

XGboost-Powered Smart Air Quality Evaluation and Prediction System

M.Nagulan, R.Thamira

Assistant Professor, Sir Issac Newton College of Engineering & Technology, Nagapattinam, Tamil Nadu, India

PG Scholar, Sir Issac Newton College of Engineering & Technology, Nagapattinam, Tamil Nadu, India

ABSTRACT: Air pollution is a major environmental and health concern, particularly in rapidly growing urban areas, as it contains harmful substances like particulate matter and gases (NO₂, CO, O₃, SO₂) that affect vulnerable populations. Traditional prediction methods often rely on limited historical data and simple regression models, which struggle to capture complex spatial and temporal relationships and can be computationally intensive. To overcome these limitations, this project proposes a smart prediction system using XGBoost combined with a deep distributed fusion network and a spatial transformation module. This system converts scattered air quality measurements into a consistent format, models the interactions between pollution sources, and predicts the Air Quality Index (AQI) accurately. By considering both spatial correlations and detailed environmental features, the proposed framework provides more reliable predictions and can guide authorities and policymakers in controlling pollution and protecting public health.

I. INTRODUCTION

Air pollution has become a critical environmental challenge in the 21st century, driven by rapid industrialization, urbanization, and population growth. Harmful pollutants, including nitrogen dioxide (NO₂), sulfur dioxide (SO₂), carbon monoxide (CO), ozone (O₃), and particulate matter (PM_{2.5} and PM₁₀), negatively affect human health, contribute to climate change, and degrade environmental quality. Monitoring and accurately predicting air quality are essential for protecting public health, guiding urban planning, and enforcing environmental regulations. Traditional monitoring systems rely on fixed sensors and statistical models to collect data from limited areas. While these provide real-time measurements, they often fail to capture spatial variations and complex, nonlinear interactions among environmental factors. Conventional machine learning approaches, such as linear regression or support vector regression (SVR), also struggle with high-dimensional, incomplete, or irregular datasets, resulting in lower prediction accuracy in dynamic urban settings. To address these limitations, this project proposes an advanced air quality prediction system using XGBoost (Extreme Gradient Boosting) combined with spatial data fusion techniques.

Objectives of this work include:

The main objective of the XGBoost-powered smart air quality evaluation and prediction system is to design and implement an intelligent, accurate, and scalable framework for monitoring and forecasting air pollution levels. The project focuses on acquiring reliable environmental and meteorological datasets, performing effective data preprocessing to handle missing and noisy values, and applying the XGBoost algorithm to model complex nonlinear relationships among various air pollutants.

II. RELATED WORK

2.1 APPLICATION OF THE LOW-COST SENSING TECHNOLOGY FOR INDOOR AIR QUALITY MONITORING

Juliana P. Sá et al. [1] focuses on the growing adoption of low-cost sensors (LCS) for indoor air quality (IAQ) monitoring. The authors explore the various sensing technologies available such as electrochemical, NDIR, metal-oxide, and optical particle counters and compare their sensitivity, selectivity, response time, and stability for key indoor pollutants. Calibration methods including laboratory, field co-location, and algorithmic correction approaches are examined in detail. Practical challenges like power requirements, placement, maintenance, and communication constraints are addressed, along with discussions on how environmental factors like humidity and temperature influence sensor accuracy.

The monitoring of indoor air quality (IAQ) using low-cost sensors (LCS) has become increasingly important due to rising concerns about air pollution and its health impacts. Low-cost sensors provide an affordable and scalable alternative to traditional, high-precision instruments, allowing continuous monitoring in homes, offices, and public buildings.

2.2 ESTABLISHING SUSTAINABLE LOW-COST AIR QUALITY MONITORING SETUP

Mannam Veera Narayana, Devendra Jalihal, and S.M.Shiva[2] presents a comprehensive overview of sustainable low-cost air quality monitoring setups from both technical and economic perspectives. It reviews modern hardware architectures, including sensors, microcontrollers, and communication modules, as well as software components such as firmware and cloud-based visualization tools. The authors evaluate sustainability factors like energy efficiency, maintainability, and environmental footprint. Trade-offs between precision, affordability, and scalability are discussed through numerous community-driven deployment examples.

Designing low-cost air quality monitoring systems that are both effective and sustainable requires a careful balance between technical performance, affordability, and environmental responsibility.

2.3 PERFORMANCE EVALUATION OF LOW-COST AIR QUALITY SENSORS

Ye Kang and colleagues [3] provide an extensive review of performance evaluation studies of low-cost air quality sensors. The paper systematically compares LCS data with that of reference-grade instruments, focusing on pollutants such as PM, CO₂, and NO₂. Key metrics such as sensitivity, precision, detection limits, linearity, and cross-sensitivity are discussed. The authors explain how temperature and humidity fluctuations introduce measurement bias and present correction techniques using machine learning and statistical algorithms. The review stresses the importance of calibration transferability and the necessity for multi-season and multi-site validation campaigns. The authors reveal that while many low-cost sensors can reliably capture temporal trends, they may fall short for regulatory monitoring. The review also critiques inconsistencies between vendor specifications and independent evaluation results.

Low-cost air quality sensors (LCS) are widely used for monitoring pollutants like PM, CO₂, and NO₂, but their accuracy can vary. Their performance is measured in terms of sensitivity, precision, detection limits, and response to environmental factors such as temperature and humidity. Calibration—either through reference instruments or algorithmic correction—is essential to improve reliability. While LCS can effectively track trends over time, they may not meet the high accuracy standards required for regulatory monitoring, and performance can differ from manufacturer specifications.

2.4 INDOOR AIR QUALITY AND HEALTH IN SCHOOLS A CRITICAL REVIEW FOR DEVELOPING THE ROADMAP FOR THE FUTURE SCHOOL ENVIRONMENT

Sasan Sadrizadeh et al. [4] explores the intricate relationship between indoor air quality (IAQ) and the health, comfort, and learning outcomes of students in schools. The authors analyze pollutants such as CO₂, PM_{2.5}, VOCs, NO₂, and biological contaminants, emphasizing the importance of proper ventilation and reduced occupant density. Numerous studies are reviewed that link poor IAQ with absenteeism, respiratory ailments, and diminished cognitive performance. Methodological challenges, including inconsistent exposure characterization and confounding factors, are discussed in depth. The paper outlines interventions like improved ventilation systems, air filtration, and behavioral strategies to mitigate pollutant exposure. Cost-effective and scalable IAQ monitoring solutions.

Indoor air quality (IAQ) greatly affects students' health and learning. Pollutants like CO₂, PM_{2.5}, VOCs, NO₂, and microbes can cause respiratory issues, absenteeism, and reduced cognitive performance. Ensuring proper ventilation, reducing classroom crowding, and using air filtration can improve IAQ. Affordable monitoring systems help schools track air quality and take timely actions to protect student health.

2.5 FEASIBILITY AND AFFORDABILITY OF LOW-COST AIR SENSORS WITH INTERNET OF THINGS FOR INDOOR AIR QUALITY MONITORING IN RESIDENTIAL BUILDINGS

Yong Yu and co-authors [5] present the practicality and cost-effectiveness of IoT-enabled low-cost sensors for residential indoor air quality monitoring. The authors consolidate findings on sensor specifications, communication technologies, and home deployment experiences. They compare performance and affordability across multiple pollutants, including PM, CO₂, and VOCs. The paper evaluates factors such as ease of installation, mobile connectivity,

and user- friendliness, which influence consumer adoption. IoT integration is discussed as a means of enabling automated ventilation and air purification systems. The authors differentiate between DIY and commercial devices in terms of reliability and technical support. Privacy concerns and data-sharing risks associated with home IoT networks are also addressed. The study presents affordability analyses, considering total ownership cost and maintenance. It highlights real-world lessons from pilot deployments regarding placement, drift, and occupant behavior effects.

IoT-enabled low-cost sensors provide an affordable and practical way to monitor indoor air quality in homes. They track pollutants like PM, CO₂, and VOCs, and can connect to mobile apps or cloud systems for real-time updates. Ease of installation, connectivity, and automation with ventilation systems improve usability. Performance depends on sensor placement, calibration, and user behavior, while privacy and data security need careful consideration.

2.6 UPDATE OF THE WHO GLOBAL AIR QUALITY GUIDELINES

Roman Perez Velasco and Dorota Jarosinska [6] introduces the updated World Health Organization (WHO) global air quality guidelines and the systematic review process behind them. The authors describe the motivation for revising the guidelines, given the increasing global burden of disease linked to air pollution exposure. The paper outlines how multidisciplinary evidence was gathered, appraised, and synthesized across pollutant categories such as PM_{2.5}, PM₁₀, NO₂, SO₂, CO, and O₃. It explains the methodological rigor involved in ensuring that epidemiological and toxicological evidence met WHO's strict inclusion criteria. The study emphasizes the integration of new scientific findings on chronic.

The WHO updated its global air quality guidelines to reflect new evidence on the health effects of pollutants such as PM_{2.5}, PM₁₀, NO₂, SO₂, CO, and O₃. The update is based on a systematic review of recent studies, emphasizing long-term exposure risks and providing stricter recommendations to protect public health worldwide.

2.7 A TECHNICAL OVERVIEW ON BETA-ATTENUATION METHOD FOR THE MONITORING OF PARTICULATE MATTER IN AMBIENT AIR

Kritika Shukla and Shankar G. Aggarwal [7] explains the principles, instrumentation, and applications of the beta-attenuation method (BAM) used to monitor particulate matter (PM) in ambient air. The authors describe the operating mechanism, where beta radiation is used to measure mass concentration based on attenuation as particles collect on a filter tape. They outline calibration procedures, operational parameters, and quality control requirements necessary to ensure reliable PM measurement. The paper compares BAM with alternative measurement techniques such as gravimetric and optical methods, emphasizing its real-time capabilities and regulatory acceptance. It also explores advancements in sensor miniaturization, data logging, and automatic filter tape movement systems that enhance BAM performance.

Kritika Shukla and Shankar G. Aggarwal describe the beta-attenuation method (BAM) as a reliable technique for measuring particulate matter in ambient air. It works by passing beta radiation through a filter where particles are collected; the reduction in radiation intensity is used to estimate particle mass concentration. The method requires proper calibration and quality control to maintain accuracy. Compared to other techniques, BAM offers continuous, real-time monitoring and is widely used in air quality assessment.

2.8 INDOOR AIR QUALITY A REVIEW OF CLEANING TECHNOLOGIES

Teresa M. Mata and colleagues present [8] a detailed review of available technologies designed to clean and improve indoor air quality. The study categorizes air cleaning methods into mechanical filtration, electrostatic precipitation, photocatalytic oxidation, activated carbon absorption, and emerging plasma-based technologies. Each method is evaluated in terms of pollutant removal efficiency, maintenance requirements, energy consumption, and secondary emission risks. The review emphasizes that mechanical filters remain the most effective and widely adopted for particulate removal, while adsorption and catalytic techniques target gaseous pollutants. The authors discuss how hybrid systems integrating multiple cleaning principles can enhance performance for complex pollutant mixtures. The paper also addresses the importance of proper sizing, filter replacement intervals, and the influence of airflow patterns on overall effectiveness. Teresa M. Mata and co-authors review different technologies used to improve indoor air quality, including filtration, electrostatic, adsorption, and catalytic methods. Mechanical filters are most effective for removing particles, while other techniques help eliminate gaseous pollutants. The study also highlights that combining multiple methods and maintaining proper airflow and system upkeep improves overall performance.

2.9 AIR QUALITY AND MANAGEMENT IN PETROLEUM REFINING INDUSTRY

Festus M. Adebisi's [9] focuses on the critical issue of air pollution within the petroleum refining industry and evaluates existing management strategies. The paper discusses the primary air pollutants emitted from refining operations, including sulfur oxides (SO_x), nitrogen oxides (NO_x), volatile organic compounds (VOCs), and particulate matter. It highlights the health and environmental impacts of these emissions, particularly in regions with high refinery density. The author examines regulatory frameworks, pollution control technologies, and emission monitoring systems currently implemented in refineries. Techniques such as flue gas desulfurization, catalytic converters, and vapor recovery units are analyzed for their effectiveness. The review also considers advancements in process optimization. Festus M. Adebisi examines air pollution in petroleum refineries and the strategies used to manage it. The study identifies major pollutants, including SO_x, NO_x, VOCs, and particulate matter, and highlights their environmental and health impacts. Current management approaches such as flue gas desulfurization, catalytic converters, vapor recovery units, and process optimization are discussed for their effectiveness in controlling emissions and complying with regulations.

2.10 INTEGRATING SMART TRAFFIC SYSTEMS WITH REAL-TIME AIR QUALITY MONITORING TO MINIMIZE EMISSIONS AND IMPROVE URBAN HEALTH

Lyazat Naizabayeva, Daniyar Zaitov, and Nurgul Seilova [10] explored the integration of smart traffic management systems with real-time air quality monitoring as a strategy to reduce urban pollution. The authors propose an IoT-enabled framework that collects air quality and traffic data from distributed sensors, analyzes it through cloud platforms, and dynamically adjusts traffic signals to control congestion and emissions. The system aims to identify pollution hotspots and implement responsive traffic rerouting to mitigate pollutant buildup. The paper discusses the role of artificial intelligence algorithms in predicting emission trends and optimizing vehicle flow in real time. It also outlines data transmission protocols, system architecture, and urban-scale implementation considerations. Case studies. Lyazat Naizabayeva, Daniyar Zaitov, and Nurgul Seilova present a system that combines smart traffic control with real-time air quality monitoring. The project uses IoT sensors to gather traffic and pollution data, which AI algorithms analyze to predict emission patterns. Traffic signals are dynamically adjusted to reduce congestion and limit pollutant buildup, helping cities manage air quality more effectively.

III. EXISTING SYSTEM

Many recent studies have applied **XGBoost**, a powerful machine-learning model, to forecast air quality metrics such as AQI and pollutant concentrations. In these systems, historical pollution data along with weather and environmental variables are collected and preprocessed to feed into XGBoost models, which learn complex, nonlinear relationships for more accurate prediction. For example, hybrid frameworks combine XGBoost with deep learning or ensemble methods to improve forecasting accuracy of PM_{2.5}, PM₁₀, NO₂, and other pollutants by capturing patterns that traditional approaches often miss. These systems demonstrate that XGBoost can reliably predict short-term air quality trends and support real-time assessment, making it suitable for smart air quality monitoring platforms that inform public health responses and environmental planning. Several existing systems have successfully applied **XGBoost** for air quality evaluation and prediction. Urban monitoring projects use XGBoost to forecast daily pollutant levels, such as PM_{2.5} and ozone, by combining historical pollution data, weather conditions, and traffic patterns. Hybrid platforms enhance accuracy by integrating XGBoost with other machine learning models, enabling the system to capture complex environmental patterns. Multi-station monitoring systems train XGBoost on data from multiple locations to account for spatial variations in pollution, supporting regional air quality management. Additionally, some real-time evaluation tools use XGBoost to analyze live sensor data.

3.1 Drawbacks

- Existing models often cannot capture complex spatial and temporal patterns of pollutants.
- Limited or inconsistent historical data reduces prediction accuracy.
- Traditional algorithms require significant computational resources for large datasets.
- Many systems struggle with missing or noisy sensor data.
- Lack of real-time prediction and adaptability limits practical usefulness.
- Accuracy is low for nonlinear or complex pollution trends.
- Sudden pollution spikes are often not predicted effectively.

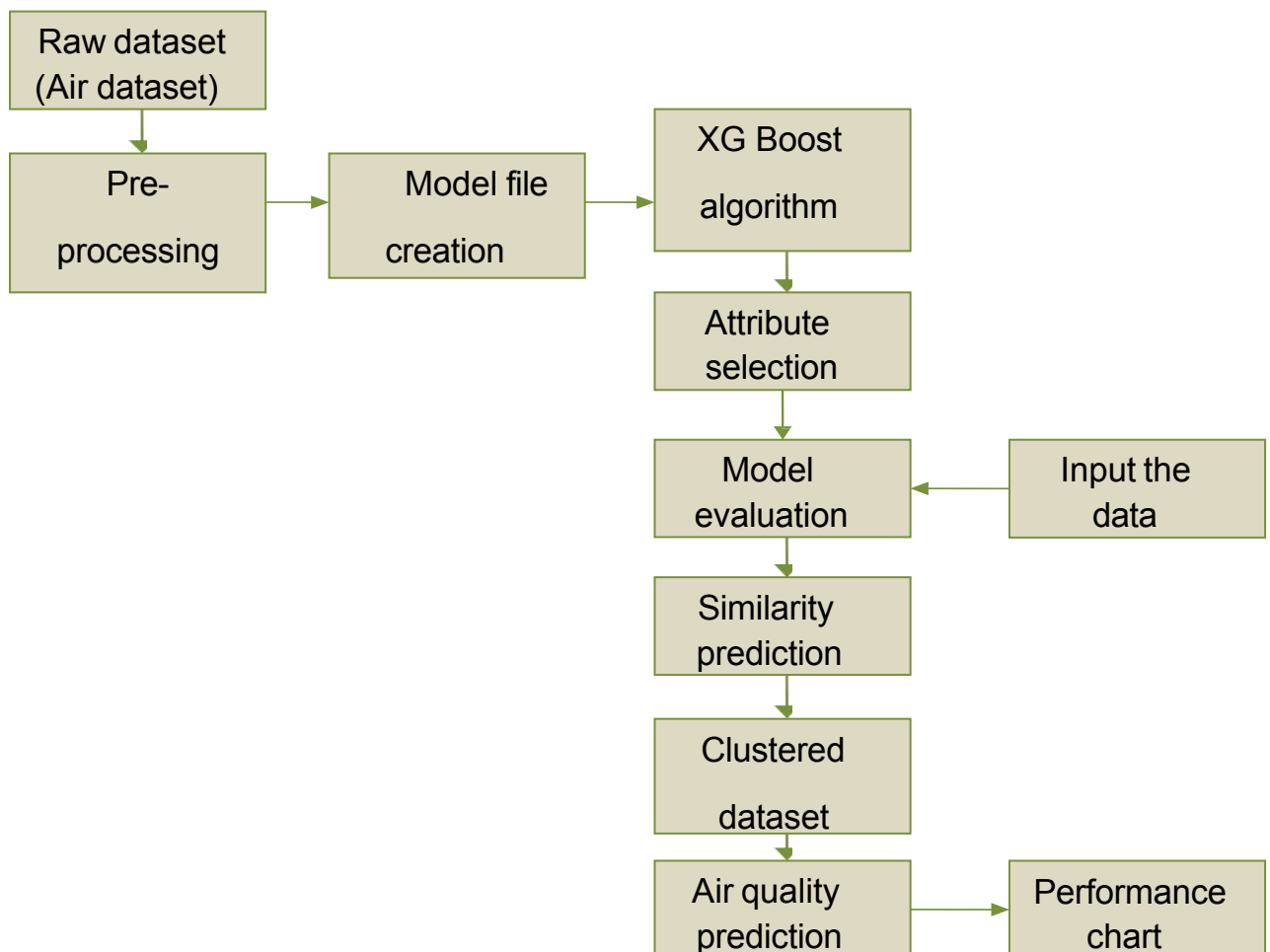
IV. PROPOSED WORK

The proposed system integrates real-time air quality monitoring with machine learning prediction using XGBoost to provide accurate, short-term forecasts of pollutant levels. Air quality sensors measure key pollutants such as PM2.5, PM10, NO₂, and O₃, while meteorological data like temperature, humidity, and wind speed are also collected. These inputs are preprocessed and fed into an XGBoost model trained to learn complex relationships between environmental factors and pollutant concentrations. The system generates continuous real-time AQI estimates, predicts pollution trends, and identifies potential hotspots. Additionally, it can support smart urban interventions, such as traffic management or public alerts, to reduce exposure and improve overall air quality. The framework also allows for scalability, enabling integration of new sensors or locations to expand coverage. The proposed system uses XGBoost to predict air quality by analyzing real-time and historical data. It identifies pollution hotspots, supports preventive actions, and can integrate with urban management tools for adaptive, real-time environmental monitoring.

4.1 Advantages:

- The system delivers accurate air quality predictions by using XGBoost to analyze both spatial and temporal data.
- It can process large and complex datasets efficiently, reducing computation and training time.
- The model handles missing or inconsistent data, ensuring dependable forecasting.
- Real-time prediction allows authorities to take timely actions to control pollution.
- The system is flexible and scalable, suitable for different cities and environmental conditions.

V. BLOCK DIAGRAM



VI. CONCLUSION

Air pollution significantly affects human health and the environment, and its impacts vary depending on the sources of pollutants. This project highlights how data mining and machine learning, particularly XGBoost, can be used to analyze environmental data and predict air quality. By considering factors like temperature, humidity, and dew point, the system can generate accurate forecasts and a reliable Air Quality Index (AQI). Using models such as linear regression and the normal equation improves prediction efficiency, while the system's performance benefits from larger, high-quality datasets. Overall, the proposed framework offers practical support for policy makers, urban planners, and the public in making informed decisions to mitigate air pollution and protect health.

VII. FUTURE WORK

Future improvements for the air quality prediction system can focus on boosting prediction accuracy and expanding the types of data it uses. Combining XGBoost with advanced models like LSTM or Graph Neural Networks can better capture spatial and temporal patterns in pollution. Adding more data sources such as satellite images, traffic density, industrial emissions, and mobile IoT sensors can enhance both coverage and reliability. The system could also incorporate adaptive learning, updating itself as new environmental trends appear. In the long term, this framework could become a fully automated, real-time decision support platform, integrated with smart city infrastructure to help manage air quality and protect urban health.

REFERENCES

- [1] Sá, Juliana P., et al. "Application of the low-cost sensing technology for indoor air quality monitoring: A review." *Environmental Technology & Innovation* 28 (2022): 102551.
- [2] Narayana, Mannam Veera, Devendra Jalihal, and SM Shiva Nagendra. "Establishing a sustainable low-cost air quality monitoring setup: A survey of the state-of-the-art." *Sensors* 22.1 (2022): 394.
- [3] Kang, Ye, et al. "Performance evaluation of low-cost air quality sensors: A review." *Science of The Total Environment* 818 (2022): 151769.
- [4] Sadrizadeh, Sasan, et al. "Indoor air quality and health in schools: A critical review for developing the roadmap for the future school environment." *Journal of Building Engineering* 57 (2022): 104908.
- [5] Yu, Yong, et al. "Feasibility and affordability of low-cost air sensors with internet of things for indoor air quality monitoring in residential buildings: Systematic review on sensor information and residential applications, with experience-based discussions." *Atmosphere* 15.10 (2024): 1170.
- [6] Velasco, Román Pérez, and Dorota Jarošnińska. "Update of the WHO global air quality guidelines: Systematic reviews—An introduction." *Environment International* 170 (2022): 107556.
- [7] Shukla, Kritika, and Shankar G. Aggarwal. "A technical overview on beta-attenuation method for the monitoring of particulate matter in ambient air." *Aerosol and Air Quality Research* 22.12 (2022): 220195.
- [8] Mata, Teresa M., et al. "Indoor air quality: a review of cleaning technologies." *Environments* 9.9 (2022): 118.
- [9] Adebiyi, Festus M. "Air quality and management in petroleum refining industry: A review." *Environmental Chemistry and Ecotoxicology* 4 (2022): 89-96.
- [10] Naizabayeva, Lyazat, Daniyar Zaitov, and Nurgul Seilova. "Integrating Smart Traffic Systems with Real-Time Air Quality Monitoring to Minimize Emissions and Improve Urban Health." *Procedia Computer Science* 251 (2024): 603-608.



INTERNATIONAL
STANDARD
SERIAL
NUMBER
INDIA



INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN SCIENCE | ENGINEERING | TECHNOLOGY

 9940 572 462  6381 907 438  ijirset@gmail.com



www.ijirset.com

Scan to save the contact details