Pollutants Detection in Eco-Friendly Cities using Sensors: Data Analysis Application

Madupu Hema[1], Madeshi Keerthana[2], Ragam Ajay[3] Dr. S. Sreekanth[4]

[1]B-Tech 4th year, Dept. of CSE(DS), Institute of Aeronautical Engineering
[2] B-Tech 4th year, Dept. of CSE(DS), Institute of Aeronautical Engineering
[3] B-Tech 4th year, Dept. of CSE(DS), Institute of Aeronautical Engineering
[4]Associate Professor, Dept. of CSE(DS), Institute of Aeronautical Engineering

---***---

Abstract - *In response to escalating environmental pollution concerns, this abstract presents an advanced sensor-based system integrated with machine learning for real-time pollutant detection and analysis. The system employs a network of strategically positioned sensors to monitor pollutants, including nitrogen dioxide (NO_2), ozone (O_3), particulate matter (PM2.5 and PM10), and volatile organic compounds (VOCs). Sensor data is transmitted to a central unit, where machine learning algorithms like SVM and CNN analyze patterns, trends, and anomalies for accurate pollutant characterization. Advanced data analysis techniques identify pollutant sources and assess risks, while adaptive learning enhances accuracy and predictive capabilities over time, optimizing pollution monitoring and management efforts.*

***Key Words*: Sensor-based system, pollutants detection, Machine learning algorithms, Data analysis techniques, Environmental monitoring, Real- time analysis, public health, Adaptive learning**

1.INTRODUCTION

The rapid growth of urbanization and industrialization has significantly increased air pollution, posing severe health and environmental challenges. Addressing these issues. Air pollution is a critical global issue, requiring advanced research on its causes, impacts, and mitigation strategies. This literature survey highlights key studies focusing on concept of eco-friendly or smart cities has emerged, emphasizing technology integration to promote healthier urban environments. A core component of such cities is real-time pollutant detection systems using advanced sensors.

These sensors monitor pollutants like nitrogen dioxide (NO_2), sulfur dioxide (SO_2), particulate matter (PM2.5 and PM10), carbon monoxide (CO), and volatile organic compounds (VOCs), which are linked to respiratory and cardiovascular diseases and environmental degradation. By collecting real-time data, cities can identify pollution hotspots, predict trends, and implement preventive measures. Smart sensors, being compact, cost-effective, and easy to deploy, offer scalable solutions for urban monitoring.

Machine learning and data visualization play a critical role, enabling data analysis, trend prediction, and actionable insights. Tools like heat maps and dashboards make pollution data accessible, empowering citizens and fostering civic engagement. Integration with technologies like IoT and cloud computing enhances data sharing and cross-departmental collaboration, enabling a dynamic response to pollution challenges.

2. LITERATURE SURVEY

Air pollution has emerged as a significant environmental and public health issue globally, necessitating comprehensive research to understand its causes, effects, and mitigation strategies. In this literature survey, we review key studies related to air pollution monitoring, forecasting, and control, with a focus on the application of machine learning techniques for air quality prediction.

Zhang et al. (2021) explored the application of machine learning techniques in air pollution prediction, particularly using Random Forest and Support Vector Machines to analyze data from smart sensors [2]. Their study showed that these models can accurately predict short-term fluctuations in pollution levels based on historical data, improving the capacity for real-time decision-making. This research highlights the potential of machine learning models in enhancing the accuracy and efficiency of pollutant detection systems in smart cities.

Cheng et al. (2020) focused on the spatial distribution of air pollution and proposed a Geographic Information System (GIS)-based approach to map pollution hotspots in urban areas using data from air quality sensors [4]. Their findings demonstrated that combining sensor data with GIS tools can help city planners and policymakers identify high-risk areas and prioritize interventions. This study underscores the need for spatial data analysis in managing pollution in eco-friendly cities.

These studies underline advancements in sensor technology, data analysis, and predictive modeling as essential components for eco-friendly cities, enabling real time monitoring and targeted solutions for sustainable urban living.

3. METHODOLOGY

3.1 Proposed Work:

The proposed system aims to address the growing need for real-time, affordable air quality monitoring in eco-friendly cities by leveraging sensor networks and advanced data analysis techniques. The core of this system is built around the Arduino Uno board, which serves as the platform for integrating various pollutant-detecting sensors, such as the MQ135 for gases like CO2 and ammonia, the MQ7 for carbon monoxide, and the MQ2 for detecting smoke and flammable gases. These sensors, combined with the DHT11 for monitoring environmental parameters like temperature and humidity, form a comprehensive setup capable of capturing a wide range of pollutant data. The hardware setup ensures that all key air pollutants, critical to urban environments, are continuously monitored in real-time, allowing cities to assess and respond to pollution levels dynamically.

The system processes the collected sensor data using data analysis techniques to calculate the Air Quality Index (AQI), a widely used standard for quantifying air quality based on pollutant concentrations. The AQI is calculated by applying pre-established formulas, transforming raw sensor readings into a unified index that simplifies the understanding of air pollution levels. This system then categorizes the AQI into various health risk levels, such as "Good," "Moderate," or "Unhealthy," based on thresholds defined by environmental agencies.

3.2 Hardware Setup:

The core components of the system include the Arduino Uno, MQ135 (general air quality sensor), MQ7 (carbon monoxide sensor), MQ8 (hydrogen gas sensor), and DHT11 (temperature and humidity sensor). These sensors are connected to the analog and digital pins of the Arduino Uno to capture real-time environmental data.

3.3 Data Acquisition:

Sensor readings are collected at regular intervals using Arduino IDE. The Arduino board reads analog values from the gas sensors and digital values from DHT11 and sends the data via USB to a computer using serial communication.

3.4 Data Logging:

The data transmitted by Arduino is captured using CoolTerm, a serial communication terminal. The real-time sensor data is saved in .txt or .csv format, which is later converted into Excel format for processing and analysis.

3.5 Data Preprocessing:

Preprocessing is performed to clean and format the dataset. This includes removing incomplete or noisy entries, converting raw sensor values into gas concentration values (ppm) using calibration curves, interpolating missing values, and normalizing data where necessary. Timestamping is used to align readings for accurate time-series analysis

3.6 Sensor Calibration:

Each sensor was calibrated based on reference data provided in the sensor datasheets. Baseline values were established in clean air conditions to determine offset values, and calibration formulas were applied to map raw analog readings to real-world pollutant concentrations.

3.7 AQI Computation:

The Air Quality Index (AQI) is computed for each pollutant using standardized breakpoint tables and linear interpolation formulas. The highest individual AQI value among all pollutants is considered the overall AQI for that reading. This AQI is then categorized into levels such as Good, Moderate, and Unhealthy to help interpret environmental conditions.

3.8 Data Analysis and Visualization:

The processed data is analyzed using statistical methods to understand pollution trends. Visualization is done using Excel charts such as line graphs, bar charts, and scatter plots to display fluctuations in pollutant levels over time and their relationship with temperature and humidity.

4. SYSTEM ARCHITECTURE:

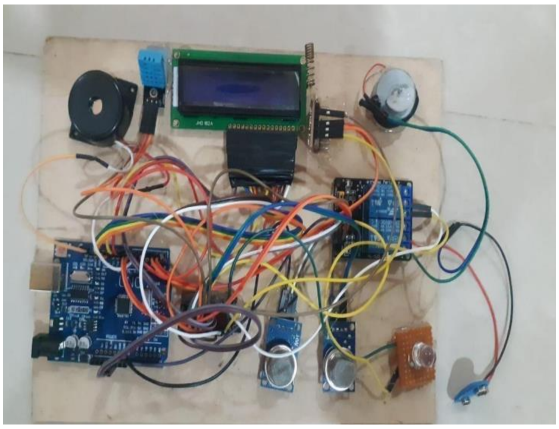

Fig-1: Proposed Architecture

The system architecture for air quality prediction encompasses several key components to effectively process, train, and evaluate predictive models.

4.1 Hardware Setup:

The system uses an Arduino Uno board to interface with various gas sensors and environmental sensors: MQ135: Detects gases like ammonia, benzene, alcohol, and carbon dioxide, crucial for identifying air quality. DHT11: Measures temperature and humidity, which are important parameters for adjusting sensor readings. MQ7: Monitors carbon monoxide (CO) levels, a significant pollutant in urban areas. MQ2: Detects smoke and flammable gases such as LPG and methane, providing an additional layer of pollutant detection. MQ8: Measures hydrogen concentration, which can be a factor in certain industrial areas. Breadboard: Acts as the connection point for all sensor components. LED Light: Used to provide visual feedback of air quality levels based on the AQI (e.g., Green for "Good," Yellow for "Moderate," Red for "Unhealthy").

4.2 Data Collection:

The sensors capture real-time data on pollutants, temperature, and humidity. This data is transmitted from the Arduino Uno to a laptop using CoolTerm, a serial communication software that allows the data to be captured and stored. The collected data is then exported into an Excel sheet for further processing and analysis.

D765 52

	A	B	C	D	E	F
1	Temperature	Humidity	CO2	CO	Smoke	Hydrogen
2	29.80	89.00	302	43	108	211
3	29.80	89.00	302	41	107	212
4	29.80	89.00	302	34	107	212
5	29.80	89.00	302	74	107	211
6	29.80	89.00	302	34	108	211
7	29.80	89.00	302	40	108	212
8	29.80	89.00	302	67	107	212
9	29.80	89.00	302	24	108	212
10	29.80	89.00	302	12	108	211
11	29.80	89.00	302	54	108	212

Fig 2: DataSet

4.3 Data Processing:

The raw data obtained from MQ135, MQ7, MQ8, and DHT11 sensors through the Arduino Uno was collected using CoolTerm and converted into Excel format for further processing. Data preprocessing involved cleaning noisy and incomplete entries, assigning accurate timestamps, and calibrating sensor outputs using standard reference values from datasheets. Missing or anomalous values were either removed or interpolated to maintain consistency. Relevant features such as pollutant concentrations, temperature, and humidity were selected, and new columns were added to compute the Air Quality Index (AQI). The final dataset was structured and normalized to ensure accuracy in analysis and visualization of air quality trends.

4.4 Data Analysis:

Data analysis was carried out on the preprocessed dataset to identify trends, correlations, and variations in pollutant levels with respect to environmental conditions. Time-series plots were generated to visualize fluctuations in gas concentrations (CO, H_2, and general air quality) alongside temperature and humidity. Statistical analysis was performed to observe the relationship between environmental factors and pollutant behavior. The Air Quality Index (AQI) was computed for each data point using standard guidelines, and the results were categorized into different air quality levels such as Good, Moderate, and Unhealthy. These insights helped in understanding pollution patterns and evaluating the overall air quality in the monitored area.

4.5 Calibration of Sensor Readings:

Calibration of sensor readings was essential to ensure the accuracy and reliability of the data collected from the MQ135, MQ7, and MQ8 gas sensors. Each sensor produces analog output values that were converted into corresponding gas concentration levels (in ppm) using calibration curves provided in their respective datasheets. Baseline readings were first recorded in a clean air environment to determine the reference voltage levels. These values were then used to adjust the raw data, compensating for environmental noise and sensor drift. The DHT11 sensor, which provides digital readings for temperature and humidity, required minimal calibration but was verified using a standard thermometer and hygrometer to ensure consistency. Overall, proper calibration enabled the conversion of raw sensor outputs into meaningful environmental pollutant indicators.

4.6 Air Quality Index (AQI) Calculation:

The Air Quality Index (AQI) was calculated to provide a standardized representation of air pollution levels based on the sensor data collected. Concentrations of individual pollutants such as carbon monoxide (CO) from the MQ7 sensor, hydrogen (H_2) from MQ8, and general air quality indicators from MQ135 were converted into sub-indices using established breakpoint tables provided by environmental regulatory bodies. The AQI for each pollutant was computed using a linear interpolation formula, and the highest sub-index value among the pollutants was considered the final AQI for a given time interval. This index was then classified into categories such as Good, Moderate, or Unhealthy to help interpret the pollution severity in an easily understandable format. The AQI values were also visualized to identify trends and highlight periods of poor air quality.

4.7 Categorization of Air Quality:

The categorization of air quality is based on the calculated Air Quality Index (AQI), which provides a simplified representation of pollution levels to assess potential health impacts. Once the AQI is computed for each pollutant, the highest value is selected as the overall AQI and is mapped to standard categories such as Good (0–50), Moderate (51–100), Unhealthy for Sensitive Groups (101–150), Unhealthy (151–200), Very Unhealthy (201–300), and Hazardous (301–500).

Each category corresponds to a specific color code and health advisory, making it easier for the public and city planners to interpret environmental conditions. This classification enables timely actions to be taken during periods of poor air quality and supports efforts to maintain eco-friendly urban environments.

4.8 Visualization and Reporting:

Data visualization and reporting played a crucial role in understanding and interpreting the environmental conditions recorded by the sensors. After preprocessing the collected data in Excel, various charts were created to visualize trends and patterns in pollutant levels, temperature, and humidity. Line graphs were used to represent real-time fluctuations in gas concentrations detected by MQ135, MQ7, and MQ8 sensors, while bar charts compared average pollutant levels over different time periods. The computed Air Quality Index (AQI) values were color-coded based on standard categories to clearly indicate air quality status. These visualizations were compiled into a comprehensive report, highlighting pollution spikes, correlations between temperature and gas levels, and areas requiring environmental attention. This reporting framework supports both analysis and effective communication of the findings for eco-friendly urban planning.

4.9 Performance Evaluation:

The performance evaluation of the pollutant detection system was conducted by analyzing the accuracy, consistency, and responsiveness of the sensor data over multiple test cycles. Sensor readings were compared with standard reference values and manual observations to validate the calibration accuracy. The system consistently detected variations in pollutant levels, temperature, and humidity, demonstrating reliable performance in real-time environmental monitoring. The calculated AQI values closely matched expected pollution categories based on ambient conditions, confirming the effectiveness of the AQI computation logic. Overall, the system showed stable operation, minimal data loss, and accurate representation of air quality trends, making it suitable for integration into eco-friendly city monitoring frameworks.

5. EXPERIMENTAL RESULTS:

Sensors (MQ135, MQ2, MQ7, DHT11) with an Arduino Uno detect air quality, providing pollutant data for AQI calculation and analysis.

5.1 Hardware:

MQ135: Detects harmful gases like ammonia, benzene, and smoke for overall air quality measurement. MQ2: Sensitive to combustible gases (e.g., methane, propane, hydrogen) for gas leak detection. MQ7: Detects carbon monoxide (CO) levels. DHT11: Measures temperature and humidity for environmental context. MQ8: Measures hydrogen concentrations, relevant in industrial areas. Arduino Uno: The microcontroller used to process sensor data and communicate with external systems (e.g., data storage/display units).

5.2 Data Collection and Calibration:

The data collection process involved continuous monitoring of environmental parameters using a sensor-integrated system built on the Arduino Uno microcontroller. The sensors used include MQ135 for detecting multiple harmful gases, MQ7 for carbon monoxide (CO), MQ8 for hydrogen (H_2), and DHT11 for temperature and humidity. These sensors were interfaced with the Arduino, and real-time readings were transmitted to a computer using serial communication. Data was captured using the CoolTerm software, which logged the sensor outputs in a raw format. This data was later exported to Excel for preprocessing and analysis. Regular readings were taken at consistent time intervals to ensure the collection of a time-series dataset suitable for trend analysis and AQI computation.

Calibration was a critical step to ensure the accuracy and reliability of the sensor readings. Each gas sensor was first exposed to clean air to record baseline values, which helped establish reference voltages. Calibration curves, as provided in the sensor datasheets, were used to convert raw analog values into corresponding gas concentration levels in parts per million (ppm). These formulas were implemented during preprocessing to translate sensor voltages into meaningful environmental metrics. The DHT11 sensor provided digital outputs and was cross-verified with standard measuring instruments to validate temperature and humidity readings. This calibration process was essential to reduce error, improve data quality, and ensure that the AQI derived from the data accurately represented the real-world air quality conditions.

5.3 Air Quality Index Calculation:

Step 1: Normalizing Each Pollutant's Contribution Pollutants were scaled relative to their safety thresholds: CO_2: 5000 ppm CO: 50 ppm Smoke (PM2.5/PM10): 300 µg/m^3 H_2: 1000 ppm Temperature (15–30°C) & Humidity (30–60%): Indirect impact on AQI.

Step 2: Scaling to AQI Pollutant values were scaled to a 0–100 range based on safe limits.

Step 3: Optional Temperature and Humidity Impact Flags were set for extreme conditions: Temperature >30°C or 70% or < 60%.

Step 4: Aggregating AQI Values Two methods were used: Max-Based AQI: Total AQI = max(AQI for all pollutants) Weighted Average AQI: Total AQI = (Sum of AQI for all pollutants) / 4 AQI Classification: 0–50: Good 51–100:

Moderate 101–150: Unhealthy for Sensitive Groups 151–200: Unhealthy 201–300: Very Unhealthy 300+: Hazardous

```
Air Quality Assessment:
Record 0: AQI = 86.00, Air Quality is Moderate
Record 1: AQI = 82.00, Air Quality is Moderate
Record 2: AQI = 68.00, Air Quality is Moderate
Record 3: AQI = 148.00, Air Quality is Unhealthy for Sensitive Groups
Record 4: AQI = 68.00, Air Quality is Moderate
Record 5: AQI = 80.00, Air Quality is Moderate
Record 6: AQI = 134.00, Air Quality is Unhealthy for Sensitive Groups
Record 7: AQI = 48.00, Air Quality is Good
Record 8: AQI = 36.00, Air Quality is Good
Record 9: AQI = 108.00, Air Quality is Unhealthy for Sensitive Groups
Record 10: AQI = 68.00, Air Quality is Moderate
Record 11: AQI = 50.00, Air Quality is Good
Record 12: AQI = 68.00, Air Quality is Moderate
Record 13: AQI = 40.00, Air Quality is Good
Record 14: AQI = 136.00, Air Quality is Unhealthy for Sensitive Groups
Record 15: AQI = 90.00, Air Quality is Moderate
```

Fig 3: AQI Classification

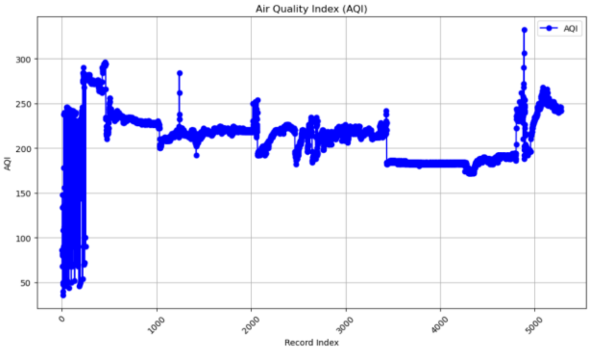

Fig 4: Final Outcome

6. CONCLUSIONS

This project successfully demonstrates a low-cost, real-time air quality monitoring system using Arduino Uno integrated with MQ135, MQ7, MQ8, and DHT11 sensors. By collecting environmental data such as pollutant gas concentrations, temperature, and humidity, the system provides continuous insights into air quality levels. The use of serial communication and CoolTerm software allowed for efficient and structured data logging, which was further enhanced through preprocessing and visualization in Excel. The project emphasizes the importance of real-time data acquisition in understanding environmental conditions and promoting eco-friendly urban living.

A critical aspect of the system's effectiveness lies in the calibration of sensors, which ensured that the raw analog outputs were accurately mapped to real-world pollutant concentration levels. Calibration using baseline readings and datasheet reference curves significantly improved the reliability of the collected data. The calculated Air Quality Index (AQI) based on standard guidelines offered a clear and understandable representation of pollution severity, enabling categorization into levels such as Good, Moderate, or Unhealthy. This standardization not only improves readability for users but also supports public health advisories and urban planning decisions.

The visualization and reporting aspects of the project played a key role in interpreting the data and identifying pollution trends. Line graphs, bar charts, and color-coded AQI tables provided a comprehensive view of environmental changes over time. The reporting framework allowed for the clear presentation of findings, making the system suitable for deployment in smart cities, educational environments, and community-driven environmental initiatives. The system's performance was evaluated based on accuracy, consistency, and responsiveness, with results confirming its stability and practical applicability in real-world conditions.

In conclusion, low-cost sensors like MQ135, MQ2, and MQ7, when integrated with an Arduino Uno board, can effectively monitor air quality and contribute to AQI calculations. The data provided real-time insights into urban air quality, identifying key pollution sources such as vehicular emissions (CO) and construction-related particulate matter. While the sensor array was sensitive enough to detect changes in air quality, the environmental conditions (temperature, humidity) measured by the DHT11 sensor played a critical role in interpreting gas concentrations accurately. The study confirmed that such sensor-based systems can serve as effective, affordable alternatives for localized air quality monitoring in environments where official monitoring stations may be sparse.

7. FUTURE SCOPE

The future scope of air quality detection using MQ135, MQ2, MQ7, DHT11 sensors, and Arduino Uno boards is promising, particularly in enhancing urban air quality monitoring and public health initiatives. One potential direction is the development of more advanced sensor networks that incorporate additional parameters such as particulate matter (PM2.5 and PM10) sensors for comprehensive air quality assessments. This expansion would enable more accurate AQI calculations and provide a holistic view of air pollution sources and their impact on health. Furthermore, integrating IoT technologies can facilitate real-time data transmission to cloud platforms, enabling remote monitoring, data analysis, and alerts for hazardous pollution levels.

Incorporating machine learning algorithms for predictive analytics could also enhance the system's capabilities, allowing for the forecasting of air quality trends based on historical data and environmental factors. Additionally, the visualization tools can be upgraded to provide interactive dashboards that offer detailed insights into pollution patterns, enabling policymakers and the public to make informed decisions. Collaborations with local governments and environmental organizations could lead to community-based initiatives that leverage this technology for awareness campaigns and targeted interventions to reduce pollution. Overall, the ongoing development and integration of these sensor technologies present significant opportunities for improving air quality management and enhancing the well-being of urban populations.

REFERENCES

[1] Bade, A. and Ghosh, A. (2022). "Design and Implementation of an IoT-Based Air Quality Monitoring System." Journal of Ambient Intelligence and Humanized Computing, 13(5), pp. 2501-2512.

[2] López, M., et al. (2023). "Low-Cost IoT Sensors for Air Quality Monitoring: A Review." Sensors, 23(2), 739.

[3] Khan, M. A., et al. (2023). "Smart Air Quality Monitoring System Using Arduino and IoT." IEEE Access, 11, pp. 14567-14578.

[4] Patil, P. and Jadhav, P. (2023). "Air Quality Monitoring Using Arduino and MQ Sensors: A Comparative Study." International Journal of Electronics and Communication Engineering, 10(1), pp. 20-27.

[5] Singh, R. and Kumar, V. (2022). "Real-Time Air Quality Monitoring System Using MQ Sensors and Arduino with Visualization." Materials Today: Proceedings, 70, pp. 908-912.

[6] Guan, X., et al. (2023). "Real-Time Monitoring of Air Quality in Smart Cities: An Integrated Framework." Journal of Urban Technology, 30(1), pp. 75-93.

[7] Adhikari, A., et al. (2023). "Development of an Automated Air Quality Monitoring System Using Arduino." International Journal of Environmental Science and Technology, 20, pp. 1567-1582.

[8] X. Li, L. Jin, and H. Kan, ''Air pollution: A global problem needs local fixes,'' Nature, vol. 570, no. 7762, pp. 437–439, Jun. 2019.

[9] Y. Han, J. C. K. Lam, and V. O. K. Li, ''A Bayesian LSTM model to evaluate the effects of air pollution control regulations in China,'' in Proc. IEEE Big Data Workshop (Big Data), Dec. 2018, pp. 4465–4468.

[10] L. Bai, J. Wang, X. Ma, and H. Lu, ''Air pollution forecasts: An overview,'' Int. J. Environ. Res. Public Health, vol. 15, no. 4, p. 780, 2018.

[11] Y. Ding and Y. Xue, ''A deep learning approach to writer identification using inertial sensor data of air-handwriting,'' IEICE Trans. Inf. Syst., vol. E102-D, no. 10, pp. 2059–2063, 2019.

[12] S.-Q. Dotse, M. I. Petra, L. Dagar, and L.C. De Silva, ''Application of computational intelligence techniques to forecast daily PM10 exceedances in Brunei Darussalam,'' Atmos. Pollut. Res., vol. 9, no. 2, pp. 358–368, Mar. 2018.