**Literature Review on Air Quality Analysis Using Machine Learning-Based Prediction**

Public health, economic production, as well as environmental sustainability, are highly influenced by air quality. Pollutants cause more than 6.6 million early deaths annually across the globe. This is with pollutant exposure linked with; chronic illness, cardiovascular illness, as well as increased death rates. Climatic regimes along with the atmosphere quality is affected by pollutants. Examples of these pollutants are; particulate matter, nitrogen dioxide, carbon monoxide (CO), ozone (O₃), and others. It is necessary to monitor and predict Air Quality Index (AQI). Monitoring as well as forecasting the Air Quality Index (AQI) is essential. Monitoring helps in designing mitigation strategies that can limit risks from pollutant exposure. The classical models forecasting AQI are imprecise because pollutant-meteorological variable relationships are poorly captured in a non-linear manner. Machine learning as well as deep learning (DL) models have made great advancements in forecasting pollutant concentration with high accuracy by harnessing vast data as well as advanced algorithmic structures. The paper explores key areas in research on air pollution that included variability in terms of seasons in AQI, geospatial research, anomaly detection methods, as well as application in prediction models. The paper highlights recent research studies from 2020 onwards that reflect AI-based developments in forecasting air quality as well as policymaking.

Pollution in the air, varies greatly with seasons due to variations in the weather as well as human activities. Variations in seasons in terms of pollution in the air are influenced by; humidity, industrial emissions, temperature, as well as winds. Studies on pollution trends in Asia, North America, as well as in Europe, confirmed that seasons in winters are more likely to have more pollution in terms of AQI because stagnant atmospheric conditions keep pollutants on ground level. Summer seasons have more ozone, though, due to photoreactive chemicals because high temperatures dominate. In China, a condition known as the "Winter Haze Phenomenon" makes pollution from PM₂. ₅ as well as from PM₁₀ more prevalent, which tends to intensify risks in terms of human health (Afifa et al., 2024). In America as well, summers have more pollution from wildfires that impact multiple states in terms of AQI. The seasons in terms of Saharan Dust impact quality in terms of amount in the air over the Atlantic that impacts regions in North America as well as in South America. The awareness regarding these seasons is crucial in order to boost prediction in terms of AQI as well as in terms of pollution measures that are season-based (Jaffe et al., 2020).

Legacy models traditionally utilized in forecasting air quality are statistical in form, employing multiple linear regression, as well as autoregressive integrated moving average (ARIMA). The models fail in capturing non-linear environmental effects as well as pollutant-pollutant relationships. Machine learning models have been more effective in enhancing forecasting accuracy in terms of pollution. Recent studies have extensively utilized a range of ML models, with Random Forest, which is effective in capturing non-linearity with high prediction accuracy, Support Vector Machine, which is utilized in classification of AQI with good generalizability, as well as Long Short-Term Memory networks that are effective in sequential AQI in forecasting time series. Combination models that merge multiple models, e.g., RF with LSTM, have also proved effective in enhancing prediction (Gao et al., 2024). In a study in Europe that compared models in forecasting AQI, LSTM proved more effective in forecasting with a 22% improvement over traditional methods. A study in India proved that a combination with CNN with LSTM enhanced prediction in terms of concentration in PM₂. ₅ by 28% (Duan et al., 2023). Furthermore, recent studies combined deep learning with satellite-based sensing data in enhancing predictions in terms of AQI. The application of Convolutional Neural Networks in satellite images enabled more accurate spatial prediction in terms of air quality. All these advancements have not addressed challenges that have constrained research in terms of computational cost, as well as limitations in data. Real-time predictions are also a requirement (Chauhan et al., 2021). In future, studies need to ensure that they make models more interpretable as this will help in improving the use of these models in policy making when it comes to air quality.

When it comes to understanding the regional trends in pollution, Geospatial analysis has been on the fore front playing a very key role. Policy makers have found it easy to analyze and pin point areas with high and critical levels of pollution, as well as come up with regional mitigation strategies. This has been facilitated by the use of; Geographic Information Systems, remote sensing from satellites, and spatial interpolation methods (Kamel Boulos & Wilson, 2023). Research that has utilized GIS mapping and satellite-based information has assessed AQI differences in urban and rural settings. Results have shown that urban settings having much greater levels of pollution due to emissions from industrial processes as well as motor vehicle emissions. Additionally, in today’s time, anomaly detection methods are being used increasingly in detecting sharp rise in air pollution (Mahmud et al., 2023). Machine learning models that, make use of anomaly detection methods such as; Isolation Forest, DBSCAN, and Autoencoders have been effective in detecting outliers in pollution. DBSCAN system has been used in detecting any anomalies in AQI data. It has shown to be very useful in detecting high levels of outliers in pollution in real time and therefore helping policy makers to take action in good time. By combining both the geospatial data with anomaly detection methods, has greatly helped in making use of air quality monitoring networks. These networks can be used in detecting the trends in pollution in real time helping in swift decision making (Mahmood Almanor & Miklós Telek, 2023).

Integrating machine learning models in air quality forecasting has important policy relevance. AI-based forecasting models can be used by policymakers to reduce; emission control, issue public health advisories, and formulate strategies for reducing pollution according to regional areas. Pollution forecasting models with Machine Learning, have influenced traffic control measures in Beijing. This has in turn lowered AQI levels by 15% within six months (Yan et al., 2021). Likewise, in India, AI-based air monitoring networks have been introduced in major urban cities. This introduction, provide real-time decision-making in controlling pollution (Rautela & Goyal, 2024). There are problems that exist, however, and include; data homogenization, interpretability, and constraints in computational resources. Future studies can investigate using explainable AI (XAI) methods to provide transparency in ML-based predictions for AQI. Additionally, government, research, and AI development groups have a very important role in working in collaboration with each other. This is in order to create improved air quality policymaking as well as in implementing machine learning models in environmental sustainability.

**References**

Afifa, Arshad, K., Hussain, N., Ashraf, M. H., & Saleem, M. Z. (2024). Air pollution and climate change as grand challenges to sustainability. *Science of the Total Environment*, *928*, 172370. https://doi.org/10.1016/j.scitotenv.2024.172370

Chauhan, R., Kaur, H., & Alankar, B. (2021). Air Quality Forecast using Convolutional Neural Network for Sustainable Development in Urban Environments. *Sustainable Cities and Society*, *75*, 103239. https://doi.org/10.1016/j.scs.2021.103239

Duan, J., Gong, Y., Luo, J., & Zhao, Z. (2023). Air-quality prediction based on the ARIMA-CNN-LSTM combination model optimized by dung beetle optimizer. *Scientific Reports*, *13*(1). https://doi.org/10.1038/s41598-023-36620-4

Gao, W., Xiao, T., Zou, L., Li, H., & Gu, S. (2024). Analysis and Prediction of Atmospheric Environmental Quality Based on the Autoregressive Integrated Moving Average Model (ARIMA Model) in Hunan Province, China. *Sustainability*, *16*(19), 8471–8471. https://doi.org/10.3390/su16198471

Jaffe, D. A., O’Neill, S. M., Larkin, N. K., Holder, A. L., Peterson, D. L., Halofsky, J. E., & Rappold, A. G. (2020). Wildfire and prescribed burning impacts on air quality in the United States. *Journal of the Air & Waste Management Association*, *70*(6). https://doi.org/10.1080/10962247.2020.1749731

Kamel Boulos, M. N., & Wilson, J. P. (2023). Geospatial techniques for monitoring and mitigating climate change and its effects on human health. *International Journal of Health Geographics*, *22*(1). https://doi.org/10.1186/s12942-023-00324-9

Mahmood Almansoori, & Miklós Telek. (2023). *Anomaly Detection using combination of Autoencoder and Isolation Forest*. 25–30. https://doi.org/10.3311/wins2023-005

Mahmud, K., Mitra, B., Mohammed Sakib Uddin, Al-Ekram Elahee Hridoy, Aina, Y. A., Ismaila Rimi Abubakar, Syed Masiur Rahman, Mou Leong Tan, & Muhammad Muhitur Rahman. (2023). Temporal assessment of air quality in major cities in Nigeria using satellite data. *Atmospheric Environment: X*, 100227–100227. https://doi.org/10.1016/j.aeaoa.2023.100227

Rautela, K. S., & Goyal, M. K. (2024). Transforming air pollution management in India with AI and machine learning technologies. *Scientific Reports*, *14*(1). https://doi.org/10.1038/s41598-024-71269-7

Yan, R., Liao, J., Yang, J., Sun, W., Nong, M., & Li, F. (2021). Multi-hour and multi-site air quality index forecasting in Beijing using CNN, LSTM, CNN-LSTM, and spatiotemporal clustering. *Expert Systems with Applications*, *169*, 114513. https://doi.org/10.1016/j.eswa.2020.114513