SEMARAK ILMU PUBLISHING

Journal of Advanced Research in Applied Sciences and Engineering Technology

Journal homepage:
https://semarakilmu.com.my/journals/index.php/applied_sciences_eng_tech/index
ISSN: 2462-1943

Monitoring and Prediction of Air Quality System using Internet of Things (IoT)

Mohammed Saad Ashraf Alrubaye[1], Azizan As'arry[1], Muhammed Amin Azman[1,*], Mohd Zuhri Mohamed Yusoff[1], Khairil Anas Md Rezali[1], Ali Zolfagharian[2]

[1] Department of Mechanical & Manufacturing, Faculty of Engineering, Universiti Putra Malaysia, 43400 Serdang, Selangor, Malaysia
[2] School of Engineering, Deakin University, Geelong 3216, Australia

ARTICLE INFO

Article history:
Received 21 December 2023
Received in revised form 7 May 2024
Accepted 9 June 2024
Available online 5 July 2024

Keywords:

Monitoring system; air quality; prediction and regression models; Internet of Things (IoT); ML.NET algorithm

ABSTRACT

Air pollution is a global issue causing 7 million deaths annually, primarily due to dangerous compounds and inhalable particles. The World Health Organization estimates that air pollution is a significant threat to human health and safety. The Internet of Things (IoT) is utilized in this research to analyse sensor data from various environmental monitoring devices. The objective is to build an IoT-based air quality system that assesses air quality conditions in specific areas and analyses air levels of various compounds. The study employs IoT technology to analyse data from various environmental sensors, aiming to create an IoT-based air quality system. The sensors include measuring NH, C6H6, VOCs (MQ135), CH4 (MQ5), and particle matter (PM2.5). The ESP32 microcontroller and thinger.io platform is used to develop the system. FastTree and Generalized Additive Model (GAM) are machine learning methods applied to predict and analyse air quality data. FastTree utilizes gradient-boosting to enhance accuracy, while GAM employs smooth parts, represented by splines, for relationship modelling. Evaluation metrics R2, RMSE, MSE, and MAE assess model performance, for MQ135 with FastTree outperforming GAM with an R2 of 91.91% compared to 78.05%. The IoT-based air quality system is user-friendly and effective, tracking harmful components against thresholds. Future work includes adding cameras and printed circuit boards for expanded analysis.

1. Introduction

Companies have grown dramatically during the last quarter-century. Such actions have resulted in a severe and difficult environmental situation [1]. The World Health Organization (WHO) has advocated imposing limits on a range of air pollutants, including ground-level ozone, nitrous oxides, and sulphur oxide, to mitigate the negative health impacts of air pollution on the general public [2]. Pollution is mostly to blame for the severe climate, which has led to changes in biological and hydrological systems, habitat loss, stratospheric ozone depletion, climate erosion, and climatic

* *Corresponding author.*
E-mail address: amin.azman@upm.edu.my

https://doi.org/10.37934/araset.48.1.6176

transition [3]. These pollutants have been related to several, including cancer, measles, asthma, respiratory issues, circulatory heart issues, and chronic cardiovascular disorders.

Since these tools will be able to quickly detect and quantify the sources of emissions that comprise dangerous compounds, the sector for environment emissions monitoring systems is growing. The contemporary automated air surveillance program makes use of laboratory analyses that need rather sophisticated infrastructure, large amounts of data, inconsistent behaviour, and expensive methods [4]. This makes the high cost and massive bulk of large structures problematic. Device data cannot predict the end emissions situation because this device will only be installed at key control sites by a chosen group of significant companies. One such thesis recommends integrating IoT technology with environmental protection to address deficiencies in conventional control and detection systems and to cut research costs [5]. Based on several previous research, this inquiry was conducted, and an outside quality control process was used [6]. Numerous substances found in the air, including O3, SO2, CO, and particulate matter, may be determined. Urban growth offers a high level of life at the price of deteriorating the environment and air quality. Carbon oxides (COx), nitrogen oxides (NOx), sulphur oxides (SOx), atmospheric particulate matter (PM) with diameters less than or equal to 10 m (PM10), and PM2.5 with diameters less than or equal to 2.5 m are examples of air pollutants (PM2.5) [7]. A number of variables, including population size, wind direction and speed, pollution dispersion, location (indoors or outdoors), and various weather conditions, have an immediate impact on the quality of the air. Moreover, countries have enacted standards and regulations to monitor air pollution and alert their citizens [8]. Indoor air pollution has gotten less attention than outdoor air pollution, even though adults spend the bulk of their time indoors. This is a result of the global shift from the manufacturing sector to the service and knowledge-based sectors, which rely on indoor office environments [9]. Pollution sources are divided into two categories: natural and man-made (human-made). Furthermore, natural pollution refers to natural incidents that have a negative influence on the environment or emit hazardous substances [10].

Natural disasters that produce enormous volumes of air pollutants such as SOx, NOx, and COx include forest conflagrations and volcanic outbursts. On the other hand, some man-made sources, such as car emissions and fuel combustion, are regarded as major drivers of air pollution [11]. Pollutants such as ozone, hydrogen, metal compounds, nitrogen, sulphur, and particulate matter may be produced. The term "atmospheric particulate matter" refers to material suspended in the atmosphere that is either liquid or solid [12].

Air pollution is mostly caused by man-made sources including fuel burning and automobile exhaust. In other hand, traditional air automated monitoring systems rely on laboratory analysis and are expensive, bulky, and equipped with complicated technology [13]. Meanwhile, data from this system cannot be utilized to anticipate overall pollution levels since it can only be installed in essential monitoring zones of a small number of significant companies [14]. In order to address the shortcomings of current monitoring systems and detection techniques while also reducing test costs, this research offers a strategy for fusing IoT technology with environmental monitoring. Moreover, systems for measuring air pollution are now being employed in cities as a result of investments made in recent decades to meet emerging knowledge about the health hazards of air pollution [15]. Furthermore, these conventional systems are container-style monitoring stations that take up a lot of area and consume a lot of power, necessitating a lot of maintenance and manufacturing expenditures. To address the drawbacks of conventional monitoring systems and detection techniques while simultaneously reducing test costs, this study offers a strategy for fusing IoT technology with environmental monitoring [16].

The Quality of the Air A quantitative indicator that describes the condition of the air quality in a particular area is post-segmentation (AQI) [17]. A score above zero indicates greater concern about the degree of air pollution and a greater threat to one's health [18]. The scale ranges from 0 to 500 [19]. As a result, as shown in Table 1, created a pollution traffic light (red, yellow, and green) that integrates the qualities listed above and accurately depicts the environmental conditions in Ibarra, Ecuador.

This study highlights the limitations of current emission parameter measurements and their implications for human health hazards [20]. It discusses the inadequacy of existing sensors in accurately measuring emissions, necessitating improvements in data collection and analysis. Current sensors are used to collect data on various emissions, but they fail to accurately predict and analyse the collected data [21].

Table 1
The Understanding of the Air Quality Index

Air Quality Index levels of health concern	Numeric Value (ppm-parts per million)	Explanations
Good	0 to 50	The quality of the air is regarded as excellent, and there is little to no risk from air pollution.
Moderate	51 to 100	The air quality is usually adequate, however certain pollutants can be harmful to vulnerable populations such as children, the elderly, and those with chronic diseases.
Unhealthy for sensitive group	101 to 150	General people may not experience any health risks while sensitive groups can be affected.
Unhealthy	151 to 200	Serious health repercussions will be felt by the sensitive group, but regular people may also encounter some health difficulties.
Very unhealthy	201 to 300	Everyone may have major health concerns at this level; thus, health awareness is necessary.
Hazardous	301 to 500	This is an emergency condition for everyone, the entire population is in serious health risk.

This research paper proposes a novel predictive modelling approach to analyse sensor data from various environmental monitoring sensors, including (MQ135, MQ5, and PM 2.5) sensors [22]. It demonstrates its effectiveness in accurately predicting emission parameters and provides insights into hazard assessment. The problem statement highlights the insufficiency of current emission parameter measurements and emphasizes the need for improved predictive modelling techniques [23]. To fulfil these aims, the research effort is segmented into the following objectives to create an Internet of Things-based air quality system to assess the region's air quality. In addition, sensors are used to monitor the air rates of various compounds, such as, CO_2, CH_4, NH_3, S, C_6H_6, and particulate matter (PM). In order, to make a forecast by analysing the air rates using two regression models using ML.NET (FastTree and GAM).

To determine the concentration of carbon dioxide in a remote place, a monitoring system is being developed. The gadget also reports the temperature, humidity, and light output of the outdoor monitoring area. In a similar vein, the author has suggested a method for monitoring urban CO_2 [24].

In other hand, the system operates outside a 100-square-kilometer metropolitan area. In various investigations, the quality of the air, water, and soil has been assessed using a monitoring system based on the Internet of Things. To monitor the quality of the air, water, and soil, several sensors are employed, such as the DHT11, MQ135, and MiCS-2610 [25]. A CO, CO_2, alcohol, temperature, and air humidity monitoring system that uses an IoT-based website as an interface to show the findings of the air quality measuring system was not examined in previous research. Rather than relying exclusively on the amount of gas in the air, believe that temperature and humidity monitoring are also necessary to monitor air quality. Consequently, this work proposes an IoT-based air quality monitoring system that can identify NH, S, CO, CO2, alcohol, particle matter, and flammable gases at certain sites and provide the results of the measurement on a website. In this study, an Arduino and a Raspberry Pi are used along with a DHT11 module that measures temperature and an MQ135 module that measures NH, S, CO, CO2, and alcohol and flammable gases [26].

2. Methodology

The proposed air quality monitoring system device includes three sensors as shown in Figure 1, The local alarm, which consists of a sound alarm via sirens and lights to alert people of the location of changing air quality, as well as a remote alarm sending text messages via email to the authorized person's device and displaying the variables that occur in the air due to gas emission or a change practical matter to provide a high response to maintain air quality. The proposed air quality monitoring system employs more current and specialized techniques to detect the presence of gases and pollution in the air and to generate local and remote alarms. To make it easier to monitor the output of sensors, the Internet of Things will be made available through the thinger.io platform on a computer webpage or a mobile app.

Through sensors, the gadget may monitor the air rates of various substances such as CO2, SO2, S, C6H6, CH4, and particle matter. The data is delivered to the cloud system, which is subsequently accessed via an Arduino WIFI module. Through a cloud Site page (Thinger.io), it is possible to access the tracking effects and monitor variables that occur in the atmosphere for the sensors used. The current model has been effectively applied, and it may be utilized to create systems in the actual world. The system's design is shown in Figure 1. The output voltage of a sensor would rise as a gas was detected in the environment, reducing the gas concentration and deoxidizing. By utilizing the ML.NET algorithm for (FastTree and GAM) regression models for (PM2.5, MQ135, and MQ5) sensors, one may determine the proportion of error during (R2, RMSE, MSE, and MAE) and obtain prediction values for them.

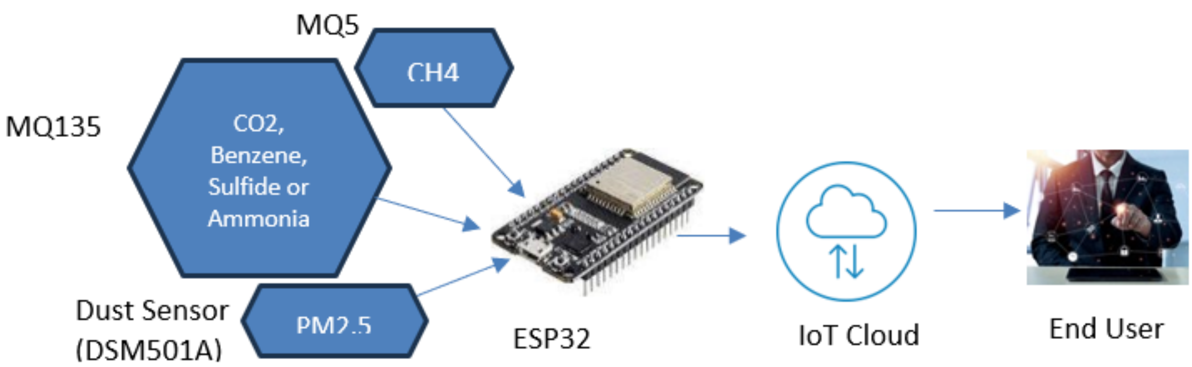


Fig. 1. Proposed IoT sensing system

2.1 Machine Learning Models Explanation

By using the ML.NET platform for regression and prediction models. Here's a step-by-step approach:

i. Data Preprocessing: The gathered data should be cleaned and prepared. In addition to handling missing values and outliers, data transformations like normalization and scaling may also be required.
ii. Feature Engineering: Extract relevant features from the sensor data that can help in predicting gas concentrations. This may involve calculating statistical features, time-based features, or creating composite features based on domain knowledge.
iii. Splitting the Data: Distinguish the pre-processed data from the training and testing sets. Regression models are usually tested on the testing set after being learned on the training set.
iv. Model Training: Train regression models using the FastTree, and GAM models. These models provide implementations of decision tree-based algorithms that can handle regression tasks efficiently.
v. Model Evaluation: Use appropriate regression assessment measures, such as mean squared error (MSE), mean absolute error (MAE), root mean square error (RMSE), and R-squared, to evaluate the trained models (R2). Analyse the models' efficiency on the testing set to find out how effectively they can estimate gas concentrations.
vi. Feature Importance: Use the feature importance capabilities provided by the FastTree and GAM modules to determine the importance of features in predicting gas concentrations. This analysis helps identify which sensors or features have the most significant impact on the predictions.
vii. Predictions and Analysis: Use the trained regression models to make predictions on new data. Analyse the predicted gas concentrations and compare them with the actual measurements. Visualize the results to gain insights into the accuracy of the predictions and any patterns or trends in the data.
viii. Relationship Analysis: To explore the relationships among the gases, calculate correlation coefficients between the predicted concentrations. Analyse the correlation matrix or create scatter plots to identify any significant relationships or dependencies between different gases.
ix. Interpretation and Insights: Interpret the results obtained from the analysis. Identify the most influential sensors or features in predicting gas concentrations and understand the relationships among the gases. Obtain important knowledge that may be used in decision-making or more research.

2.1.1 FastTree model

For classification and regression problems based on decision trees, the FastTree model is a machine learning technique utilized. It is a part of the Microsoft Machine Learning framework, specifically the ML.NET library, which is designed for developing machine learning models in .NET applications. The FastTree algorithm also builds an ensemble of decision trees to make predictions. However, there are some differences in the way the trees are constructed and the training process.

FastTree is a versatile gradient-boosting ensemble method that excels in both regression and classification tasks. It efficiently handles large-scale datasets through optimizations like approximate sorting and parallelization, making it suitable for scenarios with extensive samples and features. With built-in techniques for controlling decision tree complexity, it helps prevent overfitting and enhance generalization. FastTree also provides valuable insights into feature importance, aiding in feature selection and understanding their impact on predictions. Its tuneable parameters, including the learning rate and ensemble size, offer flexibility for optimizing model performance across diverse problem domains.

2.1.2 GAM model

Generalized Additive Models (GAMs) are statistical models that extend linear models by allowing for non-linear relationships between variables. In the context of air quality systems, GAMs are used for modelling, forecasting, and analysing air pollutant concentrations. They can capture complex and non-linear relationships between environmental variables (e.g., weather conditions, emissions, geographic factors) and pollutant levels. GAMs are valuable for trend analysis, short-term forecasting, health impact assessment, spatial analysis, data quality control, interactive visualization, and policy evaluation in air quality management. Their flexibility makes them suitable for addressing the intricate nature of air quality data and its impact on public health and environmental policies.

2.2 IoT Sensing System

The quantity of industrial emissions, automobile traffic, building activity, vegetation, and population density all significantly affect the city's air pollution levels depending on its location. For this project, sensors and data collection systems are being developed to offer real-time, precise data on the level of pollution at various locations throughout the city. A number of sensors are coupled to an ESP32 CPU to form it. The ESP32 wirelessly transfers the data to a smartphone application through a Wi-Fi module. The data is delivered from the smartphone application to the IoT cloud for monitoring and analysis. All sensor details are provided in Table 2 Parts per million (ppm) are used to measure the quantities of CO2, CO, NH3, C6H6, VOCs, and CH4, whereas grams per cubic metric (g/m3) are used to assess the amounts of particulate matter.

Table 2
Hardware Sensing Setup

S. no.	Environmental parameter	Unit	Sensor
1	NH3, C6H6, VOCs, CO2	ppm	MQ135
2	CH4	ppm	MQ5
3	Particulate Matter	μg/m3	DSM501A PM2.5

2.3 System Components

A description of the needed equipment to construct the proposed air quality monitoring system is illustrated here which includes breadboard, buzzer (beeper), jumper wires, and emitting diode (LED) of three colours (green, yellow, and red) to know the device status, device power like a lithium battery displaying a Charge Shield LED indicator (green indicates full, red indicates charging), ESP32, MQ5, MQ135, and PM2.5. Figure 2 depicts the image with all of the component connections.

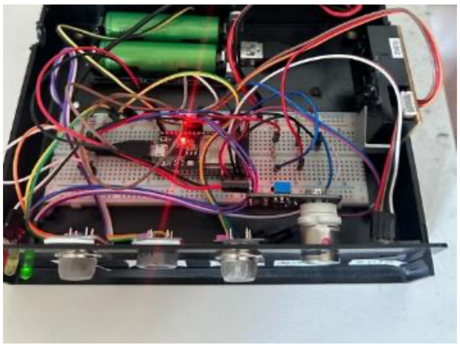


Fig. 2. The Component Connections

2.3.1 ESP32

The ESP32's temperature range of -40°C to +125°C enables it to function reliably in industrial environments. Powerful calibration circuitry allows for the dynamic removal of external circuit defects and adaptation to changing environmental circumstances.

To achieve extremely low power consumption, ESP32, which was created for portable devices, wearable technology, and Internet of Things technologies, uses a variety of various forms of proprietary software. In addition, Modern features include dynamic power scaling, different power modes, and fine-grained clock gating. The ESP32 comes with filters, analogue, low-noise receive amplifiers, an RF balun, antenna switches, and power management modules. The ESP32 increases the applications with unparalleled capacity and diversity while requiring only a small quantity of printed circuit board (PCB).

The ability of the ESP32 to function as a autonomous device or as a slave device towards the hosting MCU reduces the connection stack overhead on the main application CPU. The ESP32 can connect with a limited number of other devices through its SPI/SDIO or I2C/UART interfaces, and it also has Wi-Fi and Bluetooth capabilities. As shown in Figure 3, ESP32.

Fig. 3. ESP32

2.3.2 MQ5

As seen in Figure 4, the MQ-5 gas sensor has excellent sensitivity for detecting LPG, natural gas, and coal gas. The output voltage rises as the concentration of the measured gases rises. It offers a quick reaction and recovery time, a sensitivity adjustment, and a signal output indicator. The MQ-5 can detect a gas leak or monitor the output of your wood gas generator.

Technical Specifications: 5 V supply voltage, 160 mA supply current, Analog output type Propane/LPG (C3H8), Methane/Natural Gas (CH4), Butane (C4H10), and Coal/Town Gas were detected. Concentration: 200 - 10000 ppm200 - 10000 ppm.

Fig. 4. MQ5

2.3.3 MQ135

The MQ-135 Air Quality Sensor is a low-cost that can be utilized in several applications, as shown in Figure 5 To identify the presence of dangerous gases, tin oxide is utilized. Due to its stability, high sensitivity, quick response, and long life, it is ideal for air quality control equipment. It is capable of detecting a wide range of hazardous environmental pollutants, including smoke, CO2, nitrous oxide, ammonia, and sulphide. To test for carbon monoxide gas, this device requires a 5V AC/DC heating power supply. It also supplies current at a 160-mA distance (20–1000 ppm).

Fig. 5. MQ135

2.3.4 Dust sensor DSM501A PM2.5

A pulse signal created by the same counter and particle theory of light scattering will be collected inside the unit volume and will correspond to the exact number of particles. Figure 6 depicts the PM2.5 dust sensor.

i. Can detect smoke, dust, and other allergic housing particles.
ii. Can detect particles larger than one micron in size or 2.5 microns in size.
iii. Lightweight, compact, and simple to install.
iv. Signal processing is facilitated by a 5V input circuit.
v. Internal air generator that is unrestricted in drawing in outside air.
vi. Simple maintenance is required to maintain the sensor's long-term properties.

Fig. 6. Dust Sensor
DSM501A PM2.5

3. Results

Machine learning is being used more and more frequently in a variety of industries for data analysis and prediction. This involves connecting sensors to an IoT cloud platform. In this project, we will look into how to connect sensors to the thinger.io IoT cloud platform, as seen in Figure 7. Additionally, we will configure the platform to send notifications through Gmail and utilize machine-learning techniques to monitor and analyse the collected sensor data. In this project, we explored the process of connecting sensors (MQ5, MQ135, and PM2.5) to the thinger.io IoT cloud platform, configuring Gmail notifications, and utilizing machine learning for monitoring and analysing the collected data. By connecting the physical sensors to the cloud platform and employing machine learning techniques, we can gain valuable insights, make predictions, and take appropriate actions based on the sensor readings. This integration of IoT and machine learning opens up numerous possibilities for applications in environmental monitoring, smart buildings, and more.

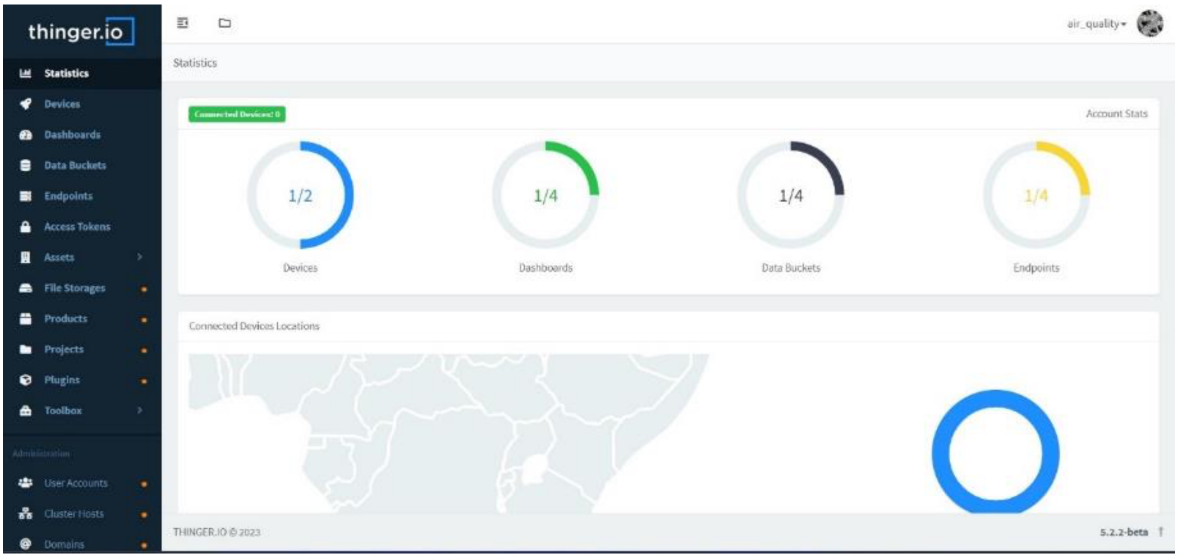


Fig. 7. Statistics in Thinger.io Platform

3.1 Analysis of Measurement Data

Once we have successfully interfaced the sensors for measuring air quality and environmental parameters with your ESP32 microcontroller, we can proceed to measure the desired variables.

3.1.1 Monitoring and analysis with machine learning

Once notifications have been set up and sensor data is streaming to thinger.io, we can move on to monitoring and ML analysis of the data that has been gathered. An open-source machine learning (ML) framework for.NET called NET is cross-platform and will load from the thinger.io platform. One way to derive meaningful insights from the data is by utilizing machine learning techniques for prediction and anomaly detection.

To create and train prediction models, will make use of machine learning libraries. The gathered sensor data is used to train the machine learning models, which may then anticipate outcomes based on fresh data and write CSharp code to execute in VisualStudio.

By continuously monitoring the sensor data, we can feed it into the trained models and obtain predictions for future air quality, temperature, humidity, and gas levels. These predictions can be used to identify patterns, detect anomalies, or trigger further actions, such as adjusting environmental controls or sending additional notifications.

3.1.2 Data collection and send data to the cloud

This section will collect the data from indoors like houses and outdoors such as public places environment by opening the thinger.io platform going to data buckets and clicking on our device name air quality can monitor the data for all sensors (MQ135, MQ5, and PM2.5). And the total data are 9844 from 01/01/2022 to 01/07/2023. As seen in Figure 8, the gadget will provide measurements of value every minute.

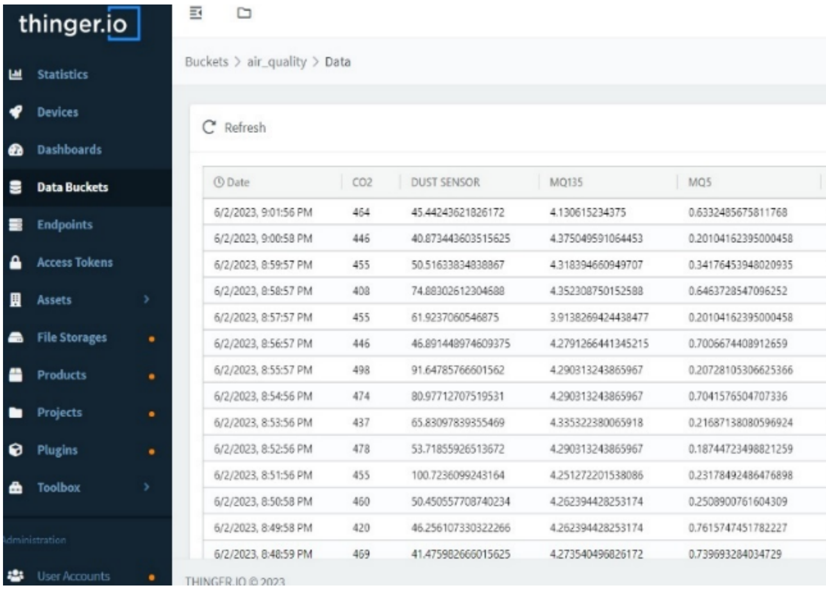


Fig. 8. Data Buckets in Thinger.io Platform

3.2 Comparison with Machine Learning Models

The R-squared (R2) statistic is a statistical measure of how much variance in a dependent variable can be accounted for by the independent predictors in the model. Lower numbers indicate greater performance when measuring a model's accuracy using the Root Mean Square Error (RMSE) measure. The mean squared error (MSE), a statistic used to assess the efficacy of regression models, quantifies the average squared discrepancies between the predicted and observed values. The mean absolute error (MAE), a measurement of the average absolute discrepancies between planned and actual values, is calculated as the average of the absolute differences.

After collecting the data from the sensors (MQ5, MQ135, and PM2.5), proceed with the analysis using the FastTree and GAM models for regression prediction for CO2, NH3, C6H6, VOCs, CH4, and particle matter. The findings from using the data from the polluting gas measurement equipment in conjunction with the chosen machine-learning approaches are shown in this subsection. Table 3 specifies the R2, RMSE, MSE, and MAE results referring to PM2.5 gas for each algorithm. As seen, no outlier values were found, hence the tests were just run once. The FastTree algorithm, followed by the GAM algorithm, produced the best results. Nevertheless, FastTree is somewhat more accurate than GAM for simulated and real alignments (where the variance reduction is used as a selection criterion from the mean squared error metric).

Since the FastTree technique requires more processing and memory resources than the other methods under discussion, it is imperative to resolve the disparities in processing times required by these algorithms. More specifically, compared to GAM, the FastTree required nearly 500 times as much processing time. The results for the FastTree algorithm as the best machine learning technique for air quality data were repeated for all the other sensors in the device. These findings are shown in Table 2 for the PM2.5 data. As can be seen, the FastTree model's R2 value of 97.15% is the best when compared to other models GAM of 80.49%. This also means that the mean error is less than that of the other models. Table 3 for the PM2.5 data, Table 4 for the MQ135 data, Table 5 for the MQ5 data. The finding is the error is less by using FastTree models compared with other models for all sensors.

Table 3
PM2.5 Average for (R2, RMSE, MSE, and MAE Analysis of FastTree, and GAM Models) by ML.NET

ML Technique	R2	RMSE	MSE	MAE
FastTree	0.9715	1.3755	1.8921	0.5232
GAM	0.8049	3.5975	12.9418	2.4443

Table 4
MQ135 Average for (R2, RMSE, MSE, and MAE Analysis of FastTree, and GAM Models) by ML.NET

ML Technique	R2	RMSE	MSE	MAE
FastTree	0.9191	0.7691	0.5915	0.3555
GAM	0.7805	1.2667	1.6045	0.8253

Table 5
MQ5 Average for (R2, RMSE, MSE, and MAE Analysis of FastTree, and GAM Models) by ML.NET

ML Technique	R2	RMSE	MSE	MAE
FastTree	0.9248	0.5535	0.3063	0.3046
GAM	0.8592	0.7577	0.5741	0.5029

Figures 9, 10, and 11 demonstrate the bar charts of the two regression models algorithm (FastTree and GAM) with the error values (R2, RMSE, MSE, MAE) for the data taken by the sensors in this device.

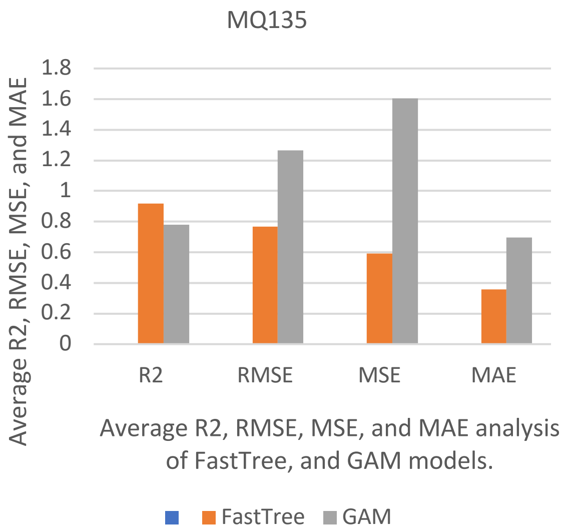


Fig. 9. PM2.5 average between R2, RMSE, MSE, MAE, and the two models

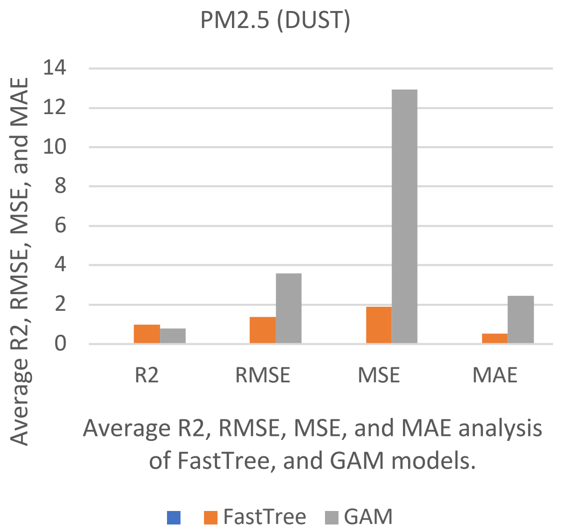


Fig. 10. MQ135 average between R2, RMSE, MSE, MAE, and the two models

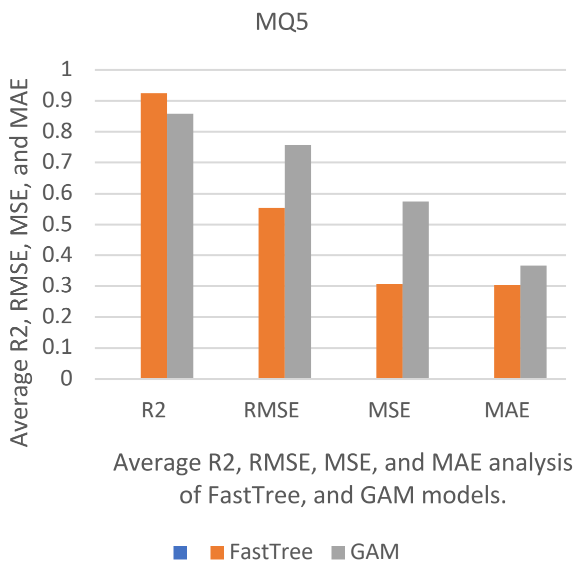


Fig. 11. MQ5 average between R2, RMSE, MSE, MAE, and the two models

These results of prediction are presented in Tables 6 and 7 for the all-sensors data by using different algorithms like FastTree and GAM models. Table 5 depicts the prediction values for all sensors using the FastTree model. Table 6 illustrates the prediction values for all sensors using the GAM model.

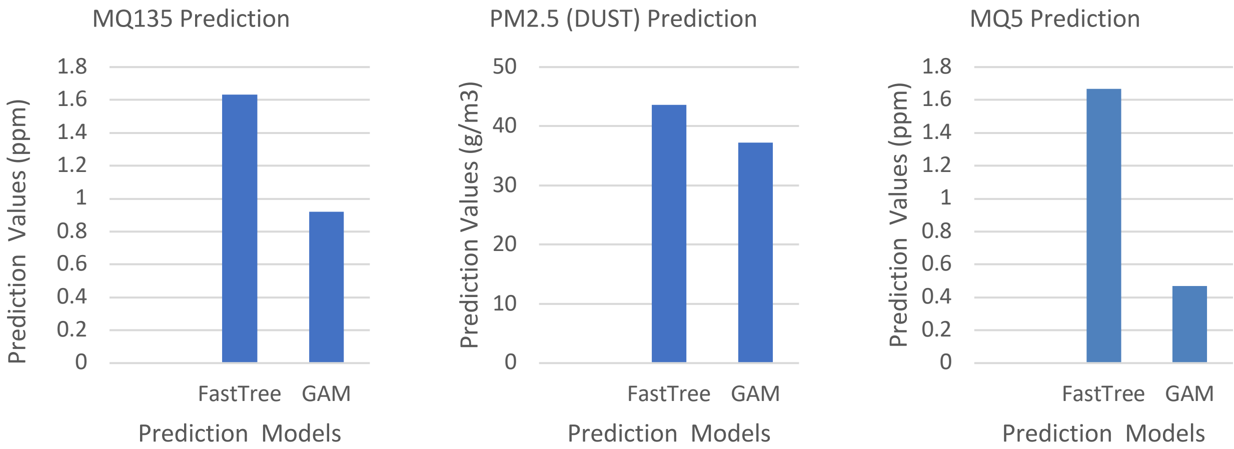
Table 6
The prediction of all gas sensors by FastTree models

Sensors	Prediction values	Units
PM2.5 (DUST)	43.5820	g/m^3
MQ135	1.6328	ppm
MQ5	1.6664	ppm

Table 7
The prediction of all gas sensors by GAM models

Sensors	Prediction values	Units
PM2.5 (DUST)	37.2089	g/m^3
MQ135	0.9219	ppm
MQ5	1.6664	ppm

Figures 12, (a) MQ135 prediction by two models (FastTree and GAM), (b) PM2.5 (Dust) Prediction by two models (FastTree, and GAM), (c) MQ5 prediction by two models (FastTree, and GAM) show the prediction value as a chart bar by using two regression models and for each sensor (PM2.5, MQ135, and MQ5), the units for MQ135 and MQ5 will be (ppm) part per million and for PM2.5 sensor will be g\m3.



(a) (b) (c)

Fig. 12. (a) The Prediction for MQ135 Using Two Regression Models, (b) The Prediction for PM2.5 Using Two Regression Models, and (c) The Prediction for MQ5 Using Two Regression Models

Figure 13, depicts the data of all sensors with the total quantity of values (PM2.5, MQ135, and MQ5) depending on the unit scale of each sensor. The proportions vary depending on the readings taken from the sensors and on the number of particles of gases found in the atmosphere. As we note, the lines appear in a straight picture, and high and low depend on the changing accrue in the air.

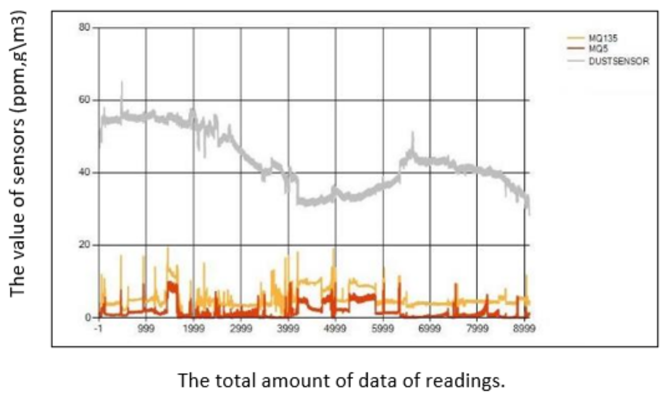


Fig. 13. The values of sensors with the total number of readings

4. Conclusions

Air pollution in metropolitan areas is a significant issue affecting the population, environment, and global economy. An IoT-based air quality monitoring system is a promising solution to address this issue. This simple, affordable, and effective system monitors dangerous components of the air, offering more benefits than previous systems. It is affordability and accurate sensor values make it a cost-effective and accessible solution for individuals and organizations. The IoT-based air quality system is a portable system that integrates Thinger.io with ESP32, allowing users to easily access and deploy it in various environments. The system combines simplicity, portability, cost, and connection via the ESP32 to be simple, inexpensive, and portable. It measures air rates of various substances, such as CH4, NH3, S, C6H6, flammable gases, and PM2.5, to detect and monitor hazardous components. The system analyses air rates using ML.NET for two regression models (FastTree and MAG), ensuring fast convergence and accurate predictions. FastTree is chosen as the primary model, ensuring precise and efficient air quality predictions. The system provides a practical and accurate approach to monitoring and predicting air quality in metropolitan areas, leveraging IoT technology to address air pollution challenges. In future works, we look forward to implementing, cameras and a printed circuit board (PCB) to improve the obtained results and test the maximum number of allowed sensors. The information acquired can then be used to develop models for predicting pollution levels and calibrating the sensors for long-term sensor drift. According to a preliminary analysis, future research will concentrate on the solar panel, power consumption, and charging, which have not undergone thorough testing.

Acknowledgment
The authors would like to thank Universiti Putra Malaysia (UPM) and the Malaysian Ministry of Higher Education (MOHE) for their continuous support in this research work. This work was supported in part by the Geran Insentif Putra Siswazah (GP-IPS) fund (grant vot. number 9697000) and the Malaysia Fundamental Research Grant Scheme (FRGS) fund (grant vot. number 5524740).

References

[1] Bălă, Gabriel-Petrică, Ruxandra-Mioara Râjnoveanu, Emanuela Tudorache, Radu Motișan, and Cristian Oancea. "Air pollution exposure—the (in) visible risk factor for respiratory diseases." *Environmental Science and Pollution Research* 28, no. 16 (2021): 19615-19628. https://doi.org/10.1007/s11356-021-13208-x
[2] Krishnamoorthi, D. M., M. Arunkumar, R. G. Chandini, T. Deepika, and J. Mohamed Yazar. "IOT based Air Quality Monitoring System." *Int. J. Res. Appl. Sci. Eng. Technol* 8, no. 7 (2020): 119-124. https://doi.org/10.22214/ijraset.2020.7022
[3] Balakrishnan, Kalpana, Sagnik Dey, Tarun Gupta, R. S. Dhaliwal, Michael Brauer, Aaron J. Cohen, Jeffrey D. Stanaway *et al.,* "The impact of air pollution on deaths, disease burden, and life expectancy across the states of India: the Global Burden of Disease Study 2017." *The Lancet Planetary Health* 3, no. 1 (2019): e26-e39.
[4] Thiyaneswaran, B., P. Elayaraja, P. Srinivasan, S. Kumarganesh, and K. Anguraj. "IoT based air quality measurement and alert system for steel, material and copper processing industries." *Materials Today: Proceedings* 81 (2023): 127-132. https://doi.org/10.1016/j.matpr.2021.02.696
[5] Kang, Insung, Anna McCreery, Parham Azimi, Amanda Gramigna, Griselda Baca, Kari Abromitis, Mingyu Wang *et al.,* "Indoor air quality impacts of residential mechanical ventilation system retrofits in existing homes in Chicago, IL." *Science of The Total Environment* 804 (2022): 150129. https://doi.org/10.1016/j.scitotenv.2021.150129
[6] Hofman, Jelle, Martha E. Nikolaou, Tien Huu Do, Xuening Qin, Esther Rodrigo, Wilfried Philips, Nikos Deligiannis, and Valerio Panzica La Manna. "Mapping air quality in IoT cities: Cloud calibration and air quality inference of sensor data." In *2020 IEEE SENSORS*, pp. 1-4. IEEE, 2020. https://doi.org/10.1109/SENSORS47125.2020.9278941
[7] Sridhar, K., P. Radhakrishnan, G. Swapna, R. Kesavamoorthy, L. Pallavi, and R. Thiagarajan. "A modular IOT sensing platform using hybrid learning ability for air quality prediction." *Measurement: Sensors* 25 (2023): 100609. https://doi.org/10.1016/j.measen.2022.100609

[8] Wu, Shui-Ping, Mei-Jun Cai, Chao Xu, Ning Zhang, Jia-Bin Zhou, Jin-Pei Yan, James J. Schwab, and Chung-Shin Yuan. "Chemical nature of PM2. 5 and PM10 in the coastal urban Xiamen, China: Insights into the impacts of shipping emissions and health risk." *Atmospheric environment* 227 (2020): 117383. https://doi.org/10.1016/j.atmosenv.2020.117383

[9] Kraakman, Norbertus Joannes Richardus, Javier González-Martín, Cristina Pérez, Raquel Lebrero, and Raúl Muñoz. "Recent advances in biological systems for improving indoor air quality." *Reviews in Environmental Science and Bio/Technology* 20 (2021): 363-387. https://doi.org/10.1007/s11157-021-09569-x

[10] Elleuch, Boubaker, Farah Bouhamed, Mabrouk Elloussaief, and Madi Jaghbir. "Environmental sustainability and pollution prevention." *Environmental Science and Pollution Research* 25 (2018): 18223-18225. https://doi.org/10.1007/s11356-017-0619-5

[11] Sassi, M. Saifeddine Hadj, and Lamia Chaari Fourati. "Deep learning and augmented reality for IoT-based air quality monitoring and prediction system." In *2021 International Symposium on Networks, Computers and Communications (ISNCC)*, pp. 1-6. IEEE, 2021. https://doi.org/10.1109/ISNCC52172.2021.9615639

[12] Misiukiewicz-Stepien, Paulina, and Magdalena Paplinska-Goryca. "Biological effect of PM10 on airway epithelium-focus on obstructive lung diseases." *Clinical immunology* 227 (2021): 108754. https://doi.org/10.1016/j.clim.2021.108754

[13] Samudro, Harida, Ganjar Samudro, and Sarwoko Mangkoedihardjo. "Phytoarchitecture Design for Public Roads to Combat Air Pollution Resulting from Transport Operations." *Journal of Advanced Research in Applied Sciences and Engineering Technology* 33, no. 1 (2023): 424-440. https://doi.org/10.37934/araset.33.1.424440

[14] Ameer, Saba, Munam Ali Shah, Abid Khan, Houbing Song, Carsten Maple, Saif Ul Islam, and Muhammad Nabeel Asghar. "Comparative analysis of machine learning techniques for predicting air quality in smart cities." *IEEE access* 7 (2019): 128325-128338. https://doi.org/10.1109/ACCESS.2019.2925082

[15] Anjum, Muhammad Shehzaib, Syeda Mahnoor Ali, Muhammad Ahmed Subhani, Muhammad Naveed Anwar, Abdul-Sattar Nizami, Umar Ashraf, and Muhammad Fahim Khokhar. "An emerged challenge of air pollution and ever-increasing particulate matter in Pakistan; a critical review." *Journal of Hazardous Materials* 402 (2021): 123943. https://doi.org/10.1016/j.jhazmat.2020.123943

[16] Barthwal, Anurag, and Debopam Acharya. "An IoT based sensing system for modeling and forecasting urban air quality." *Wireless Personal Communications* 116, no. 4 (2021): 3503-3526. https://doi.org/10.1007/s11277-020-07862-6

[17] Adam, Max G., Phuong TM Tran, and Rajasekhar Balasubramanian. "Air quality changes in cities during the COVID-19 lockdown: A critical review." *Atmospheric Research* 264 (2021): 105823. https://doi.org/10.1016/j.atmosres.2021.105823

[18] Singh, Dharmendra, Meenakshi Dahiya, Rahul Kumar, and Chintan Nanda. "Sensors and systems for air quality assessment monitoring and management: A review." *Journal of environmental management* 289 (2021): 112510. https://doi.org/10.1016/j.jenvman.2021.112510

[19] Anjum, Muhammad Shehzaib, Syeda Mahnoor Ali, Muhammad Ahmed Subhani, Muhammad Naveed Anwar, Abdul-Sattar Nizami, Umar Ashraf, and Muhammad Fahim Khokhar. "An emerged challenge of air pollution and ever-increasing particulate matter in Pakistan; a critical review." *Journal of Hazardous Materials* 402 (2021): 123943. https://doi.org/10.1016/j.jhazmat.2020.123943

[20] Maslyiak, Yurii, Andriy Pukas, Iryna Voytyuk, and Mykola Shynkaryk. "Environmental monitoring system for control of air pollution by motor vehicles." In *2018 XIV-th International Conference on Perspective Technologies and Methods in MEMS Design (MEMSTECH)*, pp. 250-254. IEEE, 2018. https://doi.org/10.1109/MEMSTECH.2018.8365744

[21] Kane, Malal. "Road polishing: High-resolution textures dataset of three mosaics made of different aggregates at different polishing stages." *Data in brief* 45 (2022): 108594. https://doi.org/10.1016/j.dib.2022.108594

[22] Rastogi, Krati, and Divya Lohani. "Context-aware IoT-enabled framework to analyse and predict indoor air quality." *Intelligent Systems with Applications* 16 (2022): 200132. https://doi.org/10.1016/j.iswa.2022.200132

[23] Jamei, Mehdi, Mumtaz Ali, Anurag Malik, Masoud Karbasi, Ekta Sharma, and Zaher Mundher Yaseen. "Air quality monitoring based on chemical and meteorological drivers: Application of a novel data filtering-based hybridized deep learning model." *Journal of Cleaner Production* 374 (2022): 134011. https://doi.org/10.1016/j.jclepro.2022.134011

[24] Nemade, Bhushankumar, and Deven Shah. "An IoT based efficient Air pollution prediction system using DLMNN classifier." *Physics and Chemistry of the Earth, Parts A/B/C* 128 (2022): 103242. https://doi.org/10.1016/j.pce.2022.103242

[25] Jabbar, Waheb A., Thanasrii Subramaniam, Andre Emelio Ong, Mohd Iqmal Shu'Ib, Wenyan Wu, and Mario A. De Oliveira. "LoRaWAN-based IoT system implementation for long-range outdoor air quality monitoring." *Internet of Things* 19 (2022): 100540. https://doi.org/10.1016/j.iot.2022.100540

[26] Hosnon, Mohd Hannas, Noor Anilizawatima Sulong, Abang Annuar Ehsan, and Khairul Nisha Mohd Kharuddin. "Indoor Environmental Quality in Micro & Nanoelectronics Laboratories at IMEN, Research Complex, UKM." *Journal of Advanced Research in Applied Sciences and Engineering Technology* 32, no. 3 (2023): 342-354. https://doi.org/10.37934/araset.32.3.342354