Social and Occupational Ergonomics, Vol. 190, 2025, 7–16
https://doi.org/10.54941/ahfe1006591

Mitigating Pakistan's Smog Crisis Through Predictive Modelling

Javed Anjum Sheikh[1], Usman Javed Sheikh[2] and Farzan Sheikh[3]

[1]Fazaia College of Education for Women, Lahore/Air University, Islamabad, Pakistan
[2]Universidad De Zaragoza, Zaragoza, Spain
[3]Superior University Lahore, Pakistan

ABSTRACT

The phenomenon of smog in Pakistan has developed into a substantial issue regarding both environmental integrity and public health, predominantly influenced by swift urbanization, industrial operations, vehicle emissions, and the periodic incineration of agricultural waste. Major cities, such as Lahore, Karachi, and Islamabad, experience severe smog in the winter season because of atmospheric temperature inversions that suppress the vertical dispersion of pollutants such as particulate matter (PM2.5), nitrogen oxides (NOx), sulfur dioxide (SO_2), carbon monoxide (CO), and ground-level ozone (O_3). In this research, we employ a predictive modeling strategy to address smog as a proactive, not a reactive, issue. Employing machine learning methods—regression analyses, time-series predictions, and classification models—combined with meteorological and emissions data, the study will strive to predict smog intensity and longevity with better precision. Incorporation of satellite data, specifically from MODIS and Sentinel-5P, with ground-based air quality measurements greatly improves model strength. Furthermore, creation of the hybrid methods that fuse artificial intelligence and numerical weather prediction models like the Weather Research and Forecasting (WRF) model has exhibited excellent potential. According to the Sustainable Development Goal 11 (Sustainable Cities and Communities) and Goal 3 (Good Health and Well-being) indicators, this study seeks to enhance real-time air quality monitoring, strengthen early warning systems, and guide long-term mitigation measures, including urban reforestation and transitioning to cleaner fuels. But chronic problems—above all, limited monitoring infrastructure and patchy data availability—underline the requirement for higher investment, capacity development, and cross-sectoral coordination to ensure effective smog reduction.

Keywords: SMOG, Pakistan, Predictive Modelling, SGD,

INTRODUCTION

The smog problem has developed into a serious environmental and public health issue in Pakistan, particularly in urban centers such as Lahore, Karachi, and Islamabad. The phenomenon is mainly driven by rapid urbanization, industrial emissions, vehicle exhaust, and seasonal burning of agricultural residue. Critical pollutants—such as fine particulate matter (PM2.5 and PM10), nitrogen oxides (NOx), sulfur dioxide (SO_2), carbon monoxide (CO), and ozone (O_3)—tend to accumulate at lower elevations during winter months due to temperature inversions, leading to longer exposure and greater health risks (Shahid, Khan and Javed, 2023). According to the World Health Organization (2020), air pollution accounts for over 4.2 million premature deaths globally each year.

In Pakistan, cardiovascular and respiratory illnesses associated with smog impose tremendous costs on the health care system and further decrease the quality of life (Gul, Arif, and Imran, 2021; Bashir, Akram, and Raza, 2024). Urban cities like Lahore consistently rank among the most polluted cities in the world, and Air Quality Index (AQI) levels routinely hit hazardous thresholds (Reuters, 2024; Le Monde, 2024). Despite growing public and institutional awareness, Pakistan lacks a nationwide, real-time smog monitoring and forecasting system. Any existing air quality monitoring stations are limited in numbers, costly to operate, and geographically limited, providing low spatial coverage and public engagement. As a result, a vast majority of the population is not provided with information on ongoing smog conditions, thereby pre-empting timely mobility-related, outdoor exposure-related, and health-related choices (AP News, 2024).

The study suggested herein seeks to surmount these obstacles through the use of advanced technological tools to improve the predictive ability in addition to accessibility of smog information. With the incorporation of real-time meteorological data, machine learning algorithms, and satellite imagery, the system aims to provide more precise and timely predictions. In addition, the project aligns with a number of United Nations Sustainable Development Goals (SDGs), including SDG 3 (Good Health and Well-being), SDG 11 (Sustainable Cities and Communities), SDG 13 (Climate Action) and SDG 15 (Life on Land). The ultimate aims to support evidence-based policy making, increase, increase the resilience of various communities, and minimize the impacts of smog in urban areas of Pakistan.

LITERATURE REVIEW

Smog in Pakistan: A Growing Crisis: Major cities like Lahore, Karachi, Islamabad, and Faisalabad are facing significant deterioration in air quality due to transportation emissions, industrial activities, and, and the seasonal burning of agricultural residues (Gul et al., 2023). Apart from this, meteorological conditions like temperature inversions and low wind speeds one side, and other side are $PM_{2.5}$, PM_{10}, NO_x, SO_2, CO, and ground-level ozone, worsening the overall air quality index (Ali et al., 2022).

Last few years, Lahore and Karachi consistently rank among the world's most polluted cities, and their air quality indices mostly crossing the limit established by the World Health Organization (WHO) (Shah et al., 2020). The main player to this persistent pollution is transportation and crop burning during harvesting seasons, and when these combined with winter weather condition, lead to dense smog (Ali et al., 2020; Shah et al., 2021; Wang et al., 2022).

Poor air quality is more than an environmental and public health concern; it is a significant barrier to achieving many of the Sustainable Development Goals (SDGs) set by the United Nations. Specifically, SDG 3 (Good Health and Well-being), SDG 11 (Sustainable Cities and Communities), SDG 13 (Climate Action) and SDG 15 (Life on Land) are directly influenced (UN, 2022). Reducing air pollution is therefore both a priority for public health and for sustainable urbanization in Pakistan.

Advances in ML Models for AQI Prediction: According to the literature review the immense potential of machine learning (ML) models for the prediction of air quality with accuracy In cities like Delhi, models such as XGBoost, when integrated with meteorological data, have demonstrated high precision in forecasting AQI (Kumar et al., 2023). Similarly, the combination of LSTM networks and RF algorithms in China has enhanced the accuracy of temporal predictions of $PM_{2.5}$ levels (Chen et al., 2021). In comparison, Pakistan's use of such advanced techniques remains limited. Most of the studies carried out locally continue to employ basic regression techniques, which do not have the ability to capture the complex, nonlinear relationships between emissions, climatic factors, and terrain (Ahmed et al., 2021). This highlights the immediate need for advanced, location-specific production technique.

Hybrid Deep Learning Models: The integration of deep learning and traditional machine learning techniques has potential in capturing complex spatial-temporal patterns in air quality data. Kim et al. (2023) introduced a hybrid CNN-LSTM model for effectively modelling (both spatial and temporal) for real-time air quality prediction. Additionally, reinforcement learning techniques optimise the adaptability and reliability of air quality alert systems (Chen et al., 2024). Hybrid models offer improved predictive accuracy, If we compare them with conventional statistical models (ARIMA or linear regression). These advantages are relevant for Pakistan, where urbanization, and inconsistent monitoring infrastructure demand intelligent perdition solutions.

Predictive Analytics in Smog Forecasting: The use of predictive analytics and machine learning techniques is increasing due the significant accuracy of smog prediction. Models like Decision Trees (DT) and Support Vector Machines (SVM) have enhanced the precision of $PM_{2.5}$ predictions in Beijing by effectively learning from historical pollution trends (Zhang et al., 2022). Similarly, in Lahore, Long Short-Term Memory (LSTM) networks have demonstrated strong predictive capabilities when incorporating air pollution indices, meteorological parameters, and temporal factors (Khan et al., 2023). Integrating these models enhances the formation and development of smog, thereby providing an improved framework for prediction and mitigating air quality deterioration (Jiang et al., 2021).

Integration of Satellite and Meteorological Data: The integration of satellite and meteorological data play vital in improving the accuracy and responsiveness of air quality prediction systems. Meteorological factors such as wind speed, humidity, and temperature inversion significantly influence pollutant in winter, when light winds and inversion layers trap smog near the surfaced (Shao et al., 2021; Wang et al., 2019). Further, Machine learning and metrological data enhance the predictive precision. Remote sensors like MODIS, OMI, and Sentinel-5P offer vital atmospheric variables like Aerosol Optical Depth (AOD), nitrogen dioxide (NO_2), and surface reflectance that support Air Quality Index (AQI) prediction (Bilal et al., 2021). In Pakistan, where ground-level measurements of air quality are sparse, satellite-derived data serves as a worthy complement. Luo et al., (2023) in Southeast Asia and in Punjab (Shah et al., 2022) also supports that integrative development yields more transferable, and stable smog prediction model.

Emission Data and Pollution Sources: Emission data form Traffic and industrial operations are primary contributors to air pollution, along with seasonal agricultural burning. Emission levels vary by region and seasonal patterns, influencing the distribution and concentration of key pollutants like nitrogen oxides (NO_x), sulphur dioxide (SO_2), volatile organic compounds (VOCs), and fine particulate matter ($PM_{2.5}$) (Shah et al., 2020; Agarwal et al., 2019). The inclusion from specific emissions into predicting models enables a more precise simulation of smog.

Satellite-Based Monitoring for Smog Detection: Satellite-based monitoring systems based on MODIS and Sentinel-5P, offer spatial coverage and real-time atmospheric data, to improve the reliability and accessibility of air quality prediction models (Hu et al., 2020; Liu et al., 2017). Li et al. (2020) combined satellite data with ground-based measurements and result was significantly improved both the spatial resolution and overall accuracy of smog predication. Satellite-based monitoring system for Pakistan is a vital tool for effective detection of smog. as ground monitoring infrastructure is limited.

Smog Prediction Models: Smog predictive models are vital tools for protect public health and guiding timely policy responses. Early warning systems enable to take precautionary measures while giving authorities enough time for effective interventions (Rahman et al., 2018). Conventional linear models fail short to represent interactions relationships between emissions, meteorology, and geographic features. It is also to be noted that machine learning approaches, time series and classification techniques offer a more adaptive understanding of smog formation and behavior (Zhao et al., 2020; Yuan et al., 2021).

Problem Statement

Pakistan's cities, especially Lahore, Karachi, Islamabad and Faisalabad, are facing toxic air quality. Whereas like $PM_{2.5}$ and PM_{10} regularly exceed safety limit, posing serious risks to public health and environmental stability. The exacerbating environmental emergency has caused an influx of respiratory and cardiovascular ailments, which represent a dire public health risk. The issue is exacerbated by the lack of an effective and real-time air quality tracking system, seriously undermining the ability of the authorities to predict and address instances of smog. While meteorological conditions—temperature inversions, wind speed, and humidity—are reported to influence pollutant dispersion (Ali, Ali and Khan, 2022), empirical research on the interaction of variables with urban microclimates and seasonal cycles in the Pakistani context is limited (Shahid, Khan and Javed, 2023).

Despite growing global application of machine learning models in real-time air quality forecasting (Kumar, Mishra and Sharma, 2023), their utilization in Pakistan is still in its nascent stage. Consequently, Pakistan does not have an effective dearth of credible forecasting tools that can ensure early warning systems, especially during winter months when smog incidences are more pronounced (Gul, Arif and Imran, 2021). Moreover, though existing research has ascertained the combined health effect of air pollution in Pakistan (Shahid, Khan and Javed, 2023; Gul, Arif and Imran, 2021), a critical gap still exists in knowing the intricate, dynamic relationships between fine particulate matter and health outcomes in different seasons, meteorological conditions, and intra-urban geographical areas.

Previous work has applied meteorological data in investigating the processes of smog formation (Ali, Ali and Khan, 2022); however, it is rare that spatial, temporal, and seasonal trends as well as health outcome data have been utilized in conjunction to develop in-depth predictive models. In addition, sophisticated machine learning methods—like Long Short-Term Memory (LSTM) networks and hybrid CNN-LSTM models, which have proven superior performance in global contexts (Khan, Shabbir and Noor, 2023; Kim, Lee and Choi, 2023)—are still not well explored in Pakistan either for public health monitoring or air pollution policy-making.

This study strives to fill these identified gaps through the creation of a hybrid machine learning-based smog prediction system that combines meteorological, satellite, pollution, and health data. The system suggested here will not only try to improve prediction accuracy but also create actionable intelligence to facilitate timely public health responses and evidence-based policy initiatives in Pakistan's cities.

Research Questions

1. To establish the relationship between $PM_{2.5}$ and PM_{10} respiratory disease in urban Pakistan, considering seasonal, meteorological, spatial, and temporal differences?
2. What ways can hybrid machine learning models be used to develop precise tools for smog prediction and air quality assessment in Pakistan?
3. How would the suggested smog prediction system enhance public health outcomes and guide air quality control policy in Pakistan?

Research Objectives

1. To establish the correlation between $PM_{2.5}$ and PM_{10} concentrations and the prevalence of respiratory illnesses in the major cities of Pakistan, according to health and air quality reports during different seasons, meteorological conditions, sites, and time periods over a 12-month observational duration.
2. Develop, design, and test hybrid machine learning models (e.g., LSTM-RF, CNN-LSTM, XGBoost) for accurate smog prediction and real-time air quality monitoring using satellite, meteorological, and ground-based data over a period of six months of model development and testing.
3. To evaluate the impact of the proposed smog prediction system on public health outcomes and policy-making, this project will entail scenario simulations, stakeholder interviews, and policy analyses, finally making recommendations for data-driven air quality management within three months of the system's deployment.

RESEARH METHODOLOGY

The solution integrates machine learning algorithms with satellite tracking and weather data to address Pakistan's air pollution problem in relation to the Sustainable Development Goals (SDGs) Smog predictive models support the following Sustainable Development Goals (SDGs).

1. **SDG 3: Good Health and Well-being** – The prediction of smog reduces exposure-related health risks, allowing better public health outcomes (Rahman et al., 2018).
2. **SDG 11: Sustainable Cities and Communities** – The better urban planning and air quality management strategies, contributing to healthier, more resilient cities (Gul et al., 2021).
3. **SDG 13: Climate Action** Monitoring pollution trends identify emission sources and supports efforts to curb climate change through informed policy actions (Li et al., 2020).

4. **SDG 15: Life on Land** – Reducing smog's impact helps ecosystems, supporting overall environmental sustainability (Agarwal et al., 2019).

Selection of Study Areas: The following four cities will be part of the study:

1. **Lahore**: The city has dense population and face winter smog, largely driven by vehicular emissions and the burning of crop in the agricultural fields (Ali et al., 2020).
2. **Karachi**: A coastal but highly industrialized city, face air pollution from factories, and ports (Gul et al., 2021).
3. **Islamabad/Rawalpindi**: These twin cities are impacted by growing nitrogen dioxide (NO_2) and carbon dioxide (CO_2) from increasing transport (Shah et al., 2020).
4. **Faisalabad**: As an industrial hub, it contributes $PM_{2.5}$ and PM_{10}, because of dense network of factories (Agarwal et al., 2019).

Data Collection and Sources: The study will use the following different datasets for analysis of air quality and smog prediction:

1. **Air Quality Data (2018–2024)**: PM2.5, PM10, NOx, SOx, CO, O_3, and AQI data will be gathered from the Pakistan Environmental Protection Agency, IQAir, OpenAQ, and local monitoring departments (Li et al., 2020).
2. **Meteorological Data**: Temperature, humidity, wind speed, rain and solar radiation will be gathered from the Pakistan Meteorological Department, where as global data from NASA's MERRA-2 and ERA5 data (Shao et al., 2021).
3. **Socioeconomic & Human Activity Data**: Traffic volume, population data, industrial production data, and crop-burning data will be obtained from the Pakistan Bureau of Statistics, Safe Cities Authority, and satellite remote sensing data from MODIS (Zhang et al., 2018).
4. **Health Data (Optional)**: Data for cardiovascular and respiratory diseases, will be obtained after following ethical approval (WHO, 2021) to examine health impacts.

Data Pre-processing: It will address data inconsistency the cleansing, harmonizing, and consolidating of different datasets:

1. **Missing Data**: Techniques such as K-Nearest Neighbors (KNN) and certain imputation techniques will fill gaps because of equipment failure and restricted sensor availability within rural settings (Ali et al., 2024; Rehman et al., 2023).
2. **Data Quality Issues**: Outlier detection and normalization methods (e.g., Z-score or IQR) will be used to pre-process the data so that the final model is reliable (Hassan et al., 2025).
3. **Satellite Data Integration**: The integration of satellite data and sparse ground data from instruments like Sentinel-5P and MODIS will improve model performance, with sparse monitoring (Iqbal et al., 2022).

Model Development: It will consider regional pollution dynamics, meteorology, and socio-economic parameters:

1. **Regional Emission Patterns**: Regional emission inventories and source pollution (e.g., crop burning in Punjab, vehicular emissions in Karachi, industrial emissions in Faisalabad) will be incorporated in machine learning

algorithms like Random Forest and XGBoost (Shah & Nawaz, 2023; Abbas et al., 2022).

2. **Weather Influence**: Hybrid modeling methods, i.e., ARIMA-LSTM will be applied to identify non-linear relationships between meteorological conditions and air pollution levels (Farooq et al., 2023; Ahmed et al., 2021).
3. **Seasonal Effects**: The seasonal winter smog trends in Lahore shall be analyzed using techniques such as Seasonal Autoregressive Integrated Moving Average (SARIMA) to address temporal correlations (Mehmood & Javed, 2023).

Feature Engineering: It will increase the model's predictive power:

1. **Local Factors**: Socio-environmental determinants, traffic, industrial complexes, and biomass burning incidence will be coded as categorical or numeric attributes (Zaheer et al., 2024; Raza & Hameed, 2023).
2. **Weather Interactions**: Interaction terms (e.g., wind direction × PM2.5 concentration) will capture pollutant dispersion, especially in urban regions like Lahore (Iqbal et al., 2022).
3. **Time-Based Features**: Time-of-day, weekday/weekend, and seasonal trends will be crucial to capture daily patterns of pollution (Usman et al., 2024).
4. **Lag Features**: Pollutant values lagged will predict the future levels of smog, as attested to by local research (Hussain et al., 2021).

Model Evaluation and Interpretation: To ensure applicability across diverse socio-environmental conditions:

1. **Accuracy vs. Interpretability**: While deep learning models like LSTM yield higher accuracy, interpretable models like Random Forests with SHAP explanations will be employed to ensure transparency for policymakers (Khan & Ashraf, 2025).
2. **Feature Attribution**: SHAP values and Partial Dependence Plots (PDPs) will measure the impact of factors such as crop burning and traffic congestion on smog levels (Tariq et al., 2023).
3. **Generalizability**: Model performance will be cross-validated for other cities (i.e., model trained from Lahore tested against other cities' datasets with ROC-AUC and RMSE metrics (Rana et al., 2023)).
4. **Health and Policy Impact**: Utilizing simulations based on historical data will aid in evaluating the capacity of the model to deliver prompt alerts and predict health outcomes, including a decrease in occurrences of respiratory diseases (Jalil & Sadiq, 2022).

Ethical Considerations: Ethical principles will be adhered:

1. **Data Privacy**: All the data will be made anonymous, and the information will be handled with ethical clearance from the relevant review committees (Sharma et al., 2022).
2. **Environmental Data Handling**: Satellite imagery will be processed according to privacy rules, and data minimization principles to protect sensitive information (Sharma et al., 2022).

Expected Outcomes: The study aims to attain the following:

1. **High-Accuracy Smog Prediction System**: A hybrid machine learning model design according to the environmental conditions of Pakistan.

2. **Public Health Benefits**: Enhanced early detection systems to reduce respiratory health issues (Ali et al., 2023).
3. **Actionable Policy Recommendations**: To support short-term interventions and long-term regulatory frameworks.
4. **Operational Dashboard**: A real-time web-based system for monitoring air quality monitoring platform and disseminating policy updates.

Challenges and Limitations: Challenges include:

1. **Limited Data Coverage**: Satellite data and ground measurements will minimize the lack of air quality sensors in rural areas (Hussain et al., 2024).
2. **Model Generalization**: Both topographic and source pollution variability can affect the transferability of models, which will be addressed using domain adaptation techniques (Rizvi et al., 2023).
3. **Public Awareness**: It is critical that the model must success to engage the public, through community workshops and local dashboards.

CONCLUSION

The goal of this study is to create a dependable smog prediction system for Pakistan based on the integration of cutting-edge machine learning algorithms, meteorological predictions, and satellite tracking to provide enhanced accuracy, real-time application, and actionable suggestions to support environmental sustainability.

REFERENCES

Abbas, S., Iqbal, M. and Rizvi, F., 2022. Regional emission patterns and machine learning models in predicting air pollution: A case study of Faisalabad. *Environmental Science & Technology*, 56(3), pp.222-230.

Agarwal, R., Jain, P. and Khan, M., 2019. Environmental impacts of smog: A case study on the agricultural and industrial sectors in Pakistan. *Environmental Monitoring and Assessment*, 191(1), pp.35-42.

Ahmed, F., Rao, R. and Naeem, S., 2021. Limitations of regression models in air quality forecasting in Pakistan. *Pakistan Environmental Journal*, 11(2), pp.202-211.

Ali, A., Ali, M. & Khan, M., 2022. Meteorological factors influencing smog in Pakistan: A comprehensive review. *Journal of Climate and Air Quality*, 29(3), pp.125-139.

Ali, A. and Hussain, S., 2020. Impact of crop residue burning on smog formation in Lahore. *Journal of Environmental Research*, 12(2), pp.58-65.

Ali, M., Rahman, A. and Zahid, S., 2023. Early warning systems for respiratory diseases linked to air pollution. *Public Health and Environment*, 15(4), pp.50-57.

Bashir, M., Akram, W. and Raza, S., 2024. Smog-related cardiovascular and respiratory illnesses in Pakistan. *Journal of Environmental Health*, 27(1), pp.88-95.

Bilal, M., Yaseen, Z. and Raza, M., 2021. Using satellite data for improving air quality predictions in Lahore. *Remote Sensing of Environment*, 250, pp.110-120.

Chen, J., Zhang, W. and Wang, Y., 2021. Hybrid LSTM-RF model for $PM_{2.5}$ forecasting: A study of China. *Science of the Total Environment*, 770, pp.145-156.

Chen, X., Zhao, R. and Li, J., 2024. Reinforcement learning for improving AQI prediction reliability. *AI in Environmental Sciences*, 12(4), pp.33-41.

Farooq, A. and Ahmed, N., 2023. ARIMA-LSTM hybrid model for predicting air quality based on meteorological data: A study on Lahore smog. *Computational Environmental Science*, 8(2), pp.90-99.

Gul, M., Arif, F. and Imran, S., 2021. Public health consequences of air pollution in Pakistan's urban regions. *Pakistan Journal of Health Sciences*, 17(4), pp.42-58.

Gul, S., Mehmood, T.& Nawaz, R., 2021. Supporting sustainable urban planning through air quality forecasting: Case of Karachi. *Sustainable Cities & Society*,14(2), pp.215-20.

Hassan, R., Akbar, A. and Shah, I., 2025. Data quality issues in environmental studies: A study on air pollution and machine learning models. *International Journal of Environmental Monitoring*, 30(1), pp.14-22.

Hu, X., Zhang, H. and Li, S., 2020. Enhancing air quality models with satellite data: A review. *Journal of Atmospheric and Oceanic Technology*, 37(5), pp.697-710.

Hussain, M. and Saeed, A., 2021. Temporal analysis of air pollution: A case study of Lahore using lagged features. *Environmental Modelling & Software*, 45(3), pp.33-40.

Iqbal, F., Zaheer, M. and Ahmad, K., 2022. Satellite data integration with ground-based measurements to improve air quality predictions. *Journal of Remote Sensing*, 10(4), pp.103-110.

Jalil, A. and Sadiq, M., 2022. Predicting health impacts of air pollution using machine learning: Simulations on respiratory diseases. *Environmental Health Perspectives*, 24(3), pp.40-46.

Jiang, X., Zhao, L. and Li, T., 2021. Time-series analysis for smog prediction using machine learning. *Atmospheric Environment*, 243, pp.117-126.

Khan, F. and Ashraf, M., 2025. Transparency in predictive air quality models using Random Forest with SHAP: A policy-oriented approach. *Journal of Environmental Informatics*, 18(2), pp.140-148.

Khan, M., Shabbir, G. and Noor, S., 2023. LSTM-based predictions for smog intensity in Lahore. *Environmental Modelling & Prediction*, 39, pp.72-83.

Kim, H., Lee, K. and Choi, S., 2023. CNN-LSTM hybrid model for real-time air quality forecasting. *Environmental Computing & Sustainable Development*, 22(6), pp.100-12.

Kumar, S., Mishra, P. & Sharma, S., 2023. Performance of XGBoost in air quality prediction: Case study of Delhi. *Environmental Modelling & Software*, 147, pp.105-15.

Le Monde, 2024. Lahore's air quality reaches hazardous levels. *Le Monde*.

Li, Q., Zhang, Z. and Song, Y., 2020. Combining satellite and ground-based data for enhanced air pollution detection. *Environmental Monitoring and Assessment*, 192(9), pp.567-580.

Li, L., Zhang, Y. and Wang, T., 2023. Evaluation of machine learning models for air quality forecasting in Asia. *Environmental Pollution*, 292, pp.1144-1152.

Luo, S., Zhang, L. and Chen, H., 2023. Satellite-based air pollution monitoring: A case study in Southeast Asia. *Journal of Environmental Science & Technology*, 58(6), pp.1545-1555.

Mehmood, S. and Javed, M., 2023. Seasonal patterns in winter smog and prediction using SARIMA models. *Atmospheric Research*, 10(2), pp.91-100.

Reuters, 2024. Lahore tops global pollution indices: *Reuters News*.

Rehman, A. and Ali, K., 2023. Missing data imputation in environmental data using K-Nearest Neighbors: A case study. *Data Science & Applications*, 9(1), pp.50-57.

Rizvi, F. and Saeed, S., 2023. Domain adaptation techniques for generalizing machine learning models in pollution forecasting. *Environmental Data Science*, 6(3), pp.170-5.

Raza, S. and Hameed, A., 2023. Feature engineering in environmental modeling: Impact of socio-environmental variables. *Environmental Modelling*, 12(4), pp.120-126.

Shahid, S., Khan, A. & Javed, M., 2023. Smog in urban centers: Impacts on air quality and health. *Environmental Research*, 213, pp.13-27.

Shah, S. and Nawaz, R., 2023. Incorporating regional emission inventories into machine learning models for pollution prediction: A case study on Islamabad. *Environmental Pollution*, 37(1), pp.55-61.

Shah, M., Khan, F. & Ahmad, S., 2020. Investigating the sources and impacts of particulate pollution in Lahore. *Environmental Science & Technology*, 54(12), pp.7118-34.

Shah, S., Bukhari, A. and Naeem, S., 2020. Vehicular pollution and urbanization in Islamabad and Rawalpindi: An environmental study. *Journal of Urban Planning*, 14(2), pp.77-84.

Shao, G., Huang, Y. and Wang, T., 2021. Meteorological data from NASA's MERRA-2 & ERA5 for air quality modelling. *Atmospheric & Climate Sciences*, 11(2), pp.215-222.

Tariq, M., Ali, S. and Zia, M., 2023. Feature attribution for air quality models using SHAP values and partial dependence plots. *Environmental Informatics*, 13(1), pp.67-75.

Usman, M. and Zaheer, M., 2024. Impact of time-based features on air quality prediction. *Environmental Data Science*, 5(3), pp.92-101.

Wang, Z., Zhang, L. and Zhang, L., 2019. Impact of temperature inversions on air quality in urban areas. *Journal of Environmental Monitoring*, 21(4), pp.118-129.

Wang, Z., Zhang, Z. and Li, M., 2022. Numerical weather prediction integration for improving air quality forecasting. *Journal of Applied Meteorology and Climatology*, 61(6), pp.800-812.

Wang, L., Zhang, Q. and Yang, X., 2022. Winter smog formation in Pakistan: Meteorological and geographical aspects. *Atmospheric Research*, 243, pp.105-118.

Yuan, T., Lin, S. and Chen, J., 2021. Machine learning approaches for air pollution forecasting: A review. *Journal of Environmental Engineering*, 145(6), pp.210-221.

Yuan, M., Song, W. and Li, X., 2022. Performance evaluation of machine learning models in air quality forecasting. *Environmental Pollution*, 278, pp.116-130.

Zhang, H., Li, Z. and Wu, X., 2022. SVM and Decision Tree models for $PM_{2.5}$ forecasting in Beijing. *Atmospheric Pollution Research*, 13(6), pp.908-917.

Zhao, S., Zhou, X. and Wu, J., 2020. Forecasting air pollution using machine learning: A study in Beijing. *Journal of Air Quality*, 13(1), pp.57-71.

Zhao, Y., Wu, C. and Zhang, R., 2023. Machine learning-based air quality prediction models: A comparison study. *Environmental Research Letters*, 17(2), pp.105-117.

Zhou, Q., Lin, H. and Zhang, M., 2022. Satellite-based monitoring of pollution sources in Lahore and Islamabad. *Environmental Science and Technology*, 56(1), pp.88-95.