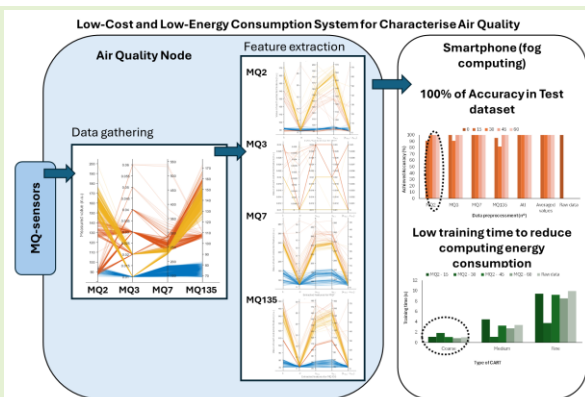


Wearable Low-Cost and Low-Energy Consumption Gas Sensor with Machine Learning to Recognise Outdoor Areas

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

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 paper 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 to 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, thanks 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; MQ sensor; Edge Computing; Fog Computing; Urban area; Rural 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 characterised 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 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 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 reference [7]. Nevertheless, 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 reference [8].

Manuscript received March 5, 2024; accepted month day, year. Date of publication month day, year. This was partially supported by the Spanish Science and Innovation Ministry through the contract

PID2022-136779OB-C31 and by "Conselleria de Educaci3n, Universidades y Empleo" through the "Subvenciones para estancias de personal investigador doctor en centros de investigaci3n radicados fuera de la Comunitat Valenciana (Convocatoria 2023)" Grant number CIBEST/2022/40.

Lorena Parra, and Jaime Lloret was with Instituto de Investigaci3n para la gesti3n Integrada de Zonas Costeras, Universitat Polit3cnica de Val3ncia, C/ Paranimf, 1, 46730 Grao de Gandia, Gandia, Valencia, Spain. (e-mail: svictud@upv.es, loparbo@doctor.upv.es, jlloret@dcom.upv.es).

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

Pascal Lorenz was with 2Network and Telecommunication Research Group, University of Haute Alsace, 34 rue du Grillenbreit, 68008 Colmar, France (e-mail lorenz@ieee.org)

Furthermore, certain compounds have the potential to react with other components in the air, generating hazardous substances [9]. Therefore, the utilisation of gas sensors holds a significant role in implementing monitoring solutions. Among the widely recognised 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 utilised for air monitoring, offering dependable and cost-efficient performance despite challenges related to calibration, as highlighted in reference [13].

Nonetheless, the individual use of these sensors is rarely seen. In most cases, 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 characterise the measured gas. The application of these sensors to discriminate between indoor and outdoor sensed data using four gas sensors was proposed in [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 characterise 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 paper is to evaluate the suitability of using a single sensor to recognise 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. Extracted features are sent via Bluetooth to a smartphone, which analyses 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 paper and their novelties compared with existing systems are the following:

- 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.
- Assess which sensor provides the best-extracted features for outdoor classification with different time buffers.
- Compare the ML models to select the one that balances performance and computing requirements.
- 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 the paper is organised as follows; Section 2 summarises the existing proposals with a limited number of sensors. Section 3 details the proposed approach and used elements. Then, the test bench is described in Section 4. In Section 5, the results that were obtained are discussed. Finally,

Section 6 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.

In 2019, 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, CO₂ 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.

In 2020, 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.

The same year, Mulyati and Sumardi [16] tested a gas sensor system employing the MQ2. An Arduino Nano was used as a microprocessor. Their research aimed to identify gas leaks, utilising 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. Multi-sensor 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 summarised.

In 2018, 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 analyse 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.

In 2019, 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 is forwarded to Thingspeak for its visualisation. 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.

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

In 2021, 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 WiFi. 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.

In 2022, 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 is transmitted using LoRa technology. On this occasion, sensors were calibrated for the following gases: CO₂, NO₂, SO₂ 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 Figure 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 minimise 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 1. 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 recognise 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 Eq. (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 the paper.

$$\bar{x} = \frac{\sum_{i=1}^N x_i}{N} \quad (1)$$

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

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 1, 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.

TABLE I
SUMMARY OF INCLUDED CART

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

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

To better understand the proposal, it has been summarised as a flow chart in Figure 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 4 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 data has been collected are described. Finally, the selected ML-based classifiers and the metrics used for the assessment

are identified.

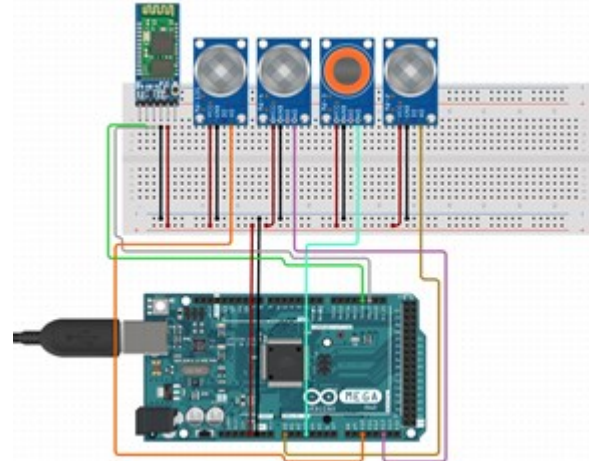


Fig. 1: Proposed wearable gas sensor device.

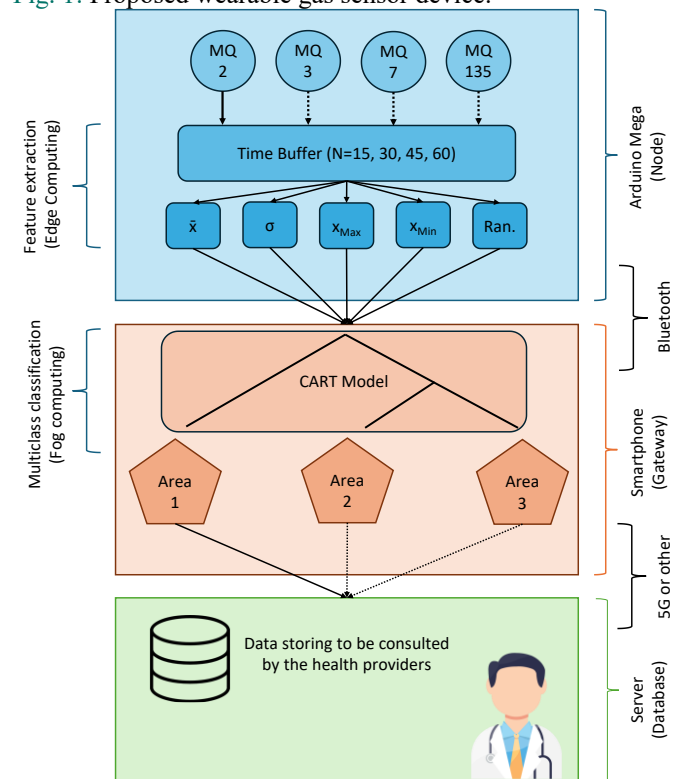


Fig. 2: Proposed wearable gas sensor device.

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

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 hours. 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 hours of data is discarded and the remaining data provides at least 400 records.

B. Measured scenarios

The measured scenarios for this paper 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 meters from a boulevard and 250 meters 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 was gathered from morning to midday in these locations, and collected at Dongying District, 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 District. The measurement point was selected on the mountain slope opposite the large residential areas. Data was 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 analyse the sensed data in different environments. Then, we evaluate which sensor is most useful to identify whether the data is sensed indoors or outdoors.

A. Characterisation 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 gathered data group in each location is plotted, see Figure 3. Raw data has 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 recognise that with raw data, using a single sensor, it is not possible to differentiate between the types of areas. While the rural area is characterised 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 analyse the distribution of extracted features for $N=15$ of all the MQ sensors. Regarding the MQ2, see Figure 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.

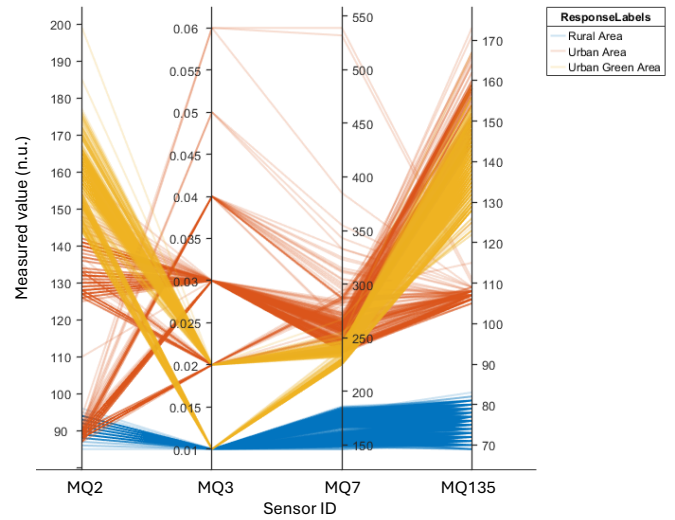


Fig. 3: Gathered data from different MQ sensors.

The urban green areas are characterised by having greater values in averaged, minimum and maximum measured data. It should be highlighted that some of the recorded values of 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 seems to be divided into two groups, one characterised 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 Figure 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 was analysed. 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.

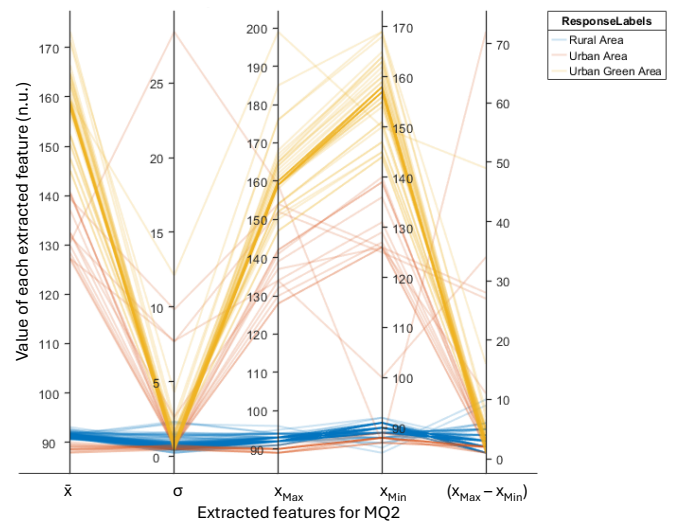


Fig. 4: Extracted features from MQ2.

Concerning MQ7 data, see Figure 6, which shows a greater dispersion than in the previous cases. In the MQ3, the dispersion was minimal. Nevertheless, MQ7 have 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 that in the rural area, there were no expected changes in the gases, 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 characterised by different responses in any of the metrics, as can be seen in MQ2 data.

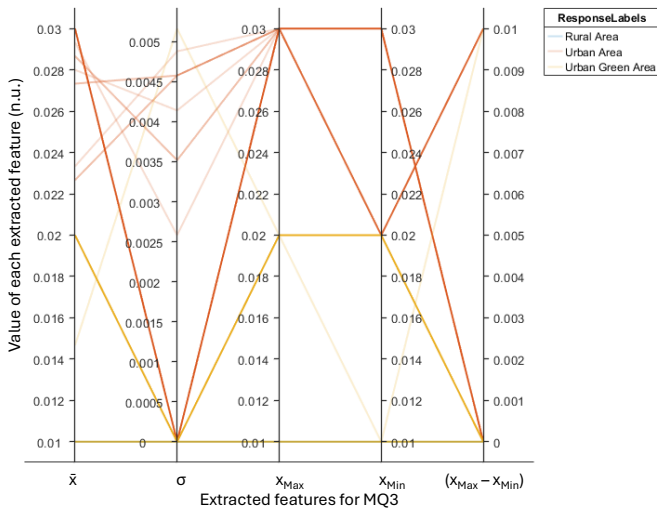


Fig. 5 Extracted features from MQ3.

Finally, the data of MQ135 is characterised by dispersion greater than that for MQ2 but lower than that for MQ7; see Figure 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. 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 Figure 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 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” consists of raw data with no preprocessing, which

includes all records as individual data inputs.

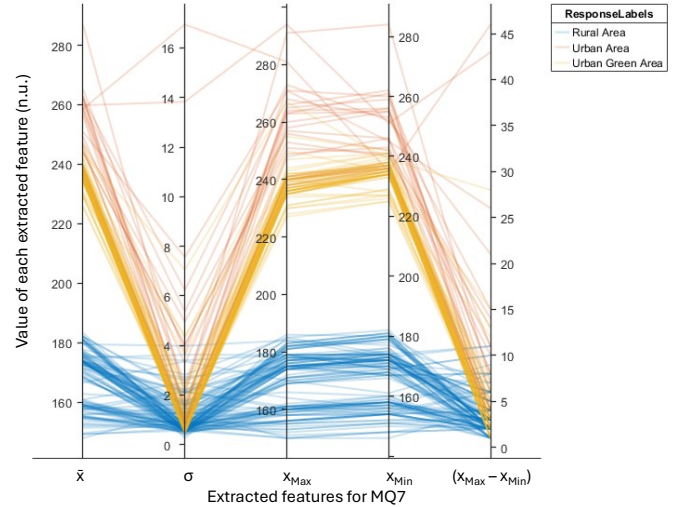


Fig. 6 Extracted features from MQ7.

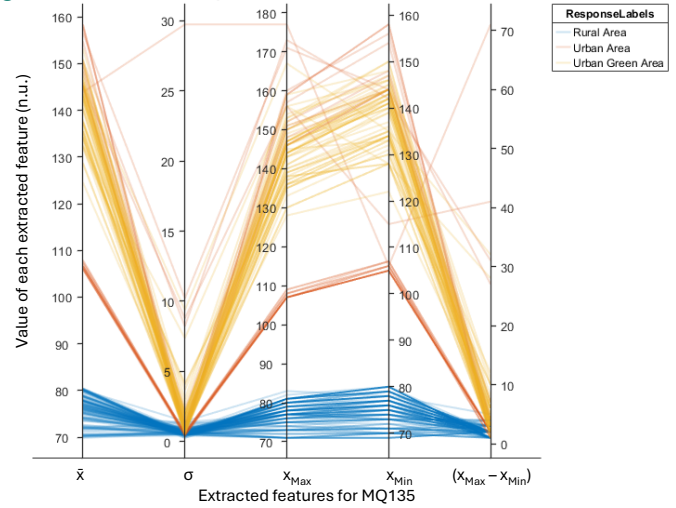


Fig. 7 Extracted features from MQ135.

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 Figure 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 Figure 10 that Coarse CART is the one with a lower training time. Using Coarse CART suppose 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.

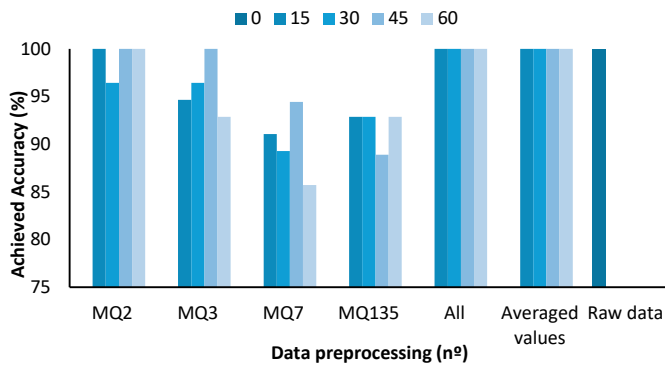


Fig. 8 Accuracy of the validation dataset.

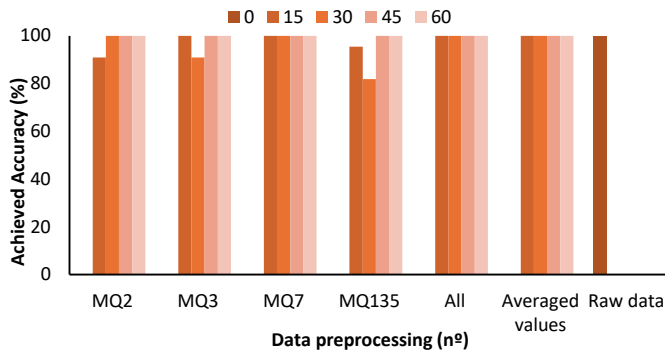


Fig. 9 Accuracy of the test dataset.

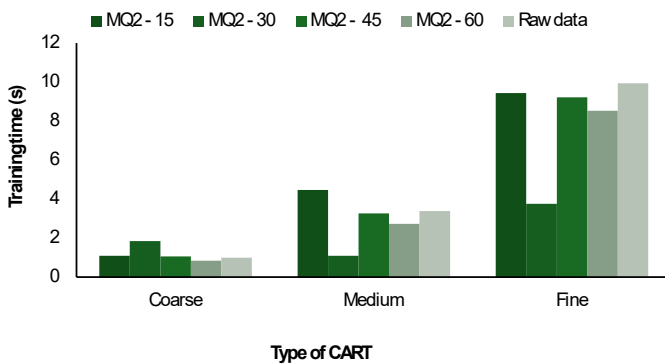


Fig. 10 Training time for generated models.

C. Comparison with existing systems

In this subsection, the performance of the proposed system is compared with 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 particulate matter 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 4 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 thanks to the use of extracted features.

On the other hand, there are two changes in how data is 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 (5 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 was 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 paper, we propose the use of a single MQ sensor to recognise 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 the MQ2.

In future work, we will analyse 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] Brauer, M., Freedman, G., Frostad, J., Van Donkelaar, A., Martin, R. V., Dentener, F., ... & Cohen, A. (2016). Ambient air pollution exposure estimation for the global

- burden of disease 2013. *Environmental science & technology*, 50(1), 79-88.
- [2] Wang, Y., Yuan, Y., Wang, Q., Liu, C., Zhi, Q., & Cao, J. (2020). Changes in air quality related to the control of coronavirus in China: Implications for traffic and industrial emissions. *Science of the Total Environment*, 731, 139133.
 - [3] Liu, Y., Zhou, Y., & Lu, J. (2020). Exploring the relationship between air pollution and meteorological conditions in China under environmental governance. *Scientific reports*, 10(1), 14518.
 - [4] Rovira, J., Domingo, J. L., & Schuhmacher, M. (2020). Air quality, health impacts and burden of disease due to air pollution (PM10, PM2. 5, NO2 and O3): Application of AirQ+ model to the Camp de Tarragona County (Catalonia, Spain). *Science of The Total Environment*, 703, 135538.
 - [5] Henneman, L. R., Liu, C., Mulholland, J. A., & Russell, A. G. (2017). Evaluating the effectiveness of air quality regulations: A review of accountability studies and frameworks. *Journal of the Air & Waste Management Association*, 67(2), 144-172.
 - [6] Huang, Y., Mok, W. C., Yam, Y. S., Zhou, J. L., Surawski, N. C., Organ, B., ... & Ong, H. C. (2020). Evaluating in-use vehicle emissions using air quality monitoring stations and on-road remote sensing systems. *Science of The Total Environment*, 740, 139868.
 - [7] Kumar, P., Omidvarborna, H., Tiwari, A., & Morawska, L. (2021). The nexus between in-car aerosol concentrations, ventilation and the risk of respiratory infection. *Environment International*, 157, 106814.
 - [8] Viciano-Tudela, S., Sendra, S., Parra, L., Jimenez, J. M., & Lloret, J. (2023). Proposal of a Gas Sensor-Based Device for Detecting Adulteration in Essential Oil of *Cistus ladanifer*. *Sustainability*, 15(4), 3357.
 - [9] Popek, R., Haynes, A., Przybysz, A., & Robinson, S. A. (2019). How much does weather matter? Effects of rain and wind on PM accumulation by four species of Australian native trees. *Atmosphere*, 10(10), 633.
 - [10] Aghoutane, Y., Moufid, M., Motia, S., Padzys, G. S., Oumouendze, L. P., Llobet, E., ... & El Bari, N. (2020). Characterisation and analysis of Okoume and aiele essential oils from Gabon by gc-ms, electronic nose, and their antibacterial activity assessment. *Sensors*, 20(23), 6750.
 - [11] Sudarmaji, A., Margiwyatno, A., & Sulisty, S. B. (2023, January). Characteristics of array MOS gas sensors in detection of adulteration on patchouli oil with candlenut oil. In *AIP Conference Proceedings* (Vol. 2586, No. 1). AIP Publishing.
 - [12] Wang, J., Zhang, C., Chang, M., He, W., Lu, X., Fei, S., & Lu, G. (2021). Optimisation of electronic nose sensor array for tea aroma detecting based on correlation coefficient and cluster analysis. *Chemosensors*, 9(9), 266.
 - [13] Wang, J., Viciano-Tudela, S., Parra, L., Lacuesta, R., & Lloret, J. (2023). Evaluation of Suitability of Low-Cost Gas Sensors for Monitoring Indoor and Outdoor Urban Areas. *IEEE Sensors Journal*.
 - [14] Irawan, Y., Wahyuni, R., Fonda, H., Hamzah, M. L., & Muzawi, R. (2021). Real Time System Monitoring and Analysis-Based Internet of Things (IoT) Technology in Measuring Outdoor Air Quality. *International Journal of Interactive Mobile Technologies*, 15(10).
 - [15] Rani, S. U., Rajarajeswari, S., Jaimon, J. G., & Ravichandran, R. O. S. H. A. N. (2020). Real-Time Air Quality Monitoring System Using MQ-135 and Thingsboard. *Journal of critical reviews*, 7(18), 4107-4115.
 - [16] Moreno-Rangel, A., Sharpe, T., Musau, F., & McGill, G. (2018). Indoor fine particle (PM2. 5) pollution and occupant perception of the indoor environment during summer of the first Passivhaus certified dwelling in Latin America. *Journal of Natural Resources and Development*, 8, 78-90.
 - [17] Stančić, P., Stojković, A., & Milić, M. Arduino-Based Gas and Smoke Detector Realized Using MQ-2 Sensor.
 - [18] Rustemli, S., Dautov, C. P., & Gazigil, L. (2018). Indoor and Outdoor Air Quality Detection using Programmable Microprocessor and Sensor Technologies. *European Journal of Engineering and Technology Research*, 3(12), 8-13.
 - [19] Rumantri, R., Khakim, M. Y. N., & Iskandar, I. (2018). Design and characterisation of low-cost sensors for air quality monitoring system. *Jurnal Pendidikan IPA Indonesia*, 7(3), 347-354.
 - [20] Sai, K. B. K., Mukherjee, S., & Sultana, H. P. (2019). Low Cost IoT based air quality monitoring setup using arduino and MQ series sensors with dataset analysis. *Procedia Computer Science*, 165, 322-327.
 - [21] Hapsari, A. A., Hajamydeen, A. I., Vresdian, D. J., Manfaluthy, M., Prameswono, L., & Yusuf, E. (2019, December). Real time indoor air quality monitoring system based on IoT using MQTT and wireless sensor network. In 2019 IEEE 6th International Conference on Engineering Technologies and Applied Sciences (ICETAS) (pp. 1-7). IEEE.
 - [22] AI-Okby, M. F. R., Neubert, S., Roddelkopf, T., & Thurow, K. (2021, November). Integration and Testing of Novel MOX Gas Sensors for IoT-based Indoor Air Quality Monitoring. In 2021 IEEE 21st International Symposium on Computational Intelligence and Informatics (CINTI) (pp. 000173-000180). IEEE.
 - [23] Jabbar, W. A., Subramaniam, T., Ong, A. E., Shu'lb, M. I., Wu, W., & de Oliveira, M. A. (2022). LoRaWAN-based IoT system implementation for long-range outdoor air quality monitoring. *Internet of Things*, 19, 100540.
 - [24] ATmega328P datasheet. Available at: https://content.arduino.cc/assets/Atmel-7810-Automotive-Microcontrollers-ATmega328P_Datasheet.pdf [Last Access : January 27, 2023]
 - [25] MQ-2 datasheet. Available at: <https://www.mysensors.org/dl/57c3ebeb071cb0e34c90057a/design/MQ2.pdf> [Last Access: January 27, 2023]
 - [26] MQ-3 datasheet. Available at: <https://www.mysensors.org/dl/57c3ebeb071cb0e34c90057a/design/MQ-3.pdf> [Last Access: January 27, 2023]
 - [27] MQ-7 datasheet. Available at: <https://www.mysensors.org/dl/57c3ebeb071cb0e34c90057a/design/MQ-7.pdf> [Last Access: January 27, 2023]
 - [28] MQ-135 datasheet. Available at: <https://datasheet4u.com/datasheet-parts/MQ-135-datasheet.php?id=605076> [Last Access: January 27, 2023]
 - [29] De Vito, S., Esposito, E., Massera, E., Formisano, F., Fattoruso, G., Ferlito, S., ... & Di Francia, G. (2021). Crowdsensing IoT architecture for pervasive air quality and exposome monitoring: Design, development, calibration, and long-term validation. *Sensors*, 21(15), 5219.
 - [30] Matlab software. Available at: https://au.mathworks.com/products/new_products/r2022b-transition.html [Last Access: January 27, 2023]
 - [31] Schneider, P., Castell, N., Vogt, M., Dauge, F. R., Lahoz, W. A., & Bartonova, A. (2017). Mapping urban air quality in near real-time using observations from low-cost sensors and model information. *Environment international*, 106, 234-247.
 - [32] Urku, D. U., & Agrawal, H. (2019). Smart real-time indoor air quality sensing system and analytics. *Int J Eng Technol (IJET)*, 10, 1484-1495.
 - [33] Shah, S. K., Tariq, Z., Lee, J., & Lee, Y. (2020, December). Real-time machine learning for air quality and environmental noise detection. In 2020 IEEE International Conference on Big Data (Big Data) (pp. 3506-3515). IEEE.
 - [34] Gambi, E., Temperini, G., Galassi, R., Senigaglia, L., & De Santis, A. (2020). ADL recognition through machine learning algorithms on IoT air quality sensor dataset. *IEEE Sensors Journal*, 20(22), 13562-13570.
 - [35] Viciano-Tudela, S., Navarro-Garcia, P., Parra, L., Sendra, S., & Lloret, J. (2023, September). Proposal and Evaluation of Collaborative Event-Triggered Algorithms in Ultra-Dense Wireless Sensor Network. In International Conference on Cooperative Design, Visualization and Engineering (pp. 1-12). Cham: Springer Nature Switzerland.