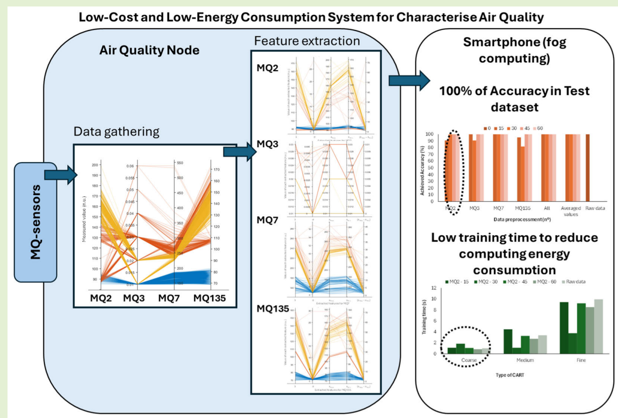
Wearable Low-Cost and Low-Energy Consumption Gas Sensor With Machine Learning to Recognize Outdoor Areas

Jianchen Wang, Lorena Parra, Raquel Lacuesta, Jaime Lloret, *Senior Member, IEEE*, and Pascal Lorenz, *Senior Member, IEEE*

***Abstract*—Urban air quality, impacted by human-made pollution, impacts health and requires continuous monitoring. MQ sensors are the preferred air quality sensors despite their high energy consumption due to their cost, requiring the use machine learning to classify different types of air. The aim of this article is to evaluate a monitoring solution with low-cost and low-energy consumption to classify urban and rural air. A single MQ sensor will be used with a network with edge and fog computing to balance the energy consumption. Edge computing was included in the node for feature extraction, and fog computing was applied in the smartphone to classify the data using machine learning. Different sensors and time buffers are compared in order to find the adequate sensor for data generation and time buffer for feature extraction. The results indicate that it has been possible to achieve accuracies of 100% using a single sensor, the MQ2, with time buffers of 45–60 measures. With this proposal, it is possible to reduce the energy consumed by data gathering to 25% of the original consumption due to the use of a single sensor, due to the reduction in the sensors used in the previous prototype. Moreover, it has been possible to reduce the energy linked to data forwarding by almost 97% due to using a time buffer.**



***Index Terms*—Air pollution, edge computing, fog computing, MQ sensor, rural area, urban area.**

I. Introduction

AIR quality is strongly affected by anthropogenic pollution, and in the last decades, the air quality in the cities due to traffic has become a health issue [1]. Several countries are boosting policies to regulate the use of vehicles in the inner parts of the cities to reduce pollution and restore urban air quality. Besides traffic, the industry is another important component in air pollution sources [2]. Meteorological conditions might play a positive or negative role in air quality according to multiple factors [3].

Air quality has a direct impact on human health [4], and therefore, it should be monitored in particular cases. Some of these cases include people with diseases or those in high-population-density areas. In comparison with rural areas, urban areas usually have worse air quality due to the low anthropogenic impact in rural areas. Nevertheless, we can distinguish different regions in the urban areas characterized by different degrees of atmospheric pollution. Areas near highways tend to have more pollution than regions surrounded by green areas without traffic. Similarly, regions close to industrial areas have more pollution than the residential parts of the city.

Several institutional efforts are being conducted beyond the regulatory frameworks to assess the air quality and know the real-time distribution of air pollution [6]. Air quality

Manuscript received 3 July 2024; accepted 9 August 2024. Date of publication 19 August 2024; date of current version 2 October 2024. This work was supported in part by the Spanish Science and Innovation Ministry under Contract PID2022-136779OB-C31; and in part by Conselleria de Educación, Universidades y Empleo through the Subvenciones para estancias de personal investigador doctor en centros de investigación radicados fuera de la Comunitat Valenciana (Convocatoria 2023) under Grant CIBEST/2022/40. The associate editor coordinating the review of this article and approving it for publication was Prof. Xiaofeng Yuan. *(Corresponding author: Jaime Lloret.)*

Jianchen Wang and Raquel Lacuesta are with the Department of Computer Science and Engineering of Systems, University of Zaragoza, 50009 Zaragoza, Spain (e-mail: 802610@unizar.es; lacuesta@unizar.es).

Lorena Parra and Jaime Lloret are with the Instituto de Investigación para la Gestión Integrada de Zonas Costeras, Universitat Politècnica de València, 46730 Valencia, Spain (e-mail: loparbo@doctor.upv.es; jlloret@dcom.upv.es).

Pascal Lorenz is with the Network and Telecommunication Research Group, University of Haute Alsace, 68100 Colmar, France (e-mail lorenz@ieee.org).

Digital Object Identifier 10.1109/JSEN.2024.3442874

monitoring solutions predominantly occur within urban cores or built-up areas, where monitoring encompasses various pollutants, such as particulate matter (PM), volatile organic compounds (VOCs), biological contaminants, and aerosols, as outlined in [7]. Nevertheless, the existing solutions have high costs and often do not provide real-time information. Thus, low spatial resolution of data is achieved, which makes it difficult to evaluate the exposition of people to pollution.

Gas sensors are frequently deployed for measuring VOCs across diverse applications, as noted in [8]. Furthermore, certain compounds have the potential to react with other components in the air, generating hazardous substances [9]. Therefore, the utilization of gas sensors holds a significant role in implementing monitoring solutions. Among the widely recognized low-cost gas sensors are those belonging to the MQ family. Other examples include the Figaro TGS [10] and FIS [11] sensors. These sensors operate based on resistance changes corresponding to the presence of specific gases. Nevertheless, recent studies suggest that the MQ family yields higher accuracies when combined with AI for data classification [8], [12]. According to recent research findings, MQ sensors are utilized for air monitoring, offering dependable and cost-efficient performance despite challenges related to calibration, as highlighted in [13].

Nonetheless, the individual use of these sensors is rarely seen. In most cases, the authors combined multiple sensors from the same group or different groups to monitor air quality. Even though multiple studies have been found, doubts about the performance of these sensors are due to the questionable specificity for the targeted molecules, which prevents their use for monitoring the exact amounts of air pollutants. Therefore, multiple sensors and ML were used to characterize the measured gas. The application of these sensors to discriminate between indoor and outdoor sensed data using four gas sensors was proposed by Wang et al. [13]. Unfortunately, the authors only collected from urban areas. In addition, the energy consumption of these sensors is high, and it must be considered to try to reduce the required sensors to characterize the studied area. As far as we are concerned, no cases in which a single sensor is used for this purpose have been proposed.

The aim of this article is to evaluate the suitability of using a single sensor to recognize diverse outdoor areas, including urban and rural areas. For that purpose, four sensors are considered: MQ2, MQ3, MQ7, and MQ135. An Arduino Mega has been selected as a microcontroller, which will perform edge computing to extract features from the MQ sensors. The extracted features are sent via Bluetooth to a smartphone, which analyzes locally thought fog computing with ML and classifies the area. Different time buffers are compared to calculate five features (average, standard deviation, minimum, maximum, and range) to extract features. The objectives of the present article and their novelties compared with the existing systems are the following.

1) Reduce the number of MQ sensors needed to provide a low-cost and low-energy consumption solution to classify types of outdoor areas by extracting features from gathered MQ data.
2) Assess which sensor provides the best-extracted features for outdoor classification with different time buffers.
3) Compare the ML models to select the one that balances performance and computing requirements.
4) Test the system in three real outdoor areas with different air quality and other environmental parameters, including a residential district, a green area, a large city, and a natural area.

The rest of this article is organized as follows. Section II summarizes the existing proposals with a limited number of sensors. Section III details the proposed approach and used elements. Then, the test bench is described in Section IV. In Section V, the results that were obtained are discussed. Finally, Section VI outlines the main conclusion and future work.

II. Related Work

In this section, a summary of various studies that have employed gas sensors is provided. These devices were used for monitoring air quality but were also found to be applicable in diverse areas. Initially, we highlight sensors' primary roles in monitoring indoor and outdoor air quality within urban settings using a single sensor. Then, the combined use of multiple sensors is outlined.

A. Single Sensor for Air Quality in Indoor and Outdoor Urban Scenarios

In this subsection, the existing proposals using single gas sensors, mainly MQ-based sensors, are outlined.

Irawan et al. [14] proposed the use of MQ135 for air quality and the presence of gas pollutants monitoring. In their proposal, the selected node was the Arduino UNO to collect the data and send it to a Raspberry Pi. The authors used the sensor to quantify the amount of CO, CO_2 gases, and alcohol in real time. Nevertheless, the authors do not compare the lecture of the sensor with a verified device or expose the sensor to known concentrations of measured gases. Thus, it is not feasible to evaluate the accuracy of the proposed system.

Rani et al. [15] used the MQ135 with an Arduino Uno to determine air quality and identify hazardous conditions in indoor areas. To evaluate the response of the sensor, they were exposed to different substances. In the first scenario, the sensor was exposed to a candle, placing it at 0.2 cm. In the second scenario, agarbhatti at 1, 0.5, and 0.2 cm was used. The authors confirmed that the response of the sensors varies when it is exposed to different substances and at different distances. Nevertheless, data were not used to classify the different scenarios, and the dataset constitutes a very specific case with no real applications to identify different environments.

Mluyati and Sadi [16] tested a gas sensor system employing the MQ2. An Arduino Nano was used as a microprocessor. Their research aimed to identify gas leaks, utilizing wireless communication via short message service. Their data highlighted the ability to detect gas leaks at levels of 52%.

Stančić et al. [17] used the MQ2 sensor to detect smoke and combustion gas inside homes. For instance, they burned paper, cigarettes, and gases, such as the ones from lighters and stoves. They concluded that the MQ2 sensor, together with the use of other sensors, could improve current smoke detectors.

B. Multisensor for Air Quality in Indoor and Outdoor Urban Scenarios

In this subsection, the current systems in which more than one MQ-based sensor or other similar sensors are included are summarized.

Rustemli et al. [18] used the MQ2, MQ4, MQ9, and MQ135 to determine the air quality by monitoring the CO, LPG, and smoke in indoor and outdoor areas. The nodes were connected to an Arduino Mega prototype. As a result, the authors provide lectures on sensors in the measured location (restaurant, office, and outdoor environments). Moreover, the authors did not control the concentration of measured gases in the areas or analyze the different lectures in diverse areas in depth. Thus, it is impossible to evaluate the proposed system's accuracy.

The same year, a prototype composed of MQ-7, Sharp GP2Y1010AU0F, and an own-designed smoke sensor, was proposed by Rumantri et al. [19]. The authors employed Arduino microprocessors, but they do not identify which. The prototype was used to measure the concentration of carbon monoxide. The authors compare the sensors measured with a verified instrument in order to provide a new calibration of the proposed system. Data were gathered by injecting carbon monoxide into a small container. A correlation has been provided, but the authors do not indicate the performance of this correlation in terms of R2 or other metrics.

Sai et al. [20] combined the MQ7 with the MQ135 to detect carbon dioxide and air quality. The authors used an Arduino UNO to power the nodes and read the measurements. Data are forwarded to Thingspeak for its visualization. Their results encompass a series of measures with the system and the analyses of a range of obtained values. Nonetheless, the authors did not compare the sensor readings with those of a validated device or evaluate the response of the sensors exposed to known concentrations of measured gas. As a result, assessing the accuracy of the proposed system is not possible.

Hapsari et al. [21] proposed a combination of MQ135 and dust sensors. The selected node, in this case, is the ESP32. Measured parameters include the concentration of carbon dioxide and the PM in the air. Data were gathered in three indoor scenarios at the university: the classroom, the canteen, and the library. The authors present and analyze the variability of data among the measured scenarios. Nevertheless, no classification of the scenarios was conducted based on the lecture on the sensors.

AI-Okby et al. [22] proposed the use of SGP40 and BME680 sensors to generate an alarm system for detecting hazardous chemical substances in indoor environments. As a microprocessor, the authors selected the WeMos D1 Mini IoT microcontroller, which serves to collect and transmit data to the cloud using Wi-Fi. The system was tested using ethanol, hexane, and acetone at different distances (25 and 40 cm) from the sensors. Even though this system can help to detect small leaks in the laboratory, their usability in real environments is limited. In addition, no relation or classification between gathered data and scenarios was conducted.

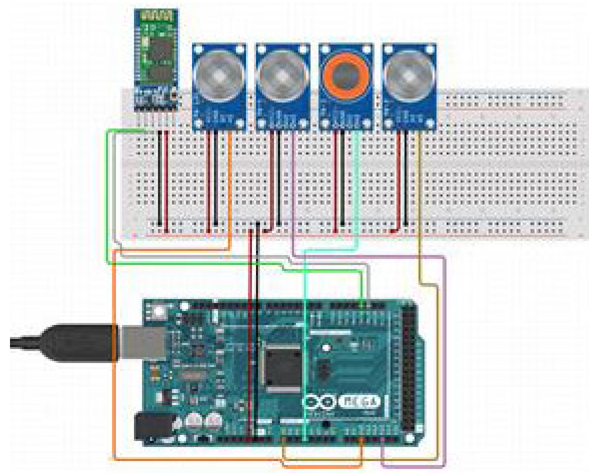


Fig. 1. Proposed wearable gas sensor device.

TABLE I
SUMMARY OF INCLUDED CART

ID	Name	Maximum number of splits	Split criterion	Surrogate decision splits
1	Fine Tree	100	*Gini's diversity index*	No
2	Medium Tree	20	*Gini's diversity index*	No
3	*Coarse Tree*	*4*	*Gini's diversity index*	*No*

Jabbar et al. [23] combined the MQ9, MQ135, MQ36, MiCS-4514, and PMS3003 sensors for monitoring air quality in outdoor scenarios. The selected node is Arduino Uno, and the data are transmitted using LoRa technology. On this occasion, sensors were calibrated for the following gases: CO_2, NO_2, SO_2, and CO. The calibrations were conducted using two known concentrations of gases. Nevertheless, in some cases, just one concentration is used. Thus, calibration might not be optimal. During the tests, differences in the concentration of gases were minimal. Obtained data were forwarded to the ThingSpeak IoT server.

III. PROPOSAL

In this section, all the elements of the proposed system are described. First of all, the sensors and the microprocessor are identified. Then, edge computing, including feature extraction, is presented. Following this, fog computing, including the ML classifier selection, is explained. Finally, the details of the conducted ML classification are outlined.

A. Wearable Sensing Device

The proposed wearable sensing devices are based on the proposal presented in [13] (see Fig. 1). Arduino Mega 2560 [24] is selected as a microprocessor due to its high number of inputs and processing capability. This computing capability is required since edge computing will be conducted for initial data preprocessing in order to minimize the consumed energy and bandwidth for data forwarding.

The sensing device is composed of four gas sensors: the MQ-2 [25], MQ-3 [26], MQ-7 [27], and MQ-135 [28]. Those sensors measure the concentration of different substances in the atmosphere, as can be seen in Table I. Even though there are currently four sensors in the tested prototype, the final prototype will have only one of them. In addition to the sensors, a Bluetooth module is incorporated to enable connectivity with the smartphone, facilitating data transfer.

B. Edge Computing and Feature Extraction

The main difference between this proposal and the existing ones [29] is the combination of edge and fog computing with the aim of reducing the number of used sensors and, thus, the consumed energy. With this proposal, data will be initially processed in the node, extracting a series of features that will be part of the input for the classification conducted by the server. Feature extraction aims to evaluate if it is feasible to recognize the area with a lower number of sensors. Thus, it will be able to reduce the energy consumption and the network requirements for the sensor operation. The extracted features include the ones described in (1) to (2) as average and standard deviation and three additional ones: the minimum, maximum, and range. The periodicity of feature extraction is one of the studied parameters in this article

$$\bar{x} = \frac{\sum_{i=1}^{N} x_i}{N} \tag{1}$$

$$\sigma = \sqrt{\frac{\sum_{i=1}^{N} (x_i - \bar{x})}{N}} \tag{2}$$

where x_i represents the data gathered at a given moment, N is the time at which metrics are calculated, $\bar{x}$ is the averaged data, and σ is the standard deviation. These features are extracted for each one of the four MQ sensors.

The obtained features are then sent to the smartphone, and fog computing is included to classify the data using an ML-based classifier.

C. Fog Computing and ML Classifier Selection

We have selected an ML classifier multiclass algorithm to classify the extracted features into three classes. To select the ML algorithm, we have considered those that are able to use nonparametric and nonlinear features since the normality of data cannot be ensured during the real application of the system. Even though the use of averaged and standard deviation cannot be affected by outliers, the minimum and maximum parameters of the extracted features are sensitive to outliers. Thus, a method which is not sensitive to outliers must be used. Finally, we searched methods usually employed with natural data. The selected method has been the classification and regression tree (CART). Three configurations for CART, which are identified in Table I, have been selected. To evaluate the performance of CART, the metrics of accuracy and cost in the validation and test datasets classification will be compared.

This classification will be conducted on the smartphone. Having the ML classification in the smartphone allows two advantages. The first one is that the user can check that the conducted classification is correctly done, allowing reinforcement learning in the future based on the inputs of the users. The second one is linked to the data forwarding and server storage capacity. If ML classification should be carried in the database, the metrics should be sent and processed. With

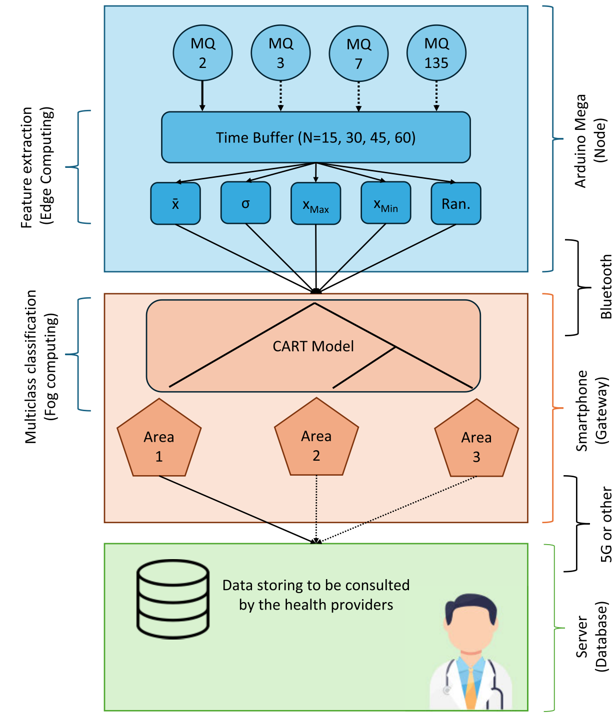


Fig. 2. Proposed architecture.

fog computing, the server will only receive the type of area, reducing the storage requirements.

To better understand the proposal, it has been summarized as a flowchart in Fig. 2. Even though four gas sensors are represented, only one will be used according to the results. In the same way, only the type of outdoor area selected by the CART is sent to the server. We can identify the tasks that have been conducted, the type of computing, and the devices.

D. Conducted ML-Based Classification

The obtained dataset was split into a training dataset (65%), a validation dataset (25%), and a test dataset (10%). We have selected the holdout validation method as a validation method, given the large datasets to avoid overfitting between training and validation datasets.

Concerning the feature extraction, we have applied four values of N (15, 30, 45, and 60). In addition, a classification will be conducted with the raw data to have the expected maximum metrics. Thus, a total of four datasets are generated for each location. Besides the data of the individual sensor, two additional datasets were generated, the first one having the extracted features for all the nodes and the second with the averaged value of each node. Table II outlines the used datasets.

To assess the classification, the obtained datasets for each feature extraction periodicity and each sensor are compared in the results. Thus, it is possible to discern which of the sensors provides more useful features to reduce the energy requirements and size of the wearable devices. MATLAB R2022b [30] is used to perform the multiclass classification.

IV. Test Bench

In this section, the test bench is fully detailed. First of all, the measured procedure is explained. Then, the areas in which

TABLE II
SUMMARY OF REGISTERS AND FEATURES IN COMBINATIONS OF SENSORS

N	MQ2	MQ3	MQ4	MQ135	All sensors		
	All features					Average	Raw
0							1922 /4
15	224 /5	224 /5	224 /5	224 /5	224 /20	224 /4	
30	112 /5	112 /5	112 /5	112 /5	112 /20	112 /4	
45	73 /5	73 /5	73 /5	73 /5	73 /20	73 /4	
60	56 /5	56 /5	56 /5	56 /5	56 /20	56 /4	

data have been collected are described. Finally, the selected ML-based classifiers and the metrics used for the assessment are identified.

A. Measurement Procedure

Following the recommendation of multiple authors and manufacturers indications, before starting the data gathering, the system was used in a controlled environment for a period of 48 h. Thus, we ensure that the sensors are clean and properly heated before starting the measurements. The measuring system is then used in the different locations, ensuring that the initial 2 h of data is discarded and the remaining data provide at least 400 records.

B. Measured Scenarios

The measured scenarios for this article include three outdoor areas with different characteristics. Of these three regions, two are located in urban areas, one being a residential district, and the other is a green area in a city. The third location is a region with natural vegetation in a rural area.

Regarding the first two locations, the residential district is located 30 m from a boulevard and 250 m from a highway. Meanwhile, the green area is located close to the boulevard but far from the highway. In both areas, traffic is high, but vegetation is a different characteristic. Data were gathered from morning to midday in these locations and collected at Dongying, Dongying, Shandong, China.

Finally, the last studied area was located at 70 km of Chongqing and at 20 km of the closest residential district, Nanchuan. The measurement point was selected on the mountain slope opposite the large residential areas. Data were gathered in the morning in this location.

V. RESULTS

In this section, the results of testing the proposed system in different environments are detailed. First, we analyze the sensed data in different environments. Then, we evaluate which sensor is most useful to identify whether the data are sensed indoors or outdoors.

A. Characterization of MQ Data in Outdoor Areas

To begin with, the comparison of obtained data in each area is shown as a parallel coordinates plot. In this graphic, every

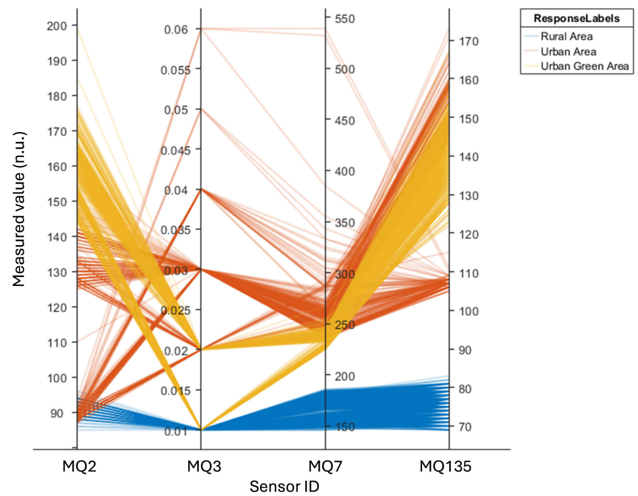


Fig. 3. Gathered data from different MQ sensors.

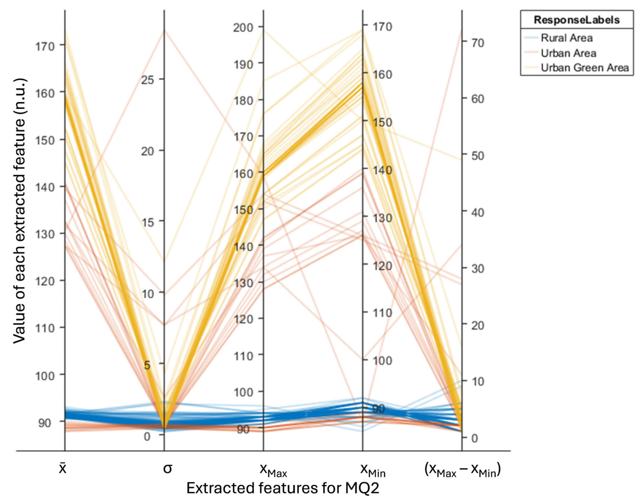


Fig. 4. Extracted features from MQ2.

gathered data group in each location is plotted (see Fig. 3). Raw data have been transformed to be plotted by adjusting the axis dimensions to the range of sensed data to allow the figure to be readable. It is possible to recognize that with raw data, using a single sensor, it is not possible to differentiate between the types of areas. While the rural area is characterized by lower values in all the sensors, the urban area and the urban green area have similar values for MQ2 and MQ135. For MQ3, the urban area is always the one with higher values.

Following, we analyze the distribution of extracted features for $N = 15$ of all the MQ sensors. Regarding the MQ2, see Fig. 4. It is possible to see that some features, such as averaged, minimum, and maximum values, are suitable for differentiating between different areas. In contrast, others, such as range or standard deviation, have similar values for all the regions. Moreover, it is possible to identify that there is a great variability of gathered data. This variability might indicate that the sensors are able to measure the small changes in air quality and other environmental factors due to the variation of gases in the atmosphere caused by human actions.

The urban green areas are characterized by having greater values in averaged, minimum, and maximum measured data. It should be highlighted that some of the recorded values of

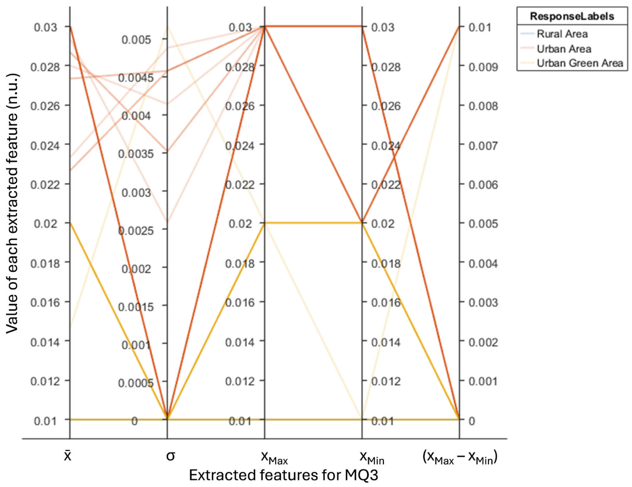


Fig. 5. Extracted features from MQ3.

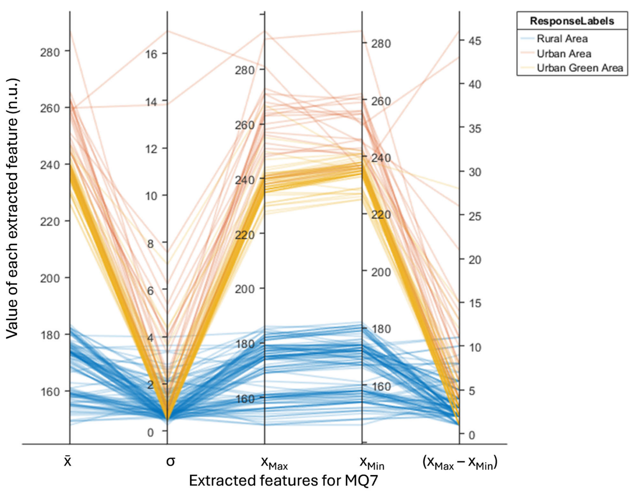


Fig. 6. Extracted features from MQ7.

MQ2 in urban areas are very similar to those from rural areas, even lower. This might indicate that even in areas strongly affected by human impact, such as traffic, there are moments in which the composition of the atmosphere is similar to that in rural areas. The data from urban areas seem to be divided into two groups, one characterized by values similar to the urban green areas and the other with values similar to the rural area.

A totally different trend is observed concerning the MQ3, which data can be seen in Fig. 5. In this case, there is an extremely low dispersion of generated data. This fact indicates that this sensor might not be suitable for identifying spatiotemporal changes in the air quality data [31]. In fact, the standard deviation is 0 in almost all the data that were analyzed. The urban area has the highest values for all the extracted features. For the MQ3 sensor, data from rural areas have no variability and are similar to some records conducted in green urban areas.

Concerning MQ7 data, see Fig. 6, which shows a greater dispersion than in the previous cases. In MQ3, the dispersion was minimal. Nevertheless, MQ7 has higher dispersion even for rural area data than MQ2. This might indicate a higher affinity of the sensor for the differences in the presence of gases in this area or a very unstable measurement. Considering

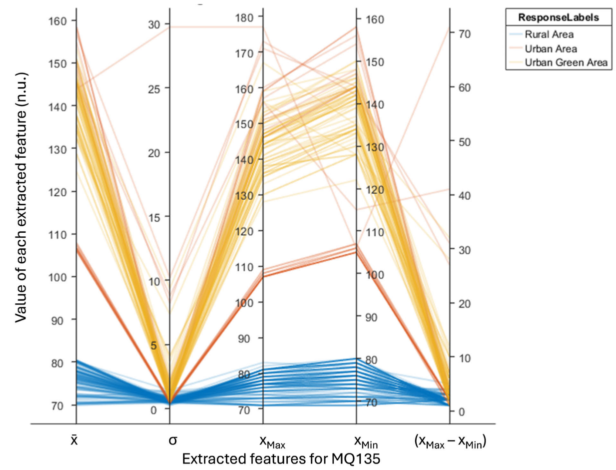


Fig. 7. Extracted features from MQ135.

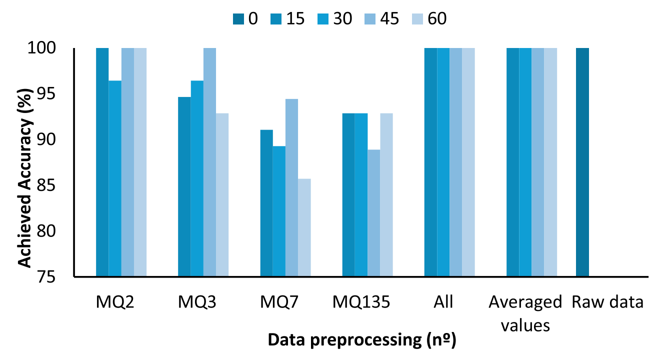


Fig. 8. Accuracy of the validation dataset.

that in the rural area, there were no expected changes in the gases, and this variability is greater than anticipated. Some metrics provide valuable differences between areas, such as average, minimum, and maximum. Nonetheless, these metrics individually have similar values for urban areas and urban green areas. The urban green area is the region with the highest value in all the metrics, while the rural area has lower values for almost all the metrics. In this case, we do not identify those two groups of registers from urban areas characterized by different responses in any of the metrics, as can be seen in MQ2 data.

Finally, the data of MQ135 are characterized by dispersion greater than that for MQ2 but lower than that for MQ7 (see Fig. 7). The dispersion of the two groups of urban areas is visible in these metrics. Nevertheless, the data from urban green areas and one of the groups of data from urban areas have similar values in most of the metrics. The records of rural areas are the ones with lower values in all the metrics.

B. Most Suitable MQ-Based Sensor and Statistical Approach for Identifying Scenarios

In the previous section, we have seen the extracted features and their apparent usability in differentiating the outdoor areas. ML is applied since it has not been feasible to use thresholds in a single feature. The results of validation and verification tests are presented in this section.

In Fig. 8, we can see the accuracy achieved in the validation for the different data used. Since the results for the three tested CART models are extremely similar, we have represented

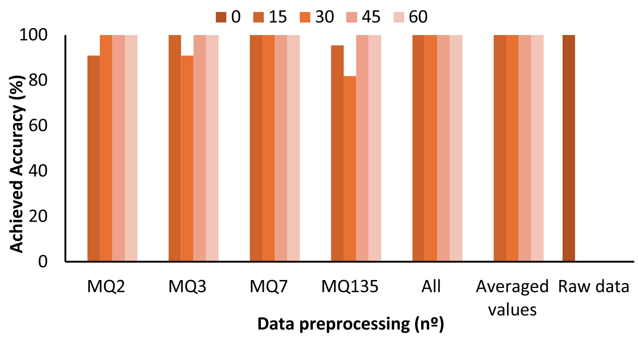


Fig. 9. Accuracy of the test dataset.

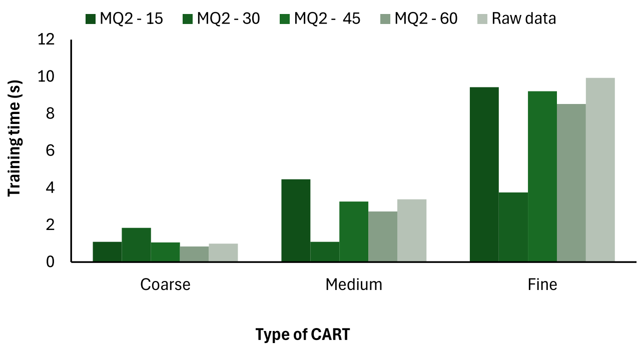


Fig. 10. Training time for generated models.

the averaged data for the three models. The initial four data processing the ones corresponding to the extracted features of individual sensors. Then, the "all" includes all the extracted features for all the sensors, while the "averaged values" include only the feature $\bar{x}$ for all the sensors. Finally, "raw data" consist of raw data with no preprocessing, which includes all records as individual data inputs.

Moreover, in this figure, it is possible to see the different results when different values of N are applied for the buffer of time. The results indicate that the accuracy is 100% for the three last options regardless of the buffer selected. Regarding the data of individual sensors, the outputs pointed out that the highest accuracy is achieved with MQ2, having 100% accuracy with almost all the time buffers. The classification with the data from MQ3 achieves an average of 95% accuracy. The average accuracy for the other two sensors for the different values is close to 90%. The results for the test dataset, as seen in Fig. 9, are significantly better than those for the validation dataset. With both results in mind, the most suitable option for wearable gas sensors is the MQ2.

Once the MQ2 has been chosen, we compare the training time metrics for each time buffer for the three different CART models. It can be seen in Fig. 10 that coarse CART is the one with a lower training time. Using coarse CART supposes a reduction of almost 85% of the required time compared with fine CART. Considering the minimal differences in the achieved accuracies and the great efficiency of coarse CART in terms of training time and computing energy consumed in the smartphone, it has been selected.

C. Comparison With the Existing Systems

In this subsection, the performance of the proposed system is compared with the existing solutions.

Even though the MQ-based sensors were combined with ML to classify samples in multiple cases, their applications to air quality are limited. None of the papers included in the related work included ML. A few examples were found, in which lectures on MQ sensors were used to classify the scenarios.

Applying ML to generated datasets using MQ7, MQ135, and PM sensors in indoor areas during daily activities achieved accuracies of 99.3% with Naïve Bayes and 99.1% with J48 algorithms, respectively [32]. Other authors used the MQ-5, MQ-131, MQ-135, MQ-136, and PMS5003 in outdoor areas to generate a dataset and multiple ML algorithms to classify data. Obtained accuracies ranged from 76% to 95% [33] using decision tree random forest and SVM. Other examples include the recognition of conducted activities in indoor areas using MQ2, MQ9, MQ135, MQ137, and MG-811, achieving accuracies from 83.4% to 86.9% with k-NN [34]. In a similar proposal, authors classify the location (indoor or outdoor) by using four MQ sensors. In this case, the authors achieved accuracies ranging from 98.22% to 99.47%, using DA and PNN [13].

As far as we are concerned, this is the only proposal that uses ML to classify different areas using a single MQ sensor. This supposes energy saving and extending the life of the wearable devices, as well as reducing their size.

D. Energy-Related Issues

The proposed system presents a considerable advantage over existing solutions regarding reduced energy consumption.

On the one hand, the energy linked to powering the sensors is reduced to 25% of the original consumption when four sensors are connected. A single sensor is powered in the proposed system due to the use of extracted features.

On the other hand, there are two changes in how data are forwarded. In the existing solution, the data of four sensors are sent. In the proposed approach, the extracted features from one sensor are sent. We can assume that the energy per packet filled with both data types can be similar since the bytes used to send the content are similar (five bytes for MQ data of four sensors and 20 bytes for five float values for the features). Meanwhile, the energy related to the data forwarding from the node to the smartphone has been reduced. With the previous approach, all gathered data were sent. In the current proposal, only one packet is sent every 45 measures, which corresponds to a decrease of 97% (1/45) in the sent packets. Thus, the energy consumption in the node for data exchange is drastically reduced due to edge computing. Nevertheless, it must be noted that the energy to power the sensors is much higher than that required for the data exchange.

VI. Conclusion

In this article, we propose the use of a single MQ sensor to recognize typologies or outdoor areas with edge and fog computing. The motivation is to reduce the energy consumption of the node compared to using multiple sensors. Moreover, the use of a buffer to process and forward the data reduced the energy consumption for the data sending.

The results indicate that using edge computing to extract features makes it possible to reach 100% accuracy when MQ2 is used. With MQ3 sensors, 100% accuracy was only reached when the time buffer was equal to 45 measured, and the rest of the sensors did not provide data for a good classification with the proposed approach. The coarse CART was the classification algorithm, which attained the best performance in terms of training time for MQ2.

In future work, we will analyze the best time interval between measures in order to reduce the energy linked to data gathering. Moreover, we will evaluate the sampling periodicity by means of adaptive event-triggered algorithms [35] in different outdoor and indoor areas.

References

[1] M. Brauer et al., "Ambient air pollution exposure estimation for the global burden of disease 2013," *Environ. Sci. Technol.*, vol. 50, no. 1, pp. 79–88, 2016.

[2] Y. Wang, Y. Yuan, Q. Wang, C. Liu, Q. Zhi, and J. Cao, "Changes in air quality related to the control of coronavirus in China: Implications for traffic and industrial emissions," *Sci. Total Environ.*, vol. 731, Aug. 2020, Art. no. 139133.

[3] Y. Liu, Y. Zhou, and J. Lu, "Exploring the relationship between air pollution and meteorological conditions in China under environmental governance," *Sci. Rep.*, vol. 10, no. 1, Sep. 2020, Art. no. 14518.

[4] J. Rovira, J. L. Domingo, and M. Schuhmacher, "Air quality, health impacts and burden of disease due to air pollution (PM_{10}, $PM_{2.5}$, NO_2 and O_3): Application of AirQ+ model to the Camp de Tarragona County (Catalonia, Spain)," *Sci. Total Environ.*, vol. 703, Feb. 2020, Art. no. 135538.

[5] L. R. F. Henneman, C. Liu, J. A. Mulholland, and A. G. Russell, "Evaluating the effectiveness of air quality regulations: A review of accountability studies and frameworks," *J. Air Waste Manage. Assoc.*, vol. 67, no. 2, pp. 144–172, Feb. 2017.

[6] Y. Huang et al., "Evaluating in-use vehicle emissions using air quality monitoring stations and on-road remote sensing systems," *Sci. Total Environ.*, vol. 740, Oct. 2020, Art. no. 139868.

[7] P. Kumar, H. Omidvarborna, A. Tiwari, and L. Morawska, "The Nexus between in-car aerosol concentrations, ventilation and the risk of respiratory infection," *Environ. Int.*, vol. 157, Dec. 2021, Art. no. 106814.

[8] S. Viciano-Tudela, S. Sendra, L. Parra, J. M. Jimenez, and J. Lloret, "Proposal of a gas sensor-based device for detecting adulteration in essential oil of cistus ladanifer," *Sustainability*, vol. 15, no. 4, p. 3357, Feb. 2023.

[9] R. Popek, A. Haynes, A. Przybysz, and S. A. Robinson, "How much does weather matter? Effects of rain and wind on PM accumulation by four species of Australian native trees," *Atmosphere*, vol. 10, no. 10, p. 633, Oct. 2019.

[10] Y. Aghoutane et al., "Characterization and analysis of okoume and aiele essential oils from Gabon by GC-MS, electronic nose, and their antibacterial activity assessment," *Sensors*, vol. 20, no. 23, p. 6750, Nov. 2020.

[11] A. Sudarmaji, A. Margiwiyatno, and S. B. Sulistyo, "Characteristics of array MOS gas sensors in detection of adulteration on patchouli oil with candlenut oil," *AIP Conf. Proc.*, vol. 2586, no. 1, Jan. 2023, Art. no. 070016.

[12] J. Wang et al., "Optimization of electronic nose sensor array for tea aroma detecting based on correlation coefficient and cluster analysis," *Chemosensors*, vol. 9, no. 9, p. 266, Sep. 2021.

[13] J. Wang, S. Viciano-Tudela, L. Parra, R. Lacuesta, and J. Lloret, "Evaluation of suitability of low-cost gas sensors for monitoring indoor and outdoor urban areas," *IEEE Sensors J.*, vol. 23, no. 18, pp. 20968–20975, Sep. 2023.

[14] Y. Irawan, R. Wahyuni, H. Fonda, M. L. Hamzah, and R. Muzawi, "Real time system monitoring and analysis-based Internet of Things (IoT) technology in measuring outdoor air quality," *Int. J. Interact. Mobile Technol.*, vol. 15, no. 10, pp. 224–240, 2021.

[15] S. U. Rani, S. Rajarajeswari, J. G. Jaimon, and R. Ravichandran, "Real-time air quality monitoring system using MQ-135 and thingsboard," *J. Crit. Rev.*, vol. 7, no. 18, pp. 4107–4115, 2020.

[16] S. Mluyati and S. Sadi, "Internet of Things (IOT) pada prototipe pendeteksi kebocoran gas berbasis MQ-2 DAN SIM800L," *Jurnal Teknik*, vol. 7, no. 2, 2019.

[17] P. Stančić, A. Stojković, and M. Milić, "Arduino-based gas and smoke detector realized using MQ-2 sensor," in *Proc. 65th Nat. Conf. Electron., Telecommun., Comput., Autom. Control Nucl. Eng. (IcETRAN)*, Sep. 2021, pp. 235–238.

[18] S. Rustemli, C. P. Dautov, and L. Gazigil, "Indoor and outdoor air quality detection using programmable microprocessor and sensor technologies," *Eur. J. Eng. Technol. Res.*, vol. 3, no. 12, pp. 8–13, Dec. 2018.

[19] R. Rumantri, M. Y. N. Khakim, and I. Iskandar, "Design and characterisation of low-cost sensors for air quality monitoring system," *Jurnal Pendidikan IPA Indonesia*, vol. 7, no. 3, pp. 347–354, 2018.

[20] K. B. K. Sai, S. Mukherjee, and H. P. Sultana, "Low cost IoT based air quality monitoring setup using Arduino and MQ series sensors with dataset analysis," *Proc. Comput. Sci.*, vol. 165, pp. 322–327, Jan. 2019.

[21] A. A. Hapsari, A. I. Hajamydeen, D. J. Vresdian, M. Manfaluthy, L. Prameswono, and E. Yusuf, "Real time indoor air quality monitoring system based on IoT using MQTT and wireless sensor network," in *Proc. IEEE 6th Int. Conf. Eng. Technol. Appl. Sci. (ICETAS)*, Dec. 2019, pp. 1–7.

[22] M. F. Ruzaij AI-Okby, S. Neubert, T. Roddelkopf, and K. Thurow, "Integration and testing of novel MOX gas sensors for IoT-based indoor air quality monitoring," in *Proc. IEEE 21st Int. Symp. Comput. Intell. Informat. (CINTI)*, Nov. 2021, pp. 173–180.

[23] W. A. Jabbar, T. Subramaniam, A. E. Ong, M. I. Shu'Ib, W. Wu, and M. A. de Oliveira, "LoRaWAN-based IoT system implementation for long-range outdoor air quality monitoring," *Internet Things*, vol. 19, Aug. 2022, Art. no. 100540.

[24] *ATmega328P Datasheet*. Accessed: Jan. 27, 2023. [Online]. Available: https://content.arduino.cc/assets/Atmel-7810-Automotive-Microcontrollers-ATmega328P_Datasheet.pdf

[25] (2023). *MQ-2 Datasheet*. Accessed: Jan. 27, 2023. [Online]. Available: https://www.mysensors.org/dl/57c3ebeb071cb0e34c90057a/design/MQ2.pdf

[26] (2023). *MQ-3 Datasheet*. Accessed: Jan. 27, 2023. [Online]. Available: https://www.mysensors.org/dl/57c3ebeb071cb0e34c90057a/design/MQ-3.pdf

[27] (2023). *MQ-7 Datasheet*. Accessed: Jan. 27, 2023. [Online]. Available: https://www.mysensors.org/dl/57c3ebeb071cb0e34c90057a/design/MQ-7.pdf

[28] *MQ-135 Datasheet*. Accessed: Jan. 27, 2023. [Online]. Available: https://datasheet4u.com/datasheet-parts/MQ-135-datasheet.php?id=605076

[29] S. De Vito et al., "Crowdsensing IoT architecture for pervasive air quality and exposome monitoring: Design, development, calibration, and long-term validation," *Sensors*, vol. 21, no. 15, p. 5219, Jul. 2021.

[30] *MATLAB Software*. Accessed: Jan. 27, 2023. [Online]. Available: https://au.mathworks.com/products/new_products/r2022b-transition.html

[31] P. Schneider, N. Castell, M. Vogt, F. R. Dauge, W. A. Lahoz, and A. Bartonova, "Mapping urban air quality in near real-time using observations from low-cost sensors and model information," *Environ. Int.*, vol. 106, pp. 234–247, Sep. 2017.

[32] D. U. Urku and H. Agrawal, "Smart real-time indoor air quality sensing system and analytics," *Int. J. Eng. Technol.*, vol. 10, no. 6, pp. 1484–1495, Dec. 2018.

[33] S. K. Shah, Z. Tariq, J. Lee, and Y. Lee, "Real-time machine learning for air quality and environmental noise detection," in *Proc. IEEE Int. Conf. Big Data (Big Data)*, Dec. 2020, pp. 3506–3515.

[34] E. Gambi, G. Temperini, R. Galassi, L. Senigagliesi, and A. De Santis, "ADL recognition through machine learning algorithms on IoT air quality sensor dataset," *IEEE Sensors J.*, vol. 20, no. 22, pp. 13562–13570, Nov. 2020.

[35] P. Navarro-Garcia, L. Parra, S. Sendra, and J. Lloret, "Proposal and evaluation of collaborative event-triggered algorithms in ultra-dense wireless sensor network," in *Proc. Int. Conf. Cooperat. Design, Vis. Eng.* Cham, Switzerland: Springer, Sep. 2023, pp. 1–12.