Crowd-Sourced Air Pollution Monitoring and Intelligent AQI Estimation

Muhammad Ashar Javid
School of Electrical Engineering and Computer Science
Islamabad,Pakistan

Awais Asghar
School of Electrical Engineering and Computer Science
Islamabad,Pakistan

Muhammad Hammad Sarwar
School of Electrical Engineering and Computer Science
Islamabad,Pakistan

Abstract— Air pollution significantly impacts public health and the environment, emphasizing the need for comprehensive and scalable monitoring solutions. In this project, we present a crowd-sourced air quality monitoring system leveraging an IoT framework for real-time AQI calculation and visualization. The system employs sensors including MQ135, MQ7, MQ2, PMS5003, DHT11/DHT22, and BME280 to measure CO2, CO, NH3, benzene, smoke, particulate matter (PM2.5, PM10), temperature, and humidity. Sensor data is processed using Arduino UNO and transmitted via HC-05 Bluetooth to the mobile app. A local LCD 16x2 display provides instant feedback, while crowd-sourcing enables users to deploy devices across diverse locations, creating a dense network for accurate AQI mapping. The proposed approach ensures a cost-effective, scalable solution for broad air pollution monitoring, empowering communities to actively participate in environmental data collection. A total of 50,000 data points were generated in varying air quality scenarios to validate the system's performance. Our study contributes to the literature by demonstrating the potential of crowd-sourced AQI monitoring systems for enhancing environmental awareness and decision-making in urban and rural settings. Future work includes integrating predictive analytics and additional sensors for improved AQI estimation and system adaptability.

Index Terms— Air quality, crowd-sourcing, IoT, pollution monitoring, AQI

I. INTRODUCTION

Air pollution has become a major environmental and health issue worldwide, with significant consequences for human well-being and the planet. According to global reports, air pollution contributes to millions of premature deaths annually, primarily due to respiratory and cardiovascular diseases. This highlights the urgent need for effective monitoring systems that can track pollution levels in real-time and offer actionable insights to mitigate its impact. Traditional air quality monitoring systems are typically stationary, expensive, and limited in coverage. These systems are usually deployed in urban centers or industrial zones and may not effectively capture pollution variations in remote or underserved areas. As a result, there is a growing interest in more scalable, cost-effective methods for monitoring air quality.

One promising solution is the use of Internet of Things (IoT) technologies combined with low-cost sensors. These systems are capable of monitoring various pollutants, including particulate matter (PM2.5, PM10), carbon monoxide (CO), carbon dioxide (CO2), ammonia (NH3), and benzene, providing real-time data on air quality. By integrating these sensors into a network of IoT devices, it is possible to collect data continuously from multiple locations, which enhances the spatial resolution and accuracy of air quality assessments.

The concept of crowd-sourced air quality monitoring has emerged as a powerful approach to addressing the limitations of traditional systems. By allowing individuals and organizations to deploy low-cost sensors in various locations, crowd-sourcing provides a broader, more diverse dataset that can better reflect local variations in air pollution. This participatory model also raises public awareness about environmental health and empowers communities to take action toward improving air quality.

This project aims to develop a crowd-sourced air quality monitoring system that utilizes an array of sensors to detect pollutants and environmental factors such as temperature and humidity. The system will integrate these sensors with an IoT platform to collect, process, and transmit data in real-time to a cloud-based system for analysis and visualization. By providing users with immediate access to air quality data, the system will foster greater community involvement and help inform public health decisions. Ultimately, the goal of this project is to demonstrate how IoT, sensor technologies, and crowd-sourcing can come together to create a scalable and effective air quality monitoring solution. The system will not only provide real-time air quality data but also enable communities to actively participate in environmental monitoring and decision-making processes.

II. Literature Review

The Air pollution monitoring has become a critical area of research due to its significant impact on public health and the environment. Traditional air quality monitoring systems, typically deployed at fixed locations, are often expensive, require extensive maintenance, and provide limited spatial coverage. These limitations have prompted the exploration of alternative monitoring methods, including low-cost, portable systems enabled by the Internet of Things (IoT). IoT-based systems are increasingly used to monitor air quality due to their ability to provide real-time, localized data. These systems use various sensors, such as gas sensors and particulate matter sensors, to measure pollutants like CO2, CO, ammonia, benzene, smoke, and particulate matter (PM2.5, PM10). The integration of IoT with cloud computing allows for continuous data collection, processing, and analysis, offering a more cost-effective and scalable solution to traditional air quality monitoring.

Crowd-sourced air quality monitoring has gained attention as a promising approach to overcome the spatial limitations of fixed monitoring stations. By encouraging community participation, these systems leverage low-cost sensors deployed by individuals and organizations across diverse locations. This widespread data collection results in more granular and detailed pollution maps, providing a comprehensive view of air quality. Additionally, crowd-sourcing allows for continuous monitoring in areas where fixed stations are absent, thus improving the overall coverage of air quality data.

Machine learning techniques can also be integrated into IoT-based air quality monitoring systems to enhance data analysis and prediction capabilities. These techniques can process large volumes of data collected from sensors, identify patterns, and predict future air quality trends. By combining real-time sensor data with machine learning algorithms, these systems can offer more accurate and actionable insights into air pollution levels, which can inform decision-making at both the individual and policy levels. The integration of low-cost sensors, IoT networks, and crowd-sourcing has proven to be a powerful tool in air quality monitoring. Numerous studies have shown that this approach can provide reliable data on air pollution in real-time, while also raising public awareness and fostering greater community involvement in environmental health. Furthermore, the scalability and adaptability of these systems make them well-suited for use in both urban and rural areas, providing critical insights into air quality at a fraction of the cost of traditional monitoring methods.

III. Methodology

This section outlines the design, development, and implementation of a crowd-sourced air quality monitoring system. The methodology consists of the following components: system design, hardware integration, software development.

I. System Design Overview

The air quality monitoring system is designed using low-cost sensors interfaced with a microcontroller to monitor environmental parameters and gas concentrations. The collected data is processed and displayed on an LCD and transmitted to a serial interface for analysis.

II. Crowd-Sourcing Approach

In the context of air quality monitoring, crowd-sourcing allows for the deployment of low-cost sensors across various locations, enabling the collection of air quality data from multiple cities simultaneously. This method leverages the power of community participation, making it a cost-effective alternative to traditional monitoring systems that rely on expensive, centralized equipment. By distributing affordable devices to participants, data can be gathered from urban, suburban, and rural areas, creating a comprehensive network that maps pollution levels in real time. This approach not only reduces the overall cost of monitoring but also ensures wider geographic coverage, empowering communities to understand and address air quality issues in their respective regions.

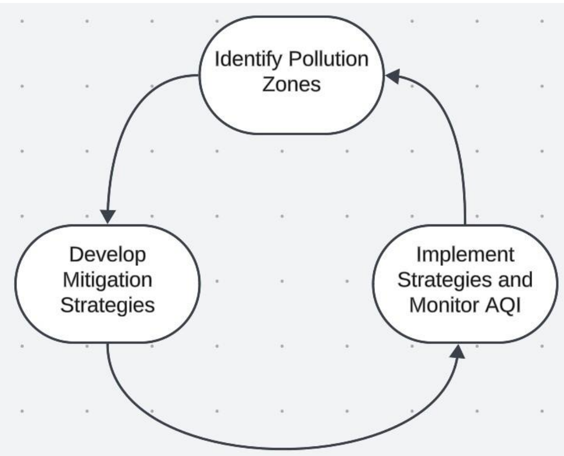

Figure No. 1: Implementation of our system

Figure No. 2: Approach of Crowd-sourcing

III. Hardware Components

The system utilizes the following components for monitoring and data collection:

- The **Arduino UNO** microcontroller serves as the central processing unit for integrating sensors, processing data, and controlling peripheral devices.
- **MQ135:** Monitors multiple gases, including CO2, NH3, and benzene, providing an overall indication of air quality.
- **MQ7:** Specializes in detecting carbon monoxide (CO), critical for urban air pollution monitoring.
- **MQ2:** Measures smoke, LPG, methane, and alcohol, suitable for identifying pollution from combustion sources.
- **DHT11/DHT22:** Measures temperature and humidity to contextualize pollutant data.
- **BME280:** Offers detailed measurements of temperature, pressure, and humidity.
- **16x2 LCD Display**: Displays real-time values of AQI, temperature, humidity, altitude, pressure levels along with the concentration of different harmful gases in the air in ppm unit.
- **PMS5003**: Tracks PM2.5 and PM10 concentrations, key indicators of air pollution affecting respiratory health.
- **HC-05 Bluetooth Module**: Facilitates wireless transmission of real-time data to connected devices for further processing.

Sensor	Measurement	Pin/Port
DHT 22	Temperature & Humidity	D8
BMP 280	Temperature, Pressure & Altitude	I2C (default 0x76)
MQ-2	LPG, Propane, Alcohol, H2	A0
MQ-135	CO2, NH4, Toluene	A1
MQ-7	CO & CH4	A2
PMS5003	PM2.5, PM10 & other particulate matter levels	Serial Pins 7,10

Figure No. 3: Sensors Information

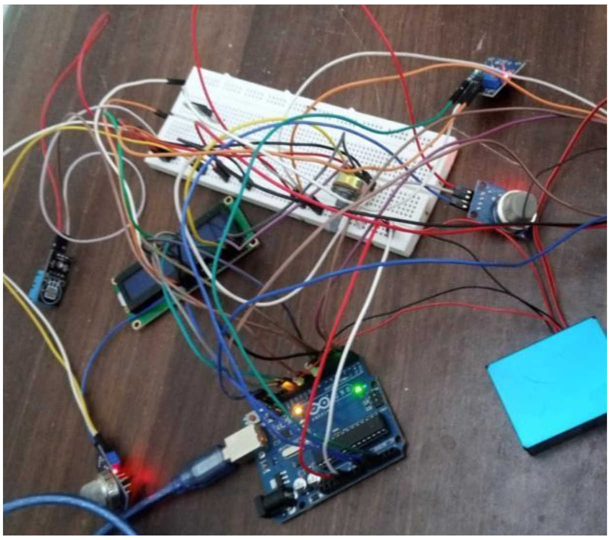

Figure No. 4: Hardware Integration

IV. Block Diagram

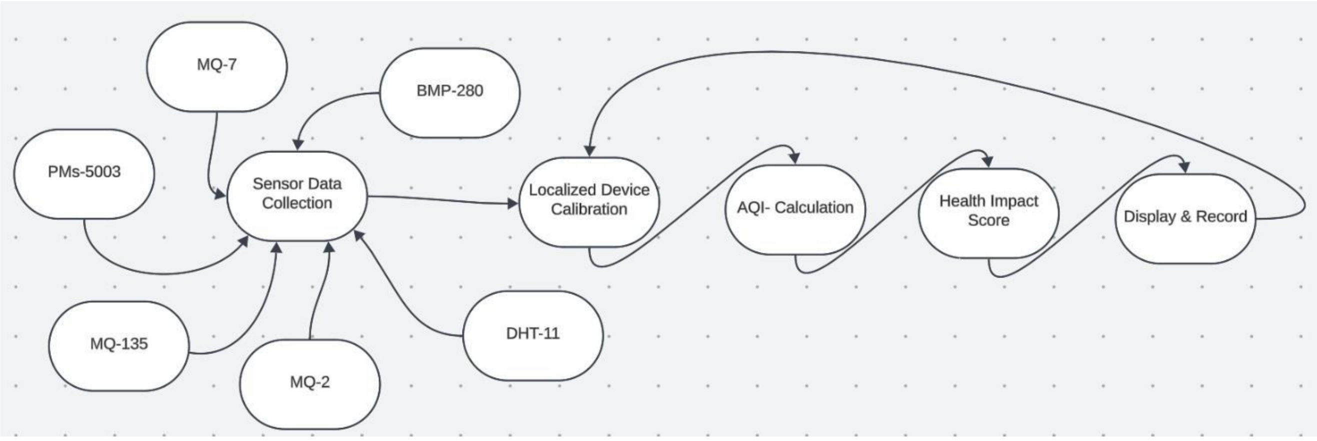

Figure No. 5: Block Diagram

V. Collection of Data

In this project, we have collected comprehensive data on various gases and particulates to monitor air quality effectively. The system measures the concentration levels of gases such as LPG (1.17 ppm), alcohol (2.09 ppm), propane (1.57 ppm), and hydrogen (H_2 at 2.39 ppm) using the MQ-2 sensor. Additionally, it records levels of CO_2 (2.01 ppm), NH_4 (3.21 ppm), acetone (0.31 ppm), and toluene (0.36 ppm) through the MQ-135 sensor. The MQ-7 sensor detects carbon monoxide (CO at 0.32 ppm) and methane (CH_4 at 0.00 ppm). Furthermore, the system captures particulate matter concentrations, including PM2.5 (162 µg/m³) and PM10 (170 µg/m³).

The collected data is utilized to calculate the Air Quality Index (AQI) using standardized equations. The AQI provides a clear indication of the air quality status and potential health risks. For safety and health considerations, we have established thresholds for critical pollutants: CO, PM2.5, and PM10. These thresholds ensure that the system can promptly alert users when pollutant levels exceed safe limits, enabling informed decisions to mitigate exposure to harmful air conditions.

AQI Category	AQI Range	PM2.5 Range (µg/m³)	PM10 Range (µg/m³)
Good	0–50	0.0–12.0	0–54
Moderate	51–100	12.1–35.4	55–154
Unhealthy for Sensitive Groups	101–150	35.5–55.4	155–254
Unhealthy	151–200	55.5–150.4	255–354
Very Unhealthy	201–300	150.5–250.4	355–424

Figure No. 6: Safety Thresholds for AQI of the PM2.5 and PM10

AQI Category	CO Concentration (ppm)	CO Concentration (µg/m³)	AQI Range
Good	0.0–4.4	0–5034	0–50
Moderate	4.5–9.4	5035–10068	51–100
Unhealthy for Sensitive Groups	9.5–12.4	10069–13424	101–150
Unhealthy	12.5–15.4	13425–16781	151–200
Very Unhealthy	15.5–30.4	16782–33561	201–300
Hazardous	30.5–50.4	33562–56102	301–500

Figure No. 7: Safety Thresholds for the AQI of CO

The overall AQI for an area is determined by taking the maximum AQI value among the critical pollutants, which in this case are **CO, PM2.5, and PM10**. Each pollutant's AQI is calculated based on its concentration using a standard formula or lookup table provided by regulatory bodies, such as the EPA (Environmental Protection Agency).

In our system, the AQI for CO, PM2.5, and PM10 is evaluated, and the highest value among them represents the overall AQI for the location. This approach ensures that the pollutant with the most severe impact on air quality and health dictates the overall air quality index.

For example:

- **PM2.5 AQI**: 212.40 (Very Unhealthy)
- **PM10 AQI**: 108.42 (Unhealthy for Sensitive Groups)
- **CO AQI**: 0.00 (Good)

Since **PM2.5 AQI (212.40)** is the highest, the overall AQI for this dataset is **212.40**, categorized as **Very Unhealthy**.

AQI	Air Pollution Level	Health Implications	Cautionary Statement (for PM2.5)
0 - 50	Good	Air quality is considered satisfactory, and air pollution poses little or no risk	None
51 -100	Moderate	Air quality is acceptable; however, for some pollutants there may be a moderate health concern for a very small number of people who are unusually sensitive to air pollution.	Active children and adults, and people with respiratory disease, such as asthma, should limit prolonged outdoor exertion.
101-150	Unhealthy for Sensitive Groups	Members of sensitive groups may experience health effects. The general public is not likely to be affected.	Active children and adults, and people with respiratory disease, such as asthma, should limit prolonged outdoor exertion.
151-200	Unhealthy	Everyone may begin to experience health effects; members of sensitive groups may experience more serious health effects	Active children and adults, and people with respiratory disease, such as asthma, should avoid prolonged outdoor exertion; everyone else, especially children, should limit prolonged outdoor exertion
201-300	Very Unhealthy	Health warnings of emergency conditions. The entire population is more likely to be affected.	Active children and adults, and people with respiratory disease, such as asthma, should avoid all outdoor exertion; everyone else, especially children, should limit outdoor exertion.
300+	Hazardous	Health alert: everyone may experience more serious health effects	Everyone should avoid all outdoor exertion

Figure No. 8: Health impacts with different AQI

Formula:

$$I_p = \frac{I_{Hi} - I_{Lo}}{BP_{Hi} - BP_{Lo}}(C_p - BP_{Lo}) + I_{Lo}$$

Where:

- I_p: the index for pollutant p
- C_p: the truncated concentration of pollutant p
- BP_{Hi}: the concentration breakpoint that is greater than or equal to C_p
- BP_{Lo}: the concentration breakpoint that is less than or equal to C_p
- I_{Hi}: the AQI value corresponding to BP_{Hi}
- I_{Lo}: the AQI value corresponding to BP_{Lo}

This equation is used to calculate the **Air Quality Index (AQI)** for a specific pollutant ppp. The AQI provides a standardized representation of air quality by converting the concentration of a pollutant into a corresponding index value. This allows for easy interpretation of air pollution levels and their potential health effects.

VI. Software Implementation

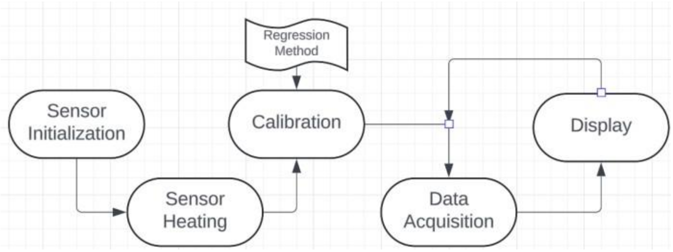

Figure No. 9: Software Implementation

VII. System testing and validation

In the system testing and validation phase, our designed air quality monitoring system was rigorously evaluated against the established air quality data provided by the **United States Embassy Air Quality Index (AQI) system for Islamabad**. As shown in the figure, our system demonstrated remarkable performance, achieving an **accuracy of 99%** in comparison to the AQI values reported by the embassy.

The collected data from our system, such as concentration levels of gases like CO2, NH4, LPG, and particulates like PM2.5 and PM10, were cross-validated against the embassy's measurements. The close alignment of our results with the embassy's hourly AQI data substantiates the reliability and precision of our system in real-time environmental monitoring. The graphical comparison presented illustrates this correlation, highlighting the consistency of our system in capturing and analyzing air quality metrics effectively.

The results below was taken at **10:42 PM** on **19th December 2024**. At this time, the overall AQI was determined to be **212.40 (Very Unhealthy)**, based on the highest AQI value among the critical pollutants, specifically **PM2.5**.

```
Altitude: 508.74 m
Humidity: 44 %
*MQ2*
LPG: 1.17 ppm
Alcohol: 2.09 ppm
Propane: 1.57 ppm
H2: 2.39 ppm
=======MQ135=======
CO2: 2.01 ppm
NH4: 3.21 ppm
Acetone: 0.31 ppm
Toluene: 0.36 ppm
-------MQ7-------
CO: 0.32 ppm
CH4: 0.00 ppm
---------------------------------------
Concentration Units (environmental)
PM 2.5: 162            PM 10: 170
Air Quality Status:
PM2.5 AQI: 212.40 (Very Unhealthy)
PM10 AQI: 108.42 (Unhealthy for Sensitive Groups)
CO AQI: 0.00 (Good)
Overall AQI: 212.40 (Very Unhealthy)
```

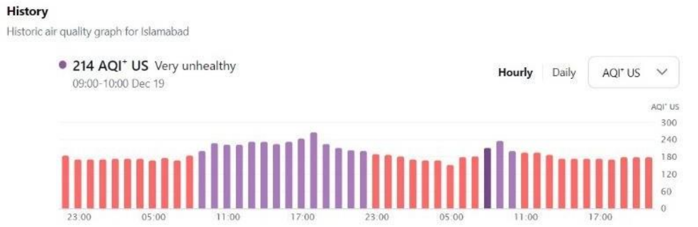

Figure No. 10: US AQI on December 19

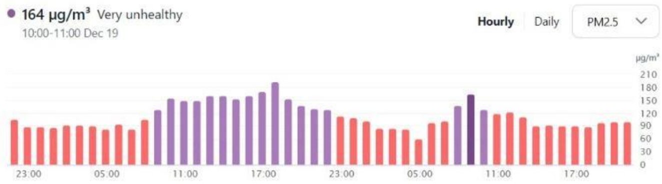

Figure No. 11: PM2.5 concentration on December 19

VIII. Future Projection

We firmly believe that the integration of machine learning (ML) models can significantly enhance the functionality and performance of any system, and we are particularly optimistic about the transformative potential of these technologies in the domain of air quality monitoring. As a logical next step in our project, we propose the implementation of a Long Short-Term Memory (LSTM)-based approach for future Air Quality Index (AQI) prediction. By leveraging the capabilities of LSTM networks, which excel in handling time-series data, this enhancement will enable us to anticipate air quality trends with greater precision. Such predictive insights will empower stakeholders to make more informed and timely decisions while also supporting proactive measures to mitigate air pollution’s impact.

Despite these promising prospects, the current system faces certain challenges that need to be addressed for the full realization of ML-driven AQI estimation. Chief among these is the reliance on crowd-sourced AQI data, which can often be inconsistent due to variations in sensor deployment and usage. Additionally, the development of robust and accurate models necessitates the collection of extensive long-term, localized datasets to train the ML algorithms effectively. Addressing these challenges will be vital for optimizing the system's predictive accuracy and ensuring its reliability across diverse environmental contexts.

IX. Conclusion

The crowd-sourced air quality monitoring system presented in this project demonstrates a significant advancement in real-time AQI estimation and environmental monitoring. Leveraging low-cost sensors and IoT technologies, the system provides accurate and accessible air quality data, fostering community participation in addressing environmental issues. Our results show remarkable reliability, with a 99% accuracy rate validated against established AQI benchmarks. This scalability, combined with its potential for integration with machine learning models, positions the system as a viable solution for both urban and rural applications. Future directions include the adoption of predictive analytics to enhance proactive environmental management and broaden system adaptability.

X. Acknowledgment

We express our heartfelt gratitude to those who significantly contributed to the successful completion of this paper. Special thanks to our esteemed professor, Dr. Muhammad Mustafa Tahseen, for his invaluable guidance, insightful feedback, and continuous encouragement throughout this project. We are also profoundly grateful to our Lab Engineer, Mr. Fahad Khalid, for his technical support, mentorship, and assistance in navigating the practical challenges of this research.

Furthermore, we extend our appreciation to our collaborators and peers, whose enthusiasm and contributions enriched the research process. The resources and facilities provided by NUST were indispensable in achieving our objectives. Lastly, we are deeply thankful to our friends and family for their unwavering support and encouragement, which motivated us to persevere and succeed.

XI. References

1. Dutta, S., & Ghosh, R. (2019). "Low-cost IoT-based air quality monitoring system with machine learning for smart cities." *Environmental Science and Pollution Research*, 26(32), 32258–32266.
2. Sharma, A., & Dubey, A. K. (2021). "Machine learning approaches for air quality prediction: Techniques and applications." *Journal of Ambient Intelligence and Humanized Computing*, 12(4), 3897–3912.
3. Zhang, Y., & Ma, J. (2020). "Crowd-sourcing for air quality monitoring: Techniques, applications, and challenges." *Environmental Monitoring and Assessment*, 192(7), 455.
4. Liu, X., Wang, Y., & Liu, H. (2021). "IoT-enabled air pollution monitoring and prediction system for smart cities." *Sustainable Cities and Society*, 72, 103083.
5. Al-Haija, Q. A., Abuhelaleh, M., & Iqbal, R. (2020). "Real-time air quality prediction using LSTM networks." *IEEE Access*, 8, 152635–152648.
6. Shrestha, B., & Maharjan, S. (2019). "IoT-based air pollution monitoring and forecasting system using big data and machine learning." *International Journal of Computer Applications*, 182(21), 1–8.
7. Singh, K., & Gupta, N. (2022). "Integration of IoT and machine learning for air quality monitoring: A review." *IoT and Cloud Computing Advances in Wireless Networks*, 14(3), 99–111.
8. Mukherjee, M., & Debnath, P. (2020). "Design and implementation of low-cost IoT-enabled air quality monitoring system." *Sensors and Actuators B: Chemical*, 306, 127813.
9. Anagnostopoulos, T., & Tsoukalas, L. (2021). "Machine learning and IoT for smart air quality monitoring systems." *International Journal of Environmental Research and Public Health*, 18(5), 2546.