International Journal of Applied Geospatial Research
Volume 15 • Issue 1 • January-December 2024

Hybrid Model for Benzene Prediction in Kuwait's Industrial Regions

Eiman Tamah Al-Shammari
https://orcid.org/0000-0002-0382-8319
College of Life Sciences, Kuwait University, Kuwait

ABSTRACT

This study introduces a novel hybrid model to enhance the prediction of benzene concentrations in three industrial regions in Kuwait, utilizing air quality data from 2022 to 2024. The hybrid model, developed through stacking techniques, integrates multiple ML algorithms to employ their collective strengths. The initial analysis involved examining pollutant trends and correlations among benzene, toluene, ethylbenzene, and xylenes (BTEX) compounds. We applied more than ten individual machine learning models to predict benzene levels. We then applied a hyperparameter, tuning the hybrid model to further enhance its prediction performance. By combining these models, the hybrid approach demonstrated superior predictive performance, evaluated using R-squared and mean squared error metrics. The results underscore the effectiveness of the hybrid model in providing accurate benzene concentration prediction, offering valuable insights for air quality management and pollution control in industrial regions.

KEYWORDS

Air Quality, Benzene Prediction, BTEX, Correlation, Environmental Technology, Gradient Boosting, Kuwait, LightGBM, Linear Regression, Machine Learning, Pollution, Random Forest, SVM, XGBoost

A HYBRID MODEL FOR BENZENE PREDICTION IN KUWAIT’S INDUSTRIAL REGIONS

Air pollution poses a significant threat to public health, particularly in urban and industrialized regions. Among the critical pollutants is benzene, a volatile organic compound (VOC) found in vehicle emissions, petroleum products, and industrial processes. Recognized as a Group I carcinogen by the International Agency for Research on Cancer, benzene is associated with serious health risks, including leukemia and other blood disorders. Its presence in the atmosphere necessitates effective monitoring and control strategies to mitigate exposure risks.

Kuwait’s rapid urbanization and industrial expansion have led to elevated benzene levels, especially in areas near major refineries and petrochemical plants. This situation underscores the need for predictive models to accurately forecast pollutant levels, facilitating timely interventions to mitigate public health impacts.

Traditional air quality models, such as Gaussian dispersion models, rely heavily on meteorological data and simplifying assumptions, which limits their effectiveness in environments with complex emission patterns. In contrast, ML models can use historical data to capture nonlinear relationships between pollutants and their sources more effectively. Individual ML models may struggle with generalization, however, particularly when pollutant emissions are variable or episodic. To address

DOI: 10.4018/IJAGR.362003

these challenges, this study developed a hybrid model using stacking techniques and integrating multiple algorithms to enhance prediction accuracy.

This study focused on three industrial monitoring stations in Kuwait—East Burgan, Al-Ahmadi, and North Pier—spanning January 2022 to June 2024. Despite the absence of meteorological data, the analysis of benzene, toluene, ethylbenzene, and xylenes (BTEX) compounds provided insights into pollutant dynamics, supporting regulatory decision-making. By employing various ML models, including gradient boosting, LightGBM, and random forest (RF), and integrating them through a hybrid stacking approach, this study aimed to improve prediction accuracy and provide actionable insights for emission control.

The findings highlight the importance of expanding air quality research across multiple sites for a comprehensive view of pollutant dynamics in Kuwait's industrial zones and underscore the value of hybrid models in air quality management, particularly in regions with diverse industrial activities.

This paper is organized as follows: The next section provides a literature review, highlighting previous research on air quality, with a focus on the health risks of benzene and other pollutants, international regulations, and related studies in Kuwait. The data description section outlines data collected from the monitoring stations, including specifics regarding the dataset and the pollutants measured. The methodology section details data preprocessing steps, the ML models employed to predict benzene concentrations, and the development of a hybrid model using stacking techniques. The results and discussion section presents the evaluation of the models, highlighting the performance of individual models and the hybrid approach. The paper concludes by summarizing key findings and offering directions for future research to enhance air quality monitoring and prediction.

LITERATURE REVIEW

Benzene, a volatile organic compound (VOC), is a significant environmental and public health concern, classified as a Group I carcinogen by the International Agency for Research on Cancer (IARC). Anthropogenic activities, such as vehicular exhaust, petroleum refining, and industrial processes, primarily emit it, with secondary contributions from natural phenomena and indoor sources (Agency for Toxic Substances and Disease Registry, 2007; Notario, Gutiérrez-Álvarez, & Adame, 2020). Benzene concentrations are notably high in industrialized and urbanized regions, particularly where refinery operations dominate, as in Kuwait. These emissions pose a dual challenge: safeguarding public health and mitigating environmental impacts.

The health risks associated with benzene are well documented. Spycher et al. (2017) demonstrated that maternal occupational exposure to benzene significantly increases the risk of childhood leukemia, particularly acute lymphoid leukemia. Other studies link benzene exposure to acute myeloid leukemia and adverse respiratory and cardiovascular outcomes, with children, pregnant women, and the elderly being particularly vulnerable (Karimi & Shokrinezhad, 2020; Giovannini et al., 2010). Indoor benzene exposure, exacerbated by limited ventilation and high temperatures in arid climates, can surpass outdoor levels, further amplifying health risks (Liu, Huang, & Li, 2020). These findings emphasize the need for accurate monitoring and predictive frameworks to inform regulatory interventions and protect at-risk populations.

Localized studies in Kuwait highlight significant variability in benzene concentrations, driven by industrial activities, vehicular emissions, and seasonal meteorological conditions. Al-Hurban, Khader, Alsaber, and Pan (2021) analyzed air quality trends from 2012 to 2017, identifying oil refineries and petrochemical facilities as primary contributors to VOC levels while highlighting the interplay between industrial sources and environmental factors like dust storms. Matar, Alsaber, and Al-Hurban (2024) reinforced these findings by revealing spatial and temporal disparities in pollutant concentrations across industrial zones, underscoring the need for geographically targeted strategies to address industrial hotspots effectively. Both studies demonstrate the critical role of continuous monitoring and adaptive interventions in managing Kuwait's air quality challenges.

Related Content