Journal of Information Systems Engineering and Management
2025, 10(34s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

Vayu – An IoT Powered Indoor Air Quality Monitoring System

Dr. M. Sunitha[1], Dr. T. Adilakshmi[2], , Manne Lakshmi Prasanna[3]
[1]*Department of Computer Science Engineering, Vasavi College of Engineering, Telangana, India.*
[2]*Department of Computer Science Engineering, Vasavi College of Engineering, Telangana, India.*
[3]*Department of Computer Science Engineering, Vasavi College of Engineering, Telangana, India.*

ARTICLE INFO

Received: 24 Dec 2024

Revised: 12 Feb 2025

Accepted: 26 Feb 2025

ABSTRACT

Air pollution is a growing global concern, impacting public health, ecosystems, and overall quality of life. Effective monitoring and forecasting are crucial for mitigating its effects and enabling proactive interventions. To address this challenge, we developed Vayu, an advanced IoT and Machine Learning-based Air Quality Monitoring System (AQMS) designed to provide real-time insights and predictive analysis. The system integrates and Arduino UNO, ESP8266 Wi-Fi Module, DHT11, MQ2 and MQ135 sensors to continuously measure key air pollutants. This data is transmitted to ThingSpeak for storage and visualization through a dedicated dashboard, which also provides Pushover notifications for timely alerts. To enhance forecasting accuracy, Machine Learning models such as ARIMA, LSTM and others analyses historical trends to predict future air quality levels. By combining real-time monitoring with intelligent forecasting, Vayu empowers individuals, policymakers and organizations with the real time data needed to take informed actions, ultimately contributing to a cleaner and healthier environment.

Keywords: air quality, application, GUI dashboard, Internet of Things, machine learning, vayu

INTRODUCTION

In 2024, World Health Organization (WHO) reported that its air quality database includes data from 7,182 human settlements in over 120 member states. The results highlighted a serious issue, indicating that 99% of the global population is inhaling air with pollutant concentrations surpassing WHO-recommended thresholds [1][8]. Air pollution has become one of the leading global health threats, leading to an increased prevalence of chronic illnesses such as cardiovascular disease, stroke, diabetes, pulmonary cancer, and Chronic Obstructive Pulmonary Disease (COPD) induced by atmospheric contamination [5][9].

Several factors contribute to the decline in air quality, with industrial emissions being a major contributor. Key pollutants such as Sulfur Dioxide (SO_2), Particulate Matter (PM), Nitrogen Oxides (NO_x), Carbon Monoxide (CO), and various VOCs are released into the atmospheric air, severely impacting air quality. Additionally, the high usage of vehicles, especially in urban areas, is a significant source of pollution. The continued reliance on coal and other fossil fuels for power generation and heating further exacerbates the problem [6]. This widespread pollution is not only a major environmental concern but also a critical public health issue that requires immediate action.

To address the growing concerns of air pollution, this project utilizes an Arduino microcontroller, an ESP8266 Wi-Fi module, and a variety of sensors to monitor air quality. The MQ135 sensor is utilized to assess the Air Quality Index (AQI), while the MQ2 sensor detects the levels of contaminants, including ammonia (NH_3), Methane (CH_4), Carbon Dioxide (CO_2), Propane (C_3H_8) and others [1]. The DHT11 sensor is used for measuring temperature and humidity levels. The sensor data is then sent to ThingSpeak, a cloud-based platform that provides real-time monitoring of air quality Through this system, users can access air quality data via an application and receive live alerts if pollutants levels exceed permissible limits.

The key innovation of this study is the development of a cloud-supported IoT monitoring framework, which not only provides real-time data but also incorporates Machine Learning (ML) models for forecasting future air quality levels [2]. This system aims to offer a comprehensive solution to monitor, predict, and manage air pollution, ultimately supporting public health and environmental protection.

Journal of Information Systems Engineering and Management
2025, 10(34s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/
Research Article

LITERATURE SURVEY

Dr. Anitha N et al. [1] developed an IoT system using Arduino, MQ-135 and Optical Dust Sensors to detect pollutants like CO_2, NH_3, smoke, and dust. When AQI exceeds 400 ppm, the system triggers LED and buzzer alerts, displaying real-time data on an LCD screen. Tested in a kitchen environment, it effectively monitored pollution variations. This low-cost, scalable solution provides real-time insights to help manage air quality and reduce health risks.

Vilas Kisanrao Tembhurne et al. [2] proposed an IoT system that tracks CO, SO_2, PM10, humidity, and temperature in real time, storing data on the cloud for remote access. The system uses an MQ-135 gas sensor to detect CO_2, NH_3, benzene, and smoke. A Raspberry Pi 3 and Arduino process and transmit the collected data, with an MCP3008 ADC chip converting analog sensor readings to digital format. The system features an LCD screen to display real-time air quality data and triggers alerts via an LED and buzzer when pollution levels exceed safe thresholds. This cost-effective solution enhances public awareness of air quality, with future improvements planned for AI-based predictive analytics.

Kamepalli Alekhya et al. [3] introduced an IoT system to track air pollution in real time and issue alerts when harmful gases exceed safe levels. The system integrates MQ135, MQ6, MQ2, and MQ7 sensors to detect CO_2, NH_3, NOx, benzene, smoke, and flammable gases. An ATmega328P microcontroller processes sensor data, which is shown on a screen and transmitted to a web server via an ESP8266 Wi-Fi module. If pollution levels exceed 400 PPM, the system triggers an alarm via a buzzer and sends an alert message through a GSM module to a smartphone. This low-cost, scalable solution enables remote air quality monitoring, aiding pollution control efforts. Future improvements include SMS/app notifications and expanded sensor capabilities for enhanced accuracy.

Farida A. Ali et al. [4] created an IoT system for indoor air quality to track real-time air pollution levels in residential and commercial buildings. The system uses MQ135 gas sensor, DHT11 temperature and humidity sensor, and GP2Y1014AU0F dust sensor to measure CO, NO_2, VOCs, PM10, temperature, and humidity. An ESP32 NodeMCU microcontroller processes the data and transmits it to a ThingSpeak cloud server via Wi-Fi. When pollution levels exceed set thresholds, the system sends alerts via WhatsApp using Twilio software and displays real-time air quality data on a web interface. The system ensures continuous air quality assessment, enabling users to take preventive actions against indoor air pollution. This cost-effective and scalable solution is designed to enhance public health by reducing exposure to harmful air pollutants.

Souptik Das et al. [5] built an IoT system to track air pollution levels in real time. The system uses an MQ-135 gas sensor to detect CO_2, NH_3, alcohol, and smoke, with data processing handled by a NodeMCU ESP8266 Wi-Fi module. The collected data is transmitted to a smartphone via the Blynk IoT platform, allowing users to remotely monitor air quality. A remote-controlled robot enhances system mobility, enabling air quality assessment in hazardous or hard-to-reach areas. The robot is powered by a 433 MHz RF transmitter-receiver, an L298N motor driver, and BO motors, ensuring flexible movement. Additionally, an LED warning system provides visual alerts.

PROPOSED IDEA

Our proposed solution seamlessly integrates Machine Learning (ML) and IoT to create an advanced Air Quality Monitoring and Forecasting System. The system consists of a compact IoT device equipped with multiple environmental sensors, including a DHT11 sensor for measuring temperature and humidity, along with MQ135 and MQ2 gas sensors for detecting pollutants such as CO_2, NH_3, NO_2, Methane, and Propane. Designed specifically for indoor environments, this device continuously captures real-time air quality data to provide accurate and up-to-date information.

The ESP8266 Wi-Fi chip serves as a bridge between the hardware and the ThingSpeak cloud platform, enabling the transmission of collected sensor data for monitoring and analysis. ThingSpeak efficiently stores and organizes the data, enabling real-time visualization, trend analysis, and historical tracking. The stored data is periodically extracted in CSV format and used as input for machine learning models to predict upcoming atmospheric quality patterns.

To enhance predictive capabilities, the system employs multiple ML models, including XGBoost, LSTM, ARIMA, VAR, and a Hybrid model (XGBoost + LSTM + MultiOutputRegressor). These models analyse historical air quality data, capture intricate patterns, and predict fluctuations in AQI (Air Quality Index) for proactive decision-making.

Journal of Information Systems Engineering and Management
2025, 10(34s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/
Research Article

A Python-based analytical framework is developed using the ThingSpeak API to automate data retrieval, perform in-depth analysis, and generate live alerts. These alerts notify users about potential air quality deterioration, allowing them to take necessary precautions in advance. Additionally, a mobile application and web-based dashboard enable end users to monitor data in real time, analyse historical trends, and predictive insights, making the system efficient, user-friendly, and impactful for pollution management and public health protection.

Frame Work

The proposed system consists of an Arduino microcontroller, an ESP8266 Wi-Fi module, and three sensors—MQ135 and MQ2 for measuring air quality index (AQI) and pollutant concentration, and DHT11 for monitoring humidity and temperature as shown in Figure. 1.

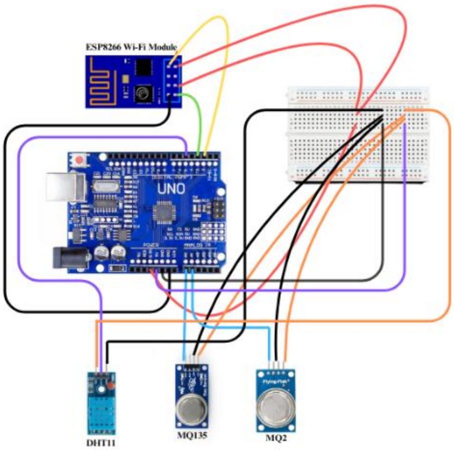

Figure. 1 Configuration of Physical Components

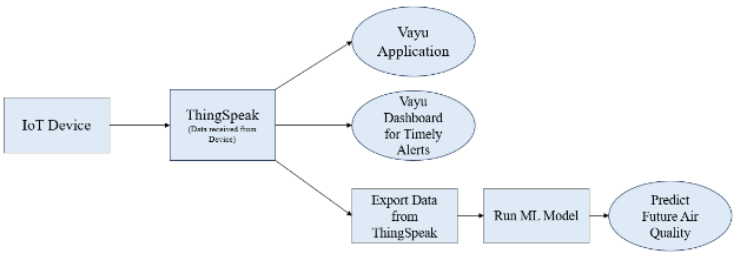

Figure. 2 Visual Representation of the Proposed System's Workflow

The Arduino is programmed using the Arduino IDE to fetch real-time sensor data, while the ESP8266 module provides internet connectivity. The acquired data is relayed to the ThingSpeak cloud platform utilizing the Write API Key of the specified channel, as illustrated in Figure. 2. Once uploaded, the data is stored and visualized on ThingSpeak, allowing users to monitor air quality trends through graphical representations. This system enables real-time tracking of environmental conditions and can be extended to send alerts or integrate with mobile applications for enhanced monitoring and user accessibility

Hardware And Software Requirements

For implementation of the proposed model, the required software and hardware include**:**

- MQ135 Sensor: It is a versatile sensor for air quality and one kind of MQ series gas sensor that has high sensitivity to a variety of gases present in ambient air like Ammonia (NH_3), Sulfur(S), Benzene(C_6H_6), CO_2

Journal of Information Systems Engineering and Management
2025, 10(34s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/
Research Article

and other poisonous gases. The sensor is subjected to operate at 5V power supply with 150mA consumption. The MQ135 sensor comes with 4 pins: VCC, GND(Ground), and separate analog and digital output terminals. It provides a digital logic output like 1(HIGH) when toxic gases concentration in the environment reaches the thresholds or 0(LOW) when no gas is detected in the air. Due to its high sensitivity and fast response time, it provides timely sensor data as soon as possible.

- MQ2 Sensor: It is an electronic sensor used for sensing the concentration of air contaminants, that includes LPG, Smoke, Methane (CH_4), Propane (C_3H_8), CO, Hydrogen. It exhibits high sensitivity and is capable of detecting a diverse spectrum of gases. It is a MOS (Metal Oxide Semiconductor) sensor also known as Chemiresistors. The MQ2 sensor operated on 5V power supply and consumes approximately 800mW. It is equipped with 4 pins: VCC, GND, an analog output pin and a digital output pin. It additionally offers a binary indication (1 or 0) of the present of dangerous gases but also an analog representation of their concentration in air.
- DHT11 Sensor: This is basic and cost-effective digital sensor for sensing Temperature and Humidity. It can be integrated with any microcontrollers such as Arduino, NodeMCU, Raspberry Pi and many more. The sensor provides temperature readings in degree Celsius, from 0 to 50 °C, and measures relative humidity as a percentage, from 20% to 90% RH.
- Arduino Board: It is an inexpensive, adaptable, easy to use and user friendly open-source microcontroller built around the ATmega328P, which can be seamlessly integrated into numerous interactive electronic applications, providing automation for smart solutions.
- ESP8266 Wi-Fi Module: A compact and economical microchip, it comes with built-in Wi-Fi support and operates on low power. It is compatible with popular development platforms kike Arduino, NodeMCU and others. It offers several GPIO pins for controlling LEDs, sensors and other peripherals. It has a built-in support for TCP/IP protocols, making it easy to integrate it into web-based applications. It facilitates serial communication, typically using UART, allowing easy connection to microcontrollers like Arduino.
- ThingSpeak: It serves as a reliable and efficient open-source IoT framework, allowing users to aggregate, process, and display data from sensors and interconnected devices through cloud integration. It offers charts, graphs and dashboards for visualizing data in real-time. It allows advanced data analysis and mathematical modeling through MATLAB and provides REST API for sending and retrieving data, making it easy to integrate with IoT devices (e.g., Arduino, ESP8266). The sensor data is represented un the format of line chart and can also be downloaded in CSV format.
- Arduino IDE Software: An essential tool for embedded development, it offers an open-source environment for programming, compiling and flashing code onto Arduino microcontroller and compatible boards. It features a simple text editor for developing Arduino programs using C/C++ programming language. It runs on multiple OS including Windows, macOS, Linux and others. It includes built-in libraries, Serial monitor for debugging, also allows users to select the correct Arduino board model and the serial port to upload the code. It provides a sketchbook feature to store all the saved projects and making it easy to organize and access our projects.

Machine Learning Models

Machine Learning models are widely used for air quality forecasting because they can efficiently analyze complex patterns in historical data, adapt to changing environmental trends, and enhance prediction accuracy. Unlike traditional statistical models, ML techniques such as XGBoost, LSTM, and hybrid approaches can effectively capture non-linear relationships between various pollutants and external factors like weather conditions, traffic emissions, and industrial activities. These models continuously improve by learning from new data, making them highly suitable for both short-term and long-term forecasting.

By providing more accurate predictions, ML models enable proactive pollution control measures, support environmental policy-making, and help safeguard public health by issuing timely alerts and recommendations.

Journal of Information Systems Engineering and Management
2025, 10(34s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/
Research Article

- ARIMA Model

It is a prevalent statistical approach for time series prediction, identifying patterns in past observations and extrapolating future trends based on underlying dependencies, making it useful for predicting air quality and other time-dependent patterns. The forecasted AQI in ppm is illustrated in Figure. 3. The ARIMA model consists of 3 main components:

AutoRegressive (AR) Component – Order (x): It represents the relationship between past and present values.

$$Y_t = \emptyset Y_{t-1} + \emptyset Y_{t-2} + - - + \emptyset p Y_{t-p} + \epsilon_t \qquad (1) Y_t = \emptyset Y_{t-1} + \emptyset Y_{t-2} + - - - - - - + \emptyset p Y_{t-p} + \epsilon_t$$

Where:

Y_t = Current value of the time series.

$\emptyset$ = AR coefficients.

ϵ_t = Error term.

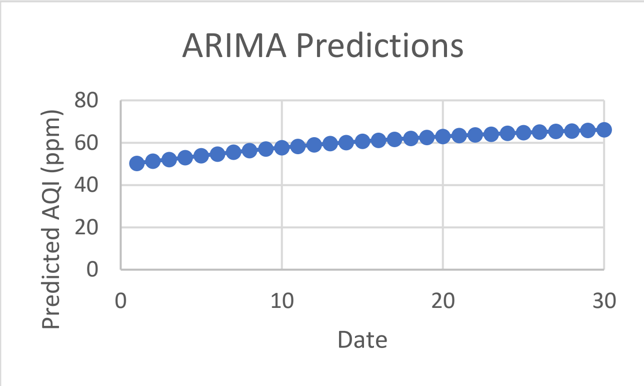

Figure. 3 Forecasted AQI Values Using the ARIMA Model

Integrated (I) Component – Order (y): It makes the time series stationary by differencing (removing trends).

$$Y_t{}' = Y_t - Y_{t-1} \qquad (2)$$

Moving Average (MA) Component – Order (z): It uses past error terms to improve predictions.

$$Y_t = \epsilon_t + \theta_1 \epsilon_{t-1} + \theta_2 \epsilon_{t-2} + - - - - - - - + \theta_q \epsilon_{t-q} \quad (3)$$

Where θ are the MA coefficients.

So, the final general ARIMA model (ARIMA(x,y,z)) equation that combines all three components:

$$Y_t{}' = \emptyset_1 Y'_{t-1} + - - + \emptyset_p Y'_{t-p} + \epsilon_t + \theta_1 \epsilon_{t-1} + - - + \theta_q \epsilon_{t-q} \; (4)$$

Where $Y_t{}'$ is the differenced series.

- LSTM Model

Long Short-Term Memory (LSTM) is an advanced form of Recurrent Neural Networks (RNNs) designed for handling sequential information by preserving essential dependencies across prolonged temporal sequences. Unlike conventional RNNs, LSTMs mitigate the vanishing gradient issue, enhancing their reliability for temporal predictions. Forecasting air quality poses a challenge due to its dependence on multifaceted influences, including meteorological patterns, industrial emissions, and cyclical environmental shifts. LSTM models can capture these dependencies and provide accurate forecasts.

An LSTM network consists of memory cells that process and store relevant information. These cells regulate data flow using three gates:

Journal of Information Systems Engineering and Management
2025, 10(34s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/
Research Article

Forget Gate (FG): Identifies and eliminates obsolete retained information.

$$f\_t \;=\; \sigma(W_f.[h_{t-1}, x_t] + b_f \qquad (5)$$

Input Gate (IG): Regulates the incorporation of novel data into the cell state.

$$i_t \;=\; \sigma(W_i.[h_{t-1}, x_t] + b_i \qquad (6)$$

$$\widetilde{C_t} \;=\; tanh(W_c.[h_{t-1}, x_t] + b_c \qquad (7)$$

Update Cell State: Combines old memory and new information.

$$C_t = f_t.C_{t-1} + i_t.\widetilde{C_t} \qquad (8)$$

Output Gate (OG): Controls the amount of processes information that is passed as the final output.

$$o_t = \sigma(W_o.[h_{t-1}, x_t] + b_o) \qquad (9)$$

$$h_t = o_t.tanh(C_t) \qquad (10)$$

Whereas:

x_t → Input data at time step t.

h_t→ LSTM output at time step t, passed to the next step.

C_t →The internal memory of the LSTM that stores long-term dependencies.

f_t → Decides which past information to forget.

i_t → Controls how much new information at add.

o_t→ Determines what part of the memory is used as output.

$\widetilde{C_t}$→ A potential update to the memory, created using the input and hidden state.

σ → An activation function that squashes values between 0 and 1 to regulate gate operations.

tanh → A function that maps values to a range between -1 and 1, often used for scaling new information.

W_f, W_i, W_c, W_o→ These matrices store the learned parameters for the forget, input, candidate, and output gates.

b_f, b_i, b_c, b_o→ Bias values added to each gate to fine-tune their operations.

The forecasted AQI in ppm using LSTM model is shown in Figure. 4.

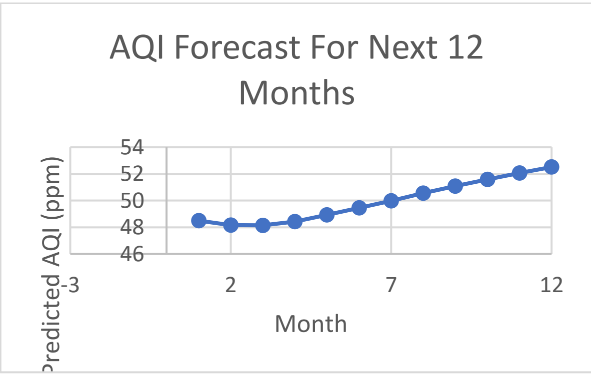

Figure. 4 Forecasted AQI Values Using the LSTM Model

- VAR Model

The Vector AutoRegression (VAR) model is a multivariate time series prediction approach that encapsulates relationships among multiple variables. Unlike single-variable models like ARIMA, VAR models interactions between

Journal of Information Systems Engineering and Management
2025, 10(34s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/
Research Article

different pollutants and environmental factors (e.g., AQI, CO2, NO2, NH3, Temperature, Humidity). The forecasted AQI in ppm using VAR model is shown in Figure. 5.

For example, AQI depends not only on past AQI values but also on pollutant levels and weather conditions. The VAR model captures these interdependencies, making it a suitable approach for air quality forecasting.

A VAR(p) where p is the lag order, mathematical formulation can be written as:

$$Y_t = Const + A_1 Y_{t-1} + A_2 Y_{t-2} + - - + A_p Y_{t-p} + e_t \quad (11)$$

Where:

Y_t = Vector of dependent variables like AQI, CO2, N02 at time t, Const = Intercept vector (constant).

A_p = Coefficient matrices (weights for past values).

e_t = Error term.

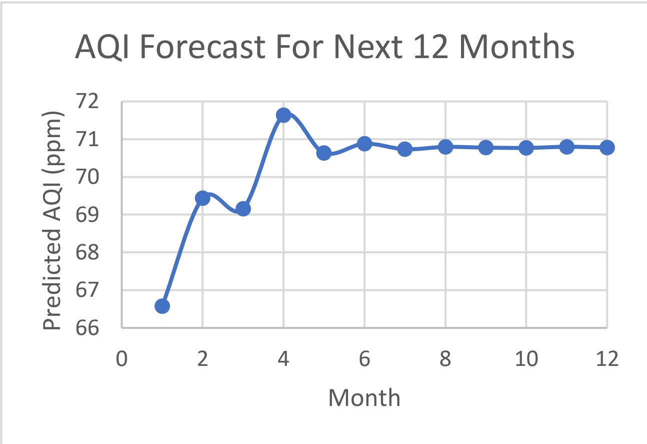

Figure. 5 Forecasted AQI Values Using the VAR Model

- XGBoost + MultiOutputRegressor Model

Extreme Gradient Boosting is a decision tree-based ensemble learning method that improves accuracy through boosting. It is an effective approach when predicting multiple correlated air quality parameters (AQI, CO2, NO2, etc.). XGBoost is a gradient boosting algorithm known for its efficiency, scalablity and high predictive power. XGBoost itself is designed for single-out regression. However, when predicting multiple air pollutants (AQI, NO2 ect.), we need a warapper model to handle multiple targets at once. MultiOutputRegressor allows XGBoost to predict multiple air quality metrics simultaneously.

For a dataset with N samples and M features, the model predicts multiple target variables Y based on input X.

$$Y_i = \sum_{k=1}^{K} f_k(X_i) + \varepsilon \qquad (12)$$

Where:

Y_i are the predicted parameters.

$f_k(X_i)$ is the ensemble of weak learners (trees).

K is total count of trees.

ε is the error term.

The Forecasted AQI in ppm using XGBoost Model is shown in Figure. 6.

Journal of Information Systems Engineering and Management
2025, 10(34s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/
Research Article

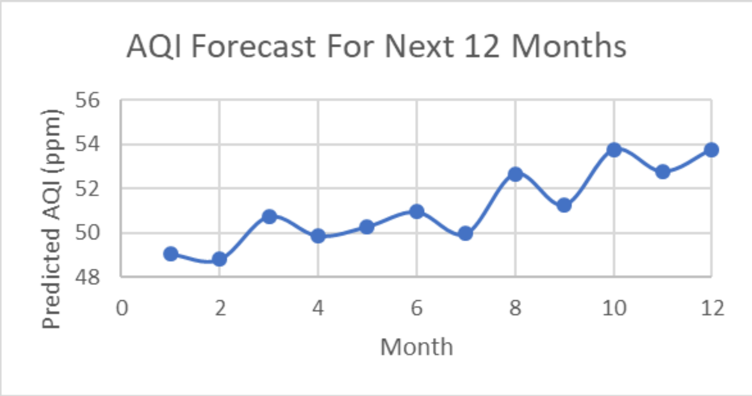

Figure. 6 Forecasted AQI Values Using the XGBoost Model

- Hybrid Model

The hybrid model combining XGBoost, LSTM, and MultiOutputRegressor leverages the strengths of each approach to enhance air quality prediction accuracy. XGBoost, a powerful gradient boosting algorithm, efficiently captures non-linear relationships in structured data, while LSTM, a deep learning model, excels at understanding temporal dependencies in time-series data. MultiOutputRegressor extends XGBoost's capabilities to predict multiple future AQI values simultaneously. The model workflow involves preprocessing data using StandardScaler, transforming time-series data into sequences for LSTM, and flattening features for XGBoost.

Both models are trained separately, and their predictions are blended using an averaging approach to obtain the final AQI forecast. This enables precise near-term (30-days) and extended-period (12-months) forecasts. The hybrid approach ensures better generalization, reduces overfitting, and effectively handles complex patterns in air quality variations. By combining XGBoost's ability to model feature interactions with LSTM's sequential learning, the model provides a robust solution for multi-step air quality forecasting, making it ideal for applications requiring high accuracy and long-term trend analysis.

The final predicted AQIN in the XGBoost + LSTM hybrid model is obtained using a weighted combination of the two models:

$$Y = \propto . Y_{XGB} + (1 - \propto). Y_{LSTM} \qquad (13)$$

Where:

Y_{XGB} is the AQI prediction from the XGBoost model.

Y_{LSTM} is the AQI prediction from the LSTM model.

$\propto$ is a blending weight factor (typically set to 0.5 for equal contribution).

The forecasted AQI in ppm for the next 12 months and 30 days is shown in Figure. 7 & Figure. 8 respectively.

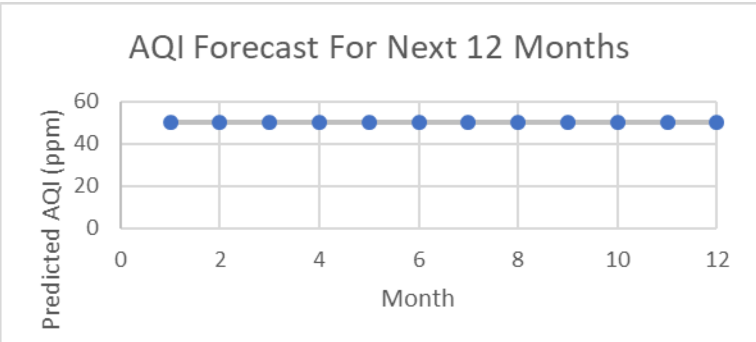

Figure. 7 Forecasted AQI Values Using the Hybrid Model

Journal of Information Systems Engineering and Management
2025, 10(34s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

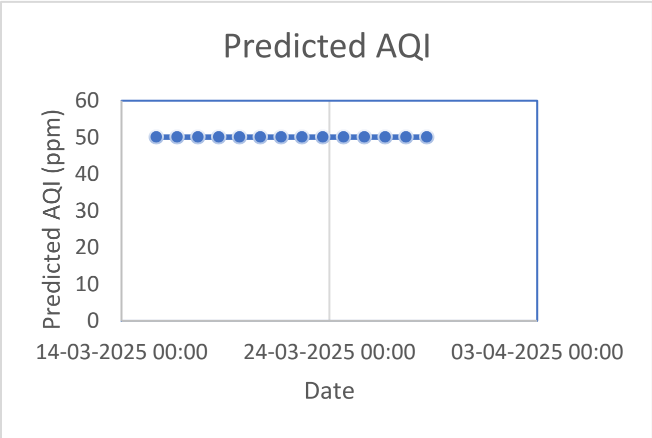

Figure. 8 Forecasted AQI Values Using the Hybrid Model

Vayu Application Interface

The application interface offers a real-time, intuitive, and visually engaging experience for monitoring air quality, ensuring users can easily interpret data at a glance as shown in Figure. 9. It features dynamic, interactive charts and graphs that update continuously based on live sensor readings, providing instant insights into air pollutant levels. Color-coded indicators help users quickly assess air quality conditions, with different shades representing safe, moderate, or hazardous levels. The interface tracks key air quality parameters, including AQI, Methane, Propane, CO2, NH3, and NO2, along with temperature and humidity trends. Designed for accessibility across devices, from smartphones to desktops, the responsive interface allows seamless monitoring anytime, anywhere, ensuring users stay informed and can take timely action when needed.

Figure. 9 Overview of the Vayu App's Design

Vayu GUI Dashboard and Timely Alerts

The Vayu system features a user-friendly graphical user interface (GUI) dashboard designed to enable live tracking of indoor atmospheric quality indicators. Accessible via a web browser, the dashboard displays current readings of various pollutants, including AQI, CO_2, NH_3, NO_2, methane, and propane, as well as temperature and humidity levels. Each parameter is color-coded based on predefined thresholds: green indicates normal levels, yellow signifies moderate concern, and red alerts to high concentrations that may pose health risks as shown in Figure. 10.

To enhance user awareness and safety, Vayu integrates live alert capabilities through Pushover notifications as shown in Figure. 11. When pollutant levels exceed safe thresholds, immediate alerts are sent to users' devices, enabling prompt action to mitigate potential hazards. This combination of real-time data visualization and instant notifications ensures that users are continually informed about their indoor air quality, promoting a healthier living environment.

Journal of Information Systems Engineering and Management
2025, 10(34s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

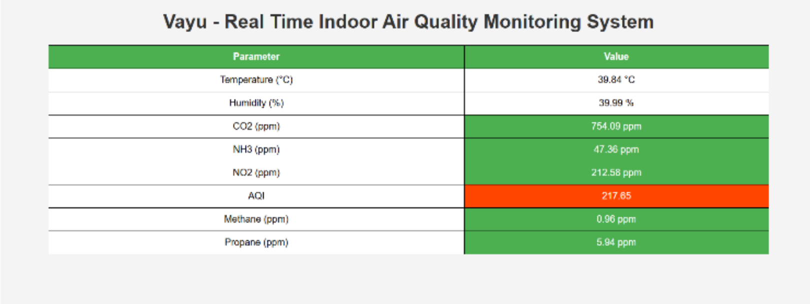

Figure. 10 Display of Real-Time Air Quality Data and Analytics

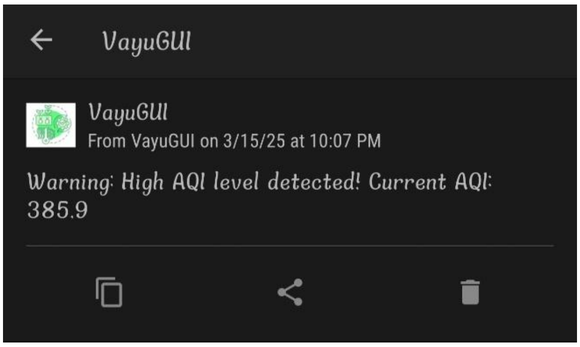

Figure. 11 Notification Sent for Critical Air Quality Conditions

COMPARATIVE ANALYSIS

Accuracy Metrics

In machine learning and statistical modeling, multiple benchmarks are employed to assess the precision of predictive models. The most common measures include Mean Absolute Error (MAE), Mean Squared Error (MSE), Root Mean Squared Error (RMSE) and Mean Absolute Percentage Error (MAPE). Every criterion imparts different insights into model performance, making them useful for different scenarios.

MAE : It assesses the typical absolute discrepancies in a dataset, without considering their directional bias.

$$MAE = \frac{1}{n}\sum_{i=1}^{n} |x_i - y_i| \qquad (14)$$

MSE: It derives the arithmetic mean of the quadratic variances between empirical and prognosticated magnitudes, amplifying the impact of substantial inaccuracies through squaring.

$$MSE = \frac{1}{n}\sum_{i=1}^{n} |x_i - y_i|^2 \qquad (15)$$

RMSE: It is derived by taking the square root of the MSE, restoring the error measurement to the same scale as the original data.

$$RMSE = \sqrt{\sum_{i=1}^{n} \frac{1}{n}\sum_{i=1}^{n} |x_i - y_i|^2} \qquad (16)$$

MAPE: It expresses deviations as a proportion of true values, facilitating seamless comparative analysis across various models and datasets.

$$MAPE = \frac{100}{n}\sum_{i=1}^{n} \frac{|x_i - y_i|}{x_i} \qquad (17)$$

Journal of Information Systems Engineering and Management
2025, 10(34s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

Whereas:

n denotes the aggregate tally of observational instances within the dataset.

x_i is the real observed value from the dataset.

$y_i$signifies the predicted magnitude generated by the forecasting framework corresponding to the actual observed value.

Performance Comparison

The hybrid model (XGBoost + LSTM + MultiOutputRegressor) outperforms all individual models in air quality forecasting due to its ability to leverage the strengths of each approach. XGBoost effectively captures non-linear relationships and feature importance, LSTM excels at recognizing temporal dependencies in time-series data, and MultiOutputRegressor ensures robust multi-variable predictions. This combination results in the lowest MAE (4.29), MSE (34.96), RMSE (5.91), and MAPE (10.76%), making it the most accurate and reliable model.

In contrast as in Table 1, ARIMA, while decent, struggles with complex patterns and has higher errors (MAE: 11.22, RMSE: 12.34). LSTM alone, though suited for sequential data, shows significant error (MAE: 21.43, RMSE: 27.72), indicating overfitting or inefficiency in handling noise. The VAR model performs the worst (MAE: 27.30, RMSE: 50.96, MAPE: 424.83%), proving ineffective for multi-variable air quality forecasting. While XGBoost alone delivers a good performance (MAE: 5.74, RMSE: 6.11, MAPE: 12.12%), integrating it with LSTM and MultiOutputRegressor in the hybrid model further refines predictions, reducing overall errors. Thus, the hybrid model emerges as the best choice for accurate air quality forecasting, effectively balancing precision and efficiency.

Table 1. Displaying the Accuracy Metrics of Various ML Models

MODEL	MAE	MSE	RMSE	MAPE
ARIMA	11.22	152.28	12.34	23.65%
LSTM	21.43	768.40	27.72	24.03%
VAR	27.30	2596.63	50.96	424.83 %
XGBoost	5.73	37.29	6.11	12.12%
Hybrid	4.29	34.96	5.91	10.76%

OBSERVATIONS AND RESULTS

Key Observations

The IoT-based system captures live air quality parameters (CO2, NH3, NO2, AQI, Methane, and Propane) using Arduino UNO, ESP8266, and multiple sensors. Data is transmitted to ThingSpeak, retrieved via Flask, and visualized on a dashboard and mobile app for real-time monitoring. Timely alerts notify users when pollutant levels exceed safe thresholds, enabling proactive measures against poor air quality. AQI (ppm) fluctuates between normal and moderate levels, indicating periodic variations in air quality. The mobile app facilitates remote tracking of air quality trends for better decision-making.

Coming to ML models, XGBoost performed well, achieving MAE: 5.74, MSE: 37.29, RMSE: 6.11, MAPE: 12.12%, making it a strong choice for structured predictions. LSTM struggled with higher errors (MAE: 21.43, RMSE: 27.72) due to its reliance on long-term sequential dependencies. Hybrid Model (XGBoost + LSTM + MultiOutputRegressor) outperformed all individual models with MAE: 4.0074, RMSE: 5.7031, and MAPE: 10.10%, proving the effectiveness of combining XGBoost's feature learning with LSTM's time-series capabilities. VAR and ARIMA models showed poor performance, especially VAR, which had extremely high MAPE (424.83%), indicating its limitations in non-stationary air quality data.

Journal of Information Systems Engineering and Management
2025, 10(34s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/
Research Article

The hybrid model provided the most precise ephemeral (30-day) and protracted (12-month) projections., closely aligning with historical pollution patterns. Separate graphs for both short-term and long-term predictions provide clearer insights into seasonal trends and pollution variations.

The system is scalable, allowing for integration with additional sensors or external meteorological data. Future enhancements include deep learning improvements, anomaly detection, and AI-driven air quality forecasting.

Experimental Results

The air quality monitoring system successfully integrates IoT-based real-time data collection with advanced machine learning models for future predictions. The sensor data, gathered using Arduino UNO, ESP8266 Wi-Fi, MQ135, MQ2, and DHT11 sensors, is transmitted to ThingSpeak and retrieved via Flask to power a dashboard for live monitoring and timely alerts. The setup is shown in Figure. 12.

For future air quality forecasting, five models were developed: LSTM, ARIMA, VAR, XGBoost, and a Hybrid XGBoost+LSTM+MultiOutputRegressor model. Among these, the hybrid model provided the most accurate predictions by leveraging the strengths of XGBoost (capturing short-term variations) and LSTM (handling long-term dependencies). The model forecasts AQI values for the next 30 days (short-term) and 12 months (long-term) with reasonable accuracy.

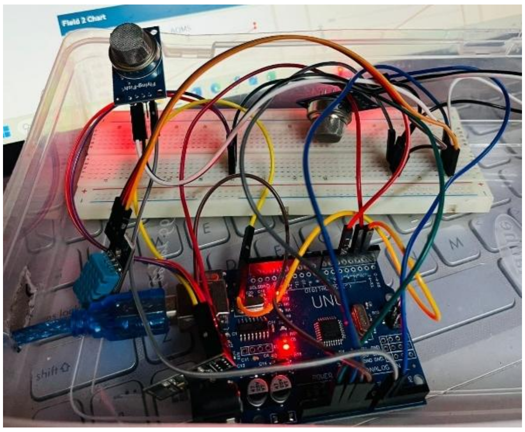

Figure. 12 Arrangement and Connection of System Components

The system's predictive performance was assessed using evaluation criteria such as MAE, MSE, RMSE, MAPE, with the hybrid model achieving the lowest error rates. The results are visualized through separate 30-days (transient) and 12 months (prolonged) forecast graphs, providing an intuitive representation of future air quality trends.

Future Prospects

Vayu has vast future potential, with significant advancements in accuracy, scalability, and impact. Expanding its sensor network to detect additional pollutants such as SO_2, O_3, and VOCs will enhance monitoring capabilities, providing a more comprehensive understanding of air quality. Deploying multiple devices across cities and regions can create a real-time air quality map, allowing individuals to make informed decisions while enabling policymakers to identify pollution hotspots and implement effective control measures. By refining forecasting models with hybrid AI techniques—integrating ARIMA, LSTM, and external factors like weather and traffic—Vayu can improve predictive accuracy, offering timely alerts and empowering users to take preventive action. Integrating smart alerts through voice assistants and automated HVAC adjustments can enhance responsiveness, ensuring indoor environments remain safe. Additionally, issuing community-wide alerts during extreme pollution episodes can help safeguard at-risk groups, including young children, senior citizens, and those with respiratory ailments.

To maximize accessibility and impact, Vayu can be integrated into a dedicated mobile and web application that provides real-time AQI monitoring, interactive pollution heatmaps, and AI-driven insights. Open data access through collaboration with municipalities, research institutions, and environmental organizations can support scientific research and influence policy-making, driving better pollution management strategies. Ensuring energy efficiency

Journal of Information Systems Engineering and Management
2025, 10(34s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

with solar-powered sensors, low-power communication protocols like LoRa and NB-IoT, and edge computing for on-device data processing can make Vayu a sustainable and scalable solution, extending its reach to remote areas. By fostering community-driven pollution control initiatives and providing accurate, real-time air quality insights, Vayu has the potential to reduce pollution exposure, prevent respiratory diseases, and contribute to a healthier environment. As climate change and urbanization continue to challenge air quality, Vayu can serve as a powerful tool for safeguarding public health, advancing research, and shaping policies for a cleaner future.

REFRENCES

[1] Dr. N. Anitha, S. Anne Caroline, P. Shalini and V.Trisha,"Air Quality Monitoring System," Proceeding of the IEEE Conference on Intelligent Computing and Control for Engineering and Business Systems (ICCEBS), 2023, doi: 10.1109/ICCEBS58601.2023.10448553.

[2] Vilas Kisanrao Tembhurne, Vinayak Dagadu Shinde, Santosh Kumar, Manish Shrimali, Gunjan Chhabra, M. N. Quadri and Vikram Rajpoot, "IoT-based Air Pollution Monitoring System to Measure Air Quality on Cloud Storage," Proceeding of the IEEE Conference on 2nd International Conference on Paradigm Shifts in Communcations Embedded Systems, Machine Leraning Signal Processing(PCEMS),"2023,doi:10.1109/PCEMS58491.2023.10136085.

[3] Kamepalli Alekhya, P. Dharithri Sravya, N. Chaitanya Naik and B. Jaya LakshmiNarayana, "Ambient Air Quality Monitoring System," Proceeding of the IEEE Conference on International Conference for Advancement in Technology (ICONAT), 2023, doi: 10.1109/ICONAT57137.2023.10080220.

[4] Farida A. Ali, Somya Prakash and Sabita Mali, "IoT-based Real-time Monitoring System for Indoor Air Quality for Human Health," Proceeding of the IEEE Conference on 2nd International Conference on Ambient Intelligence in Health Care (ICAIHC), 2023, doi: 10.1109/ICAIHC59020.2023.10431482.

[5] Mrs. A.Vani Lavanya, S. Roveno Selva and S. R. Vijay, "Envirosense Iot: Smart Air Pollution Monitoring System," Proceeding of the IEEE Conference on International Conference on Innovations and Challenges in Emerging Technologies (ICICET), 2024.

[6] Silvian-Marian Petrică, Ioana Făgărăşan, Iulia Stamatescu, Nicoleta Arghira and Radu-Nicolae Pietraru, "Low-cost air pollution monitoring systems," Proceeding of the IEEE Conference on 28th International Conference on System Theory, Control and Computing, 2024, doi: 10.1109/ICSTCC62912.2024.10744643.

[7] S. Priyadharshini, A. Vanathi, J. Cypto, Geetanjali Ganguly and M. Srinivasan, "A Logical Air Quality Monitoring System using Arduino and Smart Sensors Association," Proceeding of the IEEE Conference on International Conference on Innovative Computing, Intelligent Communication and Smart Electrical Systems (ICSES), 2023, doi: 10.1109/ICSES60034.2023.10465463.

[8] Silvian-Marian Petrica, Ioana Fagarasan, Iulia Stamatescu and Giorgian Neculoiu, "Energy Efficient IoT Air Quality Monitoring System ," Proceeding of the IEEE Conference on 24th International Conference on Control Systems and Computer Science (CSCS), 2023, doi: 10.1109/CSCS59211.2023.00086

[9] Souptik Das, Sourish Chakraborty, Dipanjan Jana, Rahul Nandy and S rijan Bhattacharya, "IoT Based Industrial Air Quality Monitoring System," Proceeding of the IEEE Conference on Second International Conference on Computer Science, Engineering and Applications (ICCSEA), 2022, doi: 10.1109/ICCSEA54677.2022.9936294.

[10] Mohammed Rakib, Sanaulla Haq, Md. Ismail Hossain and Tanzilur Rahman, "IoT Based Air Pollution Monitoring & Prediction System", Proceeding of the IEEE Conference on ICISET, 2022.

[11] Piyush Devidas Potbhare, Ketan Bhange; Gaurav Tembhare, Rahul Agrawal, Swati Sorte and Prasad Lokulwar, "IoT based Smart Air Pollution Monitoring System," Proceeding of the IEEE Conference on International Conference on Applied Artificial Intelligence and Computing (ICAAIC), 2022, doi: 10.1109/ICAAIC53929.2022.9792743.

[12] Quynh Anh Tran, Quang Hung Dang, Tung Le, Huy Tien Nguyen and Tan Duy Le, "Air Quality Monitoring and Forecasting System using IoT and Machine Learning Techniques," Proceeding of the IEEE Conference on 6th International Conference on Green Technology and Sustainable Development (GTSD), 2022, doi: 10.1109/GTSD54989.2022.9988756.

[13] Sorin lonuţ Conea, Gloria Cerasela Crişan "Green air quality monitoring system based on Arduino," Proceeding of the IEEE Conference on 14th International Conference on Electronics, Computers and Artificial Intelligence (ECAI), 2022, doi: 10.1109/ECAI54874.2022.9847421.

[14] Baboo Sivrajsingh Loderchand, Robert T. F. Ah King and Bhimsen Rajkumarsingh, "Smart Real Time System For Air Pollution Monitoring," Proceeding of the IEEE Conference on 4th International Conference on Emerging Trends in Electrical, Electronic and Communications Engineering (ELECOM), 2022, doi: 10.1109/ELECOM54934.2022.9965229.
[15] Virendra Barot and Viral Kapadia, "Air Quality Monitoring System using IoT: A Review," Proceeding of the IEEE Conference on International Conference on Computational Performance Evaluation (ComPE), 2020, doi: 10.1109/ComPE49325.2020.9200053.