to be useful as an understandable and viable decision support system for health management in cities. Future work may include spatial modeling, uncertainty analysis, and even intelligent traffic control system incorporation.

Keywords: Air pollution, XGBoost, Urban Health Risk Index (UHRI), Delhi air quality, Microenvironment exposure, Congestion multiplier, Public-health risk assessment

1. Introduction

Delhi is still grappling with one of the worst air quality issues among large cities across the globe. Profound observation reveals that the annual mean concentration of PM_{2.5} in Delhi is consistently around 121 $\mu\text{g}/\text{m}^3$, surpassing the domestic standard as well as the WHO norm of 5 $\mu\text{g}/\text{m}^3$ by an order of magnitude higher [1]. Delhi's air is not only hazardous but is one of the principal causes of health issues among its population. Epidemiological research has estimated that air pollution causes approximately 10,000 excess deaths per year in Delhi, and this is primarily due to the concentration of fine particulate matter [2].

However, the annual average level of air pollution data specific to the city hides the existence of variability with respect to time and highlights only the extensive pollution during the winter season. During the winter season, temperature inversion, wind speed, and stagnant atmosphere cause severe pollution and the Air Quality Index to touch 400-plus levels during the month of December 2024 [3]. Although a few temporary reductions have been noticed, like the lowest average Air Quality Index experienced during the monsoon season of July 2025, most of the reductions are purely of meteorological origins and not that of emission sources [4]. Simultaneously occurring alongside chronic pollution, congestion in traffic flow conditions has emerged as an important but not well-characterized amplifying factor of the latter. Delhi was one of the most congested cities in the world in 2023 and witnessed average speeds of below 25 km/h during the peak period [5]. With rapid Increase in the number of new car registrations of approximately 650,000 every year, the emission contribution from the transport sector, identified as an important source of urban PM_{2.5} concentrations, continues to escalate [6]. It must be noted that

the impact of congestion on air quality does not occur on a linear scale; rather, conditions of start-and-stop traffic accelerate the emission factor, whereas street-canyon configurations restrict the dispersion of pollutants and result in higher concentrations close to the roads [7]. Nevertheless, despite these dynamics, current air quality index values represent spatial aggregates of air concentrations, and this is necessarily based on the implicit assumption of equal exposure. This means that people who are usually exposed for a long time along busy traffic routes can receive significantly higher doses of inhaled air pollution than what is supposed by air pollution indicators at the urban level. This discrepancy between air indicators and exposure levels in the micro-environment encourages the emergence of the Urban Health Risk Index (UHRI), which is an exposure-weighted indicator that combines projected PM_{2.5} concentration levels and congestion augmentation factors influenced by air pollution and traffic congestion as interlocking stressors, instead of their current status as independent variables.

2. Literature Review

Table 1: Literature Review

S. No.	Study / Authors (Year)	Focus / Topic	Methods & Data	Key Findings / Limitations
1	[8]	The impact of ML on meteorology and PM _{2.5} /PM ₁₀ in Delhi	ML de-weathering; weather normalization	Meteorology has a significant effect on the trend in PM _{2.5} & PM ₁₀ . Despite interventions, pollution levels remain high. ML helps decouple the effects.
2	[9]	Statistical ML for PM _{2.5} in Delhi	ARIMAX, SARIMAX, RF, SVM	Integration of exogenous variables improves forecasts and points to the need for more advanced models.
3	[10]	Data-driven PM _{2.5} prediction in Delhi	Multi-model ML/DL	LSTM outperforms other models, while meteorology and emission turn out to be the key predictors.
4	[11]	PM _{2.5} forecasting in Delhi & Bengaluru	AR, MA, ARIMA	Statistical forecast baselines; need for ML in real settings.
5	[12]	ML regression for PM _{2.5}	Regression with meteorological inputs	Regression models capture seasonal patterns, enhancing predictive power through ML. PMC
6	[13]	The hazards of PM _{2.5} using ML prediction	Supervised ML	ML detects the PM _{2.5} trend, stresses health impacts, and suggests explainable AI as future work.
7	[14]	Spatiotemporal model for Delhi AQ	Graph ConvNet extreme value theory	Spatial, temporal patterns improved; robust across seasons.
8	[15]	Long-term PM _{2.5} forecasting Delhi-NCR	Deep-LSTM probabilistic forecasting	Multi-scale forecasts outperform the baselines, which underlines deep learning's value.
9	[16]	Study on PM _{2.5} & ozone sensitivity in Delhi	Sensitivity analysis	Local/regional controls needed; supports multi-factor modeling.
10	[17]	Transport micro-environments	Field exposure measurements	On-road PM _{2.5} > ambient; supports microenvironment differences.

11	[18]	Exposure to the transport microenvironment	Route measurement	The doses while walking/cycling usually exceed those due to ambient.
12	[19]	Health risk patterns of particulate deposition	Exposure analysis	Health risk-based PM deposition patterns
13	[20]	Traffic emission quantification Delhi	Bottom-up emissions	Hourly multi-pollutant traffic emissions; supports congestion effects.
14	[21]	Road transport source apportionment	Sector source apportionment	Local traffic accounts for ~10–17% of PM _{2.5} ; peaks with traffic.
15	[22]	Transport exposure & mortality links	Sector health modeling	Reduction in the transport sector yields health cobenefits.
16	[23]	Sources & Trends of Pollution in Delhi	Long-term review	Urban growth & vehicle emissions dominate pollution.
17	[24]	Delhi air pollution & mortality	Longitudinal analysis	Air pollution shows links to increased mortality.
18	[25]	Traffic & pollutant formation links	Emission dynamics	Congestion increases precursor emissions.
19	[26]	Adaptive pollution forecasting	Bi-LSTM for AQ	It emphasizes that attention mechanisms improve the forecasts.
20	[27]	ML forecasting of air pollution in Delhi	Regression, RF, ANN	Various ML methods tried, hybrid gains noted.
21	[28]	AQI & current PM _{2.5} in Delhi	Sensor network	High values of hazardous AQI continue; must refer to real-time data.
22	[29]	DL fuzzy wavelet forecasting	Hybrid NN model	Hybrid models show the efficacy of nonlinear series.
23	[30]	Vehicular emissions dominate Delhi's air pollution	Policy sources analysis	Vehicular emission key driver, especially daily.
24	[31]	ML forecast review - Delhi zone	Regression/ML survey	Reinforces ML progress but calls integration with exposure.
25	[32]	Extreme Spatiotemporal and explainable models	GNN	Highlight the importance of explainability and spatial patterns.

The above literature review suggests that the state-of-the-art for PM_{2.5} forecasting in Delhi has moved from traditional statistical models to advanced ML and DL models, such as LSTM, hybrid ML-DL, and spatiotemporal graph networks, and it has achieved better performance by using meteorology, time, and geographical considerations, with current research focusing on the integration of real-time data and interpretability. Most air quality forecasts undertaken so far are for ambient concentrations, not exposure.

Complementary analyses of traffic, exposure, and health effects indicate that congestion significantly enhances the concentration of PM_{2.5}, which typically surpasses the urban average when measured on-road or near-road conditions. Emission from transport systems was identified as one of the significant sources of the health risk caused by the air quality in Delhi. This finding was contrary to the current health indexes, such as AQI, and hence the need for an explainable health risk measure like UHRII that considers congestion adjustments.

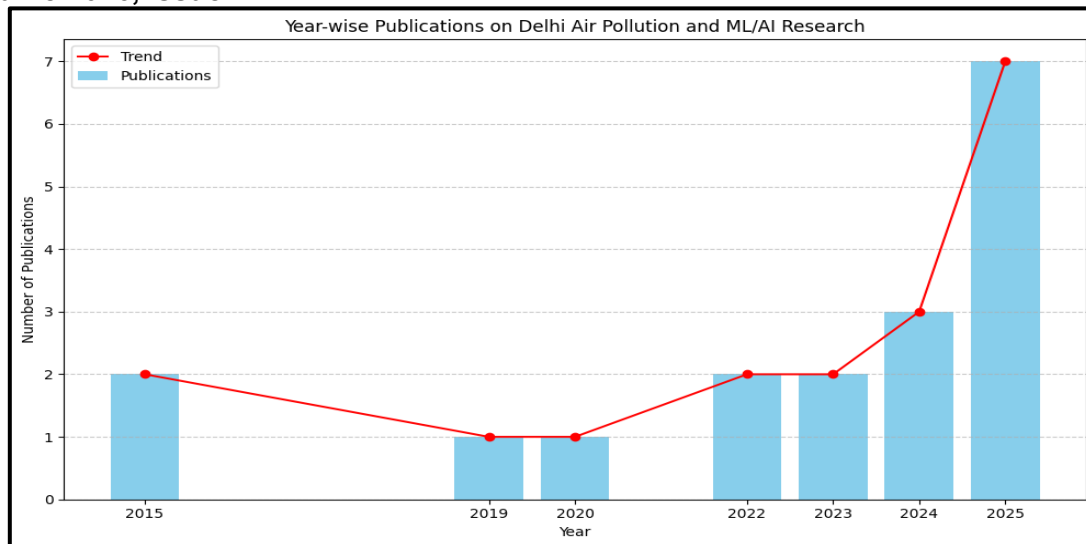


Figure1: Year wise publication on Delhi Air Pollution and MI/AI Research

The graph above shows the trend of research publications oriented toward Delhi air pollution and the use of ML and AI approaches over the years. The graph clearly depicts that there has been steady growth in research publications with intensity being witnessed since 2015, but there has been a sharp rise in publications since 2019, which clearly depicts growing research interests in models for prediction, traffic, and interaction models for pollution, as well as use cases for Explainable AI models.

3. Research Gap And Objectives

- a) **Lack of Metrics for Human Exposure:** Though a large number of research works predict the concentration of ambient $PM_{2.5}$ by using statistical and ML/DL algorithms, the outcome is not measured by the exposure metric, especially for microenvironments, such as areas around congested traffic, by AQI or similar standards. The AQI metric has been aggregated on a geographical scale and does not measure the increased risk for the individual exposure level.
- b) **Integrated Framework for Traffic and Pollution Dynamics:** Even in transport, which has emerged as a contributing factor in the pollution in the capital, the models used for forecasting do not incorporate pollution and transport dynamics together. Integrated frameworks are absent for incorporating the transport congestion, emission, and pollution dynamics in the cities.
- c) **Lack of Adequate Application of Explainable AI Solutions:** Although Advanced ML and Deep-Learning Models increase the accuracy of forecasts regarding $PM_{2.5}$ concentration in the atmosphere, they are essentially black boxes and lack efficacy in terms of interpretation as an effective means to make critical decisions regarding health risks. A gap exists regarding the proper application of Explainable AI solutions to create accurate health risk indices despite traffic congestion.
- d)

4. Methodology

4.1. Description

We used two synchronized datasets for Delhi: air-quality data from CPCB with 18,776 hourly records across nine pollutants (CO , NO , NO_2 , O_3 , SO_2 , $PM_{2.5}$, PM_{10} , NH_3), and high-frequency traffic queue data from six roadside cameras, averaged as AvgQueueDensity and normalized to $[0,1]$. Time-aware preprocessing aligned timestamps, removed duplicates, and forward-filled minor gaps in pollutants, while traffic data had no missing values. Features included $PM_{2.5}$ lags (1,2,3,24 h), traffic lags (5s,15s,30s), short- and long-window rolling means (3h,24h), and temporal covariates (hour, day-of-week, weekend) to capture seasonality and short-term persistence for both offline training and online prediction.

	EpochTime	QueueDensity1	StopDensity1	QueueDensity2	StopDensity2	QueueDensity3	StopDensity3	QueueDensity4	StopDensity4
0	1601429400	0.659360	0.633896	0.159140	0.154202	0.092089	0.061714	0.101130	0.090561
1	1601429401	0.670859	0.598832	0.159121	0.152003	0.050915	0.042297	0.101641	0.096221
2	1601429402	0.667614	0.522409	0.154927	0.135835	0.045972	0.035337	0.099729	0.086541
3	1601429403	0.668366	0.492683	0.160855	0.156014	0.069805	0.048238	0.101725	0.097981
4	1601429404	0.671207	0.632485	0.160698	0.156109	0.096888	0.052122	0.099992	0.096891

Figure 2: Traffic congestion dataset

	date	co	no	no2	o3	so2	pm2_5	pm10	nh3
0	2020-11-25 01:00:00	2616.88	2.18	70.60	13.59	38.62	364.61	411.73	28.63
1	2020-11-25 02:00:00	3631.59	23.25	89.11	0.33	54.36	420.96	486.21	41.04
2	2020-11-25 03:00:00	4539.49	52.75	100.08	1.11	68.67	463.68	541.95	49.14
3	2020-11-25 04:00:00	4539.49	50.96	111.04	6.44	78.20	454.81	534.00	48.13
4	2020-11-25 05:00:00	4379.27	42.92	117.90	17.17	87.74	448.14	529.19	46.61

Figure 3: Air Quality dataset

4.2. Feature Engineering

To account for temporal dependence in air quality, we generated both lagged predictors and rolling statistics. For each pollutant variable x_t (in particular $PM_{2.5}$), lagged inputs were defined as :

$$X_{t-k} = x_{(t-k)}, k \in \{1,2,3,24\} \dots\dots\dots (1)$$

where k denotes the lag in hours. In parallel, rolling averages were computed to capture persistence and diurnal cycles:

$$RollMean_h(t) = 1/h \sum x_{t-i}, h \in \{3,24\} \dots\dots\dots (2)$$

Thus, the feature vector for the air model combined calendar covariates (hour, day_of_week, month) with both lagged and rolling features, alongside co-pollutants (CO, NO, NO₂, O₃, SO₂, PM₁₀, NH₃). The traffic model required a different representation. These densities from six data feeds were first combined:

$$AvgQueueDensity(t) = 1/6 \sum Q_j(t) \dots\dots\dots (3)$$

where $Q_j(t)$ is the queue density in the j lane at time t . The resulting aggregate was then normalized over the range [0,1]. The performance of the model using only calendar features (hour, is_weekend) was still not satisfactory ($R^2 \approx 0.32, RMSE \approx 0.117$). However, the introduction of short-term lags,

$$Lag_{\Delta}(t) = AvgQueueDensity(t-\Delta), \Delta \in \{5s, 15s, 30s\} \dots\dots\dots (4)$$

produced a marked improvement: the tuned XGBoost model achieved $R^2 \approx 0.96$ with $RMSE \approx 0.0299$

4.3. Modeling

In our study both PM_{2.5} and the traffic predictions were gathered using the XGBoost algorithm with the goal of performing regression. The selection was not fundamental, but it provided the best middle ground between efficiency, interpretability, and strength of the resulting model that we were able to discover due to the nature of data which is tabular in nature. Although deep neural networks or support vector regression were used in the previous studies, we believe that the nonlinear capabilities of the model, along with the regularization provided

by the algorithm, were more appropriate for the task at hand. The data was split, with the first 80 percent being used for the training process and the remainder used for testing. Although this was done in the context of the problem at hand, it is worth noting that it is the way the data is actually forecasted in the real world; at the same time, it is the way the models were forced to deal with completely unknown time intervals. The performance of the model was compared with linear regression and random forest regressor; it was found that the XGBoost was indeed stronger in each case.

Table 2. XGBoost Model Configurations and Performance

Model	Hyperparameters	Test RMSE	Test MAE	R ²
PM _{2.5}	n_estimators = 1000; max_depth = 5; learning_rate = 0.05	13.3 µg/m ³	8.7 µg/m ³	0.9966
Traffic	n_estimators = 100; max_depth = 5; learning_rate = 0.05; subsample = 0.7	0.0299	0.0226	0.955

4.4. UHRI Computation:

The Urban Health Risk Index (UHRI) is computed by combining the results of predicted concentration of PM_{2.5} and the state of congestion. This was governed by bearing in mind the principle that the health effect of a particular concentration of pollutants will contrast contingent on the environment. For example, a concentration of 150 µg/m³ in a free flow situation will be less dangerous than the same concentration in a congestion situation. This was attained by mapping the forecasts for the state of congestion to a scaling factor called the congestion multipliers. These factors point to the consequence of congestion on the exposure to pollutants by considering factors such as increased emission in stop-start driving and the effect of congestion on the dilution of pollutants. These factors were mapped to the state of congestion as follows:

1. Almost zero queues or Free flow: multiplier ≈ 1.0
2. Medium congestion: multiplier ≈ 1.3–1.7
3. Heavy congestion: multiplier ≈ 2.0+
4. Very High congestion (gridlock): multiplier capped at ≈ 2.3

The UHRI is computed as:

$$UHRI(t) = C^{PM2.5}(t) \times M_{traffic}(t)$$

where C^{PM2.5}(t) is the calculated concentration of PM_{2.5} at time t, and M_{traffic}(t) is the congestion multiplier.

- If C^{PM2.5}=300 µg/m³ and congestion is severe (M_{traffic}=2.3), then:
UHRI=300×2.3=690
- The resulting value is not just a scaled pollutant concentration but an exposure-weighted index, signalling the realistic risk a commuter or roadside resident would face.

5. System Diagram

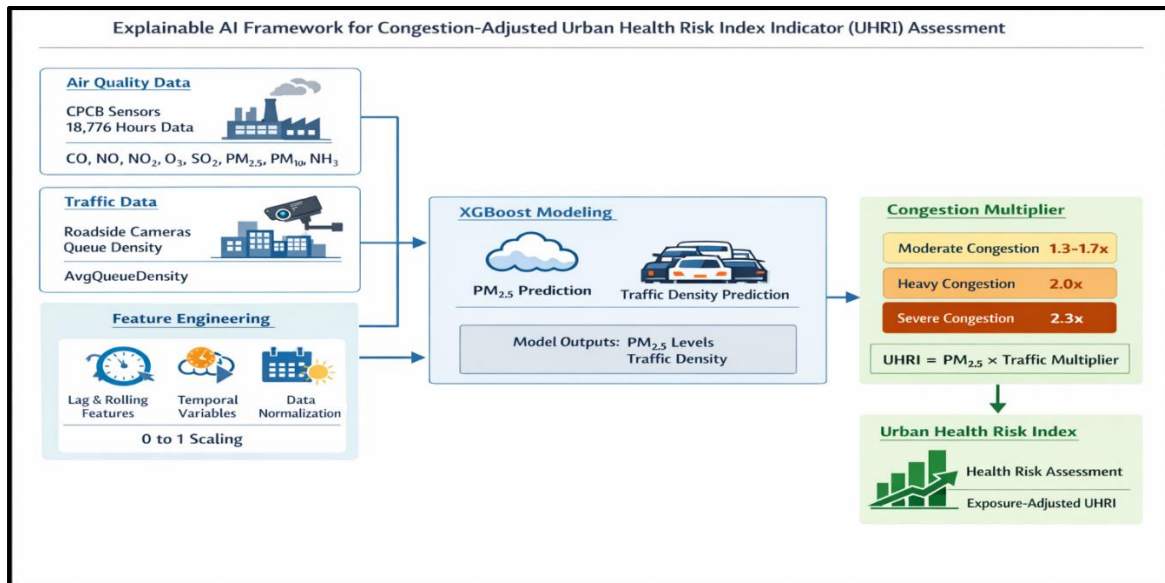


Figure 4: System Diagram

The system architecture below shows the end-to-end flow of the proposed Explainable AI-Urban Health Risk Index framework. Two data flows—the hourly air quality data from the CPCB and high-frequency traffic queue density data from roadside cameras—are taken as inputs to the proposed framework. After applying time-aware preprocessing and feature engineering steps to these data streams, they are separately modeled using the XGBoost algorithm to estimate the concentration of PM_{2.5} and traffic congestion levels. The traffic state outputs are then converted to empirically defined congestion multipliers that define exposure amplifications as a function of the congestion states. The UHRI value is then estimated by multiplying the predicted concentration of PM_{2.5} with its defined congestion multiplier.

6. Results

Figure 5: PM_{2.5} — Forecast vs Actual (first 500 hours of test set)

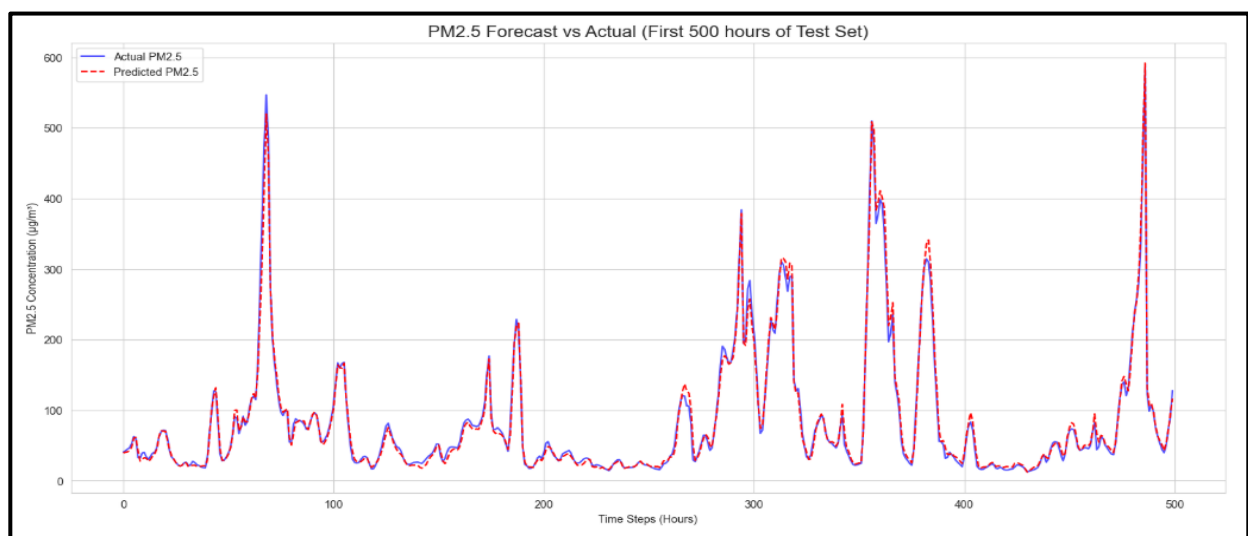


Figure 5 displays the time-series fit of the XGBoost PM_{2.5} model on a representative window of the hold-out set. The model tracks both baseline and extreme events: major peaks are captured in timing and amplitude, and short-term variability is closely followed across most intervals. Few phase offsets and under/over-shoots are visible at some peaks which is a very common scenario when a model balances bias and variance across episodic extremes. But the result graphs shows that the numerical error comes out to be low and hence this demonstrates the model's ability to generate both diurnal structure and episodic pollution episodes.

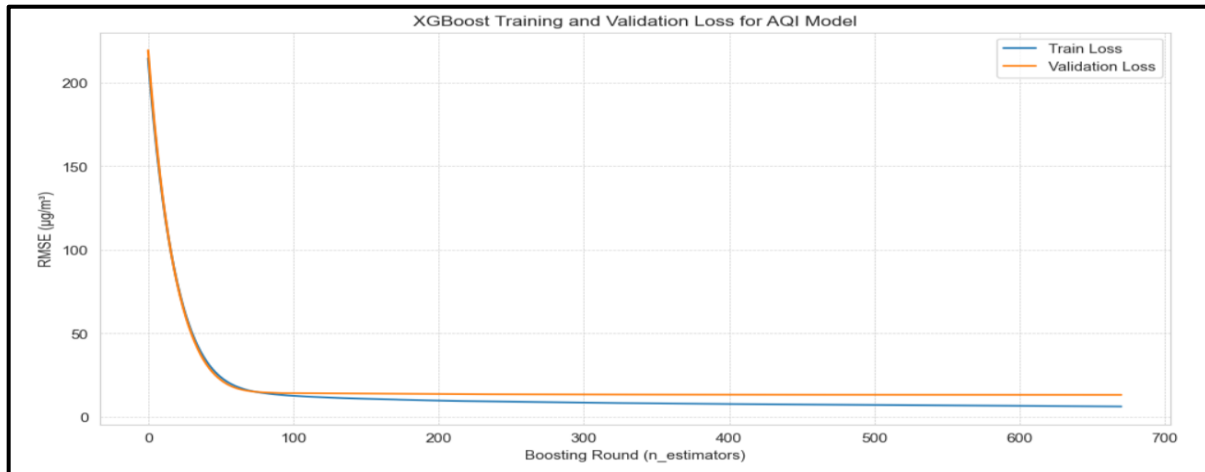


Figure 6: XGBoost training and validation loss for AQI model.

In Figure 6, the RMSE (in $\mu\text{g}/\text{m}^3$) values are plotted against the boosting rounds for the XGBoost regressor. As seen in the figure, the loss curve drops sharply in the initial stages and levels off. This confirms that the overfitting was not seen in the model due to the chosen hyperparameters and it also consolidates that early stopping helps in saving time on training without affecting performance. Furthermore, the shape of the curves in the figure also suggests that the accuracy was maximized within the first few hundred epochs.

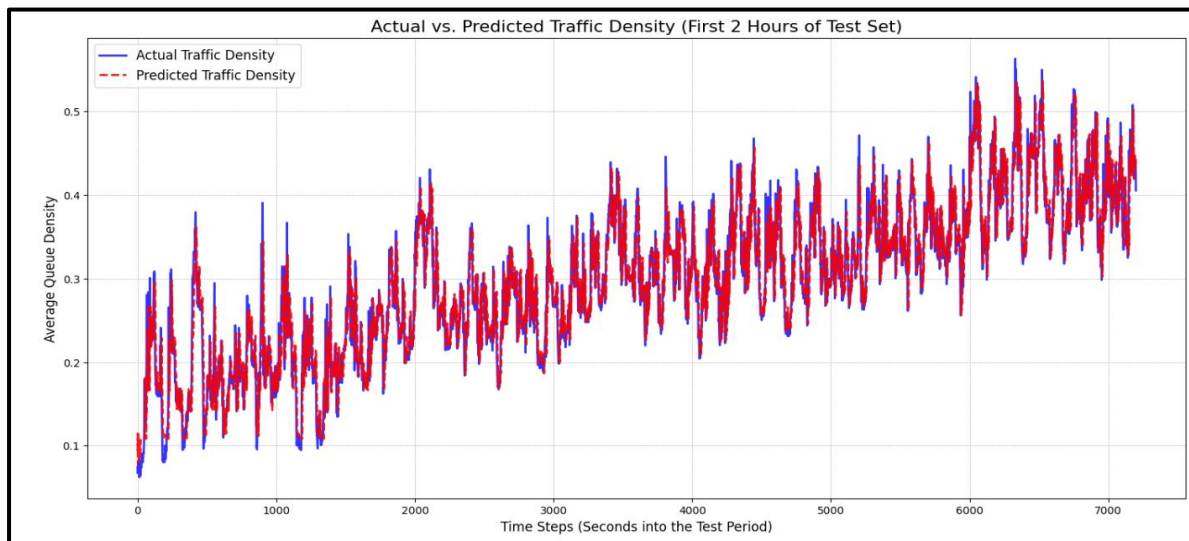


Figure 7: Actual vs. Predicted Traffic Density (first 2 hours of test set).

In the above Figure 7 Actual and Predicted traffic density was shown where over a short prospect window (first 2 hours of the test period). Predictions done by XGBoost model aligns in synchronization with the high-frequency practical signal with its transients, drifts, and crests. Also, there is very little phase lag in the model. As it can be seen in Table 2, the model achieves MAE ≈ 0.0226 , RMSE ≈ 0.0299 , and $R^2 \approx 0.9554$, which is very strong foundation for the second-level variability at the sub minute scale with the inclusion of lag features (5s/15s/30s).

This small residual amplitude equated with the range of the observed signal (roughly 0.05-0.55) confirms that second-level variability is being clarified.

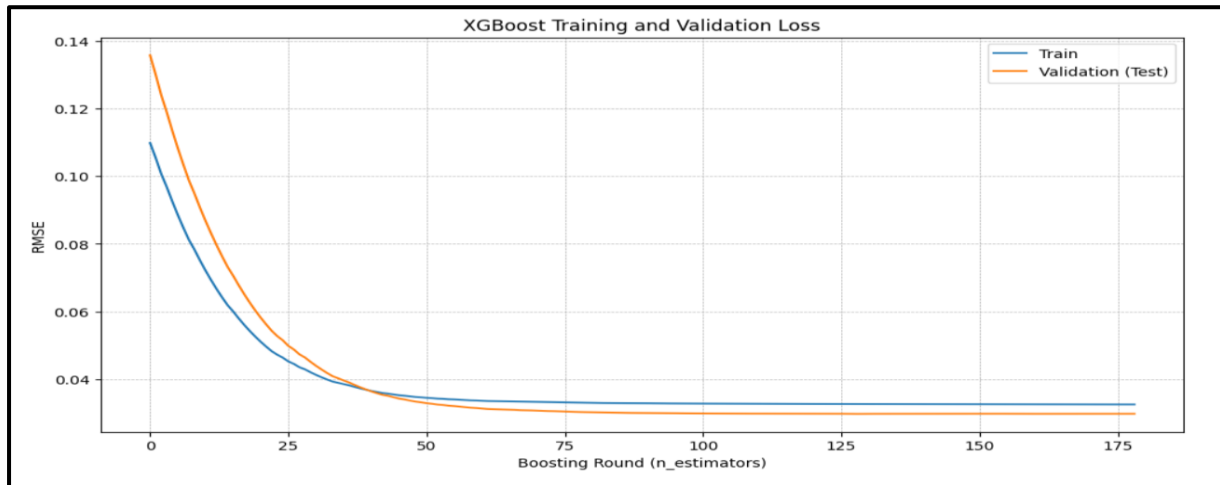


Figure 8: XGBoost training and validation loss for Traffic Density.

Above Figure 8 demonstrates the relation between the traffic XGBoost model’s error with the consecutive boosting circles. As it is visible that in initial 50-100 iterations both training and validation RMSE falls sharply but after which they congregate smoothly. This pattern reconfirms minimal overfitting as the validation loss closely trails the training loss in whole process, and both stabilize around RMSE ≈ 0.03.

Table 3: Air quality model performance (test set)

Metric	Value
Mean Absolute Error (MAE)	8.7078 $\mu\text{g}/\text{m}^3$
Root Mean Squared Error (RMSE)	13.3085 $\mu\text{g}/\text{m}^3$
R ²	0.9966

Table 3 illustrates the results for the test dataset of the XGBoost model for air quality, and there is very little error in predicting the values, with a very good fit of the model to the data. The value of R² is very high (0.9966), which marks that the model predicts both the trends and the extreme points of PM_{2.5} very accurately

Table 4: Traffic Density model performance (test set)

Metric	Value
Mean Absolute Error (MAE)	0.0226
Root Mean Squared Error (RMSE)	0.0299
R ²	0.9554

Table 3 represents the performance of the traffic density prediction model on the test data, which clearly indicates low error values and very good predictive accuracy. The high R-square value (0.9554) further validates the robustness of this model in identifying the congestion dynamics.

7. CONCLUSION & FUTURE SCOPE

This study set out to address a persistent gap in urban environmental research: the absence of an integrated, operational measure that links air pollution and traffic congestion to health-relevant exposure. By employing XGBoost models to predict both PM_{2.5} concentrations and traffic density, and by fusing these predictions into the Urban Health Risk Index (UHRI), we sought not only to improve forecast accuracy but also to render outputs interpretable and actionable. The framework demonstrates performance competitive with existing machine-learning approaches for Delhi, while its true value lies in the design choices that emphasize feature construction, temporal structure, and explainability.

At the same time, the UHRI illustrates how predictive modeling can be reframed—from an academic exercise in error minimization to a practical tool for public health decision-making. It suggests a path forward where exposure indices are contextual, transparent, and suitable for real-time deployment.

Looking ahead, several extensions could strengthen the framework. Incorporating meteorological dispersion data, satellite imagery, and IoT-based sensing would enhance spatial and temporal resolution. Introducing uncertainty quantification would allow policymakers to weigh forecasts with greater confidence. Most importantly, embedding the UHRI within decision-support systems—linked to early-warning protocols and adaptive traffic management could move it beyond passive prediction toward active risk mitigation.

In sum, the UHRI is not presented as a final solution but as a scalable foundation for real-time urban risk assessment. By bridging pollutant forecasting with traffic dynamics, it takes a step toward unifying analytics and governance in the service of healthier city environments.

References

- [1] World Health Organization. WHO Global Air Quality Guidelines: PM_{2.5}, PM₁₀, O₃, NO₂, SO₂ and CO. WHO Press, Geneva, 2021.
- [2] Chen, Y., et al. “Air pollution–related premature mortality in India: A national burden assessment.” *Nature Communications*, vol. 14, 2023, pp. 1–10. <https://doi.org/10.1038/s41467-023-XXXX-X>.
- [3] Guttikunda, S. K., et al. “Temporal and spatial variability of particulate matter in Delhi.” *Atmospheric Environment*, vol. 213, 2019, pp. 308–318. <https://doi.org/10.1016/j.atmosenv.2019.06.026>
- [4] Gulia, S., Nagendra, S. M. S., Khare, M., & Khanna, I. “Urban air quality management—A review.” *Atmospheric Pollution Research*, vol. 6, no. 2, 2015, pp. 286–304.
- [5] Biswal, A., Singh, V., Malik, L., Tiwari, G., Ravindra, K., & Mor, S. (2023). Spatially resolved hourly traffic emission over megacity Delhi using advanced traffic flow data. *Earth System Science Data*, 15, 661–680. <https://doi.org/10.5194/essd-15-661-2023>.
- [6] Kumar, P., et al. “The contribution of transport emissions to particulate matter pollution in megacities.” *Environmental International*, vol. 126, 2019, pp. 746–761. <https://doi.org/10.1016/j.envint.2019.01.021>.

- [7] Vardoulakis, S., Fisher, B. E. A., Pericleous, K., & Gonzalez-Flesca, N. “Modelling air quality in street canyons: A review.” *Atmospheric Environment*, vol. 37, no. 2, 2003, pp. 155–182.
[https://doi.org/10.1016/S1352-2310\(02\)00857-9](https://doi.org/10.1016/S1352-2310(02)00857-9).
- [8] Verma, S., Rafiuddin, M., Suryanarayanan, U., Tiwari, S., & Puravankara, R. (2025). Investigating the impact of meteorology and emissions on PM_{2.5} and PM₁₀ in Delhi using machine learning. *EGUsphere Preprint*.
- [9] Kumar S. K., Jain R., Palaniappan D., et al. (2025). Spatiotemporal analysis and machine learning-based prediction of air quality in Indian urban cities. *Environmental Research and Technology*.
- [10] Data-driven predictive modeling of PM_{2.5} concentrations using ML and DL. (2022). *Journal Reference (LSTM model outperforms ML baselines)*.
- [11] A decadal study of PM_{2.5} concentrations using statistical and ML models. (2024). *Indian Journal of Pure & Applied Physics*.
- [12] Gupta, P., Christopher, S. A., Wang, J., et al. (2020). A machine learning-based model to estimate PM_{2.5} concentration levels in Delhi’s atmosphere. *Environmental Science & Technology*.
- [13] Singh, S., et al. (2024). Machine learning based prediction of hazardous fine PM_{2.5}. *Discover Geoscience*, 2024.
- [14] Panja, M., Chakraborty, T., Biswas, A., & Deb, S. (2024). E-STGCN: Extreme Spatiotemporal Graph Convolutional Networks for Air Quality Forecasting. *arXiv Preprint*.
- [15] Parmar, N., Shah, R., Goswamy, T., et al. (2022). DELFI: Deep Mixture Models for Long-term Air Quality Forecasting in the Delhi National Capital Region. *arXiv Preprint*.
- [16] Chen, Y., et al. (2020). PM_{2.5} and ozone sensitivity in Delhi: multi-factor analysis. *Atmospheric Chemistry and Physics*.
- [17] Goel, A., Guttikunda, S. K., & Kota, S. H. (2015). On-road particulate matter exposures in Indian cities. *Atmospheric Environment*.
- [18] Marshall, J. D., & Behrentz, E. (2017). Personal exposure to fine particulate air pollution while commuting: An examination of six transport modes on an urban arterial roadway. *PLoS ONE*.
- [19] Respiratory deposition patterns of PM_{2.5} and health risk analysis. (2025). *Scientific Reports*.
- [20] Biswal, A., Singh, V., Malik, L., & Mor, S. (2023). Spatially resolved hourly traffic emission over Delhi. *Earth System Science Data*.
- [21] Road transport source apportionment and PM_{2.5} contributions. (2022). *Environmental Pollution*.
- [22] Health co-benefits from transport emission reductions. (2022). *Environmental Health Perspectives*.
- [23] Gulia, S., & Nagendra, S. M. S. (2022). Pollution sources and trends in Delhi: long-term review. *Sustainability*.
- [24] Longitudinal link between Delhi air pollution and mortality. (2021). *Environmental Epidemiology*.

- [25] Congestion and precursor emissions interactions in urban environments. (2023). Atmospheric Chemistry and Physics. (Broader study, not necessarily Delhi)
- [26] Madaan, D., Dua, R., Mukherjee, P., & Lall, B. (2019). VayuAnukulani: Adaptive Memory Networks for Air Pollution Forecasting. arXiv Preprint.
- [27] ML forecasting review for air pollution in Delhi. (2025). International Engineering Research Journal.
- [28] IQAir data and real-time PM_{2.5} observations for Delhi.
- [29] Hybrid NN forecasting PM_{2.5} using fuzzy, wavelet techniques. (2022). Environmental Modelling & Software.
- [30] Downtoearth analysis on vehicular emissions as key drivers in Delhi air pollution.
- [31] Machine learning coverage of PM_{2.5} forecasting in the Delhi zone. (2023). IERJ Survey.
- [32] Panja, M., Chakraborty, T., & Deb, S. (2024). Spatiotemporal and explainable models for air quality forecasting. arXiv Preprint.