International Journal of Environmental Sciences
ISSN: 2229-7359
Vol. 11 No. 8, 2025
https://www.theaspd.com/ijes.php

Air Quality Prediction: A Systematic Review Of Traditional Methods And Emerging Hybrid Frameworks

Abirami R[1], Mani P[2]
[1]Research Scholar, School of Computer Science Engineering and Information Systems, Vellore Institute of Technology, Vellore, Tamil Nadu, India.
Email address: abirami.r2024@vitstudent.ac.in
[2]Research Guide, School of Computer Science Engineering and Information Systems, Vellore Institute of Technology, Vellore, Tamil Nadu, India.

Abstract
Air quality is a major global challenge with pollutants like $PM_{2.5}$, PM_{10}, NO_2, SO_2, CO, and O_3 being a threat (or upset) to human health, the environment, and ultimately quality of life. The short-term forecasting of pollutant concentrations (for example, $PM_{2.5}$ and PM_{10}) can be immensely useful for early warning systems, regulatory actions, and public advisories. Fore- cast modeling has drastically evolved in the last twenty years. Although traditional modeling strategies such as linear autoregressive integrated moving average (ARIMA), multiple linear re- gression (MLR), and generalized additive models (GAM) are easy to use and interpret, they are generally limited in producing estimates of nonlinear interactions, structural breaks, and long- range dependence. Researchers have looked extensively to utilize machine learning and deep learning algorithms (i.e. Random Forests, Support Vector Machines (SVM), Long Short-Term Memory (LSTM), Gated Recurrent Units (GRU), Convolutional Neural Networks (CNN), and transformers) in recent years in an effort to overcome these limitations. These methods im- prove spatiotemporal accuracy compared to traditional statistical methods but they do not offer significant improvement in computational burden, interpretability, or timely deployment. The emergence of hybrid models, which utilize statistical, machine learning, and deep learning-type models with a variety of data sources, such as meteorological, satellite, and IoT-based sensing or other sensor data, has improved robustness and scalability. The review presents the evolution of air quality forecasting techniques, contrasts advantages and disadvantages, and underlines implications for public health and policy.
Keywords: *ARIMA, Deep Learning, LSTM, GRU, Machine Learning, Transformer, GNN.*

INTRODUCTION

Air quality is of great importance for human health, ecosystem health and functioning, and the future sustainability of life on the planet. There are five major pollutants considered critical for air quality, because of their impacts on human health and/or ecosystem health and human activities such as agriculture and forests; those pollutants include fine particles (e.g. $PM_{2.5}$ and PM_{10}), NO_2, SO_2, CO and O_3. The source of these pollutants is mostly from transportation, industry and urbanization. These effects can be chronic especially for cardiovascular and res- piratory disease and mortality and continue to diminish the quality of the environment we live and work in. The growing land use change, ongoing urbanization, and expanding economic development are responsible for increasing air quality degradation and highlight the clear need for integrated and real-time monitoring systems.

The World Health Organization (WHO) and the International Agency for Research on Cancer (IARC) classified ambient air pollution as a human carcinogen in 2013 to emphasize the impor- tance of these risks and to make emission monitoring and predictive models more imperative for public health and policy intervention towards sustainability.

These initial predictive methods relied mainly on statistical models, such as Autoregressive Integrated Moving Average (ARIMA), Multiple Linear Regression (MLR), and Generalized Additive Models (GAM). While they were easy to apply and interpret, the earlier approaches often could not account for nonlinear relationships, sudden changes, and long-distance depen- dencies [1]. To overcome these limitations, ML methods have been adopted - if you are familiar with Random Forest (RF), Gradient Boosting Decision Trees (GBDT) and Support Vector Ma- chines (SVM). These approaches improved pollutant–meteorology coupling but continued to depend on manually engineered features, limiting adaptability in dynamic settings [2].

The rise of deep learning (DL) came with more robust modelling capabilities for complex time- series and spatial structures. It was found that recurrent neural networks (RNNs) such as Long Short-Term Memory

(LSTM) or Gated Recurrent Units (GRU) were useful for learning long- term dependencies for pollutant time series [3]. Convolutional neural networks (CNNs) and Graph Neural Networks (GNNs) enhanced spatiotemporal forecasting in a network of sensors [4][5], while transformer-based architectures (e.g., Informer) recently showed robustness and efficiency when predicting long sequences [6]. Despite these advances, DL models typically maintain expensive computational costs, are hard to interpret and remain problematic for ex- traction in real time [7].

This progression is illustrated in figure 1, indicating the departure from statistical methods to ML, DL, and most recently hybrid and uncertainty-aware approaches. New methods utilize the triad of statistical, ML, and DL approaches by using convergently heterogeneous data sources including meteorological data, satellite imagery, and IoT-enabled sensor networks. In the same timeframe, cloud-native AI technologies, edge computing, and real-time streaming pipelines are being innovated to improve use-case scalability, resilience, and investment to make inten- tional deployments and AI increasingly efficient.

Old atmospheric dispersion models (e.g. Gaussian plume and puff models) are still relevant and provide theoretical basis for pollutant transport by describing wind speed, turbulence, and atmospheric stability parameters. However, their limited capability of being able to repre- sent chaotic variability in urban areas has created a space for physics-informed hybrid AI ap- proaches that combine a physical dynamic to data-driven learning to improve prediction accu- racy and reliability.

This review provides a comprehensive synthesis of air quality forecasting methodologies from a historical and forward-looking perspective. It maps the evolution from statistical models to ML and DL, discusses the development of hybrid and uncertainty-aware systems, and com- bines findings from dispersion theory. Further, it identifies strengths and weaknesses, datasets, measures of success, policy implications, and areas of research gaps. Separating it from others, this review focuses on recent developments between 2021 and 2025, with special emphasis on hybrid and uncertainty-aware frameworks, thus presenting itself as a worthy reference material for researchers and practitioners.

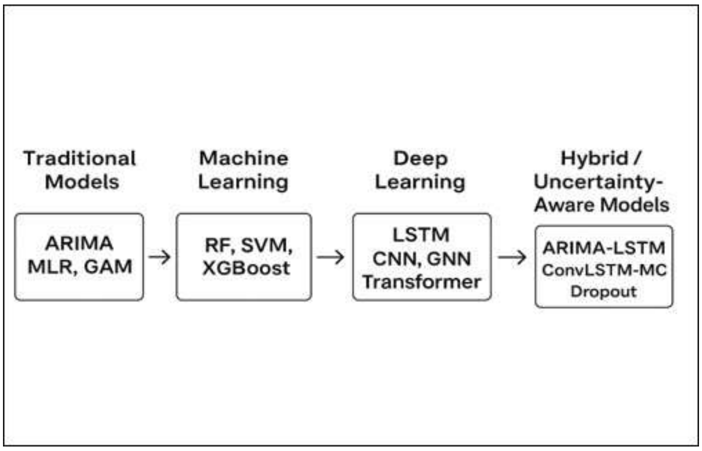

Figure 1: Progression of air quality forecasting methodologies from statistical approaches to machine learning, deep learning, and hybrid frameworks.

METHODOLOGY OF REVIEW

To ensure comprehensive and systematic coverage of the literature, this review adopted a struc- tured approach for identifying, screening, and analyzing relevant studies. Major academic databases, including Web of Science, Scopus, IEEE Xplore, ScienceDirect, and SpringerLink, were searched for publications from 2021 to mid-2025, while earlier foundational works were considered where necessary.

The search employed keywords such as "air quality forecasting," "air pollution prediction," "$PM_{2.5}$ forecasting," "machine learning," "deep learning," "transformer models," "hybrid frame- works," and "uncertainty-aware forecasting."

An initial pool of more than 120 publications was retrieved and screened by examining titles and abstracts. Studies were included if they:

- Presented original forecasting methodologies,
- reported quantitative performance metrics or comparative evaluations, or
- introduced hybrid or uncertainty-aware frameworks.

Papers were excluded if they were non-peer-reviewed, purely descriptive, or lacked method- ological contributions. Following this screening process, approximately 40 full-text articles were assessed for eligibility, and finally, 27 peer-reviewed studies were included in this review. The overall process of identification, screening, and inclusion is illustrated in figure 2.

The selected works were categorized into four broad groups:

i. Traditional statistical approaches,
ii. Machine learning methods,
iii. Deep learning architectures, and
iv. Hybrid or uncertainty-aware frameworks.

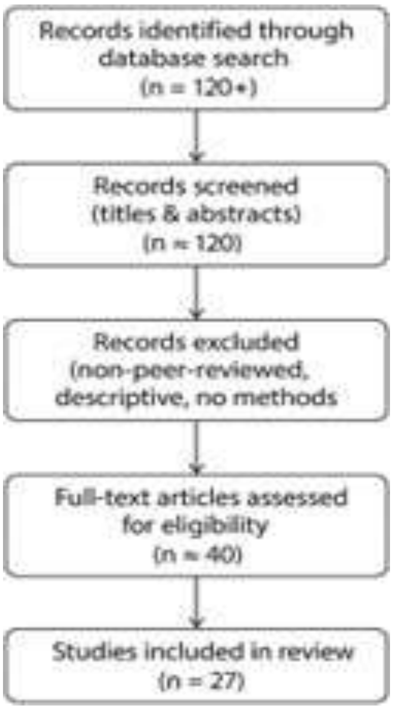

Figure 2: Flow diagram of the review methodology from identification to inclusion of 27 studies.

Traditional and Machine Learning Approaches

Traditional statistical models such as Autoregressive Integrated Moving Average (ARIMA), Multiple Linear Regression (MLR), and Generalized Additive Models (GAM) have been es- tablished for air quality in particular due to their inherent ability to be easily understood, ease of interpretation of results, and computational efficiency. They are an effective, short-term pre- diction method and provide useful baseline models when analyzing time series. However, they ignore various nonlinear dependencies, long-term temporal trends, and abrupt changes in en- vironmental factors that could potentially cause models to misrepresent forecast accuracy ([1]; [2]).

To overcome these constraints, machine learning (ML) algorithms have gaining prominence. Techniques like support vector machines (SVM), random forests (RF), and gradient-boosting decision trees (GBDT) offer more flexibility in modeling pollutant–meteorology interrelation- ships and can process high-dimensional data better than statistical techniques. For example, RF and XGBoost were found to exceed ARIMA in forecasting $PM_{2.5}$ and NO_2 concentrations in Indian and Chinese cities ([2]; [8]). However, ML approaches are largely dependent on feature engineering, which limits their scalability and tends to miss long-term sequential dependencies embedded in air quality data ([9];[10]). The important features, strengths, and limitations of the statistical methods as well as ML methods are presented in table 1.

Statistical models are still useful in terms of interpretability and computational efficiency for making predictions. In contrast, machine learning (ML) methods offer higher accuracy and flexibility in accounting for nonlinear relationships. Both types of models have their own limi- tations, which has contributed to the uptake of deep learning models for air quality forecasting.

Table 1: Comparison of statistical and machine learning approaches for air quality fore- casting, highlighting their typical models, strengths, limitations, and common use cases.

Approach	Example Models	Strengths	Limitations	Typical Use Case
Statistical Mod- els	ARIMA, MLR, GAM	Simple, interpretable, computationally efficient	Poor with nonlinearities; weak in long-term prediction	Baseline fore- casts; studies requiring interpretability
Machine Learn- ing Models	SVM, RF, GBDT	Capturenonlinearities; effective with high dimensional data; higher accuracy	Dependence on Feature engineering; weak with temporal dependencies	Complexpollutant–meteorology modeling; classification and short-term forecasts

Deep Learning–Based Approaches

Deep learning (DL) methods are presenting significant advancements to the field of air quality forecasting by eliminating some of the limitations of statistical and machine learning meth- ods. Rather than relying on features as the ML methods do, DL models find structures and hierarchies of spatial and temporal features from the raw data, which gives them advantages in modelling nonlinear and long-term dependencies.

Recurrent Neural Networks (RNNs), namely Long Short-Term Memory (LSTM) and Gated Recurrent Unit (GRU) models, have demonstrated their capability to learn sequential relation- ships in pollutants across time-series data. [3], for instance, identified that applying LSTM with hyper-heuristic optimization led to a considerable decline in RMSE (Root Mean Squared Error) of multi-hour predictions compared to ARIMA, over China, India, and the USA. Similarly,[11] reported that a Bi-GRU combined with GMVAE outperformed conventional LSTM in multi- variate pollutant prediction in Beijing.

Spatiotemporal models such as Convolutional Neural Networks (CNNs) have been employed to learn spatial correlations among pollutants, while hybrid CNN–attention frameworks have been successfully used for short-term $PM_{2.5}$prediction in highly polluted regions like New Delhi [4]. Graph Neural Networks (GNNs) extend this capacity by modelling non-Euclidean relation- ships, particularly in networks of air quality monitoring stations.[5] applied a GNN–Transformer framework to Chinese cities, achieving higher spatiotemporal accuracy than CNN–LSTM mod- els.

More recently, Transformer-based architectures such as Informer have been explored for long- sequence forecasting. [6] demonstrated that an Informer combined with XGBoost achieved R^2 = 0.961 for $PM_{2.5}$ prediction in Nanjing, significantly outperforming LSTM. Similarly, [7] used ConvLSTM with Monte Carlo Dropout to incorporate uncertainty estimation, enhancing both reliability and robustness in Beijing's air quality predictions.

Recent work has also extended these frameworks to hybrid spatio-temporal deep learning. For instance, a CNN–BiLSTM model was recently proposed for simultaneous $PM_{2.5}$ and O_3 fore- casting, achieving improved accuracy across multiple urban environments [12].Collectively, these studies highlight the way that deep learning models present considerable enhancement in prediction accuracy and generalization. Challenges remain the recognized limitations of com- putational cost, lack of interpretability, and size of dataset, underscoring the need for hybrid and uncertainty-aware forecasting frameworks. The comparative characteristics of different deep learning architectures in air quality forecasting are summarized in table 2.

Hybrid and Uncertainty-Aware Forecasting Frameworks

Hybrid models have emerged as a promising direction in air quality forecasting, aiming to combine the strengths of statistical, machine learning, and deep learning methods. The gen- eral workflow of these forecasting systems, from data acquisition and pre-processing to model

Table 2: Comparative characteristics of deep learning architectures (RNN, CNN, GNN, Transformer, Hybrid) for air quality forecasting.

Aspect	GNN (Graph Neural Network)	RNN (Recurrent Neural Network)	CNN(Convolutional Neural Network)	Transformer	Hybrid Models
Data Type	Graph-structured data (e.g., road networks, social graphs, molecules)	Sequential data (e.g., time series, speech, stock data)	Grid-like structured data (e.g., images spectrograms, 1D signals)	Long sequential data (e.g., pollutant time-series, language)	Multi-source data (e.g., sensor + satellite + mete- orology)
Core Operation	Message passing between nodes for relational learning	Temporals recurrence using hidden states	Convolutional filters to capture local spatial patterns	Attention mechanism to capture global dependencies	Combines strengths of multiple models (e.g., ARIMA + LSTM, CNN + GRU)
Temporal Dependency	Not inherently temporal; needs extensions for dynamic graphs	Captures short- and long-term dependencies	Limitedunless Integrated with RNN/temporal models	Excels at long-Range temporal dependencies	Flexible; leverages sequential + spatial + external features
Spatial Relationship	Excels at irregular/non-Euclidean spatial patterns	Weak inspatial modelling	Strong at Local spatial/hierarchical patterns	Indirect,unless Extended with spatial encoders	Captures both spatial and temporal dependencies effectively
Use Cases	Traffic flow, molecule modelling, social net- works, recommendation systems	Air quality time- series forecasting, speech recognition	Image-based pollutant dispersion, pattern extraction	Long-sequence AQ forecasting, spatiotemporal prediction	Air quality index (AQI) prediction, real-time forecasting
Model Complexity	Moderate to high; expensive for large graphs	Moderate; issues With vanishing gradients	Moderate GPU-friendly	High; requires large data and compute	Higher depends on model combination
Learning Structure	Learns from entities (nodes, edges)	Learns sequential dependencies	Learns hierarchical spatial structures	Learns contextual global patterns	Learns multi-paradigm patterns from diverse sources
Scalability	Expensive for large graphs	Harder to scale dueto sequential nature	Scales well with GPUs	Scaleswell but data-hungry	Moderate; depends on Architecture and data pipeline

training and prediction, is illustrated in figure 3. Such frameworks are particularly effective at capturing both linear and nonlinear dependencies while addressing the limitations of individ- ual models. For instance,[1] integrated Empirical Mode Decomposition (EMD) with ARIMA to mitigate error accumulation in long-term $PM_{2.5}$ prediction in Tehran, achieving greater sta- bility than ARIMA alone. Similarly,[2] combined Random Forests, XGBoost, and ARIMA across 23 Indian cities, with RF showing superior multi-step AQI prediction performance. Hy-

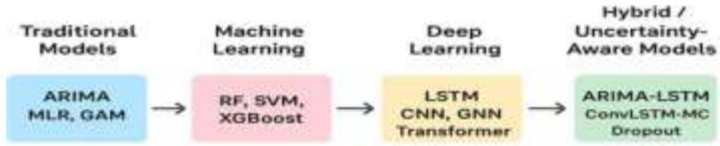

Figure 3: General framework of air quality prediction, showing data sources, prepro- cessing, modelling approaches (statistical, machine learning, deep learning, hybrid), and forecast output.

brid deep learning approaches have also demonstrated significant improvements. [13] used a decomposition–LSTM framework in Belfast, which enhanced both interpretability and ro- bustness compared to conventional deep learning models. In another study, [6] employed an Informer–XGBoost hybrid that achieved an R^2 of 0.961 for $PM_{2.5}$ in Nanjing, surpassing stan- dalone LSTM performance.

Uncertainty-aware frameworks represent a parallel advancement, emphasizing not only ac- curacy but also the reliability of forecasts. [7] implemented ConvLSTM with Monte Carlo Dropout to quantify prediction uncertainty and provide confidence intervals with forecasts for Beijing, and other recent studies have also developed probabilistic forecasting approaches, in- cluding Bayesian neural networks and quantile regression, to better inform policymakers about the uncertainty risk of extreme pollution events. Hybridization of CNN and recurrent networks has also proven effective.[14]developed a CNN–RNN hybrid framework that outperformed conventional deep learning models in short-term $PM_{2.5}$ forecasting.

In recent empirical studies in various regions, hybrid and uncertainty-aware frameworks have been evaluated; these methods and findings are listed in table 3. In general, these models ex- hibit improvements in prediction robustness, interpretability, and flexibility. As such, they will continue to be a relevant option for operational deployment in air quality monitoring systems. However, the computational cost and complexity of these models indicate a need for more optimization for real-time scalability.

Air Quality Datasets

Reliable datasets form the backbone of air quality forecasting studies, as they provide the basis on which predictive models can be derived, trained, validated, and tested. Datasets would usu- ally include measurements of pollutant concentrations, meteorological variables, traffic emis- sions, and satellite observations of atmospheric variables. Reliable datasets can be classified as ground-based monitoring datasets, satellite datasets, or multi-source or hybrid datasets.

table 4 summarizes key datasets, their sources, locations, and applications in recent forecasting studies.

Ground-Based Monitoring Datasets

Ground station data remain the most used resource for air quality prediction.figure 4. Workflow of ground monitoring stations for air quality forecasting, showing the process of measuring pol- lutants (e.g., $PM_{2.5}$, PM_{10}, NO_2) and translating them into pollutant concentrations and AQI dashboards. For instance, [2] utilized CPCB datasets covering 23 Indian cities from 2015–2020 to predict $PM_{2.5}$, NO_2, and SO_2 concentrations, applying RF and XGBoost models. [19] used continuous air quality monitoring station (CAAQMS) data from Jaipur to evaluate models in- cluding MLR, SVR, ANN, GRU, and CNN for $PM_{2.5}$ forecasting. Likewise, [5] utilized mon- itoring station data from Beijing–Tianjin–Hebei (BTH) to study hybrid models with weather features, reporting an improved forecasting performance.

Strength: Ground-based datasets can be trusted and interpreted.

Table 3: Recent studies on hybrid and uncertainty-aware frameworks for air quality fore- casting, including study areas, methods, and key performance findings.

Year	Author(s)	Study Area	Method	Key Findings
2024	P. A. Gupta [3]	China, India, USA	LSTM + Hyper-Heuristic Optimization	Reduced RMSE by 75% compared to ARIMA for multi-hour forecasts.
2023	L. Fang et al. [15]	Kiev, Ukraine	LSTM + EFO (satellite in- tegration)	Improved $PM_{2.5}$ predic- tion using satellite NO_2 and CO data.
2024	J. Wang et al. [11]	Beijing, China	Bi-GRU + GMVAE	Outperformed LSTM in multivariate pollutant fore- casting.
2023	C. Zhang & B. Han [6]	Nanjing, China	Informer + XGBoost	Achieved R^2 = 0.961 for $PM_{2.5}$, surpassing LSTM.
2023	H. Lin et al. [5]	Chinese cities	GNN + Transformer	Higher spatiotemporal ac- curacy than CNN–LSTM.
2022	M. Xu & J. Yang [4]	New Delhi, India	CNN + Attention	Outperformed LSTM in short-term $PM_{2.5}$ fore- casts.
2021	M. Lee & K. Chen [7]	Beijing, China	ConvLSTM + MC Dropout	Quantified uncertainty, improved reliability of forecasts.
2024	S. Gupta et al. [9]	Global smart cities	RF, XGBoost, CNN	CNN best captured spatial pollution patterns.
2023	K. Y. Chen [16]	London, UK	ELM + GA	LSTM more robust in volatile environments.
2023	K. Fernandez et al. [17]	Jakarta, Indonesia	XGBoost, AdaBoost, LightGBM	XGBoost achieved highest $PM_{2.5}$ classification accu- racy.
2024	H. Wu et al. [13]	Belfast, N. Ireland	Decomposition + LSTM	Enhanced interpretability and robustness vs. conven- tional DL.
2019	S. Hu & R. Nelson [18]	Beijing, China	SVM + Feature Selection	Effective for short-term

				NO_2 prediction with tun- ing.
2024	Y. Fang & P. Sen [8]	Indian smart cities	GBDT, SVM, XGBoost	XGBoost performed best for AQI prediction.
2024	Y. Liang et al. [1]	Tehran, Iran	EMD + ARIMA	Reduced error accumula- tion, improved long-term $PM_{2.5}$ stability.
2023	P. K. Sharma et al. [2]	23 Indian cities	RF, XGBoost, ARIMA	RF achieved best multi- step AQI prediction.

Limitation: Limited spatial coverage, high infrastructure cost, and potential missing values in harsh environments.

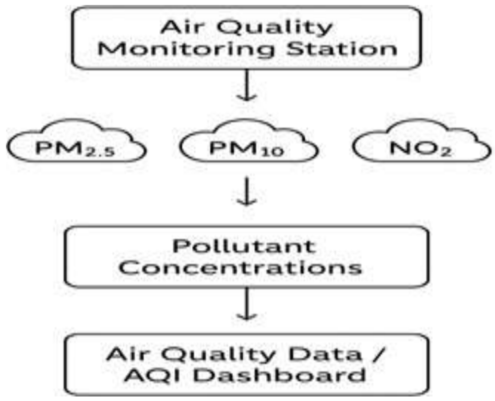

Figure 4: Workflow of ground monitoring stations for air quality forecasting, showing pollutant measurement, data processing, and AQI reporting.

1.1 Satellite-Based Datasets

Satellite observations provide valuable spatial coverage in regions with sparse ground moni- toring. [8] combined CPCB data with satellite-derived pollutant information to assess $PM_{2.5}$ and CO impacts in Delhi. Likewise, [20] integrated Copernicus satellite products, LiDAR, and weather data in Sweden and the Netherlands for advanced forecasting tasks. These datasets are especially effective for capturing large-scale pollution transport phenomena, complementing localized ground-based measurements.

Strength: Wide spatial coverage, useful for global and regional-scale studies.

Limitation: Lower temporal resolution, dependence on cloud-free conditions, and indirect estimation of ground-level concentrations.

1.2 Multi-Source and Hybrid Datasets

Combining ground, satellite, and meteorological data has become a common practice to en- hance predictive performance. [21] used CPCB and Kaggle datasets along with optimization techniques for AQI improvement across Indian cities. [17] employed Jakarta's AQ network with hourly AQ and weather data for $PM_{2.5}$ classification, highlighting the benefits of integrat- ing multiple sources. [1] further explored global IoT sensor networks for real-time AI-based forecasting in smart cities.

Strength: Rich and diverse, capturing pollutant dynamics from multiple perspectives. **Limitation:** Integration challenges, differences in spatial/temporal resolution, and potential calibration errors across sources.

2 Evaluation Metrics

To evaluate air quality forecasting performance, a range of robust and diverse metrics must be employed to assess accuracy, generalization, and reliability. Depending on whether the task chosen will be a regression of pollutant concentration prediction or a classification of the AQI (Air Quality Index) categories, a combination of items will need to be figured out. In figure 5, we observe the classification of pollutants (such as $PM_{2.5}$, PM_{10}, and NO_2) based on designated ”safe” and ”unsafe” thresholds for the AQI score. It can illustrate that the thresholds to quantify and diagnose these models have an observable visual distinction separating safe from unsafe. Additionally, the thresholds of accuracy and performance metrics are suitable in that they accurately represent an important matter of public health.

2.1 Traditional Statistical Metrics

Root Mean Square Error (RMSE), Mean Absolute Error (MAE), and the Coefficient of Deter- mination (R^2) are the most widely adopted metrics. These quantify prediction error magnitude

Table 4: Representative air quality datasets used in forecasting studies, with locations, periods, data sources, and applications.

S.No	Title of the Paper	Location	Period	Data Source	Brief Information
1	A. Smith et al. [22]	Urban Areas	2015–2022	Ground-Based Sensors	Real-time pollutant data ($PM_{2.5}$, PM_{10}, NO_2, CO, SO_2, O_3) for im- proved prediction.
2	D. Williams et al. [23]	London, UK	2016–2021	Sensors + Weather	Urban AQ with meteorological fea- tures for better input.
3	S. Gupta et al. [9]	Sweden	1990–2023	CO_2 & Energy Use	Long-term analysis of energy im- pact on AQ over 34 years.
4	J. Liu et al. [21]	Indian Cities	2015–2023	CPCB, Kaggle	ML + optimization for AQI im- provement across Indian cities.
5	P. K. Sharma et al. [2]	23 Indian Cities	2015–2020	CPCB Dataset	RF, XGBoost used to predict $PM_{2.5}$, NO_2, SO_2 levels.
6	X. Zhao et al. [19]	Jaipur, India	2019–2023	CAAQMS (CPCB)	Compared MLR, SVR, ANN, GRU, CNN for $PM_{2.5}$.
7	H. Lin et al. [5]	BTH, China	2015–2022	AQ Stations	Integrated weather and AQ for hy- brid model performance.
8	K. Fernandez et al. [17]	Jakarta, Indone- sia	2016–2021	Jakarta AQ Network	Hourly AQ + weather data used for $PM_{2.5}$ classification.
9	M. R. Hassan et al. [20]	Sweden, Nether- lands	2017–2023	OSM, Copernicus	LiDAR, satellite, and weather inte- gration for AQ forecasting.
10	L. Jiang & T. Brown [24]	Global Cities	2015–2024	IoT Sensors	Real-time data for AI- based AQ prediction in smart cities.
11	K. Y. Chen [16]	China	2019–2021	AQI GitHub Repo	GA-KELM model surpassing CMAQ, SVR, DBN-BP in

					AQI.
12	R. Mohan [25]	Madrid, Spain	Jan–Jun2019, 2022	Madrid Open Data	A3T-GCN using AQ, traffic, and weather data; beat LSTM, GRU.
13	S. Kumar & P. Das [10]	Pune, India	2019	Smart City Dataset	Compared RF, LR, DT; DT had best results.
14	Z. Lin et al. [26]	Tehran, Iran	2001–2021	Tehran AQCC	MLR and GAM performed best; meteorology a key driver.
15	Y. Fang & P. Sen [8]	Delhi, India	2016–2022	CPCB	RF, XGBoost found $PM_{2.5}$ and CO most impactful.

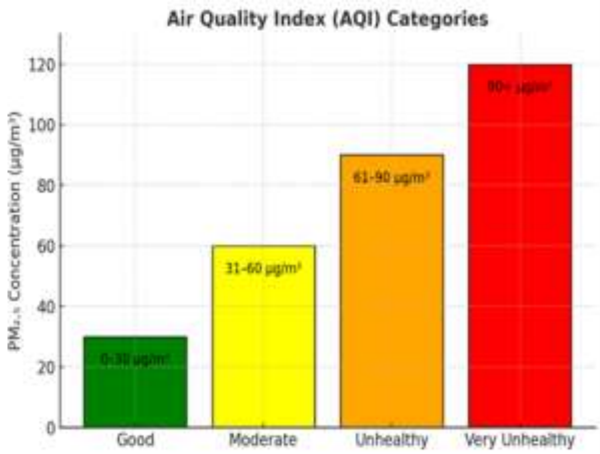

Figure 5: Classification of pollutants into safe and unsafe AQI categories, illustrating threshold-based evaluation.

and goodness of fit. For example, [2] used RMSE and MAE to evaluate RF and XGBoost pre- dictions for Indian cities, while [1] employed R^2 to assess improvements of EMD–ARIMA over traditional ARIMA in Tehran. [27] RAQP outperformed direct forecasting methods, achieving lower RMSE values and higher PLCC or R^2 scores.

2.2 Percentage-Based Metrics

Relative measures such as Symmetric Mean Absolute Percentage Error (SMAPE) and Mean Absolute Percentage Error (MAPE) normalize errors relative to actual pollutant levels, mak- ing them particularly useful when comparing across datasets of varying scales. [11] applied SMAPE in Beijing while comparing Bi-GRU + GMVAE with LSTM models. Similarly, [15] utilized percentage-based metrics to evaluate LSTM–EFO frameworks in Kiev, demonstrating improved $PM_{2.5}$ accuracy.

2.3 Classification Metrics for AQI Categories

Models that make predictions of discrete AQI categories instead of the concentration levels of pollutants utilize classification-based metrics (F1-score, precision, recall, and Matthews Cor- relation Coefficient [MCC]). [17] used accuracy and F1-scores to assess XGBoost, AdaBoost, and LightGBM for $PM_{2.5}$ classification in Jakarta. Likewise, [10] evaluated RF, LR, and DT in Pune utilizing classification metrics.

2.4 Uncertainty-Aware Metrics

As probabilistic and hybrid models increase, uncertainty-aware metrics are becoming more significant. Among these are Prediction Interval Coverage Probability (PICP) and Mean Pre- diction Interval Width (MPIW), which ascertain confidence interval representation of true vari- ability. [7] used Monte Carlo

Dropout in ConvLSTM on Beijing, showing better uncertainty quantification. [13] also emphasized interpretability by employing decomposition-based LSTM with reliability evaluations.

2.5 Benchmarking and Comparative Evaluation

Recent studies emphasize benchmarking new methods against traditional baselines. For in- stance, [9] demonstrated that their LSTM + Hyper Heuristic Multi-Chain model reduced RMSE by 75% compared to ARIMA across China, India, and the USA. [6] compared Informer + XG- Boost with LSTM in Nanjing, reporting $R^2 = 0.961$ for $PM_{2.5}$. Such comparative evaluations ensure fair assessment of methodological improvements across models and datasets.

3 CHALLENGES AND OPPORTUNITIES

Even with exciting advances in predictive modelling, air quality forecasting is still struggling with persistent obstacles that impact the number of scaling and deployable projects. One of those barriers is data quality: a considerable amount of the air quality data is plagued with missing values, different sampling frequencies, and sensor errors related to calibration. For in- stance, [22] highlighted how urban ground-sensor datasets often contain noisy and incomplete entries, requiring preprocessing and imputation for effective forecasting. Similarly, [9] noted that long-term datasets from Sweden showed variability due to inconsistent CO_2 and energy use records, complicating analysis.

- Model interpretability is another pressing concern. While deep learning architectures such as CNNs, GNNs, and Transformers offer high accuracy, their "black box" nature limits adoption by policymakers who require transparent reasoning behind predictions.
  [5] demonstrated that hybrid models integrating meteorological variables improved ac- curacy yet still lacked interpretability for end-users. This gap has motivated increasing interest in explainable AI approaches, such as decomposition frameworks [13] and atten- tion mechanisms [4], which enhance transparency while retaining predictive power.
- Computational cost and scalability: Also present significant barriers. Many high-performing models like Informer + XGBoost [6] and ConvLSTM with uncertainty quantification [7] require considerable GPU resources, making real-time deployments for low-resource en- vironments or IoT-based applications challenging. Research on lightweight frameworks that are edge-computing ready provides other promising routes to consider for reducing latency while maintaining accuracy.
- Transferability and generalization: remain unresolved challenges. Models trained on a specific region often fail to generalize to other geographies with different climatic or emission profiles. For example, [8] showed that models calibrated on Delhi data un- derperformed when tested on other Indian smart cities. Similarly, Tehran-based studies
  [5] emphasized the importance of meteorological drivers in AQI prediction, highlighting difficulties in transferring models across distinct environments.

Despite the difficulties we have mentioned, difficulties that can even be taken as chances for innovations in the domain. Current examples such as hybrid models (e.g., statistical baselines combined with deep neural networks as in [2]; [21] show how it is also achievable to be both ro- bust and interpretable alongside accuracy. While the best practices in applications and research keep growing, it is heartening to notice the fast adoption of explainable AI tools ([7]e.g. uncer- tainty quantification and attention-based tools) that may facilitate further progress in the process of moving from scientific projections to decision-making by policymakers. The connectivity provided by Edge-AI beachheads and IoT solutions ([24],[16]) is a truly thrilling aspect of how remotely situated sites can perform real-time forecasts. The research on uncertainty-aware models [18] and how probabilistic forecasts can be communicated together with numerical outputs to better convey confidence intervals to clients/users/decision-makers is something we must look into in the future.

In summary, the possibilities we examined explainable AI, hybrid architectures, scalable IoT solutions, and uncertainty-aware forecasting will also include solving the issues we listed: costs of computing, transferability if we are working to create the next-generation systems of air quality forecasting that are actionable and useful, data quality, and interpretability.

4 Policy Implications and Health Impact

Air quality forecasting involves more than just modelling science; it can be a key method for protecting public health and supporting evidence-based and impact-influenced policy making. Pollutants such as $PM_{2.5}$, PM_{10}, NO_2, SO_2, CO, and O_3 are linked to respiratory illnesses, cardiovascular disease, and premature mortality, and in 2013 the International Agency for Re- search on Cancer (IARC) (WHO) classified outdoor air pollution as a human carcinogen, sug- gesting a serious level of public health concern [22].

Many countries utilize the Air Quality Index (AQI) when attempting to convert complex model outputs into practical information. The AQI takes pollutant concentrations and places them into standard health-based ranges. The categories provide a communication pathway from a tech- nical forecast to another technical advisory and then to the public. The standard AQI ranges and categories, along with expected health effects, are summarized in table **??** for these vari- ous studies to reference, as most air quality communication frameworks globally are based on these. For example, studies completed in London [23] and Madrid [25] clearly indicated that the inclusion of AQI advisories into urban processes substantially assisted in regulating traffic and industrial emissions. Agencies in India, such as [8], highlighted that combining CPCB data in time with AQI categorization provided alerts that were timely indicators in Delhi, influ- encing advisories to the public and the readiness of hospitals. Several forecasting frameworks

Table 5: AQI Categories and Associated Health Implications

AQI Range	Category	Health Implications
0–50	Good	Air quality is satisfactory; air pollution poses little or no risk.
51–100	Moderate	Acceptable quality; some pollutants may pose minor risks for sensitive groups.
101–150	Unhealthy for Sensitive Groups	Sensitive populations may experience health effects; general public less affected.
151–200	Unhealthy	Everyone may begin to experience health effects; sensitive groups more seri- ously.
201–300	Very Unhealthy	Health alert: serious health effects likely across the population.
301–500	Hazardous	Emergency conditions: entire population likely to be affected.

explicitly link predictive modeling to policy applications. [17]used Jakarta's AQ network for $PM_{2.5}$ classification, which supported smart-city decision-making regarding air quality man- agement. Similarly, [24] emphasized how IoT-based real-time sensors could help build policy- ready frameworks for smart cities globally. Overall, these two case studies put into context how an accurate prediction model, coupled with an AQI communications framework, can en- able timely interventions like vehicle restrictions, closing schools, and shutting down industries during pollution events.

At the same time, health impact assessments play a key role in validating policy interven- tions. [5]demonstrated how long-term meteorological and AQ datasets from Tehran could guide strategic planning by linking exposure levels with population health trends. [20] fur- ther integrated Copernicus

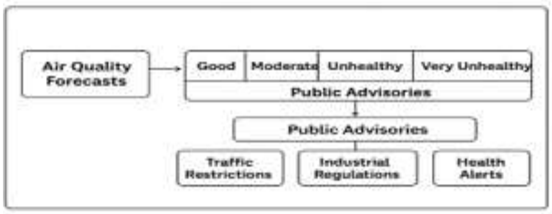

Figure 6: Conversion of forecasting outputs into AQI categories and their integration into policy and

public health advisories.
satellite and LiDAR data to capture regional transport of pollutants, enabling policymakers in Europe implement cross-border air quality agreements.
In summary, the integration of forecasting systems with AQI categories ensures that predictive outputs are interpretable, actionable, and relevant for public health. This alignment between science and policy allows for adaptive responses in urban planning, emission control, and healthcare preparedness. figure 6 illustrates how forecasting outputs are converted into AQI categories and subsequently used to inform policies and health advisories.

5 FUTURE RESEARCH DIRECTIONS

The future wave of air quality forecasting science will develop in various dimensions. In the domain of machine learning (ML), it is possible to create adaptive algorithms that can process heterogeneous data as well as utilize semi-supervised or active learning methods to minimize the reliance on large amounts of labelled data. For deep learning (DL), the direction for fu- ture research should be on developing explainable and lightweight architectures with poten- tial applications in real-time, incorporating physics-informed neural networks (PINNs), and merging state-of-the-art models like Transformers with Graph Neural Networks (GNNs) to ex- ploit temporal and spatial relationships. In terms of datasets and data fusion, advances will depend on building multi-source datasets covering ground monitoring, satellite data, meteo- rological parameters, traffic flows, and land use, and applying transfer learning to facilitate under-monitored areas and feed real-time IoT sensor streams. A second area of focus includes uncertainty-aware models, more specifically through Bayesian neural networks, ensemble tech- niques, and quantile regression methods, with hybrid frameworks being constructed to directly quantify uncertainty and increase operating reliability. Collectively these are taking steps to- ward establishing reliable, interpretable, and actionable forecasting systems that can inform real-time public health and policy decisions.

6 CONCLUSION

Air pollution remains a threat to public health, the environment, and sustainable development, necessitating the need for accurate forecasting. In this review, the development of forecasting techniques from classical statistical methods to machine learning and deep learning models was followed. While the ARIMA, MLR, and GAM models provided some interpretability, they were often inadequate for capturing nonlinear correlations or long-range relations. Machine and deep learning models such as Random Forests, SVM, LSTM, GRU, CNNs, and trans- formers improved spatiotemporal prediction, but they are invariably still constrained by high computational feasibility, low interpretability and challenge for real-time applications. Among reviewed approaches, hybrid frameworks are the most promising answer. The combination of statistical, machine learning, and deep learning models combined with a varied, heterogeneous mix of data such as weather data, satellite imagery, and IoT sensor networks, will provide robustness, flexibility, and scalability needed to utilize hybrid techniques. Uncertainty-aware hybrids offer a trade-off between predictive accuracy and comprehensibility, who are some of these key things are in place to trust and use in policy evidence-based. In the future of air qual- ity forecasting, the goal is to have new hybrid systems that are interpretable, computationally lightweight, and can be used for real-time applications. Bridging methodology progress and social needs, such hybrid models can provide early warnings, support effective regulations, and play a central role in sustainable air quality management.

REFERENCES

[1] Y. Liang, P. Deng, and X. Liu. A hybrid air quality prediction model based on empirical mode decomposition. *Tsinghua Science and Technology*, 29(1):118–130, 2024.

[2] P. K. Sharma, L. B. Hassan, and N. R. Mehta. Air pollution prediction with machine learning: A case study of indian cities. *International Journal of Environmental Science and Technology*, 20:5333–5348, 2023.

[3] P. A. Gupta. Future air quality prediction using long short-term memory based on hyper heuristic multi-chain model. *IEEE Access*, 12, 2024.

[4] M. Xu and J. Yang. A hybrid air quality index prediction model based on cnn and attention gate unit. *IEEE Access*, 10:110921–110937, 2022.

[5] H. Lin, B. Wang, and L. Yu. Semi-supervised air quality forecasting via self-supervised hierarchical graph neural network. *IEEE Transactions on Knowledge and Data Engineer- ing*, 35(5):5230–5243, 2023.

[6] C. Zhang and B. Han. A novel coupled optimization prediction model for air quality. *IEEE Access*, 11, 2023.

[7] M. Lee and K. Chen. Uncertainty-aware deep learning architectures for highly dynamic air quality prediction. *IEEE Access*, 9:39142–39155, 2021.

[8] Y. Fang and P. Sen. Integrating machine learning techniques for air quality index fore- casting and insights from pollutant-meteorological dynamics in sustainable urban envi- ronments. *Earth Science Informatics*, 17:3733–3748, 2024.

[9] S. Gupta and R. Patel. Enhancing air quality forecasting using machine learning tech- niques. *IEEE Access*, 12:197290–197299, 2024.

[10] S. Kumar and P. Das. Comparative analysis study for air quality prediction in smart cities using regression techniques. *IEEE Access*, 11:128325–128338, 2023.

[11] J. Wang and Y. Li. Predicting multivariate air pollution: A gaussian-mixture nested facto- rial variational autoencoder approach. *IEEE Geoscience and Remote Sensing Letters*, 21, 2024.

[12] Y. Ö zü pak, F. Alpsalaz, and E. Aslan. Air quality forecasting using machine learning: comparative analysis and ensemble strategies for enhanced prediction. *Water, Air, & Soil Pollution*, 236:464, 2025.

[13] H. Wu, J. Sun, and T. Zhao. Two-stage feature engineering to predict air pollutants in urban areas. *IEEE Access*, 12, 2024.

[14] A. Utku, U. Can, M. Alpsü lü n, H. C. Balıkç ı, A. Amoozegar, A. Pilatin, and A. Barut. Advancing air quality monitoring: Deep learning-based cnn-rnn hybrid model for pm2.5 forecasting. *Atmosphere*, 16(9):1003, August 2025.

[15] L. Fang, Z. Wei, and R. Zhang. Forecasting air quality in kiev during 2022 military con- flict using sentinel 5p and optimized machine learning. *IEEE Transactions on Geoscience and Remote Sensing*, 61, 2023.

[16] K. Y. Chen. Air quality index forecasting via genetic algorithm-based improved extreme learning machine (ga-kelm). *IEEE Access*, 11:67086–67097, 2023.

[17] K. Fernandez, A. T. Lim, and R. Gupta. Boosting algorithm to handle unbalanced classifi- cation of pm2.5 concentration levels in jakarta-indonesia. *IEEE Access*, 11:35680–35696, Apr 2023.

[18] S. Hu and R. Nelson. A predictive data feature exploration-based air quality prediction approach. *IEEE Access*, 7:224331–224346, 2019.

[19] X. Zhao, Y. Sun, and K. Chen. Machine learning and deep learning approaches for pm2.5 prediction: A study on urban air quality in jaipur, india. *Earth Science Informatics*, 18(97), 2025.

[20] M. R. Hassan, T. S. Kumar, and W. Li. Machine learning framework for high-resolution air temperature downscaling using lidar-derived urban morphological features. *Urban Climate*, 57, 2024.

[21] J. Liu, M. Wong, and T. Zhou. Optimized machine learning model for air quality index prediction in major cities in india. *Scientific Reports*, 14(6795), 2024.

[22] A. Smith, B. Johnson, and C. Lee. Urban air pollution monitoring system with forecasting models. *IEEE Sensors Journal*, 16(8):2598–2606, 2016.

[23] D. Williams, E. Martinez, and F. Kim. Data-driven air quality characterization for urban environments: A case study. *IEEE Access*, 6:77996–78006, 2018.

[24] L. Jiang and T. Brown. Comparative analysis of machine learning techniques for predict- ing air quality in smart cities. *IEEE Access*, 7:128325–128338, 2019.

[25] R. Mohan. Graph neural network for air quality prediction: A case study in madrid. *IEEE Access*, 11:2729–2742, 2023.

[26] Z. Lin, Y. Gao, and X. Ren. Application of machine learning and deep learning techniques in modeling the associations between air pollution and meteorological parameters in urban areas of tehran metropolis. *Environmental Monitoring and Assessment*, 196(994), 2024.

[27] T. Suzuki and M. Inoue. Recurrent air quality predictor based on meteorology- and pollution-related factors. *IEEE Transactions on Industrial Informatics*, 14(9):4211–4223, Sept 2018.