Comparative Analysis of ARIMA and LSTM Architectures for Urban Air Quality Prediction: A 2024–2025 Case Study of Delhi- NCR

Ashish Yadav[1], Ankit Mishra[2], Harsh Chaudhary[3], Anupama Pandey[4], Ms. Priya Tyagi[5]

[1] Department of Computer Science and Engineering (AI&ML), Nitra Technical Campus, Raj Nagar, Ghaziabad, UP, India
Email: ashishyadav639158@gmail.com

[2] Department of Computer Science and Engineering (AI&ML), Nitra Technical Campus, Raj Nagar, Ghaziabad, UP, India
Email: 2004akmishra@gmail.com

[3] Department of Computer Science and Engineering (AI&ML), Nitra Technical Campus, Raj Nagar, Ghaziabad, UP, India
Email: harshch131@gmail.com

[4] Department of Computer Science and Engineering (AI&ML), Nitra Technical Campus, Raj Nagar, Ghaziabad, UP, India
Email: pandeyanupama255@gmail.com

[5] Department of Computer Science and Engineering, Nitra Technical Campus, Raj Nagar, Ghaziabad, UP, India

How to Cite this Article:
Tyagi, P., Pandey, A., Chaudhary, H., Mishra, A. & Yadav, A. (2026). Comparative Analysis of ARIMA and LSTM Architectures for Urban Air Quality Prediction: A 2024–2025 Case Study of Delhi- NCR. International Journal of Creative and Open Research in Engineering and Management, <i>02</i>(04).
https://doi.org/10.55041/ijcope.v2i4.969

https://doi.org/10.55041/ijcope.v2i4.969

Abstract

Air pollution is one of the most critical environmental challenges facing modern urban centers, with the Air Quality Index (AQI) serving as a vital metric for public health management and municipal governance. This research provides an exhaustive investigation into the two primary paradigms of time-series forecasting for environmental data: the statistical Autoregressive Integrated Moving Average (ARIMA) and the deep learning-based Long Short-Term Memory (LSTM) network. While ARIMA offers a robust, interpretable framework for linear patterns through the Box-Jenkins methodology, LSTM architectures excel at capturing the complex, non- linear dependencies characteristic of atmospheric pollutants over extended sequences. Using a high-resolution dataset from the Central Pollution Control Board (CPCB) and meteorological ERA5 reanalysis for Delhi-NCR (2019–2025), we evaluate these models across diverse seasonal cycles. Results indicate that while ARIMA is computationally efficient for short-term, stable trends, the LSTM model achieves a significantly higher R^2 score (0.96) during extreme pollution events, such as the "Severe Plus" episode of November 2024 and the sustained hazardous spikes of October 2025. This paper aims to establish a benchmark for selecting the appropriate predictive architecture to support early-warning systems like the Graded Response Action Plan (GRAP).

Keywords: Air Quality Index (AQI), Time Series Forecasting, Autoregressive Integrated Moving Average (ARIMA), Long Short-Term Memory (LSTM), Deep Learning, Machine Learning, Environmental Prediction, PM2.5, Urban Air Pollution, Delhi-NCR, ERA5 Reanalysis, Central Pollution Control Board (CPCB), Graded Response Action Plan (GRAP), Non-linear Modeling, Spatiotemporal Analysis.

1. Introduction

The escalating volatility of urban atmospheres, driven by rapid industrialization, increasing vehicular emissions, uncontrolled construction dust, and seasonal biomass burning, has made real-time environmental prediction an urgent necessity. In cities like New Delhi, where the AQI frequently breaches the "Hazardous" threshold of 450, predictive models form the backbone of emergency environmental governance. According to the World Health Organization (WHO), prolonged exposure to fine particulate matter (PM_{2.5}) is linked to severe cardiovascular and respiratory diseases, contributing to millions of premature deaths globally each year. In Delhi alone, annual average exposure levels consistently exceed both the Indian National Ambient Air Quality Standards (NAAQS) and the WHO's recommended safety limits.

Traditional numerical methods, while theoretically sound, often struggle with the computational overhead and localized precision required for hyperlocal, hourly forecasting. Furthermore, standard regression models fail to account for the autocorrelation inherent in time-series environmental data, where current pollution levels are highly dependent on previous states. To bridge this gap, data-driven intelligence has emerged as a superior alternative. This paper rigorously compares a classical statistical approach (ARIMA) with a modern recurrent neural network architecture (LSTM). By predicting concentrations of major pollutants, these models allow for proactive health advisories, potentially saving lives and minimizing economic disruptions caused by reactive lockdowns.

Reference: World Health Organization (WHO). (2023). WHO Ambient Air Quality Database, 2022 Update.

Commission for Air Quality Management (CAQM). (2025). Graded Response Action Plan (GRAP) for Delhi-NCR: Revised Guidelines.

2. Problem Statement

The primary challenge in air quality forecasting is the extreme non-stationarity and high variance of the data. AQI values are influenced by a myriad of dynamic factors including wind speed, boundary layer height, temperature inversions, and sudden anthropogenic spikes (e.g., Diwali firecrackers or agricultural stubble burning in neighboring states). Existing monitoring systems typically provide current values, but without reliable 24-to-72-hour forecasts, emergency response remains reactive. Therefore, developing an efficient predictive model capable of processing multi-pollutant dependencies and non-linear seasonal trends is critical for urban sustainability.

References: Ansari, A., & Quaff, A. R. (2024). "Advanced Machine Learning Techniques for Precise hourly Air Quality Index Prediction," *Int. Journal of Environmental Research*.

3. Objectives of the Study

The core objectives of this research are defined as follows:

1. To conduct a rigorous spatiotemporal analysis of Delhi’s historical AQI and meteorological data from 2019 through 2025.
2. To mathematically formulate, identify, and estimate a baseline ARIMA model using the Box-Jenkins methodology for linear time-series forecasting.
3. To develop, tune, and optimize an LSTM deep learning network capable of retaining long- term temporal dependencies and capturing non-linear pollution spikes.

4. To evaluate and compare the predictive performance of both models using standard statistical metrics including Root Mean Square Error (RMSE), Mean Absolute Error (MAE), and Coefficient of Determination (R^2).
5. To identify the superior model for real-world integration into Delhi's Graded Response Action Plan (GRAP) forecasting framework.

4. Literature Review

The application of time-series forecasting to environmental parameters has evolved significantly from basic moving averages to complex deep learning ensembles.

Statistical and Linear Models: Traditional methods like ARIMA and SARIMA have been widely utilized due to their strong mathematical foundations. A study by Atoui et al. (2022) evaluating AQI levels using Exponential Smoothing and SARIMA found that these models are highly effective at capturing seasonal variations in regions with stable climatic conditions. However, researchers have repeatedly noted that ARIMA struggles with the "high-variance" nature of pollutants in megacities, as it fundamentally assumes a linear relationship between past and future values.

Machine Learning and Deep Learning: To overcome the limitations of linear models, the focus has shifted toward Artificial Neural Networks (ANNs) and Recurrent Neural Networks (RNNs). Sidhu et al. (2024) investigated predictive modeling of AQI across India, highlighting that deep learning models better capture the impact of regional transmission paths, such as the transport of smoke from stubble burning in Punjab to Delhi. Similarly, studies leveraging the LSTM architecture emphasize its unique gating mechanisms, which solve the vanishing gradient problem found in standard RNNs, allowing the network to "remember" long-term seasonal patterns while reacting to sudden short-term anomalies. Recent comparative analyses across Indian metro cities have consistently demonstrated that LSTM models outperform traditional statistical methods in both RMSE and MAE when applied to multi-variate, non-linear environmental datasets.

5. Environmental Context and Data Acquisition

5.1 The Delhi Case Study and GRAP Framework

Delhi's air quality is heavily dictated by its landlocked geography and extreme seasonal meteorology. During the winter months (November to February), dropping temperatures and low wind speeds create a shallow boundary layer that traps pollutants near the surface—a phenomenon known as winter inversion. This is exacerbated by post-monsoon agricultural residue burning in neighboring states, which can account for up to 45% of Delhi's pollution during peak harvest season.

To manage this, the Commission for Air Quality Management (CAQM) enforces the Graded Response Action Plan (GRAP), categorizing interventions into four stages based on AQI:

- **Stage I (Poor, AQI 201-300):** Strict enforcement of dust control and waste management.
- **Stage II (Very Poor, AQI 301-400):** Targeted actions to reduce traffic and diesel generator usage.
- **Stage III (Severe, AQI 401-450):** Ban on non-essential construction and demolition activities.
- **Stage IV (Severe Plus, AQI >450):** Entry bans for commercial trucks and potential closure of schools and offices. The effectiveness of GRAP relies entirely on accurate 3-day advance predictions to pre-emptively invoke these stages before pollutants accumulate.

5.2 Dataset Description and Preprocessing

The dataset fuses ground-level sensor data with meteorological reanalysis:

1. **Pollutant Data:** Sourced from the Central Pollution Control Board (CPCB) stations (e.g., Anand Vihar, RK Puram) comprising hourly concentrations of PM_{2.5}, PM_{10}, NO_2, SO_2, CO, and O_3 from 2019 to 2025.
2. **Meteorological Data:** Sourced from the ECMWF ERA5 reanalysis dataset, providing hourly metrics for 2m temperature, relative humidity, boundary layer height, and 10m U/V wind components at a 0.25° x

0.25° spatial resolution.

Preprocessing Pipeline:

- *Imputation:* Missing values resulting from sensor calibration downtime were handled using k-Nearest Neighbors (kNN) imputation for minor gaps, and linear interpolation for longer sequences to preserve temporal flow.
- *Normalization:* Deep learning models are highly sensitive to input scale. All features for the LSTM model were transformed using Min-Max scaling to a range of $$ to ensure rapid convergence during backpropagation.

- *Stationarity (For ARIMA):* The Augmented Dickey-Fuller (ADF) test was applied. Non- stationary variables were differenced (d) until the p-value fell below the 0.05 threshold of significance.

5.3 Mathematical Calculation of the Indian AQI

The AQI is not a raw measurement but a calculated index. In India, the CPCB mandates that the final AQI is determined by calculating individual sub-indices for at least three pollutants (one of which must be $PM_{2.5}$ or PM_{10}) and selecting the maximum sub-index.

The sub-index (I) for a given pollutant concentration (C) is calculated using linear interpolation between defined breakpoints:

Where:

- C = Measured concentration of the pollutant.
- C_{HI}, C_{LO} = Concentration breakpoints greater than/equal to or less than/equal to C.
- I_{HI}, I_{LO} = AQI index values corresponding to C_{HI} and C_{LO}. The final AQI is the maximum of all sub-indices: $AQI = \max(I_{PM2.5}, I_{PM10}, I_{NO2},..., I_n)$.

References: Central Pollution Control Board (CPCB). (2014). National Air Quality Index (NAQI) Calculation Methodology. Hersbach, H., et al. (2020). "The ERA5 global reanalysis," Quarterly Journal of the Royal Meteorological Society.

6. Theoretical Framework of Predictive Models

6.1 ARIMA (AutoRegressive Integrated Moving Average)

ARIMA is a classical stochastic time-series model that characterizes data based on its own past values and past forecast errors. It is universally denoted by three non-negative parameters: (p, d, q).

6.1.1 Mathematical Formulation

The generalized equation for predicting a value y_t at time t is a combination of Autoregressive (AR) and Moving Average (MA) components on differenced data:

Where:

- p: Autoregressive order (number of lag observations).
- ϕ_i: Autoregressive coefficients.
- d: Integrated order (number of non-seasonal differences required for stationarity).
- q: Moving Average order (size of the moving average window of past errors).
- θ_j: Moving average coefficients.
- ϵ_t: White noise error term at time t.

References: Box, G. E. P., & Jenkins, G. M. (1970). Time Series Analysis: Forecasting and Control. Sidhu, et al. (2024). "Predictive modelling of AQI across various cities: Impact of stubble burning," Int. Journal for Multidisciplinary Research. Hochreiter, S., & Schmidhuber, J. (1997). "Long Short-Term Memory," Neural Computation.

6.1.2 The Box-Jenkins Methodology

To identify the optimal (p, d, q) parameters, this study adheres to the Box-Jenkins methodology, an iterative three-stage approach:

1. **Model Identification:** The data is tested for stationarity. If non-stationary, it is differenced. The Autocorrelation Function (ACF) and Partial Autocorrelation Function (PACF) plots are analyzed. The lag where the PACF cuts off dictates p, while the lag where the ACF cuts off dictates q.
2. **Parameter Estimation:** Maximum Likelihood Estimation (MLE) is used to compute the
ϕ_i and θ_j coefficients that minimize the Akaike Information Criterion (AIC) and Bayesian Information Criterion (BIC).
3. **Diagnostic Checking:** Residuals are analyzed to ensure they resemble white noise. The Ljung-Box test is applied to confirm the absence of significant autocorrelation in the error terms.

6.2 LSTM (Long Short-Term Memory) Networks

Standard Recurrent Neural Networks (RNNs) suffer from the "vanishing gradient" problem when backpropagating errors across long time sequences, rendering them unable to connect distant atmospheric events (e.g., summer conditions affecting winter baselines). The LSTM architecture was designed to overcome this by introducing an internal memory mechanism regulated by logic gates.

6.2.1 Architectural Components and Gate Equations

An LSTM cell maintains a **Cell State (C_t)**, the long-term memory, and a **Hidden State (h_t)**, the short-term working memory. The flow of information is regulated by three distinct gates:

1. **The Forget Gate (f_t):** Evaluates h_{t-1} and the current input x_t to determine what information from the past cell state should be discarded using a sigmoid (σ) activation:
2. **The Input Gate (i_t) and Candidate Memory ($\tilde{C}_t$):** The input gate decides which new information to store, while a $\tanh$ layer creates a vector of new candidate values: *Cell State Update:* The old state is updated by forgetting the irrelevant past and adding the scaled new candidates:
3. **The Output Gate (o_t):** Determines what part of the updated cell state will be output as the new hidden state:

(Note: W represents weight matrices and b represents bias vectors, optimized via the Adam optimizer during training).

6.2.2 Hyperparameter Optimization

An LSTM's performance relies heavily on its configuration. The network was optimized using grid search methodologies to determine the optimal number of hidden layers, units per layer, learning rate, and dropout rate (to prevent overfitting by randomly disabling neurons during training).

References: Mishra, A., & Gupta, Y. (2024). "Comparative analysis of AQI prediction using deep learning algorithms," *Spatial Information Research*.

- Sharma, G., Khurana, S., et al. (2024). "Comparative Analysis of ML Techniques in AQI prediction," *Int. Journal of System Assurance Engineering*.

7. Experimental Setup and Evaluation Metrics

The dataset was partitioned sequentially: 80% for training (2019–2023) and 20% for testing (2024–2025) to evaluate the models on the most recent, severe pollution episodes.

The models were evaluated using the following statistical criteria:

1. **Root Mean Square Error (RMSE):** Penalizes larger forecasting errors more severely, which is critical when predicting hazardous health limits.
2. **Mean Absolute Error (MAE):** Calculates the average magnitude of absolute errors.
3. **Coefficient of Determination (R^2):** Measures the proportion of variance in the dependent variable predictable from the independent variables.

(Where y_i is the actual value, \hat{y}_i is the predicted value, and \bar{y} is the mean of observed values).

8. Results and Discussion

8.1 Model Metrics Comparison

Training both models on the Delhi dataset yielded the following results for PM_{2.5} and overall AQI prediction over a 24-to-72 hour horizon:

Model Specification	RMSE (\mu g/m^3)	MAE (\mu g/m^3)	R^2 Score	Computational Cost
ARIMA (8,1,3)	42.15	28.60	0.81	Low (Seconds)
Vanilla LSTM	12.45	9.80	0.94	High (Hours)
Optimized LSTM	7.89	4.27	0.96	High (Hours)

8.2 Analysis of Residuals and Behavior

- **ARIMA Performance:** The ARIMA model performed adequately during the monsoon and summer months (March–September), where Delhi's AQI generally fluctuates linearly within the "Moderate" (101–200) range. However, residual analysis showed significant autocorrelation in error terms during winter. ARIMA could not "flatten" the seasonal spikes using basic differencing, causing it to consistently underestimate pollution peaks.
- **LSTM Performance:** The LSTM model demonstrated a superior ability to map the non- linear relationship between declining wind speeds, dropping temperatures, and exponential PM accumulation. Its residuals closely resembled white noise, proving the internal gating mechanisms successfully captured the underlying environmental logic.

8.3 Delhi AQI Trends (2024–2025): A Practical Evaluation

The superiority of the deep learning approach is best illustrated by its response to actual crisis events:

The November 2024 Crisis: On November 18, 2024, Delhi recorded a 24-hour AQI of **491** ("Severe Plus"). The ARIMA model predicted a gradual slope based on the previous week's trend. The LSTM model, analyzing the sudden drop in the ERA5 boundary layer height alongside historical November crop-burning patterns, accurately predicted the exponential spike 48 hours in advance, validating its utility for triggering GRAP Stage-IV.

October 2025 – The Worst in Three Years: Data analysis reveals October 2025 had the highest average AQI (232) of the past three Octobers, with a peak of 392. The LSTM accurately forecasted the sustained accumulation of pollutants that failed to disperse, unlike previous years where brief spikes quickly subsided.

Annual 2025 Trends: Despite intense winter smog, continuous data shows 2025 had 79 days of "Good" to "Satisfactory" air (the highest since 2018, excluding 2020), and only 8 days of "Severe to Severe+" air. However, the zero occurrences of truly "Good" air (AQI 0-50) throughout the year highlight the chronic baseline of urban emissions.

9. Conclusion

This study provides a comprehensive comparative analysis of time-series forecasting paradigms for urban environmental management. The Box-Jenkins ARIMA model serves as an efficient, low-overhead baseline suitable for stable meteorological periods. However, for megacities like Delhi—characterized by extreme seasonal volatility and multi-source emission profiles—the classical statistical approach is insufficient.

The LSTM architecture overwhelmingly outperforms ARIMA across all error metrics (R^2 of 0.96 vs 0.81). By utilizing its specialized forget and input gates, the LSTM successfully processes high-dimensional, non-linear meteorological dependencies, providing precise 72-hour forecasts even during unprecedented pollution spikes. Integrating such deep learning models into municipal frameworks is essential for transitioning systems like the Graded Response Action Plan (GRAP) from a reactive protocol to a proactive, predictive defense mechanism.

10. Future Scope

The transition to intelligent environmental governance opens several avenues for future research:

1. **Hybrid Architectures:** Developing CNN-LSTM or spatial-temporal Graph Neural Networks (GNNs) to combine the temporal memory of LSTMs with the spatial feature extraction of CNNs, allowing for kilometer-level grid forecasting across the NCR.
2. **Explainable AI (XAI):** Implementing SHAP (SHapley Additive exPlanations) to interpret the LSTM's "black-box" decisions, allowing policymakers to identify exactly which feature (e.g., wind speed vs. traffic volume) is driving a specific pollution spike.
3. **IoT Sensor Integration:** Transitioning from sparse, fixed CPCB stations to dense arrays of low-cost IoT sensors mounted on public transport to feed high-resolution, real-time data into the neural network.

ACKNOWLEDGEMENTS

The authors would like to express their sincere gratitude to the faculty members and mentors for their invaluable guidance, constructive feedback, and continuous support throughout the course of this research. Their expertise in time-series analysis and machine learning significantly contributed to the development and refinement of this study.

The authors also acknowledge the Central Pollution Control Board (CPCB) for providing access to comprehensive air quality data, and the European Centre for Medium-Range Weather Forecasts (ECMWF) for the ERA5 reanalysis dataset, which served as essential inputs for this research.

The authors are grateful to the broader research community for their foundational contributions in the domains of ARIMA modeling, deep learning, and environmental forecasting, which have informed and guided this work.

Furthermore, the authors would like to thank their institution for providing the necessary infrastructure and academic environment conducive to conducting this research.

Finally, the authors extend their appreciation to their families and peers for their encouragement and support throughout the completion of this study.

11. References

[1] World Health Organization, *WHO Ambient Air Quality Database, 2022 Update: Status Report*. Geneva, Switzerland, 2023.

[2] G. E. P. Box and G. M. Jenkins, *Time Series Analysis: Forecasting and Control*. San Francisco, CA, USA: Holden-Day, 1970.

[3] Central Pollution Control Board (CPCB), *National Air Quality Index (NAQI) Calculation Methodology and Breakpoints*. New Delhi, India: Ministry of Environment, Forest and Climate Change, Government of India, 2014.

[4] S. Hochreiter and J. Schmidhuber, "Long Short-Term Memory," *Neural Computation*, vol. 9, no. 8, pp. 1735–1780, 1997.

[5] H. Hersbach *et al.*, "The ERA5 global reanalysis," *Quarterly Journal of the Royal Meteorological Society*, vol. 146, no. 730, pp. 1999–2049, 2020.

[6] Sidhu *et al.*, "Predictive modelling of AQI across various cities and states in India: Impact of stubble burning," *International Journal for Multidisciplinary Research*, 2024.

[7] A. Ansari and A. R. Quaff, "Advanced Machine Learning Techniques for Precise Hourly Air Quality Index Prediction in Azamgarh, India," *International Journal of Environmental Research*, 2024.

[8] M. Atoui *et al.*, "Evaluating AQI levels using Naïve models, Exponential Smoothing, TBATS, and SARIMA in Lebanon," *Environmental Monitoring Studies*, 2022.

[9] W. Mao *et al.*, "Modeling Air Quality Prediction Using a Deep Learning Approach," *Journal of Environmental Science*, 2021.

[10] G. Sharma, S. Khurana *et al.*, "Comparative Analysis of Machine Learning Techniques in Air Quality Index (AQI) Prediction in Smart Cities," *International Journal of System Assurance Engineering and Management*, 2024.

[11] Commission for Air Quality Management (CAQM), *Graded Response Action Plan (GRAP) for Delhi-NCR: Revised Guidelines*, 2025.

[12] Q. Liu, B. Cui, and Z. Liu, "Air Quality Class Prediction Using Machine Learning Methods Based on Monitoring Data and Secondary Modeling," *Atmosphere*, 2024.

[13] H. Kumar *et al.*, "Artificial Intelligence Predictive Algorithms to Forecast Fine Particulate Matter," *Taiwan Environmental Monitoring*, 2021.

[14] A. Dikshit *et al.*, "Explainable AI (XAI) in Hydrometeorological Hazard Modeling," *Environmental Science*, 2024.

[15] A. Mishra and Y. Gupta, "Comparative Analysis of Air Quality Index Prediction Using Deep Learning Algorithms," *Spatial Information Research*, 2024.