International Journal of Recent Development in Engineering and Technology
Website: www.ijrdet.com (ISSN 2347-6435 (Online) Volume 15, Issue 06, June 2026)

Deep Learning, GANs, and XAI Paradigms for Sustainable Air Quality Management" A Comprehensive Survey

Rajneesh Pachouri[1], Dr. Vikas Sakalle[2], Dr. Sonal Sakalle[3]

[1]*Research Scholar, LNCT University Bhopal (M.P), India*
[2]*Professor & Head,* [3]*Associate Professor (Computer science & Engineering), LNCT University Bhopal (M.P), India*

***Abstract*—Abstract—Air quality management stands as a paramount sustainability goal in the era of Industry 5.0, yet real-world environmental datasets are continually plagued by severe class imbalances, data vulnerabilities, and a fundamental lack of model transparency. To address these challenges, this paper presents a reliable, generative, and interpretable multi-class framework that classifies the Air Quality Index (AQI) into six distinct categorical levels. The proposed architecture uniquely integrates the capabilities of Generative Adversarial Networks (GANs), Machine Learning (ML)/Deep Learning (DL) classifiers, and Explainable AI (XAI) paradigms. Originally, a set of traditional machine learning models were tested such as Decision Tree (DT), Random Forest (RF), Adaptive Boosting (AdaBoost), Gradient Boosting (GB), and Logistic Regression (LR). The Random Forest model resulted in baseline performance metrics with accuracy and precision of 0.99. To overcome critical data limitations such as missing values, noise, data bias, and structural class imbalances (where dominant classes such as "Satisfactory" or "Moderate" skew the predictions away from critical edge cases such as "Severe"), a GAN was deployed to synthesize high quality, realistic environmental scenario data. The augmentation of the classification pipeline with the GAN-generated synthetic records allowed the system to reach optimized classification performance with overall accuracies closer to 100%. To overcome the trade-off between high accuracy and the "black box" characteristic of deep neural networks, the framework has rigorous local and global interpretability mechanics. Local Interpretable Model-agnostic Explanations (LIME) are applied to estimate local surrogacy plots for individual instances while SHapley Additive exPlanations (SHAP) provide global surrogacy insights to outline exact feature importance and multi-pollutant variable interactions. Ultimately, this unified GAN-AI-XAI paradigm provides an exceptionally robust, transparent, and trustworthy model for multi-class AQI forecasting, ensuring that the predictive outcomes are fully explainable and actionable for public environmental stakeholders and policy developers.**

***Keywords*— Air Quality Index, Explainable AI, Generative Adversarial Networks, Artificial Intelligence, Multi-class Classification.**

I. INTRODUCTION

Air quality management has emerged as one of the most critical sustainability goals in the era of Industry 5.0. As rapid urbanization, industrial expansion, and vehicular emissions continue to escalate, the magnitude of air pollution and the impact of drastic environmental pollutants increase day by day. Global estimates indicate that air pollution is responsible for approximately 7 million premature deaths annually. Furthermore, nearly 99% of the global population resides in neighborhoods where the concentrations of air toxins exceed standard atmospheric guidelines , presenting severe risks to human respiratory and circulatory systems , alongside monumental economic costs.The environmental degradation is generally indicated by some key pollutants such as Particulate Matter ($PM_{2.5}$ and PM_{10}), Ozone O_3, Carbon Monoxide (CO, Nitrogen Dioxide (NO_2), and Sulfur Dioxide (SO_2). To overcome these hazards, the environmental monitoring networks and the IoT-enabled sensors have been extensively deployed to publish real-time air eminence data. However, passive monitoring is not enough for proactive public health intervention. Precise and timely forecasting of the Air Quality Index (AQI) is very much needed. Accurate predictive modeling helps the municipalities to identify the high-risk periods or highly polluted micro-environments beforehand empowering communities and stakeholders to implement protective measures and execute sustainable policy choices.For decades, statistical and statistics-based methods have been the backbone of air quality prediction. Early systems relied heavily on traditional statistical models, such as the Autoregressive Integrated Moving Average (ARIMA) and Multivariable Linear Regression (MLR). While these mathematical methods can handle basic time-series dependencies adequately, they are inherently based on linear assumptions and face challenges when dealing with highly nonlinear or non-stationary environmental dynamics.

International Journal of Recent Development in Engineering and Technology
Website: www.ijrdet.com (ISSN 2347-6435 (Online) Volume 15, Issue 06, June 2026)

To address this limitation, Traditional Machine Learning (TML) algorithms—including Decision Trees (DT), Random Forests (RF), Support Vector Regression (SVR), and Gradient Boosting (GB)—were introduced, demonstrating significant effectiveness in capturing nonlinear relationships and extracting latent features from input variables. However, TML models are limited by their relatively shallow architectures, which restricts their ability to perform automated, deep data mining when handling large-scale, multi-site meteorological operations.The advent of Deep Learning (DL) has changed this domain forever, with multi-layer neural networks capable of handling complex, high-dimensional and large-scale datasets. Sophisticated DL architectures such as Convolutional Neural Networks (CNNs), Graph Convolutional Networks (GCNs), and Graph Attention Networks (GATs) are extremely powerful at extracting spatial dependencies from regional multi-site stations. Meanwhile, temporal feature architectures such as Long Short-Term Memory (LSTM) networks, Transformers, and Bidirectional Encoder Representations from Transformers (BERT) model complex spatio-temporal relationships over long horizons. Despite these algorithmic leaps, modern predictive air quality management suffers from two critical roadblocks: data vulnerability and the "black-box" dilemma. 1. Data Vulnerability and Class Imbalance: Real-world environmental monitoring data is continuously prone to missing values, noise, sensor faults, and severe class imbalances.Because extreme atmospheric anomalies (e.g., "Severe" AQI conditions) occur less frequently than "Good" or "Satisfactory" conditions, classifiers suffer from data bias, severely damaging accuracy during the most dangerous pollution spikes. Generative Adversarial Networks (GANs) have emerged as a disruptive solution to this problem.GANs simulate real-world atmospheric conditions to successfully produce high-quality synthetic data, balancing unevenly distributed classes, mitigating noise, and filling data gaps. 2. The Black-Box Dilemma: As models become more complex, especially hybrid GAN-CNN and Deep Learning frameworks, their inner workings are completely opaque. Stakeholders in high-stakes sustainable urban planning and policy implementation cannot blindly trust a black-box model. This has spurred the urgent integration of Explainable Artificial Intelligence (XAI).XAI frameworks, utilizing local surrogates like Local Interpretable Model-agnostic Explanations (LIME) and global surrogates like SHapley Additive exPlanations (SHAP), unveil the exact patterns driving predictions. XAI details feature importance, ensuring that predictive pipelines remain transparent, reliable, accountable, and trustworthy for public policymakers.

While individual studies have begun integrating pieces of these methodologies—such as utilizing GANs for missing data generation or applying XAI to standard random forests —there remains a distinct research gap in comprehensively synthesizing the synergistic interaction of Deep Learning, GAN architectures, and XAI paradigms specifically applied to environmental sustainability. Contributions and Scope of this Survey This survey paper provides a comprehensive, state-of-the-art overview of how deep generative models and explanation frameworks are converging to build next-generation, sustainable air quality management systems. The key objectives and structure of this survey are outlined as follows:

- *Taxonomy of Predictors:* We categorize and analyze the evolution of air quality models from statistical and traditional machine learning baselines to advanced spatio-temporal Deep Learning networks (CNNs, LSTMs, Transformers**).**
- *Generative Data Engineering:* We discuss the ability of Generative Adversarial Networks (GANs) to mimic complex structures of environmental data, and their potential to solve multi-class imbalances, edge-case data bias, and missing historical records.
- *The Role of Explainability (XAI):* We dissect local and global interpretability paradigms (including LIME and SHAP), demonstrating how they illuminate the impact of complex meteorological and chemical parameters on AQI classifications.
- *Synergistic Frameworks & Future Horizons:* Finally, we evaluate the emerging trade-off between predictive accuracy and interpretability , proposing an integrated paradigm map and outlining open research directions to reduce computational complexity and enable reliable real-time forecasting for future smart cities

II. Related Work

The development of predictive systems for Air Quality Index (AQI) forecasting and environmental informatics has transitioned across four key epochs: conventional statistical mechanics, traditional machine learning (TML), multi-layered deep learning (DL), and trust-enabling explainable frameworks.This literature survey details the foundational milestones and recent paradigm integrations that guide modern sustainable air quality management. **Statistical and Traditional Machine Learning Paradigms** Early environmental modeling relied profoundly on parametric statistical architectures.

International Journal of Recent Development in Engineering and Technology
Website: www.ijrdet.com (ISSN 2347-6435 (Online) Volume 15, Issue 06, June 2026)

Authors have extensively used the Autoregressive Integrated Moving Average (ARIMA) model to map time-series data dependencies. ARIMA is effective in local linear modelling, but the choice of parameters demands extensive domain knowledge and the forecasting performance is severely limited over a long-term window. To consider multi-aspect environmental features, Multivariable Linear Regression (MLR) models were proposed. However, MLR models are based on linear assumptions and cannot accommodate the non-stationary property of atmospheric variations.

Researchers used Traditional Machine Learning (TML) algorithms for modelling non-linear interactions. Significant progress includes the adoption of Support Vector Regression (SVR) with weighted Absolute Percentage Error (APE) metrics in grey multivariable regression models to improve trace pollutant extraction.To accelerate optimization schedule, Backpropagation Neural Networks (BNN) were combined with Particle Swarm Optimization (PSO) to calculate air toxin shifts with limited computational fluid dynamics simulations . Although these single-tier TML arrangements are flexible, they are structurally limited, which impedes the feature-mining capability when large high-dimensional datasets are encountered.

Advanced Spatio-**Temporal Deep Learning Frameworks:** Deep Learning (DL) has largely overcome the limits of shallow architectures by assembling multi-layer networks that learn hierarchically. Managing spatial correlations across multi-site regional networks led to the deployment of Convolutional Neural Networks (CNNs) and Graph Convolutional Networks (GCNs) that group spatial features using station adjacency data. To account for changing spatial impacts, Graph Attention Networks (GATs) use adaptive attention layers to determine the dynamic importance of neighboring sensor nodes. For temporal extraction, Long Short-Term Memory (LSTM) networks are widely used to maintain long-term historical memory across time series. More recently, self-attention architectures—specifically Transformer and Bidirectional Encoder Representations from Transformers (BERT) blocks—have been implemented to resolve long-term temporal tracking issues and navigate complex spatio-temporal dynamics.

Additionally, hybrid architectures have emerged to balance spatial and temporal components, such as combining multi-scale temporal modeling with hierarchical spatial division to dramatically cut Mean Absolute Error (MAE) compared to classic baselines.

Generative Adversarial Networks (GANs) for Data Engineering: Real-world atmospheric monitoring datasets suffer from structural data vulnerabilities. Sensor degradation, hardware failure, and communication drops lead to noisy environments, data bias, and missing sequences. Moreover, extreme environmental crises (e.g., "Severe" or "Hazardous" AQI states) represent clear edge cases, triggering severe multi-class data imbalance where normal conditions dominate the training distribution.

Generative Adversarial Networks (GANs) have introduced a robust path for data engineering in this domain. Because air metrics shift continuously, GANs excel at simulating high-dimensional atmospheric scenarios.The generator is trained using game-theoretic minimax optimization to generate synthetic data similar to real meteorological and chemical data.Supplementing highly imbalanced real samples with GAN-generated records balances minority groups, mitigates historical gaps, and removes structural classification bias. Explainable AI (XAI) and Trust Paradigms

While modern hybrid architectures (such as GAN-CNN or deep ensemble systems) reach high predictive accuracy, their nested non-linear transformations turn them into uninterpretable "black boxes". This lack of transparency restricts their use in real-world policy development and urban management. To address this gap, Explainable AI (XAI) frameworks have been integrated into environmental workflows.

XAI paradigms are broadly divided into local and global surrogates.Local Interpretable Model-agnostic Explanations (LIME) train local linear surrogates around certain data points to explain individual classification events. In contrast, SHapley Additive exPlanations (SHAP) use cooperative game theory to assign uniform weights of feature importance to the entire model.By mapping variable interactions and showing the explicit influence of criteria like Carbon Monoxide (CO), Ozone (O3), and Particulate Matter (PM2.5), XAI uncovers predictive patterns. This transparency makes high-performance deep networks accountable and actionable for public environmental stakeholders

International Journal of Recent Development in Engineering and Technology
Website: www.ijrdet.com (ISSN 2347-6435 (Online) Volume 15, Issue 06, June 2026)

Table1: Summary Matrix of Reviewed Literature

Reference Citation	Core Methodology / Models Used	Primary Focus / Environmental Application	Key Findings & Performance Gains	Identified Limitations
[1]	ARIMA (Autoregressive Integrated Moving Average)	Localized time-series forecasting of atmospheric parameters.	Delivers stable and mathematically consistent short-term baseline metrics.	Requires expert parameter tuning; accuracy degrades over long horizons.
[2]	MLR (Multivariable Linear Regression)	Multi-pollutant tracking and parameter analysis.	Simple implementation that scales well across multiple basic weather inputs.	Restrained by linear assumptions; struggles with non-stationary dynamics.
[3]	SVR + Grey Multivariable Regression	Fine particulate matter ($\text{PM}_{2.5}$) and trace dilution estimation.	Utilizing APE weights successfully increased regression precision.	High sensitivity to initial feature weighting constraints.
[4]	BNN (Backpropagation Neural Network) + PSO	Fast air quality classification and wind simulation optimization.	Concurrently minimizes fluid dynamics steps and speeds up processing time.	Susceptible to local minima trapping during parameter optimization.
[5]	CNN (Convolutional) & GCN (Graph Convolutional Networks)	Regional spatial correlation mapping across multi-site sensor networks.	Extracts geographic patterns using station adjacency matrices.	Rigid graph architectures struggle with dynamic sensor node dropouts.
[6]	GAT (Graph Attention Networks)	Dynamic spatial neighborhood weighting.	Attention maps adaptively score the significance of neighboring stations.	High computational overhead during large graph scale-ups.
[7]	LSTM, Transformers, and BERT Networks	Long-term temporal and sequential dependency modeling.	Self-attention blocks overcome exploding/vanishing gradients over long time series.	Massive training data demands; prolonged convergence phases.
[8]	Hybrid Temporal-Hierarchical Spatial Networks	Joint spatio-temporal feature fusion.	Effectively captures complex weather-pollutant relationships, minimizing MAE.	High architectural design and engineering complexity.
[9]	GAN (Generative Adversarial Networks)	Synthetic data generation and class imbalance engineering.	Effectively mitigates minority class gaps (e.g., "Severe" AQI conditions).	Prone to mode collapse and training instabilities during adversarial play.
[10]	XAI Paradigms (LIME & SHAP Surrogates)	Black-box model interpretation and post-hoc feature evaluation.	Provides local instance charts and global game-theoretic importance scoring.	Explanations act as post-hoc surrogates rather than showing internal mechanics.

III. Problem Statement

Despite continuous advancements in environmental data collection via IoT networks and smart city sensory arrays, developing a highly accurate, robust, and trustworthy multi-class forecasting model for the Air Quality Index (AQI) remains a significant challenge. This overarching problem stems from several interlinked technical and operational bottlenecks:

Structural Data Vulnerabilities and Missing Sequences

Real-world environmental monitoring systems operate in open, often harsh physical conditions, making them highly prone to sensor drift, communication dropouts, and hardware degradation. Consequently, the raw spatial-temporal data streams collected from regional stations are heavily plagued by missing sequences, anomalies, and background noise. Traditional data-cleaning or simple imputation methods (such as mean substitution or linear interpolation) fail to capture the complex, non-linear correlations between distinct atmospheric pollutants, leading to corrupted data that degrades the training quality of downstream classifiers.

Severe Class Imbalances and Edge-Case Data Biases

The distribution of historical AQI records is naturally and fundamentally skewed. In most urban centers, catastrophic or highly hazardous pollution events (classified as "Severe" or "Hazardous") occur far less frequently than standard operational days ("Good" or "Satisfactory").

This creates a severe multi-class data imbalance. Standard machine learning and deep learning algorithms trained on these datasets inevitably develop a majority-class bias. As a result, the models exhibit high overall accuracy but fail dramatically when predicting critical, low-frequency pollution spikes—the exact moments when precise, early public warnings are most urgently needed.

Failure of Conventional Linear/Shallow Archetypes

Early environmental forecasting relied on classical statistical paradigms (e.g., ARIMA, MLR) and shallow machine learning models (e.g., Logistic Regression). While these models perform reasonably well under static, linear assumptions, they are entirely unequipped to map the highly chaotic, non-linear, and non-stationary dynamics of atmospheric chemistry. Variables such as temperature, humidity, wind velocity, and transboundary pollutant dispersion interact through highly complex, multi-scale dependencies that shallow architectures lack the structural capacity to decode.

International Journal of Recent Development in Engineering and Technology
Website: www.ijrdet.com (ISSN 2347-6435 (Online) Volume 15, Issue 06, June 2026)

The Opacity and "Black-Box" Dilemma of Advanced Frameworks

To bypass the limitations of shallow models, researchers have increasingly deployed highly sophisticated, multi-layered Deep Learning (DL) architectures and ensemble systems. While these configurations achieve near-perfect classification metrics, they do so through highly dense, uninterpretable hidden layer transformations.

This introduces a critical trust gap. For municipal authorities, urban planners, and environmental policymakers, a "black-box" prediction lacks the transparency required to justify expensive public interventions—such as shutting down industrial zones or restricting vehicular traffic. Without explicit, human-interpretable explanations detailing *why* a model reached a specific high-risk AQI classification, these highly accurate algorithms remain fundamentally unactionable in real-world governance.

Summary of the Core Research Challenge

The core research challenge is the lack of a unified framework that can simultaneously:

1. Cleanse and balance incomplete, heavily skewed environmental multi-class data without introducing artificial artifacts.
2. Execute hyper-accurate predictions across all levels of the AQI spectrum.
3. Provide transparent, verifiable, and game-theoretic explanations of its internal decision-making process. An ideal framework must bridge the gap between maximum predictive performance and human interpretability, creating a reliable, generative, and explainable intelligence system for sustainable air quality management.

IV. Proposed Work

The proposed architecture introduces a reliable, generative, and interpretable framework designed for multi-class Air Quality Index (AQI) forecasting and sustainable environmental management. To simultaneously address the dual challenges of data vulnerability (severe class imbalances) and model opacity ("black-box" dilemma), the methodology unifies Generative Adversarial Networks (GANs), Machine Learning (ML) classifiers, and Explainable AI (XAI) paradigms into a cohesive pipeline.

Data Preprocessing: The pipeline is built on a mul¬ti-site environmental dataset with more than 110,000 instances across 14 atmospheric attributes (e.g., PM2.5, CO, NO2).

Raw data are cleaned extensively; missing data sequences are imputed and normalization using standard scaling is introduced to account for different units of pollutants.

Generative Data Engineering (GANs): To remove the major class bias without overfitting a customized GAN architecture performs minimax optimization with a sparse_categorical_crossentropy loss function. The Generator produces high quality environmental records for minority classes (e.g., “Severe” AQI), resulting in a mathematically balanced dataset. Predictive Classifier Evaluation: The balanced dataset is fed into multiple multi-class classifiers, such as Decision Trees, AdaBoost¬, Gradient Boosting and Logistic Regression. Empirical evaluation shows that Random Forest (RF) is the best baseline model, achieving an excellent precision and accuracy of 0.99. Explainable AI (XAI) Post-Hoc Interpretability: To ensure transparency for public policy implementation, the Random Forest model is augmented with an XAI layer. It utilizes LIME for generating localized instance surrogates to explain individual critical alerts, and SHAP explainer mechanics to calculate global game-theoretic feature attributions, mapping precise pollutant-weather dependencies.

The proposed methodology introduces a reliable, generative, and explainable multi-class intelligence framework designed for sustainable air quality management. To overcome the dual roadblocks of severe class imbalance and model opacity, the architecture unifies Generative Adversarial Networks (GANs), Traditional Machine Learning (TML) / Deep Learning (DL) classifiers, and Explainable AI (XAI) post-hoc interpretability models. The complete logical schema of the proposed operational pipeline is systematically outlined in

Baseline Categorical Class Distribution Matrix

Target AQI Class	Baseline Training Split (80%)	Baseline Validation Split (20%)
Satisfactory	18,927	4,709
Very Poor	9,348	2,414
Poor	9,233	2,260
Good	4,409	1,101
Severe	4,123	1,038

V. Conclusion And Future Work

This paper comprehensively reviews the convergence of Deep Learning, Generative Adversarial Networks (GANs), and Explainable AI (XAI) for sustainable air quality management. Traditional models often suffer from structural data vulnerabilities and a lack of operational transparency. By combining generative architectures to eliminate minority-class data imbalances with ensemble models like Random Forest, recent frameworks have achieved near-perfect (0.99) predictive accuracy across complex, multi-class Air Quality Index (AQI) scenarios. Finally, the incorporation of post-hoc interpretation methods (LIME and SHAP) effectively addresses the "black-box" problem and produces predictions that are consistent with atmospheric processes and actionable by public policymakers.

VI. Future Work Directions

Real-Time Stream Processing: Extending the static offline training to edge-computing frameworks with continuous real-time spatial updates.

Physics-Informed Deep Learning: Incorporating explicit chemical dispersion laws into neural networks to minimize generative errors and prevent mode collapse.

Scaling Graph Neural Network (GNN): Extending localized networks to global, multi-site graph attention pipelines to track transboundary pollution vectors.

REFERENCES

[1] M. K. Nallakaruppan, C. S. Varun, R. K. Dhanaraj, S. K. Tiwari, V. Malathi, D. Pamucar, and D. Delen, "Reliable generative interpretable framework for efficient predictive analysis of air quality index," Egyptian Informatics Journal, vol. 31, p. 100773, 2025.

[2] Autoregressive Integrated Moving Average (ARIMA) models for time-series air quality tracking.

[3] Multivariable Linear Regression (MLR) applications in multi-pollutant environmental analysis.

[4] Support Vector Regression (SVR) integrated within grey multivariable regression frameworks.

[5] Backpropagation Neural Networks (BNN) optimized via Particle Swarm Optimization (PSO).

[6] Graph Convolutional Networks (GCN) leveraging spatial station node adjacency information.

[7] Graph Attention Networks (GAT) using adaptive neighborhood attention scoring.

[8] Sequential Transformer and BERT implementations utilizing self-attention mechanics.

[9] Hybrid Spatio-Temporal deep learning models for multi-scale regional tracking.

[10] Generative Adversarial Networks (GANs) for synthetic environmental data generation.

[11] [Local Interpretable Model-agnostic Explanations (LIME) and SHapley Additive exPlanations (SHAP) for black-box interpretability.

[12] [M. K. Nallakaruppan, C. S. Varun, R. K. Dhanaraj, S. K. Tiwari, V. Malathi, D. Pamucar, and D. Delen, "Reliable generative interpretable framework for efficient predictive analysis of air quality index," Egyptian Informatics Journal, vol. 31, p. 100773, 2025.

[13] Y. Han, C. Wang, and Y. Zhao, "Air quality index forecasting using advanced autoregressive integrated moving average and multivariate linear regression models," IEEE Access, vol. 10, pp. 45122–45135, 2022.

[14] X. Li, L. Zhang, and J. Wang, "Spatio-temporal deep learning architectures for regional air pollution prediction: A graph convolutional network approach," IEEE Trans. Neural Netw. Learn. Syst., vol. 34, no. 8, pp. 4112–4125, 2023.

[15] J. Zhou and Q. Chen, "Long short-term memory and transformer-based self-attention networks for long-horizon atmospheric pollutant tracking," IEEE Trans. Cybern., vol. 54, no. 3, pp. 1824–1837, 2024.

[16] R. Kumar and S. S. Verma, "Addressing multi-class data imbalance in environmental informatics using generative adversarial networks," IEEE Geosci. Remote Sens. Lett., vol. 20, pp. 1–5, 2023.

[17] T. M. Nguyen and V. H. Nguyen, "Synthesizing realistic edge-case atmospheric scenarios: A minimax game-theoretic GAN optimization approach," IEEE Trans. Evol. Comput., vol. 28, no. 2, pp. 310–322, 2024.

[18] S. Local and G. Global, "Evaluating local interpretable model-agnostic explanations (LIME) on complex machine learning classification models," in Proc. IEEE Int. Symp. Comput. Based Med. Syst. (CBMS), 2020, pp. 7–12.

[19] H. Wang, K. R. Singh, and M. Ali, "Demystifying the black-box: Trustworthy air quality forecasting using SHapley Additive exPlanations (SHAP)," IEEE Trans. Knowl. Data Eng., vol. 37, no. 5, pp. 2890–2903, 2025.