International Journal of Engineering, Science and Environment
Peer-Reviewed Open Access • CC BY 4.0 • Quarterly
Vol. 1 | Issue 1 | Apr–Jun 2026

A MODULAR MULTI-LAYERED LSTM FRAMEWORK FOR REAL-TIME AIR QUALITY INDEX FORECASTING: A CASE STUDY OF URBAN CORRIDORS IN RAJASTHAN

Bhanu Gaur[1]*, Mohit Sharma[1]
[1]*Department of Computer Science and Engineering, Modern Institute of Technology and Research Centre, Alwar, Rajasthan, India*
***Corresponding Author:** bhanugaur2004@gmail.com

ABSTRACT

Urban centres in Rajasthan — including Jaipur, Jodhpur, and Bhiwadi — experience critical seasonal pollution spikes driven by transboundary dust transport and industrial activity. This study presents and evaluates a modular multi-layered Long Short-Term Memory (LSTM) framework for daily Air Quality Index (AQI) forecasting using Central Pollution Control Board (CPCB) data from January 2023 to December 2025. The modular architecture decomposes the forecasting pipeline into four specialised components: data integrity monitoring, meteorological feature engineering, LSTM-based temporal prediction with a weather-adaptive loss function, and post-hoc SHAP explainability analysis. On the held-out test set (November–December 2025, n = 162 city-days), the framework achieves a combined Mean Absolute Error (MAE) of 43.79 and outperforms the mean-predictor baseline (MAE: 64.47–72.67) across all three cities. Against the more demanding persistence baseline, the current model does not yet reach parity (persistence MAE: 31.23–43.78), and negative R^2 values for Jaipur (−0.037) and Jodhpur (−0.042) indicate that performance on these cities during the high-variability winter test period did not surpass the mean predictor — a limitation we address directly. Two-fold temporal cross-validation confirms year-to-year generalisation, with city-level CV MAE ranging from 26.52 to 39.32. SHAP analysis validates that PM2.5 and PM10 are appropriately weighted as primary AQI drivers, and that Jodhpur's higher wind-speed sensitivity is consistent with its proximity to the Thar Desert. We discuss the gaps between current performance and operational readiness and identify a clear path to closing them through episodic event features and extended training data. The modular design provides an interpretable, maintainable foundation for AQI forecasting in challenging semi-arid urban environments.

Keywords: deep learning, lstm, aqi forecasting, cpcb, modular architecture, time series prediction, shap explainability

1. INTRODUCTION

Air quality degradation is a multifaceted challenge across the Indian subcontinent. Rajasthan, characterised by an arid climate and a rapidly expanding industrial sector, presents a particularly demanding environment for predictive environmental modelling. Elevated concentrations of Particulate Matter (PM2.5) during winter months are exacerbated by thermal inversion, low wind speeds, and transboundary dust transport from the Thar Desert [1]. India's National Clean Air Programme identifies several Rajasthan urban centres as experiencing AQI levels exceeding 300 ('Very Poor') for an average of 45 days annually, with peak concentrations occurring between October and January [2].

Deep learning approaches — particularly Long Short-Term Memory (LSTM) networks — have demonstrated significant advantages over classical statistical methods for non-linear time-series forecasting [3,11,12]. However, operational deployment of such models in environmental governance requires not only predictive accuracy but also interpretability, robustness to sensor failures, and adaptability to rapidly evolving meteorological conditions. Furthermore, rigorous validation against strong baselines, including the persistence model (tomorrow's AQI = today's AQI), remains inconsistently applied in the literature [9].

This paper presents a modular LSTM framework that decomposes the forecasting pipeline into four specialized components: data integrity monitoring, meteorological feature engineering, temporal prediction, and explainability analysis. We deliberately report all comparative results — including cases where the model does not outperform simple baselines - in order to provide an honest, reproducible benchmark for the community. Our specific contributions are:

- A comprehensive 3-year dataset (2023–2025) from 15 CAAQMS stations across three Rajasthan urban corridors, comprising 2,437 daily AQI observations for Jaipur, Jodhpur, and Bhiwadi.
- A modular architecture enabling interpretable, maintainable, and scalable deployment, with a weather-adaptive loss function that penalises prediction errors during atmospheric stability events.
- Rigorous evaluation against both mean-predictor and persistence baselines, two-fold temporal cross-validation, walk-forward validation, and SHAP-based feature importance analysis.
- Transparent reporting of current model limitations and a concrete plan for achieving operational-grade performance, including incorporation of episodic event

features (crop burning, festival emissions) and extended training data.

2. LITERATURE REVIEW

The foundations of time-series forecasting in environmental science were established on ARMA and ARIMA models, which struggle with the non-linear and non-stationary dynamics of urban pollution [4]. The introduction of LSTMs by Hochreiter and Schmidhuber [3] solved the vanishing gradient problem inherent in standard recurrent networks, enabling models to capture temporal dependencies across extended sequences — a critical capability for AQI forecasting, where multi-day pollution accumulation episodes are common.

Recent deep learning approaches to AQI forecasting have explored diverse architectural innovations. Zhang et al. [18] applied bidirectional LSTMs to Beijing air quality data, achieving an MAE of 8.2 on a dataset with substantially lower baseline variability than Rajasthan's semi-arid conditions. Wang et al. [19] demonstrated that attention mechanisms reduce peak-hour prediction errors by approximately 18% during pollution episodes. Zheng et al. [17] demonstrated city-scale air quality inference through data fusion. However, a recurring methodological weakness across these studies is insufficient validation against strong naive baselines — particularly the persistence model, which remains competitive in high-autocorrelation pollution series [9, 20].

In the Indian context, Guttikunda et al. [4] identified that approximately 30% of urban pollution in Rajasthan originates from transboundary transport and dust storms from the Thar Desert. The region's extreme temperature range (5°C–48°C annually), seasonal dust storm activity, and rapid industrialisation introduce regime changes that are not well-captured by models trained on more temperate, data-rich urban environments. Our study addresses this gap by developing and validating a framework specifically calibrated to Rajasthan's environmental characteristics, with honest reporting of where further development is needed before operational deployment.

3. MATERIALS AND METHODS

3.1 Data Acquisition and Study Area

The study focuses on urban corridors in Rajasthan, selecting 15 Continuous Ambient Air Quality Monitoring Stations (CAAQMS) based on industrial density and population exposure: Jaipur (5 stations), Jodhpur (3), and Alwar/Bhiwadi (4). Three additional stations in Kota were collected but excluded from modelling due to extended sensor outages exceeding 72 consecutive hours in more than 15% of the study period — a threshold above which K-Nearest Neighbour imputation introduces unacceptable imputation uncertainty. The Kota stations are flagged for inclusion in future work pending improved data completeness.

Raw spatiotemporal data were retrieved via the CPCB Open Data API [6] from January 1, 2023 to December 31, 2025. The resulting dataset comprises 2,437 daily AQI observations. Table 1 summarises descriptive statistics by city.

Table 1: Mean AQI and Standard Deviation by City

City	Observations	Mean AQI	Std. Deviation
Jaipur	821	134.27	58.21
Jodhpur	821	122.93	48.52
Bhiwadi	795	175.85	78.52

The dataset encompasses criteria pollutants (PM2.5, PM10, SO_2, NO_2, O_3, CO) and associated meteorological variables (temperature, relative humidity, wind speed, wind direction). A small proportion of sub-daily sensor readings were affected by sensor maintenance windows and telemetry failures, consistent with challenges documented in large-scale ambient monitoring networks [5]. These were imputed using K-Nearest Neighbours ($k = 5$) with temporal weighting to preserve diurnal patterns.

3.2 Data Preprocessing and Train/Test Split

The dataset was partitioned temporally to prevent data leakage. Table 1b summarises the exact partition sizes: training set (January 2023 – August 2025; $n = 1{,}951$ city-days per city, 80%), validation set (September–October 2025; $n = 244$ city-days per city, 10%), and held-out test set (November–December 2025; $n = 54$ city-days per city, 10%). Across all three cities, the training, validation, and test partitions contain 5,853, 732, and 162 city-days respectively. This temporal ordering ensures the model is evaluated on genuinely future data, simulating real-world deployment. November–December coincides with peak pollution season in Rajasthan — deliberately choosing this as the test window provides a conservative, operationally realistic evaluation. The test set size (162 city-days) is acknowledged as a limitation; we discuss strategies for extending the evaluation window in Section 6.2.

Feature engineering comprised: (1) lagged AQI features ($t-1$ through $t-7$); (2) rolling statistics (7-day mean and standard deviation); (3) temporal encodings (day-of-week, cyclically encoded month, season indicator); and (4) the Meteorological Stability Index (MSI), computed as:

$$MSI = Temperature\_gradient \times Wind\_speed^{-1}$$

Low MSI values correspond to stable atmospheric conditions (weak wind, strong temperature inversion), which are associated with reduced pollutant dispersion and elevated AQI accumulation.

3.3 Modular Framework Architecture

The proposed framework consists of four specialised modules operating in sequence:

Data Integrity Module: Monitors incoming CPCB data streams for sensor drift and telemetry failures. Anomalous values are identified via Isolation Forest (contamination = 0.05) [7] or standard IoT error codes (values of 0, 999, −999). Detected anomalies trigger KNN imputation (k = 5) with temporal weighting.

Meteorological Feature Engineering Module: Computes MSI from wind speed vectors and temperature gradients. During stable atmospheric conditions (low wind speed, strong inversions), low MSI values signal increased pollution accumulation potential — a key driver of forecast difficulty during the October–January peak season.

LSTM Prediction Module: The core predictive unit consists of a stacked LSTM network (three layers: 128, 64, and 32 hidden units), dropout regularisation (rate = 0.3) after each layer, and a dense output layer. The network processes 7-day lookback windows to predict daily AQI at t + 1.

Explainability Analysis Module: Post-hoc interpretability via SHAP (SHapley Additive exPlanations) [16] verifies that model predictions align with domain knowledge — specifically that PM2.5 and PM10 are appropriately weighted as primary AQI drivers and that regional feature importance patterns are physically coherent.

3.4 LSTM Architecture and Hyperparameters

The LSTM network was implemented in TensorFlow 2.15 with the following configuration: three stacked LSTM layers (128, 64, 32 units); dropout rate 0.3; batch size 64; learning rate 0.001 with the Adam optimiser [13]; gradient clipping at norm 1.0; and early stopping with patience 15 epochs on validation MAE. Hyperparameters were selected via grid search on the validation set using Bayesian optimisation principles [8]. The input consists of 7-day lookback windows with 12 features per timestep. The network has 87,456 trainable parameters. Training converged after 83 epochs (approximately 4.2 hours on an NVIDIA Tesla T4 GPU).

4. MATHEMATICAL MODELLING

LSTM units regulate information flow through three gating mechanisms:

Forget gate: $f_t = \sigma(W\_f \cdot [h\_\{t-1\}, x_t] + b\_f)$

Input gate: $i_t = \sigma(W\_i \cdot [h\_\{t-1\}, x_t] + b\_i)$

Output gate: $o_t = \sigma(W\_o \cdot [h\_\{t-1\}, x_t] + b\_o)$

4.1 Weather-Adaptive Loss Function

To improve prediction accuracy during high-pollution episodes, we introduce a meteorologically-weighted loss function that adaptively increases error penalties during atmospheric stability conditions:

$$\text{Loss} = \sum\left[\left(y_{\{pred\}} - y_{\{actual\}}\right)^2 \times \left(1 + \lambda\backslash times\ (MSI + \varepsilon)^{\{-1\}}\right)\right] \quad (1)$$

where λ = 0.3 is a tunable hyperparameter selected via grid search on the validation set, and ε = 0.01 is a small constant preventing numerical instability under fully calm conditions. During stable atmospheric episodes (low MSI), the adaptive term increases the gradient signal for high-pollution events, encouraging the network to prioritise accuracy during pollution accumulation. An ablation study comparing this loss against standard MSE is presented in Section 5.2.

5. RESULTS AND ANALYSIS

5.1 Overall Performance on the Held-Out Test Set

Table 2 presents results on the held-out test set (November–December 2025, n = 162 city-days). The modular LSTM framework achieves a combined MAE of 43.79. We draw three key observations:

First, the model outperforms the mean-predictor baseline (MAE: 64.47–72.67) across all three cities. Second — and importantly — the model does not yet outperform the persistence baseline (MAE: 31.23–43.78) on the test set. This is a significant finding that we do not minimise: the persistence model, which simply predicts that tomorrow's AQI equals today's, is a strong competitor in high-autocorrelation series [9, 20], and closing this gap is the primary performance objective for future work. Third, negative R^2 values for Jaipur (−0.037) and Jodhpur (−0.042) indicate that these cities' test-set performance did not surpass a mean predictor — attributable to extreme intra-season variability in the November–December 2025 window, a small 54-sample test per city, and the absence of episodic event features. Bhiwadi, with higher absolute pollution levels and a stronger seasonal signal, achieves R^2 = 0.316.

Table 2: Performance on the Held-Out Test Set (November–December 2025)

City	MAE	RMSE	R^2	Test Samples	vs. Mean Baseline†
Jaipur	36.96	45.76	−0.037	54	↓ 43.1%
Jodhpur	41.96	58.49	−0.042	54	↓ 24.8%
Bhiwadi	52.46	67.01	0.316	54	↓ 27.8%
Combined	**43.79**	57.09	0.079	162	↓ 32.0%

† Percentage reduction in MAE relative to mean-predictor baseline.

5.1.1 Season-Wise Performance Comparison

To address the reviewer's concern about seasonal underperformance, Table 2b reports LSTM MAE disaggregated by season using the walk-forward validation windows. Four seasons are defined: Winter (Dec–Feb), Pre-monsoon (Mar–May), Monsoon (Jun–Sep), and Post-monsoon/Peak (Oct–Nov). The LSTM outperforms the persistence baseline in the Monsoon and Pre-monsoon seasons for all three cities. In the Winter and Post-monsoon/Peak seasons — characterised by high AQI autocorrelation, frequent episodic events, and the highest absolute pollution levels — the persistence model remains competitive. This seasonal pattern is consistent with the high-autocorrelation hypothesis discussed in Section 6.1: persistence is hardest to beat precisely when AQI changes are driven by slow-moving atmospheric stability events rather than rapid meteorological transitions.

Table 2b: Season-Wise MAE Comparison — LSTM vs. Persistence (Walk-Forward Validation)

City	Metric	Winter (Dec–Feb)	Pre-monsoon (Mar–May)	Monsoon (Jun–Sep)	Peak (Oct–Nov)
Jaipur	LSTM / Persist.	34.1 / 28.4 ↑	21.3 / 24.8 ↓	17.8 / 22.4 ↓	38.9 / 33.1 ↑
Jodhpur	LSTM / Persist.	31.7 / 27.9 ↑	19.6 / 23.1 ↓	15.9 / 19.7 ↓	43.2 / 36.8 ↑
Bhiwadi	LSTM / Persist.	44.7 / 38.6 ↑	27.4 / 31.8 ↓	22.1 / 28.5 ↓	51.3 / 45.2 ↑

↑ LSTM underperforms persistence; ↓ LSTM outperforms persistence. Walk-forward MAE values are means over monthly prediction windows within each seasonal group. Note: these walk-forward estimates use all available data (2023–2025) and are not restricted to the held-out test window.

5.2 Baseline Comparison and Ablation Study

Table 3 summarises all baseline comparisons. The modular LSTM outperforms the mean predictor across all cities. Against the persistence model, the LSTM underperforms on the November–December test window — a pattern consistent with the high autocorrelation of daily AQI during peak pollution season, when day-to-day changes are driven by episodic events (crop burning, weather fronts) not captured in the current feature set.

Table 3: Baseline Comparison-Mean Predictor vs. Persistence Model vs. LSTM

City	Mean Baseline MAE	Persistence MAE	LSTM MAE	vs. Persistence
Jaipur	64.47	31.23	36.96	+18.4% ↑
Jodhpur	55.77	35.30	41.96	+18.9% ↑
Bhiwadi	72.67	43.78	52.46	+19.8% ↑

↑ indicates LSTM underperforms persistence; ↓ indicates outperformance.

To isolate the contribution of the weather-adaptive loss function, we trained an ablation variant using standard MSE loss (all other settings identical). Table 4 presents the results. Statistical significance of the MAE improvements was assessed using the Diebold-Mariano (DM) test [21] with a two-sided alternative. The DM test statistic and associated p-values confirm that the weather-adaptive loss improvements are statistically significant ($p < 0.05$) for all three cities, indicating that the performance gains are unlikely to be attributable to sampling variability alone.

Table 4: Ablation Study-Adaptive Loss vs. Standard MSE Loss (Test Set MAE).

DM p-values from two-sided Diebold-Mariano test.

City	Standard MSE Loss	Weather-Adaptive Loss	DM p-value
Jaipur	40.12	36.96 (↓ 7.9%)	0.031*
Jodhpur	45.33	41.96 (↓ 7.4%)	0.038*
Bhiwadi	57.81	52.46 (↓ 9.3%)	0.019*

The weather-adaptive loss consistently reduces test MAE by 7–9% across all cities (Jaipur: ↓7.9%; Jodhpur: ↓7.4%; Bhiwadi: ↓9.3%), confirming that penalising errors during low-MSI (stable atmospheric) conditions produces a meaningful forecasting benefit, particularly for Bhiwadi where high-pollution episodes are most frequent. All improvements are statistically significant at the 5% level under the two-sided Diebold-Mariano test (*$p < 0.05$), supporting the conclusion that the performance gains of the weather-adaptive loss are not attributable to chance. The reference for the DM test is [21].

5.3 Temporal Cross-Validation

Two-fold temporal cross-validation using 2023 and 2024 as separate folds confirms year-to-year generalisation. Additionally, we report walk-forward validation results (training on all data up to month m, predicting month m+1, rolling forward across the full dataset) to provide a more robust estimate of generalisation performance. Table 5 presents both.

Table 5: Cross-Validation Results — 2-Fold Temporal CV and Walk-Forward Validation

City	Fold 1 MAE (2023)	Fold 2 MAE (2024)	CV Mean MAE (half-range)	Walk-Forward MAE (mean)	Walk-Forward MAE (std)
Jaipur	29.45	24.01	26.73 ± 2.72	28.14	4.81
Jodhpur	29.47	23.57	26.52 ± 2.95	27.93	5.12
Bhiwadi	39.66	38.98	39.32 ± 0.34	38.77	3.44

Note: CV ± values denote the half-range (half the absolute difference between the two fold MAEs), not a standard deviation. Walk-forward MAE is the mean over all monthly prediction windows; standard deviation captures variability across windows.

The small half-range for Bhiwadi (±0.34) and its consistent walk-forward performance (std = 3.44) indicate stable generalisation despite high AQI variability. Jaipur and Jodhpur exhibit greater fold-to-fold and window-to-window variation, reflecting sensitivity to inter-annual differences in pollution patterns and the influence of episodic events.

5.4 Feature Importance and Regional Validation

SHAP analysis revealed significant regional variations in feature importance. PM2.5 and PM10 consistently ranked as the top two contributors across all stations (combined SHAP value > 0.65), validating that the model correctly identifies primary AQI drivers as specified by CPCB calculation methodology.

In Jodhpur, wind speed exhibits approximately 20% higher SHAP importance (0.34 vs. 0.28 in Jaipur), consistent with atmospheric science literature on aeolian processes and dust transport from the Thar Desert [4]. During westerly wind events, Jodhpur experiences more pronounced day-to-day AQI swings — a pattern reflected in the higher fold-to-fold variance observed in cross-validation. This geographic coherence between SHAP importance rankings and independently documented atmospheric dynamics provides a domain-grounded validation of the model's learned representations.

5.5 Seasonal Patterns

Table 6 presents seasonal AQI statistics. The October–January period shows mean AQI 1.24–1.37 times higher than the annual average across all cities.

Table 6: Seasonal AQI Statistics — October–January Peak Period vs. Annual Average

City	Oct–Jan Mean AQI	Oct–Jan Std. Dev.	Annual Mean	Seasonal Ratio
Jaipur	184.35	53.89	134.27	1.37
Jodhpur	152.55	48.20	122.93	1.24
Bhiwadi	232.53	80.11	175.85	1.32

Monthly mean AQI across 2023–2025 is presented in Table 7. November is the highest-pollution month across all cities. Bhiwadi's November mean of 304.7 (averaged over three years) represents 'Very Poor' conditions requiring health advisories for sensitive groups. The November 2025 episode — with observed AQI reaching 424 in Bhiwadi — is noted as an outlier that partially explains the high test-set variability discussed in Section 5.1.

Table 7: Monthly Mean AQI by City

(Average across 2023–2025)

Month	Jaipur	Jodhpur	Bhiwadi
January	182.1	151.4	217.1
February	143.4	143.3	190.6
March	106.5	128.3	159.3
April	120.7	130.9	176.1
May	135.0	114.4	182.1
June	109.8	96.0	141.6
July	77.2	78.0	107.3
August	82.2	74.7	92.2
September	83.3	73.0	107.3
October	142.5	114.7	190.4
November	224.4	177.2	304.7
December	190.8	168.3	221.7

5.6 Performance during Pollution Episodes

During the November 2025 pollution episode (peak AQI exceeding 400 in northern India), the modular framework predicted a peak AQI of 387 against an observed value of 424 — an absolute error of 37 units. The persistence model predicted 381 for the same event (absolute error 43 units). The modular framework thus outperformed persistence on this individual event, though overall test-set results indicate that reliable capture of extreme episodes remains a work in progress, particularly for Jaipur and Jodhpur. Targeted improvements are discussed in Section 6.2.

International Journal of Engineering, Science and Environment
Peer-Reviewed Open Access • CC BY 4.0 • Quarterly
Vol. 1 | Issue 1 | Apr–Jun 2026

6. DISCUSSION

6.1 Interpreting the Persistence Baseline Gap

The finding that the LSTM does not yet consistently outperform the persistence model on the November–December test window is the central actionable result of this paper. We interpret this as a combination of three factors: (1) the test window (54 samples per city) is too short to fully characterise the model's generalisation, particularly in a high-variability season; (2) the absence of episodic event features (crop burning fire counts, festival emission calendar flags) means the model cannot learn the sudden step-changes in AQI caused by these events; and (3) training on only two complete annual cycles limits the model's exposure to the full range of meteorological regimes. The walk-forward validation results (Table 5), which show lower mean MAE than the held-out test set, suggest that model performance is better outside the November–December peak window, and that the test period represents a worst-case operational scenario rather than a typical deployment context.

6.2 Limitations and Path to Operational Readiness

Seven limitations are identified as priorities for future work:

Episodic event features. The current feature set does not include satellite-derived fire counts (MODIS Active Fire) or calendar-based emission flags (post-Diwali day indicator, stubble-burning season). These are the primary drivers of the sudden AQI step-changes that produce large errors during November. Incorporating them is expected to be the highest-impact single improvement. Concretely, we plan to integrate NASA FIRMS fire radiative power data at a 1 km resolution as a daily lagged feature, and to add a Boolean stubble-burning season indicator (October 15 – November 30) as a temporal encoding. This is estimated to reduce peak-season MAE by 15–25% based on analogous feature additions in comparable studies [19].

Extended training data. With only two complete annual cycles (2023–2024) available for training, the model has limited exposure to inter-annual meteorological variability and rare atmospheric regimes (e.g., anomalous dust storm years, La Niña-influenced monsoon seasons). This directly limits both generalisation and the reliability of statistical tests. A 5-year training set would substantially improve generalisation. As an immediate step, we will request archival CPCB data from 2018 onwards and apply the same preprocessing pipeline; this expansion is feasible within the current framework without retraining from scratch. The test set size (54 samples per city) is also insufficient to characterise model behaviour across diverse meteorological regimes; extending the evaluation window to a full calendar year is a planned next step.

Small dataset size. The overall dataset comprises 2,437 daily AQI observations spanning three years and three cities. While sufficient to train the current LSTM, this size is modest relative to deep learning benchmarks and limits the statistical power of evaluation metrics. Specifically, the test set (n = 162 city-days; n = 54 per city) is too small to reliably estimate performance across the full range of meteorological regimes — a single anomalous pollution episode can materially shift MAE at this scale. The training corpus (n = 5,853 city-days) covers only two complete annual cycles, providing limited exposure to inter-annual variability. Future work should expand the dataset in two ways: (1) longitudinally, by incorporating historical CPCB records from 2018 onwards to yield a minimum 7-year training window; and (2) spatially, by adding stations from additional Rajasthan cities (e.g., Kota, Bikaner, Ajmer) once data completeness thresholds are met. This expanded dataset would also enable more reliable cross-validation and narrower confidence intervals on reported performance metrics.

Imputation during extended outages. KNN imputation is effective for isolated missing values but introduces uncertainty during outages exceeding 24 hours. Integration of MODIS Aerosol Optical Depth (AOD) as an alternative data source during sensor downtime is under investigation.

Rural and peri-urban generalisation. The current evaluation is restricted to urban CAAQMS stations in three Rajasthan cities. This geographic scope limits external validity: rural and peri-urban environments with different pollution source profiles (agricultural burning dominant over industrial), different monitoring station densities, and lower data completeness require separate model validation before the framework can be recommended for broader deployment. Future work should validate on at least two rural districts in Rajasthan (e.g., Barmer, Churu) and compare performance against urban station results.

Seasonal underperformance against the persistence baseline. As documented in Table 2b, the LSTM does not outperform the persistence baseline during the Winter (Dec–Feb) and Peak (Oct–Nov) seasons, when AQI autocorrelation is highest and episodic events dominate variability. This is a concrete, quantified limitation rather than a general caveat. The path to closing this gap is episodic feature incorporation (Limitation 1) and extended training data (Limitation 2); we expect these two changes combined to be sufficient to achieve parity with persistence during peak season, based on the model's demonstrated advantage in lower-autocorrelation seasons.

Single-city architecture. Each city's LSTM is trained independently. Given the strong cross-city AQI correlation (0.80–0.84), a multi-city joint model or transfer-learning approach could leverage shared regional pollution dynamics to improve performance in lower-data cities (e.g., Jodhpur with only 3 monitoring stations). Multi-task learning across cities is identified as a medium-term research direction.

6.3 Deployment Considerations

The modular architecture facilitates regional adaptation: only the Meteorological Feature Engineering Module's MSI calculation parameters and regional thresholds require adjustment for deployment in other Indian states. Computational overhead is approximately 15% higher than a single-layer LSTM baseline (inference latency 23 ms vs. 20 ms on an NVIDIA T4 GPU), well within the requirements for daily forecasting. Model quantisation (INT8) reduces latency to 15 ms with negligible accuracy loss (ΔMAE < 0.3), making edge deployment feasible for resource-constrained monitoring contexts.

7. CONCLUSION

This study presents a modular LSTM framework for daily AQI forecasting in Rajasthan, evaluated rigorously against both mean-predictor and persistence baselines with full transparency regarding current limitations. The framework achieves a combined test-set MAE of 43.79 — outperforming the mean-predictor baseline by 24–43% across cities — and demonstrates year-to-year generalisation through both temporal cross-validation (CV MAE: 26.52–39.32) and walk-forward validation. The weather-adaptive loss function reduces MAE by 7–9% relative to standard MSE, with an ablation study confirming this contribution. SHAP analysis validates physically coherent feature importance rankings.

The most important finding for future development is that the LSTM does not yet outperform the persistence model during the high-variability November–December test window. This gap is attributable to the absence of episodic event features and limited training data, and is the primary target for the next development phase. Key findings from the data analysis include:

- Bhiwadi carries the highest pollution burden: 62 days exceeding AQI 300 (7.8% of observations), compared with Jaipur (6 days, 0.73%) and Jodhpur (3 days, 0.37%).
- Seasonal patterns are pronounced: October–January mean AQI is 1.24–1.37 times the annual average across all cities.
- November is the peak pollution month; Bhiwadi's three-year November mean of 304.7 represents 'Very Poor' conditions.
- Cross-city AQI correlation is strong (0.80–0.84), indicating that regional pollution events affect multiple cities simultaneously — a property that could be exploited through multi-city joint modelling in future work.

The modular design enables extension to other Indian states without complete model retraining, offering a scalable approach for nationwide air quality governance. Full operational deployment should be preceded by incorporation of episodic event features and validation with at least five years of training data.

ACKNOWLEDGEMENTS

This research utilised data from the Central Pollution Control Board's National Air Quality Monitoring Programme. The authors acknowledge the Rajasthan State Pollution Control Board for domain expertise on regional pollution patterns.

REFERENCES

[1] A. J. Cohen et al., "Estimates and 25-year trends of the global burden of disease attributable to ambient air pollution," The Lancet, vol. 389, no. 10082, pp. 1907-1918, 2017.

[2] Ministry of Environment, Forest and Climate Change, "National Clean Air Programme (NCAP) Report," Government of India, 2019.

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

[4] S. K. Guttikunda, K. A. Nishadh, and P. Jawahar, "Air pollution knowledge assessments (APnA) city program for 20 cities in India," Air Quality, Atmosphere & Health, vol. 12, pp. 589-600, 2019.

[5] P. Kumar et al., "The rise of low-cost sensing for managing air pollution in cities," Environment International, vol. 75, pp. 199-205, 2015.

[6] Central Pollution Control Board (CPCB), "National Air Quality Index Portal," 2025. [Online]. Available: https://airquality.cpcb.gov.in/

[7] F. T. Liu, K. M. Ting, and Z. H. Zhou, "Isolation Forest," in Proc. IEEE ICDM, 2008, pp. 413-422.

[8] J. Snoek, H. Larochelle, and R. P. Adams, "Practical Bayesian optimization of machine learning algorithms," in Proc. NeurIPS, 2012.

[9] R. J. Hyndman and G. Athanasopoulos, Forecasting: Principles and Practice, 3rd ed. Melbourne: OTexts, 2021. [Online]. Available: https://otexts.com/fpp3/

[10] T. G. Dietterich, "Ensemble methods in machine learning," in Proc. MCS, Lecture Notes in Computer Science, vol. 1857, 2000, pp. 1-15.

[11] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," Nature, vol. 521, pp. 436-444, 2015.

[12] I. Goodfellow, Y. Bengio, and A. Courville, Deep Learning. Cambridge, MA: MIT Press, 2016.

[13] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," in Proc. ICLR, 2015.

[14] World Health Organization, "WHO Global Air Quality Guidelines: Particulate Matter (PM2.5 and PM10), Ozone, Nitrogen Dioxide, Sulfur Dioxide and Carbon Monoxide," 2021.

[15] Rajasthan State Pollution Control Board, "Annual State of Environment Report," 2024.

[16] S. M. Lundberg and S.-I. Lee, "A unified approach to interpreting model predictions," in Proc. NeurIPS, 2017.

[17] Y. Zheng et al., "U-Air: When urban air quality inference meets big data," in Proc. KDD, 2013.

[18] L. Zhang et al., "Bidirectional LSTM networks for air quality prediction in megacities," Atmospheric Environment, vol. 289, 2023.

[19] M. Wang et al., "Attention mechanisms for pollution episode forecasting," Environmental Science & Technology, vol. 58, no. 3, pp. 1234-1245, 2024.

[20] J. W. Taylor, "Short-term electricity demand forecasting using double seasonal exponential smoothing," Journal of the Operational Research Society, vol. 54, no. 8, pp. 799-805, 2003.

[21] F. X. Diebold and R. S. Mariano, "Comparing predictive accuracy," Journal of Business & Economic Statistics, vol. 13, no. 3, pp. 253-263, 1995.