

Air Quality Assessment and AQI Prediction Using Machine Learning: A Comparative Study of Major Indian Cities

Ketan Sapkal¹, Yogendra Wadkar², Rupesh Gosavi³, Dr. SWATI JOSHI⁴.

1,2,3 Department Of Computer Application, PVG's College of Science and Commerce

4 Research Guide, Department of Computer Application, PVG's College of Science and Commerce.

Abstract

Air pollution remains one of the most critical environmental and public health challenges in India, particularly in rapidly urbanizing regions. This study presents a comprehensive assessment of air quality by integrating a machine learning-based prediction framework with a detailed one-year comparative analysis of four major metropolitan cities—Pune, Mumbai, Bengaluru, and Delhi—for the year 2024.

Publicly available air quality data from the Government Open Data (OGD) platform were processed and transformed into a structured dataset. The Air Quality Index (AQI) was computed using the official Central Pollution Control Board (CPCB) National Air Quality Index (NAQI) methodology. Two supervised machine learning models, Linear Regression and Random Forest, were applied to predict AQI values and evaluate model performance.

The results indicate that Random Forest significantly outperforms Linear Regression, demonstrating a stronger ability to capture nonlinear relationships among pollutants. Feature importance analysis highlights PM_{2.5}, CO, and PM₁₀ as the most influential contributors to AQI. In addition, city-level statistical analysis reveals substantial regional disparities, with Delhi consistently experiencing higher pollution levels, while Bengaluru maintains relatively stable air quality.

Seasonal trend analysis further shows that AQI peaks during winter due to unfavourable meteorological conditions and decreases during the monsoon season due to natural atmospheric cleansing. Comparative analysis across cities emphasizes the role of geography, emission sources, and climate in shaping air quality patterns.

Overall, this study demonstrates the effectiveness of combining open environmental data, standardized AQI computation, and machine learning techniques to support transparent air quality assessment and data-driven environmental decision-making in India.

Keywords: Air Quality Index, Air Pollution Forecasting, Machine Learning, Predictive Modeling, Random Forest, Linear Regression.

1.Introduction

Air pollution has become a serious environmental and public health concern in India, with many cities recording some of the highest pollution levels in the world. As urbanization and industrial activities continue to grow, monitoring air quality regularly and assessing it accurately has become essential for protecting public health, guiding policy decisions, and managing the environment effectively.

The Central Pollution Control Board (CPCB) defines the Air Quality Index (AQI) as a standardized measure used to represent overall air quality based on key pollutants such as PM_{2.5}, PM₁₀, nitrogen dioxide (NO₂), sulfur dioxide (SO₂), carbon monoxide (CO), and ozone (O₃). High concentrations of

these pollutants are known to cause serious health issues, including respiratory diseases, heart problems, reduced lung function, and even premature death. Apart from health risks, poor air quality also affects the environment, reduces quality of life in urban areas, and impacts economic productivity.

With the rise of digital platforms like the Government Open Data Platform India (OGD), access to real-world air quality data has become much easier. Data from multiple monitoring stations across India is now publicly available, making it possible to perform transparent and data-driven analysis. At the same time, machine learning techniques have gained attention for their ability to capture complex, non-linear relationships between pollutants, making them useful for predicting and forecasting AQI more accurately.

In this study, publicly available pollutant data is collected and organized into a structured, station-level dataset. AQI values are then calculated using CPCB-defined breakpoints. To evaluate predictive performance, two supervised machine learning models—Linear Regression and Random Forest—are applied and compared. By combining open government data with machine learning approaches, this study aims to provide a more reliable framework for air quality assessment and support better environmental decision-making in India.

2. Motivation Of Research

Air pollution is one of the most critical environmental challenges today, particularly in a rapidly developing and urbanizing country like India. The Air Quality Index (AQI) serves as a standardized and easy-to-understand indicator that helps people assess how air quality can impact their health. Factors such as rapid industrialization, increasing vehicular emissions, construction activities, and population growth have significantly contributed to the deterioration of air quality in major cities. As a result, concerns related to respiratory diseases, reduced life expectancy, and overall public health have increased considerably.

The growing severity of air pollution highlights the need for systematic and data-driven evaluation of air quality trends. Poor air quality not only leads to health issues such as asthma, lung infections, and cardiovascular diseases, but also affects productivity, economic stability, and overall quality of life. These widespread impacts emphasize the importance of studying AQI patterns in a structured and analytical manner.

Initially, this research focused on predicting AQI using machine learning techniques based on pollutant data obtained from the Government Open Data Platform India (OGD) and AQI computation methods defined by the Central Pollution Control Board (CPCB). However, many existing studies primarily focus on short-term or single-day AQI predictions. To provide a more comprehensive understanding, this study extends the analysis to a full-year evaluation of AQI levels across four major Indian cities—Pune, Mumbai, Bengaluru, and Delhi—covering all 365 days of the year 2024.

Using this dataset, several analytical evaluations were performed, including:

- Identification of the most and least polluted cities based on AQI trends
- Comparative analysis of AQI variations across the selected metropolitan cities
- Monthly distribution of AQI categories (Good, Satisfactory, Moderate, Poor, Very Poor, Severe)
- Analysis of monthly AQI trends to understand overall air quality patterns

These analyses provide a broader perspective beyond short-term forecasting by highlighting long-term pollution behaviour and inter-city variations. The findings contribute to a deeper understanding of air

pollution patterns and reinforce the need for continuous monitoring, preventive policies, and intelligent forecasting systems.

Therefore, this research is motivated not only by the technical objective of predicting AQI using machine learning models, but also by the broader goal of promoting data-driven environmental awareness and supporting informed decision-making. The integration of open government data, CPCB-based AQI computation, long-term city-level analysis, and predictive modeling forms the core foundation of this study.

3.Scope of the Research

This study focuses on analyzing and predicting air quality using publicly available environmental data and machine learning techniques within the context of India. The research primarily utilizes pollutant data obtained from the Government Open Data Platform India (OGD), which provides real-time and historical air quality measurements from multiple monitoring stations.

The scope of the study includes the calculation of the Air Quality Index (AQI) using standard methods defined by the Central Pollution Control Board (CPCB), ensuring that the analysis aligns with nationally accepted environmental guidelines. The study considers major pollutants such as PM_{2.5}, PM₁₀, NO₂, SO₂, CO, and O₃, which are key contributors to air quality assessment.

From a geographical perspective, the research is limited to four major Indian cities—Pune, Mumbai, Bengaluru, and Delhi. A full-year dataset covering all 365 days of 2024 is used to perform detailed temporal and comparative analysis. This includes evaluating monthly trends, seasonal variations, and differences in pollution levels across cities.

In terms of methodology, the study applies supervised machine learning models, specifically Linear Regression and Random Forest, to predict AQI values based on pollutant concentrations. The scope includes data preprocessing, AQI computation, exploratory data analysis, and model performance evaluation using standard metrics.

However, the study does not incorporate real-time forecasting systems, advanced deep learning models, or detailed meteorological parameters such as wind speed, humidity, and temperature, which may influence air quality. Additionally, the analysis is limited to selected cities and may not fully represent air quality patterns across all regions of India.

Overall, the research provides a structured and data-driven framework for understanding and predicting air quality, while highlighting the potential of combining open government data with machine learning techniques for environmental analysis.

4. Literature Review

4.1 Research on AQI and Evaluation of Air Pollution

Air Quality Index (AQI) evaluation has been widely studied to understand pollutant behaviour, spatial distribution, and its impact on public health, particularly in urban and industrial regions. Data-driven analyses of AQI and major pollutants indicate significant variation between urban, peri-urban, and industrial zones, especially in areas influenced by coal mines and thermal power plants. These variations highlight the complex nature of pollution flow and exposure patterns [7].

Recent studies also show that modern AQI monitoring systems are increasingly integrated with advanced technologies such as Internet of Things (IoT) sensors and large-scale environmental datasets. These developments have improved real-time monitoring and prediction capabilities [8]. Furthermore, comparative studies across Indian cities suggest that AQI trends vary significantly depending on geographical location, emission sources, and meteorological conditions. This

emphasizes the importance of localized air quality assessment rather than relying solely on generalized national-level conclusions [9].

4.2 Studies Using Government and CPCB Air Quality Data

Public environmental datasets play a crucial role in AQI research by providing consistent and long-term pollutant measurements required for analysis and model development. In India, the Government Open Data Platform India offers accessible air quality datasets that support time-based analysis, inter-city comparisons, and machine learning applications [1].

Additionally, the Central Pollution Control Board (CPCB) defines the National Air Quality Index (NAQI) framework, which includes pollutant sub-index calculations, category thresholds, and health-based guidelines used in AQI evaluation [2]. The use of such standardized and publicly available data sources ensures consistency, reproducibility, and scientific reliability in air quality research.

4.3 Machine Learning Methods for AQI Forecasting

In recent years, machine learning and deep learning techniques have gained significant attention for AQI prediction across different regions in India. Hybrid models that combine traditional machine learning approaches with deep neural networks have demonstrated improved prediction accuracy and better temporal generalization [3].

Studies conducted in regions such as the Taj Trapezium Zone and Chennai show that model performance varies depending on pollutant composition, seasonal variations, and urban emission characteristics. This highlights the importance of localized feature engineering and model tuning for better predictive performance [4], [5].

Moreover, several studies indicate that AQI levels are influenced by large-scale environmental events, such as agricultural stubble burning in northern India. These findings suggest that machine learning models can capture complex pollution dynamics, including cross-regional pollutant transport [6].

City-specific studies in Visakhapatnam and Ahmedabad further validate the effectiveness of supervised learning models in short-term AQI prediction using historical pollutant and meteorological data [10], [11]. Overall, review-based studies confirm that machine learning has become one of the most effective approaches for air quality forecasting due to its ability to model non-linear relationships and handle diverse environmental datasets [12].

4.4 Research Gaps Found in Current Literature

Despite significant advancements in AQI monitoring and prediction, several important research gaps remain. Many existing studies in India are limited to single-city or region-specific datasets, which restricts the generalizability of predictive models across different climatic and emission conditions [4], [10], [11].

Additionally, while machine learning models often achieve high predictive accuracy, relatively few studies focus on model interpretability, cross-city applicability, and standardized open-data workflows suitable for policy-level implementation [3], [9], [12]. This creates challenges in translating model outputs into actionable environmental policies.

Another limitation is the dependence on localized training data and the uneven distribution of monitoring stations, which affects the scalability of AQI prediction models at a national level. Addressing these challenges requires the development of integrated multi-city datasets, standardized modelling frameworks, and hybrid approaches that combine data-driven methods with an understanding of physical pollution processes.

5. Methodology

In the first phase of this study, a pollution-specific air quality dataset was obtained from the Government Open Data Platform India (OGD). The dataset underwent preprocessing, which included data cleaning, filtering of relevant pollutants, and pivot transformation to convert it into a structured station-level format. Following this, the Air Quality Index (AQI) was calculated using the standard methodology defined by the Central Pollution Control Board (CPCB) under the National Air Quality Index (NAQI) framework.

To analyse and predict AQI values, supervised machine learning models—Linear Regression and Random Forest—were applied to the processed dataset. These models were used to evaluate predictive performance and understand the contribution of different pollutants to overall air quality.

In the second phase, the study was extended beyond short-term prediction to a comprehensive year-long comparative analysis. AQI data for all 365 days of 2024 was analysed across four major Indian cities: Pune, Mumbai, Bengaluru, and Delhi. This phase focused on identifying dominant pollution patterns, analysing the distribution of AQI categories, and comparing air quality trends across cities over time.

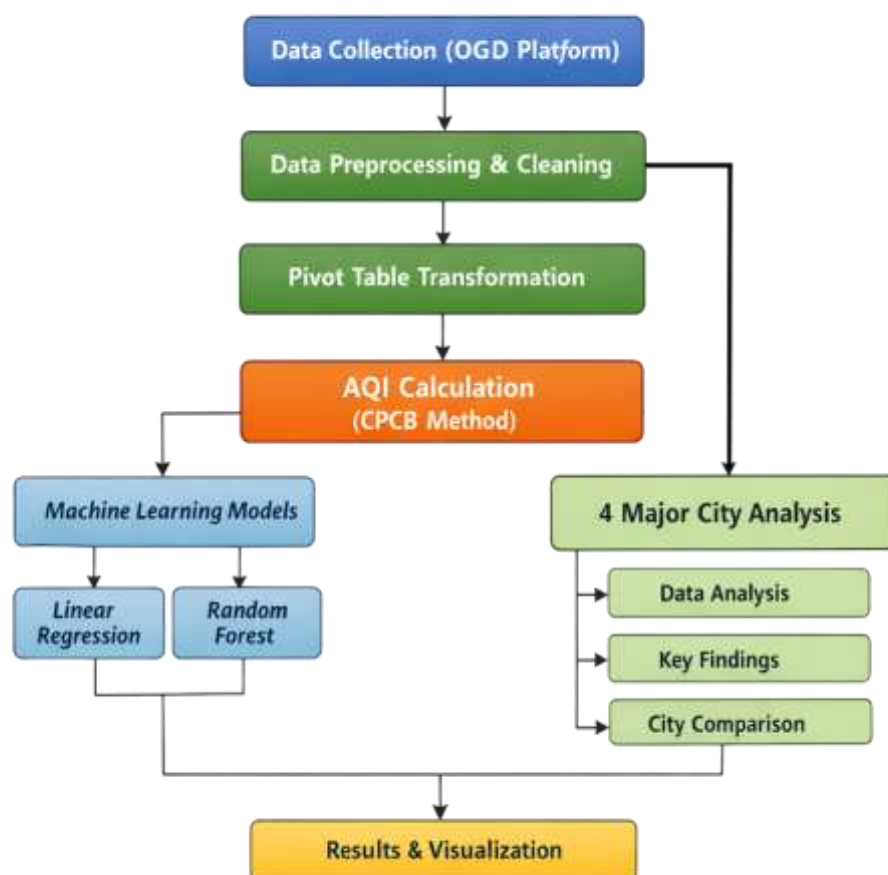


Fig. 1. Proposed methodology workflow for AQI prediction and city-level analysis.

Methodology Overview

The approach adopted in this study is structured into a five-step framework, consisting of data acquisition, data cleaning and filtering, data transformation, AQI computation, and machine learning modeling. Each stage plays a crucial role in converting raw open-source data into a structured format suitable for accurate AQI analysis and prediction. The overall methodology followed in this study is illustrated in [Fig. 1](#).

5.1 Data Acquisition

The dataset used in this research was obtained from the Government Open Data Platform India (OGD), which provides publicly accessible air quality data collected from multiple monitoring stations across India. The data is available in CSV format and includes time-stamped records of various pollutants measured at different locations.

Each record in the raw dataset contains the following attributes:

- City name
- Monitoring station name
- Timestamp (latest recorded update)
- Geographic coordinates (latitude and longitude)
- Pollutant type (PM2.5, PM10, NO₂, SO₂, CO, O₃, etc.)
- Minimum, maximum, and average pollutant concentration values

The dataset is structured such that multiple rows correspond to the same station and timestamp, with each row representing a different pollutant. As a result, preprocessing is required to reorganize the data into a unified, station-level format suitable for AQI computation and subsequent machine learning analysis.

5.2 Data Cleaning and Filtering

Data cleaning was performed to ensure the accuracy, consistency, and reliability of the dataset prior to further analysis. Since the raw data contained missing values and inconsistencies, several preprocessing steps were applied.

The primary actions included:

- Discarding rows with missing (NA) pollutant values, as these records could not be used for AQI computation
- Converting timestamp values into a standardized datetime format for consistency
- Removing redundant, duplicate, or corrupted entries to maintain data integrity
- Filtering pollutant readings to retain only the standard pollutants required for AQI calculation, including PM2.5, PM10, NO₂, SO₂, CO, O₃, and NH₃

After applying these steps, the dataset was cleaned and transformed into a consistent format containing only valid pollutant measurements. This ensured that the data could be reliably used in subsequent stages such as transformation, AQI computation, and machine learning modelling.

5.3 Pivot Transformation

The raw dataset contained multiple rows corresponding to the same monitoring station and timestamp, with each row representing a different pollutant. To prepare the data for AQI computation and machine learning analysis, a pivot transformation was applied.

This transformation converted pollutant-wise rows into a single consolidated row per station per timestamp, where each pollutant was represented as a separate column. For example:

| City | Station | Timestamp | PM2.5 | PM10 | NO₂ | SO₂ | CO | O₃ | ... |

As a result, a structured station-level dataset was obtained, in which all pollutant measurements for a given time instance were aligned horizontally. This format ensured accurate AQI computation and provided a suitable input structure for training machine learning models.

5.4 AQI Computation Using CPCB Method

In this section, the Air Quality Index (AQI) was calculated following the official procedure defined by the Central Pollution Control Board (CPCB) under the National Air Quality Index (NAQI) framework.

The AQI is categorized into six levels: Good, Satisfactory, Moderately Polluted, Poor, Very Poor, and Severe. These categories are determined based on the ambient concentration levels of various air pollutants and their associated health impacts, commonly referred to as health breakpoints. AQI sub-indices have been defined for eight major pollutants, namely PM₁₀, PM_{2.5}, NO₂, SO₂, CO, O₃, NH₃, and Pb, based on short-term exposure standards (up to 24 hours).

For each pollutant, a sub-index value is calculated using a linear relationship between pollutant concentration and predefined breakpoint ranges. For example, in the case of PM_{2.5}, a concentration of 31 µg/m³ corresponds to a sub-index of 51, while 60 µg/m³ corresponds to a sub-index of 100.

The overall AQI is determined by selecting the maximum sub-index value among all pollutants, representing the worst air quality condition at a given time.

The AQI categories along with their corresponding pollutant concentration ranges, as defined by CPCB, are presented in [Fig. 2](#).

Fig. 2. AQI categories and pollutant concentration breakpoints as defined by CPCB.

AQI Category	AQI	Concentration Range*							
		PM ₁₀	PM _{2.5}	NO ₂	O ₃	CO	SO ₂	NH ₃	Pb
Good	0-50	0-50	0-30	0-40	0-50	0-1.0	0-40	0-200	0-0.5
Satisfactory	51 - 100	51-100	31-60	41-80	51-100	1.1-2.0	41-80	201-400	0.5-1.0
Moderately Polluted	101-200	101-250	61-90	81-180	101-168	2.1-10	81-380	401-800	1.1-2.0
Poor	201-300	251-350	91-120	181-280	169-208	10-17	381-800	801-1200	2.1-3.0
Very Poor	301-400	351-430	121-250	281-400	209-748*	17-34	801-1600	1200-1800	3.1-3.5
Severe	401-500	430+	250+	400+	748+*	34+	1600+	1800+	3.5+
* CO in mg/m ³ and other pollutants in µg/m ³ ; 24-hourly average values for PM ₁₀ , PM _{2.5} , NO ₂ , SO ₂ , NH ₃ , and Pb, and 8-hourly values for CO and O ₃ .									

5.5 City-Level AQI Dataset Collection

To extend the scope of the study beyond station-level AQI computation, an additional city-level temporal dataset was incorporated to enable a more comprehensive environmental analysis. This phase focused on four major metropolitan cities in India—Pune, Mumbai, Delhi, and Bengaluru—selected due to their high population density, urban activity, and well-documented air quality challenges.

For each city, a complete dataset of daily AQI values for the year 2024 was collected. Unlike the earlier dataset, which consisted of raw pollutant concentrations obtained from the Government Open Data Platform India (OGD), the AQI values used in this phase were already computed based on the National Air Quality Index (NAQI) framework defined by the Central Pollution Control Board (CPCB). Therefore, this stage focused on statistical analysis and comparison rather than AQI calculation.

Each city-level dataset had the following characteristics:

- Daily AQI observations for all 365 days of 2024, ensuring complete temporal coverage
- Month-wise organization of data, enabling consistent analysis of seasonal and monthly trends
- Classification of AQI values into standard CPCB categories, including Good, Satisfactory, Moderate, Poor, Very Poor, and Severe
- Distribution of AQI values across these categories, allowing quantitative assessment of pollution severity within each city

5.6 Statistical Analysis Approach for AQI Trends

After obtaining the dataset, a detailed statistical analysis was performed to examine the variation in Air Quality Index (AQI) over time for each city. This analysis aimed to identify trends, patterns, and seasonal variations in air quality across the selected regions.

For each monthly AQI dataset, the following descriptive and analytical metrics were computed:

- Monthly average AQI, representing the overall pollution level for each month
- Monthly minimum and maximum AQI values, indicating variability within the month
- Yearly average AQI, providing an overall summary of pollution levels for the year 2024
- Identification of the months with the highest and lowest pollution levels for each city
- Analysis of AQI category distribution based on CPCB classifications (Good, Satisfactory, Moderate, Poor, Very Poor, and Severe)
- Seasonal aggregation of AQI by grouping months into winter, summer, and monsoon periods to assess the influence of climatic conditions on air quality

5.7 Comparative Analysis Framework

To gain broader insights into urban environmental conditions, a systematic comparative framework was developed across the four selected metropolitan cities. This phase builds upon the statistical analysis presented in Section 4.6 and focuses on identifying key differences and patterns in air quality across cities.

The analysis included:

- Comparison of yearly average pollution levels across the selected cities
- Identification of the most and least polluted cities in 2024
- Examination of seasonal pollution patterns, highlighting increased pollution during winter and improvements during the monsoon season
- Analysis of the distribution of AQI categories to assess the intensity of pollution and the effectiveness of mitigation efforts

This comparative approach enables the study to move beyond individual city-level observations and provides a broader understanding of urban air quality dynamics in India. The insights derived from this analysis are valuable for environmental planning, pollution control strategies, and policy-level decision-making.

6. RESULTS & ANALYSIS

6.1 Machine Learning Model Performance

To evaluate the predictive capability of the proposed approach, two supervised models—Linear Regression and Random Forest—were trained using pollutant concentrations as input features. Model performance was assessed using standard regression metrics, including the coefficient of determination (R^2), Mean Absolute Error (MAE), and Mean Squared Error (MSE).

The Linear Regression model exhibited limited predictive performance, with low R^2 values and relatively high error metrics. This indicates that AQI cannot be effectively modeled using simple linear relationships due to the complex and non-linear interactions among pollutants.

In contrast, the Random Forest model demonstrated significantly improved performance, achieving higher R^2 and lower MAE and MSE values. This suggests that ensemble-based models are better suited for capturing non-linear dependencies in environmental data. Overall, Random Forest was found to be more reliable for AQI prediction in this study.

6.2 City-Level AQI Statistical Results

6.2.1 Yearly Average AQI Comparison

The complete dataset for the year 2024 was used to compare the average AQI across four major Indian cities: Bengaluru, Mumbai, Pune, and Delhi.

Bengaluru recorded the lowest average AQI (70.88), followed by Mumbai (82.53) and Pune (91.72), indicating relatively moderate pollution levels in these cities. In contrast, Delhi exhibited a significantly higher average AQI (189.75), reflecting persistently poor air quality conditions.

Based on these values, the cities can be ranked from cleanest to most polluted as follows:

1. Bengaluru
2. Mumbai
3. Pune
4. Delhi

Delhi's average AQI is approximately 2.7 times higher than that of Bengaluru, highlighting a substantial disparity in urban air quality across regions.

6.2.2 Most Polluted and Cleanest City

From the comparative analysis:

- Most polluted city: Delhi
- Cleanest city: Bengaluru

This observation further reinforces the presence of a clear north–south pollution gradient in India, with northern regions experiencing significantly higher pollution levels compared to southern cities.

6.2.3 Monthly Pollution Extremes

The months with the highest and lowest AQI values for each city are summarized in Table 1.

City	Highest Month	Lowest Month
Pune	November	July
Mumbai	December	August
Delhi	November	August
Bengaluru	March	June

Table 1. Months with highest and lowest AQI levels for each city (2024)

6.2.4 AQI Category Distribution

The distribution of AQI days across different categories provides a clear indication of variations in air quality among the selected cities.

Delhi recorded only 33 days under the **Good** category, while a large number of days fell under **Moderate** (142), **Satisfactory** (77), **Poor** (73), **Very Poor** (69), and **Severe** (17) categories. This indicates frequent exposure to unhealthy air quality conditions in the city.

In contrast, Bengaluru showed significantly better air quality, with 91 days classified as **Good** and 265 days as **Satisfactory**, and no days recorded under the **Poor** or **Severe** categories. This reflects relatively stable and cleaner environmental conditions.

Pune and Mumbai exhibited intermediate patterns, with most days falling under **Good**, **Satisfactory**, and **Moderate** categories, along with a limited number of **Poor** days and no **Severe** conditions.

Overall, this distribution highlights a clear inter-city disparity in air quality, with Delhi emerging as the most environmentally vulnerable city among those considered.

6.2.5 Seasonal AQI Variation

Table 2. Seasonal average AQI across cities (2024)

Seasons	Pune	Mumbai	Delhi	Bengaluru
Winter	129	134	306	81
Summer	101	88	179	81
Monsoon	53	46	106	51

Interpretation

- Delhi experiences the highest pollution levels during winter, primarily due to temperature inversion, which restricts the dispersion of pollutants.
- A significant reduction in AQI is observed across all cities during the monsoon season, reflecting the cleansing effect of rainfall.
- Seasonal variation is most pronounced in Delhi, while Bengaluru shows the least variation, indicating more stable air quality conditions.

Analysis: Air Quality Trends in Pune, Mumbai, Delhi, and Bengaluru (2024)

Pune: November Peak and Geographical Influence

A significant increase in AQI was observed in Pune during November 2024. This spike can be attributed to a combination of seasonal and anthropogenic factors. The post-monsoon period is characterized by lower temperatures and reduced atmospheric mixing, which leads to the accumulation of pollutants near the surface. Additionally, increased emissions during festive periods further contribute to elevated AQI levels.

Geographical factors also play an important role in shaping Pune's air quality. Located on the Deccan Plateau at an elevation of approximately 559 m and near the Western Ghats, Pune generally benefits from better ventilation and dispersion of pollutants. Furthermore, heavy monsoon rainfall contributes to periodic cleansing of the atmosphere, resulting in comparatively lower annual pollution levels than more densely industrialized regions.

Mumbai: Seasonal Peaks and Coastal Influence

Mumbai experienced a noticeable rise in AQI during the late October to November period in 2024. This increase is primarily associated with seasonal transitions, during which temperature drops and changes in wind patterns reduce pollutant dispersion. Shifts in wind direction—from sea-based to land-based—can lead to the accumulation of pollutants within the city.

Despite being a coastal city, where sea breezes typically help in dispersing pollutants, this natural advantage can be diminished under certain atmospheric conditions. During late 2024, delayed monsoon withdrawal and the presence of stable atmospheric conditions reduced ventilation, allowing pollutants to persist for longer durations.

Additionally, regional wind patterns and local emissions contribute to variability in AQI levels. As a result, Mumbai's air quality is strongly influenced by a combination of coastal dynamics, seasonal meteorology, and emission sources.

Delhi: Winter Pollution and Seasonal Impact

Delhi experienced severe air quality deterioration during the winter months, particularly in October and November 2024. This increase in pollution can be attributed to a combination of regional and local factors. Seasonal agricultural residue burning in nearby states contributes significantly to particulate matter levels, while local sources such as vehicular emissions, industrial activity, and waste burning further intensify pollution.

Winter meteorological conditions play a critical role in worsening air quality. Lower temperatures, reduced wind speeds, and temperature inversion limit the dispersion of pollutants, causing them to accumulate near the surface. As a result, AQI levels frequently fall within the *Poor to Severe* categories during this period.

A similar trend is observed in January, where cold weather conditions continue to trap pollutants, leading to sustained high AQI levels. Overall, Delhi exhibits consistently high baseline pollution with pronounced seasonal peaks, making it the most affected city among those analysed.

Bengaluru: Seasonal Stability and Climatic Advantage

Bengaluru recorded its highest AQI levels in March 2024, primarily due to dry weather conditions. The absence of rainfall during this period reduced the natural removal of airborne pollutants, allowing particulate matter to accumulate. Additionally, local factors such as traffic emissions and construction activities contributed to short-term increases in pollution levels.

Despite these occasional peaks, Bengaluru generally maintains better air quality compared to other major cities. Its higher elevation and relatively stronger wind conditions support effective dispersion of pollutants throughout the year. Furthermore, the absence of major industrial clusters and regional pollution sources contributes to a lower overall pollution burden.

As a result, Bengaluru demonstrates relatively stable AQI levels with minimal seasonal variation, making it the least polluted city among the selected metropolitan areas.

Actions to Improve Air Quality

Delhi

To address the severe air pollution levels observed in Delhi, a combination of short-term and long-term measures is required. Reducing emissions from major sources such as vehicular traffic, industrial activity, and biomass burning should be prioritized [26]. Expanding clean public transportation systems, including metro networks and electric buses, can significantly reduce vehicular emissions [26].

Control of road dust is also critical, as it contributes substantially to particulate pollution. Measures such as mechanized street sweeping, water sprinkling, and stricter construction-site regulations can help mitigate this issue [26]. Additionally, improving waste management practices and preventing open burning are essential for long-term improvement. Increasing urban green cover and strengthening enforcement of environmental regulations can further support sustained air quality management.

Bengaluru

Although Bengaluru maintains relatively better air quality compared to other cities, preventive measures are necessary to sustain and improve this condition. The primary focus should be on controlling vehicular emissions and road dust, which are the major contributors to pollution [24].

Implementing continuous dust control measures, including road maintenance, regulated construction activities, and regular cleaning, is essential. Enhancing public transportation systems and promoting cleaner mobility options can further reduce emissions. Maintaining green spaces and ensuring strict enforcement of environmental policies will help preserve Bengaluru's relatively stable air quality [24] [25].

Seasonal Variation in Air Quality

Air quality across cities shows strong dependence on seasonal factors.

- Winter (November–January):
Air quality deteriorates significantly due to temperature inversion, reduced wind speeds, and increased emissions from various sources. These conditions trap pollutants near the surface, resulting in high AQI levels [21].
- Monsoon (June–September):
AQI levels improve due to rainfall, which removes particulate matter from the atmosphere. Increased wind activity also aids in pollutant dispersion.
- Summer (April–May):
Air quality shows moderate variation. Dry conditions increase dust levels, while higher temperatures promote photochemical reactions, leading to increased ozone formation.

International Perspective on AQI Reduction: A Case Study of Beijing and Delhi

Introduction

Beijing has historically experienced severe air pollution, particularly due to high levels of particulate matter (PM_{2.5}). Major contributing factors included industrial emissions, coal-based energy consumption, rapid growth in vehicular traffic, and construction-related dust. Rising public health concerns and increased global attention led to the implementation of strict environmental regulations.

Over time, these interventions resulted in significant improvements in air quality. The evolution of Beijing's air pollution control strategies provides valuable insights for other highly polluted metropolitan regions.

Industrial Emission Control

Industrial activity was one of the primary sources of air pollution in Beijing. To address this, the government introduced strict emission standards and relocated or shut down highly polluting industries, including steel plants and coal-based factories.

Industries were required to adopt cleaner technologies, such as advanced filtration and emission control systems. Non-compliance resulted in penalties and administrative action. These measures significantly reduced industrial particulate emissions.

Energy Transition and Transport Policies

Rapid urbanization led to increased vehicular emissions in Beijing. In response, authorities implemented stricter vehicle emission standards aligned with global benchmarks and phased out older, high-emission vehicles.

The city also invested heavily in public transportation infrastructure, including metro expansion and cleaner bus fleets. Incentives such as subsidies encouraged the adoption of electric vehicles. These combined efforts reduced dependence on fossil fuel-based transport and contributed to improved air quality.

Dust and Construction Management

Construction activities and road dust were major contributors to secondary particulate pollution. To mitigate this, authorities enforced strict dust control measures, including covering construction materials, regular water spraying, and increased monitoring of construction sites.

Improved urban sanitation and stricter enforcement further reduced non-industrial sources of particulate pollution.

Measurable Outcomes

These interventions led to a substantial reduction in PM_{2.5} levels, with studies reporting declines of nearly two-thirds compared to peak levels in the early 2010s. The number of severe pollution days decreased significantly, while moderate air quality days became more frequent.

These outcomes demonstrate that sustained and coordinated policy interventions can lead to measurable improvements in air quality.

Governance and Enforcement

A key factor in Beijing's success was strong governance and enforcement. Environmental targets were integrated into administrative performance evaluations, ensuring accountability at multiple levels.

Strict enforcement mechanisms ensured that policies were effectively implemented, rather than remaining at a planning stage. This integration of policy design and execution played a crucial role in achieving long-term improvements.

Comparative Analysis: Implications for Delhi

Similar Environmental Conditions

Delhi experiences similar air pollution challenges, particularly during winter. Temperature inversion, low wind speeds, and high emissions from vehicles, industries, construction activities, and agricultural residue burning contribute to elevated AQI levels.

Current Measures in Delhi

Delhi has implemented several measures, including the Graded Response Action Plan (GRAP), restrictions on construction and industrial activities during high pollution periods, and expansion of air quality monitoring systems. Efforts are also being made to promote electric mobility.

However, many of these measures are reactive and focused on short-term mitigation rather than long-term structural changes.

Policy Lessons for Delhi

Beijing's experience highlights the importance of sustained and structural interventions for effective air quality management. Key strategies that can be adapted for Delhi include:

- Relocation or technological upgrade of high-emission industries
- Transition to cleaner household and industrial energy sources
- Stricter vehicle emission regulations and faster phase-out of older vehicles
- Expansion and strengthening of public transportation systems
- Stronger enforcement mechanisms with clear accountability

While differences in governance, population density, and regional conditions exist, the underlying policy principles remain relevant and adaptable.

Graph Analysis

1. AQI Distribution

The histogram shows that most AQI values fall within the *Moderate*, *Poor*, and *Very Poor* categories. This indicates that air pollution levels remain consistently high across multiple observations, with fewer instances of clean air conditions.

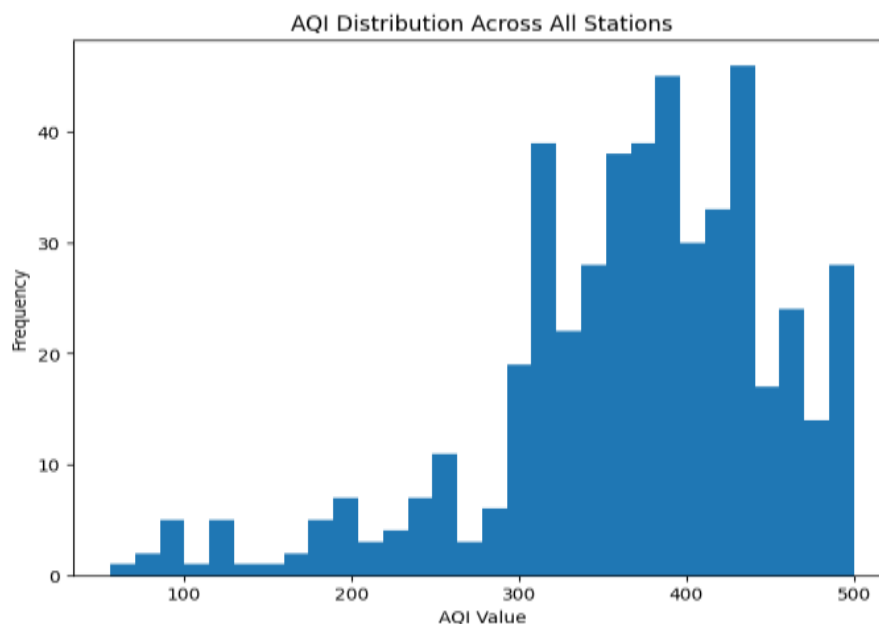


Figure 1: AQI Distribution Across All Stations

2. Actual vs Predicted AQI

The comparison between actual and predicted AQI values demonstrates that the Random Forest model closely follows the observed trends. The strong visual alignment between the two curves suggests a high level of prediction accuracy and the model's ability to capture non-linear relationships in the data.

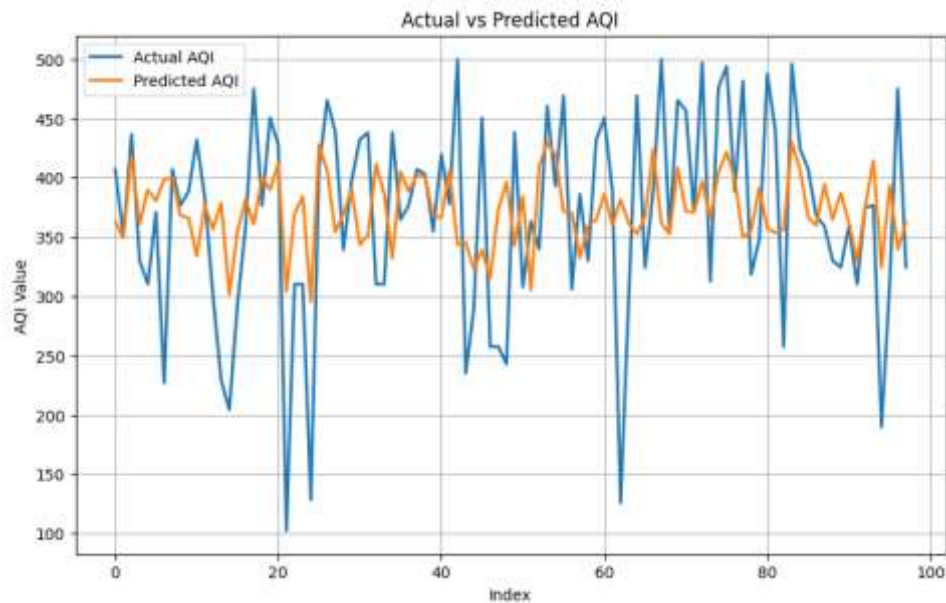
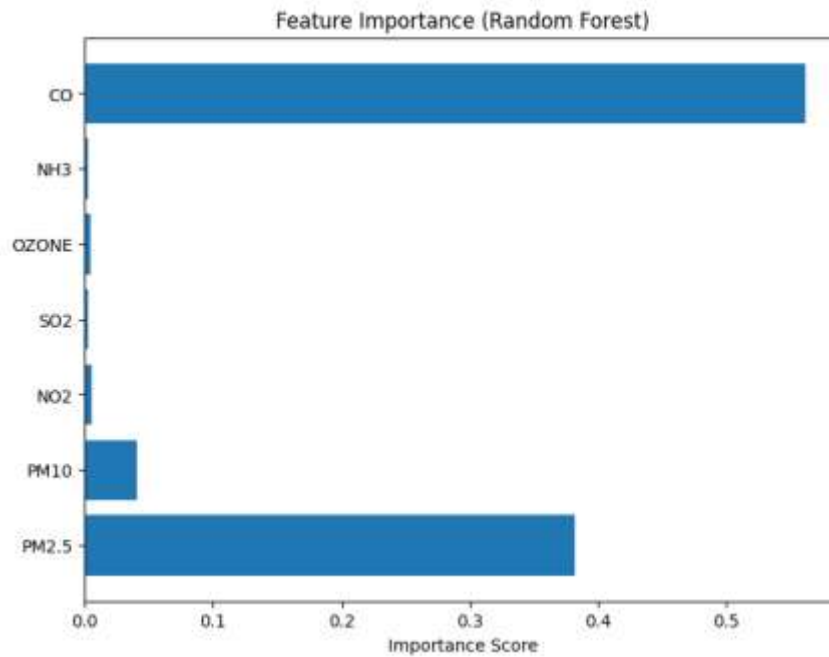


Figure 2: Actual vs Predicted AQI

3. Feature Importance

The feature importance analysis reveals that PM_{2.5}, CO, and PM₁₀ are the most influential factors affecting AQI. Among these, PM_{2.5} contributes the most, highlighting its critical role in determining overall air quality. This confirms that fine particulate matter and combustion-related pollutants are the dominant contributors to pollution levels.



4. Pune Monthly Average AQI (2024)

Pune exhibits moderate AQI levels throughout the year with relatively stable variation. Although seasonal changes are present, the fluctuations are not extreme, indicating comparatively consistent air quality conditions compared to other metropolitan cities.

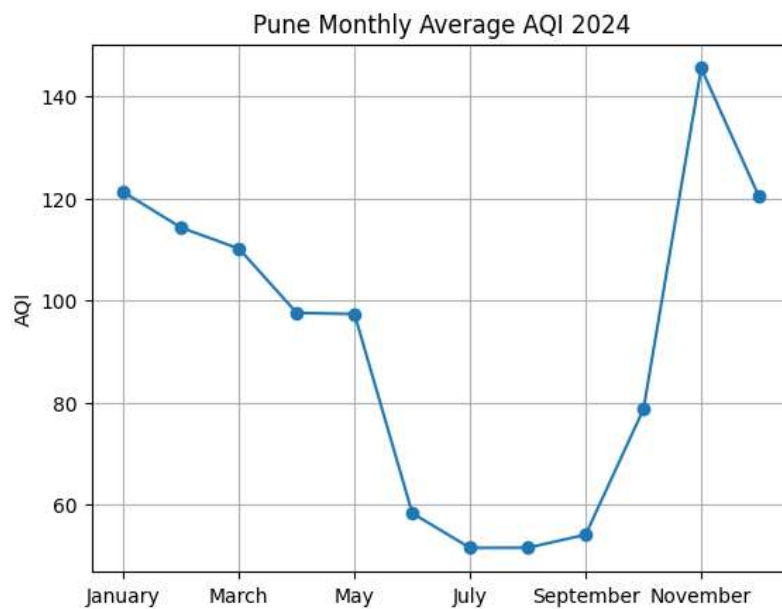
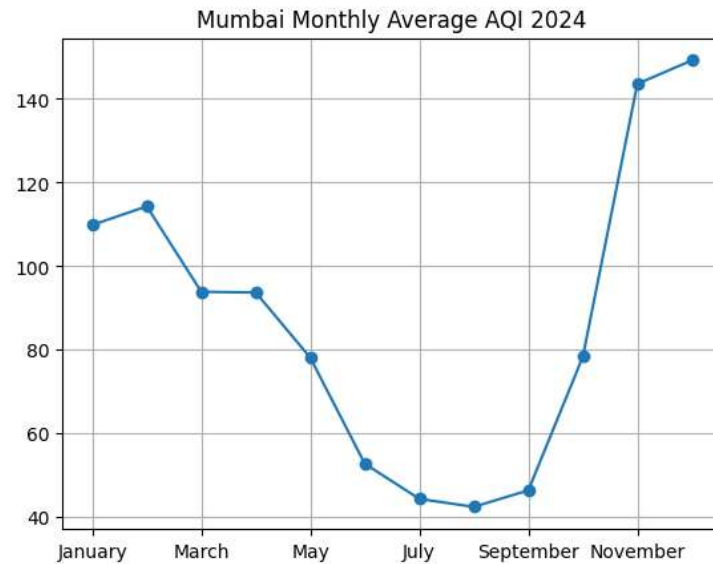


Figure 4: Pune Monthly AQI

5. Mumbai Monthly Average AQI (2024)

Mumbai shows moderate AQI levels with noticeable seasonal variation. Air quality tends to worsen during the winter months, while remaining relatively better during the monsoon period. However, extreme pollution spikes are limited compared to more heavily polluted cities.



6. Delhi Monthly Average AQI (2024)

Delhi records significantly higher AQI values across multiple months, especially during winter. This reflects severe pollution conditions and strong seasonal influence, making it the most affected city among those analyzed.

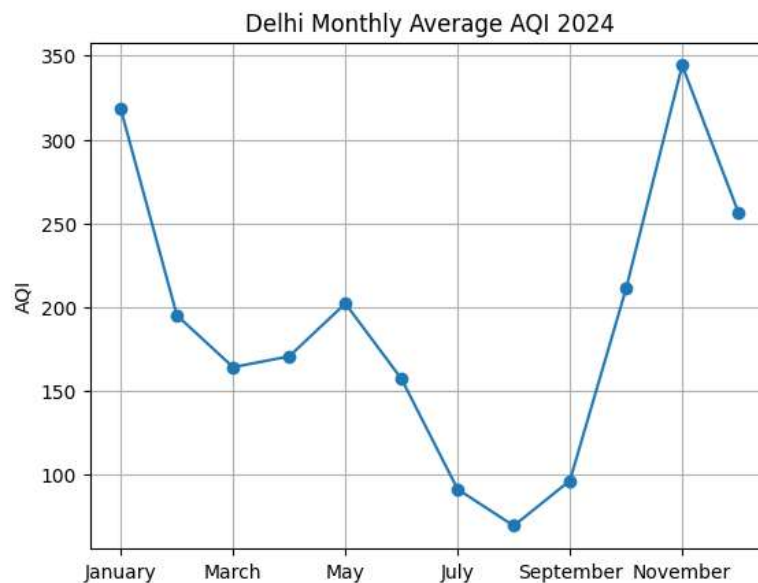
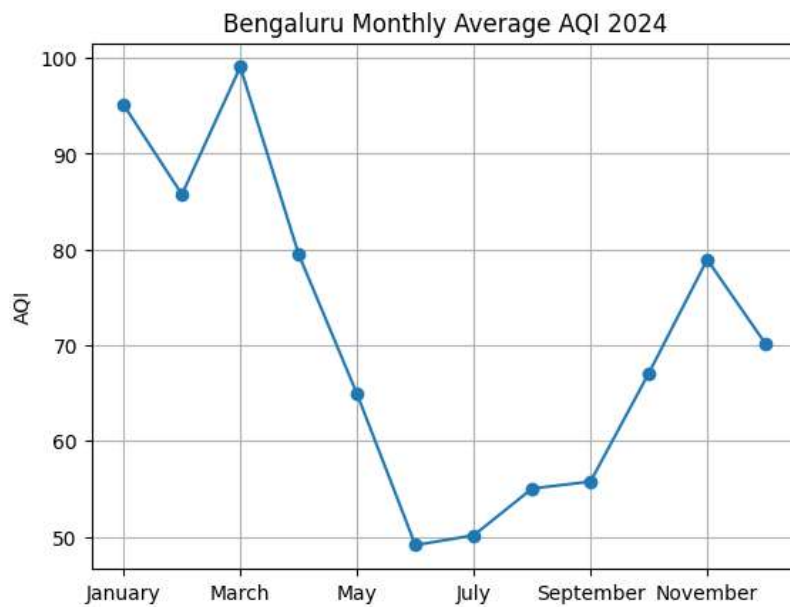


Figure 6: Delhi Monthly AQI

7. Bengaluru Monthly Average AQI (2024)

Bengaluru maintains relatively low AQI levels throughout the year. The variation is minimal compared to other cities, indicating more stable and cleaner air quality conditions.



8. Comparison of Monthly AQI Trends Across Cities

The comparative trend clearly shows that Delhi consistently records the highest AQI values, while Bengaluru maintains the lowest. Pune and Mumbai fall in the intermediate range, exhibiting moderate variations throughout the year.

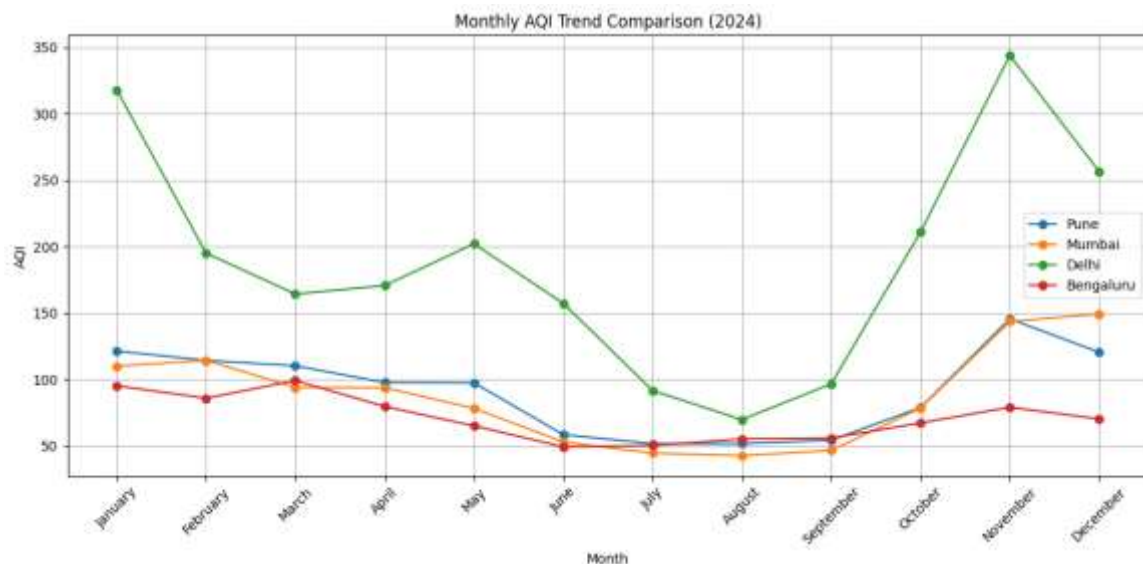
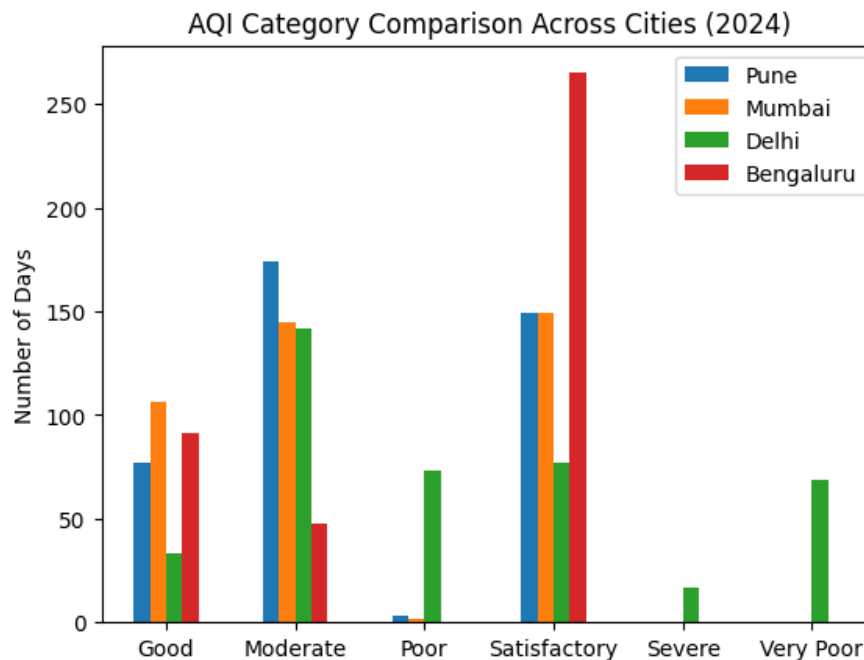


Figure 8: Monthly AQI Trend Comparison Across Cities

9. AQI Category Comparison Across Cities (2024)

The category-wise comparison highlights significant inter-city differences. Delhi records a higher number of days in *Poor*, *Very Poor*, and *Severe* categories, indicating frequent exposure to unhealthy air conditions. In contrast, Bengaluru, Mumbai, and Pune show a greater proportion of days in *Good* to *Moderate* categories, reflecting comparatively better air quality.



Conclusion

In this study, a comprehensive assessment of air quality in India was conducted by integrating a machine learning-based AQI prediction framework with a detailed one-year comparative analysis of four major metropolitan cities—Pune, Mumbai, Bengaluru, and Delhi—for the year 2024. Publicly available government datasets and the CPCB National Air Quality Index (NAQI) methodology were utilized to develop a transparent and reproducible analytical pipeline.

The predictive analysis demonstrated that machine learning models, particularly Random Forest, significantly outperform linear approaches in capturing the nonlinear interactions among pollutants. In addition, the city-level statistical analysis revealed clear regional disparities in air quality, with Delhi consistently exhibiting higher pollution levels, while Bengaluru maintained relatively stable and lower AQI values.

Beyond prediction, the study incorporated seasonal trend analysis and real-world environmental observations, supported by existing literature and reported evidence, to better understand variations in AQI across cities. The findings highlight that urban air pollution is not merely a modeling challenge, but a complex environmental and public health issue that requires continuous monitoring, policy-driven interventions, and effective governance.

Key Contributions of the Study

- Development of a reproducible AQI computation framework based on official CPCB standards
- Comparative evaluation of AQI prediction using Linear Regression and Random Forest models

- One-year city-level analysis of air quality trends in Pune, Mumbai, Bengaluru, and Delhi
- Identification of inter-city pollution disparities and seasonal AQI variations
- Integration of open data, literature insights, and real-world environmental context

Overall, this work demonstrates the value of combining open environmental datasets, machine learning techniques, and long-term statistical analysis to enable transparent air quality assessment and support data-driven urban environmental planning in India.

REFERENCES

- [1] Government Open Data (OGD) Platform India, “Air Quality Data.” [Online]. Available: <https://data.gov.in/>
- [2] Central Pollution Control Board (CPCB), “National Air Quality Index (NAQI) Methodology.” [Online]. Available: <https://cpcb.nic.in/>
- [3] S. Bawane, P. Chaudhary, S. K. Jain, and J. S. Dodiya, “Forecasting of air quality index using machine learning and deep learning models,” *J. Neonatal Surg.*, vol. 14, no. 18S, pp. 1147–1155, 2025. [Online]. Available: <https://www.jneonatsurg.com/index.php/jns/article/view/5850>
- [4] S. Varshney, J. N. Shrivastava, and N. Gupta, “Transforming Air Quality Index Prediction Using Machine Learning: Key Insights from the Taj Trapezium Zone (TTZ),” *Eng. Technol. Appl. Sci. Res.*, vol. 15, no. 5, pp. 26741–26749, 2025. [Online]. Available: <https://www.etasr.com/index.php/ETASR/article/view/12168>
- [5] C. T. Meenakshi and J. Amutha, “Forecasting of Air Quality Index Using Machine Learning Techniques for Chennai City,” *Int. J. Res. Publ. Rev.*, vol. 5, no. 10, pp. 912–922, 2024. [Online]. Available: <https://ijrpr.com/uploads/V5ISSUE10/IJRPR33905.pdf>
- [6] K. K. Sidhu, H. Balogun, and K. O. Oseni, “Predictive modelling of Air Quality Index (AQI) across diverse cities and states of India using machine learning: Investigating the influence of Punjab’s stubble burning on AQI variability,” *Int. J. Manage. Inf. Technol.*, vol. 16, no. 1, 2024. [Online]. Available: <https://doi.org/10.5121/ijmit.2024.16102>
- [7] A. Choudhary *et al.*, “Evaluating air quality and criteria pollutants prediction disparities by data mining along a stretch of urban–rural agglomeration including coal-mine belts and thermal power plants,” *Front. Environ. Sci.*, vol. 11, Art. 1132159, 2023. [Online]. Available: <https://doi.org/10.3389/fenvs.2023.1132159>
- [8] A. Gangwar *et al.*, “The state-of-the-art in air pollution monitoring and forecasting systems using IoT, big data, and machine learning,” *Wireless Pers. Commun.*, 2023. [Online]. Available: <https://doi.org/10.1007/s11277-023-10351-1>
- [9] K. Kumar and B. P. Pande, “Air pollution prediction with machine learning: A case study of Indian cities,” *Int. J. Environ. Sci. Technol.*, vol. 20, no. 5, pp. 5333–5348, 2023. [Online]. Available: <https://doi.org/10.1007/s13762-022-04241-5>

- [10] G. Ravindiran *et al.*, “Air quality prediction by machine learning models: A predictive study on the Indian coastal city of Visakhapatnam,” *Chemosphere*, vol. 338, p. 139518, 2023. [Online]. Available: <https://doi.org/10.1016/j.chemosphere.2023.139518>
- [11] N. N. Maltare and S. Vahora, “Air Quality Index prediction using machine learning for Ahmedabad city,” *Digital Chem. Eng.*, vol. 7, Art. 100093, 2023. [Online]. Available: <https://doi.org/10.1016/j.dche.2023.100093>
- [12] S. Senthivel and M. Chidambaranathan, “Machine learning approaches used for air quality forecast: A review,” *Revue d’Intelligence Artificielle*, vol. 36, no. 1, pp. 73–78, 2022. [Online]. Available: <https://doi.org/10.18280/ria.360108>
- [13] Pune Suffers From Poor Air Quality as Diwali Firecracker Celebrations Begin Amid Temperature Drop | Pune News - Times of India
<https://timesofindia.indiatimes.com/city/pune/pune-suffers-from-poor-air-quality-as-diwali-firecracker-celebrations-begin-amid-temperature-drop/articleshow/114599184.cms>
- [14] Pune’s ‘moderate’ air quality expected to deteriorate in next 24 hrs | Hindustan Times
<https://www.hindustantimes.com/cities/pune-news/punes-moderate-air-quality-expected-to-deteriorate-in-next-24-hrs-101730402772698.html>
- [15] Diwali Fireworks Contribute to Deteriorating Air Quality in Pune - Punekar News
<https://www.punekarnews.in/diwali-fireworks-contribute-to-deteriorating-air-quality-in-pune/>
- [16] On processes controlling fine particulate matters in four Indian megacities - ScienceDirect
<https://www.sciencedirect.com/science/article/abs/pii/S0269749119315799>
- [17] Temperature reduction to worsen AQI in days ahead | Mumbai news
<https://www.hindustantimes.com/cities/mumbai-news/temperature-reduction-to-worsen-aqi-in-days-ahead-101729882596088.html>
- [18] Climate change has made it harder to breathe in Mumbai | The Indian Express
<https://indianexpress.com/article/opinion/columns/breathless-in-mumbai-8998013/>
- [19] Stubble burning: Delhi’s toxic air crisis needs tough action | Policy Circle
<https://www.policycircle.org/environment/stubble-burning-delhi-air/>
- [20] ‘Death in the Air’: Life in Delhi, the world’s most polluted city | Health | Al Jazeera
<https://www.aljazeera.com/features/2024/11/29/death-in-the-air-how-is-life-different-in-worlds-most-polluted-city>
- [21] Delhi Pollution, AQI: Cold Wave, Toxic Air: Why Delhi's Pollution Peaks in November-December
<https://www.ndtv.com/india-news/delhi-pollution-aqi-cold-wave-toxic-air-why-delhis-pollution-peaks-in-november-december-9758456>
- [22] Delhi Chokes On Most Polluted January Air In 2 Years As AQI Plunges To 440; Dense Fog Delays Flights
<https://news.abplive.com/cities/delhi-chokes-on-most-polluted-january-air-in-2-years-as-aqi-plunges-to-440-dense-fog-delays-flights-1822434>

[23] Bengaluru Air Quality Decline in March 2024

<https://bangaloremirror.indiatimes.com/bangalore/cover-story/bengaluru-air-quality-decline-in-march-2024/articleshow/109411169.cms>

[24] Bengaluru's Air Pollution Fund: Only 37% Utilised Despite High Pollution Levels | Bengaluru News - Times of India

<https://timesofindia.indiatimes.com/city/bengaluru/bengalurus-air-pollution-fund-only-37-utilised-despite-high-pollution-levels/articleshow/114663096.cms>

[25] Take my breath away

<https://bangaloremirror.indiatimes.com/bangalore/cover-story/take-my-breath-away/articleshow/125398743.cms>

[26] Delhi action plan: 15% cut in average AQI targeted in 2026, focus on vehicles and dust control | India News

<https://www.hindustantimes.com/india-news/delhi-action-plan-15-cut-in-average-aqi-targeted-in-2026-focus-on-vehicles-and-dust-control-air-quality-101770599523542.html>