Int. j. adv. multidisc. res. stud. 2024; 4(2):309-317

Received: 15-01-2024
Accepted: 05-03-2024

International Journal of Advanced Multidisciplinary Research and Studies

ISSN: 2583-049X

Air Quality Index Prediction in Punjab using Genetic Algorithm and Artificial Intelligence

[1] Dr. Rachhpal Singh, [2] Dr. Parvinder Kaur, [3] Amandeep Kaur
[1] AP, Khalsa College, Amritsar, Punjab, India
[2, 3] Assistant Professor, Guru Nanak College, Batala, Punjab, India

Corresponding Author: **Dr. Rachhpal Singh**

Abstract
Stubble burning is today's big problem in Punjab and its surrounding areas. This happens due to the huge quantity of stubble generated after harvesting crops. The requirement of the next crop sowing within a period creates the stubble burning problem that harms our ecosystem and ecology. Stubble burning increases the PM (Particulate Matter) which disturbs our AQI (Air Quality Index) in the environment. To control this, we can make some predictions using AI (Artificial Intelligence) techniques like ML (Machine Learning), DL (Deep Learning) etc. Many conventional algorithms were used for the prediction of AQI like SVM (Support Vector Machine), RF (Random Forest), ANN (Artificial Neural Network), and regression/classification. A hybrid approach of GA (Genetic Algorithm) with RF (Random Forest) as the proposed technique gave better prediction than traditional techniques. In the past five years, air pollution data analyzed from different cities in Punjab has helped estimate and forecast PM levels by using our proposed hybrid RG technique and a comparison with other conventional techniques.

Keywords: Artificial Intelligence, Machine Learning, Air Quality Index, $PM_{2.5}$, PM_{10}, Support Vector Machine, Artificial Neural Network, Random Forests Correlation and Regression Analysis

1. Introduction

Stubble burning is the agricultural field residue burning after harvesting rice or wheat crops [1]. Due to short time in harvesting and sowing of rice/wheat crops in their sessions, stubble burning is done by farmers [2]. This burning reduces nutrients in soil and disturbs our ecosystem. Burning smoke emitted with ammonia, nitrogen oxide and methane gases pollute the environment. These gases affect our lungs that exacerbate asthma which further make a risk in chronic bronchitis [3]. Burning of crop residue also create a pollution in Ozone layer. Modernization, vehicle emissions, population growth, deforestation and industrialization also are the pollution cause by releasing hazardous gases like sulphur dioxide (SO_2), lead (Pb), ozone (O_3), nitrogen dioxide (NO_2) and carbon monoxide (CO) that increases particulate matter $PM_{2.5}$, PM_{10} having some airborne particles in solid and liquid compositions [4]. These compositions have some organic particles, SO_4 and NO_3 molecules [5]. The particulate matter (PM) less than 2.5m dimension values is $PM_{2.5}$ which is most harmful for our environment [6]. $PM_{2.5}$ is hazardous for humanity that irritates lungs by damaging lungs [7]. $PM_{2.5}$ also develops skin cancer and lung cancer [8]. Tiny particles in lungs halts our respiratory system which further develops COVID-19 coronavirus infection [9]. An analysis and by doing research on particulate materials, it was found that it create a bad effect on human health [10]. In North region of India (Punjab, Haryana and Delhi) this stubble burning create a worst situation. During this period air pollution in Punjab increases to 39%, in Haryana it is 43% and in Delhi and its related NCR area it reaches to 48% according to government meteorologists which increases AQI in environment [11].

Governments in states have taken strict decisions and actions in past years to stop or reduce stubble burning. In 1981, IPC (Indian Penal Code) having Section 188 implemented for controlling stubble burning with unadorned punishment. Production of paddy straw in Punjab increases to 180 lakh tones every year [12]. Many industries started utilizing raw material of stubble for paper mills manufacturing, biogas plants utilization, power plants implementation, worm farms creation, packing the materials for different purposes, biomass power production, fuel boilers management etc. Punjab Pollution Control Board issued burnt data (in hectares) from 2019 to December 2023 in Table 1 below.

Table 1: Burning Report by PPCB

Years wise	2019	2020	2021	2022	2023
Stubble burn (in lakh hectares)	19.12	8.13	16.01	17.98	15.99

Table 2 is about the fire farm cases year wise in Punjab that was increasing year by year and making the situation worst.

Table 2: Stubble burning cases in Punjab (By PPCB)

Years wise	2019	2020	2021	2022	2023
Stubble Burning cases in Punjab	52,991	36,117	21,998	17,999	17,234

Sangrur district has highest 5,565 stubble burn cases. Ferozepur has 3,322, Moga has 2,596, Bathinda has 2,900 and at end Barnala has 2,266. PAU (Punjab Agricultural University, Ludhiana) shown 12,124 stubble burn cases till 30th November 2023. Punjab's Faridkot district has predicted highest worst air pollution. These cases effect the AQI level. Different district in Punjab showing AQI level in Table 3 (A day in November 2023).

Table 3: Different Districts of AQI levels

City	Sangrur	Bathinda	Moga	Ferozepur	Barnala
AQI	391	224	208	187	181

Burning crops residue made it of low fertile and also increase the AQI level [13]. A complete chart having various AQI levels with its AQI range and effects on health as shown in Table 4.

Table 4: Impact on health with various AQI levels

AQI Value	AQI Health effects	Meaning
0-50	Good	*Little Risk*
51-100	Moderate	*It create impact on those persons who have effect of poor air quality*
101-150	Unhealthy for Sensitive persons	*It create impact on sensitive children, unhealthy persons and old age person*
151-200	Unhealthy for all	*Everyone feel impact*
201-300	Very Dangerous	*Lungs and Heart problems occurs for all persons who passes through this level*

We looked for AQI-specific factors in the PPCB dataset in order to forecast air quality. The National Ambient Board developed six levels for AQI classification: First, acceptable (0 to 50), second, bearable (51 to 100), third, moderate (101 to 150), fourth, poor (151 to 200) and fifth, severe (201 to 300). Further, we can say that AI (Artificial Intelligence) forecasted AQI in better way. Certain prediction parameters used in auto learning methods work similarly to those used in pure statistics. Some of AI techniques as ANN (Artificial Neural Networks) gave AQI prediction in better form [14]. A hybrid approach by combining AI with GA offers a policy for deep and precise prediction. In order to obtain some correct initial values with some threshold values that apply to expedite training, the forecasting output from GA has been further optimized. The purpose of this work proposal is to forecast the concentration of air pollutants in Punjab so that hybrid approach-based preventive actions can be implemented to save human lives. (See Fig 1) Displays the entire AQI view, which is provided by website: https://www.iqair.com/in-en/india/punjab.

2. Literature Review

A complete study of related work is described in below table 5 (literature survey):

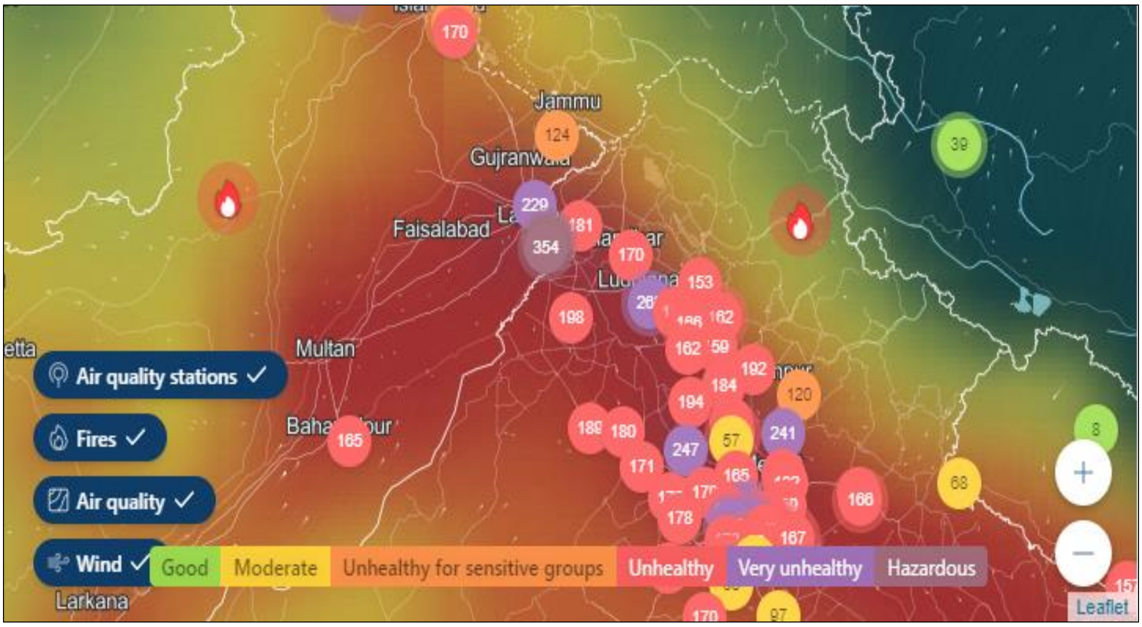

Fig 1: A full AQI view (From website iqair.com)

Table 5: Survey of Literature

Authors	Methodology used with results
Kaur *et al.* [15]	Stubble as a Renewable Source of Energy, A complete study of Stubble Burning and Crisis of Environmental Degradation in Punjab, India that shows the results having 80% accuracy than previous methods.
Kumar *et al.* [16]	Assistance of PEDA (Punjab Energy Development Agency), produce electricity from agricultural waste of paddy straw as biomass fuel that decreased global warming and the greenhouse effect and gave 80% accuracy than previous.
Bellinger *et al.* [17]	By employing data mining and machine learning approaches, the air pollution databases for China, the USA, and Europe and gave 83.2% accuracy than previous.
Rybarczyk and Zalakeviciute [18]	In order to achieve the highest level of accuracy in pollution control, a hybrid model with traffic density correlation using an air pollution database that gave 85.9% accuracy than previous.
Sharma *et al.* [19]	Data on Delhi City's air quality to track different levels of pollution that show better output as compared to previous databases.
Sweileh *et al.* [20]	Air pollution and health literature by looking through a number of Scopus articles from 1990 to 2017 that gave more accurate data.
Dua *et al.* [21]	Delhi region's air pollution data for forecasting analyzed and provided an online air pollution prediction system in real time.
Kumar and B. P. Pande [22]	Machine learning method to identify highly polluted Indian cities and predicts the pollution levels of SO2 in the environment of Maharashtra state that gave more accuracy as compared to traditional techniques.
Mahalingam *et al.* [23]	A support vector machine model of several Indian cities was used to predict the AQI with a high degree of accuracy.
Singh *et al.* [24]	By gathering data from the ITO station in Delhi, RPART and C5.0 supervised machine learning algorithms were used to classify and predict the model for atmospheric pollution.
Castelli *et al.* [25]	Created a machine learning-based predicted model to determine the air quality in California by taking into account various contaminants and particulate levels. Regression using Support Vectors.
Bamrah *et al.* [26]	Calculated the AQI with 81% accuracy by taking into account different levels of pollution concentration and utilizing a variety of regression techniques and machine learning.
Kumar *et al.* [27]	Predicted levels of PM2.5 concentration in the Delhi regions using regression and time series analysis, varying the parameters to determine the effectiveness of the suggested strategy.
Harishkumar *et al.* [28]	Using Taiwan's air pollution information, machine learning prediction methods were examined with respect to expected and actual values.
Liang *et al.* [29]	Evaluated Taiwan's AQI data for forecasting using a variety of machine learning classifiers.
Madan *et al.* [30]	Analyzed data on the amount of pollutants using machine learning, taking into account many parameters and the accuracy of the anticipated air pollution
Madhuri *et al.* [31]	Determined the concentrations of air contaminants using an RF machine learning technique.
Monisri *et al.* [32]	Created an air quality forecast model by combining data on air pollution from many sources.
Patil *et al.* [33]	Obtained machine learning AQI data for use in artificial neural networks, logistic regression, and forecasting.
Chhapariya *et al.* [34]	Found Punjabi locations for stubble burning databases using machine learning and fuzzy approaches to identify and determine the optimal classifier strategy.
Sanjeev [35]	Analyzed datasets of contaminants and used the Random Forest classifier to estimate the quality of the air.
Arif *et al.* [36]	Compiled the history of forest fires and used machine learning techniques to forecast their detection in burned areas.
Barthwal *et al.* [37]	Developed a model to predict PM concentrations at NCR locations, and assessed the VIR (variable importance ranking) to determine the mean, absolute, and root mean square errors.
Kaur *et al.* [38]	AQI was forecasted utilizing data visualizations, correlation, and statistical outliers with a machine learning approach, based on perceived air quality data of Indian cities.
Pardasani and Raghav [39]	Examined using multiple regression analysis the effects of stubble burning in Punjab on the NCR area.
Keil *et al.* [40]	Analyzed different burn data approaches and recommended that farmers use no-burn methods such as "Happy Seeder." Moreover, Happy Seeder was used to analyze cost management and identification of all contributing aspects.
Sangwan and Deswal [41]	Compared artificial intelligence techniques such as Support Vector Machine, Random Forest, and Artificial Neural Network on a range of pollution characteristics, taking into account the Rohtak area during stubble burning, in order to study PM2.5 modeling.
Pant *et al.* [42]	Concentrated on AQI prediction using supervised machine learning approaches in Dehradun, where pollutants included SO2, PM10, NO2, PM2.5, and so on. The prediction accuracy was determined to be 98.63%.
Aruna *et al.* [43]	Streamlined the tasks of collecting and disposing of stubble.

3. Methodology

Air pollution is currently posing a serious threat to agricultural areas and negatively affecting several Indian towns, especially those in the Punjab region. Elevated AQI is detrimental to health and impedes India's economic development. Major sources of pollution include increased industrial energy production, traffic on roads, power plants, open burning of waste, waste incineration, and so forth. The Punjab Pollution Control Board [44] provided the data needed for the research study, which covered the years 2018–2023. The current analysis is based on air pollution data [45] from the Punjab Pollution Control Board. This dataset has observations from January 2018-November 2023 and 10 attributes with 29,112 samples from five different cities in Punjab. Table 6 below provides a brief description of the pollutants/particles exhibiting AQI from the available data-set.

Table 6: Statistical data showing various AQI values compared with Pollutants

Statistic Values → / Pollutant Values ↓	Count	Mean	Standard Deviation	Minimum Value	Maximum Value
$PM_{2.5}$	23,888	66,889	62991	0.041	9988
PM_{10}	18,112	19,119	89999	0.011	1012
NO	24,891	17,399	23100	0.021	422
NO_2	24,211	27,999	24111	0.011	488
NH_3	18,898	21,998	25222	0.011	34989
CO	26,999	2190	6999	0.223	18799
SO_2	27,001	15200	19200	0.012	19111
O_3	25,102	34,299	23111	0.013	25100
Toluene	22,123	82145	19898	0.600	451.0
Benzene	23,200	3332	15112	0.122	423.00

Table 7 below illustrates an exact connection of AQI with each value from the provided dataset:

Table 7: Correlation or association between pollutants and AQI

Pollutants	**PM_{10}**	**$PM_{2.5}$**	**CO**	**NO_2**	**SO_2**	**NH_3**	**O_3**	**Xylene**	**Benzene**	**Toulene**
Correlation	0.8103	0.6523	0.6811	0.53111	0.52000	0.25121	0.20298	0.16698	0.04111	0.27999

It should be mentioned that skewed distribution is aided by a normal distribution of the data, which makes AI algorithms work better. Therefore, it is essential for identifying Skewness from current features that are derived from data sets. These were mapped, and vital changes were made based on this Skewness to turn the skewed distribution into a normal one [46]. (See Fig 2) Illustrates the highly skewed values of the properties of CO, Benzene, Xylene, and Toluene etc.

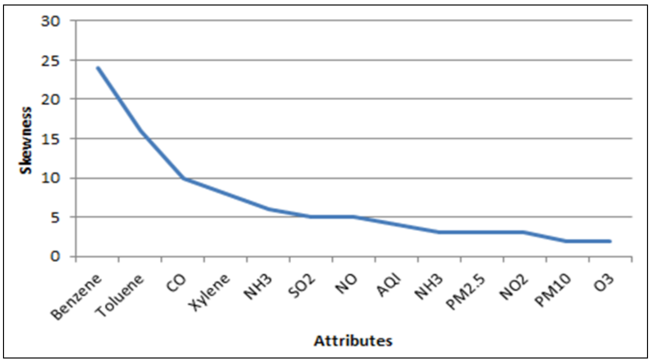

Fig 2: Various dataset attributes showing Skewness

A) Problem Definition:

The amount of fine particles in the air is an important component after precise measurement. Other factors that impact its level are wind direction, wind speed, and solar illumination, among others. The $PM_{2.5}$ particle type is significant for predicting pollution levels. $PM_{2.5}$ concentration levels were determined using a variety of methods, but it is a very challenging task to determine such levels precisely because of their variable dependence and time-dependent behavior on a number of additional factors in Punjab and its surrounding areas, such as vehicle's CO_2 emissions and burning of stubble. As a result, time-based regression was evaluated as the main task at a higher level, and it becomes useful for measuring continuous $PM_{2.5}$ prediction. It depends on historical datasets or $PM_{2.5}$ readings in addition to a variety of meteorological factors with time-series format records. The issue of accurate $PM_{2.5}$ measurements in Punjab and the surrounding areas is especially addressed by the proposed plan. AI systems are trained on historical data and run through GA to generate future predictions. Following the initial phase of data collection, pre-processing was carried out, followed by an application of the analysis process and a discussion of GA and AI (Feature selection –RF algorithm). It is shown in (See Fig 3).

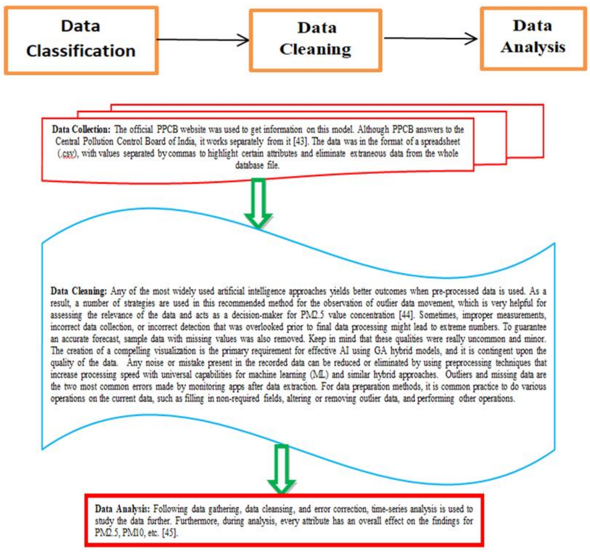

Fig 3: Data Processing

B) Data Analysis:

Using the provided data, the following analysis of varying PM concentrations in different seasons is discussed as follow:

a) Evaluation of Sessional PM concentrations:

Seasonal patterns influence PM levels in the Punjab region because of geographic and climatic factors. Three seasons were used to examine full-year pollutant concentrations in accordance with different risk ratings. 16th October to 31st January (winter), 01st February to 30th June (spring and summer) and 01st July 15 October (monsoons) are three sessions in Punjab climate. The daily average PM2.5 and PM10 concentrations from PAU and PPCB for the months of December 2022 to December 2023 are used to generate the forecast models. The PM time-series with the highest, lowest, mean, median, mode, standard deviation (SD), and range values are shown in Table 8.

Table 8: Statistics of PM concentration of winter session

PMs → Factors ↓	$PM_{2.5}$	PM_{10}
Highest	732.12	989.99
Lowest	5.1	13.11
Mean	121.99	245.12
Median	91.98	212.11
Mode	51.11	110.98
SD	101.99	150.01
Range	711.21	924.02

On May 15, 2023, the maximum daily average $PM_{2.5}$ and PM_{10} value for the spring and summer seasons are recorded. The PM time-series with the highest, lowest, mean, median, mode, standard deviation (SD), and range values are shown in Table 9.

Table 9: Statistics of PM concentration of spring and summer session

PMs → Factors ↓	$PM_{2.5}$	PM_{10}
Highest	311.98	892.01
Lowest	15.01	89.11
Mean	98.02	277.00
Median	90.12	249.11
Mode	40.99	99.00
SD	48.00	120.11
Range	235.11	899.00

Table 10 presents the basic statistics on PM variations during the monsoon season.

Table 10: Statistics of PM concentration of monsoon session

PMs → Factors ↓	$PM_{2.5}$	PM_{10}
Highest	191.11	472.01
Lowest	5.1	16.00
Mean	50.01	147.11
Median	40.02	113.00
Mode	41.00	96.99
SD	38.00	99.00
Range	185.11	450.01

b) PM values Vs Environmental Factors:

Meteorological variables such as Relative Humidity (RH), Rainfall (RF), Temperature (TEMP), Solar Radiation (SR), Wind Direction (WD), Wind Speed (WS), Atmospheric Pollutants, and Atmospheric Pressure (AP) such as Sulphur Dioxide (SO_2), Nitrogen Dioxide (NO_2), Carbon Monoxide (CO), and Ozone (O_3) are known to have an impact on a region's PM levels.

In order to more precisely estimate future PM concentrations, the prediction models in this work incorporate pollutant inputs in addition to meteorological data as predictor variables. The basic statistics of the 11 parameters that were inputted into the forecast model for this study are listed in Table 11.

Table 11: Statistics of predicted factors (From 01st January 2023 - 15th December 2023)

Environmental Factors	RF	RH	TEMP	WS	SR	WD	CO	AP	O_3	SO_2	NO_2
Max	949	97.99	43.00	9.2	312.1	277.1	7.1	750.00	451	121.00	122.0
Min	01	21.98	13.00	0.2	9.8	67.1	0.31	723.00	1.2	7.1	12.01
Mean	66.01	63.01	26.91	2.1	89	169	1.99	712.01	57	34.00	56.11
SD	156.02	14.99	7.01	0.91	45	45.01	0.71	6.99	69	13.00	23.01

Air pollutants such as $PM_{2.5}$, PM_{10}, CO, NO_2, O_3, SO_2, etc. were analyzed using a hybrid technique of AI and GA for improving AQI prediction. The entire methodological procedure combining AI and GA is depicted in (See Fig 4).

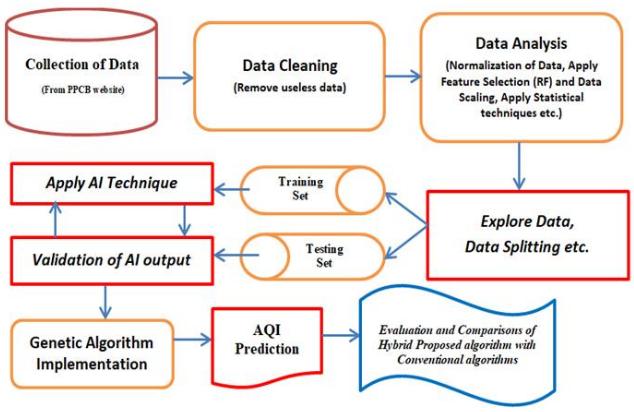

Fig 4: Proposed methodology

c) Apply RF Algorithm (Feature selection):

Research in this area showed that predictions can be made more accurate by removing some input variables and using less computer power. The best values for pollutants as input variables were found in the current study by comparing each pair of initial and final variables using a feature selection method that uses correlation. Be advised that lot of AI techniques are very susceptible to these anomalies. To choose the most important features, a correlation analysis was done between the value sets of the AQI features and those of other pollutants.

Random Forests (RFs) are found to be more suitable for prediction than ANN and SVM since they are fully non-parametric and do not require knowledge of the distribution of different input parameters. Furthermore, only Random Forests can accept both category and numerical inputs in order to reflect non-linear relationships between accessible classes and features that take missing data into account. Additionally, RFs are preferred because of their clear values and basic comprehensible regression structure. Using training data, a collaborative regression tree is created using the Random Forest technique and values are then selected for predicting. Every tree in the ensemble grows according to random vectors. Decision trees with regression and classification are used to handle continuous data values. Keep in mind that every decision tree seems to be a tree structure with an attribute test running on each node. Each branch transmits the test result and each terminal node contains a feature label. All initial input data is contained in the root node. Next, a variety of splitting factors are used to divide the data set into child nodes. The decision tree approach calculates entropy, predicted entropy and information gain in order to evaluate the splitter based on input variable values and to determine whether to split the inputted nodes further. The Random Forest algorithm is elaborated in (See Fig 5) below:

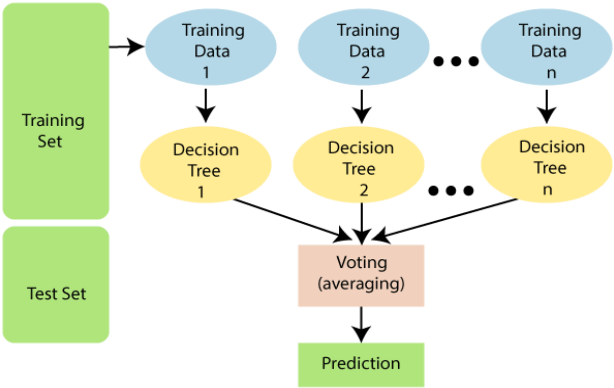

Fig 5: RF elaboration

The two stages of Random Forest's operation are the creation of the random forest from the combination of N decision trees and the prediction of each tree generated in the first phase. The stages below can be used to demonstrate the working process:

Step1: Choose K data points at random from the training set.

Step2: Construct the decision trees linked to the chosen data points (subsets).

Step3: When building decision trees, select the number N.

Step4: Repeat Step1 and Step2.

Step5: Locate each decision tree's predictions for the new data points and then allocate them to the category with the majority of votes.

d) Applying GA:

GA [47] is an enhanced technique that uses a mechanism based on population inheritance and natural selection to simulate hybridization, reproduction, and mutation. Inheritance and natural selection processes provide the basis for the application of bio-inspired operators for effective solutions to search and optimization problems. People with GA may benefit from a treatment that was "chromosome" encoded. Every possible person is a member of the population in the further solution domain. From this population, a fitness function with a feasible solution is calculated. Once each individual's fitness value has been assessed using the predetermined fitness function, individual selection is generated in the subsequent generation using the designated fitness function. The goal of selection is to preserve the strong and destroy the weak. The values of these selected individuals were subsequently developed into a new generation set of values by applying two additional processes, crossover and mutation. People from the younger generation succeed more than those from the older generation, who were gradually moving towards the greatest possible result. The younger generation inherits the good value sets from the older generation.

4. Experimental Findings and Talk

This paper discusses an experimental setup and empirical analysis for AQI value predicting based on airborne pollutants. The dataset including air pollution data is divided into two subsets before being evaluated using GA: A training subset with a weight of 75% and a testing subset with a weight of 25%. Python programs run on the Google Pro-cloud environment, which features a Tesla P50 processor with an Intel core clocked at 2.2GHz, 32GB RAM and 512GB SSD external memory. For data processing, some Python libraries are taken into consideration, such as Seaborn, NumPy, Scikit-learn and Pandas etc. Next, an analysis of the dataset is conducted to ascertain the overall AQI values with respect to those pollutants that have a substantial effect on raising the AQI value.

The methodology used to determine the accuracy and efficiency of AQI forecasting that explains how AI-based methodology is useful for evaluating AQI. AI frequently overlooks the problem of imbalanced datasets which could lead to poor classification and prediction performances. The classes are split unevenly because of some missing values in the target attribute and AQI Bucket. AQI level may be predicted using four well-known AI models—SVM, RF, ANN and proposed hybrid RG model. Tables 12 and table 13 depict the performance of used AI models in terms of recall, F1-score, accuracy and precision throughout the training phase and the testing phase. Precision is the percentage of relevant instances that are present in the retrieved instances, whereas recall is the percentage of relevant examples that have been recovered. Accuracy is defined as the ratio of correctly detected attributes to the whole set of variables. The F1-score is a weighted average of recall and precision. Note that the RG had the best accuracy and the SVM model had the lowest in training phase.

Table 12: Comparison of model results in the training set

Algorithms	Accuracy	Training time	Recall	F1 score	Precision
ANN	88	0.111	88	91	93
SVM	81	0.261	93	88	89
RF	93	0.532	95	91	97
RG	95	0.112	89	98	99

Recall indicates the percentage of pertinent examples that have been retrieved, whereas precision indicates the percentage of appropriate instances that are available in recovered instances. The ratio of accurately identified attributes to all other variables is known as accuracy. The precision and recall numbers are weighted averaged to determine the F1-score. Remember that the SVM model had the lowest accuracy throughout the testing phase, even though the RG model had the highest accuracy.

Table 13: Comparison of various methods outputs in testing set

Algorithms	Accuracy	Prediction time	Recall	F1 score	Precision
ANN	86	0.021	91	92	92
SVM	78	0.031	91	83	90
RF	91	0.040	97	94	98
RG	93	0.019	99	96	89

Eleven inputs (RF, RH, TEMP, SR, WS, WD, AP, CO, O_3, SO_2, and NO_2) are used as explanatory variables in the construction of the forecast models. The forecast models are used to produce short-term (7-day) forecasts. Using seasonal concentrations, the $PM_{2.5}$ and PM_{10} levels are forecasted for seven days during the same season. (See Fig 6, See Fig 7 and See Fig 8) Display the PM levels that are expected and actual for each season.

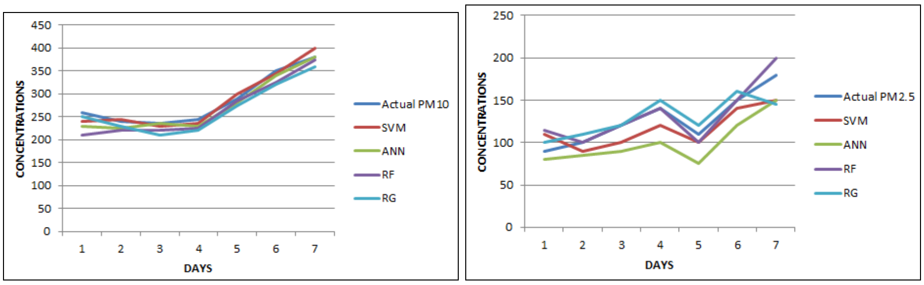

Fig 6: Forecast of PM_{10} and $PM_{2.5}$ Level (7 days Monsoon Data)

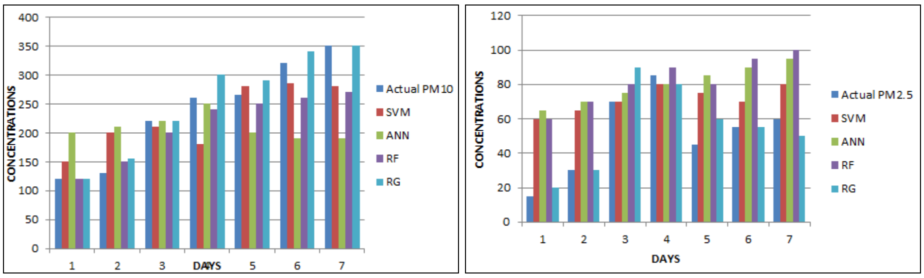

Fig 7: Forecast of PM_{10} and $PM_{2.5}$ Level (7 days Spring-Summer Data)

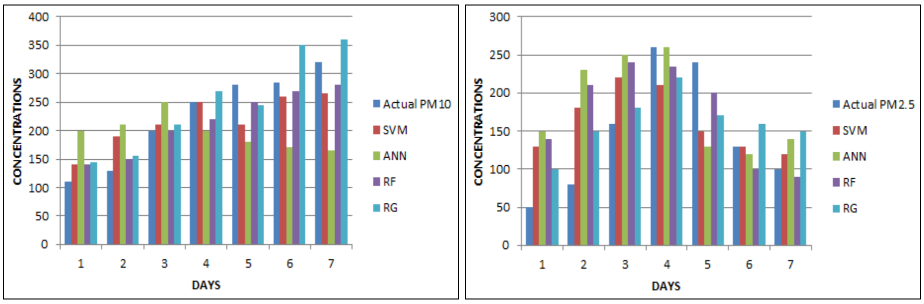

Fig 8: Forecast of PM_{10} and $PM_{2.5}$ Level (7 days Winter Data)

(See Fig 6) Displays the seven-day predictions by suggested method for $PM_{2.5}$ and PM_{10} concentrations using the Monsoon data-set as input. It has been noted that when it comes to predicting PM levels during the monsoon, RG outperforms other models. (See Fig 7) Illustrates the forecasting of PM levels for the summer-spring session. During the spring and summer sessions, RG and RF outperform other models in terms of $PM_{2.5}$ and PM_{10} levels. The seven-day forecast for PM levels during the winter sessions is displayed in (See Fig 8). Keep in mind that ANN is at its lowest level while RG and RF show the best results for all seasonal outcomes.

5. Conclusion

It is particularly difficult to anticipate air quality because of the unpredictable nature of contaminants, dynamic environment and changes in time and place. Because of the detrimental consequences that air pollution has on humans, animals, plants, climate, environment and historical sites, developing countries constantly monitor and analyzed the quality of the air. Researchers' attention has been largely focused on the Indian AQI forecast. Here, five years' worth of air pollution statistics for a few Punjabi cities was examined. Following the filling of all NAN dataset values, outliers are resolved and data values are normalized through first cleaning and then preprocessing. Using a correlation methodology combined with a feature selection method, pollutants that affect the AQI are filtered for additional study and skewed features are transformed. Results of AI techniques for train-test data subsets are examined together with metrics such as recall, accuracy, precision and F1-score. Using train-test sets, the proposed hybrid RG achieved the highest accuracy while SVM had the lowest accuracy. Additional work on forecasting PM_{10} and $PM_{2.5}$ concentrations using RG, SVM, ANN, and RF approaches was conducted. In future, further research can be done with deep learning with meta-heuristic techniques.

6. References

1. Chanana Ishita, Aparajita Sharma, Pradeep Kumar, Lokender Kumar, Sourabh Kulshreshtha, Sanjay Kumar, *et al*. Combustion and Stubble Burning: A Major Concern for the Environment and Human Health. Fire. 2023; 6(2):79. Doi: https://doi.org/10.3390/fire6020079
2. Abdurrahman Muhammad Isa, Sukalpaa Chaki, Gaurav Saini. Stubble burning: Effects on health & environment, regulations and management practices. Environmental Advances. 2020; 2. Doi: https://doi.org/10.1016/j.envadv.2020.100011
3. Singh Abhijeet, Ashok Singh Vishnoi, Anwar Hameed Banday, Pratyashee Bora, Puneeta Pandey. Influence of stubble burning on air quality of Northern India: A case study of Indo-Gangetic plains of India. Environmental Monitoring and Assessment. 2023; 195(4):487. Doi: https://doi.org/10.1007/s10661-023-11027-w
4. Singh Jabrinder, Naveen Singhal, Shailey Singhal, Madhu Sharma, Shilpi Agarwal, Shefali Arora.

Environmental implications of rice and wheat stubble burning in north-western states of India. In Advances in Health and Environment safety, pp. 47-55. Springer, Singapore, 2018. Doi: https://doi.org/10.1007/978-981-10-7122-5_6
5. Pöschl Ulrich. Atmospheric aerosols: Composition, transformation, climate and health effects, Angewandte Chemie International Edition. 2005; 44(46):7520-7540. Doi: https://doi.org/10.1002/anie.200501122
6. Kam Winnie, Kalam Cheung, Nancy Daher, Constantinos Sioutas. Particulate matter (PM) concentrations in underground and ground-level rail systems of the Los Angeles Metro. Atmospheric Environment. 2011; 45(8):1506-1516. Doi: https://doi.org/10.1016/j.atmosenv.2010.12.049
7. Shaughnessy William J, Mohan M Venigalla, David Trump. Health effects of ambient levels of respirable particulate matter (PM) on healthy, young-adult population. Atmospheric Environment. 2015; 123:102-111. Doi: https://doi.org/10.1016/j.atmosenv.2015.10.039
8. Feng Shaolong, Dan Gao, Fen Liao, Furong Zhou, Xinming Wang. The health effects of ambient PM2. 5 and potential mechanisms, Ecotoxicology and environmental safety. 2016; 128:67-74. Doi: https://doi.org/10.1016/j.ecoenv.2016.01.030
9. Shereen Muhammad Adnan, Suliman Khan, Abeer Kazmi, Nadia Bashir, Rabeea Siddique. COVID-19 infection: Emergence, transmission, and characteristics of human coronaviruses. Journal of Advanced Research. 2020; 24:91-98. Doi: https://doi.org/10.1016/j.jare.2020.03.005
10. Ali Muhammad Ubaid, Siyi Lin, Balal Yousaf, Qumber Abbas, Mehr Ahmed Mujtaba Munir, Audil Rashid, *et al*. Pollution characteristics, mechanism of toxicity and health effects of the ultrafine particles in the indoor environment: Current status and future perspectives. Critical Reviews in Environmental Science and Technology. 2022; 52(3):436-473. Doi: https://doi.org/10.1080/10643389.2020.1831359
11. Nagar Pavan K, Mukesh Sharma. A hybrid model to improve WRF-Chem performance for crop burning emissions of PM2. 5 and secondary aerosols in North India," Urban Climate. 2022; 41:101084. Doi: https://doi.org/10.1016/j.uclim.2022.101084
12. Kaur Hardeep, Manvendra Singh. An Assessment of Environmental Pollution and Policy Initiatives in Punjab, India: A Review, 2022. Doi: 10.13189/eer.2022.100401
13. Tiwari Rajesh Kumar, Tapan Kumar Dey. A Novel Approach for Analysis of Air Quality Index before and After Covid-19 Using Machine Learning. In Applications of Artificial Intelligence and Machine Learning, pp. 139-149. Springer, Singapore, 2022. Doi: https://doi.org/10.1007/978-981-19-4831-2_12
14. Jaiswal Harsh, Aman Singh, Akhilesh Chauhan, Dicksha Sharma. Air Quality Index Prediction using Machine Learning, 2022.
15. Kaur Hardeep, Manvendra Singh. Stubble as a Renewable Source of Energy: A Study of Stubble Burning and Crisis of Environmental Degradation in Punjab, India. In Renewable Energy Optimization, Planning and Control: Proceedings of ICRTE 2022, pp. 311-323. Singapore: Springer Nature Singapore, 2023. Doi: https://doi.org/10.1007/978-981-19-8963-6_28
16. Kumar Parmod. Energy Generation by Use of Crop Stubble in Punjab, In Climate Change Challenge (3C) and Social-Economic-Ecological Interface-Building, pp. 507-518. Springer, Cham, 2016. Doi. https://doi.org/10.1007/978-3-319-31014-5_30
17. Bellinger Colin, Mohomed Shazan Mohomed Jabbar, Osmar Zaïane, Alvaro Osornio-Vargas. A systematic review of data mining and machine learning for air pollution epidemiology. BMC Public Health. 2017; 17(1):1-19. Doi: https://doi.org/10.1186/s12889-017-4914-3
18. Zalakeviciute Rasa, Yves Rybarczyk, Jesús López-Villada, Maria Valeria Diaz Suarez. Quantifying decade-long effects of fuel and traffic regulations on urban ambient PM2. 5 pollution in a mid-size South American city. Atmospheric Pollution Research. 2018; 9(1):66-75. Doi: https://doi.org/10.1016/j.apr.2017.07.001
19. Sharma Nidhi, Shweta Taneja, Vaishali Sagar, Arshita Bhatt. Forecasting air pollution load in Delhi using data analysis tools. Procedia computer science. 2018; 132:1077-1085. Doi: https://doi.org/10.1016/j.procs.2018.05.023
20. Sweileh Waleed M, Samah W Al-Jabi, Sa'ed H Zyoud, Ansam F Sawalha. Outdoor air pollution and respiratory health: A bibliometric analysis of publications in peer-reviewed Journals (1900–2017). Multidisciplinary Respiratory Medicine. 2018; 13(1):1-12. doi: 10.1109/ESCI56872.2023.10099645
21. Dua Radhika Dua, Divyam Madaan Madaan, Prerana Mukherjee Mukherjee, Brejesh Lall Lall. Real time attention based bidirectional long short-term memory networks for air pollution forecasting, In 2019 IEEE fifth international conference on Big Data computing service and applications (BigDataService), pp. 151-158. IEEE, 2019. Doi: 10.1109/BigDataService.2019.00027
22. Kumar K, Pande BP. Air pollution prediction with machine learning: A case study of Indian cities, International Journal of Environmental Science and Technology, 2022, 1-16. Doi: https://doi.org/10.1007/s13762-022-04241-5
23. Mahalingam Usha, Kirthiga Elangovan, Himanshu Dobhal, Chocko Valliappa, Sindhu Shrestha, Giriprasad Kedam. A machine learning model for air quality prediction for smart cities. In 2019 International conference on wireless communications signal processing and networking (WiSPNET), pp. 452-457. IEEE, 2019. Doi: 10.1109/WiSPNET45539.2019.9032734
24. Singh Anish, Raja Kumar, Nitasha Hasteer. Comparative Analysis of Classification Models for Predicting Quality of Air. In 2020 IEEE 5th International Conference on Computing Communication and Automation (ICCCA), pp. 7-11. IEEE, 2020. Doi: 10.1109/ICCCA49541.2020.9250805
25. Castelli Mauro, Fabiana Martins Clemente, Aleš Popovič, Sara Silva, Leonardo Vanneschi. A machine learning approach to predict air quality in California, Complexity, 2020. Doi: https://doi.org/10.1155/2020/8049504
26. Bamrah Sumneet Kaur, Saiharshith KR, Gayathri KS. Application of Random Forests for Air quality estimation in India by adopting terrain features. In 2020

4th International Conference on Computer, Communication and Signal Processing (ICCCSP), pp. 1-6. IEEE, 2020. Doi: 10.1109/ICCCSP49186.2020.9315252

27. Kumar Saurabh, Shweta Mishra, Sunil Kumar Singh. A machine learning-based model to estimate PM2. 5 concentration levels in Delhi's atmosphere, Heliyon. 2020; 6(11). Doi: https://doi.org/10.1016/j.heliyon.2020.e05618
28. Harishkumar KS, Yogesh KM, Ibrahim Gad. Forecasting air pollution particulate matter (PM2. 5) using machine learning regression models, Procedia Computer Science. 2020; 171:2057-2066. Doi: https://doi.org/10.1016/j.procs.2020.04.221
29. Liang Yun-Chia, Yona Maimury, Angela Hsiang-Ling Chen, Josue Rodolfo Cuevas Juarez. Machine learning-based prediction of air quality. Applied Sciences. 2020; 10(24):9151. Doi: https://doi.org/10.3390/app10249151
30. Madan Tanisha, Shrddha Sagar, Deepali Virmani. Air quality prediction using machine learning algorithms–a review. In 2020 2nd International Conference on Advances in Computing, Communication Control and Networking (ICACCCN), pp. 140-145. IEEE, 2020. Doi: 10.1109/ICACCCN51052.2020.9362912
31. Madhuri R, Sistla S, Srinivasa Raju K. Application of machine learning algorithms for flood susceptibility assessment and risk management. Journal of Water and Climate Change. 2021; 12(6):2608-2623. Doi: https://doi.org/10.2166/wcc.2021.051
32. Monisri PR, Vikas RK, Rohit NK, Varma MC, Chaithanya BN. Prediction and analysis of air quality using machine learning. Int J Adv Sci Technol. 2020; 29(5):6934-6943.
33. Patil Miss Priti Ashok, Ms Ashwini Salunkhe. Comparative analysis of construction cost estimation using artificial neural networks. J Xidian Univ. 2020; 14:1287-305. Doi: https://doi.org/10.37896/jxu14.7/146
34. Chhapariya Koushikey, Anil Kumar, Priyadarshi Upadhyay. A fuzzy machine learning approach for identification of paddy stubble burnt fields. Spatial Information Research. 2021; 29(3):319-329. Doi: https://doi.org/10.1007/s41324-020-00356-4
35. Sanjeev Dyuthi. Implementation of machine learning algorithms for analysis and prediction of air quality. International Journal of Engineering Research & Technology (IJERT). 2021; 10(3):533-538. ISSN: 2278-0181
36. Arif M, Alghamdi KK, Sahel SA, Alosaimi SO, Alsahaft ME, Alharthi MA, *et al*. Role of machine learning algorithms in forest fire management: A literature review. J Robotics Autom. 2021; 5(1):212-226. Doi: 10.36959/673/372
37. Barthwal Anurag, Debopam Acharya, Divya Lohani. Prediction and analysis of particulate matter (PM2. 5 and PM10) concentrations using machine learning techniques. Journal of Ambient Intelligence and Humanized Computing, 2021, 1-16. Doi: https://doi.org/10.1007/s12652-021-03051-w
38. Kaur Rajveer, Puneeta Pandey. Air pollution, climate change, and human health in Indian cities: A brief review," Frontiers in Sustainable Cities. 2021; 3:705131. Doi: https://doi.org/10.3389/frsc.2021.705131
39. Pardasani Raghav. Rethinking Rice Cultivation: A Multiple Regression Analysis of Factors Influencing the Prevalence of Stubble Burning in Punjab, 2021.
40. Keil Alwin, Krishnapriya PP, Archisman Mitra, Mangi L Jat, Harminder S Sidhu, Vijesh V Krishna, *et al*. Changing agricultural stubble burning practices in the Indo-Gangetic plains: Is the Happy Seeder a profitable alternative? International Journal of Agricultural Sustainability. 2021; 19(2):128-151. Doi: https://doi.org/10.1080/14735903.2020.1834277
41. Sangwan V, Deswal S. PM 2.5 modelling during paddy stubble burning months using artificial intelligence techniques. Journal of Achievements in Materials and Manufacturing Engineering. 2022; 110(1). Doi: 10.5604/01.3001.0015.7027
42. Pant Alka, Sanjay Sharma, Mamta Bansal, Mandeep Narang. Comparative Analysis of Supervised Machine Learning Techniques for AQI Prediction. In 2022 International Conference on Advanced Computing Technologies and Applications (ICACTA), pp. 1-4. IEEE, 2022. Doi: 10.1109/ICACTA54488.2022.9753636
43. Aruna Ch, Sai Charitha B, Mahitha G, Bhuvana Ch, Bhargavi D, Lakshmi Mounika D. Agri-Stubble Aggregation and Disposal. 2022; 5(6). Doi: 10.15680/IJMRSET.2022.0506011
44. Bhawan Parivesh, East Arjun Nagar. Central Pollution Control Board, 2020.
45. Kumar Rajesh, Suresh K Sharma, Thakur JS, Lakshmi PVM, Sharma MK, Singh T. Association of air pollution and mortality in the Ludhiana city of India: A time-series study. Indian Journal of Public Health. 2010; 54(2):98. Doi: 10.4103/0019-557X.73278
46. Gupta Shreekant, Shalini Saksena, Omer F Baris. Environmental enforcement and compliance in developing countries: Evidence from India. World Development. 2019; 117:313-327. Doi: https://doi.org/10.1016/j.worlddev.2019.02.001
47. Mirjalili Seyedali. Genetic algorithm. In Evolutionary algorithms and neural networks, pp. 43-55. Springer, Cham, 2019.