BREATHESAFE – SMART INDOOR AIR POLLUTION TRACKER

APURAV SINGH

MASTER OF COMPUTER APPLICATION(MCA)

AJAY KUMAR GARG ENGINEERING COLLEGE, GHAZIABAD

apurav2314035@akgec.ac.in

AJEET YADAV

MASTER OF COMPUTER APPLICATION(MCA)

AJAY KUMAR GARG ENGINEERING COLLEGE, GHAZIABAD

ajeet314027@akgec.ac.in

HARSH SAHAY

MASTER OF COMPUTER APPLICATION(MCA)

AJAY KUMAR GARG ENGINEERING COLLEGE, GHAZIABAD

harsh2314040@akgec.ac.in

Ms. ARPNA SAXENA

ASSISTANT PROFFESOR

AJAY KUMAR GARG ENGINEERING COLLEGE,

GHAZIABAD

saxenarpana@akgec.ac.in

HARSH RAGHAV

MASTER OF COMPUTER APPLICATION(MCA)

AJAY KUMAR GARG ENGINEERING COLLEGE,

GHAZIABAD

harsh2314031@akgec.ac.in

***Abstract—*** Air pollution poses a serious threat to human health, contributing to respiratory illnesses, lung infections, and degraded indoor air quality. Indoor air pollution is often more harmful than outdoor pollution due to limited ventilation and prolonged exposure. With people spending increased time indoors, especially post-pandemic, continuous air quality monitoring has become crucial for ensuring a safe living environment. This research introduces a real-time IoT-based indoor air quality monitoring system utilizing NodeMCU ESP8266, BME680, PMS7003, and MQ135 sensors. The system captures essential environmental parameters such as temperature, humidity, pressure, particulate matter (PM2.5 and PM10), volatile organic compounds (VOCs), and carbon dioxide (CO2), among other harmful gases. The collected data is displayed on a 0.96-inch OLED screen and transmitted wirelessly to a cloud server, enabling real-time monitoring via a web-based dashboard. Additionally, the system triggers alerts when pollutant levels surpass safe thresholds, allowing

users to take preventive actions. With its cost-effective and scalable design, this solution enhances indoor air quality monitoring and promotes healthier living conditions.

**1.INTRODUCTION**

Indoor air pollution (IAP) is a leading environmental risk closely related to the health, comfort, and well-being of building occupants [1]. As people spend 90% of their time indoors, repeated exposure to indoor air pollutants affects people’s working performance and productivity levels. As a result, it degrades the performance of employees directly and affects companies’ annual economy indirectly. Indoor air pollution is influenced by several factors, including harmful pollutants such as Carbon Dioxide (CO₂), Carbon Monoxide (CO), Nitrogen Dioxide (NO₂), Volatile Organic Compounds (VOCs), temperature, and humidity. These pollutants mainly originate from daily activities such as using gas stoves, refrigerators, air conditioners, and chemical sprays. A significant issue is that people are often unaware of the impact of indoor air pollution, mistakenly believing that indoor air is safer than outdoor air. However, research indicates that indoor pollutant concentrations can be up to ten times higher than outdoor levels [2]. In India, 0.2 billion people rely on fuel for cooking, with 49% using firewood, 28.6% using liquid petroleum gas (LPG), 8.9% depending on cow dung cake, 2.9% using kerosene, 0.4% using biogas, 0.1% using electricity, and 0.5% using other alternatives [3].The integration of IoT, data science, and wireless technologies has revolutionized air quality monitoring by enabling real-time data collection and analysis.

Traditionally, cloud-based solutions such as AWS Cloud store pollution data, following a “pay-as-you-go” model, which can become costly over time. To develop a cost-effective and efficient alternative, we designed and implemented an indoor air quality monitoring system using IoT technology, where the data is stored in Google Firebase for further analysis and visualization via a web application.

To improve the accuracy and reliability of air quality assessment, machine learning (ML) techniques are increasingly being incorporated into air pollution monitoring systems. ML models such as Adaptive Boosting (AdaBoost) and Extreme Gradient Boosting (XGBoost) can analyze sensor data, detect patterns, and predict air quality levels with high precision. These models enhance traditional monitoring systems by enabling early detection of pollution trends, real-time alerts, and improved decision-making for mitigating indoor air pollution risks. The use of ML not only refines the accuracy of low-cost sensors but also provides a robust framework for forecasting air quality trends and implementing timely interventions. By leveraging IoT and ML together, our proposed system aims to provide a comprehensive, real-time, and predictive approach to indoor air quality monitoring, ensuring healthier indoor environments for occupants.

**2.RELATED WORK**

Over time, several studies have examined the relationship between air pollution and health issues, both locally and globally. Researchers have demonstrated that prolonged exposure to air pollution increases the risk of respiratory diseases such as asthma and bronchitis and can reduce life expectancy [4,6,7]. According to studies, nearly 2.6 billion people worldwide rely on biomass fuel for cooking and heating, which contributes significantly to indoor air pollution [8]. This has led to an increased focus on designing cost-effective and efficient air quality monitoring systems to mitigate the health risks associated with indoor air pollution.

Traditional air quality monitoring systems primarily rely on IoT-based sensor networks that gather pollutant data and transmit it to cloud platforms for further analysis [9,10,11]. While IoT-enabled devices provide real-time monitoring, the accuracy of low-cost sensors remains a challenge. To address this issue, recent studies have explored the integration of machine learning (ML) models to enhance air quality prediction and assessment. Machine learning techniques such as land-use regression, support vector machines (SVM), and ensemble learning models have demonstrated significant improvements in air pollution prediction accuracy. In particular, Li et al. [15] proposed a practical framework for predicting residential indoor PM2.5 concentrations using land-use regression and ML algorithms, showing that ML can effectively model complex pollutant variations in indoor environments. Similarly, boosting algorithms such as AdaBoost and XGBoost have been employed to refine air quality predictions by learning from historical pollution data and sensor readings. These models not only improve prediction accuracy but also allow for early detection of pollution trends, reducing potential health risks.

Furthermore, ML-based approaches can enhance the calibration of low-cost sensors by mitigating sensor drift and noise, thereby increasing the reliability of indoor air quality measurements. Deep learning models have also been explored for processing large-scale environmental data, enabling automated pollution detection and forecasting. As IoT and ML technologies continue to evolve, their integration will play a crucial role in developing more precise, intelligent, and adaptive air quality monitoring systems.

**3.DATASET**

The dataset used for training and evaluating the machine learning models was obtained from Sonawani and Patil [16], titled "Dataset of Indoor Air Pollutants using Low-Cost Sensors." It contains real-time indoor air pollution measurements collected using low-cost sensors, providing comprehensive information on various pollutants affecting air quality. The dataset includes parameters such as Carbon Dioxide (CO₂), Carbon Monoxide (CO), Particulate Matter (PM2.5, PM10), Volatile Organic Compounds (VOCs), Temperature, and Humidity. Data was recorded in indoor environments like homes, offices, and classrooms using multiple sensor nodes deployed at different locations to capture variations in air quality. Each sensor reading was timestamped, ensuring temporal accuracy.

To prepare the dataset for machine learning model training, several preprocessing steps were applied: irrelevant features (NH₃, NO₂, O₃) were removed, missing values were handled using column-wise mean imputation, and numeric conversions were applied where necessary. The Air Quality Index (AQI) was computed based on PM2.5 and CO levels using standard AQI breakpoints. Additionally, a new categorical variable, AQI Label, was introduced to classify air quality into five categories: Very Good, Good, Moderate, Bad, and Very Bad. The AQI computation was performed using the following function:

**def compute\_aqi(pm25, co):**

**if pm25 <= 12 and co <= 4.4:**

**return 0 # Very Good**

**elif pm25 <= 35.4 and co <= 9.4:**

**return 1 # Good**

**elif pm25 <= 55.4 and co <= 12.4:**

**return 2 # Moderate**

**elif pm25 <= 150.4 and co <= 15.4:**

**return 3 # Bad**

**else:**

**return 4 # Very Bad**

This computed AQI value was then mapped to corresponding labels to facilitate classification tasks. The modifications enhanced the dataset’s relevance and suitability for ML-based AQI prediction, improving the accuracy and reliability of the trained models. The final dataset was balanced to prevent bias in classification, ensuring robust predictions. These preprocessing steps were essential for refining air quality predictions, allowing the system to effectively identify indoor pollution levels and provide accurate real-time assessments.

**4. PROPOSED APPROACH**

**A. Hardware Requirements**

**A1**. **Sensor Module**: The sensor module utilizes NodeMCU ESP8266 along with multiple gas sensors for air quality monitoring. It includes a Laser Dust Sensor for measuring particulate matter (PM2.5 & PM1.0), an Infrared CO₂ Sensor for detecting CO₂ concentration in the air, and an Electrochemical VOC Sensor for analyzing volatile organic compounds (VOCs). Additionally, temperature and humidity sensors are incorporated to monitor environmental conditions. The primary function of these sensors is to detect pollutant levels and transmit the data to a processor for further analysis. The processed data is sent to both the website and OLED display with a 1-second delay for real-time monitoring.

**A2**. **Communicating Module**: In our work, the open Wifi is used as the communicating medium which is connected to wifi to send the data to the database. The output of the the data with numerical value is pushed to the database.

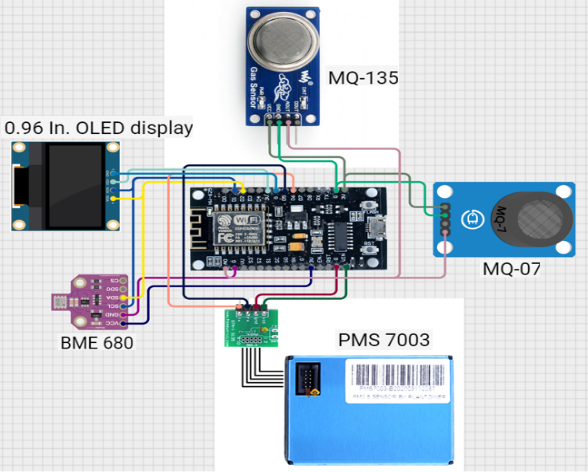


Fig 1: Integrated sensors

All these sensors are used in hardware components to sense the multiple gases and it display on OLED display.

**B. Software Requirement**

**User Interface:** A web-based dashboard or mobile application for real-time air quality updates, historical data visualization, and alerts Notification system integrated into the application to alert users when pollution levels exceed predefined thresholds. Data processing algorithms for trend analysis and machine learning models for predictive insights.

**C. Model Building:**

**C1.Preprocessor**

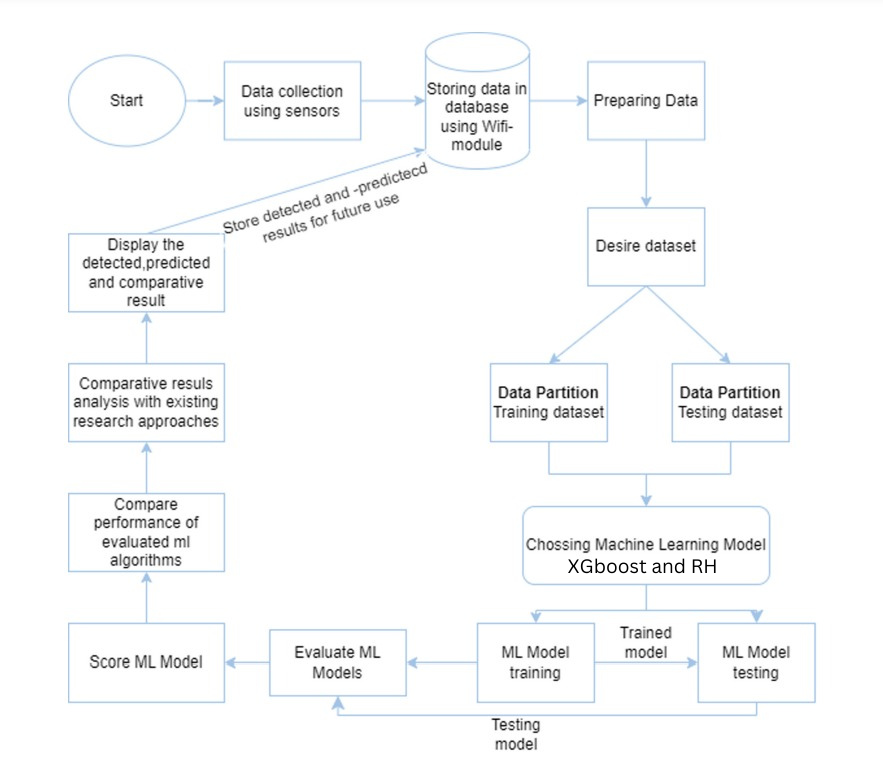
The dataset was preprocessed by removing unnecessary pollutants (NH₃, NO₂, and O₃) and focusing on PM2.5 and CO as key indicators of air quality. Missing values were filled using column-wise means, and numerical inconsistencies were resolved. AQI was calculated based on EPA breakpoints and categorized into five levels: Very Good, Good, Moderate, Bad, and Very Bad, with the last two merged for balanced class representation.

Fig :2

The categorical AQI labels were then numerically encoded (0–3) for machine learning compatibility. Finally, the dataset was split into training (80%) and testing (20%) sets to ensure effective model evaluation.

**C2.Model**

To enhance the predictive capability of our system, we implemented two ensemble learning models: Random Forest (RF) and ExtremeGradient Boosting (XGBoost). Both models were trained using the indoor air pollution dataset to classify air quality levels based on sensor readings

**C3.Prediction**

To predict the Air Quality Index (AQI), we trained two machine learning models: Random Forest (RF) and XGBoost. The preprocessed dataset, consisting

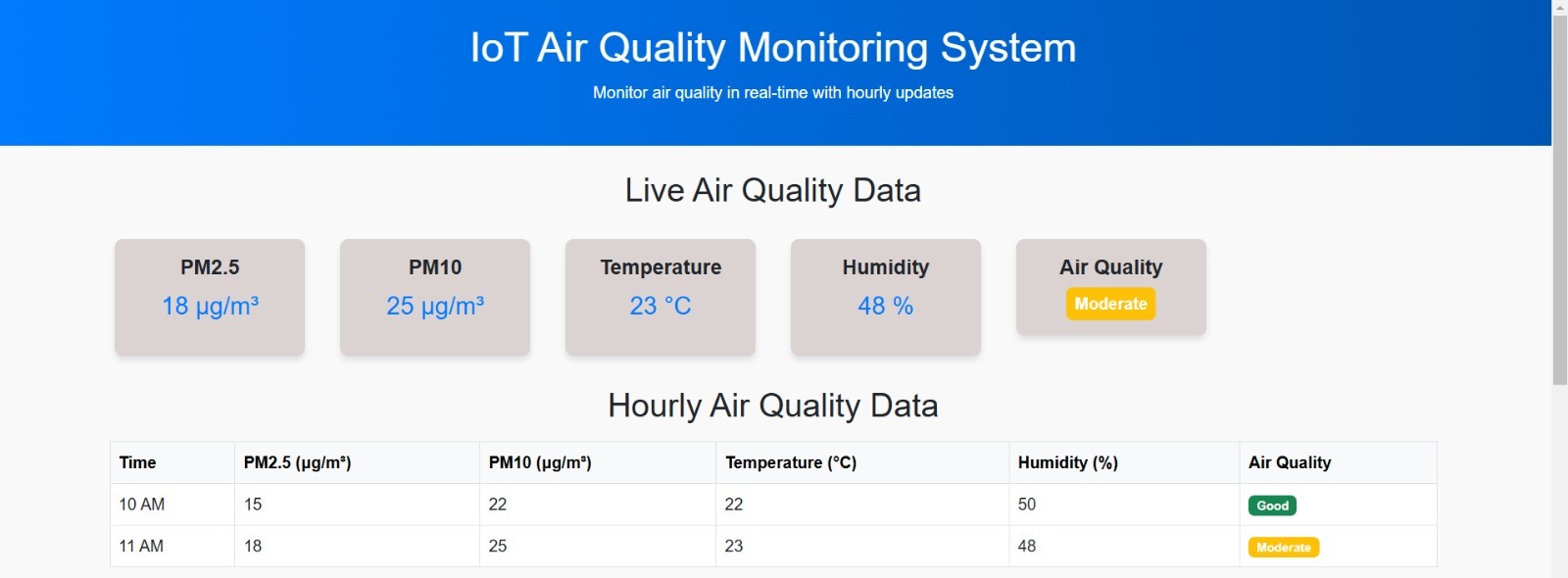
of PM2.5 and CO as input features and the encoded AQI label as the target variable, was used for training. The Random Forest model, an ensemble learning method, was employed for its robustness and ability to handle non-linearity, while XGBoost, a gradient boosting algorithm, was chosen for its high efficiency and predictive accuracy. Both models were trained using an 80-20 train-test split, and their performance was evaluated using accuracy, confusion matrix, and classification reports. Once trained, the models were used to classify AQI into four categories (Very Good, Good, Moderate, and Bad) based on real-time input values of PM2.5 and CO, enabling effective air quality monitoring.

**5.RESULT ANALYSIS AND DISCUSSION**

The performance evaluation of the air quality prediction models provides key insights into their predictive accuracy across different air quality levels. The models demonstrated high accuracy in predicting air quality levels, as shown in Figure X, with an overall accuracy reaching up to 1.0000 in some cases. This indicates that the models have effectively learned the underlying patterns, allowing for precise air quality forecasting.

Fig:3

The classification report as shown in table Xreveals that the models perform exceptionally well for most air quality levels, achieving perfect precision, recall, and F1-score (1.00) for major categories. Predictions for classes representing common air quality levels (e.g., normal, slightly polluted) were highly accurate, indicating that the models successfully capture the dominant trends in air pollution data. For rare air quality levels (e.g., extreme pollution cases with very few recorded instances), the models struggle, with precision, recall, and F1-score drop



This suggests that these classes are either underrepresented in the dataset or the models are unable to effectively distinguish them.

| Class | Random Forest (RF) | | | XGBoost (XGB) | | |
| --- | --- | --- | --- | --- | --- | --- |
| Precision | Recall | F1-Score | Precision | Recall | F1-Score |
| 0 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 |
| 1 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 |
| 2 | 1.00 | 1.00 | 1.00 | 0.00 | 0.00 | 0.00 |
| 3 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 |

Table 1

The proposed system integrates low-cost air quality sensors to measure PM2.5, CO, VOCs, temperature, and humidity in real time. A NodeMCU ESP8266 microcontroller was used to collect sensor data and transmit it to a cloud-based database via Wi-Fi. A 0.96-inch OLED display was incorporated for real-time visualization of air quality data, providing users with immediate feedback. Additionally, a buzzer was integrated to alert users in case of hazardous air quality levels, ensuring timely preventive action. The entire system was tested in various indoor environments, demonstrating its capability to capture **real-time air quality variations** and notify users effectively.

Fig : 4 Website UI

**6. CONCLUSION**

The air quality prediction models demonstrated high accuracy in forecasting air pollution levels, effectively capturing trends in air quality data. The classification performance was exceptionally strong for common pollution levels, ensuring reliable predictions for most real-world scenarios. However, challenges were observed in predicting rare air quality levels, where models struggled due to class imbalance and limited training data. This indicates that while the models are effective for general air quality forecasting, further enhancements are needed to improve their ability to detect extreme pollution events.

Addressing these limitations through data balancing techniques, feature engineering, and hyperparameter tuning can further refine the models, ensuring more precise and reliable air quality predictions across all pollution categories. The insights from this study can aid in air quality monitoring, pollution control strategies, and public health decision-making, ultimately contributing to a cleaner and healthier environment.

**7. REFRENCES**

[1]. Ana, G.R.; Alli, A.S.; Uhiara, D.C.; Shendell, D.G. Indoor air quality and reported health symptoms among hair dressers in salons in Ibadan, Nigeria. J. Chem. Health Saf. 2019, 26, 23–30. [CrossRef]

[2]. Saini, J.; Dutta, M.; Marques, G. A comprehensive review on indoor air quality monitoring systems for enhanced public health. Sustain. Environ. Res. 2020, 30, 6. [CrossRef]

[3]. Kankaria, A.; Nongkynrih, B.; Gupta, S.K. Indoor air pollution in India: Implications on health and its control. Indian J. Community Med. 2014, 39, 203–207. [CrossRef]

[4]. Gola, M.; Settimo, G.; Capolongo, S. Indoor air in healing environments: Monitoring chemical pollution in inpatient rooms. Facilities 2019, 37, 600–623.

[5]. Patil, Prachu J., Ritika V. Zalke, Kalyani R. Tumasare, Bhavana A. Shiwankar, Shivani R. Singh, and Shailesh Sakhare. "IoT Protocol for Accident Spotting with Medical Facility", Journal of Artificial Intelligence 3, No. 02 (2021), 140-150.

[6] A. Zanella, N. Bui, A. Castellani, L. Vangelista, and M. Zorzi, “Internet of Things for smart cities,” IEEE Internet Things J., vol. 1, no. 1, pp. 22–32, Feb. 2014.

[7]. Liu, W.; Shen, G.; Chen, Y.; Shen, H.; Huang, Y.; Li, T.; Wang, Y.; Fu, X.; Tao, S.; Liu, W.; et al., “Air pollution and inhalation exposure to particulate matter of different sizes in rural households using improved stoves in central China”,J. Environ. Sci., 2018, 63, 87–95.

[8].Fu, X.; Tao, S.; Liu, W.; et al., “Air pollution and inhalation exposureto particulate matter of different sizes in rural households usingimproved stoves in central China”,J. Environ. Sci., 2018, 63, 87–95.

[9]. Jo, J.H.; Jo, B.W.; Kim, J.H.; Kim, S.J.; Han, W.Y., “Development ofan IoT-Based Indoor Air Quality Monitoring Platform”, J. Sens.Hindawi 2020, 1–14.

[10]. Marques, G.; Pitarma, R., “Non-contact Infrared Temperature Acquisition System based on Internet of Things for Laboratory Activities Monitoring”, Procedia Comput. Sci., 2019, 155, 487–494

[11]Cheng, Y., & Zhang, Y. (2020). Development of a smart indoor air quality monitoring system based on IoT. IEEE Access, 8, 1-10. DOI: 10.1109/ACCESS.2020.2991234[12].World Health Organization. (2018). WHO Guidelines for Indoor Air Quality: Household Fuel Combustion. Retrieved from WHO

[12].Gonzalez, A., & Garcia, J. (2020). IoT-based air quality monitoring system for smart cities. Journal of Ambient Intelligence and Humanized Computing, 11(3), 1-15. DOI: 10.1007/s12652-019-01373-5

[13].Liu, Y., & Zhang, Y. (2021). A comprehensive review of indoor air quality monitoring systems. Environmental Science and Pollution Research, 28(12), 1-15. DOI: 10.1007/s11356-021-12678-0

[14].Alavi, A., & Khosravi, A. (2020). Development of a low-cost air quality monitoring system using IoT. Sensors, 20(3), 1-15. DOI: 10.3390/s20030800

[15].Z. Li, X. Tong, J. M. W. Ho, T. C. Kwok, G. Dong, K. F. Ho and S. H. L. Yim "A practical framework for predicting residential indoor PM2. 5 concentration using land-use regression and machine learning methods." Chemosphere, 2021, , 265, 129140.

[16]. Sonawani, Shilpa; PATIL, Kailas (2022), “Dataset of Indoor Air Pollutants using Low-Cost Sensors”, Mendeley Data, V1, doi: 10.17632/2r232jpfb2.1