IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 5, Issue 10, June 2025
Impact Factor: 7.67

# Spatiotemporal Patterns in Air Pollution: A Hybrid Machine Learning for Composite AQI Prediction

**Sreerama Samartha JG, Dayananda GK, Vayusutha M**
Faculty of Electronics & Communication Engineering
Canara Engineering College, Benjanapadavu Manglore, KA, India
jgssb85@gmail.com, dayanand.gk@gmail.com, vayu.sutha@gmail.com

**Abstract**: *The research proposes a hybrid machine learning model to predict a composite Air Quality Index (AQI) based on weather variables, pollutant concentrations, human mobility measures, and geospatial clusters. The models, trained using XGBoost, LightGBM, and CatBoost, showed excellent predictive performance, with CatBoost showing the lowest MSE and MAE. The study highlights the importance of temporal cross-validation in avoiding overfitting to time-series and CatBoost's strength in air quality prediction. This approach integrates interpretable machine learning into environmental policy-making, providing actionable recommendations for pollution reduction. Results exhibited excellent predictive performance for all models: CatBoost recorded lowest MSE (4.73) and MAE (1.68), implying highest stability, whereas XGBoost produced maximum $R^2$ (0.967), which depicts outstanding explanatory power. LightGBM lagged behind marginally ($R^2$: 0.928), implying compromise between speed and precision. SHAP analysis indicated pollutant concentrations (PM2.5, O3), geospatial cluster labels, and the interaction factor Population_Not_Staying_at_Home × mil_miles were key drivers of AQI variation, with wind speed variance and humidity playing an important role. The research illustrates the importance of temporal cross-validation in avoiding overfitting to time-series and highlights CatBoost's strength in air quality prediction. These results move the field forward by integrating interpretable machine learning into environmental policy-making, providing actionable recommendations for reducing hotspots of pollution with spatially focused interventions. The adaptability of the framework to multi-pollutant AQI systems makes it a scalable tool for urban air quality management*

**Keywords**: Air Quality Index (AQI); spatiotemporal patterns; ensemble machine learning; SHAP analysis; temporal cross-validation; geospatial clustering