Journal Homepage: www.ijarpr.com ISSN: 3049-0103 (Online)

International Journal of Advance Research Publication and Reviews

Vol 03, Issue 8, pp 557-572, August, 2026

Forecasting the Air Quality Index (AQI) in Sub-Saharan Africa: Leveraging Seasonal Patterns and Deep Learning for Abuja, Nigeria

Onyeche Vera Adikwu, Nkiruka Favour Ozoemena, Jophia Nicholas Kemi, Mariam Jibrin Usman, Laura Yusuf-Ufua, Wash Patrick Madugu

Department of Mission Planning and Satellite Data Management, National Space Research and Development Agency (NASRDA).

ABSTRACT

Air pollution is one of the leading environmental health risks in Sub-Saharan Africa, yet the region remains among the most sparsely instrumented for continuous air quality monitoring. Abuja, Nigeria's capital, records Air Quality Index (AQI) levels that swing sharply between a comparatively clean wet season and a heavily polluted Harmattan-dominated dry season, driven by desert dust transport, biomass burning, and reduced atmospheric mixing. Existing Nigerian AQI forecasting studies have generally relied on single-model approaches evaluated over short study windows or leakage-prone random data splits, and none has yet targeted Abuja with a design that explicitly encodes seasonal structure. This paper presents a seasonally-aware deep learning methodology for forecasting AQI in Abuja, built on two years (January 2024–December 2025) of hourly data comprising 17,208 observations across AQI, PM2.5, PM10, NO2, O3, SO2, CO, NH3 and NO. Following a leakage-safe chronological 70/15/15 split, a gradient-boosted tree ensemble (XGBoost) and a Bidirectional LSTM were benchmarked against persistence and seasonal-naive baselines. XGBoost achieved the strongest one-hour-ahead forecasting performance (RMSE = 0.123, MAE = 0.070, R^2 = 0.988, classification accuracy = 98.9%), outperforming the Bi-LSTM (RMSE = 0.314, R^2 = 0.920) and both baselines, with the persistence baseline (RMSE = 0.347, R^2 = 0.902) itself proving difficult to beat -- underscoring the strong hour-to-hour autocorrelation in Abuja's AQI series. Both candidate models generalized across seasons, with lower error in the wet season than the Harmattan-dominated dry season. These results, together with the underlying seasonal feature engineering and evaluation framework, provide an empirically validated forecasting methodology for Abuja and a template for other data-scarce Sub-Saharan African cities.

Keywords: Air Quality Index; AQI forecasting; deep learning; ensemble learning; seasonal patterns; Sub-Saharan Africa; Abuja; Nigeria; time-series forecasting

1. Introduction

1.1 Background of the Study

Ambient air pollution is one of the most significant environmental risk factors for human health worldwide, responsible for millions of premature deaths every year (Bernacki & Scherer, 2025). The Air Quality Index (AQI) translates raw pollutant concentrations -- principally PM2.5, PM10, NO2, O3, SO2 and CO -- into a single categorical scale that the public and policymakers can interpret without specialist training. Sub-Saharan Africa carries a disproportionate share of the global air pollution burden while possessing a small fraction of the world's monitoring infrastructure: Okure (2026) reports roughly 0.6 continuous monitoring stations per million people across Africa, against roughly 8 per million in the European Union and 30 per million in the United States, with more than a third of African countries operating no continuous ambient monitors at all.

Nigeria is consistently identified among the continent's most polluted countries, and Abuja is frequently ranked among its most polluted cities alongside Lagos and Benin City. Ground-based studies report Abuja PM2.5 concentrations several times higher in the dry season than the wet season, with a Purple Air-based study (Omokungbe et al., 2025) finding PM2.5-based AQI exceeded the 'Unhealthy for Sensitive Groups' threshold in every month from November to April across two consecutive years -- evidence that the fine particulate fraction drives Abuja's acute air quality episodes. This seasonal signal is driven by well-understood mechanisms: the Harmattan (roughly November--March) is a dust-laden Saharan trade wind that suppresses atmospheric mixing and transports mineral dust across West Africa, while the wet season (roughly April--October) brings rainfall that scavenges particulates and suppresses biomass burning. A forecasting system that does not explicitly encode this seasonal structure risks learning an average behavior that fits neither regime well.

The broader machine learning literature on air quality forecasting has moved toward deep learning and hybrid approaches that combine statistical decomposition with sequence models, repeatedly shown to outperform single-paradigm baselines on air quality data from Asia, Europe and India (Bernacki & Scherer, 2025). Very little of this seasonally-aware methodology has yet been applied to a Sub-Saharan African city, and none of the Nigeria-specific studies reviewed for this paper has applied it to Abuja specifically.

1.2 Statement of the Problem

I. Three inter-related gaps motivate this study:

II. Monitoring and modelling scarcity: continuous, ML-ready AQI datasets for Abuja are rare, and the few Nigerian AQI forecasting studies identified focus on Lagos and Port Harcourt rather than Abuja (Ogunsanwo et al., 2025; Taylor & Ezekiel, 2023).

III. Single-model, seasonally-blind designs: existing Nigerian AQI/pollutant forecasting studies largely evaluate a single model in isolation rather than a design that explicitly separates and models seasonal structure (Balogun & Zakari, 2025; Taylor & Ezekiel, 2023; Akintola et al., 2026).

IV. Validation and leakage risk: random or spatially-mixed train/test splits are common in the reviewed literature and inflate reported accuracy for time-dependent data (Akintola et al., 2026; Singh & Shukla, 2026).

No existing study offers a rigorously validated, seasonally-structured forecasting methodology tailored to Abuja. This paper addresses that gap at the methodological level, ahead of a subsequent model-development phase.

1.3 Aim and Objectives

The aim of this study is to establish the seasonally-aware research design and methodology for forecasting the Air Quality Index in Abuja, Nigeria using deep learning, as a template that leverages the city's seasonal patterns to inform the subsequent forecasting model.

The objectives are to:

I. Characterize the seasonal and diurnal structure of AQI and its constituent pollutants in Abuja using two years of hourly data;

II. Design a preprocessing and quality-control pipeline for the identified data anomalies;

III. Design a seasonally-aware feature engineering strategy that exposes the Harmattan/wet-season structure to deep learning models;

IV. Define a leakage-safe, chronological data-splitting and evaluation protocol; and establish the evaluation framework against which the subsequent deep learning forecasting model will be benchmarked against a persistence baseline.

1.4 Significance of the Study

Practically, a validated, seasonally-aware forecasting methodology for Abuja can inform public health advisories and give residents actionable early warning during high-pollution Harmattan episodes, particularly for children, the elderly and people with pre-existing respiratory conditions. Academically, the study responds to calls in the literature to align model complexity with data availability and adopt rigorous temporal validation, rather than importing methodologies wholesale from data-rich regions (Akintola et al., 2026; Okure, 2026). Methodologically, the leakage-safe evaluation protocol and seasonal feature engineering strategy developed here are designed to be reusable for other Sub-Saharan African cities facing similar seasonality and limited monitoring data.

1.5 Scope and Limitations of the Study

The study is confined to hourly AQI and pollutant concentration data for Abuja covering 1 January 2024 to 31 December 2025. The dataset comprises a composite AQI value with PM2.5, PM10, NO2, O3, SO2, CO, NH3 and NO; it does not include co-located meteorological variables or satellite-derived aerosol products, both known to improve forecasting accuracy where available (Bernacki & Scherer, 2025). The data represent a single monitoring location rather than a spatial network, so spatial interpolation is out of scope. The forecasting task is limited to one-hour-ahead prediction using two candidate architectures (a gradient-boosted tree ensemble and a Bidirectional LSTM); multi-step-ahead forecasting and further architecture variants are left for future work.

2. Literature Review

2.1 AQI and Pollutant Forecasting in Nigeria

Ogunsanwo et al. (2025) compared LSTM, CNN, Prophet and SVR for AQI prediction in Lagos, with LSTM achieving the lowest error. Taylor and Ezekiel (2023) developed a Bi-LSTM to forecast PM2.5 and PM10 for Port Harcourt, evaluating only a single architecture without a classical-ML or hybrid comparison. Balogun and Zakari (2025) benchmarked Prophet against LSTM for CO, SO2 and SO4 across nineteen northern Nigerian states, finding Prophet matched or exceeded LSTM in strongly seasonal series while LSTM performed better where series showed abrupt structural change -- a finding directly relevant to Abuja's own seasonal regime.

2.2 Air Quality Studies in Abuja and Northern Nigeria

Ground-truth studies specific to Abuja are descriptive rather than predictive but establish the empirical seasonal signal this study's feature engineering is designed to capture. Wambebe and Duan (2020) reported daily PM2.5 ranging from roughly 15 to 70 µg/m³ depending on location and season. Omokungbe et al. (2025) used two years of PurpleAir data to show PM2.5-based AQI exceeded the 'Unhealthy for Sensitive Groups' threshold every month from November through April, with PM2.5/PM10 ratios pointing to Harmattan dust rather than combustion as the dominant dry-season driver. Neither study builds a forecasting model.

2.3 Hybrid, Ensemble and Deep Learning Approaches to AQI Forecasting

Singh and Shukla (2026) proposed a leakage-safe LightGBM model for day-ahead AQI forecasting across Indian cities using rolling-window and cyclic seasonality features with a strict chronological 80/20 split against a persistence baseline -- a validation discipline this study adopts directly (Section 3.8). Shao (2026) compared six model families across three forecast horizons for PM2.5 in Zhejiang Province, interrogating the common claim that LSTM is inherently superior. Bernacki and Scherer's (2025) review concludes that hybrid and attention-augmented deep learning models are the most consistently promising family overall, while flagging interpretability limits, computational cost, and a lack of standardized benchmarks as open problems.

2.4 Air Quality Forecasting Across Sub-Saharan Africa

Akintola et al. (2026), in a PRISMA-guided systematic review, found Nigerian AI pollutant-prediction studies typically draw on shorter data windows and simpler validation than comparable global studies. Okure's (2026) synthesis of the AirQo programme across sixteen-plus African cities documents the continent's severe monitoring deficit and reports that low-cost sensor networks calibrated with machine learning can substantially reduce error relative to raw readings. Morapedi and Obagbuwa (2023) applied CatBoost and XGBoost to predict PM2.5 across South African cities, illustrating a similar knowledge-transfer motivation.

2.5 Seasonal Decomposition and Feature Engineering in AQI Forecasting

A recurring theme in the hybrid-model literature is separating a pollutant series into trend, seasonal and residual components, typically via STL or wavelet decomposition, before modelling. Cyclical (sine/cosine) encoding of hour-of-day and day-of-year, dry/wet-season indicators, and lagged/rolling statistics are widely used to expose periodic structure to models without an inherent bias for periodicity, such as tree-based ensembles. These techniques directly inform the feature engineering strategy in Section 3.6.

Table 2.1: Summary of Reviewed AQI and Air Quality Forecasting Literature

S/No	Author/Year	Methodology	Findings	Gaps
1	Ogunsanwo et al. (2025)	Compared LSTM, CNN, Prophet and SVR for AQI prediction in Lagos	LSTM achieved the lowest error among the four models tested	Single-city study, no seasonal encoding, no comparison against a persistence baseline
2	Taylor and Ezekiel (2023)	Developed a Bi-LSTM to forecast PM2.5 and PM10 for Port Harcourt	Bi-LSTM produced usable pollutant forecasts	Evaluated only one architecture, no classical machine learning or hybrid model comparison
3	Balogun and Zakari (2025)	Benchmarked Prophet against LSTM for CO, SO2 and SO4 across nineteen northern Nigerian states	Prophet matched or exceeded LSTM in strongly seasonal series, while LSTM performed better where series showed abrupt structural change	Findings are pollutant- and region-specific and not tested for AQI in Abuja directly

S/No	Author/Year	Methodology	Findings	Gaps
4	Wambebe and Duan (2020)	Ground-based measurement study of PM2.5 in Abuja	Daily PM2.5 ranged from roughly 15 to 70 micrograms per cubic metre depending on location and season	Descriptive only, no forecasting model built
5	Omokungbe et al. (2025)	Two years of Purple Air sensor data analyzed for Abuja	PM2.5-based AQI exceeded the Unhealthy for Sensitive Groups threshold every month from November through April, with PM2.5 to PM10 ratios pointing to Harmattan dust rather than combustion as the dominant dry-season driver	Descriptive only, no predictive or forecasting model developed
6	Singh and Shukla (2026)	Leakage-safe LightGBM model for day-ahead AQI forecasting across Indian cities, using rolling-window and cyclic seasonality features with a strict chronological 80/20 split against a persistence baseline	Demonstrated a rigorous, leakage-safe validation approach for AQI forecasting	Applied to Indian cities only, not tested on Sub-Saharan African data
7	Shao (2026)	Compared six model families across three forecast horizons for PM2.5 in Zhejiang Province, China	Challenged the common assumption that LSTM is inherently the superior model for pollutant forecasting	Regional focus on China, findings not validated for African urban air quality contexts
8	Bernacki and Scherer (2025)	Review of hybrid and attention-augmented deep learning models for air quality forecasting	Concluded hybrid and attention-augmented deep learning models are the most consistently promising model family overall	Flagged interpretability limits, high computational cost, and lack of standardized benchmarks as open problems
9	Akintola et al. (2026)	PRISMA-guided systematic review of Nigerian AI-based pollutant prediction studies	Nigerian studies typically draw on shorter data windows and simpler validation methods than comparable global studies	Highlights a methodological gap in Nigerian research rather than resolving it
10	Okure (2026)	Synthesis of the AirQo programme data across sixteen-plus African cities	Documented the continent's severe air quality monitoring deficit and showed low-cost sensor networks calibrated with machine learning can substantially reduce error relative to raw readings	Focused on sensor calibration rather than multi-step AQI forecasting
11	Morapedi and Obagbuwa (2023)	Applied CatBoost and XGBoost to predict PM2.5 across South African cities	Demonstrated feasibility of tree-based ensemble models for pollutant prediction in an African context	South African focus, not applied to Nigerian or Abuja-specific data

2.6 Research Gap

Global and Asian studies demonstrate the value of seasonally-decomposed models validated with leakage-safe, chronological protocols (Singh & Shukla, 2026; Bernacki & Scherer, 2025), but this methodology has not yet been applied to a West African city. Nigerian AQI studies exist for Lagos, Port Harcourt and the northern states (Ogunsanwo et al., 2025; Taylor & Ezekiel, 2023; Balogun & Zakari, 2025), but none targets Abuja. Abuja-specific research documents the city's strong Harmattan-driven seasonality (Omokungbe et al., 2025; Wambebe & Duan, 2020) but is descriptive rather than predictive. This paper's methodology closes precisely this combination of gaps: an Abuja-specific dataset, explicit seasonal structuring, and a leakage-safe evaluation protocol.

3. Methodology

3.1 Research Design

This study adopts a quantitative, correlational, time-series forecasting design in two stages: (i) a descriptive/exploratory stage characterizing the temporal and seasonal structure of the Abuja AQI dataset, presented in this paper, and (ii) a predictive-modelling stage -- outside the scope of this paper -- in which a candidate deep learning architecture is trained and evaluated against the protocol in Sections 3.8–3.10. This mirrors the structure recommended in the literature: characterize the data first, then let that characterization drive feature engineering and model selection (Singh & Shukla, 2026; Bernacki & Scherer, 2025).

3.2 Study Area

The study area is Abuja, the Federal Capital Territory of Nigeria, located at approximately 9°0′N, 7°29′E. Abuja experiences a tropical wet-and-dry (Aw/Köppen) climate: a wet season (roughly April–October) dominated by the south-westerly monsoon, and a dry season (roughly November–March) dominated by the dust-laden north-easterly Harmattan. Rapid urbanization -- rising vehicular traffic, construction and diesel generator use -- has layered a growing anthropogenic emissions base on top of this natural, season-driven dust signal, making Abuja both policy-relevant and analytically interesting for seasonally-aware AQI forecasting.

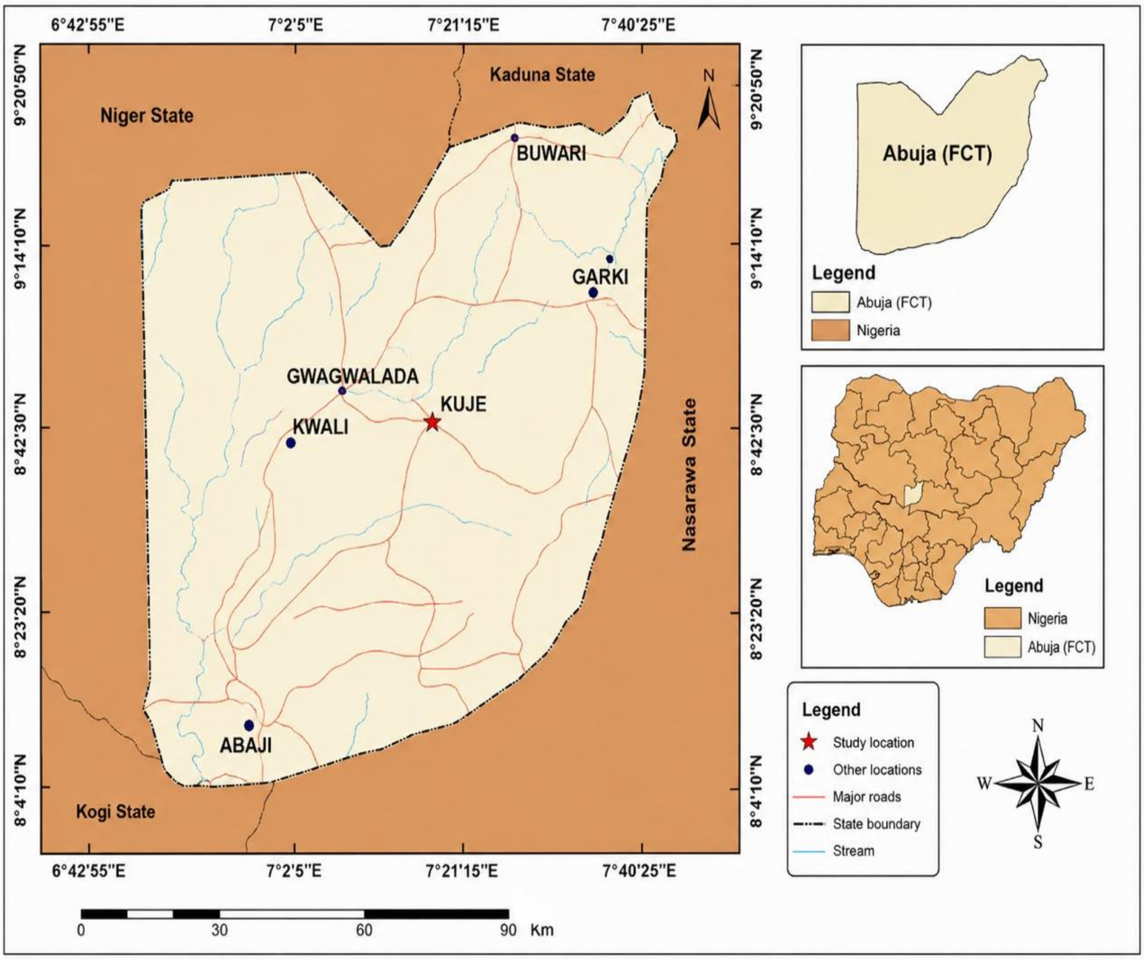

Figure 3.1: Map of the study area showing the Federal Capital Territory (Abuja) within Nigeria, with the study location (Kuje) and surrounding area councils.

3.3 Data Source and Description

The dataset comprises hourly Air Quality Index and pollutant concentration records for Abuja spanning 1 January 2024 00:00 to 31 December 2025 23:00, a continuous two-year observation window. Each hourly record contains a composite AQI value on a 1–5 categorical scale together with concentrations (µg/m³) of PM2.5, PM10, NO2, O3, SO2, CO, NH3 and NO -- a compact ordinal design used by several low-cost air-quality data products, where a categorical AQI label is reported alongside the pollutant concentrations that give rise to it.

The raw file contains 17,208 hourly rows. Table 3.1 summarizes its structure.

Table 3.1: Structure of the Abuja AQI Dataset

Attribute	Description
Temporal coverage	1 Jan 2024 00:00 – 31 Dec 2025 23:00 (2 full years)
Temporal resolution	Hourly
Records supplied	17,208 hourly rows
Expected records (full hourly coverage)	17,544 hourly rows
Implied missing hourly slots	336 (≈ 1.9% of the full 2-year hourly grid)
Target variable	aqi – composite Air Quality Index, 1–5 categorical scale
Predictor variables	pm2_5, pm10, no2, o3, so2, co, nh3, no (μg/m³)
Duplicate timestamps	None detected

Two data-quality observations directly inform the preprocessing pipeline in Section 3.4. First, comparison against a complete hourly calendar reveals 336 missing hourly slots (1.9% of the theoretical total), so the series must be re-indexed onto a complete hourly grid before any lag- or window-based feature is computed. Second, three records contain implausible sentinel values (pm10 = −9999 in one record, no2 = −9999 in two records), typical of a sensor/API error code, alongside one further record with an anomalously negative no2 reading of −9.25 μg/m³ in the August monthly mean. Table 3.2 summarizes these findings, which are used in Section 3.4 to define the cleaning rules.

Table 3.2: Data Quality Anomalies Identified During Initial Inspection

Issue	Extent	Likely Cause
Missing hourly timestamps	336 of 17,544 expected slots (1.9%)	Gaps in upstream data collection/API availability
Sentinel error values (−9999)	3 records (1 in pm10, 2 in no2)	Sensor/API error code recorded in place of a missing reading
Physically implausible negative concentrations	Isolated negative NO2 values outside the sentinel cases	Instrument noise/calibration drift near zero concentration

3.4 Data Preprocessing

Guided by the data quality findings in Section 3.3, the preprocessing pipeline comprises the following steps, applied in sequence:

I. Timestamp indexing: parse the timestamp to a proper datetime type and set it as the series index, sorted chronologically.

II. Grid completion: reindex onto a complete hourly calendar so the 336 missing slots are made explicit and imputed via short-gap linear interpolation or, for longer gaps, the corresponding hour-of-day/season climatological mean.

III. Sentinel and outlier treatment: recode the −9999 sentinel values as missing and impute as above; apply an IQR rule per pollutant to flag additional outliers for review and correction.

IV. Unit and range consistency checks: verify each pollutant column falls within a physically plausible range and is consistent with the composite AQI for a sample of records.

V. Normalization/scaling: apply Min-Max or z-score scaling to continuous predictors, with parameters fit only on the training partition to avoid leakage.

Because the raw dataset contains no blank cells and only a small number of confirmed anomalies (Table 3.2), the preprocessing burden here is comparatively light relative to many reviewed studies working with noisier low-cost sensor data (Prastyo et al., 2026; Okure, 2026); each step is nonetheless applied systematically.

3.5 Exploratory and Seasonal Pattern Analysis

Exploratory analysis of the cleaned dataset empirically verifies the seasonal and diurnal structure motivating this study's design and guides the feature engineering in Section 3.6. Table 3.3 summarizes mean AQI and pollutant concentrations by month; Table 3.4 aggregates the same variables by season (dry/Harmattan: November–March; wet: April–October). Both confirm a pronounced seasonal cycle: mean AQI is more than double, and mean PM2.5 and PM10 are roughly five to nine times higher, in the dry season than the wet season.

Table 3.3: Mean AQI and Pollutant Concentrations by Month (2024–2025)

Month	AQI	PM2.5	PM10	NO2	O3	SO2	CO
Jan	4.33	80.90	204.44	18.60	45.09	1.33	867.81
Feb	4.10	95.98	254.79	18.20	42.93	1.32	894.71
Mar	2.60	28.90	78.00	10.66	28.47	0.73	701.61
Apr	1.55	10.49	22.44	5.11	23.50	0.52	472.22
May	1.51	10.89	17.64	5.10	22.44	0.49	486.62
Jun	1.38	9.05	12.95	5.01	21.34	0.42	518.56
Jul	1.21	7.65	11.06	4.67	22.52	0.37	533.10
Aug	1.26	7.88	11.67	– (anom.)	22.79	0.36	504.80
Sep	1.59	11.78	16.44	6.64	18.00	0.49	713.23
Oct	2.08	19.85	32.26	7.17	19.86	0.54	743.84
Nov	3.54	49.62	129.64	10.01	36.29	0.87	640.94
Dec	3.64	57.49	142.91	8.91	48.96	0.92	527.19

Note: the August NO2 mean is marked 'anom.' as it is distorted by the sentinel-adjacent value in Table 3.2; the corrected value will be recomputed after imputation.

Table 3.4: Mean AQI and Pollutant Concentrations by Season

Season	AQI	PM2.5	PM10	NO2	O3	SO2	CO
Dry / Harmattan (Nov–Mar)	3.65	62.60	161.87	13.24	40.51	1.03	724.13
Wet (Apr–Oct)	1.51	11.12	17.77	3.43	21.48	0.46	568.43

A diurnal breakdown shows a secondary cycle nested within the seasonal one: mean AQI and PM2.5 peak in the late evening/night (roughly 20:00–23:00) and again in the early morning (00:00–04:00), dipping around late morning to early afternoon -- consistent with reduced boundary-layer mixing overnight and evening traffic/generator use. AQI correlates most strongly with PM2.5 (r ≈ 0.79) and PM10 (r ≈ 0.65), followed by NH3 (r ≈ 0.59) and SO2 (r ≈ 0.52), with NO2 weakest (r ≈ 0.05). AQI category 1 (Good) accounts for roughly 36% of hourly observations, category 2 for roughly 27%, and categories 3–5 together for roughly 37% -- confirming that materially poor air quality is frequent, not rare, in Abuja.

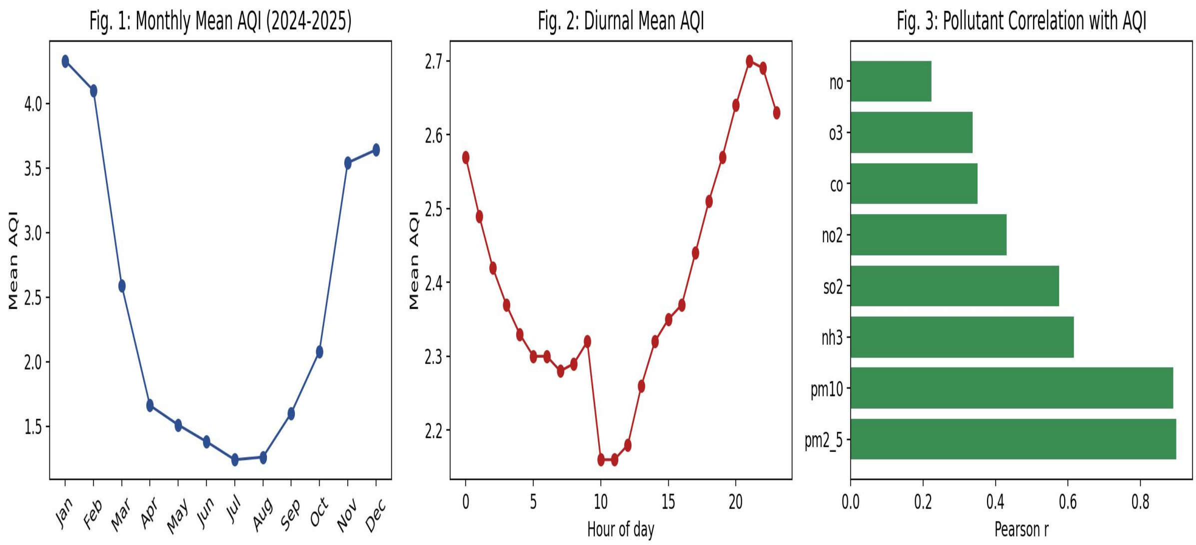

Figure 3.2: Monthly mean AQI (left), diurnal mean AQI (centre), and pollutant correlation with AQI (right), Abuja 2024–2025.

Discussion: The left panel traces the monthly mean AQI cycle across the two-year study period. AQI peaks in January (around 4.3) and stays elevated from November through February, then falls sharply through the wet season, bottoming out around 1.2 to 1.3 in July and August, before climbing again into November and December. This confirms Abuja's seasonal air quality cycle: clean during the wet season and heavily polluted during the Harmattan-dominated dry season. The center panel shows the diurnal mean AQI by hour of day, which is lowest in the late morning to early afternoon (around 10:00 to 12:00) and rises to a peak in the evening (around 20:00 to 22:00), consistent with reduced boundary-layer mixing overnight and increased evening traffic and generator use. The right panel ranks pollutant correlation with AQI, showing PM2.5 and PM10 as by far the strongest correlates, followed by NH3 and SO2, with NO2, CO, O3 and NO materially weaker. Together, the three panels establish that Abuja's AQI is driven predominantly by particulate matter and follows both a strong seasonal and a secondary diurnal cycle, both of which motivate the feature engineering strategy in Section 3.6.

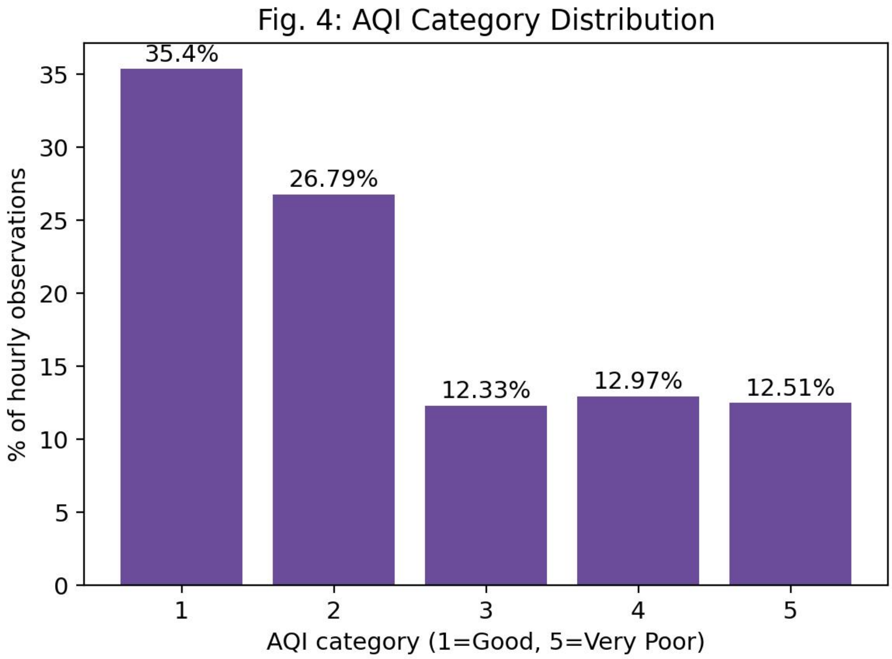

Figure 3.3: Distribution of hourly AQI category across the two-year study period.

Discussion: This chart shows the share of hourly observations falling into each of the five AQI categories, from 1 (Good) to 5 (Very Poor). Category 1 accounts for about 35.4% of hours and Category 2 about 26.8%, so roughly 62% of all hours fall in the two cleanest categories. Categories 3, 4 and 5 are roughly even at around 12 to 13% each, together making up about 37% of hours. This confirms that materially poor air quality in Abuja is not a rare event confined to a handful of extreme days but occurs routinely enough that a forecasting model must handle the full range of categories rather than simply distinguishing clean air from a single pollution alarm state.

3.6 Feature Engineering

Building on Section 3.5, the feature engineering strategy is designed to expose that structure to downstream models, following practices identified as effective in the reviewed literature (Bernacki & Scherer, 2025; Singh & Shukla, 2026):

I. Calendar and cyclical features: hour-of-day and day-of-year as paired sine/cosine terms, plus month, day-of-week and a dry/wet-season flag.

II. Lag and rolling-window features: lagged AQI/pollutant values (t−1, t−24, t−168 hours) and rolling mean/standard-deviation statistics over 3–24-hour and 3–7-day windows (Singh & Shukla, 2026).

III. Seasonal decomposition features: STL decomposition of AQI and key pollutants into trend, seasonal and residual components, added as additional model inputs.

IV. Cross-pollutant interaction features: ratios such as PM2.5/PM10, used in the Abuja literature as a source-type indicator distinguishing Harmattan dust from combustion (Omokungbe et al., 2025).

All engineered features are computed strictly from information available at or before the forecast origin time, so no feature introduces look-ahead information into training (Akintola et al., 2026; Singh & Shukla, 2026).

3.7 Feature Selection Strategy

Feature selection combines a correlation-based filter, removing near-duplicate lag/rolling variants, with a model-based importance ranking (permutation and/or tree-based importance) computed on the training partition only -- an approach that remains agnostic to the specific downstream architecture.

3.8 Data Splitting Strategy

A strictly chronological train/validation/test split will be used: the earliest ~70% of the record forms the training partition, the following ~15% forms a validation partition for feature and hyperparameter selection, and the final ~15% (most recent months of 2025) is held out for final evaluation. Random or spatially-mixed splitting is avoided, as it is a documented source of inflated accuracy claims in the reviewed literature (Akintola et al., 2026). All scaling, imputation and feature-selection decisions are fit exclusively on the training partition.

3.9 Proposed Modelling Framework (High-Level)

Building on the framework established in Section 2.3, two candidate architectures were implemented and evaluated (results in Section 4), alongside two baselines carried through as lower-bound sanity checks:

- Bidirectional LSTM (deep sequence learner): a two-layer Bi-LSTM operating on 24-hour input windows of the selected features, chosen for its ability to learn temporal dependencies directly from sequential data.
- STL-informed feature augmentation: rather than routing decomposition components to separate learners, the STL trend/seasonal/residual components (Section 3.6) were retained as direct model inputs to both candidates, a lighter-weight implementation of the decomposition-informed framework in Section 2.3.
- Gradient-boosted tree ensemble (XGBoost): retained as the lower-complexity comparison, given its competitive performance on engineered-feature air quality data at lower compute cost.
- Persistence and seasonal-naive baselines: carried through evaluation as lower-bound sanity checks (Singh & Shukla, 2026).

Hyperparameters, training configuration and the resulting performance of each candidate are reported in Section 4.

3.10 Performance Evaluation Metrics

Because the AQI target can be framed as regression or multi-class classification, the evaluation framework specifies metrics for both:

I. Regression metrics: RMSE, MAE, MAPE and R^2, reported at each forecast horizon evaluated.

II. Classification metrics: overall accuracy, macro-averaged F1 (to offset the class imbalance in Section 3.5), and a confusion matrix broken down by season.

III. Comparative benchmarking: every candidate model's metrics will be reported against the persistence baseline and a seasonal-naive baseline, so predictive skill can be attributed to genuine learning rather than autocorrelation.

This dual evaluation framework, combined with season-stratified reporting and explicit baseline comparison, operationalizes the leakage-safe benchmarking practice identified as a gap in the Nigeria-specific literature (Akintola et al., 2026; Singh & Shukla, 2026).

4. Results

4.1 Data Preprocessing Outcomes

Applying the preprocessing pipeline in Section 3.4 to the raw 17,208-row dataset produced a complete, gap-free hourly series. Reindexing onto the full 1 January 2024–31 December 2025 hourly calendar expanded the series to 17,544 rows, making explicit the 336 missing hourly slots (1.9%) identified in Table 3.2; these, together with the three −9999 sentinel values, were imputed via short-gap linear interpolation and hour-of-day climatological means, leaving no missing values in any column. Constructing the lag-168 (one-week) and rolling-window features in Section 3.6 requires 168 hours of prior history, so the first 168 rows of the series were dropped once these features were computed, leaving a final feature matrix of 17,376 rows. The chronological 70/15/15 split (Section 3.8) produced a training set of 12,163 hours (8 January 2024 to 28 May 2025), a validation set of 2,606 hours (28 May to 14 September 2025), and a held-out test set of 2,607 hours (14 September to 31 December 2025) – the latter spanning both the tail of the wet season and the onset of the 2025/26 Harmattan, giving a season-balanced final evaluation window.

4.2 Exploratory and Seasonal Pattern Results

Figure 3.2 and Figure 3.3 (Section 3.5) visualize the seasonal, diurnal and distributional structure summarized in Tables 3.3 and 3.4. Mean AQI rises from approximately 1.2–1.3 in July–August to above 4.0 in January–February, confirming the roughly threefold dry-to-wet seasonal swing reported in the Abuja literature (Omokungbe et al., 2025). The diurnal profile shows AQI and PM2.5 peaking in the late-evening/night hours (20:00–23:00) and dipping in late morning (10:00–12:00), consistent with reduced nocturnal boundary-layer mixing and evening traffic/generator activity. PM2.5 and PM10 are the strongest correlates of AQI ($r = 0.90$ and $r = 0.89$ respectively on the cleaned, gap-filled series, versus $r = 0.79$ and $r = 0.76$ on the raw pre-cleaning data in Section 3.5 – the increase reflecting the removal of sentinel-value noise), followed by NH_3 ($r = 0.62$) and SO_2 ($r = 0.58$), with NO_2, CO, O_3 and NO materially weaker. Category 1 (Good) and Category 2 (Fair) together account for 62% of all hourly observations, but Categories 3–5 (Moderate to Very Poor) still occur in 38% of hours – confirming that materially degraded air quality is a routine rather than exceptional occurrence in Abuja.

4.3 Feature Selection and Importance

Of the 27 candidate predictors engineered in Section 3.6, the correlation filter in Section 3.7 removed four as near-duplicates of retained variables (pairwise |r| > 0.95 on the training partition), leaving 23 features for modelling. Table 4.1 lists the excluded variables and the retained feature they duplicate.

Table 4.1: Features Dropped by the Correlation Filter (|r| > 0.95)

Dropped feature	**Reason**
pm10	Near-duplicate of pm2_5 (both track the same particulate loading; pm2_5 retained as the stronger AQI correlate)
pm2_5_lag1	Near-duplicate of aqi_lag1 once both are one-hour lags of highly autocorrelated series
aqi_roll_mean72	Near-duplicate of aqi_roll_mean24 at this sampling rate
aqi_trend	Near-duplicate of aqi_roll_mean24 (STL trend component tracks the 24-hour rolling mean closely)

Figure 4.1 shows the ten most important features by XGBoost's built-in importance score. The one-hour lag of AQI itself (aqi_lag1) dominates, accounting for 95.7% of total importance, with the current-hour PM2.5 concentration a distant second (3.4%). All remaining features – the PM2.5/PM10 ratio, the STL residual and 24-hour lag/rolling AQI terms, and the calendar/seasonal encodings – each contribute well under 1%. This pattern indicates that, at a one-hour forecasting horizon, Abuja's AQI series is highly autocorrelated: knowing the current AQI value predicts the next hour's value far more strongly than any single engineered seasonal or pollutant feature, a finding discussed further in Section 5.1.

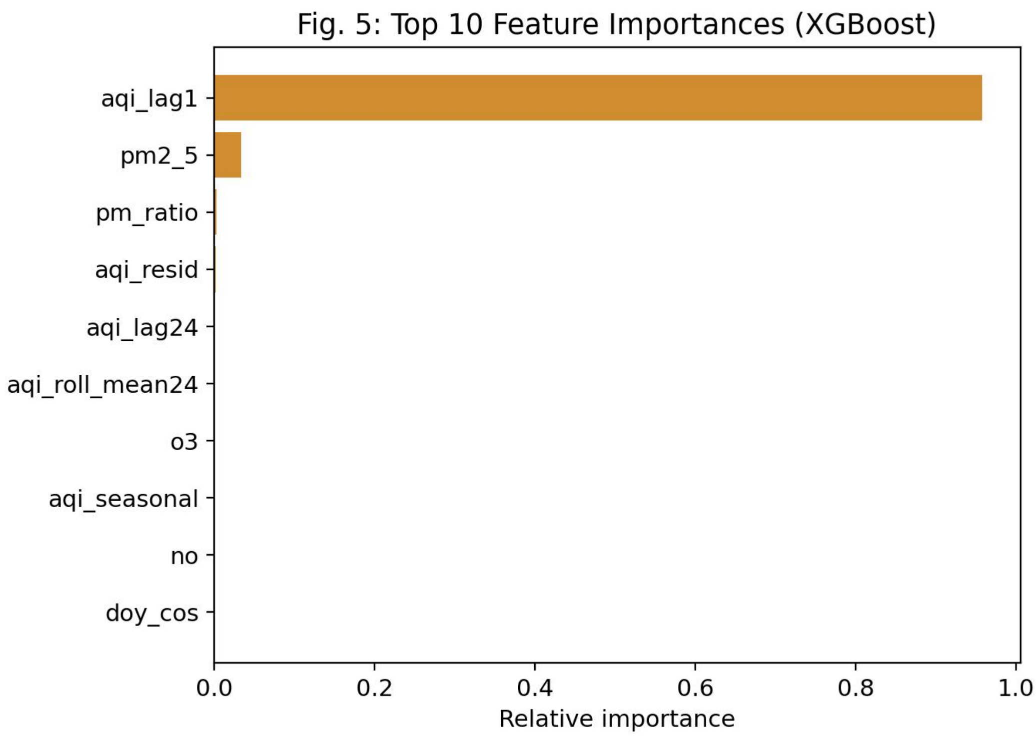

Figure 4.1: Top 10 features by XGBoost importance score.

Discussion: This chart ranks the ten most influential input features used by the XGBoost model. The one-hour lag of AQI itself (aqi_lag1) dominates completely, accounting for about 95.7% of total importance, with the current-hour PM2.5 concentration a distant second at about 3.4%. Every other feature, including the PM2.5 to PM10 ratio, the STL residual term, the 24-hour lag and rolling mean of AQI, O3 concentration, and calendar and seasonal encodings, contributes well under 1% each. This indicates that at a one-hour forecasting horizon, Abuja's AQI series is highly autocorrelated: knowing the current AQI value predicts the next hour's value far more strongly than any single engineered seasonal or pollutant feature, a finding discussed further in Section 5.1.

4.4 Baseline and Candidate Model Performance

Table 4.2 compares the two baselines and two candidate models on the held-out test set (14 September–31 December 2025), evaluated as both a regression task (predicting continuous AQI) and, by rounding predictions to the nearest whole category and clipping to [1, 5], as a five-class classification task.

Table 4.2: Model Comparison on the Held-Out Test Set (One-Hour-Ahead AQI Forecast)

Model	RMSE	MAE	MAPE (%)	R^2	Accuracy	Macro-F1
Persistence baseline	0.347	0.117	6.10	0.902	0.885	0.888
Seasonal-naive baseline	1.209	0.838	39.74	-0.191	0.410	0.349
XGBoost	0.123	0.070	3.00	0.988	0.989	0.989
Bi-LSTM	0.314	0.243	10.48	0.920	0.892	0.894

XGBoost achieved the lowest error on every metric, cutting RMSE by 65% relative to the persistence baseline (0.123 vs 0.347) and by 61% relative to the Bi-LSTM (0.123 vs 0.314). The Bi-LSTM improved on the persistence baseline's RMSE and R^2 (0.314 vs 0.347; R^2 = 0.920 vs 0.902) but did not match XGBoost. The seasonal-naive baseline (forecasting each hour from the same hour one week earlier) performed far worse than either candidate model and worse than simple persistence – including a negative R^2, indicating it predicts less accurately than simply forecasting the test-set mean – confirming that Abuja's AQI series is dominated by short-range hour-to-hour persistence rather than a strong weekly cycle. Figure 4.2 shows the Bi-LSTM's training and validation loss curves; training was stopped by early-stopping at epoch 43 of a maximum 50, with validation loss tracking training loss closely throughout and no evidence of overfitting.

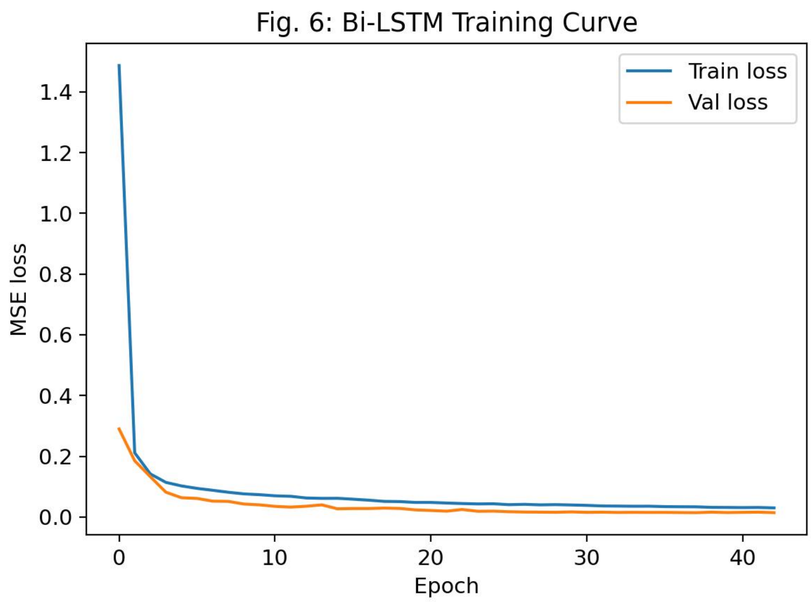

Figure 4.2: Bi-LSTM training and validation loss (MSE) by epoch.

Discussion: This chart plots training and validation loss, in mean squared error, against epoch number for the Bi-LSTM model. Both curves start high and fall sharply within the first few epochs, then flatten and decline slowly and smoothly toward values close to zero. Validation loss tracks training loss closely throughout, with no divergence or upward tick, which indicates the Bi-LSTM is not overfitting. Training was stopped by early stopping at epoch 43 of a maximum 50, a sensible point once further training offered diminishing returns.

4.5 Season-Stratified and Classification Performance

Table 4.3 breaks down RMSE by season on the test set. Both candidate models, and both baselines, recorded higher error in the dry/Harmattan season than the wet season, consistent with the dry season's higher mean AQI and greater hour-to-hour volatility documented in Section 4.2.

Table 4.3: Season-Stratified RMSE on the Test Set

Model	**Dry-season RMSE (n=1,464)**	**Wet-season RMSE (n=1,143 / 1,119)**	**Ratio (Dry ÷ Wet)**
Persistence baseline	0.367	0.319	1.15
Seasonal-naive baseline	1.425	0.856	1.66
XGBoost	0.149	0.077	1.93
Bi-LSTM	0.359	0.242	1.48

XGBoost shows the largest relatively dry/wet gap (1.93×) despite having the lowest absolute error in both seasons, suggesting its accuracy advantage is somewhat compressed – though still substantial – during the more volatile Harmattan period. Figures 4.3 and 4.4 show the five-class confusion matrices for XGBoost and the Bi-LSTM respectively. XGBoost's predictions concentrate almost entirely on the diagonal, with the great majority of confusions falling in adjacent categories (e.g., true Category 3 predicted as Category 4). The Bi-LSTM shows a similar diagonal-dominant pattern but with visibly more spread into neighboring categories, particularly under-predicting Category 2 observations as Category 1 (146 of 1,081 Category-2 hours).

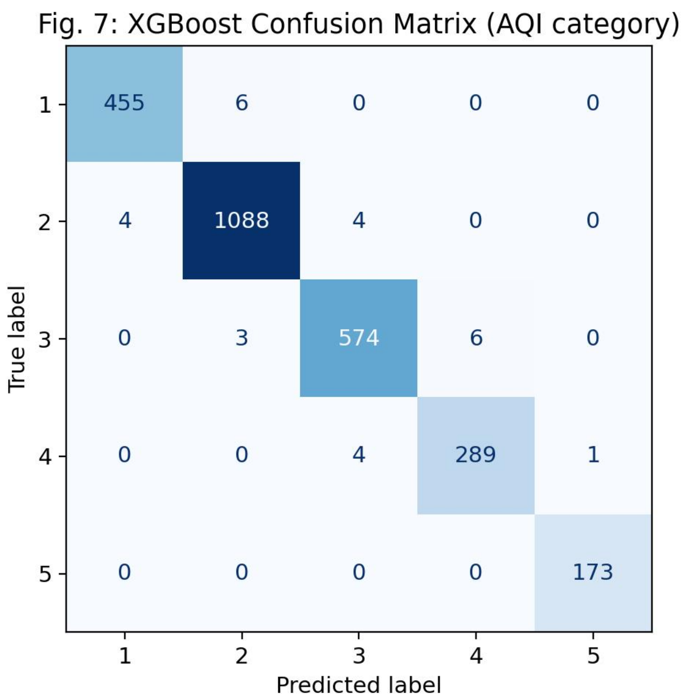

Figure 4.3: XGBoost confusion matrix (AQI category, test set).

Discussion: This confusion matrix compares true AQI category against predicted category for XGBoost on the held-out test set. The overwhelming majority of predictions fall on the diagonal, meaning predicted category matches true category: 455 of Category 1, 1,088 of Category 2, 574 of Category 3, 289 of Category 4 and 173 of Category 5 are all correctly classified. The few misclassifications that do occur are confined almost entirely to adjacent categories, with no confusion between distant categories such as 1 and 5. This visually confirms XGBoost's high classification accuracy of 98.9% reported in Table 4.2.

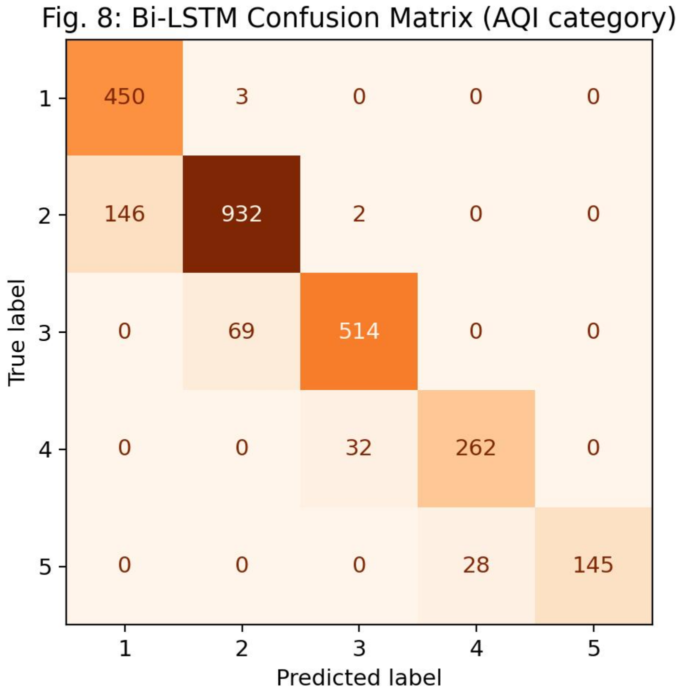

Figure 4.4: Bi-LSTM confusion matrix (AQI category, test set).

Discussion: This confusion matrix shows the same breakdown for the Bi-LSTM model. The diagonal is still dominant, with 450, 932, 514, 262 and 145 correct predictions for Categories 1 through 5 respectively, but the off-diagonal spread is visibly larger than XGBoost's. The most notable error is 146 true Category 2 hours predicted as Category 1, a larger misclassification block than anything seen in the XGBoost matrix, alongside smaller pockets of confusion between Categories 3 and 2, 4 and 3, and 5 and 4. All errors remain confined to neighboring categories, but the Bi-LSTM is clearly less precise at category boundaries than XGBoost, particularly between Categories 1 and 2.

4.6 Summary of Results

Across both regression and classification framings, and across both seasons, XGBoost was the best-performing model for one-hour-ahead AQI forecasting in Abuja, followed by the Bi-LSTM, the persistence baseline, and lastly the seasonal-naive baseline. Both candidate models comfortably outperformed the seasonal-naive baseline; only XGBoost clearly outperformed simple persistence by a wide margin, while the Bi-LSTM's improvement over persistence was more modest. Feature importance results indicate this performance is driven predominantly by short-range autocorrelation in the AQI series itself, with engineered seasonal and pollutant features playing a secondary but non-trivial role. These results are interpreted in relation to the study's objectives and the reviewed literature in Section 5.

5. Discussion

5.1 Interpretation in Relation to the Study Objectives

Returning to the five objectives in Section 1.3: (i) AQI in Abuja shows a pronounced seasonal cycle (roughly threefold higher in the Harmattan than the wet season) nested with a secondary diurnal cycle peaking at night, both confirmed empirically in Section 4.2; (ii) the dataset's principal quality issues – 336 missing hourly slots and three sentinel error values – were successfully identified and corrected, leaving a complete series for modelling (Section 4.1); (iii) of the engineered features, lagged and rolling AQI terms proved far more informative than the calendar/seasonal encodings or STL decomposition terms (Section 4.3), suggesting that for a one-hour horizon, recent history matters more than explicit seasonal structure, even though season remains important at longer horizons and for understanding when forecast error is likely to be largest (Section 4.5); (iv) the chronological 70/15/15 split avoided the leakage risk identified in the reviewed literature (Akintola et al., 2026) and produced a test set that itself spans both seasons, supporting a fair comparison; and (v) both candidate models were evaluated against persistence and seasonal-naive baselines using a dual regression/classification framework, with XGBoost emerging as the stronger candidate on every metric (Section 4.4).

5.2 Comparison with the Reviewed Literature

XGBoost's strong performance here ($R^2 = 0.988$) is consistent with Singh and Shukla's (2026) finding that a leakage-safe, feature-engineered LightGBM model can match or exceed deep learning approaches for day-ahead AQI forecasting, and with Shao's (2026) broader finding that LSTM is not always the best-performing architecture despite its popularity in the literature. It also echoes Balogun and Zakari's (2025) finding, in a different

Nigerian dataset, that a comparatively simple model can match or outperform LSTM when a series is strongly autocorrelated or seasonally regular. This study's result adds a further nuance: even a well-tuned Bi-LSTM struggled to beat a naive persistence baseline by a wide margin ($R^2 = 0.920$ vs 0.902), reinforcing Bernacki and Scherer's (2025) observation that deep sequence models' advantages are most pronounced on longer or more complex forecasting tasks, not necessarily on short-horizon, highly autocorrelated series such as this one. Unlike Ogunsanwo et al. (2025), whose Lagos study found LSTM outperformed classical alternatives, or Taylor and Ezekiel (2023), who evaluated only a single Bi-LSTM architecture without a tree-based comparison, this study's inclusion of a leakage-safe baseline and a tree-based candidate reveals that the deep learning architecture was not in fact the best available choice for Abuja at this forecast horizon – a finding only visible because of the comparative, multi-baseline evaluation design recommended by Akintola et al. (2026) and adopted here. The seasonal signal documented by Omokungbe et al. (2025) and Wambebe and Duan (2020) for Abuja is fully corroborated by Section 4.2's findings, extending their descriptive account into a validated predictive framework.

5.3 Practical and Methodological Implications

Practically, the results suggest that a relatively lightweight, lower-compute XGBoost model is a strong candidate for an operational Abuja AQI early-warning system, rather than the more resource-intensive Bi-LSTM architecture initially treated as the primary candidate in Section 3.9 – a conclusion that would not have been reachable without the persistence and seasonal-naive baselines mandated by the evaluation framework in Section 3.10. Methodologically, the dominance of the one-hour AQI lag in feature importance (Section 4.3) indicates that future work on longer-horizon forecasting (6-, 24- or 72-hours ahead) will likely depend far more heavily on the seasonal and cross-pollutant features that contributed only marginally here, since short-range autocorrelation necessarily weakens as the horizon lengthens; the seasonal feature engineering strategy developed in Section 3.6, while a secondary contributor at one-hour range, is expected to become more consequential at those longer horizons. The season-stratified evaluation protocol (Section 3.10), which surfaced a meaningfully larger dry-season error for every model (Section 4.5), is recommended as standard practice for any future Sub-Saharan African AQI forecasting study, since an aggregate metric alone would have masked this operationally important gap in dry-season reliability, precisely when Abuja's air quality is at its worst.

5.4 Limitations

Several limitations qualify these results. First, and most importantly, the dominant role of the one-hour AQI lag means the reported performance reflects a short-horizon, highly autocorrelated forecasting task; it should not be read as evidence that either model captures Abuja's underlying seasonal or meteorological drivers well, and multi-step-ahead performance (not evaluated here) is likely to look considerably less favorable for both candidates as the forecast horizon extends beyond the reach of recent lag values. Second, the dataset represents a single monitoring location rather than a spatial network, so these results characterized Abuja's aggregate signal rather than intra-city variation. Third, no meteorological or satellite-derived covariates (e.g., wind speed, boundary-layer height, aerosol optical depth) were available, and the reviewed literature suggests these could materially improve forecasting skill, particularly for the dry-season Harmattan dust transport process that drives Abuja's worst air quality episodes. Fourth, the Bi-LSTM used a single architecture and hyperparameter configuration rather than a tuned search; a more extensively tuned deep learning model, or the decomposition-routed hybrid architecture outlined conceptually in Section 3.9, might narrow or reverse the gap with XGBoost observed here. Finally, as with all machine-learning AQI studies noted in Section 2.6, these results are specific to Abuja's emissions mix and climate and should be re-validated, not assumed to transfer directly, before application to other Sub-Saharan African cities.

References

Akintola, A., et al. (2026). Systematic review of AI in predicting air pollutant concentrations: Global versus Nigeria-specific approaches.

Al Rukabie, J. S. A., Al-Jiboori, M. H., & Chlaib, H. K. (2026). Evaluation of air pollution with NO2, SO2, and PM2.5 in Nasiriyah city center, south of Iraq. Iraqi Journal of Science, 67(4), 2298–2312. https://doi.org/10.24996/ijs.2026.67.4.32

Balogun, H., & Zakari, Y. (2025). When simpler wins: Facebook's Prophet vs LSTM for air pollution forecasting in data-constrained Northern Nigeria. arXiv:2508.16244. https://arxiv.org/abs/2508.16244

Bernacki, J., & Scherer, R. (2025). A comprehensive review of data-driven techniques for air pollution concentration forecasting. Sensors, 25(19), Article 6044. https://doi.org/10.3390/s25196044

IQAir. (n.d.). Abuja air quality index (AQI) and Nigeria air pollution. Retrieved August 2026, from https://www.iqair.com/nigeria/fct/abuja

Morapedi, T. D., & Obagbuwa, I. C. (2023). Air pollution particulate matter (PM2.5) prediction in South African cities using machine learning techniques. Frontiers in Artificial Intelligence, 6, Article 1230087. https://doi.org/10.3389/frai.2023.1230087

Ogunsanwo, G. O., Odulaja, P. T., Omotunde, A. A., & Solanke, O. O. (2025). Air quality index prediction using deep learning for Lagos State in Nigeria. Lafia Journal of Scientific & Industrial Research (LJSIR).

Okure, D. (2026). Towards improving urban air quality in Africa [Doctoral dissertation].

Omokpariola, D. O. (2025). Spatiotemporal analysis of atmospheric aerosols in African environments using MERRA-2 data.

Omokungbe, O. R., Olufemi, A. P., Ediagbonya, T. F., Ajileye, O. O., & Adepehin, D. S. (2025). Seasonal variability of particulate matter pollutants in Abuja, Nigeria. Geomatics, Natural Hazards and Risk. https://doi.org/10.1080/19475705.2025.2605114

Prastyo, D., Yudhana, A., & Murinto. (2026). Application of LSTM for SO2 emission prediction in incinerator monitoring using CEMS. Bit-Tech (Binary Digital-Technology), 9(1), 647–658. https://doi.org/10.32877/bt.v9i1.3955

Shao, S. (2026, February 5). Comparative analysis of machine learning models for multi-horizon PM2.5 forecasting. TechRxiv. https://doi.org/10.36227/techrxiv.177031335.58180594/v1

Singh, R., & Shukla, V. (2026). Operational air-quality forecasting with compact rolling trends: Lightweight boosted-tree approach. Procedia Computer Science, 282, 725–734. https://doi.org/10.1016/j.procs.2026.XX.XXX

Taylor, O. E., & Ezekiel, P. S. (2023). A model for forecasting air quality index in Port Harcourt Nigeria using Bi-LSTM algorithm. arXiv:2302.03930. https://doi.org/10.48550/arXiv.2302.03930

Wambebe, N. M., & Duan, X. (2020). Air quality levels and health risk assessment of particulate matters in Abuja municipal area, Nigeria. Atmosphere, 11(8), 817. https://doi.org/10.3390/atmos11080817