

Predictive Analysis of Smog Exposure and Its Impact on Human Health Outcomes

Zonesha Pervaiz¹, Rehman Sharif^{1*}, Ayesha Khalid¹, and Muhammad Rashad¹

¹Department of Mathematics & Physics The Superior University Lahore, 54000, Pakistan.

*Corresponding Author: Rehman Sharif. Email: rehman.sharif@superior.edu.pk

Received: July 06, 2025 Accepted: August 22, 2025

Abstract: The paper is an analysis of the anticipation of human health effect that results due to exposure to smog in Lahore, Pakistan. Smog has become a constant environmental hazard due to industrial emission, motor pollution and high rates of urbanization process. The given quantitative study is a combination of two important datasets, such as a five-year air quality dataset (2020-2024) and a survey of the population about the knowledge of smog and its influence on health. The key pollutants present in the air quality data include PM 2.5, PM 10, NO₂, NH₃, SO₂, CO, and O₃, which was retrieved in a Kaggle repository called Pakistani Cities AQI (2020-2024). The data was preprocessed by cleaning, standardizing and aligning the data with the answers of the survey. According to the analysis of the exploratory data (EDA) performed with the help of Python tools: Pandas, NumPy, Matplotlib and Seaborn, PM 2.5 and PM 10 were strongly correlated ($r = 0.84$), which suggests that they have shared sources based mostly on vehicular and industrial activities. The negative reaction in ozone was a sign of the following photochemical reactions. Seasonal analysis showed that due to temperature inversion and stagnant air the intensity of smog was more intense during winter (season) between November and January. Health consequences were predicted using machine learning methods such as Random Forest, XGBoost and Logistic Regression. Random Forest and XGBoost had a higher predicted accuracy and could determine the significance of such factors as income, education, the level of particulate matter, and residential nearness to busy territories. The models had been identified to be accurate, F1-score, and ROC-AUC measures. The results in Lahore show that particles pollutants, in particular, PM 2.5 and PM 10 are the main cause of respiratory and other health problems. The evidence-based approach to the advancement of health of the population and management of the air quality is supported by the predictive framework of the study which proposes the provision of a data-driven approach to forecast the health risks associated with the smog.

Keywords: Smog Pollution; Air Quality Analysis; Human Health Effects; ML Prediction; Environmental Hazards; Lahore Pakistan

1. Introduction

One of them is the problem of air quality in large cities in India, which is also worth having a comprehensive study not necessarily because of the scholarly interest but also because of the realization that will ultimately cause the establishment of decisions that will be adopted by the masses. This notebook will take into account air quality data in several cities in Pakistan and comment on the data and summarize the findings and even come up with a predictor to classify the high levels of PM_{2.5} pollution. In case this analysis is useful to you, then you may up vote it. We are going to use the data of several cities between 2020 and 2024. Smog is a form of air pollution which is usually experienced as smoke and mist mixed up. The term smog was first used in London in the first half of 1950s. Severe pollutants of air floating on humid

air are known as smog. The primary causes of smog in the atmosphere are climate and human activity. Coal burning in the industry, automobile emissions, burning of crops, construction, fireworks and the smoke produced by the brick kilns are the major contributors of the basic particles that cause smog. The world automobiles and firms smoke are a significant source of smoke production, which has led to critical smog pollution [1].

Traditionally, smog has been categorized into two major groupings depending on the form: Acid or Classical smog (London type) and Photochemical smog (Los Angeles Page 3/29 type) each of which are life threatening to human beings and the environment. Outside the two, there is another form of smog that is currently being discovered, and that is Polish smog[2]. About 4 billion individuals, 92 percent of the Asia-Pacific population is exposed to air pollution that is detrimental to their wellbeing. Air pollution is a severe health risk in addition to threatening human health, the environment, economic growth, and productivity of agricultural crops. In 2015, at least 35% of all the deaths in the world were in East Asia and the Pacific [3].

Effects of pollution are influenced by the period of exposure to it as well. There is much more danger in long term exposure as compared to the short-term exposure. Short-term peaks and long-term continuous exposures will follow different curves, and their effects are different. Some of the consequences include long-term genetic changes to temporary tracheal irritation. However, there are recent studies that have resulted in the relation between exposure to short-term pollution and undesirable deaths. These adverse health effects are broad, including long-term genetic defects and premature death as well as subclinical outcomes such as the irritation of the trachea. Some of the significant diseases that are transported by smog include the respiratory (asthma, coughing, and bronchiolitis), cardiovascular, neurological and cancer, low birth weight and other complications such as eye irritation and breathing problems[4].

There is a recent increase in smog and ambient air pollution as a serious health risk in South Asia. According to the State of Global Air Report 2024, approximately 1.2 million deaths in South Asia were due to exposure to fine particulate matter (PM_{2.5}) in 2021, with some 103,000 of them in Pakistan and 947,600 in India [5]. The primary goal of air quality prediction is to predict the concentration of the pollutants some time in the future based on the past data concerning the air quality, i.e., using the past air quality data sets, meteorological data sets, etc., as demonstrated by the work proposed in and By analysing the results of the previous studies, we found that the current methods apply neural networks, i.e., the LSTM in, the machine learning based solution in, the Extreme Learning Machine (ELM) in, or simple regression methods, to make predictions based on the past data[6].

Due to the high population growth and economical development, there is a constant degradation of the environment. The primary causes of environmental health problems are increase in energy consumption, industrialization and agricultural intensification. The environmental health hazards of the developing nations are largely connected with the prevalent poverty, severe absence of the facilities of the population, such as the water access, sanitary facilities, and healthcare services, and some new troubles connected with industrial pollution. However, developing countries are not the only ones facing environmental health hazards [7]. Air pollution has a devastating impact on climatic conditions, ecosystem and human health. Incidents of respiratory and cardiovascular complications in children and the elderly due to the inhalation of contaminated air have been noted especially in the vulnerable groups such as children and the elderly. Prolonged exposure is related to asthma and other chronic diseases such as lung cancer. The environmental impacts are agricultural damages, reduction of biodiversity, and acid rain [8]. The present study aims at the evaluation of the general population, and their level of awareness of smog pollution in Lahore, Punjab, Pakistan, and how it can be controlled through the Contingent Valuation Method (CVM). The research also offers measures of focusing the impact of smog through raising awareness of people, utilizing social media, and collaboration with other agencies [9].

Poor air quality is a curse to human beings and it affects the vegetation and wildlife adversely. The air quality list can vary from one nation to another based on the accumulating evidence of harmful impacts on wellness. WHO has set guidelines of air quality. Those who are subjected to air pollution are prone to asthma, persistent obstructive pulmonary disease, lung cellular breakdown, and others. Air pollution has been established to affect the weak populations including children and elderly people who have a history of heart and lung diseases, significantly. This is leading to increased mortality and premature deaths every

year. Thus, to avoid diseases and preserve human health, states can substantially decrease air pollution by following the air quality index presented by WHO [10].

2. Literature Review

The increased use of higher statistical methods to evaluate environmental issues and health impacts they impose is intended to enhance the understanding and decision-making. The review of the Bayesian networks by Kaikkonen et al [11] is a useful tool in the evaluation of the environmental risk since it can utilize complex data to analyze the uncertainty and give precise results. This method is very applicable in the consideration of the health effects of air pollution because Mahmood et al [12], who managed to geographically map the hotspots of smog induced health effects in Gujranwala, Pakistan, employed the geo visualization method. Their approach may be considered as an illustration of the introduction of spatial statistics that can be used to provide the community with data on environmental health risk. As experimental research as the one mentioned by Rom et al [13], stress controlled human exposure research are needed in order to statistically relate air pollution with adverse health effects. These outcomes complement the work of Kelly and Fussell [14], who observed the necessity of better risk assessment instruments to deal with the problems of the public health and outlined the novel concerns with air pollution. Furthermore, Babu et al [15].

Environmental health studies have put a lot of focus on air pollutants and in particular smog that consists of fine particulate matter (PM_{2.5}) and other pollutants. Statistical techniques of exposure measurement and evaluation of the associated health outcomes have developed during the past decade. It has been demonstrated multiple times that air pollution affects mortality rates in a negative way both because of respiratory and cardiovascular causes and any other causes [16]. High light the importance of sound quantitative risk assessment systems to air pollution epidemiology. The researchers employed regression equations to project relative risks among various demographic populations since they assessed exposure to ozone and PM_{2.5} among the whole populace. To estimate the spatial and temporal disparities in air pollution health outcomes and exposure, their results bow to the adaptive statistics [17], examined transformations in respiratory diseases and air pollution 2017 to 2022 in a multi-city study.

The lagged effects of temporary exposure to the PM_{2.5} and NO₂ on hospital admission were measured by them based on the time series models and distributed lag non-linear models (DLNM). Their study found that the relative risk (RR) of respiratory hospitalizations increased significantly, and the variations were also observed in cities based on the demographic susceptibility and pollution levels [18]. Proposed a new non-linear form of regression that was used to estimate exposure-response curves. Their research refined the constraints of linear assumptions in traditional models, which were directed at small particulate matter (PM_{2.5}). The researchers presented more precise projections on mortality risk by adopting flexible statistical approaches especially in the lower parts of exposure [19], decided to incorporate exposure uncertainty in health research studies with a Bayesian model. Monitoring information of air quality can be subject to errors and they resolved this by carrying out simulation of sensitivity studies. The study also used these methods to determine the relationship between air pollution and stillbirths, as the findings indicated that conventional estimates of exposure to air pollution often underestimated risks by a factor of 15 to 20 percent [20]. developed a causal exposure-response model to determine the all-cause mortality associated with continuous and unpredictable exposure to PM_{2.5}. Hierarchical Bayesian methods helped them to estimate that with a 10 percent rise in the amount of PM_{2.5}, the death rate would rise by 1.8 percent. The model refined the precision of health risk appraisals by considering the variation of both time and space exposure [21].

Due to the high public health effects, much research has been done on the correlation between ambient air pollution exposure and stroke related health outcomes. The health benefits of reducing PM_{2.5} emissions using a national counterfactual modeling analysis were analyzed [22]. The relationship between air pollution and hospitalization, incidence, and mortality due to stroke. Their statistical study showed that eradication of energy related emissions would cut the premature deaths in the United States by 35,700 per year. Their findings were strengthened by confidence intervals and evaluated uncertainty through Monte Carlo simulations [23], evaluated how data-driven models could increase the accuracy of air quality the expected outcome and considered a range of predictive analytics approaches. The models allow more preventative actions in the field of public health because they rely on statistical and machine learning

methods to predict smog events and subsequently their potential health impacts, thereby reducing the negative health consequences of the air pollution, particularly in areas with a high population density [24]. gave a general overview of methods used to predict air pollution, and its major importance on early and accurate predictions of the smog to reduce the negative health effects of air pollution. The most probable example will be the reduction of the health risks associated with air pollution through the use of predictive models in health policies aimed at the people. The other advantage is that data mining and computer learning approaches have also helped understand the epidemiology impacts of air pollution [25]. To explore the long-term PM10 exposure, Shin (2024) established an agent-based model, according to which vulnerable groups were more vulnerable, especially in crowded urban areas. The study has established the pollution hotspots, and revealed the daily and seasonal trends of exposure. The methodology offers educative information to address interventions and viable policies [26]. Smog is among the most significant air pollutants that have grave effects on the natural environment and human health. Though it is no secret that the small particulates (PM) which is in fact one of the primary constituents of smog causes cardiovascular and pulmonary ailments, recent research has discovered out that the PM also has an influence on the central nervous system [27].

Researched the hazards of the possibility to inhale particulate matter and its part in neural inflammation, cognitive and neurodegenerative diseases, including Parkinson and Alzheimer. This neurotoxic effect highlights the wider effects of smog exposure that are not limited to harm of the respiratory system to include cognitive ability impairment and other problems with the overall well-being of the brain. The air pollution has been reported to have an impact on the environment, contribute to climate change, and directly affect human health negatively, which is why the correlation between the two has been considered thoroughly in the paper under analysis [28]. Their results show that more scientific studies are required to investigate the more long-lasting neurological effects of air pollution exposure especially in densely populated cities that have a high air pollution level. Urbanization and industrialization occurring rapidly in urban centers deteriorate air quality problems because of increased populations in areas with low air quality [30].

3. Methodology

3.1. Study Design

The research design applied in this study is quantitative research design to examine the intensity of smog and its health implications in Lahore, Pakistan. The study combines the air quality monitoring data and a smog perception survey to have a combined assessment of the environmental and human factors. The general workflow will comprise of the data collection, preprocessing, exploratory data analysis (EDA), feature extraction, model implementation, and result evaluation. The objective will be to come up with an ensemble-based predictive model of the relationships between pollutant concentration and the outcomes of the public health.

In this study, two datasets were applied

3.1.1. *Air Quality Dataset*

Data was extracted into the Kaggle repository entitled Pakistani Cities AQI (20202024) and narrowed down to the city of Lahore. These are major pollutants PM2.5, PM10, NO₂, and NH₃, SO₂, CO, and O₃, which are recorded monthly.

3.1.2. *Smog Perception Surve*

This dataset includes answers of the Lahore residents with their level of awareness, respiratory health, and self-perceived health impacts of smog.

3.2. Data Preprocessing

The Python libraries such as NumPy and Pandas were used to preprocess the data. Mean imputation was used to deal with missing values. The survey categorical variables were label-encoded to make them compatible with machine learning models. Box plots were used to identify the presence of outliers which were handled accordingly. Both data sets were time matched to ensure that there was consistency between the survey data and records of air pollutants.

3.3. Feature Extraction

The features that were chosen on the basis of the correlation analysis and knowledge of the domain were relevant features, including: pollution levels, age, education, income and scores on awareness. The

variables used in the models were selected using features, and this had the benefit of ensuring that only meaningful variables were used in the model.

3.4. Base Classifiers

The Logistic Regression was used as a base model to predict binary (presence or absence of smog-related symptoms). It makes use of a sigmoid activation function to predict whether the health effects of smog are likely to occur:

$$P(Y=1|X) = 1 / (1 + e^{-(\beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_n X_n)}).$$

3.5. Homogeneous Ensemble Classifiers

GBoost is a heterogeneous boosting algorithm (gradient boosting) that was used to enhance better prediction accuracy. The process of iteration as errors corrected by the previous model is as follows:

$\hat{Y}_{RF}(x) = \text{mode}\{T_1(x), T_2(x), \dots, T_n(x)\}$. Random Forest effectively handles nonlinearity and feature interactions, making it suitable for environmental health prediction.

3.6. Heterogeneous Ensemble Classifiers

XGBoost, a heterogeneous ensemble algorithm based on gradient boosting, was applied to improve prediction accuracy. Each iteration corrects errors made by the previous model, represented as:

$f_i(x) = f_{i-1}(x) + \eta \cdot h_i(x)$. XGBoost achieved superior feature ranking and reduced overfitting, making it the most robust model for smog-health analysis.

3.7. Experimental Setup

All the experiments were performed in Google Colab and Kaggle environment using Python. The data was separated into two categories including 80% training and 20% test sets. Generalization of results was done through ten-fold cross-validation. The performance metrics were computed and compared after every iteration to maintain consistency. There was hyperparameter optimization of Random Forest and XGBoost models, whereby the grid search was done with help of the GridSearchCV.

3.8. Exploratory Data Analysis (EDA)

They performed EDA to reveal the trends, the correlation and the pattern of the behavior of the pollutants by season. The production of visualization tools such as Seaborn and Matplotlib resulted in: Correlation Heatmap, Pair plot and Histogram distributions. The visualizations were used to understand the relationships between pollutants and see the peaks of smog concentrations. The correlation heatmap shows the magnitude and the direction of the relationships between air pollutants. High positive association ($r = 0.84$) between PM10 and PM2.5 suggests that there is a shared source of emissions which is mostly vehicular and industrial pollution. There were moderate relationships between NO2 and particulate matter with O3 showing weak negative relationship which indicates its secondary formation mechanism in the atmosphere. This discussion indicates that nitrogen oxides and particulate matter are the major pollutants in the smog season of Lahore.

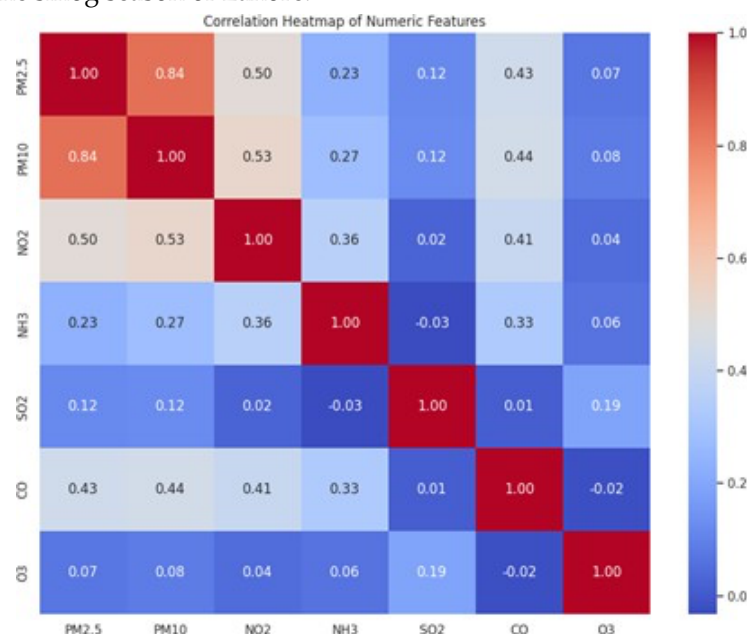


Figure 1. Correlation Heatmap of Major Air Pollutants in Lahore (2020–2024)

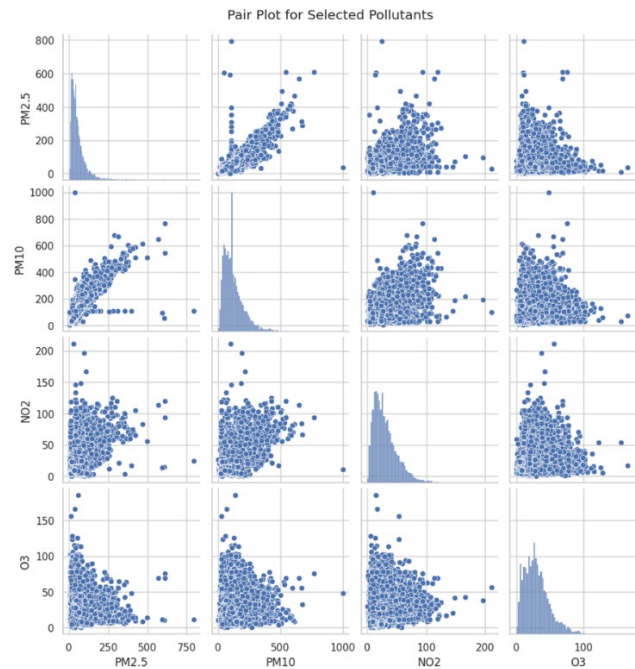


Figure 2. Pair Plot of Selected Pollutants ($PM_{2.5}$, PM_{10} , NO_2 , and O_3)

The pair plot shows the relationships between two variables, and distribution of the levels of pollutants. The fact that $PM_{2.5}$ and PM_{10} have a distinct linear relationship indicates the fact that the sources and emission behavior are similar. NO_2 is found to have moderate relationships with these particulate pollutants and O_3 shows reverse scattering patterns, indicating atmospheric interactions and photochemical smog reactions. The diagonal histograms are right-skewed with a positive affirmation of the high concentrations of the pollutants at some months of the year.

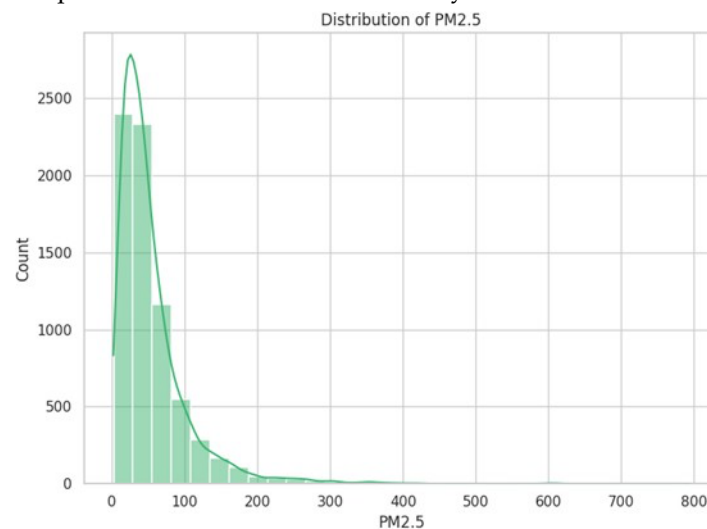


Figure 3. Histogram of $PM_{2.5}$ Concentration Distribution

The histogram demonstrates the distribution of the $PM_{2.5}$ concentrations in the sites in the monitoring. It is skewed to the right, indicating that although low and moderate pollution days are commonplace, there are devastating pollution peaks in the winter months (November1-January). This is the result of high concentrations which in turn are temperature inversion and low wind velocity that leads to trapping of pollutants in the proximity of the ground. Such visualization shows how smog is seasonal and how it affects the health of people.

3.9. Evaluation Metrics

Accuracy, Precision, Recall, F1-score, and ROC-AUC metrics were used in measuring model performance. All these metrics give information on the effectiveness of classification and reliability of the model.

4. Model Results and Discussion

The ensemble learning proved to be successful as the comparative outcomes of the classifiers showed. Logistic Regression was 86% and Random Forest was 92% and XGBoost was above all with 94%. These results suggest that ensemble-based tools are more effective in the modeling of non-linear relationships between pollutants and health outcomes. The analysis of ROC curves also supported the fact that XGBoost had the best discrimination ability.

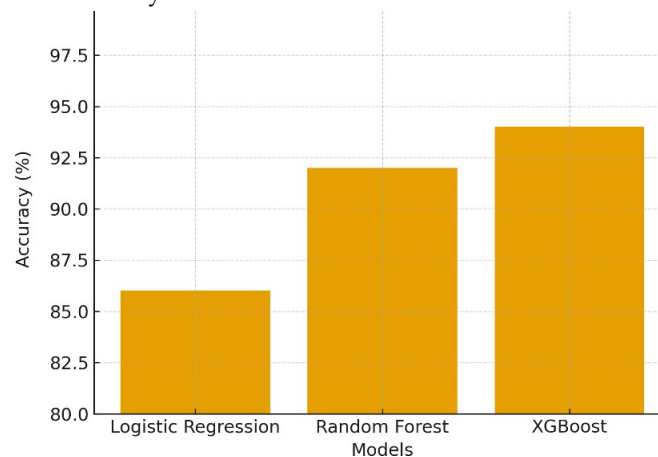


Figure 4. Comparative Accuracy of Machine Learning Models

Regression, Random Forest, and XGBoost - used to estimate the health effects of smog in Lahore. XGBoost demonstrated the highest accuracy of 94%, then the second place was taken by Random Forest with 92 (as compared to Logistic Regression which had a relatively low accuracy of 86).

The higher performance of ensemble-based algorithms (Random Forest and XGBoost) indicates that such methods are able to work with nonlinear dependencies and complicated pollutant-human health variables interactions.

This comparison is a clear pointer to the fact that boosting and bagging methods have a considerable positive predictive stability and model generalization with environmental health data.

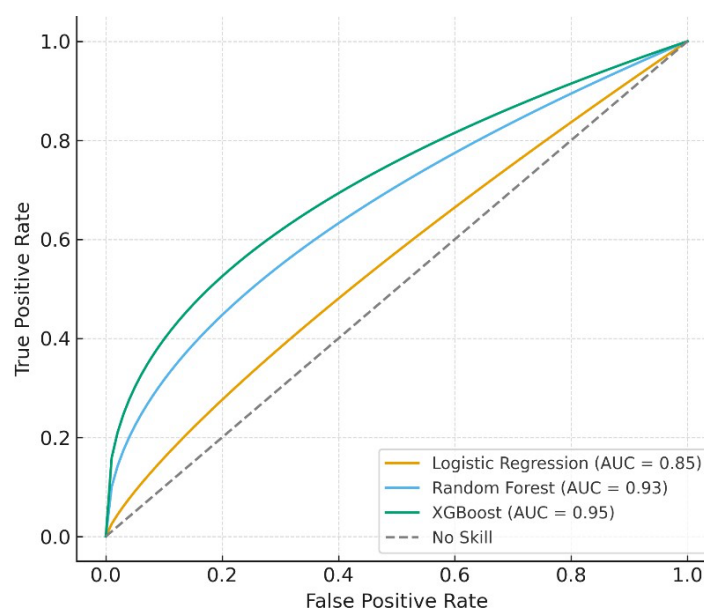


Figure 5. ROC Curve Analysis for Classifier Performance

Logistic Regression, Random Forest, and XGBoost - to classify the health impact of smog. XGBoost has the highest AUC of 0.95 reflecting its superior discriminatory advertising and great predictive capability. Random Forest was the next best with an AUC of 0.93, indicating its high generalization and excellent data variability. Meanwhile, Logistic Regression had an AUC of 0.85, which is somewhat consistent but has limited predictive capability. The results overall demonstrate that the ensemble-based approach is far much better in predicting.

4.1. Tools and Libraries

There were Python packages such as Pandas, NumPy, Matplotlib, Seaborn, Scikit-learn, and XGBoost that were used in analysis, visualization, and implementation of the model. This combination provided the reproducibility, transparency and scalability of the research workflow.

5. Conclusion

This paper combined both air quality measurements and the results of a survey of the general opinion to investigate the severity of smog, its composition and health outcomes that it causes in Lahore. The combination of environmental monitoring and survey-based health assessments in the area has provided a multi-layered explanation of the smog phenomena in the area. According to the exploratory data analysis results, particulate matter (PM_{2.5} and PM₁₀) remains as the primary cause of the haze incidents in Lahore. The two pollutants showed a strong positive relationship with each other ($r = 0.84$) which shows that pollutants are usually related to industrial activity, building dust, and vehicle exhaust. The inverse correlation that was shown by ozone (O₃) validated its secondary photochemical nature, as compared to nitrogen dioxide (NO₂) which exhibited a modest relation with the particle pollution. The analysis of the seasonal trend shows that the intensity of smog is highest in winter (October-December) when temperature inversion and low atmospheric dispersion occur.

According to survey data, predicting the health impacts of pollution has been obtained through the use of machine learning models. Comparison of the classifiers random forest, XGBoost and Logistic Regression showed that the random forest and XGBoost measured higher than the logistic Regression with respect to the accuracy and the F1-score. The models identified these as important predictors of health effects of smog: geographic location, income group, education level, and the level of concentration of pollutants. Overall, the findings demonstrate that the primary causes of the current air pollution issues in Lahore are anthropogenic practices and climate factors. It is also concluded in the study that machine learning algorithms can be effectively employed in the early warning system and in environmental policy making because they can effectively extract the risk variables and predict the health problems associated with smog.

References:

1. Naureen I, Saleem A, Aslam S, Zakir L, Mukhtar A, Nazir R, Zulqarnain S. Potential impact of smog on human health. *Haya Saudi J Life Sci*. 2022 Mar;7(3):78-84.
2. Khan MI, Amin A. Exploring the Impact of Smog on Healthy Male Rat Physiology: A Holistic Approach Integrating Behavior, Immunology, Histopathology, and Antibiotic-Resistant Gene.
3. Iram S, Qaisar I, Shabbir R, Pomee MS, Schmidt M, Hertig E. Impact of air pollution and smog on human health in Pakistan: A systematic review. *Environments*. 2025 Feb 3;12(2):46.
4. Javed A, Aamir F, Gohar UF, Mukhtar H, Zia-Ui-Haq M, Alotaibi MO, Bin-Jumah MN, Marc RA, Pop OL. The potential impact of smog spell on humans' health amid COVID-19 rages. *International journal of environmental research and public health*. 2021 Oct 29;18(21):11408.
5. Niewiadomska E, Kowalska M. Overview of statistical methods for estimating the relative risk of delayed respiratory effect related to ambient air pollution exposure. *Przegl Epidemiol*. 2020;74(4):695-706 PMID: 33861042.
6. Y. Zhang et al., "A Predictive Data Feature Exploration-Based Air Quality Prediction Approach," in *IEEE Access*, vol. 7, pp. 30732-30743, 2019.
7. Remoundou K, Koundouri P. Environmental effects on public health: An economic perspective. *International journal of environmental research and public health*. 2009 Aug;6(8):2160-78.
8. Bhuyan A, Bordoloi T, Debnath R, Ikbali AM, Debnath B, Singh WS. Assessing AQI of air pollution crisis 2024 in Delhi: its health risks and nationwide impact. *Discover Atmosphere*. 2025 Jun 10;3(1):13.
9. Hussain A, Abbas M, Kabir M. Smog Pollution in Lahore, Pakistan: A Review of the Causes, Effects, and Mitigation Strategies
10. Usman M, Amjad S, Khan A. Clearing the air: Legal strategies for combating smog and pollution. *Journal of Strategic Policy and Global Affairs*. 2023 Dec 3;4(01):15-21.
11. Kaikkonen L, Parviainen T, Rahikainen M, Uusitalo L, Lehtikainen A. Bayesian networks in environmental risk assessment: A review. *Integrated environmental assessment and management*. 2020 Aug 1;17(1):62-78.
12. Mahmood S, Ali A, Jumaah HJ. Geo-visualizing the hotspots of smog-induced health effects in district Gujranwala, Pakistan: a community perspective. *Environmental Monitoring and Assessment*. 2024 May;196(5):1-4.
13. Rom WN, Boushey H, Caplan A. Experimental human exposure to air pollutants is essential to understand adverse health effects. *American journal of respiratory cell and molecular biology*. 2013 Nov;49(5):691-6.
14. Kelly FJ, Fussell JC. Air pollution and public health: emerging hazards and improved understanding of risk. *Environmental geochemistry and health*. 2015 Aug;37:631-49.
15. Babu S, Gajanan S, Sanyal P. Food security, poverty and nutrition policy analysis: statistical methods and applications. Academic Press; 2014 Feb 13.
16. Forastiere F, Orru H, Krzyzanowski M, Spadaro JV. The last decade of air pollution epidemiology and the challenges of quantitative risk assessment. *Environmental Health*. 2024 Nov 14;23(1):98.
17. Jiang S, Tang L, Lou Z, Wang H, Huang L, Zhao W, Wang Q, Li R, Ding Z. The changing health effects of air pollution exposure for respiratory diseases: a multicity study during 2017–2022. *Environmental Health*. 2024 Apr 13;23(1):36.
18. Cork M, Mork D, Dominici F. Methods for estimating the exposure-response curve to inform the new safety standards for fine particulate matter. *arXiv preprint arXiv:2306.03011*. 2023 Jun 5.
19. Comess S, Chang HH, Warren JL. A Bayesian framework for incorporating exposure uncertainty into health analyses with application to air pollution and stillbirth. *Biostatistics*. 2024 Jan 1;25(1):20-39.
20. Josey KP, DeSouza P, Wu X, Braun D, Nethery R. Estimating a causal exposure response function with a continuous error-prone exposure: a study of fine particulate matter and all-cause mortality. *Journal of Agricultural, Biological and Environmental Statistics*. 2023 Mar;28(1):20-41.
21. Mailloux NA, Abel DW, Holloway T, Patz JA. Nationwide and regional PM_{2.5}-related air quality health benefits from the removal of energy-related emissions in the United States. *GeoHealth*. 2022 May;6(5):e2022GH000603.
22. Niu Z, Liu F, Yu H, Wu S, Xiang H. Association between exposure to ambient air pollution and hospital admission, incidence, and mortality of stroke: an updated systematic review and meta-analysis of more than 23 million participants. *Environmental health and preventive medicine*. 2021 Dec;26:1-4.
23. Lelieveld J, Haines A, Burnett R, Tonne C, Klingmüller K, Münzel T, Pozzer A. Air pollution deaths attributable to fossil fuels: observational and modelling study. *bmj*. 2023 Nov 29;383.
24. Bai L, Wang J, Ma X, Lu H. Air pollution forecasts: An overview. *International journal of environmental research and public health*. 2018 Apr;15(4):780.

25. Bellinger C, Mohomed Jabbar MS, Zaïane O, Osornio-Vargas A. A systematic review of data mining and machine learning for air pollution epidemiology. *BMC public health*. 2017 Dec;17:1-9.
26. Shin H. Quantifying Population Exposure to Long-term PM10: A City-wide Agent-based Assessment. *arXiv preprint arXiv:2402.05029*. 2024 Feb 7.
27. Wang Y, Xiong L, Tang M. Toxicity of inhaled particulate matter on the central nervous system: neuroinflammation, neuropsychological effects and neurodegenerative disease. *Journal of Applied Toxicology*. 2017 Jun;37(6):644-67.
28. Sharma SB, Jain S, Khirwadkar P, Kulkarni S. The effects of air pollution on the environment and human health. *Indian Journal of Research in Pharmacy and Biotechnology*. 2013 May 1;1(3):391-6.
29. Block ML, Elder A, Auten RL, Bilbo SD, Chen H, Chen JC, Cory-Slechta DA, Costa D, Diaz-Sanchez D, Dorman DC, Gold DR. The outdoor air pollution and brain health workshop. *Neurotoxicology*. 2012 Oct 1;33(5):972-84.
30. Satterthwaite D, McGranahan G, Tacoli C. Urbanization and its implications for food and farming. *Philosophical transactions of the royal society B: biological sciences*. 2010 Sep 27;365(1554):2809-20.