STUBBLE BURNING AND ITS IMPACT IN DELHI'S AIR POLLUTION OF INDIA: PREDICTIVE APPROACH USING MACHINE LEARNING

AKILA, R.* – BALASAKTHIPRIYAN, M.

Department of Mathematics, Anna University, Chennai – 600025, Tamilnadu, India (e-mail: balasakthipriyan2030@gmail.com; phone: +91-7708-589-929; ORCID: 0009-0008-1265-7022)

**Corresponding author
e-mail: akilarajamanickam@yahoo.com; phone: +91-9677-123-627; ORCID: 0000-0002-5757-8132*

(Received 8th Apr 2025; accepted 16th Jun 2025)

Abstract. Stubble burning, a common agricultural practice in northern India, particularly in Punjab and Haryana, has been identified as a major contributor to Delhi's air pollution of India. This paper uses a predictive method based on machine learning to examine and measure how stubble burning affects Delhi's air pollution. Using the K-means algorithm, the strategy clusters Air Quality Index(AQI) data over a period of years stretching from March 2015 to January 2025, hence revealing a clear cluster with much higher pollution levels that often shows during the stubble burning season. Moreover, a study of fire count data and Air Quality Index trends for the years from March 2015 to January 2025 verifies a clear relationship that when fire count rises, Air Quality Index levels also increase. A Gradient Boosting Regression model is trained on pre-stubble burning data and tested on post-stubble burning data, with comparisons between expected and actual Air Quality Index values in Delhi, forecasting the extent of stubble burning affecting air quality. The findings show the AQI change per 1% fire count increase varies between 0.08% and 0.38%, showing a consistent but varying impact. This confirms that emissions from stubble burning in Punjab and Haryana contribute substantially to seasonal pollution spikes in Delhi. In addition to evaluating sources' contributions, research has proposed mitigating strategies to lower total air pollution levels in urban areas.
Keywords: *fire count, Delhi air quality, gradient boosting regression, K-means clustering, air quality index prediction*

Introduction

With Delhi being among the most impacted cities, air pollution has become one of the most serious environmental issues in India. Particularly in the post-monsoon months, when a significant rise in pollutant levels drives the Air Quality Index (AQI) into dangerous ranges, the city suffers concerning levels of air pollution. Among the various causes, stubble burning in the surrounding states of Punjab, Haryana, and parts of Uttar Pradesh significantly aggravates Delhi's air quality problem.

Farmers often choose this economical and efficient agricultural technique, which burns crop waste following paddy harvest. But it causes significant emissions of pollutants like fine particle matter ($PM_{2.5}$ and PM_{10}), carbon monoxide (CO), nitrogen oxides (NOx), and greenhouse gasses including ozone (O_3) and carbon dioxide (CO_2), etc.

In light of the difficulties in measuring PM at ground level, characterizing its chemical makeup, and the time and resources needed for this process, researchers have also looked into using chemical dispersion modeling to simulate the movement of smoke and other air pollutants (Beig et al., 2019; Ghude et al., 2020; Sembhi et al., 2020; Rahman et al., 2022; Govardhan et al., 2023a,b). Moreover, aerosol movement was tracked by examining satellite data in the form of AOD values, meteorological data, and ground-

http://www.aloki.hu ● ISSN 1589 1623 (Print) ● ISSN1785 0037 (Online)
DOI: http://dx.doi.org/10.15666/aeer/2304_79357956

level $PM_{2.5}$ concentrations (Chauhan and Singh, 2017; Jethva et al., 2018; Nair et al., 2020a). These contaminants greatly impair air quality, hence promoting smog and having major environ mental and health effects.

Stubble burning has become an annual event; pollution levels in Delhi rise nearly immediately following burning activity escalation in Punjab and Haryana. post-2015, taking into account the extreme air pollution incident that occurred in November 2016; initially documented and depicted in Delhi (Mukherjee et al., 2018; Sawlani et al., 2019; Ganguly et al., 2019; Kanawade et al., 2020). Since the first episode was seen in 2016 and revealed a dense layer of smog covering the city, the number of studies has steadily increased after 2016–17. Since then, it has been treated as a yearly occurrence during this time, and it is frequently linked to fire count information obtained from Punjabi and Haryana's agricultural fields. According to Dey et al. (2012), the most widely used evaluation instruments for this correlation were meteorological variability, pollutant concentrations tracked at CAAQMS, and aerosol optical depth (AOD). Chauhan and Singh (2017) investigated the impact of crop residue burning activities and the accompanying Diwali festival on air quality in the IGP Region for the year 2016 following the initial onset of an episodic situation. The ground-based air quality index (AQI) and MODIS satellite data (AOD and Angstrom Exponent, AE) were utilized to determine the atmospheric phenomenon's variability. October through November saw higher AOD (average range of 0.79–1.05) and AE (average range of 0.90–1.14) values, which served as the basis for reports of smoke plume transport.

According to one study by Beig et al. (2019), emissions from distant areas related to smoke and dust were responsible for 65% of the total, according to the simulations. The majority of the dust, according to the study, comes from transboundary (outside of India) sources, which raises the emissions from stubble burning over Punjab and Haryana at a height of 3000 meters. It also claimed that such an event causes a sharp increase and rapid decrease in concentration. Nevertheless, the emissions from burning stubble are ground-based and lack sufficient buoyancy to raise them above 3000 meters. Later, Beig et al. (2020) measured the effects of biomass-burning activities in Delhi using air quality data, satellite data, and ground-based incidents. Because of its strong reliance on the weather-controlled air mass transportation pathway, the contribution was observed to fluctuate daily, peaking at 58%. When local emissions rose as a result of festivals in the first week of November, the contribution peaked. Ghude et al. (2020) recently used $PM_{2.5}$ simulations to make Early Warning predictions for Delhi City.

The research used k-means clustering techniques to segment pollution data based on similarity, providing a deeper understanding of pollution-related health risks (Riches et al., 2022). A work Investigated air pollution forecasting in Delhi using data analysis tools. Their study emphasized the impact of rapid urbanization and industrialization on deteriorating air quality. The researchers analyzed historical pollution levels, highlighting the necessity for improved policy measures and real-time monitoring strategies to mitigate rising air pollution (Sharma et al., 2018). An article Examined the effects of crop residue burning on air quality in India. Their study found that biomass burning significantly contributed to the increase in $PM_{2.5}$ and PM_{10} concentrations, exacerbating pollution levels, particularly in regions such as Punjab, Haryana, and Uttar Pradesh. They suggested alternative residue management techniques to address this problem (Lan et al., 2022).

The study looked at how firecrackers and stubble burning affected Delhi's air quality. While firecracker emissions during festivals caused a temporary spike in pollution, their

http://www.aloki.hu ● ISSN 1589 1623 (Print) ● ISSN1785 0037 (Online)
DOI: http://dx.doi.org/10.15666/aeer/2304_79357956

research indicated that stubble burning added 50–75% to the rise in $PM_{2.5}$ levels. Their results underlined the pressing need for rigorous regulatory policies and sustainable substitutes to reduce sporadic pollution incidents (Khan et al., 2023). The paper used machine learning techniques to examine Delhi's air pollution. Their study underlined the significance of predictive analytics in projecting air quality from past data. Using random forest, support vector machine, regression, and classification algorithms, they examined major pollutants including $PM_{2.5}$, PM_{10}, CO, NO_2, SO_2, and O_3 to forecast ambient air pollution levels (Sinha and Singh, 2020). The paper looked at how well K-Means and K-Medoids clustering techniques forecasted air pollution. Focusing on NO_2 and SO_2 levels, their research drew on air quality data from India's National Centre for Medium Range Weather Forecasting. The findings showed that both clustering methods successfully grouped pollution data, hence supporting pattern recognition and forecasting. The study highlighted the need of data mining methods in air pollution research (Suganya et al., 2020).

Regression models were used in an article's analysis of air quality index prediction. Their results demonstrated how well machine learning methods like random forest regression and multivariate linear regression can forecast trends in air pollution. Their research showed that data-driven methods could enhance strategies for managing air quality (Suroshe et al., 2022). Using time-series analysis, the study looked into historical correlation models of air quality. They developed the Gaussian Hidden Markov Model (GHMM) to analyze AQI data, which successfully captures temporal dependencies in air pollution levels. The necessity of time-series forecasting in air quality monitoring was highlighted by their study (Liu et al., 2024). Using hybrid models that combine grey wolf optimization with decision tree regression and deep learning techniques like long short-term memory (LSTM) networks, the study proposed the best machine learning model for AQI prediction. In important Indian cities, their study demonstrated that hybrid AI-driven models improved prediction accuracy (Natarajan et al., 2024). The report included a case study that examined the impact of stubble burning on Delhi's air quality. According to their findings, burning crop residue significantly impacted seasonal increases in $PM_{2.5}$ and PM_{10} levels, resulting in significant air pollution events during the post-monsoon months. The authors suggested technological alternatives and legislative measures to mitigate the adverse effects of burning biomass (Singh et al., 2021).

It can be concluded that despite extensive efforts to ascertain the potential causes of elevated concentrations in Delhi, India, during the post-monsoon season, significant variability and specific limitations persist within the respective studies. The studies definitively established the impact of stubble burning activities occurring in the neighboring urban areas; however, only a limited number of studies examined fire count data or fire emissions to derive conclusions. Most of them were broad and looked at the relationship between fire count, weather, and PM levels in general. Additionally, most studies either used AOD, AE, and SSA driven data to back up their results or used trajectory analysis to figure out how air masses moved to find their source. Some studies have shown that the WRF-Chem model isn't very good at predicting PM concentrations and that an extra statistical analysis is needed to get good results. No research has corroborated the fire count data with ground-level recorded data within the state. The Delhi NCR region has a huge network of strong and calibrated continuous monitoring stations that collect data every hour. This data will be looked at and compared for both stubble and non-stubble periods, as well as for stubble and non-stubble burning areas. To understand and lessen the effect of stubble burning on Delhi's air quality, which is a long-

http://www.aloki.hu ● ISSN 1589 1623 (Print) ● ISSN1785 0037 (Online)
DOI: http://dx.doi.org/10.15666/aeer/2304_79357956

term problem, we need a comprehensive, data-driven plan. This study employs machine learning techniques to analyze and predict the extent of pollution resulting from stubble burning, utilizing data from various sources, including AQI values and satellite-derived fire count data.

This paper uses a predictive method based on machine learning to examine and measure how stubble burning affects Delhi's air pollution. Using the K-means algorithm, the strategy clusters Air Quality Index data across a period of years stretching from March 2015 to January 2025, hence exposing a clear cluster with much higher pollution levels that often shows during the stubble burning season. Moreover, a study of fire count data and Air Quality Index trends for the years from March 2015 to January 2025 verifies a clear relationship that when fire count rises, Air Quality Index levels also increase.

The aims of the study are to use Machine learning technics Gradient Boosting Regression to predictively model the effect of stubble burning on Delhi's air quality. The difference between forecasted and actual AQI values is a measure of the contribution of external pollution sources by training the model on pre-stubble burning AQI data and evaluating it against post-burning data. This predictive modeling technique emphasizes the effect of stubble burning of Punjab and Haryana into the degree of the air pollution in Delhi. The results of this study emphasize the pressing need for policy actions to minimize the negative consequences of stubble burning on air quality. Important actions to lower pollution levels during the stubble burning season are promoting sustainable farming practices, increasing access to alternative residue management technology, and enforcing tighter policies. Furthermore, public awareness initiatives and financial incentives for farmers to switch to non-burning techniques have to be enhanced to promote long-term behavioral change.

Materials and methods

The post-monsoon season is the most polluted time of year in Delhi NCR, India. There are some days when pollution levels are very high, which happens during festivals and when people in Punjab and Haryana burn more stubble. From October to November, there are many festivals, starting with Dussehra and continuing with Dhanteras, Diwali, and Chhatt puja. During this time, local activities and sources start to grow more quickly (Ghei and Sane, 2018; Daga et al., 2019; Ganguly et al., 2019; Singh and Srivastava, 2020). Researchers have previously noted that biomass combustion and household emissions substantially influence ambient $PM_{2.5}$ concentrations. Some studies utilized fire emissions data to quantify the $PM_{2.5}$ load and its carbon content, which were correlated with the combustion of crop residues. These were additionally linked to numerical models such as WRF-Chem, utilized for predicting and forecasting pollutant concentration on a regional scale (Chowdhury et al., 2007; CPCB, 2010; Sharma and Dikshit, 2016; ARAI-TERI, 2018).

Later, Liu et al. (2018) determined the seasonal influence of regional biomass burning on PM concentrations in Delhi. The research examined meteorological variables and satellite-derived fire radiative power, establishing correlations with PM concentrations. Nagar and Sharma (2022) recently did a study to see how burning crop waste affects $PM_{2.5}$ in parts of Northern India. They used WRF-Chem and a hybrid model (HM) based on linear regression. The contribution of crop residue burning to $PM_{2.5}$ was 31 ± 16% (in 232 ± 36 PM $\mu g/m^3$) and Secondary Organic Aerosol (SOA) in CRB contributed $PM_{2.5}$ was 18±9%.

http://www.aloki.hu ● ISSN 1589 1623 (Print) ● ISSN1785 0037 (Online)
DOI: http://dx.doi.org/10.15666/aeer/2304_79357956

Bisht et al. (2015) conducted PM 2.5 monitoring in Delhi in 2015, examining its organic carbon (OC) and elemental carbon (EC) content. The high OC/EC value during the day (7.09) and at night (4.55), as well as the strong correlation (0.74 R 2) between the two, suggest that burning agricultural waste and using dirty fuels like wood are possible sources. Concentration weighted trajectory (CWT) analysis confirmed the findings, showing that smoke from the fields of Punjab and Haryana and local activities were the main causes during the post-monsoon season (Bisht et al., 2015). Recently, Singh et al. (2023) performed a source apportionment analysis by tracking volatile organic compounds during the paddy stubble-burning period in the Mohali district of Punjab, recognized as the preeminent stubble-burning state in the nation. They used a positive matrix factorization (PMF) model and said that burning fresh paddy stubble at the monitoring site made up 6% of the total.

Jethva et al. (2018) performed a synergistic analysis of agricultural fires and their effects on air quality in Northern India. The research utilized NASA's A-train satellite data, terrestrial measurements, and back trajectory computations to derive conclusions. A notable impact was indicated by the observed northwesterly wind flow and the variations in fire count and rising AOD derived from satellite data. Bhadauriya et al. (2020) investigated the correlation between the incineration of paddy straw and the escalation of air pollution in Northwestern India in 2020. The authors examined AOD, AE, Single Scattering Albedo, meteorological variables, trajectories, fire count, and ground data to comprehend the synergy. The authors employed the Residue to Grain Ratio (RGR) to quantify emissions from the combustion of paddy residue in Punjab and Haryana, establishing a correlation between rising emissions and increasing concentrations in Delhi.

Data sources

Reliable sources' several datasets are needed to properly examine how stubble burning affects Delhi's air pollution of India. The main data sources for this study are the Air Quality Index (AQI) figures for Delhi, India gathered from other government sources and the Central Pollution Control Board (CPCB). Key pollutants including $PM_{2.5}$, PM_{10}, NO_2, SO_2, CO, and O_3 have hourly and monthly values included in this data. Historical AQI trends in Punjab and Haryana over several years allow us to compare pollution levels before, during, and after the stubble-burning season. Fire count data in Punjab and Haryana is obtained from NASA's VIIRS (Visible Infrared Imaging Radiometer Suite) and MODIS (Moderate Resolution Imaging Spectroradiometer) satellite datasets. This information on fire incidents, their intensity, and geographical position serves to monitor stubble-burning activity in Punjab and Haryana. Fire counts and pollution levels are correlated by analyzing the data during the same time frame as AQI.

Data cleaning

Analysis should be preceded by data cleaning and preprocessing to guarantee correctness and consistency. Sensor failures or holes in data collecting could cause missing AQI and meteorological data points. Missing values are filled using interpolation methods like linear interpolation and forward-fill techniques. In cases where a large portion of data is missing, those records are removed to prevent biased analysis. Since AQI data, fire count data, and meteorological data are obtained from different sources, their timestamps may differ. All datasets are standardized to Indian Standard Time (IST) to maintain uniformity. For different studies, data is resampled to hourly, daily, and

http://www.aloki.hu ● ISSN 1589 1623 (Print) ● ISSN1785 0037 (Online)
DOI: http://dx.doi.org/10.15666/aeer/2304_79357956

monthly intervals. Sensor mistakes, unexpected extreme pollution occurrences, or data recording problems can cause AQI data outliers. Anomalies are found and removed using the z-score approach and the Interquartile Range (IQR) technique. Outlier removal is visualized and validated using box plots and histograms. Smoothing methods like rolling averages are employed to get significant trends since pollution levels might vary with abrupt changes in industry activity or weather. This clarifies the seasonal and recurring impacts of stubble burning on air quality. The study may provide more accurate and consistent findings by guaranteeing high-quality, standardized, and cleansed data, hence improving understanding of the influence of stubble burning on Delhi's air pollution in India.

Clustering data using K-means clustering algorithm

An unsupervised machine learning technique called the K-means clustering algorithm divides a dataset into k separate, non-overlapping clusters. When examining how stubble burning in Punjab and Haryana affects air pollution in Delhi, K-means aims to cluster comparable pollution levels ($PM_{2.5}$, PM_{10}, etc.) together depending on their temporal and pollutant traits so that one can find times of high pollution, especially following stubble burning. K-means clustering partitions the dataset into k clusters, each containing data points comparable to one another based on specific characteristics including air pollution levels, time of day, and month of the year. It iteratively modifies the cluster centroids and uses the Euclidean distance to define similarity between points. Each cluster has a centroid, which is the mean of all the points within the cluster. K-means uses Euclidean distance to assign data points to the closest centroid and update centroids.

Algorithm

1. Initialize Centroids:

Select k random centroids: $C_1, C_2, \ldots., C_k$.

2. Repeat until convergence:

a) Assignment:

Assign each point x_1 to the nearest centroid C_k:

$j = arg \min_{1 \le j \le k} d(x_i - C_j)$

where the Euclidean distance is:

$$d(x_i - C_j) = \sqrt{\sum_{l=1}^{n} (x_{i_l} - y_{j_l})^2}$$

b) Update:

Update centroids:

$$C_j = \frac{1}{|C_j|} \sum_{x_i \in C_j} x_i$$

3. Convergence:

Stop if centroid changes are below a threshold ϵ:

http://www.aloki.hu ● ISSN 1589 1623 (Print) ● ISSN1785 0037 (Online)
DOI: http://dx.doi.org/10.15666/aeer/2304_79357956

$$\sum_{j=1}^{k} \| Cj^{(new)} - Cj^{(old)} \|^2 < \epsilon$$

Or If maximum iterations reached.

4. End:

Return final centroids and cluster assignments.

The technique groups Air Quality Index data in Delhi from March 2015 to January 2025 using the K-means algorithm. This makes it easy to see a definite cluster with substantially higher pollution levels that commonly appears during the stubble burning season.

The K-means algorithm's convergence and cluster formation allow us to read the outcomes as the final centroids signifying the "average" of every data point in that cluster. These centroids help to grasp the general traits of every cluster—e.g., average pollution level, average temperature, or time-related data like hour of the day or month. Examining the pollution values—e.g., $PM_{2.5}$, PM_{10}—in each cluster helps us to determine which clusters match greater pollution times. A cluster can be classified as one indicating post-stubble burning pollution if it has high pollution levels and it relates to the months following stubble burning (October to December). Clusters that exhibit high pollution levels in Delhi during particular months or hours—for example, late fall or evening hours—suggest that some times of the year or day are more prone to pollution spikes in Delhi caused by stubble burning in Punjab and Haryana or higher vehicle emissions. The frequency of various pollution patterns is better understood by the number of data points in each cluster. While smaller clusters can signify uncommon but important pollution events, larger clusters could reflect typical pollution trends. A strong tool for finding trends in pollution data is k-means clustering. K-means clustering lets us group times with comparable pollution traits in the framework of examining how stubble burning in Punjab and Haryan affects air pollution in Delhi, hence highlighting the times of high pollution that follow stubble burning. Temporal and seasonal trends in air pollution levels, the identification of particular times (e.g., months or hours) when stubble burning most affects pollution levels.

Principal component analysis (PCA), reduces the number of dimensions in large datasets to principal components that retain most of the original information. It does this by transforming potentially correlated variables into a smaller set of variables, called principal components and it is used for plotting the result. Consistent segmentation of the data into meaningful groups and successful analysis of pollution trends are possible by means of K-means clustering with a predetermined number of clusters. Particularly at important times like the post-stubble burning months, this approach offers insights that can direct policy interventions and decisions to solve air quality concerns.

Prediction of AQI Data in Delhi of India using Gradient Boosting Regression. This methodology employs Gradient Boosting Regression (GBR) to predict the impact of stubble burning in Punjab and Haryana on air pollution levels in Delhi. The training data comes from the period before stubble burning, and the test data is from after stubble burning. The model is then trained on the pre-stubble burning data and assessed by contrasting the projected values with the actual observed values from the test data (post-stubble burning period). An ensemble learning method called gradient boosting generates weak learners—often decision trees—sequentially, each trying to fix the mistakes of the

http://www.aloki.hu ● ISSN 1589 1623 (Print) ● ISSN1785 0037 (Online)
DOI: http://dx.doi.org/10.15666/aeer/2304_79357956

last one. The objective is to minimize the loss function iteratively. Algorithm of Gradient Boosting Regression is given below.

Start with an initial guess. Begin by predicting the average value of the target:

$$F_0(X) = \frac{1}{N}\sum_{i=1}^{N} y_i \quad \text{(Eq.1)}$$

In *Equation (1)*, $F_0(X)$ is the initial prediction, and y_i is the true value for each data point.

Equation (2) calculate the error (residuals) by finding the difference between the true value and the predicted value (the error or residual):

$$r_i = y_i - F_0(X_i) \quad \text{(Eq.2)}$$

where r_i is the error for the i-th data point, and $F_0(X_i)$ is the initial prediction.

Train a simple model (weak learner). Fit a simple model (like a small decision tree) to predict these errors. This model is called a weak learner $h_m(X)$.

Equation (3) calculate the step size (γ_m) which helps to compute how much to change the model based on the weak learner:

$$\gamma_m = \frac{\sum_{i=1}^{N} r_i . h_m(X_i)}{\sum_{i=1}^{N}(h_m(X_i))^2} \quad \text{(Eq.3)}$$

where γ_m controls the size of the update.

Equation (4) gives the update of the model. Add the weak learner's predictions, scaled by γ_m, to the current model:

$$F_m(X) = F_{m-1}(X) + \gamma_m . h_m(X) \quad \text{(Eq.4)}$$

where $F_m(X)$ is the updated prediction.

Repeat steps 2-5. Keep calculating new errors, training new weak learners, and updating the model for a set number of iterations.

Equation (5) gives Final prediction. After M steps, the final model is:

$$F(X) = F_0(X) + \sum_{m=1}^{M} \gamma_m . h_m(X) \quad \text{(Eq.5)}$$

where $F_0(X)$ is the initial prediction and $F(X)$ is the final prediction after all updates.

Model validation

Model performance is evaluated by comparing the Predicted values with the actual observed values.

The MAE in *Equation (6)* measures the average absolute difference between predicted and actual values:

$$MAE = \frac{1}{n}\sum_{i=1}^{n} |y_i - \hat{y}_i| \quad \text{(Eq.6)}$$

where y_i is the actual value and $\hat{y}_i$ is the predicted value.

http://www.aloki.hu ● ISSN 1589 1623 (Print) ● ISSN1785 0037 (Online)
DOI: http://dx.doi.org/10.15666/aeer/2304_79357956

The MSE in *Equation (7)* measures the average squared difference between actual and predicted values:

$$MSE = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2 \tag{Eq.7}$$

Equation (8) is used to calculate RMSE is the square root of the MSE and provides an error metric in the same units as the target variable:

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2} \tag{Eq.8}$$

Equation (9) gives the R^2value indicates how much variance in the target variable is explained by the model:

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2} \tag{Eq.9}$$

where ȳ is the mean of the actual values. A higher R^2indicates that the model explains more of the variance in the target variable.

Equation (10) gives the correlation coefficient assesses the strength and direction of the linear relationship between the expected and actual data. Ranging from -1 (perfect negative correlation) to +1 (perfect positive correlation), 0 denotes no linear relationship:

$$r = \frac{\sum_{i=1}^{n}(y_i - \bar{y})(\hat{y}_i - \hat{\bar{y}})}{\sum_{i=1}^{n}(y_i - \bar{y})^2 \sum_{i=1}^{n}(\hat{y}_i - \hat{\bar{y}})^2} \tag{Eq.10}$$

where ȳ is the mean of the actual values and $\hat{\bar{y}}$ is the mean of the predicted values. A correlation coefficient closer to 1 indicates that the predictions are highly correlated with the actual values.

Analyzing the impact of fire count on air quality index

With an eye on areas like Haryana and Punjab, the approach for this project is meant to thoroughly investigate the interaction between fire count data, Air Quality Index (AQI), and meteorological variables including wind patterns in order to understand how stubble burning and other influences degrade the air quality in Delhi. To evaluate the impact of both local and regional fire events on the AQI, this study combines several data sources and uses several methods including time-series analysis, statistical assessment, and meteorological modeling. A major component of this study is the division of data into two separate times: before and after the stubble burning season. A focused study of how fire activity in Punjab and Haryana (where stubble burning is common) affects AQI levels in Delhi is made possible by this temporal divide. Relatively constant fire counts and moderate AQI values define the first period—before stubble burning. This is the starting point for comparison and analysis of the typical air quality situation in Delhi. Usually, the second time after stubble burning starts shows a notable rise in fire activity and a related decline in air quality. This period is critical for identifying whether the spike in fire activity leads to a corresponding increase in AQI levels, especially due to particulate matter ($PM_{2.5}$ and PM_{10}) generated by the fires. The specific months associated with stubble burning are identified based on historical data or agricultural calendars. In most

http://www.aloki.hu ● ISSN 1589 1623 (Print) ● ISSN1785 0037 (Online)
DOI: http://dx.doi.org/10.15666/aeer/2304_79357956

cases, stubble burning peaks in late October to early November, making it easy to categorize the before and after periods.

Also, looking at fire count data in Punjab and Haryana and Air Quality Index patterns in Delhi from March 2015 to January 2025 shows that there is a definite link between the two: when fire count goes up, Air Quality Index levels go up also.

Once the data is visualized, statistical techniques are used to evaluate the relationship between fire count and AQI levels. This study aims to measure how fire events affect air quality and to determine whether the increase in fire activity directly connects to the increase in AQI values. The average fire count and AQI values for the "before" and "after" periods are calculated to reveal the larger trends. This provides a picture of the overall rise in fire activity and air pollution levels between the two times. The percentage change is computed using the equation to measure the relative shift in both fire count and AQI between the two times.

$$\text{Percentage Change} = \frac{\text{Value After} - \text{Value Before}}{\text{Value Before}} \text{x}100 \qquad \text{(Eq.11)}$$

where Value After is the average fire count or AQI after the stubble burning season and Value Before is the average fire count or AQI before the stubble burning season.

Equation (11) helps to measure the rise in fire occurrences in Punjab and Haryana and air pollution levels in Delhi during the stubble burning season in relation to the baseline period. The intensity and direction of the link between fire count and AQI is evaluated by means of a correlation study. To measure the linear relationship between the two variables, the Pearson correlation coefficient (r) is determined. A high positive connection would imply that rising fire activity is directly linked to a decline in air quality.

Studying AQI patterns throughout certain times of the year helps one to see if increases in air pollution correspond with good wind conditions for moving smoke and particulate matter from Punjab and Haryana. Verifying the theory that pollution from stubble burning considerably adds to Delhi's air quality problems during the winter season depends on this association. Statistical techniques include correlation analysis or regression models help to identify the intensity and character of these links, hence enabling a quantitative study of the relationship between fire counts and AQI values.

Results and discussion

Python is used in this project to examine how stubble burning affects Delhi's air pollution of India. K-means clustering identifies pollution patterns using hourly and monthly data; Fire count and AQI data are compared to show a correlation; and Gradient Boosting Regression forecasts how stubble burning will affect air quality by comparing the expected and actual values.

Particularly stressing the influence of stubble burning, *Figure 1* is the Principal Component Analysis(PCA) scatter plots provide an in-depth visualization of how Delhi's air quality of India has evolved using K-Means Clustering over a period of years spanning from March 2015 to January 2025, particularly highlighting the impact of stubble burning. AQI data has been classified into six separate pollution levels—"Good," "Satisfactory," "Moderate," "Poor," "Very Poor," and "Severe" using the K-means clustering technique. Projected onto two main components obtained from dimensionality reduction, the depicted data points show hourly or daily AQI values. One important

http://www.aloki.hu ● ISSN 1589 1623 (Print) ● ISSN1785 0037 (Online)
DOI: http://dx.doi.org/10.15666/aeer/2304_79357956

finding throughout all years is the seasonal movement in cluster distribution. A notable amount of the data falls in the green (Good), yellow (Satisfactory), and orange (Moderate) categories during the pre-stubble burning months—March to September—indicating fairly clean air. In the post-stubble burning months (October–February), where red (Poor), dark red (Very Poor), and black (Severe) clusters predominate, nevertheless, a notable shift is clear. This suggests that the stubble burning season coincides with a notable increase in pollution levels.

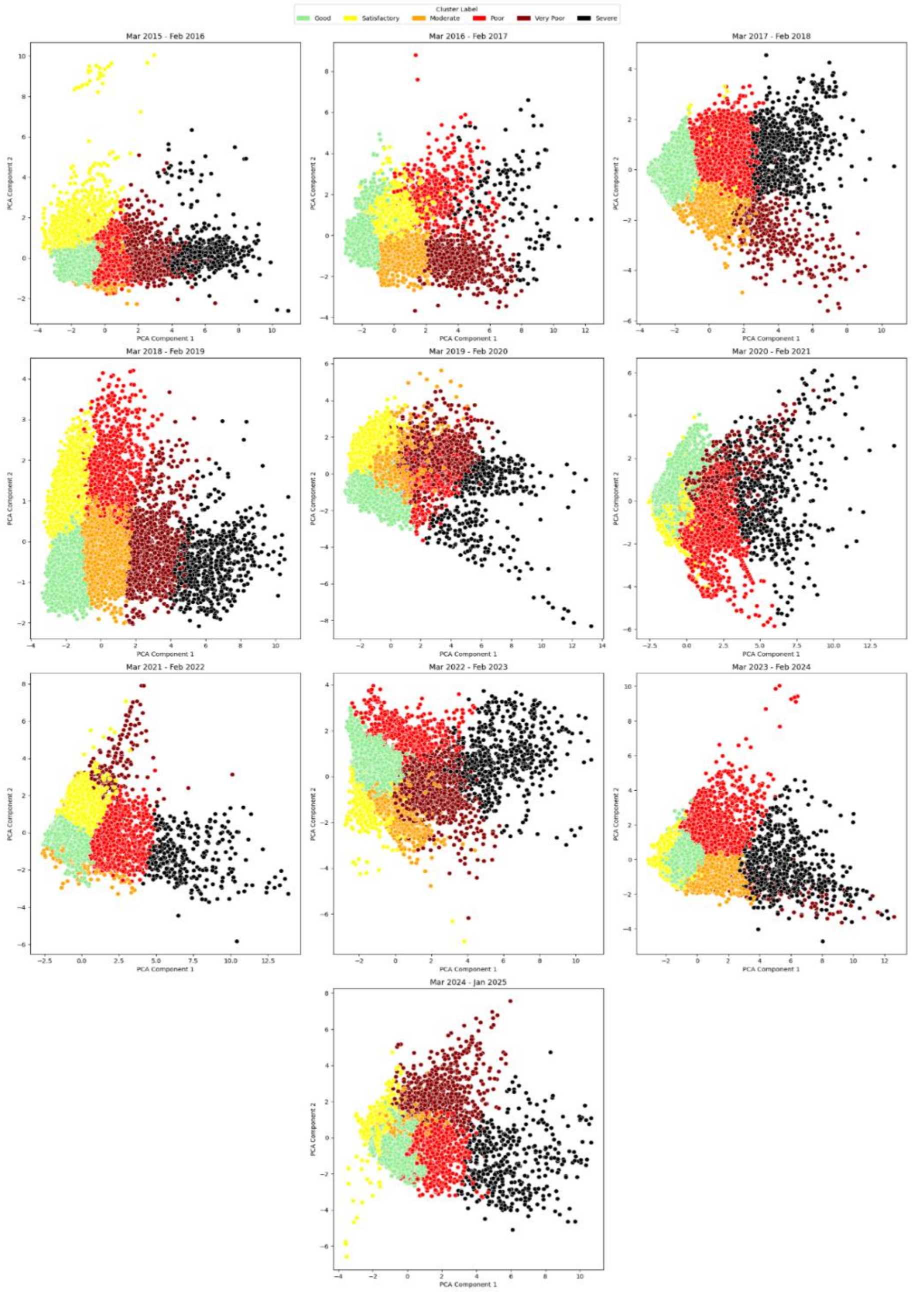

Figure 1. *PCA scatter plots provide an in-depth visualization of Delhi's air quality of India has evolved by classify the AQI data in Delhi into six pollution levels "good" (green), "Satisfactory" (yellow), "Moderate" (orange), "Poor" (red), and "very poor"(dark red) and "Severe"(black) using K-means clustering over a period of years spanning from March 2015 to January 2025*

http://www.aloki.hu ● ISSN 1589 1623 (Print) ● ISSN1785 0037 (Online)
DOI: http://dx.doi.org/10.15666/aeer/2304_79357956

The development of these plots over time also shows an increasing seriousness in air pollution. While pollutant levels rose post-stubble burning, the distribution was well controlled in the initial years (2015–2017). But starting in 2018, the distribution of high-pollution clusters has grown, with black (Severe) pollution events occurring more often and spread over a larger PCA area. This suggests that the intensity and frequency of severe pollution episodes have increased over time. The persistence of these high-pollution clusters well beyond the stubble burning period further indicates that pollutant dispersion is being hindered by meteorological factors such as temperature inversions and stagnant winter air. This phenomenon exacerbates the pollution crisis, leading to prolonged periods of hazardous air quality, which poses severe health risks, particularly for vulnerable populations such as children, the elderly, and individuals with respiratory conditions.

Figure 2 provide a comprehensive month-wise Stacked Bar Graph to analyse the pollution levels in Delhi over a period of years spanning from March 2015 – January 2025, reinforcing the trends observed in the PCA scatter plots. multiple years, reinforcing the trends observed in the PCA scatter plots. These charts clearly depict a recurring seasonal pattern, where air quality remains relatively better during the summer and monsoon months (March–September), but deteriorates sharply in the post-monsoon and winter months (October–February). This seasonal trend aligns with known pollution sources, including the onset of stubble burning in Punjab and Haryana during October–November, as well as the adverse meteorological conditions in winter that trap pollutants closer to the surface.

In the earlier years (2015–2017), the proportion of "Severe" and "Very Poor" AQI days was relatively lower, and the months of October and November showed a moderate increase in pollution levels. However, from 2018 onwards, there is a visible rise in the number of highly polluted days, with the months of October, November, and December consistently showing a sharp surge in "Very Poor" and "Severe" air quality days. Changes in wind patterns, humidity, and temperature can have a big effect on the amount of air pollution. In 2018, bad weather, like low wind speeds, made it easier for pollutants to build up. In 2018, it was found that 4.1 tons of PM2.5 came from the smoke of farm fires in Delhi. This was based on ground-based observations, fire counts, and FRP. The Centre for Science and Environment (CSE) (2022) used SAFAR data on stubble farm fire smoke to Delhi's $PM_{2.5}$ and $PM_{2.5}$ levels to make this estimate. According to WRF-Chem, ground-based observations, and fire counts, stubble burning contributed to air pollution in Delhi in 2018. The amount of pollution changed from day to day, reaching a peak of 58%. This is because the amount of pollution depends on the transportation path, which is affected by the weather. The highest contribution was on November 5, two days after Dhanteras, when local emissions were higher because of festivals. The contribution on November 7, one day after Diwali, was only 8–10% (Beig et al., 2020). In Delhi in 2018, WRF-Chem, AOD, and fire count showed that 25–65% of the pollution came from fires, depending on how active they were and the weather conditions. The mean model bias during Diwali was 170µg/m3 (RMSE =270), which is about 51% of the corresponding observed concentration. This shows how uncertain the WRF-Chem model is on days with high pollution when there is no accurate local emission inventory (Ghude et al., 2020). According to WRF-Chem, FINNEI, and ground-based observations, it was determined that fire emissions accounted for approximately 50–75% (80–120 µg/m^3) of $PM_{2.5}$ in Delhi during 2018, utilizing only a global emission inventory and not a local emission inventory (Kulkarni et al., 2020).

http://www.aloki.hu ● ISSN 1589 1623 (Print) ● ISSN1785 0037 (Online)
DOI: http://dx.doi.org/10.15666/aeer/2304_79357956

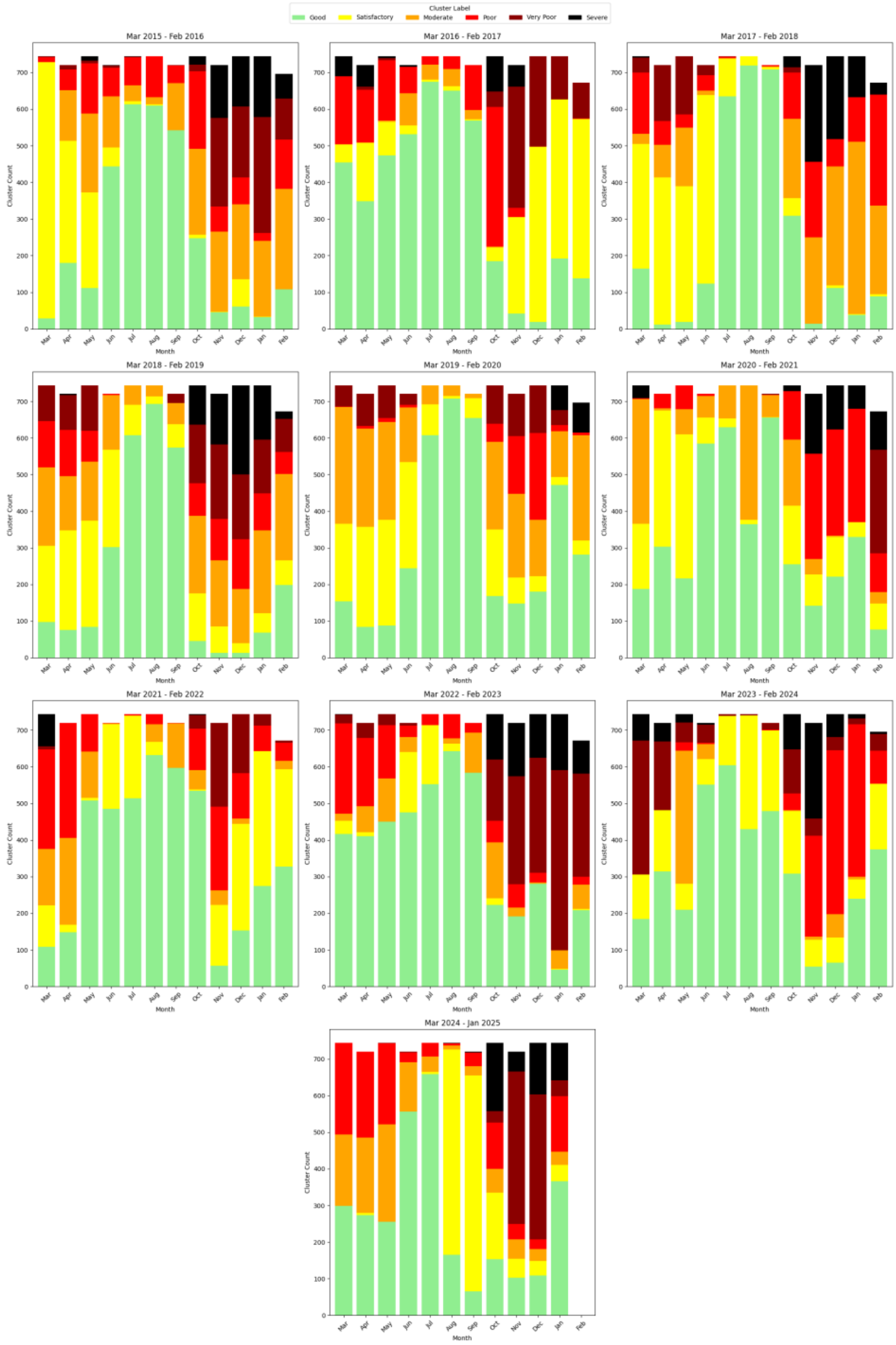

Figure 2. *Comprehensive stacked bar graph on month wise clusters analysis of pollution levels in Delhi over a period of years spanning from March 2015 – January 2025*

Additionally, a critical observation is that pollution levels do not normalize immediately after the stubble burning season ends. Instead, the elevated levels persist well into January and February, suggesting that particulate matter and other pollutants accumulate and remain suspended in the atmosphere for extended periods due to unfavorable winter dispersion conditions. This extended pollution duration highlights the long-term impact of stubble burning, where its effects are not confined to just a few weeks but continue to degrade air quality for months. Furthermore, a comparative analysis

http://www.aloki.hu ● ISSN 1589 1623 (Print) ● ISSN1785 0037 (Online)
DOI: http://dx.doi.org/10.15666/aeer/2304_79357956

across years indicates that while some short-term measures may have been implemented to curb pollution, their effectiveness remains limited, as the overall trend points toward worsening air quality. This necessitates a multi-faceted approach that includes stricter enforcement of stubble burning regulations, large-scale adoption of sustainable agricultural practices, enhanced pollution control measures in urban areas, and improved meteorological forecasting to preemptively mitigate pollution peaks. The evidence presented in these stacked bar charts underscores the urgency of addressing Delhi's air pollution crisis with sustained policy interventions and innovative technological solutions.

The *Figure 3* clearly show a strong link between fire count and air pollution (AQI) in Delhi over multiple years. Each plot represents a different year, where the red line shows the average AQI, and the blue dashed line represents the fire count. The green-shaded area marks the months before stubble burning, while the gray-shaded area shows the period after stubble burning. From March to September, fire counts remain low, and AQI stays at moderate levels. However, starting in October, fire count increases sharply, reaching a peak in November, which matches a significant rise in AQI levels. This pattern repeats every year, proving that stubble burning in Punjab and Haryana plays a major role in increasing Delhi's pollution during winter.

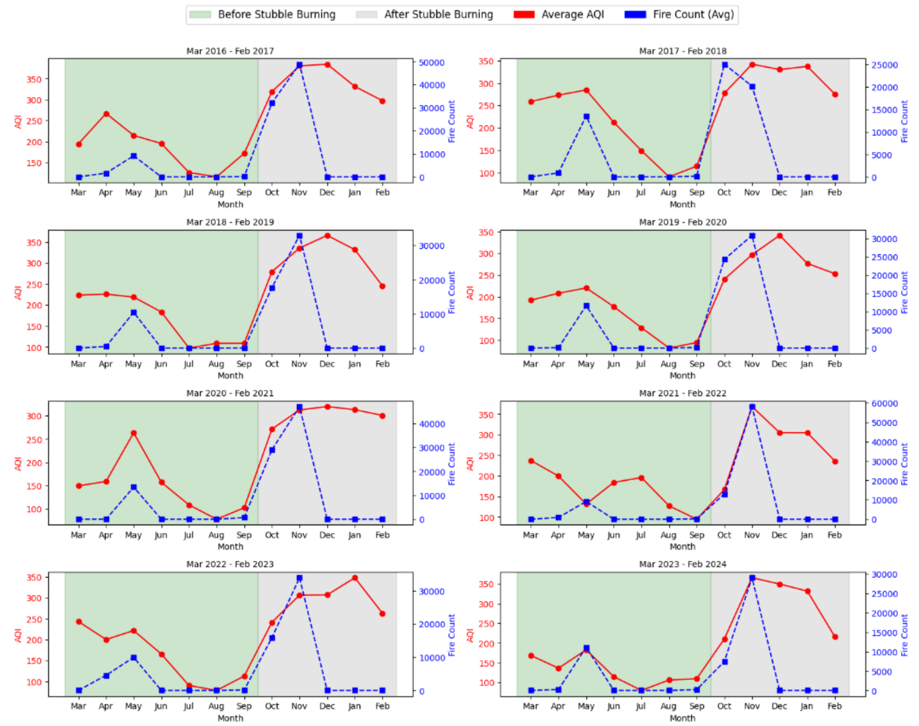

Figure 3. *Yearwise comparison of air quality index and number of fire counts in Punjab during before stuble burning and after stubble burining from the year March 2016 – Febuary 2024*

A key methodology in this analysis is the creation of dual-axis visualizations to compare fire count and AQI over time. This approach allows for the simultaneous display of both fire activity and AQI trends, making it easier to visually assess any relationships between the two variables. The AQI data is plotted against the primary y-axis, providing a measure of the air quality in Delhi. The AQI values indicate the concentration of

http://www.aloki.hu ● ISSN 1589 1623 (Print) ● ISSN1785 0037 (Online)
DOI: http://dx.doi.org/10.15666/aeer/2304_79357956

pollutants, and by plotting the AQI values over time, trends can be observed. A significant increase in AQI during the after-stubble burning period can be indicative of the influence of fire activity on air quality. The fire count data is plotted on a secondary y-axis, allowing for the visualization of the number of fires over the same period. This helps in understanding whether the increase in fire events coincides with the deterioration in air quality. In the graph, the before and after periods of stubble burning are highlighted using different colors (e.g., green for before and gray for after) to distinguish the two phases. This visual cue makes it easy to identify whether there are any marked changes in fire activity and air quality as the stubble burning season progresses.

Over the years, both fire counts and AQI levels have increased, especially in the year 2019, 2020, 2021, and 2022, where fire counts exceeded 40,000–60,000 incidents, leading to very poor or severe AQI levels (above 300-400). This suggests that either stubble burning has intensified, or weather conditions (like low wind speed) have made pollution worse by trapping pollutants in the air. Even after fire counts drop in December and January, pollution remains high, likely due to cold weather and weak winds that keep pollutants from dispersing. The data in the year 2023 and 2024 shows the problem is still ongoing, meaning current efforts to reduce stubble burning have not been fully effective.

Statistical tests are conducted over eight years, from 2016 to 2024, to verify a correlation between the Air Quality Index in Delhi and fire counts in Punjab and Haryana. From 2016 to 2024, the Air Quality Index in Delhi and the number of fires in Punjab and Haryana were related by $r = 0.485$. This demonstrates a moderate positive linear correlation between the Air Quality Index in Delhi and the fire counts in Punjab and Haryana. The inconclusive statistical result is because fires in Punjab probably affect AQI in Delhi with a delay of 1 to 2 weeks, which isn't shown in data from the same month. Fire counts only go up in October and November, but AQI stays high for longer, so a linear/monthly analysis is too simple.

This analysis confirms that stubble burning is a major reason for Delhi's winter air pollution. To fix this, better policies, stricter enforcement, and effective crop residue management alternatives are needed. Also, predictive models using meteorological data could help issue early pollution warnings and take action before the situation worsens. The analysis reveals a strong positive correlation between fire count and AQI levels, indicating the impact of stubble burning on Delhi's air quality. Across multiple years, fire count spikes significantly post-stubble burning, leading to a notable rise in AQI. On average, fire count increases by 300-900%, while AQI rises by 58-130% in response. The AQI change per 1% fire count increase varies between 0.08% and 0.38%, showing a consistent but varying impact. This confirms that emissions from stubble burning in Punjab and Haryana contribute substantially to seasonal pollution spikes in Delhi.

The *Figure 4* time-series plots show how Gradient Boosting Regression, a machine learning method that improves predictive accuracy by combining multiple weak learners one after the other, can be used to predict the Air Quality Index (AQI). We designed our method to accurately capture the seasonal changes in air pollution dynamics. The reason for training the model on AQI data from March to September (before stubble burning) and testing it from October to February (after stubble burning) is based on the following: The time before stubble burning is a good way to see how much air pollution there is without the direct effect of stubble burning. The model can learn how pollution levels change over time due to things like traffic, industries, and the weather by using AQI data before the stubble-burning season. This is separate from stubble burning. Testing the model after stubble burning helps us see how AQI levels change because of stubble

http://www.aloki.hu ● ISSN 1589 1623 (Print) ● ISSN1785 0037 (Online)
DOI: http://dx.doi.org/10.15666/aeer/2304_79357956

burning and see how well the model works at predicting pollution levels during the most important months. This method lets us see the extra effect of stubble burning instead of having it mixed in with other seasonal patterns. If we only trained the model during the stubble-burning period, it might become too sensitive to that event and have trouble making long-term predictions after that time. To avoid overfitting and capture general trends, our method makes sure the model is generalized and strong. It does this by accurately capturing trends before the season starts and predicting how pollution levels change when stubble burning starts.

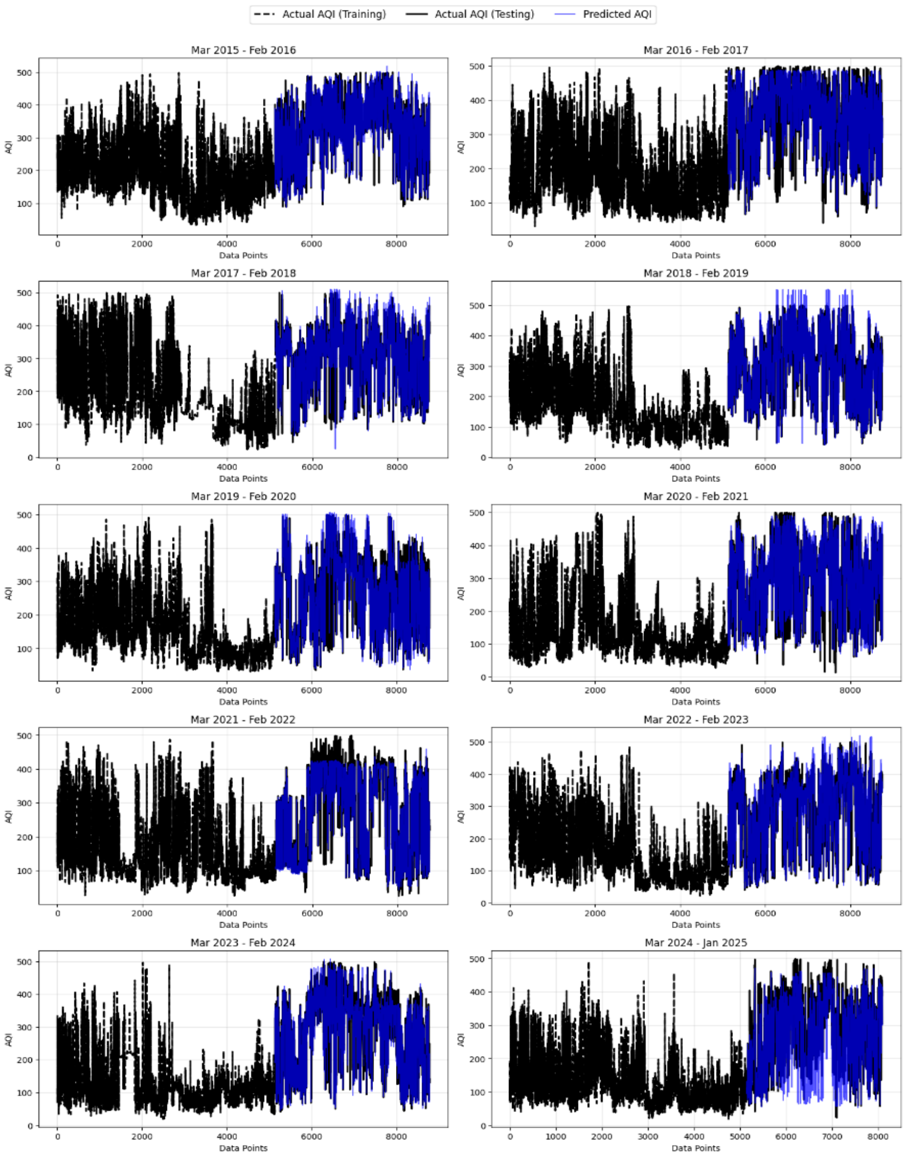

Figure 4. *Prediction of air quality index data over a period of years spanning from March 2019 – January 2025*

The model was trained on AQI data from March to September (before the stubble burning period) and tested on data from October to February (after the stubble burning period) to see how well it could predict pollution levels during the important stubble burning months. The prediction period did not coincide with the test period, as the model was intended to generate out-of-sample forecasts instead of merely confirming its accuracy on historical data. The predicted period was shorter than the test period because of Data Limitations and Model Stability. The plots show three important things for each year from 2015 to 2025: the actual AQI values used for training (black dashed lines), the actual AQI values used for testing (solid black lines), and the predicted AQI values (blue lines). This structured method makes it easy to compare predicted and actual AQI levels side by side, showing the model's strengths and weaknesses when it comes to tracking pollution trends. The testing period from October to February coincides with times when pollution is at its worst, and many things, like burning stubble, affect air quality. A clear seasonal pattern can be seen in all years, with AQI levels rising sharply after September, which is when stubble burning is at its peak in Punjab and Haryana. The model's ability to predict pollution spikes caused by stubble burning and changes in the weather is confirmed by how well it did at predicting pollution during this important time.

In the *Figure 4* the black line represents the actual AQI for the testing period show pronounced spikes, indicating severe air quality deterioration. While the blue line in *Figure 4* represent the predicted AQI lines follow a similar upward trajectory, they often underestimate extreme pollution spikes, suggesting that the model struggles to fully capture the impact of external pollution sources like stubble-burning emissions and meteorological variations. The years from 2015 to 2017 exhibit a closer match between actual and predicted AQI values, whereas from the year 2018 onward shows a significant divergence, especially during the peak pollution months of October to December. This discrepancy suggests that pollution events have become more severe and unpredictable in recent years, making it harder for traditional models to forecast AQI levels accurately. A noticeable trend is the increasing prediction error over time, particularly between the years 2019 and 2024, where the model consistently underestimates extreme pollution levels. This discrepancy arises due to several contributing factors, including rising stubble-burning emissions, worsening meteorological conditions such as air stagnation and temperature inversions, and increasing pollution sources such as vehicular traffic and industrial emissions. In contrast, during the period from 2020 to 2021, a temporary dip in AQI levels is observed, likely due to COVID-19 lockdowns that reduced emissions from transportation and industry.

Interestingly, in this period, the model predictions align more closely with actual AQI values, suggesting that reduced variability in pollution sources improves predictive accuracy. However, post-2022, the AQI spikes become more extreme, surpassing previous years, indicating a post-pandemic rebound in pollution levels, which the model fails to capture adequately. This reinforces the need for continuous model updates with real-time emission data and meteorological inputs to maintain forecasting reliability.

The increasing gap between actual and predicted AQI values highlights a crucial limitation: the model relies primarily on pre-stubble-burning AQI data without explicitly incorporating meteorological variables such as wind speed, humidity, temperature, and atmospheric pressure, which significantly influence pollution dispersion and accumulation. In years with stronger winds (period from the year 2015 to 2017), pollution events were less severe, and the model performed better, while in years with weaker winds and stagnant conditions (during the period from the year 2019 to 2024), AQI levels remained elevated for extended periods, leading to underestimation by the model. This underscores the necessity of integrating meteorological predictors and real-time emission sources into future models for improved accuracy. From a policy perspective, these

http://www.aloki.hu ● ISSN 1589 1623 (Print) ● ISSN1785 0037 (Online)
DOI: http://dx.doi.org/10.15666/aeer/2304_79357956

findings emphasize the urgent need for more sophisticated predictive approaches and stronger mitigation strategies. The growing divergence between actual and predicted AQI suggests that current pollution control measures, including restrictions on stubble burning, have been insufficient. Future models should incorporate real-time satellite data on fire counts, adaptive machine learning techniques that adjust to evolving pollution trends, and meteorological factors to enhance predictive power. Additionally, policy interventions such as stricter enforcement of anti-burning laws, incentivization of alternative residue management practices, and enhanced regional cooperation are crucial in reducing air pollution spikes. Given that pollution levels remain elevated for several months post-stubble burning, public health measures such as real-time AQI monitoring, early warning systems, and improved urban air filtration strategies should be prioritized to mitigate long-term health risks.

Table 1 gives the detailed year-wise analysis reveals fluctuations in model accuracy, with a general trend of increasing errors in recent years. The 2015–2016 period shows the best predictive performance, with low MAE (10.43), RMSE (18.49), and MSE (342.03), along with a high R^2 (0.96) and correlation (0.98), indicating strong agreement between predicted and actual AQI values. This suggests that during the initial years, pollution patterns were relatively stable, making them easier to predict. However, in the 2016–2017 period, a sharp decline in model accuracy is observed, with MAE more than doubling to 21.53, RMSE surging to 57.05, and MSE spiking to 3255.1, while R^2 drops significantly to 0.66. This suggests the emergence of more severe and unpredictable pollution events, likely due to increasing stubble-burning emissions and changing meteorological conditions, making AQI forecasting more challenging. Between 2017 and 2021, the model performance stabilizes, with MAE ranging between 9.76 and 14.47, RMSE between 24.92 and 31.71, and consistently high R^2 values (above 0.90 except for 2019–2020, where it drops to 0.83). The 2019–2020 period stands out with increased prediction errors (MAE: 26.68, RMSE: 46.88, MSE: 2197.85, and R^2: 0.83), suggesting a significant pollution anomaly, possibly influenced by meteorological factors like stagnant air and lower wind speeds, reducing pollutant dispersion. The 2020–2021 period, despite the COVID-19 pandemic reducing emissions, still shows substantial prediction errors (MSE: 1005.48, R^2: 0.91), implying that even during lockdown, pollution spikes remained unpredictable. From 2022 onwards, the model's predictive accuracy starts deteriorating again, with MAE increasing to 20.1 (2022–2023) and 14.58 (2023–2024), while RMSE and MSE remain high. However, the most alarming trend is seen in 2024–2025, where MAE jumps drastically to 35.07, RMSE reaches 64.66, and MSE skyrockets to 4181.35, with R^2 dropping to 0.6—one of the lowest values recorded. This signifies that the model struggles significantly to capture recent AQI trends, likely due to an unprecedented increase in pollution levels, extreme meteorological variations, and possibly unaccounted-for emission sources.

The correlation coefficient remains consistently high across all years (above 0.83), indicating a strong linear relationship between actual and predicted AQI values. However, the sharp decline in R^2 and the rising error metrics in 2024–2025 suggest that the model is increasingly failing to capture pollution variability, emphasizing the need for a more adaptive approach. The worsening prediction accuracy in later years suggests that traditional models trained on past data are struggling to account for evolving pollution patterns, necessitating real-time updates, inclusion of meteorological variables, and possibly deep learning-based techniques for improved forecasting. In conclusion, the model performed well in earlier years but has progressively struggled to predict extreme

http://www.aloki.hu ● ISSN 1589 1623 (Print) ● ISSN1785 0037 (Online)
DOI: http://dx.doi.org/10.15666/aeer/2304_79357956

pollution events, especially in recent times. This reinforces the growing severity and unpredictability of pollution in Delhi, particularly due to stubble burning and worsening atmospheric conditions. Future improvements should focus on integrating meteorological factors, updating training datasets frequently, and employing more robust predictive techniques such as ensemble learning or hybrid deep learning models to ensure better forecasting accuracy.

Table 1. *Gradient boosting regression performance metrics*

Period	MAE	RMSE	MSE	R^2	Correlation
Mar 2015 - Feb 2016	10.43	18.49	342.03	0.96	0.98
Mar 2016 - Feb 2017	21.53	57.05	3255.1	0.66	0.83
Mar 2017 - Feb 2018	10.75	25.66	658.38	0.93	0.96
Mar 2018 - Feb 2019	14.09	24.92	620.81	0.95	0.98
Mar 2019 - Feb 2020	26.68	46.88	2197.85	0.83	0.92
Mar 2020 - Feb 2021	14.47	31.71	1005.48	0.91	0.96
Mar 2021 - Feb 2022	9.76	25.7	660.43	0.96	0.98
Mar 2022 - Feb 2023	20.1	34.16	1166.61	0.91	0.95
Mar 2023 - Feb 2024	14.58	29.81	888.56	0.93	0.96
Mar 2024 - Jan 2025	35.07	64.66	4181.35	0.6	0.86

Conclusions

With Delhi being among the most impacted cities, air pollution has become one of the most serious environmental issues in India. Particularly in the post-monsoon months, when a significant rise in pollutant levels drives the Air Quality Index (AQI) into dangerous ranges, the city suffers concerning degrees of air pollution. Among the various causes, stubble burning in the surrounding states of Punjab and Haryana significantly aggravates Delhi's air quality problem.

This project examines the impact of stubble burning on air pollution in Delhi using a Gradient Boosting Regression model. The approach begins by clustering air particle data, such as PM2.5, PM10, CO, SO_2 etc. and other pollutants, into 6 clusters using K-means clustering. The clustering revealed that certain clusters with high pollution levels appeared predominantly during the post-stubble burning period, indicating a direct link between fire activity and elevated air pollution. The project further analyzed the relationship between fire count and AQI, showing a positive correlation: as fire count increases, AQI levels also rise. The Gradient Boosting model was trained on data from the pre-stubble burning period and tested on the post-stubble burning period, accurately predicting AQI fluctuations. This model helps predict how air quality will change in Delhi and helps create better plans for dealing with pollution. The results show that the AQI changes between 0.08% and 0.38% for every 1% increase in fire count. This project can predict changes in Delhi's air quality based on fire count data, which can help people get early warnings and take steps to reduce pollution. It also talks about how machine learning can help with environmental studies and how it can help make decisions about policies that will improve public health when pollution levels are high. Government policies, farmer education programs, subsidies for machinery that helps with managing residue, and stricter enforcement of anti-burning laws are all reasons why stubble burning has gone down. It could include incorporating more granular, real-time data for better accuracy, along with refining the model to consider additional factors like traffic and industrial emissions.

http://www.aloki.hu ● ISSN 1589 1623 (Print) ● ISSN1785 0037 (Online)
DOI: http://dx.doi.org/10.15666/aeer/2304_79357956

REFERENCES

[1] ARAI-TERI (2018): Source Apportionment of $PM_{2.5}$ and PM_{10} of Delhi-NCR for Identification of Major Sources. – Joint study by Automotive Research Association of India (ARAI) & The Energy and Resources Institute (TERI), p. 560.

[2] Beig, G., Srinivas, R., Parkhi, N. S., Carmichael, G. R., Singh, S., Sahu, S. K., Maji, S. (2019): An anatomy of the winter 2017 air quality emergency in Delhi. – Science of the Total Environment 681: 305-311.

[3] Beig, G., Sahu, S. K., Singh, V., Tikle, S., Sobhana, S. B., Gargeva, P., Murthy, B. S. (2020): Objective evaluation of stubble emission of North India and quantifying its impact on air quality of Delhi. – Science of the Total Environment 709: 136126.

[4] Bhadauriya, S., Chaudhary, N., Mamatha, S., Ray, S. S. (2020): Relationship between rice residue burning and increasing air pollution in North-west India. – The International Archives of the Photogrammetry Remote Sensing and Spatial Information Sciences XLIII-B3-2020: 1423-1430.

[5] Bisht, D. S., Dumka, U. C., Kaskaoutis, D. G., Pipal, A. S., Srivastava, A. K., Soni, V. K., Attri, S. D., Sateesh, M., Tiwari, S. (2015): Carbonaceous aerosols and pollutants over Delhi urban environment: Temporal evolution, source apportionment, and radiative forcing. – Science of the Total Environment 521-522: 431-445.

[6] Chauhan, A., Singh, R. P. (2017): Poor air quality and dense haze/smog during 2016 in the Indo-Gangetic plains associated with the crop residue burning and Diwali festival. – Conference: IGARSS 2017 - 2017 IEEE International Geoscience and Remote Sensing Symposium.

[7] Chowdhury, Z., Zheng, M., Schauer, J. J., Sheesley, R. J., Salmon, L. G., Cass, G. R., Russell, A. G. (2007): Speciation of ambient fine organic carbon particles and source apportionment of PM2.5 in Indian cities. – Journal of Geophysical Research: Atmospheres 112(D15).

[8] Daga, M. K., Mawari, G., Bharali, D., Banker, H., Mehar, P., Saluja, P., Kumar, D., Kumar, D., Jha, M. K., Gargava, P. (2019): Assessment of the air quality and its impact on health and environment in India. – Indian Journal of Medical Specialities 10(3): 117.

[9] Dey, S., DiGirolamo, L., van Donkelaar, A., Tripathi, S. N., Gupta, T., Mohan, M. (2012): Variability of outdoor fine particulate (PM2.5) concentration in the Indian Subcontinent: A remote sensing approach. – Remote Sensing of Environment 127: 153-161.

[10] Ganguly, N. D., Tzanis, C. G., Philippopoulos, K., Deligiorgi, D. (2019): Analysis of a severe air pollution episode in India during Diwali festival-a nationwide approach. – Atmósfera 32(3).

[11] Ghei, D., Sane, R. (2018): Estimates of air pollution in Delhi from the burning of firecrackers during the festival of Diwali. – PLoS One 13(8): e0200371.

[12] Ghude, S. D., Kumar, R., Jena, C., Debnath, S., Kulkarni, R. G., Alessandrini, S., Rajeevan, M. (2020): Evaluation of PM2.5 forecast using chemical data assimilation in the WRF-Chem model: A novel initiative under the Ministry of Earth Sciences Air Quality Early Warning System for Delhi, India. – Current Science 118(11): 1803-1815.

[13] Govardhan, G., Ambulkar, R., Kulkarni, S., Vishnoi, A., Yadav, P., Choudhury, B. A., Khare, M., Ghude, S. D. (2023a): Stubble-burning activities in north-western India in 2021: Contribution to air pollution in Delhi. – Heliyon 9(6): e16939.

[14] Govardhan, G., Ghude, S. D., Kumar, R., Sharma, S., Gunwani, P., Jena, C., Yadav, P., Ingle, S., Debnath, S., Pawar, P., Rajeevan, M. (2023b): Decision Support System version 1.0 (DSS v1.0) for air quality management in Delhi, India. – Geoscientific Model Development 17(7): 2617-2640.

[15] Jethva, H. T., Chand, D., Torres, O., Gupta, P., Lyapustin, A., Patadia, F. (2018): Agricultural burning and air quality over northern India: A synergistic analysis using NASA's A-train satellite data and ground measurements. – Aerosol and Air Quality Research 18(7).

http://www.aloki.hu ● ISSN 1589 1623 (Print) ● ISSN1785 0037 (Online)
DOI: http://dx.doi.org/10.15666/aeer/2304_79357956

[16] Kanawade, V. P., Srivastava, A. K., Ram, K., Asmi, E., Vakkari, V., Soni, V. K., Varaprasad, V., Sarangi, C. (2020): What caused severe air pollution episode of November 2016 in New Delhi? – Atmospheric Environment 222: 117125.

[17] Khan, A. A., Garsa, K., Jindal, P., Devara, P. C. S. (2023): Effects of stubble burning and firecrackers on the air quality of Delhi. – Environmental Monitoring and Assessment 195: 1170

[18] Krishna, K. R., Beig, G. (2018): Influence of meteorology on Particulate Matter (PM) and vice versa over two Indian metropolitan cities. – Open Journal of Air Pollution 7(3).

[19] Kulkarni, S. H., Ghude, S. D., Jena, C., Karumuri, R. K., Sinha, B., Sinha, V., Khare, M. (2020): How much does large-scale crop residue burning affect the air quality in Delhi? – Environmental Science and Technology 54(8): 4790-4799.

[20] Lan, R., Eastham, S. D., Liu, T., Norford, L. K., Barrett, S. R. H. (2022): Air quality impacts of crop residue burning in India and mitigation alternatives. – Nature Communications 13: 6537.

[21] Liu, T., Marlier, M. E., DeFries, R. S., Westervelt, D. M., Xia, K. R., Fiore, A. M., Milly, G. (2018): Seasonal impact of regional outdoor biomass burning on air pollution in three Indian cities: Delhi, Bengaluru, and Pune. – Atmospheric Environment 172: 83-92.

[22] Liu, Y., Wen, L., Lin, Z., Xu, C., Chen, Y., Li, Y. (2024): Air quality historical correlation model based on time series. – Scientific Reports 14: 22791.

[23] Mukherjee, T., Asutosh, A., Pandey, S. K., Yang, L., Gogoi, P. P., Panwar, A., Vinoj, V. (2018): Increasing potential for air pollution over megacity New Delhi: A study based on 2016 Diwali episode. – Aerosol and Air Quality Research 18(9).

[24] Nagar, P. K., Sharma, M. (2022): A hybrid model to improve WRF-Chem performance for crop burning emissions of PM2.5 and secondary aerosols in North India. – Urban Climate 41: 101084.

[25] Nair, M., Bherwani, H., Kumar, S., Gulia, S., Goyal, S., Kumar, R. (2020): Assessment of contribution of agricultural residue burning on air quality of Delhi using remote sensing and modelling tools. – Atmospheric Environment 230: 117504.

[26] Natarajan, S. K., Shanmurthy, P., Arockiam, D., Balusamy, B., Selvarajan, S. (2024): Optimized machine learning model for air quality index prediction in major cities in India. – Scientific Reports 14: 6795

[27] Rahman, M. H., Singh, N., Kundu, S., Datta, A. (2022): Potential areas of crop residue burning contributing to hazardous air pollution in Delhi during the post-monsoon season. – Journal of Environmental Quality 51(2): 181-192.

[28] Riches, N. O., Gouripeddi, R., Payan-Medina, A., Facelli, J. C. (2022): K-means cluster analysis of cooperative effects of CO, NO_2, O_3, $PM_{2.5}$, PM_{10}, and SO_2 on incidence of type 2 diabetes mellitus in the US. – Environ Res. 212: 113259.

[29] Sawlani, R., Agnihotri, R., Sharma, C., Patra, P. K., Dimri, A. P., Ram, K., Verma, R. L. (2019): The severe Delhi SMOG of 2016: A case of delayed crop residue burning, coincident firecracker emissions, and atypical meteorology. – Atmospheric Pollution Research 10(3): 868-879.

[30] Sembhi, H., Wooster, M., Zhang, T., Sharma, S., Singh, N., Agarwal, S., Khaiwal, R. (2020): Post-monsoon air quality degradation across Northern India: Assessing the impact of policy-related shifts in timing and amount of crop residue burnt. – Environmental Research Letters 15(10): 104067.

[31] Sharma, M., Dikshit, O. (2016): Comprehensive Study on Air Pollution and Greenhouse Gases (GHGs) in Delhi. – Final Report, prepared by IIT Kanpur, sponsored by Delhi Pollution Control Committee, New Delhi, p. 298.

[32] Sharma, N., Taneja, S., Sagar, V., Bhatt, A. (2018): Forecasting air pollution load in Delhi using data analysis tools. – Procedia Computer Science 132: 1077-1085.

[33] Singh, A. K., Srivastava, A. (2020): The impact of fireworks emissions on air quality in Delhi, India. – Environmental Claims Journal 32(4): 289-309.

http://www.aloki.hu ● ISSN 1589 1623 (Print) ● ISSN1785 0037 (Online)
DOI: http://dx.doi.org/10.15666/aeer/2304_79357956

[34] Singh, G., Kumar, A., Vaid, D., Sharma, P. (2021): Stubble burning and its impact on air quality in Delhi NCT: A case study. – Journal of University of Shanghai for Science and Technology 23(6): 58-68.
[35] Singh, R., Sinha, B., Hakkim, H., Sinha, V. (2023): Source apportionment of volatile organic compounds during paddy-residue burning season in north-west India reveals a large pool of photochemically formed air toxics. – Environmental Pollution 338: 122656.
[36] Sinha, S., Singh, S. (2020): Review on air pollution of Delhi zone using machine learning algorithm. – Journal of Air Pollution and Health 4(4): 259-272.
[37] Suganya, S., Meyyappan, T., Kumar, S. S. (2020): Performance analysis of K-means and K-medoids algorithms in air pollution prediction. – International Journal of Recent Technology and Engineering 8(5): 3573-3577.
[38] Suroshe, S., Dharpal, S. V., Ingole, N. W. (2022): Prediction of air quality index using regression models. – GIS Science Journal 9(8): 576-578.

http://www.aloki.hu ● ISSN 1589 1623 (Print) ● ISSN1785 0037 (Online)
DOI: http://dx.doi.org/10.15666/aeer/2304_79357956