Journal of Mechatronics and Artificial Intelligence
Homepage: http://ejournal.upi.edu/index.php/jmai/
p-ISSN 3062-729X e-ISSN 3048-4227

Air Quality Classification System using Random Forest Algorithm using MQ-7 and MQ-135 Sensors with IoT-based

Aisyah Aira Putri Maharani[1*], Rizky Hamdani Sakti[1], Muhamad Fajar Imanul Haq[1], Muhamad Ajis[1], Abdu Malikh Silaban[1]

[1] Department of Mechatronics and Artificial Intelligence, Universitas Pendidikan Indonesia, Bandung, Indonesia
*Correspondence author email: aisyahairaputrim@upi.edu

ARTICLE INFO

Article History:

Submitted/Received 30 Oct 2024
First Revised 20 Nov 2024
Accepted 29 Nov 2024
Publication Date 01 Dec 2024

Keywords:
Air Quality Classification,
Random Forest,
MQ-7 Sensors,
Mq-135 Sensors,

ABSTRACT

Air is one of the essential elements in human life besides water and soil. However, currently the air quality in Indonesia is getting worse. Therefore, an air quality classification system using Random Forest algorithm based on CO and CO2 levels using IoT-based MQ7 and MQ13 sensors is needed as a smart solution. The workflow of this system begins with detecting air quality using the MQ-7 sensor for CO gas and the MQ-135 sensor for CO2 gas. Then, the classification process is carried out with the Machine Learning Random Forest algorithm by utilizing a number of training data that has been stored in the program to classify the gas sensor detection results into three types of classes, namely “Good”, “Bad”, or “Toxic”. The final output of this system is a website display that can be accessed on a PC/Laptop monitor in real time. From the results of the Random Forest machine learning algorithm classification testing process, 1 unsuitable data was found from a total of 100 trials that have been carried out. Therefore, the Random Forest machine learning algorithm can be said to be successful in detecting air levels in the surrounding environment well because it provides an accuracy value of 99%.

1. Introduction

Air is one of the important elements in human life besides water and soil [1]. This element is the main component supporting human survival in the world to carry out various activities in it. Air contains various substances such as oxygen, carbon monoxide, carbon dioxide, and so on that are needed by humans. The levels of these substances must of course be at the proper normal limits. According to the World Health Organization (WHO), substance levels that are not in accordance with normal limits can be harmful to health [2].

Currently, air quality in Indonesia is getting worse. This is supported by data from databooks which states that the Central Kalimantan area has an Air Pollution Standard Index (ISPU) value that reached 240 on October 5, 2023 [3]. According to the Directorate of Air Pollution Control (MoEF), ISPU is a unitless number used to describe air quality conditions at a particular location based on the impact on human health.

based on the impact on human health, aesthetic value, and other living things. The calculation of the ISPU value is based on the measurement results of seven air pollutant parameters namely PM10, PM2.5, NO2, SO2, CO, O3, and HC [4]. Based on the Minister of Environment and Forestry Regulation No. 14 of 2020 concerning the Air Pollutant Standard Index, ISPU in the range of 0-50 has good air quality, the range of 51-100 means moderate air quality, and the range of 101-200 unhealthy air quality which is detrimental to humans, animals, and plants. Therefore, the ISPU value that reaches 240 in one of the regions in Indonesia is a condition that must be a concern so that an innovation is immediately sought as a solution.

existence of a monitoring system, the Indonesian people can know what steps to take appropriately in these conditions. This has been done by previous researchers who created similar systems with their respective methods [5]–[9].In the study, they utilized MQ-7, MQ-135 and MG-811 sensors. From the test results, the system built from the research successfully detected certain levels contained in the air. One of the existing studies [10]–[12] has also utilized the Internet of Things (IoT) to send the detected air quality values to an IoT platform to be monitored in real-time by users.

Based on these previous studies, no one has combined an air level detection system using IoT with a Machine Learning classification algorithm. In fact, these two things can become a complex system that is an innovation in preventing the decline in air quality in Indonesia. Therefore, this research carries the Random Forest machine learning algorithm to be able to classify CO and CO2 value data obtained previously from the measurement results using IoT-based MQ-7 and MQ-135 sensors as a novelty. The CO and CO2 values were chosen to represent the existing air quality. Then, the Machine Learning Random Forest algorithm was chosen to be able to classify the values obtained from the sensor into three states namely “good”, “bad”, and “toxic”. Thus, this system is expected to be a useful system for its users because it can display real-time air quality data that can be accessed anytime through the website.

2. Methods

2.1. *Block Diagram of the System*

A basic explanation of how the system works as a whole can be seen in the block diagram in Figure 1 below:

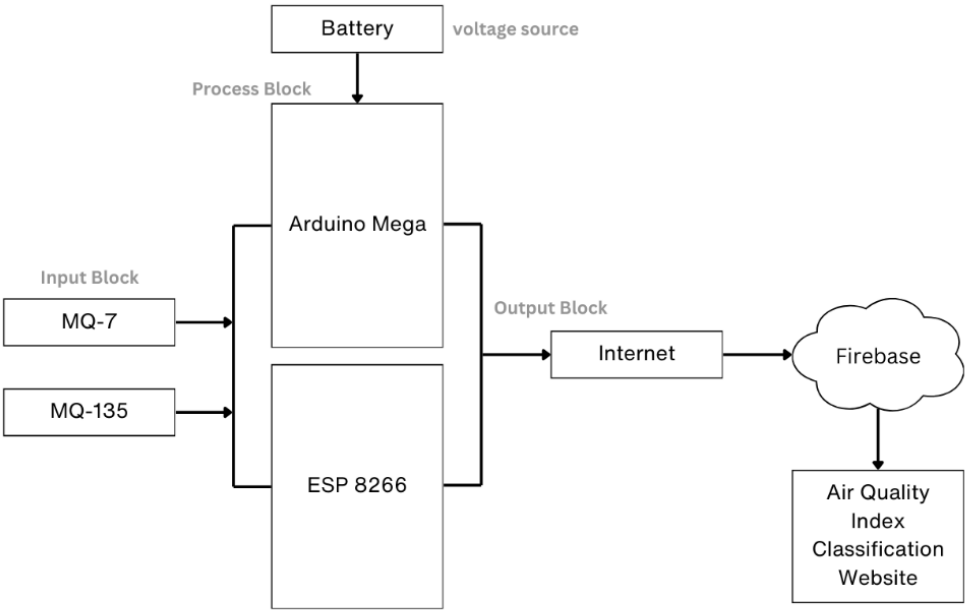


Figure 1. Block Diagram of the System.

2.2. *Random Forest Algorithm*

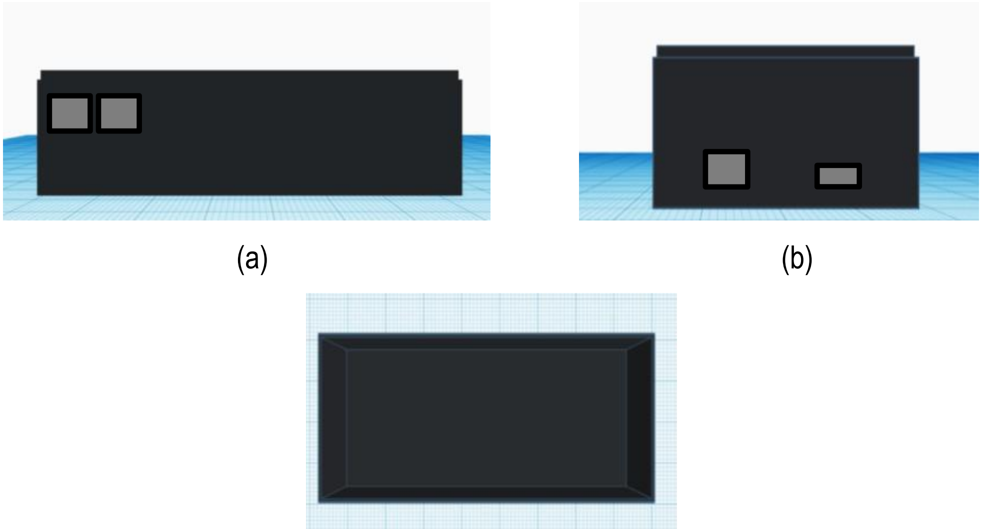
A Random Forest (RF) model was designed to classify air quality based on MQ-7 and MQ-135 sensor data. RF was used due to its reliable ability to handle noisy data and interacting features. The hyperparameters used in this model can seen in Table 1. The selection of hyperparameter values is done through grid search with k-fold cross-validation (k=5) to obtain the optimal configuration. These results ensure the RF model can effectively classify air quality into 'Good', 'Medium', and 'Poor' categories based on sensor data.

Table 1. Random Forest Hyperparameters

Parameters	Description	Amount
n_estimators	to maintain a balance between accuracy and computation time	100
max_depth	to prevent overfitting of the training data	10
min_samples_split	to avoid the information of significant nodes.	5
Min_samples_leaf	To improve the generelizabity of the model	2

2.3. *Design of the System*

A comprehensive system description based on the packaging design can be seen in Figure 2 below:



(a) (b)

(c)

Figure 2. Overview of Packaging System (a) Front View (b) Side View (c) Inside View.

The system design will be packaged in a box measuring 13x9x7cm. The entire system wiring will be placed inside the box as shown in Figure 2 (c). Then, there are two round holes in the front of the box as shown in Figure 2 (a) as a place for MQ-7 and MQ-135 sensors to detect CO and CO2, as well as two box holes on the side as shown in Figure 2 (b) to place the USB cable so that the Arduino uno microcontroller and ESP32 can be connected to the laptop. The system schematic design can be seen in Figure 3 below:

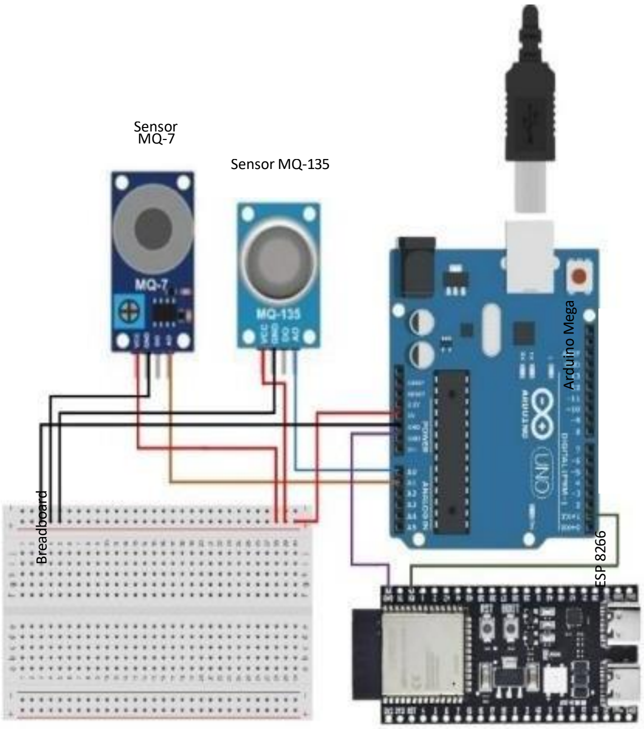


Figure 3. Schematic System.

Based on the system schematic in Figure 3 above, the main control of the system is on the Arduino Uno and ESP32 which have been connected with a number of jumper cables to a number of hardware components. There are two gas sensors, MQ-7 and MQ-135, connected to the system. The hardware is programmed in a software design in Figure 4 below.

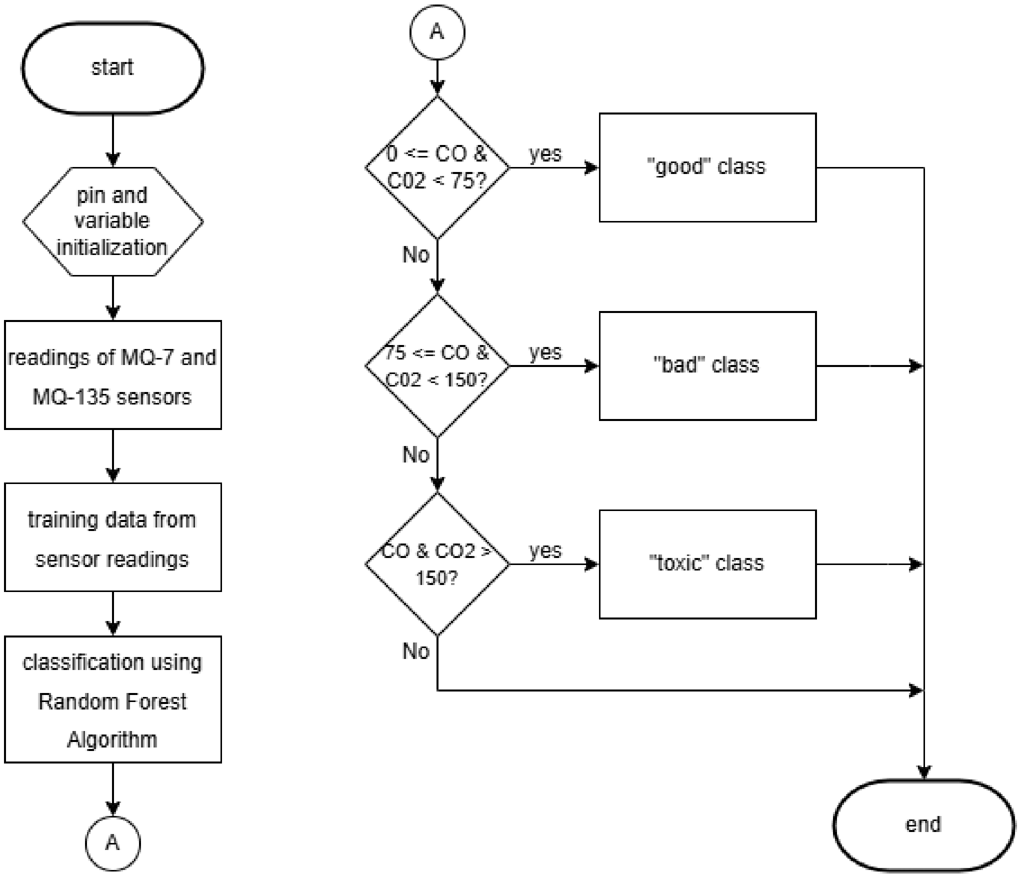
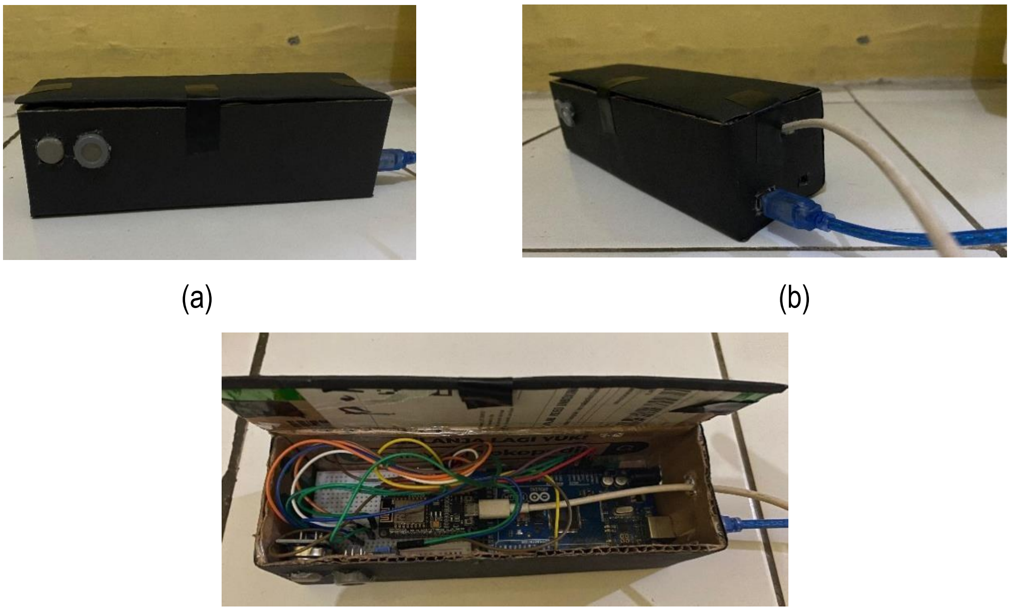


Figure 4. System Workflows.

The system work stage begins with the initialization of certain pins and variables followed by reading the MQ-7 and MQ135 sensors. The results of the reading of the two sensors are in the form of part per million or ppm values. This set of readings will later be used as test data to be classified by the Random Forest machine learning algorithm. But before that, we first prepare training data or training data which is divided into three types of classes, namely the “Good”, “Bad”, and “Toxic” classes. The amount of training data for each class is 1200 data. After the training data is trained, random forest classification can be carried out by utilizing test data obtained from the detection results of MQ-7 and MQ135 sensors. Random Forest classification is a class condition called “Good”, “Bad”, and “Toxic”.

2.4. *System Implementation*

The implementation of the system design produces a prototype that can be seen in Figure 5.



(a) (b)

(c)

Figure 5. System Implementation (a) Front View (b) Side View (c) Inside View.

The entire hardware implementation is interconnected with jumper cables, with the use of Bread board as a means of inter-pin wiring of each component so that the results are more neatly organized.

2.5. *Website Implementation*

The implementation of the final system design results in a website display that can be seen in Figure 6.

Figure 6. Website Implementation.

The website display is designed using the Figma application. In this display, there are features of location, time, CO levels, CO2 levels, and class status that can facilitate users in seeing air conditions in the surrounding environment.

3. Results and Discussion

Testing and analysis is divided into 2 main stages, namely the testing stage for each MQ-7 and MQ-135 gas sensor, and the accuracy testing stage of the Random Forest classification.

3.1. *MQ-7 and MQ-135 Sensor Testing*

This test was carried out with the aim of obtaining the ppm value of the MQ-7 and MQ-135 sensors from the detected test environment conditions. Data was collected in the UPI environment over a 2-minute period, with sensor readings taken every second. The total dataset obtained was 1200 overall data covering air quality variations ranging from 'Good', 'Medium', to bad'.

Table 2. MQ-7 and MQ-135 Sensor Testing.

Time	CO Value (ppm)	CO2 Value (ppm)	Condition
15:02:02	167	360	Toxic
15:02:11	30	133	Good
17:54:11	64	196	Good
13:31:29	155	155	Bad
17:54:32	64	195	Good
….	….	….	…
18:44:37	82	177	Bad
18:44:59	89	182	Bad
18:45:21	115	297	Bad
18:45:43	82	253	Bad
18:46:04	70	223	Good

Table 1 shows the changes in ppm values of the MQ-7 and MQ-135 sensor detection results from the test environment. The ppm value of the test varies with a tendency to be in a "bad" condition. This indicates that the test environment does not have good enough air conditions for humans to breathe.

3.2. *Random Forest Machine Learning Algorithm Accuracy Testing*

This test aims to determine the amount of accuracy obtained from the Random Forest machine learning algorithm in classifying the detected environmental conditions. There are 100 trials in Random Forest testing. The results of the Random Forest machine learning algorithm classification test accuracy in Table 3 below.

Table 3. Random Forest Machine Learning Algorithm Accuracy Testing.

Trial to-	Clasification Result	Real Condition
1	Toxic	Toxic
2	Good	Good
3	Good	Good
4	Bad	Bad
5	Good	Good
6	Good	Good
7	Bad	Bad
8	Bad	Bad
9	Bad	Bad
10	Bad	Toxic
…	**….**	….
91	Good	Good
92	Toxic	Toxic
93	Bad	Bad
94	Good	Good
95	Toxic	Toxic
96	Toxic	Toxic
97	Bad	Bad
98	Bad	Bad
99	Good	Good
100	Bad	Bad

Based on the results of the test experiment, 1 data classification result of the Random Forest machine learning algorithm was obtained which did not match the original test conditions. Therefore, a calculation is needed to determine the accuracy value using the following equation (1).

$$\text{Accuracy} = \frac{\text{Number of Suitable Classification}}{\text{Total Test Data}} \times 100\% \tag{1}$$

Based on equation 1 above, the accuracy value of the Random Forest machine learning algorithm classification testing experiment is 99%.

4. Conclusion

The design of this air quality detection system is carried out by utilizing 2 types of gas sensors, namely MQ-7 and MQ-135 based on ESP8266 and Arduino Mega microcontrollers. This system also uses Random Forest machine learning algorithm to classify CO and CO2 gas levels in ppm detected by the system. The classification results will be divided into 3 classes, namely “good”, “bad”, and “toxic”.

In the process of implementing the Random Forest machine learning algorithm classification test, 1 data was found that was not suitable from a total of 100 trials that had been carried out. Therefore, the Random Forest machine learning algorithm can be used well in detecting air levels in the surrounding environment because it successfully provides an accuracy value of 99%.

References

[1] I. Manisalidis, E. Stavropoulou, A. Stavropoulos, and E. Bezirtzoglou, "Environmental and Health Impacts of Air Pollution: A Review," *Front. Public Heal.*, vol. 8, no. February, pp. 1–13, 2020, doi: 10.3389/fpubh.2020.00014.

[2] M. Szabados *et al.*, "Indoor air quality and the associated health risk in primary school buildings in Central Europe – The InAirQ study," *Indoor Air*, vol. 31, no. 4, pp. 989–1003, 2021, doi: 10.1111/ina.12802.

[3] B. Sidabutar, E. Kurniati, and K. Adelia, "Evaluasi Kualitas Udara Emisi Gas Buang/Ambien di PLTD Kahayan Baru," *J. Phys. It's Appl.*, vol. 3, no. 1, pp. 226–230, 2023.

[4] A. A. Rosa, B. A. Simon, and K. S. Lieanto, "Sistem Pendeteksi Pencemaran Udara Portabel Menggunakan Sensor MQ-7 dan MQ-135," *Ultim. Comput. J. Sist. Komput.*, vol. 12, no. 1, pp. 23–28, 2020, doi: 10.31937/sk.v12i1.1611.

[5] J. Saini, M. Dutta, and G. Marques, "A comprehensive review on indoor air quality monitoring systems for enhanced public health," *Sustain. Environ. Res.*, vol. 30, no. 1, pp. 1–12, 2020, doi: 10.1186/s42834-020-0047-y.

[6] D. Singh, M. Dahiya, R. Kumar, and C. Nanda, "Sensors and systems for air quality assessment monitoring and management: A review," *J. Environ. Manage.*, vol. 289, no. November 2020, p. 112510, 2021, doi: 10.1016/j.jenvman.2021.112510.

[7] A. Albertraj and J. Vijila, "Design of Indoor Air Quality Monitoring System to Ensure a Healthy Universe," in *Proceedings - International Conference on Smart Electronics and Communication, ICOSEC 2020*, 2020, pp. 1123–1127. doi: 10.1109/ICOSEC49089.2020.9215297.

[8] U. Rahardja, Q. Aini, D. Manongga, I. Sembiring, and I. D. Girinzio, "Implementation of Tensor Flow in Air Quality Monitoring Based on Artificial Intelligence," *Int. J. Artif. Intell. Res.*, vol. 6, no. 1, pp. 2579–7298, 2023, doi: 10.29099/ijair.v6i1.430.

[9] M. Rogulski and A. Badyda, "Current trends in network based air quality monitoring systems," in *IOP Conference Series: Earth and Environmental Science*, 2019. doi: 10.1088/1755-1315/214/1/012085.

[10] G. Marques, C. R. Ferreira, and R. Pitarma, "Indoor Air Quality Assessment Using a CO 2 Monitoring System Based on Internet of Things," *J. Med. Syst.*, vol. 43, no. 3, 2019, doi: 10.1007/s10916-019-1184-x.

[11] J. Buelvas, D. Múnera, D. P. Tobón V, J. Aguirre, and N. Gaviria, “Data Quality in IoT-Based Air Quality Monitoring Systems: a Systematic Mapping Study,” *Water. Air. Soil Pollut.*, vol. 234, no. 4, 2023, doi: 10.1007/s11270-023-06127-9.

[12] J. Jo, B. Jo, J. Kim, S. Kim, and W. Han, “Development of an IoT-Based indoor air quality monitoring platform,” *J. Sensors*, vol. 2020, pp. 13–15, 2020, doi: 10.1155/2020/8749764.