International Journal of Technology and Applied Science (IJTAS)

E-ISSN: 2230-9004 ● Website: www.ijtas.com ● Email: editor@ijtas.com

Machine Learning Based Air Quality Assessment of Jabalpur City Madhya Pradesh India

Priya Mishra[1], Dr. D.C. Rahi[2]

[1]Research scholar, ME (Environmental Engineering) Jabalpur Engineering College, Jabalpur (M.P) 482011, India.
[2]Assistant Professor, Department of Civil Engineering, Jabalpur Engineering College, Jabalpur (M.P) 482011, India

Abstract:
Air pollution is a major environmental and public health concern, making accurate air quality assessment and prediction essential. This study evaluates the ambient air quality of Jabalpur, Madhya Pradesh, using data from five CAAQMS stations: Govind Bhavan Colony, Gupteshwar, Marhatal, Suhagi, and Vijay Nagar. Major pollutants (PM_{10}, $PM_{2.5}$, SO_2, NO_2, CO, NH_3, and O_3) were analyzed for 2024–2026 using descriptive statistics and CPCB-based Air Quality Index (AQI) calculations. Six machine learning algorithms in WEKA—Gaussian Process, Linear Regression, IBk, M5Rules, M5P Model Tree, and Random Tree—were evaluated using CC, MAE, and RMSE. The M5P Model Tree achieved the best performance, with a CC of 0.9794, MAE of 3.4989, and RMSE of 5.1670. PM_{10} and $PM_{2.5}$ were the dominant pollutants across the monitoring stations, while winter (November–February) recorded the highest AQI due to atmospheric inversion, low wind speed, and reduced pollutant dispersion. The findings demonstrate the potential of machine learning for reliable AQI prediction and effective air quality management in Jabalpur.

Keywords: Air Pollution, Air Quality Index (AQI), PM_{10}, $PM_{2.5}$, CAAQMS, Machine Learning, M5P Model Tree.

1. Introduction:

Air pollution is one of the major environmental challenges affecting human health, ecosystems, and sustainable urban development. Rapid urbanization, industrialization, increasing vehicular traffic, construction activities, and the growing demand for energy have resulted in increased emissions of various pollutants into the atmosphere. Air pollutants can adversely affect respiratory and cardiovascular health and may also cause damage to vegetation, soil, water resources, buildings, and biodiversity. The increasing concentration of pollutants in urban environments has therefore created a need for continuous monitoring and reliable assessment of ambient air quality.

Air pollution consists of both primary and secondary pollutants. Primary pollutants are emitted directly from sources such as vehicles, industries, power plants, construction activities, and biomass burning. Important primary pollutants include particulate matter (PM_{10} and $PM_{2.5}$), sulphur dioxide (SO_2), nitrogen oxides (NO_x), carbon monoxide (CO), ammonia (NH_3), and volatile organic compounds (VOCs). Secondary pollutants are formed in the atmosphere through chemical reactions involving primary pollutants and meteorological factors. Ground-level ozone (O_3), sulphuric acid, nitric acid, and peroxyacetyl nitrate are examples of secondary pollutants. Among these pollutants, particulate matter is of particular concern because fine particles can remain suspended in the atmosphere and penetrate deeply into the respiratory system.

The sources of air pollution can generally be classified as point, line, and area sources. Industrial plants, power stations, and large emission stacks are examples of point sources, while vehicular traffic and transportation corridors represent line sources. Domestic activities, construction, agricultural operations, and other widely distributed activities may act as area sources. The contribution of these sources varies spatially and temporally and is also influenced by meteorological conditions such as wind speed, temperature, atmospheric stability, and rainfall. In urban areas, poor dispersion conditions can result in the accumulation of pollutants near the ground, leading to deterioration of ambient air quality.

Air pollution also produces several environmental impacts. Sulphur and nitrogen compounds can contribute to acid deposition, which may alter soil chemistry and damage aquatic ecosystems and vegetation. Excess nitrogen deposition can contribute to eutrophication, while ground-level ozone can cause injury to plant tissues and reduce photosynthetic activity and agricultural productivity. Particulate matter can reduce visibility, interfere with solar radiation reaching vegetation, and transport toxic substances and heavy metals. Therefore, monitoring individual pollutants alone may not provide an easily understandable representation of the overall air quality experienced by the public.

The Air Quality Index (AQI) provides a simplified numerical representation of air pollution by combining information from different pollutants into a single value. In India, the National Air Quality Index (NAQI) was developed to communicate air quality conditions in an understandable form. The AQI considers major pollutants such as $PM_{2.5}$, PM_{10}, NO_2, SO_2, CO, O_3, NH_3, lead, and benzene. AQI categories help indicate whether air quality is satisfactory or poses increasing levels of health concern. The Central Pollution Control Board (CPCB) plays an important role in air quality monitoring and assessment in India.

Continuous Ambient Air Quality Monitoring Stations (CAAQMS) provide valuable pollutant concentration data for assessing temporal and spatial variations in urban air quality. However, air quality is highly variable because pollutant concentrations are affected by emission sources, seasonal changes, atmospheric conditions, and local activities. In particular, winter conditions can promote pollutant accumulation because of lower wind speeds, reduced atmospheric mixing, and temperature inversion. Consequently, conventional monitoring provides information about existing conditions, but accurate prediction of AQI can further support early assessment and effective environmental management.

In recent years, machine learning techniques have gained importance in environmental research because of their ability to identify complex relationships between multiple variables and develop predictive models from historical observations. Algorithms such as regression models, nearest-neighbour methods, decision trees, and model trees can be applied to air quality datasets to predict AQI and evaluate the influence of different pollutants. Such approaches can complement conventional monitoring and provide useful information for pollution management and urban planning.

The present study focuses on the ambient air quality of Jabalpur, Madhya Pradesh, using pollutant data obtained from five CAAQMS locations: Govind Bhavan Colony, Gupteshwar, Marhatal, Suhagi, and Vijay Nagar. The study analyzes PM_{10}, $PM_{2.5}$, SO_2, NO_2, CO, NH_3, and O_3 for the period 2024–

2026. Descriptive statistical analysis and CPCB-based AQI assessment are used to examine the spatial and seasonal characteristics of air quality. Furthermore, six machine learning algorithms—Gaussian Process, Linear Regression, IBk, M5Rules, M5P Model Tree, and Random Tree—are evaluated for AQI prediction using standard performance measures. The study aims to identify the major pollutants influencing air quality in Jabalpur and determine a suitable machine learning approach for reliable AQI prediction.

2.OBJECTIVE

To assess the ambient air quality of Jabalpur using data collected from Real-Time Monitoring (RTM) and manual monitoring stations. To analyze the concentration levels of major air pollutants such as $PM_{2.5}$, PM_{10}, SO_2, NO_2, CO, O_3, and NH_3. To calculate and evaluate the Air Quality Index (AQI) based on CPCB guidelines. To preprocess and prepare the air quality dataset for machine learning analysis using WEKA software. To develop and compare different machine learning models available in WEKA for air quality prediction and classification. To evaluate the performance of machine learning algorithms using suitable statistical measures such as accuracy, precision, recall, F-measure, and ROC analysis. To identify seasonal variations and pollution trends in different monitoring locations of Jabalpur. To provide recommendations for improving air quality management and supporting environmental decision-making in Jabalpur.

3.SCOPE

The present study focuses on assessing the ambient air quality of Jabalpur city, Madhya Pradesh, using machine learning techniques implemented through WEKA software. The study involves the collection and analysis of air quality data from Real-Time Monitoring (RTM) stations and manual monitoring stations. Major air pollutants such as $PM_{2.5}$, PM_{10}, SO_2, NO_2, CO, O_3, and NH_3 are considered for evaluating the Air Quality Index (AQI).

The collected data are preprocessed to remove missing values, duplicate records, and outliers before being used for machine learning analysis. Various classification and prediction algorithms available in WEKA are applied to identify pollution patterns, compare model performance, and evaluate the accuracy of AQI prediction. The study also examines seasonal variations in pollutant concentrations and identifies major pollution hotspots within Jabalpur.

The findings of this research will contribute to understanding the air quality status of Jabalpur and demonstrate the usefulness of machine learning techniques in environmental monitoring. The results may support policymakers, environmental agencies, and urban planners in developing effective air pollution control strategies and sustainable environmental management practices.

4. STUDY AREA

The present study was conducted in Jabalpur, Madhya Pradesh, an important administrative, educational, industrial, commercial, and cultural centre of central India. Jabalpur is located in the Mahakoshal region and has experienced rapid urbanization, industrial development, increasing vehicular traffic, and construction activities. These factors have contributed to growing environmental and air-quality concerns, making the city suitable for studying ambient air quality and applying machine learning techniques for AQI prediction.

Five monitoring locations representing different land-use categories were selected: Gupteshwar

(Residential), Marhatal (Commercial), Govind Bhavan Colony (Silence Zone), Vijay Nagar (Residential), and Suhagi (Industrial Area). Gupteshwar is primarily a residential locality with moderate traffic, while Marhatal is a busy commercial area characterized by shops, markets, offices, and heavy vehicular movement. Govind Bhavan Colony is classified as a silence zone because of nearby educational and sensitive establishments. Vijay Nagar is a developing residential area with houses, apartments, schools, and commercial facilities. Suhagi represents an industrial environment with manufacturing units, warehouses, transportation activities, and relatively high vehicle movement.

Jabalpur lies in the eastern part of Madhya Pradesh, approximately between 23°10′ and 23°17′ N latitude and 79°52′ and 80°02′ E longitude. The district covers about 5,210 km², while the urban area covers approximately 263 km². It is centrally connected to several important cities, including Bhopal, Nagpur, Prayagraj, and Raipur. The region contains plains, hills, plateaus, and river valleys, with elevations ranging from approximately 390 to 420 m. The Narmada Valley is a major geographical feature of the region.

The geology of Jabalpur includes Precambrian rocks, Vindhyan formations, Lameta beds, Deccan Trap basalts, and alluvial deposits. The Narmada River is the principal water body and is supported by tributaries such as the Hiran, Gour, and Bargi. The soils include black cotton, red and yellow, alluvial, and sandy loam types, supporting agriculture and natural vegetation.

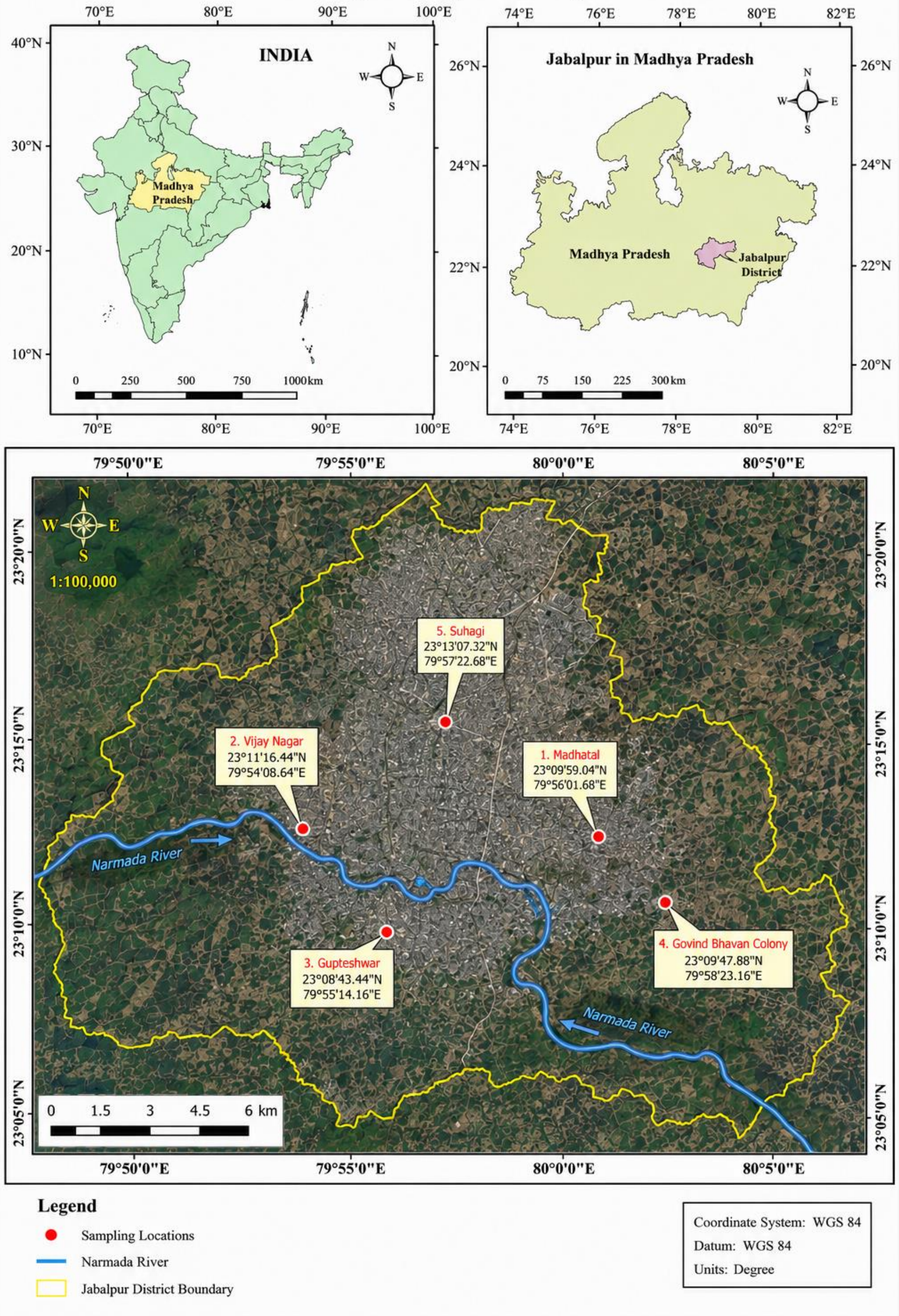
Jabalpur has a tropical wet and dry climate with three major seasons: hot and dry summer, monsoon, and cool winter. Summer temperatures may reach 30–45°C, while winter temperatures generally range from 8–25°C. Annual rainfall is approximately 1,200–1,400 mm, mainly received during the monsoon. Meteorological factors such as temperature, wind speed, wind direction, humidity, and rainfall strongly influence pollutant dispersion and accumulation.

Air quality in Jabalpur is influenced by vehicular emissions, industrial activities, construction dust, and road dust. $PM_{2.5}$ and PM_{10} concentrations are particularly important in industrial and traffic-intensive areas. Air-quality data from the five selected monitoring stations were used to analyze pollutant trends, seasonal variations, AQI, and machine-learning-based prediction for supporting effective environmental management and planning.

International Journal of Technology and Applied Science (IJTAS)

E-ISSN: 2230-9004 ● Website: www.ijtas.com ● Email: editor@ijtas.com

Fig. 1 Study Area Map



International Journal of Technology and Applied Science (IJTAS)
E-ISSN: 2230-9004 ● Website: www.ijtas.com ● Email: editor@ijtas.com

5.METHODOLOGY

The study follows a systematic methodology for assessing and predicting the ambient air quality of Jabalpur using Real-Time Monitoring (RTM)/Continuous Ambient Air Quality Monitoring Stations (CAAQMS) and using manual methods and assessment using machine learning techniques. The major steps include data collection, validation, preprocessing, AQI calculation, dataset preparation, machine learning modelling, performance evaluation, and interpretation of results.

Air quality data were collected from four RTM stations and one manual monitoring station in Jabalpur. The monitoring stations continuously measure major pollutants, including $PM_{2.5}$, PM_{10}, SO_2, NO_2, CO, O_3, and NH_3, along with relevant meteorological parameters such as temperature, relative humidity, wind speed, wind direction, rainfall, and solar radiation. RTM stations automatically transmit recorded observations to a centralized system, where the data are validated and used for AQI assessment.

Before analysis, the collected data were checked for completeness and accuracy. Duplicate records, missing values, invalid measurements, abnormal observations, and outliers were identified and treated during preprocessing. The cleaned dataset was then organized into a structured format suitable for statistical analysis and machine learning. The dataset was prepared in formats such as CSV/ARFF for use in WEKA.

Data Collection and Quality Control

RTM stations use automated analyzers to continuously measure pollutant concentrations and transmit data to a centralized database. Different analytical principles are used for different pollutants, such as beta attenuation for particulate matter, ultraviolet fluorescence for SO_2, chemiluminescence for NO_2 and NH_3, non-dispersive infrared methods for CO, and UV photometry for O_3. Manual monitoring was also conducted using standard instruments and methods, including Respirable Dust Sampler (RDS) and $PM_{2.5}$ samplers. Proper calibration, maintenance, sampling procedures, laboratory analysis, and quality-control practices were followed. Monitoring sites were selected to provide adequate airflow and avoid direct interference from buildings, trees, or nearby pollution sources. Data were further checked for abnormal values, detection-limit issues, and consistency between particulate measurements.

AQI Calculation

The Air Quality Index (AQI) was calculated according to CPCB guidelines. Individual pollutant sub-indices were calculated using the respective concentration breakpoints. For AQI calculation, valid data for at least three pollutants are required, including either $PM_{2.5}$ or PM_{10}, with a minimum of 16 hours of valid observations. CO and O_3 are calculated using 8-hour averages, while the remaining pollutants generally use 24-hour averages. The final AQI is determined by the highest pollutant sub-index, representing the pollutant having the greatest influence on the overall air quality.

The AQI values are classified into categories such as Good, Satisfactory, Moderately Polluted, Poor, Very Poor, and Severe, allowing the air quality condition and associated health impacts to be communicated more effectively.

$$I_p = \frac{I_{HI} - I_{LO}}{B_{HI} - B_{LO}}\left(C_p - B_{LO}\right) + I_{LO}$$

They just mean:

- **Cp** = your measured concentration = **63.5**
- **BLO** = lower concentration breakpoint = **51**
- **BHI** = upper concentration breakpoint = **100**

- **ILO** = lower AQI breakpoint = **51**
- **IHI** = upper AQI breakpoint = **100** Put those numbers into the formula:

$$I_p = \frac{100-51}{100-51}(63.5-51)+51$$

Since:

$$\frac{49}{49} = 1$$

we get:

$$I_p = 1(12.5) + 51$$
$$I_p = 63.5$$

Rounded:

$$\boxed{PM_{10}\ Sub\text{-}Index = 64}$$

Think of it this way

The breakpoint table is basically a **conversion table**.

You start with:

Pollutant concentration → find its range → convert it to an AQI sub-index

You repeat that for PM_{10}, $PM_{2.5}$, SO_2, NO_2, CO, etc.

Then:

$$\boxed{\text{Final AQI} = \text{highest sub-index}}$$

PM_{10} wins with **64**, so the whole site's AQI becomes **64 (Satisfactory)**.

Machine Learning Using WEKA

After preprocessing, the dataset was imported into WEKA (Waikato Environment for Knowledge Analysis) for machine learning analysis. WEKA provides tools for data visualization, attribute selection, classification, clustering, and model evaluation. Feature selection was performed to identify important pollutant parameters, remove irrelevant variables, reduce computational complexity, minimize overfitting, and improve prediction performance.

Several machine learning algorithms can be applied to the prepared air quality dataset, including J48 Decision Tree, Random Forest, Naïve Bayes, Support Vector Machine (SMO), and Multilayer Perceptron. In the broader AQI prediction analysis, models were trained using historical pollutant data to establish relationships between pollutant concentrations and AQI.

Model Evaluation

The prepared dataset was divided for model training and testing, with 10-fold cross-validation used for reliable performance evaluation. Predicted AQI values were compared with observed values. Model performance was assessed using measures such as accuracy, precision, recall, F-measure, confusion matrix, MAE, RMSE, and ROC-AUC, depending on the type of model and analysis.

The model providing the most reliable predictions with lower error was considered the most suitable for AQI prediction. Finally, the model outputs were interpreted to identify pollution patterns, important

International Journal of Technology and Applied Science (IJTAS)

E-ISSN: 2230-9004 ● Website: www.ijtas.com ● Email: editor@ijtas.com

pollutants, seasonal trends, and the usefulness of machine learning for air quality monitoring and environmental decision-making.

Overall Methodological Workflow:

RTM &Manual Monitoring Stations

↓

Air Quality Data Collection

↓

Data Validation and Quality Checking

↓

Data Preprocessing

↓

Removal of Missing/Duplicate/Invalid Data

↓

Dataset Preparation

↓

AQI Calculation Using CPCB Guidelines

↓

Feature Selection

↓

Import Dataset into WEKA

↓

Machine Learning Model Development

↓

Training and Cross-Validation

↓

Performance Evaluation using Accuracy, MAE, RMSE, etc.

↓

AQI Prediction and Classification

↓

Interpretation of Pollution Patterns and Results

International Journal of Technology and Applied Science (IJTAS)
E-ISSN: 2230-9004 ● Website: www.ijtas.com ● Email: editor@ijtas.com

6.RESULT & DISCUSSION

the results and discussion of the ambient air quality analysis conducted at selected monitoring stations in Jabalpur. The study examined major air pollutants, including PM_{10}, $PM_{2.5}$, SO_2, NO_2/NOx, CO, O_3, and NH_3. A dataset containing 216 instances and 9 attributes was analyzed to understand pollutant concentration, spatial variation, seasonal patterns, and overall Air Quality Index (AQI). AQI was calculated according to CPCB guidelines to assess air quality and identify the pollutants mainly responsible for deterioration.

Machine learning techniques available in WEKA were also applied to predict AQI. Six algorithms—Gaussian Process, Linear Regression, IBk (K-Nearest Neighbours), M5Rules, M5P, and Random Tree—were evaluated using correlation coefficient, Mean Absolute Error (MAE), and Root Mean Squared Error (RMSE). Correlation indicates how closely predicted values match actual AQI, while lower MAE and RMSE indicate better prediction accuracy. Among the models, M5P performed best, achieving a correlation coefficient of 0.9794, MAE of 3.4989, and RMSE of 5.1670. IBk and M5Rules also performed well, while Gaussian Process showed comparatively higher errors. M5P was therefore selected as the most suitable model for AQI prediction because it provided the highest correlation and lowest prediction errors.

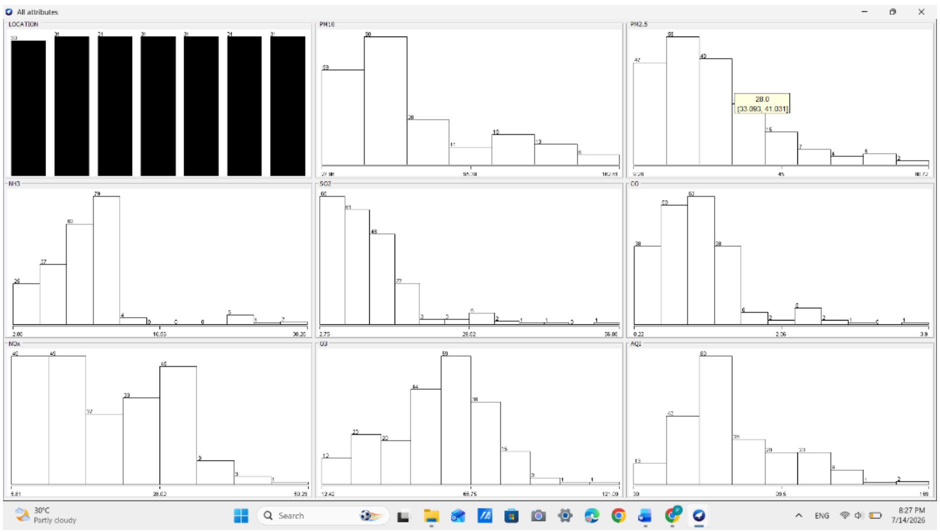
The spatial analysis showed considerable differences in pollutant concentrations among the five monitoring locations: Govind Bhavan Colony (GBC), Gupteश्वर, Marhatal, Suhagi, and Vijay Nagar. PM_{10} was identified as the most frequently prominent pollutant across all locations in the 2026 analysis. High particulate concentrations were associated mainly with road dust, construction activities, exposed soil, traffic-related resuspension, commercial activity, domestic fuel use, and waste or biomass burning.

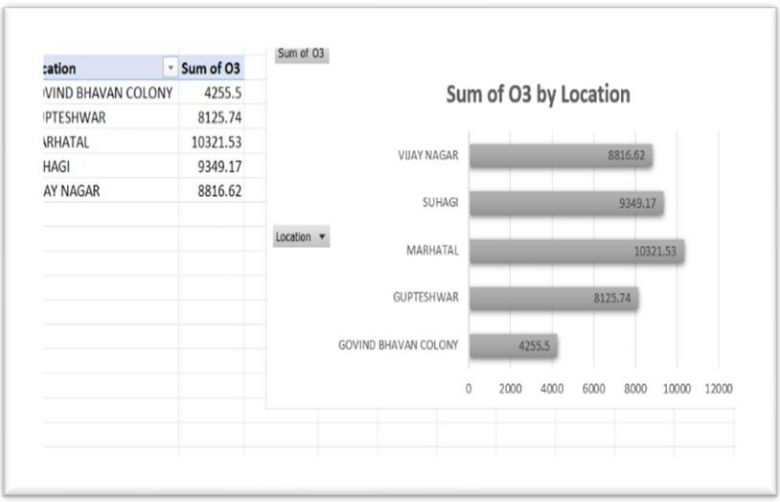
Seasonal analysis demonstrated a clear relationship between AQI and meteorological conditions. Winter generally experienced the poorest air quality because low temperatures, calm winds, morning fog, and temperature inversions restrict pollutant dispersion and cause accumulation near the ground. Summer is characterized by high temperatures, low humidity, moderate winds, increased dust generation, and enhanced photochemical activity, which can increase ozone formation. During the monsoon, heavy rainfall and high humidity remove particulate matter through wet deposition, resulting in substantially lower PM_{10}, $PM_{2.5}$, and AQI values.

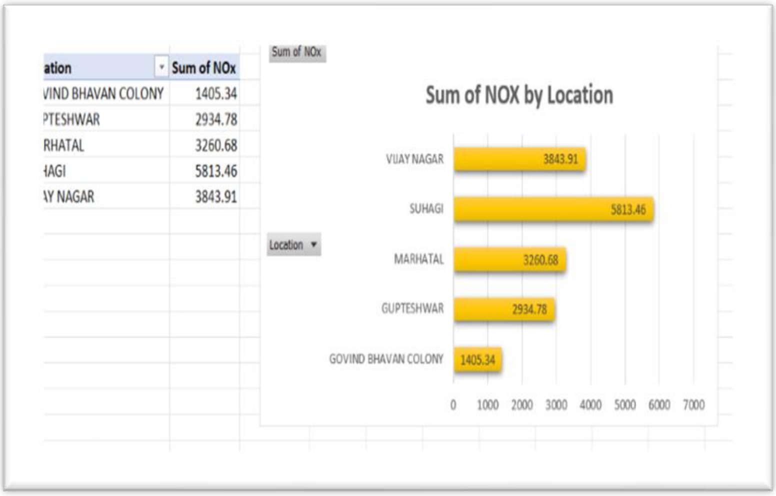
Comparison of monthly AQI values for 2024 and 2025 showed a recurring seasonal pattern. Marhatal generally recorded the highest AQI, indicating comparatively poorer air quality, while Govind Bhavan Colony generally recorded the lowest AQI. In 2025, Marhatal recorded an average AQI of approximately 105–110, whereas GBC remained around 61. The highest values generally occurred during winter, particularly in November and December, while the lowest values occurred during the monsoon months.

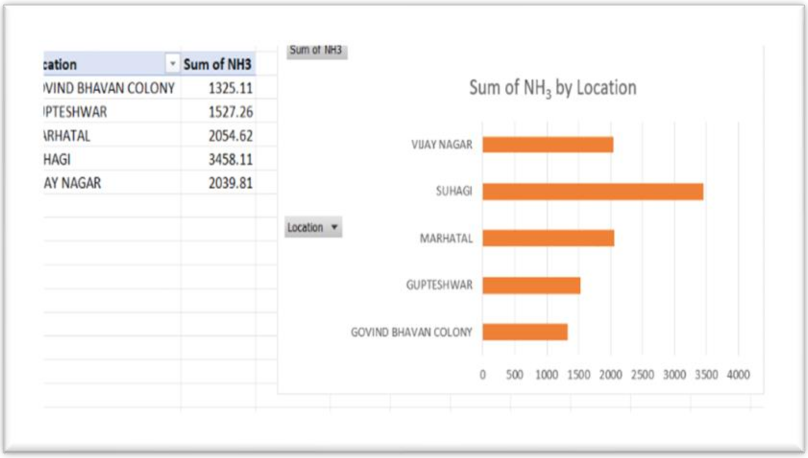
The 2025 analysis further indicated that PM_{10} dominated most locations, while $PM_{2.5}$ became more significant during winter and post-monsoon periods. O_3 occasionally became prominent during summer because of enhanced photochemical reactions, while CO showed occasional dominance during certain monsoon months.

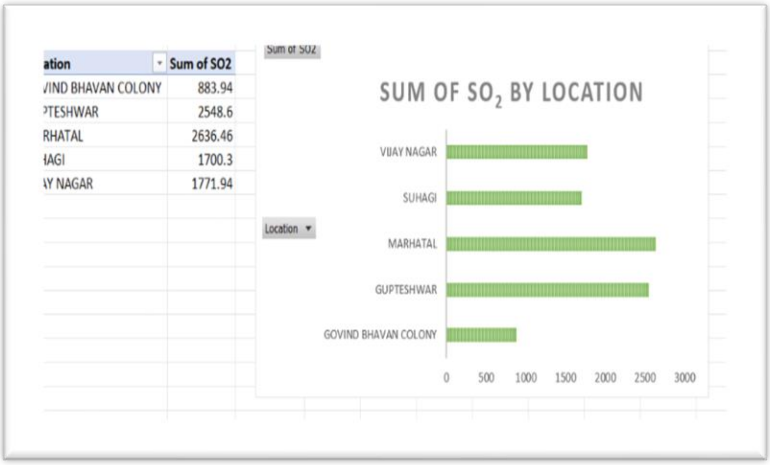
Overall, the study concludes that meteorological conditions control the broad seasonal pattern of AQI, while local anthropogenic activities determine the magnitude and persistence of pollution. Traffic emissions, road dust, construction, industrial activity, domestic fuel combustion, biomass and waste burning, population density, and urban expansion are important contributors. The findings emphasize the need for location-specific pollution-control measures, including dust suppression, traffic management, improved waste handling, control of construction emissions, and continuous air-quality monitoring. The successful performance of the M5P model also demonstrates the potential of machine learning for accurate AQI prediction and air-quality forecasting.



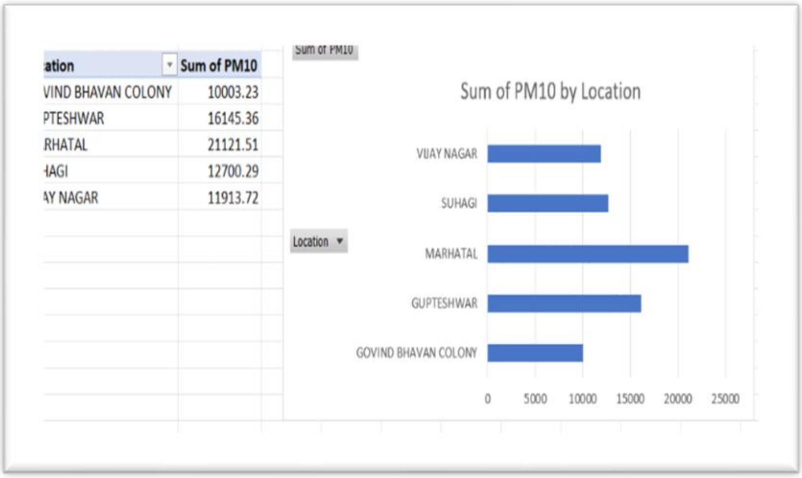



ation
Sum of NOx
VIND BHAVAN COLONY 1405.34
PTESHWAR 2934.78
RHATAL 3260.68
HAGI 5813.46
AY NAGAR 3843.91
Sum of NOx
Sum of NOX by Location
Location
VIJAY NAGAR 3843.91
SUHAGI 5813.46
MARHATAL 3260.68
GUPTESHWAR 2934.78
GOVIND BHAVAN COLONY 1405.34
0 1000 2000 3000 4000 5000 6000 7000

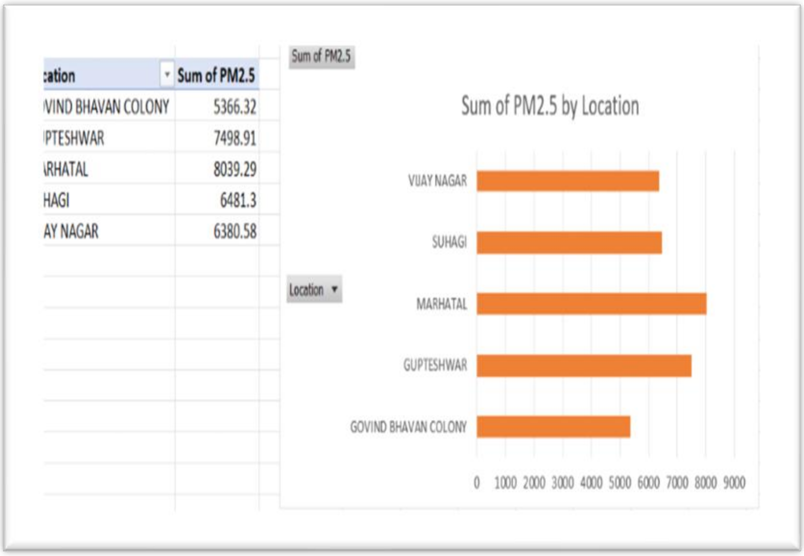

cation
Sum of NH3
VIND BHAVAN COLONY 1325.11
IPTESHWAR 1527.26
ARHATAL 2054.62
HAGI 3458.11
AY NAGAR 2039.81
Sum of NH3
Sum of NH3 by Location
Location
VIJAY NAGAR
SUHAGI
MARHATAL
GUPTESHWAR
GOVIND BHAVAN COLONY
0 500 1000 1500 2000 2500 3000 3500 4000


ation
Sum of SO2
VIND BHAVAN COLONY 883.94
PTESHWAR 2548.6
RHATAL 2636.46
HAGI 1700.3
AY NAGAR 1771.94
Sum of SO2
SUM OF SO2 BY LOCATION
Location
VIJAY NAGAR
SUHAGI
MARHATAL
GUPTESHWAR
GOVIND BHAVAN COLONY
0 500 1000 1500 2000 2500 3000

International Journal of Technology and Applied Science (IJTAS)
E-ISSN: 2230-9004 ● Website: www.ijtas.com ● Email: editor@ijtas.com

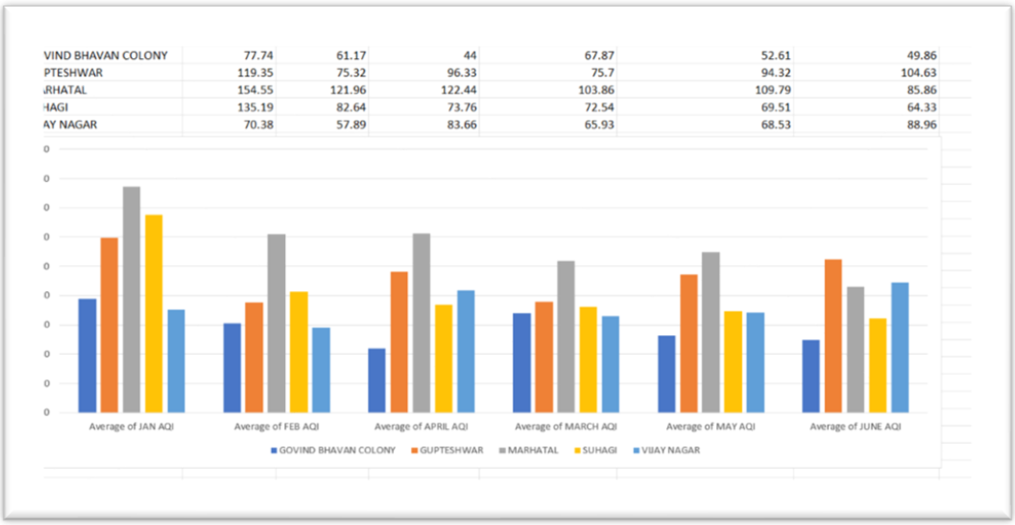


ation	Sum of PM10
VIND BHAVAN COLONY	10003.23
PTESHWAR	16145.36
RHATAL	21121.51
HAGI	12700.29
AY NAGAR	11913.72



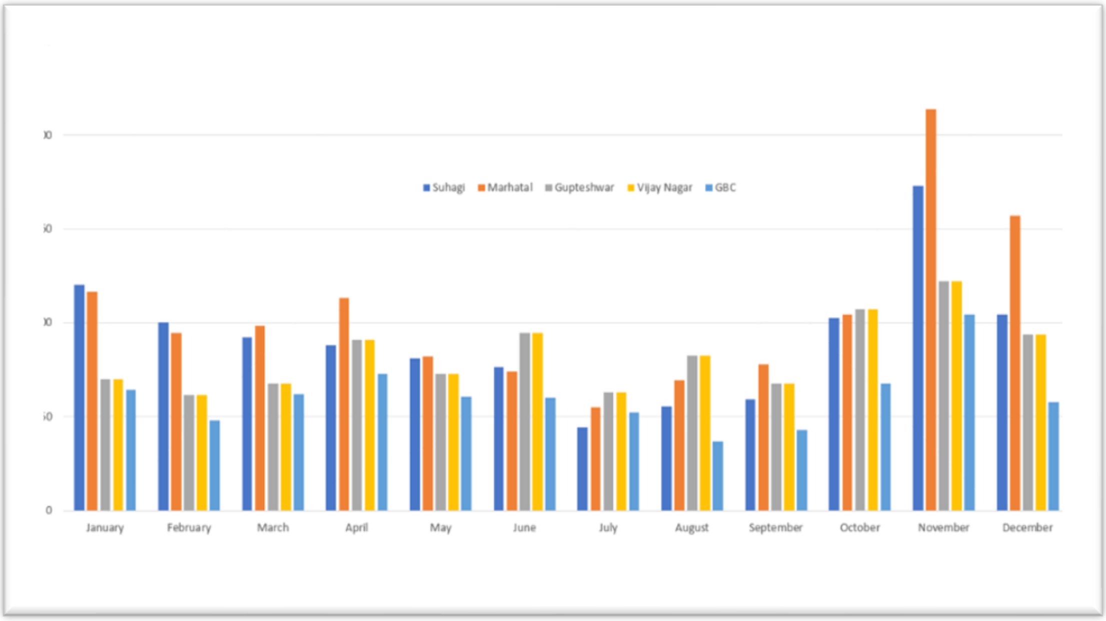
ation	Sum of PM2.5
VIND BHAVAN COLONY	5366.32
PTESHWAR	7498.91
RHATAL	8039.29
HAGI	6481.3
AY NAGAR	6380.58

2026



VIND BHAVAN COLONY	77.74	61.17	44	67.87	52.61	49.86
PTESHWAR	119.35	75.32	96.33	75.7	94.32	104.63
RHATAL	154.55	121.96	122.44	103.86	109.79	85.86
HAGI	135.19	82.64	73.76	72.54	69.51	64.33
AY NAGAR	70.38	57.89	83.66	65.93	68.53	88.96

2025


Suhagi
Marhatal
Gupteshwar
Vijay Nagar
GBC
0
January
February
March
April
May
June
July
August
September
October
November
December

2024

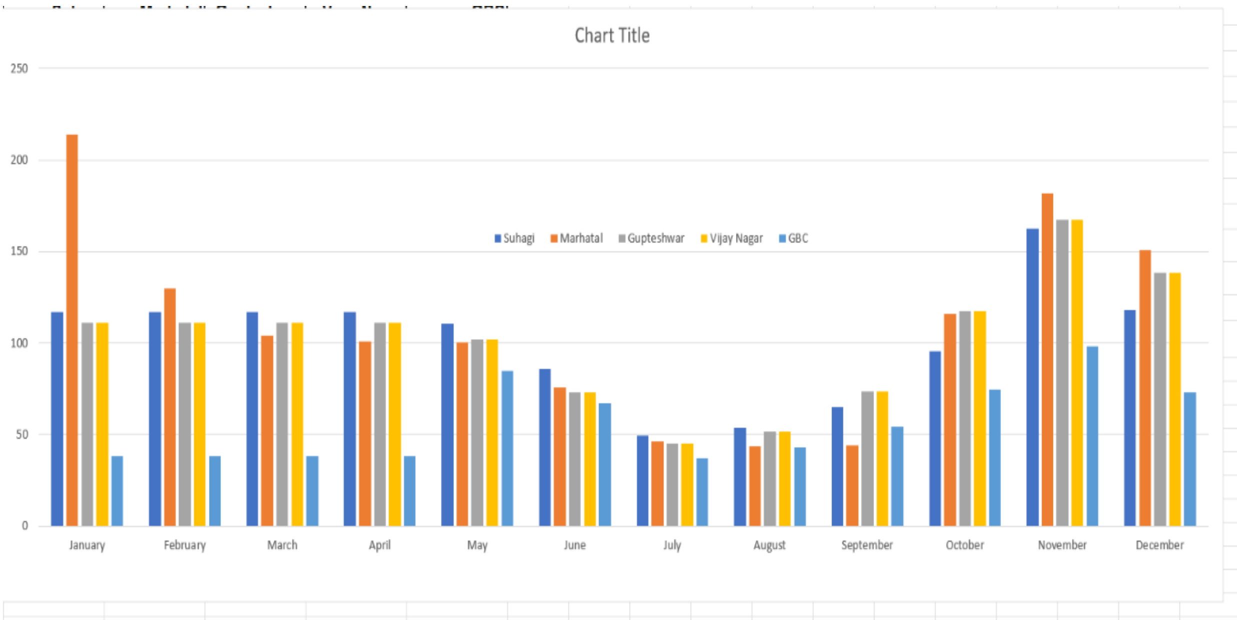


Chart Title
250
200
150
100
50
0
Suhagi
Marhatal
Gupteshwar
Vijay Nagar
GBC
January
February
March
April
May
June
July
August
September
October
November
December

7.CONCLUSION

The study successfully evaluated ambient air quality in Jabalpur using data from five monitoring stations and analyzed major pollutants such as PM10, PM2.5, SO_2, NO_2, CO, NH_3, and O_3. PM10 and PM2.5 were the dominant pollutants, contributing most to AQI deterioration. Higher pollution was associated with traffic, construction, commercial activities, road dust, and residential emissions. M5P Model Tree performed best among the six machine-learning algorithms, achieving the highest correlation (0.9794), lowest MAE (3.4989), and lowest RMSE (5.1670). It outperformed IBk, M5Rules, Linear Regression, Gaussian Process Regression, and Random Tree.

Overall Performance Comparison

Algorithm	Correlation (↑ Higher is Better)	MAE (↓ Lower is Better)	RMSE (↓ Lower is Better)	Performance
Gaussian Process	0.9195	7.1612	10.1781	★ Good
Linear Regression	0.9442	5.8740	8.3775	★★ Better
kNN IBk	0.9500	5.4306	8.0476	★★★ Very Good
M5Rules	0.9758	3.8430	5.5655	★ Good
M5P	0.9794	3.4989	5.1670	Best
Random Tree	0.9301	5.5780	9.4470	★ Good

Note: ↑ Higher correlation is better. ↓ Lower MAE and RMSE are better.

The study found a strong seasonal pattern:

Winter: Highest AQI because of low temperature, weak winds, and atmospheric inversion, which trap pollutants. Pre-monsoon: Moderate pollution, with road dust, construction, traffic, and increased temperature contributing to pollution. Monsoon: Lowest AQI and cleanest period, mainly because rainfall removes particulate matter and improves atmospheric mixing. Post-monsoon: AQI gradually worsens due to reduced rainfall, lower wind speeds, biomass burning, and renewed accumulation of particles.

Marhatal consistently had the highest pollution/AQI, largely because of heavy traffic, commercial activity, construction, and dense urban activity. Govind Bhavan Colony generally had the lowest AQI, reflecting comparatively cleaner surroundings and fewer emission sources. 2024–2026 comparison showed that the overall seasonal pattern remained consistent, although pollution levels varied by location and year. The available January–June 2026 data also followed the established pattern. Overall, meteorological conditions strongly control seasonal air-quality variation, while human activities determine the intensity and persistence of pollution at individual locations.

The study concludes that PM10 and PM2.5 are the major causes of poor air quality in Jabalpur, pollution is highest in winter and lowest during the monsoon, Marhatal is the most polluted location, Govind Bhavan Colony is the cleanest, and the M5P Model Tree is the most effective AQI prediction model.

“The recommendations of the study focus on strengthening environmental policies, controlling pollution sources, promoting clean energy and sustainable transportation, improving air-quality monitoring, and using meteorological data for better AQI forecasting. The main limitations are the availability of limited and sometimes inconsistent historical data, the restriction to a specific region and time period, and the exclusion of certain important factors such as traffic, industrial activities, biomass burning, and population growth. In the future, the study can be extended by using larger multi-year and multi-location datasets, advanced machine-learning and deep-learning algorithms, real-time IoT-based monitoring, and AI-integrated web, mobile, and smart-city applications to achieve more accurate, reliable, and practical AQI prediction."

REFERENCES:

1. Agrawal, Jitendra. Analysis of Clustering Algorithm of WEKA Tool on Air Pollution Dataset. International Journal of Computer Applications, June 2017.
2. Cai, M., Yin, Y., & Xie, M. (2009). Prediction of hourly air pollutant concentrations near urban arterials using artificial neural network approach. Transportation Research Part D: Transport and Environment, 14(1), 32–41.
3. Cannon, J., & Lord, E. R. (2000). Forecasting Summertime Surface-Level Ozone Concentrations in the Lower Fraser Valley of British Columbia: An Ensemble Neural Network Approach. Journal of the Air & Waste Management Association, 50(3), 322–339.
4. Department of the Environment and Heritage. (2005). Ground-level ozone (O3) – Air quality fact sheet.
5. Gomez-Sanchis, J., Martín-Guerrero, J., Soria-Olivas, E., Vila-Francés, J., Carrasco, J., & Valle-Tascón, S. (2006). Neural networks for analysing the relevance of input variables in the prediction of tropospheric ozone concentration. Atmospheric Environment, 40(32), 6173–6180.
6. Google Earth.
7. Jiang, N., & Riley, M. L. (2015). Exploring the Utility of the Random Forest Method for Forecasting Ozone Pollution in Sydney. Journal of Environment Protection and Sustainable Development, 1(5), 245–254.
8. Kolehmainen, M., Martikainen, H., & Ruuskanen, J. (2001). Neural networks and periodic components used in air quality forecasting. Atmospheric Environment, 35(5), 815–825.
9. Kurt, A., Gulbagci, B., Karaca, F., & Alagha, O. (2008). An online air pollution forecasting system using neural networks. Environment International, 34(5), 592–598.
10. Luna, S., Paredes, M. L. L., de Oliveira, G. C. G., & Corrêa, S. M. (2014). Prediction of ozone concentration in tropospheric levels using artificial neural networks and support vector machine at Rio de Janeiro, Brazil. Atmospheric Environment, 98, 98–104.
11. Madhya Pradesh Pollution Control Board.
12. Microsoft Excel.com
13. Mohamed Shakir, & N. Rakesh. (2018). Investigation on Air Pollutant Data Sets using Data Mining Tool. Proceedings of the Second International Conference on I-SMAC (IEEE).
14. Noorani Subramanian, Varun. (2016). Data Analysis for Predicting Air Pollutant Concentration in Smart City Uppsala.
15. Pérez, P., Trier, A., & Reyes, J. (2000). Prediction of PM2.5 concentrations several hours in advance using neural networks in Santiago, Chile. Atmospheric Environment, 34(8), 1189–1196.
16. Schlink, U., Dorling, S., Pelikan, E., Nunnari, G., Cawley, G., Junninen, H., Greig, A., Foxall, R., Eben, K., Chatterton, T., Vondráček, J., Richter, M., Dostal, M., Bertucco, L., Kolehmainen, M., & Doyle, M. (2003). A rigorous inter-comparison of ground-level ozone predictions. Atmospheric Environment, 37(23), 3237–3253.

17. Singh, K. P., Gupta, S., & Rai, P. (2013). Identifying pollution sources and predicting urban air quality using ensemble learning methods. Atmospheric Environment, 80, 426–437.
18. Subramanian, Varun Noorani. (2016). Data Analysis for Predicting Air Pollutant Concentration in Smart City Uppsala.
19. Tüfekci, P. (2014). Prediction of full load electrical power output of a base load operated combined cycle power plant using machine learning methods. International Journal of Electrical Power & Energy Systems, 60, 126–140.
20. U.S. Environmental Protection Agency. (2003). Guidelines for Developing an Air Quality (Ozone and PM2.5) Forecasting Program.
21. Vitkar, Swati. (2017). Comparative Analysis of Various Data Mining Prediction Algorithms, Demonstrated using Air Pollution Data of Navi Mumbai. Research Journal of Chemical and Environmental Sciences.
22. Wan, F., & Lei, K. S. (2008). Adaptive Neuro-Fuzzy Approach to Air Pollution Prediction in Macau.
23. Wang, W., Men, C., & Lu, W. (2008). Online prediction model based on support vector machine. Neurocomputing, 71(4–6), 550–558.
24. WEKA – The University of Waikato. WEKA 3: Data Mining with Open-Source Machine Learning Software in Java.
25. Environment Australia. Ground-level ozone (O3) Fact Sheet. Available at: http://www.environment.gov.au/protection/publications/factsheet-ground-level-ozone-o3
26. WEKA Project. WEKA Software. Available at: Ahmed, A.A.M., Jui, S.J.J., Sharma, E., Ahmed, M.H., Raj, N., Bose, A., 2024.