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

Smog in Pakistan—driven by urbanization, industrialization, vehicular emissions, and crop residue burning—poses critical threats to public health, the environment, and economic sustainability. Major urban centres like Lahore, Karachi, and Islamabad face worsening smog, especially during winter months when temperature inversions trap pollutants such as PM2.5, NOx, SO_2, CO, and O_3 near ground level. This study focuses on predictive modelling as a proactive approach to mitigate smog. Machine learning techniques—regression, time-series forecasting, and classification algorithms—are employed alongside meteorological and emission data to predict smog intensity and duration. Integration of satellite data (MODIS, Sentinel-5P) with ground monitoring enhances model accuracy. Additionally, hybrid models combining AI with numerical weather prediction (e.g., WRF) show promising results. Aligned with SDG 11 (Sustainable Cities) and SDG 3 (Good Health), this research supports real-time air quality monitoring, early warnings, and long-term strategies like urban reforestation and clean fuels. Challenges such as limited monitoring infrastructure and inconsistent data demand greater investment, capacity-building, and cross-sector collaboration for effective mitigation.

Keywords: Smog, Pakistan, Predictive modelling, SGD,

INTRODUCTION

Smog has become an escalating environmental and public health concern in Pakistan, particularly in urban centres such as Lahore, Karachi, and Islamabad. The phenomenon is primarily driven by rapid urbanization, industrial emissions, vehicular exhaust, and agricultural residue burning. These pollutants—including fine particulate matter (PM2.5, PM10), nitrogen oxides (NOx), sulphur dioxide (SO_2), carbon monoxide (CO), and ozone (O_3)—are often trapped near ground level during winter months due to temperature inversions, resulting in prolonged exposure and significant health risks (Shahid, Khan and Javed, 2023).

According to the WHO (2020), air pollution contributes to over 4.2 million premature deaths globally each year. In Pakistan, smog-related respiratory and cardiovascular illnesses continue to strain health systems and deteriorate overall quality of life (Gul, Arif and Imran, 2021; Bashir, Akram and Raza, 2024). Cities such as Lahore frequently top global pollution indices, with Air Quality Index (AQI) readings often reaching hazardous levels (Reuters, 2024; Le Monde, 2024).

Despite increasing awareness, Pakistan still lacks a real-time, accessible smog monitoring and forecasting infrastructure. Existing air quality stations are few, expensive, and static, offering limited geographic coverage and minimal public engagement. Consequently, the general public remains largely uninformed about real-time smog levels, which affects decision-making related to travel, outdoor activity, and health precautions (AP News, 2024).

By employing advanced technologies, the proposed system aims to overcome the limitations of traditional monitoring tools, improving both prediction accuracy and data accessibility. Furthermore, the research aligns with several United Nations Sustainable Development Goals (SDGs), notably SDG 3 (Good Health and Well-being), SDG 11 (Sustainable Cities and Communities), and SDG 13 (Climate Action). Ultimately, this initiative seeks to empower communities, inform policy decisions, and mitigate the adverse impacts of smog in Pakistan's urban landscape.

LITERATURE REVIEW

Smog in Pakistan: A Growing Crisis: Urban centres in Pakistan—particularly Lahore, Karachi, and Islamabad—are experiencing severe air pollution due to vehicular emissions, industrial activities, and seasonal crop residue burning (Gul et al., 2023). Meteorological phenomena such as temperature inversions and stagnant winds worsen these conditions by trapping pollutants near the ground (Ali et al., 2022). Consequently, harmful substances like $PM_{2.5}$, PM_{10}, NO_{mx}, SO_2, CO, and ground-level O_3 accumulate in the atmosphere.

Lahore frequently ranks among the world's most polluted cities, with $PM_{2.5}$ levels often exceeding WHO safety limits (Shah et al., 2020). The issue is compounded by over four million registered vehicles and widespread crop residue burning during harvest seasons (Ali et al., 2020; Shah et al., 2021). Geographical and climatic factors especially in Punjab's winter smog events, resulting in long-term exposure to toxic air (Wang et al., 2022). These developments underscore the need for real-time forecasting & regulatory measures.

The increasing air pollution problem in Pakistan has far-reaching consequences for public health and the environment, as well as for achieving the United Nations Sustainable Development Goals (SDGs). Specifically, SDG 3 (Good Health and Well-being) and SDG 11 (Sustainable Cities and Communities) are directly impacted, as air pollution contributes to respiratory diseases and hinders the creation of healthy, livable urban spaces (UN, 2022). Therefore, addressing air pollution is crucial not only for public health but also for sustainable urban development in Pakistan.

Advances in ML Models for AQI Prediction: Recent global research has demonstrated significant success in ML-based AQI prediction. For instance, XGBoost models, integrated with meteorological data, achieved high prediction accuracy in Delhi (Kumar et al., 2023). In China, a hybrid LSTM-RF model improved temporal accuracy in $PM_{2.5}$ forecasting (Chen et al., 2021). However, applications in Pakistan remain underdeveloped, with most studies relying on simplistic regression models that fail to capture dynamic environmental interactions (Ahmed et al., 2021). This reveals a critical research gap in localized, advanced forecasting approaches.

Hybrid Deep Learning Models: Hybrid models, which combine deep learning and traditional ML methods, are gaining popularity for their improved adaptability and accuracy. Kim et al. (2023) proposed a CNN-LSTM hybrid for real-time air quality prediction, which effectively captured spatial and temporal variations. Reinforcement learning methods have also been applied to optimize AQI alerts and enhance reliability (Chen et al., 2024). These advances demonstrate how hybrid frameworks can outperform conventional models in air pollution forecasting.

Predictive Analytics in Smog Forecasting: ML is also being effectively applied to smog-specific forecasting. In Beijing, SVM and Decision Tree models have improved $PM_{2.5}$ predictions (Zhang et al., 2022), while in Lahore, LSTM models have been used to predict smog intensity with high accuracy (Khan et al., 2023). These models incorporate pollution levels, meteorological conditions, and time-series features to better understand smog formation patterns (Jiang et al., 2021).

Integration of Satellite and Meteorological Data: Meteorological variables—such as wind speed, temperature inversion, and humidity—play a critical role in the dispersion and accumulation of air pollutants. Empirical studies have demonstrated that low wind speeds and the presence of temperature inversion layers, particularly during winter, exacerbate smog conditions by trapping pollutants near the surface (Shao et al., 2021; Wang et al., 2019). Integrating real-time meteorological data into smog prediction models substantially improves forecasting accuracy and responsiveness.

In parallel, satellite-based observations offer a valuable complement, especially in regions like Pakistan where ground-based monitoring infrastructure is sparse. Remote sensing platforms such as MODIS, OMI, and Sentinel-5P provide essential atmospheric indicators including Aerosol Optical Depth (AOD), nitrogen dioxide (NO_2), and surface reflectance. Bilal et al. (2021) highlighted the significance of AOD and NO_2 retrievals in enhancing Air Quality Index (AQI) prediction capabilities. Moreover, synergistic integration of satellite data with meteorological parameters has been shown to yield more robust and generalizable prediction models, as evidenced by Shah et al. (2022) in Punjab, Pakistan, and Luo et al. (2023) across various regions of Southeast Asia.

Emission Data and Pollution Sources: Accurate emission inventories are foundational to forecasting models. In Pakistan, vehicular and industrial emissions, coupled with agricultural burning, are primary pollution sources. Emissions of NO_x, SO_2, VOCs, and $PM_{2.5}$ vary by region and season, affecting air quality outcomes (Shah et al., 2020; Agarwal et al., 2019). Models incorporating source-specific emissions can better simulate pollution trends and smog episodes.

Satellite-Based Monitoring for Smog Detection: Due to limited on-ground infrastructure, satellite-based monitoring is crucial in Pakistan. Satellites like MODIS and Sentinel-5P offer broad spatial coverage and real-time data, enhancing prediction models' reliability (Hu et al., 2020; Liu et al., 2017). Li et al. (2020) found that combining satellite and ground data significantly improved model resolution and accuracy.

The Need for Smog Prediction Models: Early smog warnings allow vulnerable populations to take preventive measures and help decision-makers implement timely interventions (Rahman et al., 2018). Traditional linear models struggle with the complex interplay of emissions, weather, and topography, while ML-based time series and classification models offer a more nuanced understanding (Zhao et al., 2020; Yuan et al., 2021).

Problem Statement

Urban centres in Pakistan, notably Lahore ia experiencing hazardous levels of air pollution, with concentrations of fine particulate matter ($PM_{2.5}$ and PM_{10}) far exceeding international safety thresholds. This escalating environmental crisis has led to a surge in respiratory and cardiovascular illnesses, posing a serious public health concern. The challenge is exacerbated by the absence of a comprehensive and real-time air quality monitoring infrastructure, which hampers authorities' ability to forecast and manage smog events effectively. While meteorological parameters—such as temperature inversions, wind speed, and humidity—are known to influence pollutant dispersion (Ali, Ali & Khan, 2022), limited research has been conducted on how these factors interact with seasonal cycles and urban microclimates in the Pakistani context (Shahid, Khan & Javed, 2023).

Although the global research community has increasingly adopted machine learning models for real-time air quality forecasting (Kumar, Mishra & Sharma, 2023), their application in Pakistan remains limited. Consequently, the country lacks reliable forecasting tools that could alert communities to impending smog events and help mitigate associated health risks—especially during the winter season when smog episodes peak in intensity (Gul, Arif & Imran, 2021).

Furthermore, while prior studies have established the general health impacts of air pollution in Pakistan (Shahid, Khan & Javed, 2023; Gul, Arif & Imran, 2021), significant gaps remain in understanding the dynamic interplay between fine particulate matter ($PM_{2.5}$ and $PM_{1}o$) and health outcomes across different seasons, meteorological conditions, and geographic zones within urban areas. Existing studies have utilized meteorological data to explore smog formation (Ali, Ali & Khan, 2022), yet few have incorporated spatial, temporal, and seasonal variations alongside health data to build comprehensive predictive models. Additionally, the potential of advanced machine learning techniques—such as LSTM and CNN-LSTM hybrids, which have shown promising results internationally (Khan, Shabbir & Noor, 2023; Kim, Lee & Choi, 2023)—remains largely untapped in the Pakistani context for both public health management and air quality policy formulation.

This research proposes to fill this critical gap by developing a robust, hybrid machine learning-based smog prediction system that integrates meteorological, satellite, pollution, and health datasets. Such a system aims not only to enhance prediction accuracy but also to provide actionable insights for timely public health interventions and evidence-based policymaking in urban Pakistan.

Research Questions

1. To find the relationship between $PM_{2.5}$ and PM_{10} respiratory illnesses in urban Pakistan, considering seasonal, meteorological, spatial, and temporal variations?
2. How can hybrid machine learning models be used to develop accurate tools for smog prediction and air quality monitoring in Pakistan?
3. How can the proposed smog prediction system improve public health outcomes and inform air quality management policies in Pakistan?

Research Objectives

1. To assess the relationship between $PM_{2.5}$ and PM_{10} concentrations and the incidence of respiratory illnesses in major urban centers of Pakistan, by analysing health and air quality data across different seasons, weather patterns, locations, and time periods within a 12-month study period.
2. To design, develop, and validate hybrid machine learning models (e.g., LSTM-RF, CNN-LSTM, XGBoost) for accurate smog forecasting and real-time air quality monitoring, using satellite, meteorological, and ground-based datasets, over a six-month model development and testing phase.
3. To evaluate the impact of the proposed smog prediction system on public health outcomes and policy planning, by conducting scenario simulations, stakeholder interviews, and policy analysis, and delivering recommendations for data-driven air quality management within three months after system deployment.

RESEARH METHODOLOGY

The methodology integrates machine learning models with satellite-based monitoring and meteorological data to address Pakistan's air pollution crisis, aligning with the Sustainable Development Goals (SDGs):

1. **SDG 3: Good Health and Well-being –** Predicting hazardous smog levels to improve health outcomes (Rahman et al., 2018).
2. **SDG 11: Sustainable Cities and Communities –** Supporting urban planning and air quality management (Gul et al., 2021).
3. **SDG 13: Climate Action –** Contributing to climate change mitigation by understanding pollution patterns (Li et al., 2020).
4. **SDG 15: Life on Land –** Promoting ecosystem health by mitigating smog's environmental impact (Agarwal et al., 2019).

Selection of Study Areas: Four major urban centres in Pakistan:

1. **Lahore**: Known for high population density, severe winter smog, and significant crop residue burning (Ali et al., 2020).

2. **Karachi:** A coastal city with high industrial emissions and diverse pollution sources (Gul et al., 2021).
3. **Islamabad/Rawalpindi**: Characterized by vehicular pollution and rapid urbanization (Shah et al., 2020).
4. **Faisalabad:** An industrial hub contributing to elevated particulate matter levels (Agarwal et al., 2019).

Data Collection and Sources: The study will utilize diverse datasets:

1. **Air Quality Data (2018–2024)**: Concentrations of PM2.5, PM10, NOx, SOx, CO, O_3, and AQI data will be collected from sources like the Pakistan Environmental Protection Agency (Pak-EPA), IQAir, OpenAQ, and local monitoring stations (Li et al., 2020).
2. **Meteorological Data**: Temperature, humidity, wind speed, rainfall, and solar radiation will be sourced from the Pakistan Meteorological Department & NASA's MERRA-2 and ERA5 datasets (Shao et al., 2021).
3. **Socioeconomic & Human Activity Data**: Traffic volumes, population density, industrial output, and crop-burning data will be sourced from the Pakistan Bureau of Statistics (PBS), Safe Cities Authority, and MODIS satellites (Zhang et al., 2018).
4. **Health Data (Optional)**: Hospital admission data on respiratory and cardiovascular conditions, after ethical approval (WHO, 2021).

Data Pre-processing: It addresses the inherent challenges in Pakistan, such as inconsistent data quality and coverage:

1. **Missing Data**: Techniques like K-Nearest Neighbors (KNN) and multiple imputation methods will fill gaps caused by equipment failures & limited sensor coverage, in rural regions (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 clean the data, ensuring the reliability of the final model (Hassan et al., 2025).
3. **Satellite Data Integration**: Combining satellite imagery with limited ground data from sources like Sentinel-5P and MODIS will improve model performance, with sparse monitoring (Iqbal et al., 2022).

Model Development: It will consider regional pollution dynamics, meteorological conditions, and socio-economic factors:

1. **Regional Emission Patterns**: The study will incorporate regional emission inventories and pollution sources (e.g., crop burning in Punjab, vehicular emissions in Karachi, industrial pollutants in Faisalabad) into machine learning models like Random Forest and XGBoost (Shah & Nawaz, 2023; Abbas et al., 2022).
2. **Weather Influence:** Hybrid models (e.g., ARIMA-LSTM) will be used to capture non-linear interactions between meteorological factors and air pollution (Farooq et al., 2023; Ahmed et al., 2021).

3. **Seasonal Effects:** The cyclical nature of winter smog in Lahore will be modeled using techniques such as SARIMA to account for temporal dependencies (Mehmood & Javed, 2023).

Feature Engineering: It will enhance the model's predictive capability:

1. **Local Factors**: Socio-environmental variables, traffic data, industrial zones, & biomass burning incidents, will be encoded as categorical or numerical features (Zaheer et al., 2024; Raza & Hameed, 2023).
2. **Weather Interactions**: Interaction features (e.g., wind direction × PM2.5 concentration) will capture pollutant dispersion, especially in cities like Lahore (Iqbal et al., 2022).
3. **Time-Based Features**: Features such as time-of-day, weekday/weekend, and seasonal patterns will be crucial for detecting daily pollution fluctuations (Usman et al., 2024).
4. **Lag Features**: Lagged pollutant values will help predict future smog levels, as confirmed by local studies (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 offer higher accuracy, interpretable models such as Random Forests with SHAP explanations will be prioritized to ensure transparency for policymakers (Khan & Ashraf, 2025).
2. **Feature Attribution**: SHAP values and Partial Dependence Plots (PDPs) will quantify the influence of factors like crop burning and traffic congestion on smog levels (Tariq et al., 2023).
3. **Generalizability**: The model's performance will be validated across different regions (e.g., Lahore trained model tested on other cities datasets using metrics like ROC-AUC and RMSE (Rana et al., 2023).
4. **Health and Policy Impact**: Simulations on historical data will help assess the model's ability to provide timely alerts and forecast health outcomes, such as reduced respiratory disease incidents (Jalil & Sadiq, 2022).

Ethical Considerations: Ethical standards will be adhered:

1. **Data Privacy**: All data will be anonymized, and data will be handled with ethical clearance from relevant review boards (Sharma et al., 2022).
2. **Environmental Data Handling**: Satellite imagery will be processed following privacy regulations, and a data minimization approach will be implemented (Sharma et al., 2022).

Expected Outcomes: The study aims to achieve the following:

1. **High-Accuracy Smog Prediction System**: A hybrid machine learning framework tailored to Pakistan's unique environmental conditions.
2. **Public Health Benefits**: Improved early warning systems to reduce the incidence of respiratory diseases (Ali et al., 2023).
3. **Actionable Policy Recommendations:** Insights into pollution trends will guide short-term interventions and long-term regulatory frameworks.

4. **Operational Dashboard**: A real-time web-based tool for monitoring air quality and providing policy updates.

Challenges and Limitations: Challenges include:

1. **Limited Data Coverage**: Fusion of satellite data with ground measurements will mitigate the lack of air quality sensors in rural areas (Hussain et al., 2024).
2. **Model Generalization**: Variability in topography and pollution sources may affect model transferability, which will be addressed using domain adaptation techniques (Rizvi et al., 2023).
3. **Public Awareness**: The effectiveness of the model will depend on public engagement and policy adoption, facilitated through community workshops and localized dashboards.

CONCLUSION

By integrating advanced ML, meteorological forecasting, and satellite-based monitoring, this research aims to create a robust smog prediction system for Pakistan. The methodology ensures high accuracy, real-time usability, and policy-relevant insights, contributing to long-term 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 centres: 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.