Spectrum of Engineering Sciences
ISSN (e) 3007-3138 (p) 3007-312X Volume 4, Issue 5, 2026

XAI-AIRNET: EXPLAINABLE SPATIOTEMPORAL MACHINE LEARNING FOR 24-HOUR AIR QUALITY FORECASTING ACROSS TEN PAKISTANI CITIES

Bushra Shaikh[*1], Abdullah Maitlo[2], Hina Kareem[3], Manahil Shaikh[4]

[*1,2,3,4]*Faculty of Physical Sciences, Institute of Computer Science, Shah Abdul Latif University, Khairpur Mirs, Sindh, Pakistan*

[*1]shykhbushra23@gmail.com, [2]abdullah.maitlo@salu.edu.pk, [3]hinakareem.mahar@gmail.com, [4]manahilshaikh36@gmail.com

DOI: https://doi.org/10.5281/zenodo.20376417

Keywords
air quality forecasting; Air Quality Index; PM2.5; Pakistan; explainable artificial intelligence; SHAP; Random Forest; XGBoost; LightGBM; spatiotemporal machine learning

Article History
Received: 24 March 2026
Accepted: 03 May 2026
Published: 25 May 2026

Corresponding Author: *
Bushra Shaikh

Abstract
Short-term air pollution forecasting can help environmental agencies, health services, and residents prepare for harmful pollution episodes. This study presents XAI-AirNet, an explainable artificial intelligence (XAI) framework for 24-hour-ahead Air Quality Index (AQI) forecasting across ten Pakistani cities. The analysis uses 21,840 hourly observations collected from 6 November 2025 to 4 February 2026. The records include PM10, PM2.5, carbon monoxide, nitrogen dioxide, sulfur dioxide, ozone, dust, temperature, humidity, precipitation, wind speed, wind direction, and pressure. A numeric AQI target was derived from PM2.5 and PM10 through breakpoint-based interpolation. The category target used the reported AQI class shifted 24 hours ahead. The pipeline produced 135 predictors from city identifiers, time variables, lagged pollutant values, and rolling-window summaries. A chronological 80/20 split was used to test Ridge, Random Forest, XGBoost, and LightGBM regressors. Logistic Regression, Random Forest, XGBoost, LightGBM, and a multilayer perceptron were tested for category prediction. Random Forest gave the best AQI value forecast, with MAE = 29.00, RMSE = 39.25, and R2 = 0.609. LightGBM was close, with MAE = 28.38, RMSE = 39.31, and R2 = 0.608. For category forecasting, Logistic Regression gave the best macro-F1 score of 0.506. This result shows that simple interpretable boundaries can remain useful when AQI classes are imbalanced. SHapley Additive exPlanations (SHAP) and model feature importance identified recent AQI, rolling AQI, PM2.5, PM10, temperature, and humidity as major predictors. The results show that transparent ensemble learning can support next-day air quality warning systems in Pakistan. They also show the need for longer, multi-season datasets to improve category-level warnings.

1. INTRODUCTION

Air pollution is a major environmental and public-health risk. Fine particulate matter is linked with respiratory disease, cardiovascular disease, and avoidable mortality [1], [2]. These risks are acute in fast-growing South Asian cities, where traffic, industry, dust, seasonal inversions, and local weather can combine over short time scales. City-level forecasting is therefore a practical need, not only a technical exercise.

Air quality prediction has used atmospheric transport models, statistical time-series methods, and hybrid systems. Recent work shows that machine-learning models can learn nonlinear links between pollutants, meteorology, time, and location from observational data [6]-[15]. Tree ensembles and gradient-boosting methods are useful for tabular environmental data because they can process mixed predictors and provide interpretable importance scores [16]-[21]. Explainability matters because public agencies need to know which conditions led to a predicted unhealthy episode.

Pakistan is a suitable setting for a multi-city AQI forecasting benchmark. Major cities face repeated particulate pollution, winter smog, urban dust, and meteorological trapping. Yet open hourly benchmarks that cover several Pakistani cities remain limited. This study addresses that gap by using a three-month hourly dataset from ten cities and by evaluating both AQI value prediction and AQI category prediction at a 24-hour horizon.

The study makes four contributions. First, it organizes 21,840 hourly pollutant and weather records into a structured multi-city benchmark. Second, it defines both numeric AQI forecasting and categorical risk prediction for the next day. Third, it compares linear, ensemble, gradient-boosting, and neural-network baselines under the same chronological split. Fourth, it uses feature importance and SHAP analysis to connect model behavior with environmental drivers such as recent AQI, particulate matter, and meteorological conditions.

2. Related Work

Short-term air quality forecasting is difficult because pollutant concentrations are nonlinear, time-dependent, and sensitive to weather. PM2.5 and PM10 often show persistence, but they can change quickly with wind, rain, humidity, atmospheric stability, traffic patterns, industrial emissions, and regional transport. Machine learning is well suited to this setting because it can combine pollutant, weather, temporal, and spatial variables without requiring a full chemical transport model.

Many recent studies have evaluated tree ensembles, support vector machines, gradient boosting, and neural networks for AQI, PM2.5, PM10, and ozone prediction [6]-[15]. XGBoost is widely used because it provides regularized boosted trees for strong tabular prediction [16]. LightGBM is efficient on large tabular datasets because it uses histogram-based learning and leaf-wise tree growth [17]. Random Forest remains a strong baseline because it averages many decision trees and reduces sensitivity to noisy individual variables [18], [19]. Deep models, including long short-term memory and convolutional recurrent hybrids, are also common when long sequential datasets are available [10], [13], [22], [23].

A related development is the move from accuracy-only reporting to interpretable forecasting. SHAP assigns feature-level contributions to predictions and is widely used for tree-based and other machine-learning models [20], [21]. In air quality studies, such explanations help identify whether a forecast is driven by pollutant persistence, weather, city effects, or time patterns. This improves trust when forecasts are used for public advisories.

Evaluation design is also important. Recent work on filtered retrieval benchmarks shows that ground truth must be defined within the same candidate set used for evaluation; otherwise, performance estimates can be optimistic or inconsistent [26]. Although air quality forecasting is a different task, the same principle supports clear target construction, consistent train-test separation, and transparent reporting. Leakage-aware benchmarking also shows that hidden duplicates across data splits can inflate reported accuracy [29]. This study therefore uses a chronological split and does not allow future observations to enter training.

3. Materials and Methods

XAI-AirNet follows a chronological forecasting design. The raw dataset contains hourly observations for ten Pakistani cities from 6 November 2025 to 4 February 2026. Each record includes pollutants, meteorological variables, coordinates, and time descriptors. The pipeline includes preprocessing, AQI target construction, feature engineering, chronological splitting, model training, evaluation, and model explanation.

The numeric AQI target was estimated from PM2.5 and PM10 using breakpoint-based interpolation aligned with public AQI communication practice [3], [4]. For each city and timestamp, separate PM2.5 and PM10 sub-indices were calculated. The larger sub-index was then used as the estimated AQI. The classification target used the reported AQI labels: Good, Moderate, Unhealthy for Sensitive Groups, Unhealthy, Very Unhealthy, and Hazardous.

For 24-hour forecasting, the estimated AQI value and the AQI category were shifted forward by 24 hourly steps within each city. This design makes each row predict the same city one day later and avoids mixing information across cities. Feature engineering produced 135 predictors. These predictors included city and coordinate variables, time indicators, pollutant lags at 1, 3, 6, 12, and 24 hours, and rolling means over 3, 6, 12, and 24 hours. Following dataset documentation practices in recent public benchmark releases, the manuscript reports sample counts, city summaries, variable summaries, and modeling details to make the analysis easier to audit [27].

The first 80% of observations were used for training, and the final 20% were reserved for testing. This chronological split better matches deployment than a random split because the model must forecast later observations from earlier data. Regression models were evaluated with mean absolute error (MAE), root mean squared error (RMSE), and R2. Category models were evaluated with accuracy, macro-precision, macro-recall, and macro-F1. Macro-F1 was emphasized because the AQI categories are imbalanced. The Good and Hazardous categories are much less frequent than Moderate and Unhealthy categories.

Figure 1. XAI-AirNet workflow for 24-hour AQI forecasting, including preprocessing, feature engineering, chronological evaluation, and explainable model analysis.

Spectrum of Engineering Sciences
ISSN (e) 3007-3138 (p) 3007-312X Volume 4, Issue 5, 2026

Table 1. Dataset and experimental design summary.

Item	Value
Rows after cleaning	21840
Rows used for modeling	21600
Cities	10
Time period	2025-11-06 00:00:00 to 2026-02-04 23:00:00
Forecast horizon	24 hours ahead
Engineered features	135
Regression target	Estimated AQI from PM2.5/PM10
Classification target	AQI category shifted 24 hours ahead
Train-test split	80% chronological train / 20% test

Table 2. City-wise descriptive summary of estimated AQI, particulate matter, and meteorology.

city	Records	Mean_AQI	Median_AQI	Max_AQI	Mean_PM25	Mean_PM10	Mean_Temperature	Mean_Humidity
Faisalabad	2184	216.870	198.090	490.620	155.740	159.080	13.820	75.160
Multan	2184	200.640	188.950	451.190	136.830	140.610	14.960	71.170
Lahore	2184	192.060	184.870	403.250	126.760	129.250	14.120	72.980
Sialkot	2184	183.470	176.300	409.980	115.190	117.840	13.120	78.950
Rahim Yar Khan	2184	148.720	157.710	249.360	72.360	76.020	15.870	68.410
Islamabad	2184	129.830	134.490	197.830	56.130	58.900	11.310	67.620
Rawalpindi	2184	129.830	134.490	197.830	56.130	58.900	11.410	67.620
Peshawar	2184	124.950	133.500	206.950	53.730	56.430	12.930	49.410
Karachi	2184	105.210	101.740	171.860	37.710	46.420	20.750	53.320
Quetta	2184	64.320	62.800	335.590	19.100	40.730	8.930	42.770

Table 3. Distribution of observed AQI categories in the full dataset.

AQI category	Records	Share (%)
Good	759	3.480
Moderate	4861	22.260
Unhealthy for Sensitive Groups	4199	19.230
Unhealthy	8766	40.140
Very Unhealthy	2617	11.980
Hazardous	638	2.920

Table 4. Descriptive statistics for selected pollutants and meteorological variables.

Variable	count	mean	std	min	25%	50%	75%	max
estimated_aqi	21840.000	149.590	65.090	2.080	98.530	154.670	180.430	490.620
pm10	21840.000	88.420	66.010	0.500	40.500	68.500	117.800	488.000
pm2_5	21840.000	82.970	65.740	0.500	34.700	62.400	112.500	486.200
carbon_monoxide	21840.000	1580.860	1145.630	91.000	741.000	1296.500	2113.250	11482.000
nitrogen_dioxide	21840.000	39.470	32.290	0.000	13.400	30.400	57.800	185.100
ozone	21840.000	63.800	50.400	0.000	21.000	55.000	100.000	265.000
dust	21840.000	9.860	32.530	0.000	1.000	3.000	8.000	711.000
temperature	21840.000	13.720	5.040	-5.900	10.200	13.300	16.900	30.400
humidity	21840.000	64.740	20.270	4.000	52.000	64.000	81.000	100.000
wind_speed	21840.000	5.340	3.920	0.000	2.900	4.200	6.600	33.400
pressure	21840.000	1020.050	3.490	1006.400	1017.600	1020.000	1023.400	1030.800

Table 5. Models evaluated in the XAI-AirNet benchmark.

Task	Model	Role
AQI value forecasting	Ridge	Linear baseline
AQI value forecasting	Random Forest	Nonlinear ensemble
AQI value forecasting	XGBoost	Gradient boosting
AQI value forecasting	LightGBM	Gradient boosting
AQI category forecasting	Logistic Regression	Interpretable classifier
AQI category forecasting	Random Forest	Nonlinear ensemble
AQI category forecasting	XGBoost	Gradient boosting
AQI category forecasting	LightGBM	Gradient boosting
AQI category forecasting	Multilayer Perceptron	Neural-network benchmark

4. Results

The 24-hour numeric AQI forecast was moderately predictable from recent pollutant levels, weather, time variables, and city information. Random Forest achieved the lowest RMSE of 39.254 and the highest R2 of 0.609. LightGBM produced the lowest MAE of 28.377 and nearly the same R2 of 0.608. XGBoost and Ridge were slightly weaker but remained close to the two best models. These results suggest that much of the next-day AQI signal is captured by recent pollutant persistence and rolling exposure patterns.

AQI category forecasting was more difficult. Logistic Regression achieved the highest macro-F1 score, 0.506, and the highest macro-recall, 0.608. Random Forest had the highest accuracy, 0.584, but its macro-F1 was lower at 0.428. This difference shows the effect of class imbalance. A model can score well on the common classes while doing less well on rare classes such as Good and Hazardous.

Table 6. Final regression comparison for 24-hour-ahead estimated AQI forecasting.

Model	MAE	RMSE	R2
Random Forest	29.003	39.254	0.609
LightGBM	28.377	39.309	0.608
XGBoost	29.143	40.287	0.588
Ridge	30.757	40.468	0.585

Table 7. Final classification comparison for 24-hour-ahead AQI category forecasting.

Model	Accuracy	Precision_macro	Recall_macro	F1_macro
Logistic Regression	0.543	0.476	0.608	0.506
LightGBM	0.566	0.474	0.484	0.473
XGBoost	0.578	0.501	0.461	0.467
CPU Neural Network	0.533	0.472	0.437	0.443
Random Forest	0.584	0.461	0.424	0.428

Table 8. Interpretable key predictors identified from feature-importance and SHAP evidence.

Feature	Interpretation
estimated_aqi_lag1h	Recent AQI persistence
estimated_aqi_roll24h	Daily smoothed AQI memory
pm2_5_lag1h	Recent fine-particle state
pm10_lag1h	Recent coarse-particle state
estimated_aqi_lag24h	Same-hour previous-day AQI
pm2_5_roll24h	Daily PM2.5 exposure memory
pm10_roll24h	Daily PM10 exposure memory
temperature_roll24h	Thermal context
humidity_roll24h	Moisture and aerosol context
carbon_monoxide_lag1h	Combustion-related persistence
ozone_roll24h	Photochemical background
wind_speed_roll24h	Dispersion-related meteorology

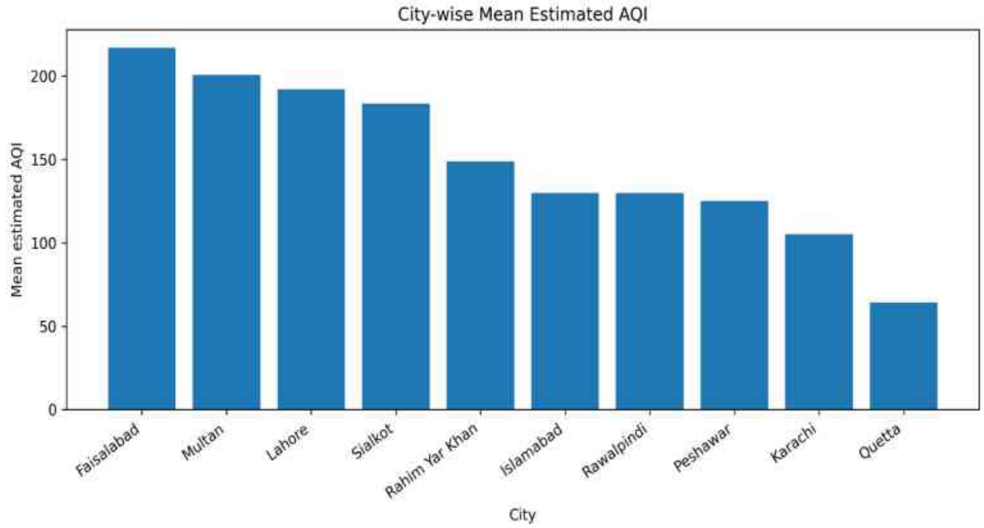


Figure 2. City-wise mean estimated AQI across the ten Pakistani cities, showing the strongest average burden in Faisalabad, Multan, Lahore, and Sialkot.

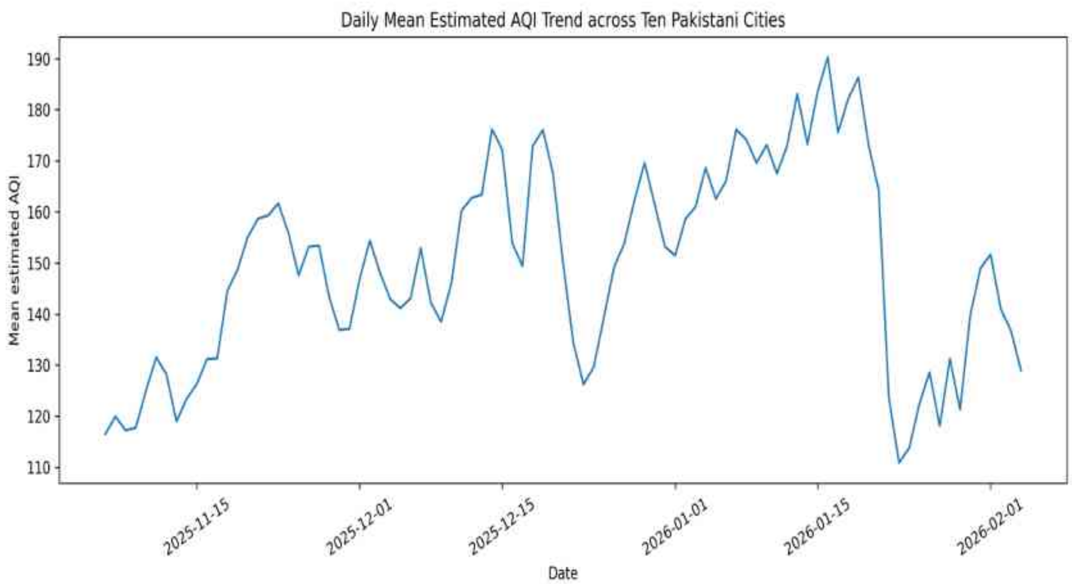


Figure 3. Daily mean estimated AQI during the study period, showing short-term variation across the three-month observation window.

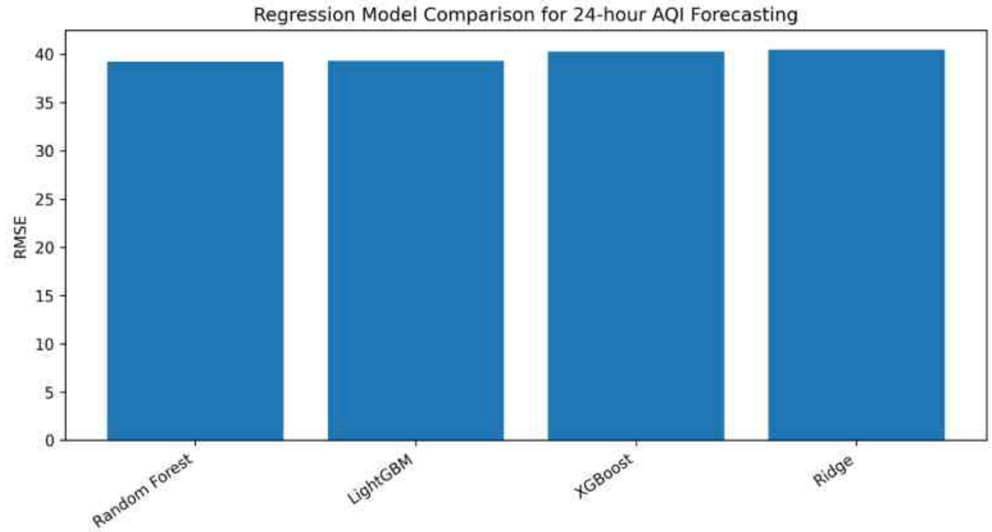
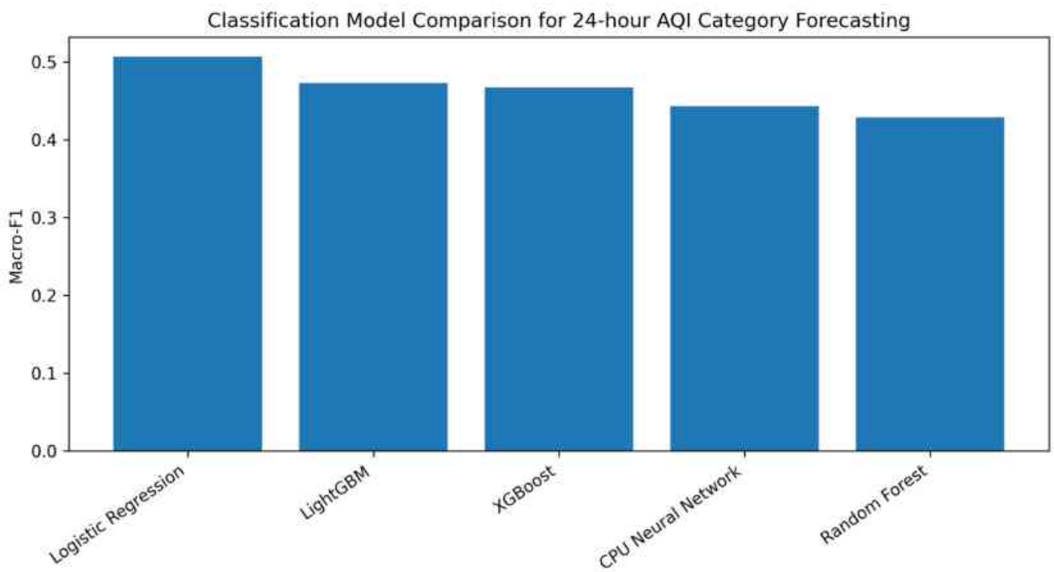


Figure 4. Regression model comparison by RMSE, where lower values indicate more accurate AQI value forecasts.



***Fi*gure 5. Classification model comparison by macro-F1, highlighting performance under category imbalance.**

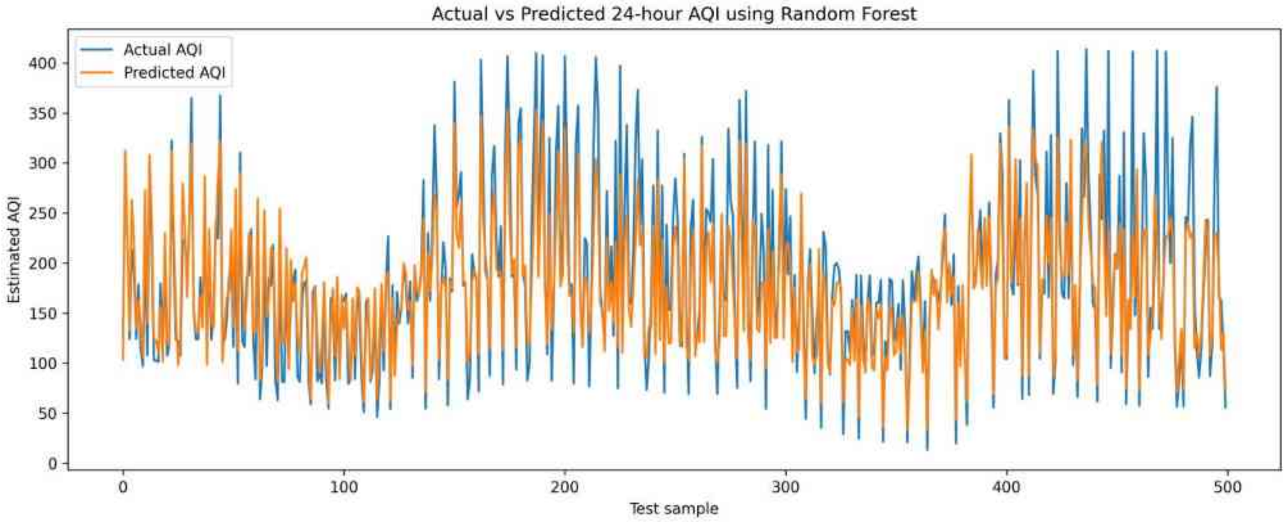


Figure 6. Actual and predicted estimated AQI for the best regression model, illustrating how closely the model tracks observed variation.

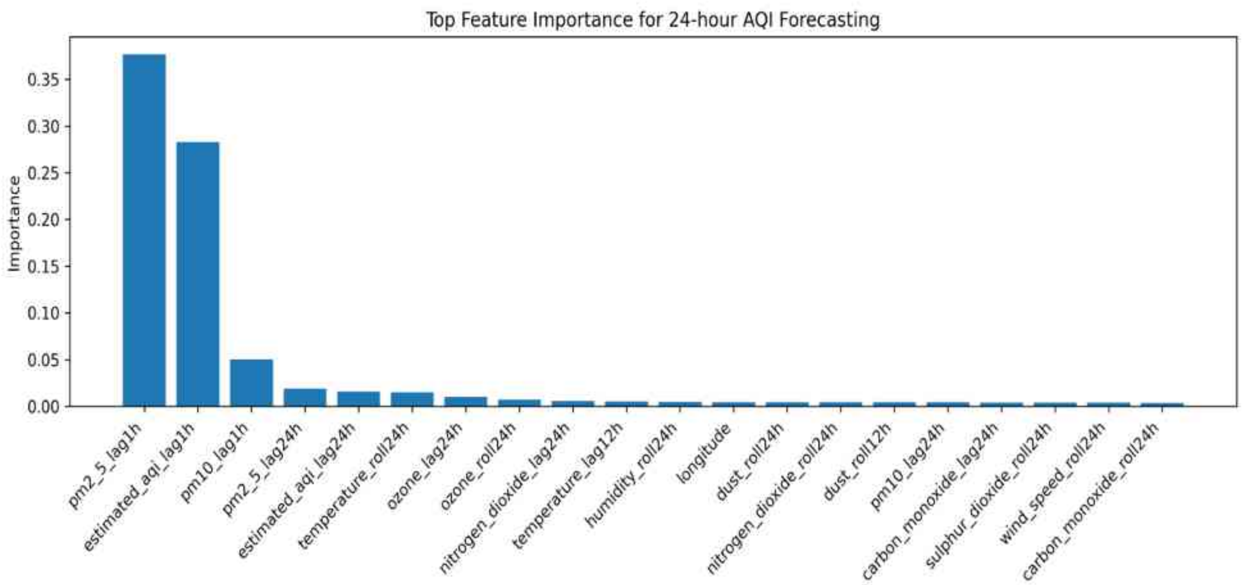


Figure 7. Feature-importance ranking for 24-hour AQI forecasting, with lagged and rolling AQI features among the strongest predictors.

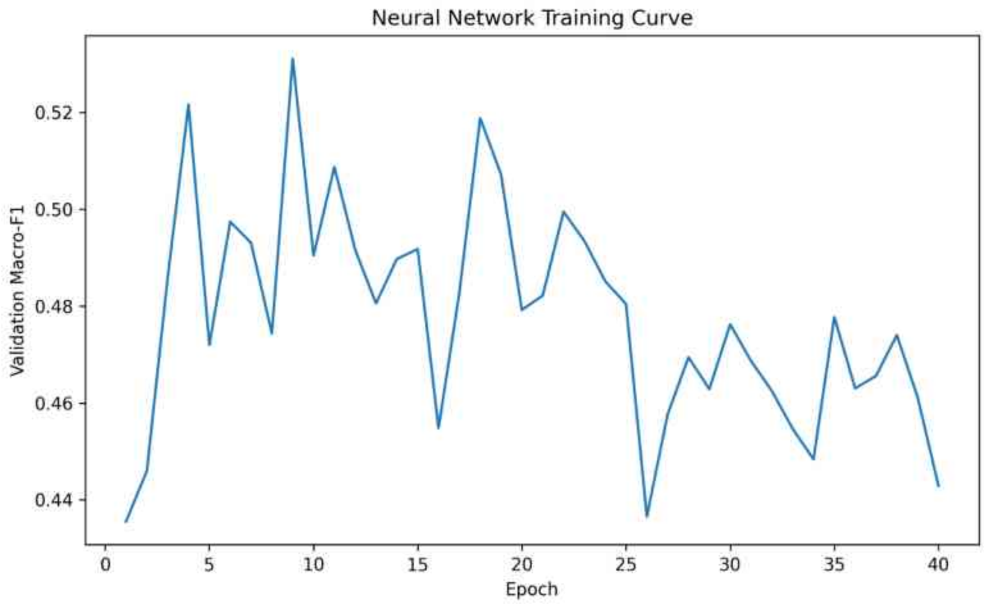


Figure 8. Neural-network training curve based on validation macro-F1 across training epochs.

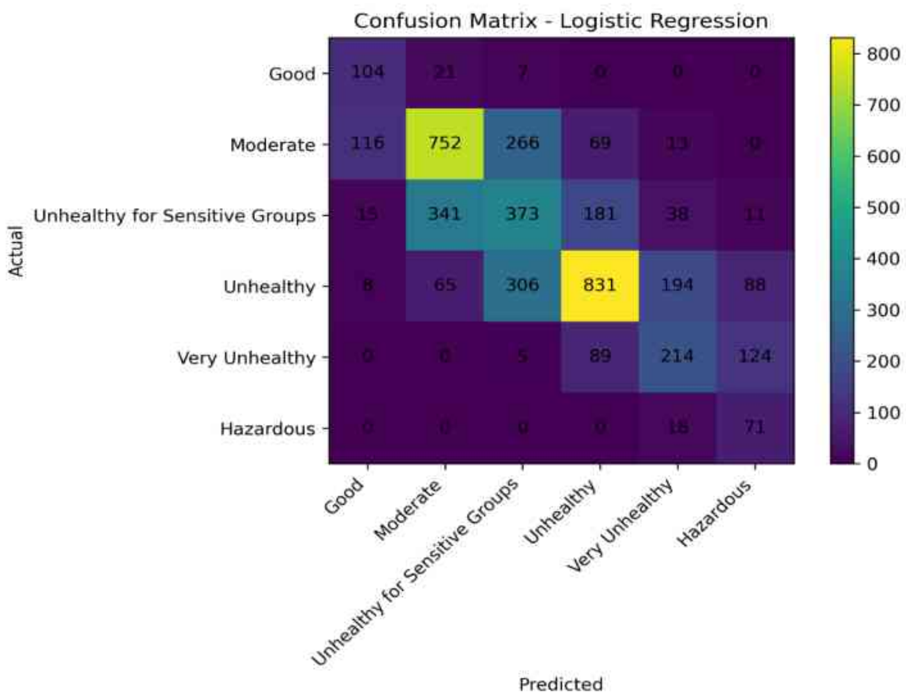


Figure 9. Confusion matrix for the best AQI category forecasting model, showing where category-level errors are concentrated.

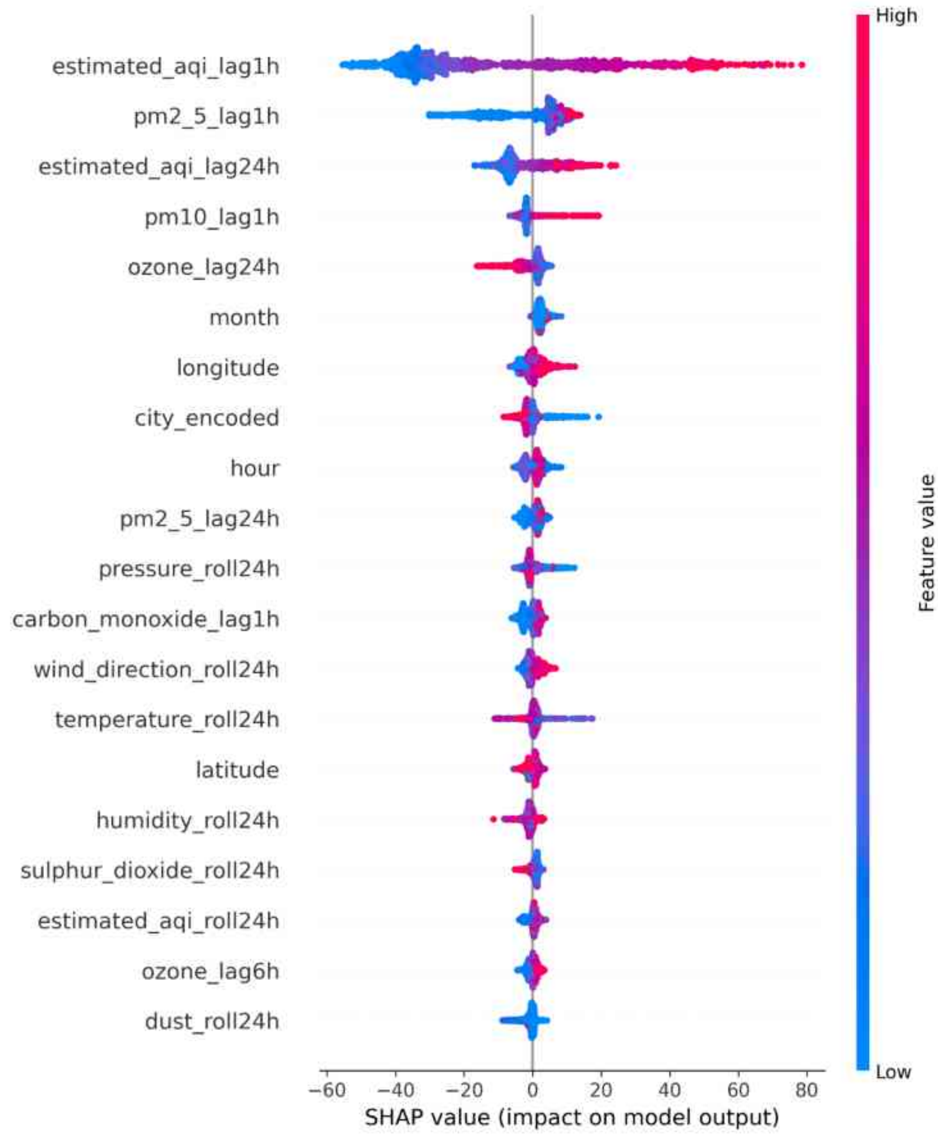


Figure 10. SHAP summary plot for AQI value prediction, showing how major features contribute to higher or lower forecasts.

5. Discussion

The results lead to three main findings. First, AQI forecasts are strongly influenced by temporal persistence. Recent AQI values, daily rolling AQI, and recent PM2.5 and PM10 appear among the most informative predictors. This is consistent with atmospheric behavior because particulate episodes often last for several hours or days when dispersion is weak.

Second, Random Forest and LightGBM gave the strongest regression results. Their similar performance suggests that nonlinear interactions among pollutants, weather, city differences, and recent history are important. The Ridge baseline was weaker, but not by a large margin. This indicates that part of the signal is approximately linear, while ensemble models capture additional nonlinear structure.

Third, category prediction remains harder than numeric AQI forecasting. Logistic Regression gave the best macro-F1 despite not having the highest accuracy. This result is plausible because macro-F1 gives equal weight to rare and common categories. Under imbalanced AQI labels, a simpler classifier can sometimes provide more balanced decision boundaries than a more flexible model. This does not mean that linear models are always better. It means that category-level forecasting needs class-imbalance handling, calibrated probabilities, and longer datasets.

The explainability outputs are environmentally reasonable. SHAP and feature-importance results identify recent AQI, rolling AQI, PM2.5, PM10, temperature, humidity, carbon monoxide, ozone, and wind speed as important drivers. These features match known pollution mechanisms. Particulate persistence reflects accumulated emissions and limited dispersion. Humidity can affect aerosol behavior and visibility. Wind speed can dilute or transport pollutants. Temperature can be linked with boundary-layer conditions and seasonal heating patterns.

The study also supports a broader lesson for environmental machine learning. Forecasts should not be judged by a single average metric. Model selection should consider value prediction, class-level warning behavior, calibration, and interpretability. Work on unseen-regime prediction in other data-driven systems shows that strong ranking performance may not guarantee robust decision quality when conditions shift [30]. Air quality forecasting faces a similar challenge when future seasons, emissions, or weather regimes differ from the training period.

6. Policy and Practical Implications

XAI-AirNet can support practical air quality management in several ways. Municipal agencies could use next-day city-level forecasts to prepare advisories, plan temporary control actions, and warn sensitive groups. Health services could use forecasts to support asthma and cardiovascular-risk communication during expected high-pollution days. Schools and workplaces could use early warnings to adjust outdoor activities when unhealthy categories are likely.

The explainable part of the framework is important for policy use. A forecast is more useful when it is accompanied by the main drivers behind it. If the forecast is driven by high PM2.5 persistence, agencies may focus on local emission sources and public-health warnings. If wind or humidity patterns are important, agencies can communicate meteorological contributions more clearly. Interpretable machine-learning work on structured questionnaire data shows that explanation-based models can turn feature patterns into actionable recommendations [28]. A similar approach can improve air quality communication by linking forecasts to understandable risk factors.

The results also point to data priorities. Hourly resolution is valuable, but the current record covers only about three months. Future systems should include longer monitoring, seasonal cycles, official station measurements, satellite indicators, emission proxies, road activity, and boundary-layer variables. Such additions would help models distinguish routine persistence from unusual pollution events.

7. Limitations and Future Work

This study has several limitations. First, the dataset covers a three-month period, which limits seasonal generalization. Second, the numeric AQI target was estimated from PM2.5 and PM10 rather than taken directly from an official monitoring platform. Breakpoint-based AQI estimation is useful, but official reporting may use additional pollutant rules, averaging windows, and local conventions. Third, category prediction is affected by class imbalance, especially for the Good and Hazardous classes. Fourth, the neural-network baseline was a multilayer perceptron trained on engineered tabular features, not a sequential recurrent model.

Future work should extend the data period across multiple seasons and years. It should also test spatial transfer by training on some cities and evaluating on others. Longer records would allow models such as long short-term memory networks, temporal convolutional networks, and transformer-based forecasters to be compared fairly. The evaluation should also include calibration, alert thresholds, city-wise fairness, and rare-event sensitivity. Distribution shift across unseen operating regimes remains a central problem in predictive systems, as shown in federated SLA-risk forecasting under unseen 6G RAN regimes [30]. Similar unseen-regime testing is needed for air quality models when future weather, emissions, or monitoring conditions change.

The current benchmark remains useful because it provides a transparent starting point for 24-hour AQI prediction across Pakistani cities. The limitations should guide future extensions rather than weaken the main result: recent pollutant history and weather features can provide informative next-day forecasts when evaluated with a chronological design.

8. Conclusion

This paper presented XAI-AirNet, an explainable spatiotemporal machine-learning framework for 24-hour-ahead AQI forecasting across ten Pakistani cities. The study used 21,840 hourly pollutant and meteorological records, created 135 engineered features, and evaluated regression and classification models under a chronological split. Random Forest achieved the strongest numeric AQI forecast, while Logistic Regression achieved the best macro-F1 for AQI category prediction. Explainability analysis identified recent AQI, rolling AQI, PM2.5, PM10, and meteorological variables as the main drivers. The findings show that transparent machine-learning models can support short-term air quality early warning in Pakistan. They also show that longer multi-seasonal datasets and stronger class-imbalance methods are needed for better category-level forecasting.

REFERENCES

[1] World Health Organization, "Ambient (outdoor) air pollution," WHO Fact Sheet, Oct. 2024.

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

[3] U.S. Environmental Protection Agency, Technical Assistance Document for the Reporting of Daily Air Quality: The Air Quality Index (AQI), EPA-454/B-24-002, May 2024.

[4] U.S. Environmental Protection Agency, "National Ambient Air Quality Standards for Particulate Matter," EPA, 2024.

[5] Health Effects Institute, State of Global Air 2024. Boston, MA, USA: Health Effects Institute, 2024.

[6] S. Tirink, "Machine learning-based forecasting of air quality index under long-term environmental patterns: A comparative approach with XGBoost, LightGBM and SVM," PLOS ONE, 2025.

[7] R. Rakholia et al., "Accurate PM2.5 urban air pollution forecasting using a multivariate multi-step ensemble machine learning model," Chemosphere, 2024.

[8] I. E. Agbehadji and I. C. Obagbuwa, "Systematic review of machine learning and deep learning techniques for spatiotemporal air quality prediction," Atmosphere, vol. 15, no. 11, p. 1352, 2024.

[9] S. Zhou et al., "Deep-learning architecture for PM2.5 concentration prediction: A review," Environmental Data Science and Applications, 2024.

[10] X. Wang et al., "Air quality forecasting using a spatiotemporal hybrid deep learning model," Scientific Reports, vol. 14, 2024.

[11] W. Yang et al., "Spatio-temporal feature interpretable model for air quality forecasting," Ecological Informatics, 2024.

[12] J. Ma et al., "Application of the XGBoost machine learning method in PM2.5 prediction: A case study of Shanghai," Aerosol and Air Quality Research, vol. 20, pp. 128-138, 2020.

[13] N. Zaini et al., "PM2.5 forecasting for an urban area based on deep learning and decomposition methods," Scientific Reports, vol. 12, 2022.

[14] Y. Ozupak et al., "Air quality forecasting using machine learning: Comparative analysis of regression models," Water, Air, & Soil Pollution, 2025.

[15] A. Yenkikar et al., "Explainable forecasting of air quality index using a hybrid machine-learning framework," Environmental Modelling and Assessment, 2025.
[16] T. Chen and C. Guestrin, "XGBoost: A scalable tree boosting system," in Proc. 22nd ACM SIGKDD Int. Conf. Knowledge Discovery and Data Mining, 2016, pp. 785-794.
[17] G. Ke et al., "LightGBM: A highly efficient gradient boosting decision tree," in Advances in Neural Information Processing Systems, vol. 30, 2017.
[18] L. Breiman, "Random forests," Machine Learning, vol. 45, no. 1, pp. 5-32, 2001.
[19] T. K. Ho, "Random decision forests," in Proc. 3rd Int. Conf. Document Analysis and Recognition, 1995, pp. 278-282.
[20] S. M. Lundberg and S.-I. Lee, "A unified approach to interpreting model predictions," in Advances in Neural Information Processing Systems, vol. 30, 2017.
[21] S. M. Lundberg et al., "From local explanations to global understanding with explainable AI for trees," Nature Machine Intelligence, vol. 2, pp. 56-67, 2020.
[22] S. Hochreiter and J. Schmidhuber, "Long short-term memory," Neural Computation, vol. 9, no. 8, pp. 1735-1780, 1997.
[23] I. Goodfellow, Y. Bengio, and A. Courville, Deep Learning. Cambridge, MA, USA: MIT Press, 2016.
[24] F. Pedregosa et al., "Scikit-learn: Machine learning in Python," Journal of Machine Learning Research, vol. 12, pp. 2825-2830, 2011.
[25] C. R. Harris et al., "Array programming with NumPy," Nature, vol. 585, pp. 357-362, 2020.
[26] S. Bibi, F. A. Rajput, M. Younis, S. Bibi, and B. Raza, "Vector+SQL retrieval with selectivity workloads: Measuring tail latency and quality under filtered Top-K," VFAST Transactions on Software Engineering, vol. 14, no. 1, pp. 335-349, 2026.
[27] B. Raza, S. Bibi, S. Bibi, and A. Nawaz, "SADA COLOR DATASET (SCD): 9 paper colors x 4 illumination conditions for robust color vision evaluation," Spectrum of Engineering Sciences, vol. 4, no. 2, pp. 871-887, 2026.
[28] B. Raza, S. Rajper, H. Shaikh, Z. H. Shar, S. Bibi, I. Hyder, and S. Bibi, "Generative AI confidence modeling in nursing education based on interpretable machine learning: An analytical framework for structured questionnaire data," Spectrum of Engineering Sciences, vol. 4, no. 2, 2026.
[29] B. Raza, S. Bibi, S. Bibi, and A. Nawaz, "Touchless document interaction with leakage-aware benchmarking: A four-command static hand-gesture study," Spectrum of Engineering Sciences, vol. 4, no. 4, 2026.
[30] P. Mangi, S. Bibi, A. Nawaz, and S. Bibi, "When clients drift: Federated SLA-risk forecasting across unseen 6G RAN regimes," Spectrum of Engineering Sciences, vol. 4, no. 4, 2026.