Journal of Physical Chemistry and Functional Materials

Home Page of Journal: https://dergipark.org.tr/jphcfum

Decision Trees Based Odor Detection Method with Fusion Data Model Using Semiconductor Gas Sensor Array

Irfan Kilic[a]*, Kubra Yaman[b] , Mutlu Kundakci[b],
[a]*Firat University, Department of Information Technologies, Elazig, Türkiye*
[b]*Ataturk University, Faculty of Science, Department of Physics, Erzurum, Türkiye*
** Corresponding author: E-mail: irfankilic@firat.edu.tr*

ABSTRACT

Odor, one of our senses, is an important sense of life. In some cases, smell can be vitally important. A smell can also help in the correct recognition of a substance. Odor is composed of gas molecules. In this study, 8 simple semiconductor gas sensors were used to detect the odor of various substances. A gas sensor fusion setup was created with gas sensors and a candidate data set was created by collecting sensor data with the help of Arduino Mega embedded system. With the help of this data set, the odor of 7 different cleaning agents and similar substances was detected with the help of Decision Trees (DT). The results obtained from the decision tree (DT) classifier using the data set obtained from the fusion setup (95.44%) are close to the state-of-the-art results. As a result of the study, the feasibility of an embedded odor detection device has been demonstrated.

ARTICLE INFO

Keywords:
Odor detection,
Semiconductor gas sensor,
Decision trees (DT),
Embedded system,
Classification

Received: 2024-11-11
Accepted: 2024-11-20
ISSN: 2651-3080
DOI: 10.54565/jphcfum.1582917

1. Introduction

The effect caused by one or more chemical gas molecules is called odor. The chemical gas molecules interact with the odor receptors in the nose and are processed in the brain and olfaction takes place. Many studies have been conducted on odor in the fields of chemistry, biology, and neuroscience. Olfaction begins when volatile molecules that can be mixed into the air bind to receptors sensitive to different odors in the olfactory epithelium in our nose. In humans, approximately 400 different odor receptor genes detect different odor molecules. These receptors interact with specific molecules, discriminate odor molecules, and transmit the odor signal to the brain. This transmission is described by the lock-and-key model [1], [2]. The sense of smell is directly related to memory and emotional responses [3], [4]. In particular, the amygdala and hippocampus limbic system play a role in establishing a strong link between odor and memories [5]. People have genetic differences in their ability to smell. These genetic differences can cause people to perceive certain odors more or less strongly [6].

Related Works

Some recent literature studies using gas sensors and gas sensor assemblies for odor detection are given in Table 1. Table 1 shows that a gas sensor array is generally used for odor detection. From the data obtained from these sensors, feature extraction was performed with algorithms such as Principle Component Analysis (PCA), Relief, and Linear Regression. Deep Learning based classifiers (MLP, BiLSTM, CNN) were frequently used as classifiers.

Table 1 Literature on odor detection using gas sensors

Reference, Year	Aim of Study	Method	Performance Metrics
Kim E. et al. [7], 2012	Improving the sensitivity of gas sensor assemblies	Neural-Genetic Classification Algorithm (NGCA), ANN (Artificial Neural Network) and GA (Genetic Algorithm)	Accuracy (Acc): 95%
Yang, Z. et al. [8], 2018	Developing a robot with a high-speed gas sensor module based on localized surface plasmon resonance (LSPR)	LSPR with gas-sensitive substrate	Distance: less than 2 cm
Chanonsirivorakul and Nimsuk [9], 2020	Investigating odor perception in humans with the response of an ammonia gas sensor for the standardization of odor measurement	Sensor reading method	%0,0025 ammonia concentration
Yatabe, R. et al. [10], 2021	Developing an odor sensor system with multi-channel data output	Machine Learning	Acc: 97.9%
Zhang, W. et al. [11], 2021	Qualitative and quantitative artificial sense of smell detection of gas	Deep Learning, WCCNN-BiLSTM- many-to-many GRU	R^2: 97,12% (Average for 4 gases)
Al Isyrofie et al. [12], 2022	Determining the freshness of meat	PCA, Deep Learning	Acc: 98.70%
Dobrzyniewski et al. [13], 2022	Odor Air Quality Index (OAQI) Determination of Odor Annoyance	OAQI, PCA (Principle Component Analysis), Multiple Linear Regression (MLR)	R^2: 98% RMSE: 4.22
Aleixandre, M. et al. [14], 2022	Improved online active odor detection procedure	Online learning active sensing loop, Blend odor, Measure odor	p= 0.023 (RRMSE)
Wen, J. et al. [15], 2022	Detecting whether the fruits are spoiled or not by odor	ReliefF Algorithm, Multi-Layer Perception (MLP)	ReliefF + MLP Acc: 100%
Qian, J. et al. [16], 2023	Multisensor Odor Detection System for Odor Classification	Chebyshev polynomial-based weighted discriminant	Acc: 96.52%

Motivation and Contributions

Semiconductor gas sensors developed today are made by mimicking the odor receptors of living organisms. Each sensor is more sensitive to different molecules. Therefore, it is possible to detect different odors more accurately by using sensors sensitive to different odors. Gas sensors can be used alone but can only recognize a small range of odors. By using a gas sensor array, the odor range is expanded and a large number of odors can be recognized. Therefore, gas sensor arrays with different sensitivities should be used. Studies in the literature are mainly focused on detecting fresh air, classifying the odor of meat and other foods, and detecting spoilage [17]. Single-sensor studies have also been carried out for precise gas measurement [18]. In this study, Decision Trees (DT) are used for the classification of cleaning and similar items by using only data normalization without using feature extractors. As a result of the study, the following contributions are expected to be obtained:

(1) Adding odor data set for cleaning agents to the literature
(2) Effective classification without using feature extractors and deep learning
(3) Demonstrating that decision trees are effective in classifying odor data
(4) Demonstrating the feasibility of an embedded odor detection

2. Experimental Setup and Collection of Data Set

There have been studies in the literature for basic odor detection using low-cost sensors. Low-cost gas sensors are generally used for the detection of carbon dioxide, smoke, alcohol, metal gas, LPG, propane, isobutane, hydrogen, carbon monoxide, and similar predatory gases. In this study, a fission dataset was created using low-cost MQ gas sensors. Within the scope of the study, an experimental setup was created using a total of 8 different MQ gas sensors. Semiconductor gas sensors used in the data set collection and their specifications are summarized in Table 2 [19].

In this study, 8 different Semiconductor gas sensors were used for odor detection. The models of these gas sensors, their application areas, and detection rates are presented. These sensors can detect limited gas types on their own. Within the scope of the study, a sensor fusion was created by combining 8 different gas sensors. Thus, it is aimed to detect different odors for each sensor. The electronic setup for sensor fusion is given in Figure 1.

Table 2 Semiconductor gas sensors used for data set collection and their specifications

Sensor No	Sensor Model	Application	Rate*
1	MQ135	- Ammonia (NH_3) - Carbon Dioxide (CO_2) - Benzene (C_6H_6) - Alcohol (ethanol - C_2H_5OH) - Smoke - - Air quality	10 - 1000 ppm
2	MQ3	- Alcohol (ethanol- C_2H_5OH) - Benzene (C_6H_6) - Methanol (CH_3OH) - Propane (C_3H_8) - Carbon Monoxide (CO)	0.05 - 10 mg/L
3	MQ4	- Methane (CH_4) - Natural Gas - Propane (C_3H_8)	200 - 10000 ppm
4	MQ5	- Natural gas (CH_4) - LPG (propane, C_3H_8; butane, C_4H_{10}) - H_2 (Hydrogen)	200 - 10000 ppm
5	MQ6	- LPG (propane, C_3H_8; butane, C_4H_{10}) - Alcohol (ethanol) - Carbon Monoxide (CO)	200 - 10000 ppm
6	MQ7	- Carbon Monoxide (CO) - Hydrogen (H_2)	20 - 2000 ppm
7	MQ8	- Hydrogen (H_2)	100 - 10000 ppm
8	MQ9	- Carbon Monoxide (CO) - Flammable Gases (LPG, methane - CH_4)	10 - 1000 ppm

*ppm: parts per million

As seen in Figure 1, signals are received from 8 different sensors through an analog-digital converter using Arduino Mega. These collected signals are combined to create a data set. The Arduino-based experimental setup is fixed in a closed box. Substances of different odors were placed in this closed box and data was collected from the sensors for about 2 minutes for each substance. In the experimental setup, substances such as detergent, cologne, perfume, wet wipes, vinegar, and adhesive spray were dropped into the box and their odors were collected. In addition to these 6 different odors, the experimental setup was left empty and the odor of normal ambient air was also obtained. In total, odor values were obtained for 7 different situations. The number of samples collected for the collected data set is summarized in Table 3.

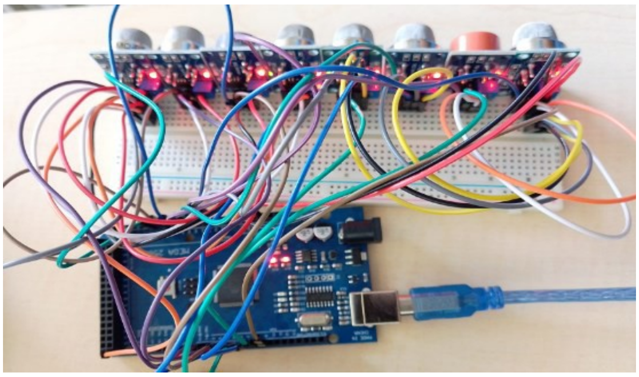

Figure 1 Electronic setup for data set collection

Table 3 Characteristics of the collected data set

No	Class	Number of data
1	Detergent	1264
2	Air	1282
3	Cologne	1272
4	Perfume	1798
5	Wet Wipes	1648
6	Vinegar	1303
7	Adhesive Spray	1227

As seen in Table 3, a minimum of 1227 and a maximum of 1798 samples were collected for 7 different odor types. The number of sample data collected for each odor class is generally close to each other.

3. Proposed Odor Detection Method

In this study, low-cost Semiconductor gas sensors are used to detect different odors that we encounter in daily life. The general block diagram of the sensor fusion-based odor detection method is summarized in Figure 2. As shown in Figure 2, the proposed method consists of 4 basic steps. In the first step, analog data from 8 different sensors were converted into digital data. The data collected for each sensor is normalized within itself. The normalized data were combined to obtain features. A total of 9794x8 feature matrices were created for 8 different sensor types. The features obtained from the sensors of these odor types are discriminative. The odor data obtained from MQ135 and MQ3 sensors are shown in Figure 3 with a scatter plot graph.

When the features extracted from the MQ135 and MQ3 sensors are analyzed in Figure 3, it is seen that each odor has distinctive features. The same is also seen in the relationships between the other sensors. The generated dataset was classified using the decision tree algorithm. Decision Tree algorithm is an algorithm that can work faster than other machine learning algorithms. For this reason, the Decision Tree algorithm is preferred in the proposed method [20], [21]. The hyperparameters of the

Decision Tree classifier model used are presented in Table 4.

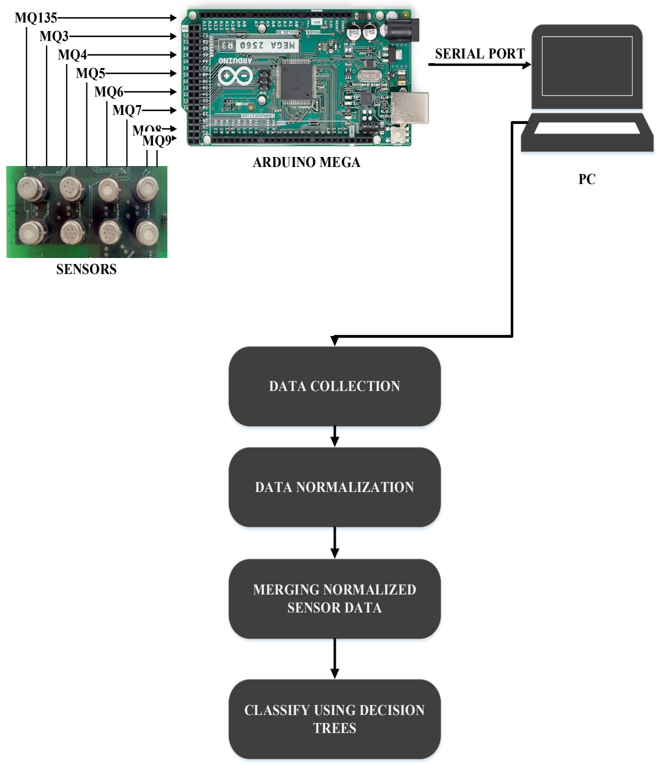

Figure 2 General block diagram of sensor fusion-based odor detection method

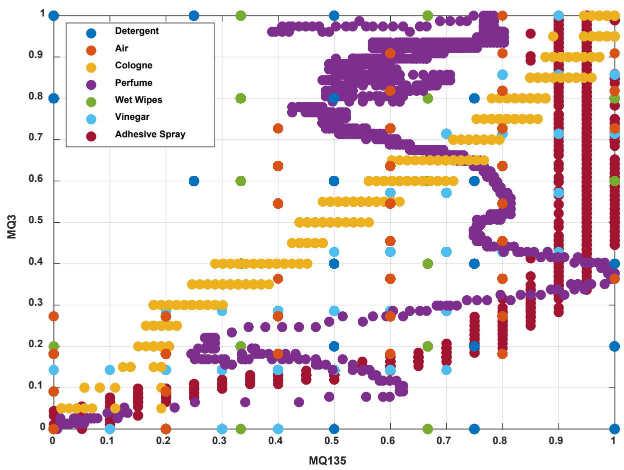

Figure 3 Scatter plot of the obtained features

Table 4 Hyperparameters of the proposed DT classifier model

No	Parameters	Value
1	Preset	Fine DT
2	Maximum number of splits	100
3	Split criterion	Gin's diversity index
4	Surrogate decision splits	Off

4. Experimental Results

The method proposed in this study was developed in the MATLAB platform. The normalized and merged odor data were classified by a 10-fold crossover with the Decision Tree. The confusion matrix obtained with the proposed method is presented in Figure 4.

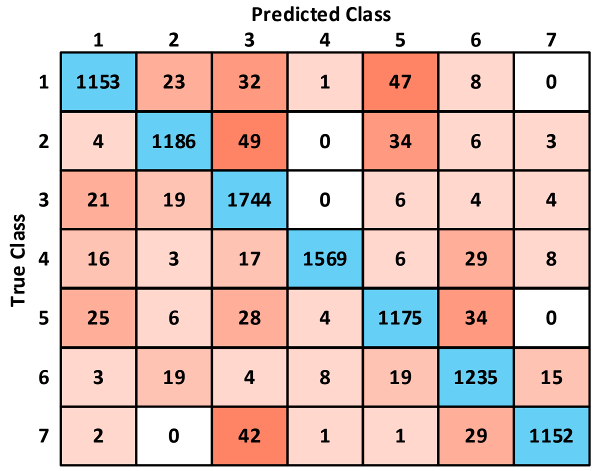

Figure 4 Confusion matrix obtained with the proposed method

When the neighboring matrix given in Figure 4 is examined, it is seen that the accuracy value of all classes is high in general. Using the results in the complexity matrix, class-based accuracy values for 7 different classes are calculated as shown in Figure 5.

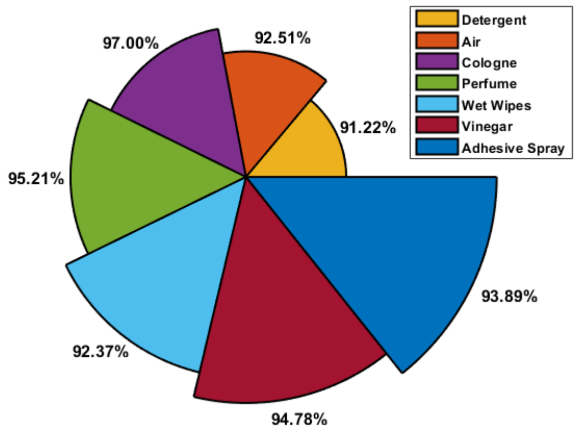

Figure 5 Class-based accuracy results

When the class-based accuracy rates calculated in Figure 5 are examined, the highest result is calculated as 97% with cologne odor and the lowest classification result is detergent odor with 91.22%. It appears that the classification success is high for all odors. The class-based ROC curve and AUC values of the proposed method are presented in Figure 6.

In Figure 6, the AUC value is calculated for each class and the highest result is 0.9956 with the adhesive spray odor and the lowest result is 0.9905 with the detergent odor. It is clearly seen that the overall AUC values for each class are high. Accuracy (%), Precision (%), Recall (%), Geometric Mean (%), and F- Score (%) metrics are used to obtain the statistical results of the proposed method. The proposed Decision Tree classifier was run for 100 iterations. The maximum, minimum, mean, and standard deviation statistics of the obtained results are calculated and listed in Table 5.

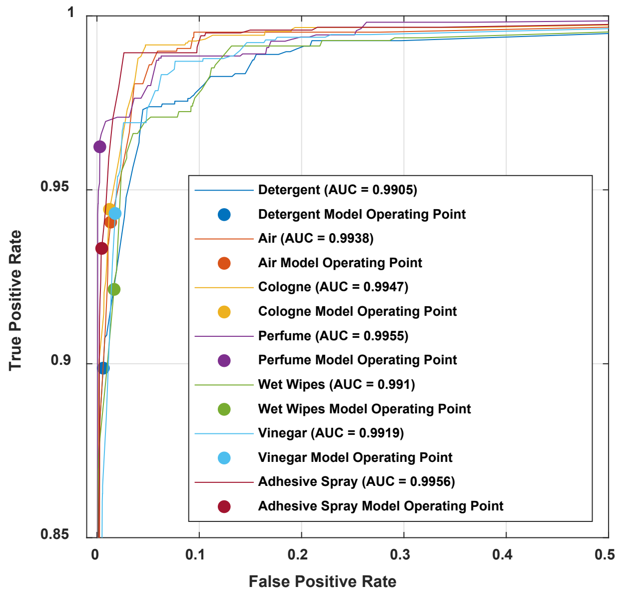

Figure 6 ROC curves and AUC values obtained with the proposed method

Table 5 100 iteration results using the DT classifier

Statistics	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)
Max	**95.44**	95.71	95.33	95.52
Min	92.27	92.65	91.93	92.29
Mean	93.83	93.99	93.60	93.79
Std	0.602	0.570	0.639	0.599

The descriptions of the performance metrics in Table 5 and the calculation equations (1-4) are provided below. The performance rates were calculated using five metrics: accuracy, precision, recall, and F1-score. These metrics were calculated using the number of true positives (TP), false positives (FP), true negatives (TN), and false negatives (FN). The mathematical notations of the used performance metrics are shown in Eqs. 1-5.

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \qquad (1)$$

$$\Pr ecision = \frac{TP}{TP + FP} \qquad (2)$$

$$\mathrm{Re}call = \frac{TP}{TP + FN} \qquad (3)$$

$$F1 - Score = \frac{2TP}{2TP + FP + FN} \qquad (4)$$

As seen in Table 5, the highest accuracy rate for the proposed odor classification model was 95.44% and the lowest accuracy rate was 92.44% after 100 iterations. In order to test the success of the sensor fusion used in the proposed method, odor classification results were calculated using each sensor separately. The calculated accuracy results are presented in Figure 7.

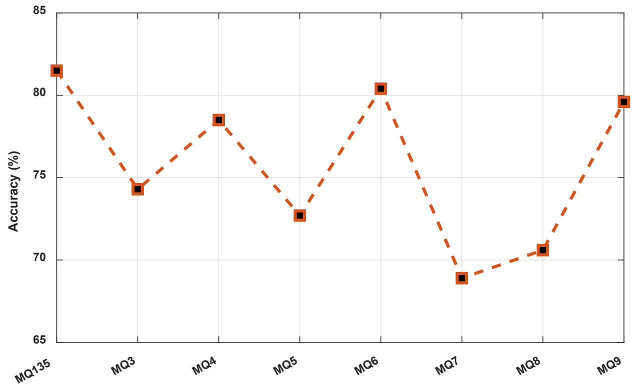

Figure 7 Sensor-based classification accuracies

When the sensor-based classification performance is analyzed in Figure 7, it is seen that the highest accuracy is calculated with the MQ135 sensor at 81.5% and the lowest accuracy is calculated with the MQ7 sensor at 68.9%. It is seen that the performance results will be low if the features obtained from only one or two sensors are used while collecting the data set. In this case, sensor fusion was preferred and the collected dataset was able to distinguish odors with much higher accuracy.

5. Conclusion

In this study, low-cost semiconductor gas sensors used in the literature were used to classify odors that we frequently encounter in daily life. The use of low-cost sensors with high-accuracy results demonstrated the contribution of this study. In addition, a sensor fusion dataset was obtained by combining these sensors, which alone detect odors with low accuracy. The creation of this fusion dataset is another contribution of this study. As a result of the preprocessing with the collected dataset, 95.44% accuracy was calculated for 7 different odors using the Decision Tree classifier. The calculated precision, recall, geometric mean, and F-score metrics confirm the success of the proposed method. In future studies, it is planned to develop an embedded system-based real-time method for the detection of at least 20 different odors.

Competing interests

The authors declare that they have no competing interests.

References

[1] L. Buck and R. Axel, "A novel multigene family may encode odorant receptors: A molecular basis for odor recognition," *Cell*, 1991, doi: 10.1016/0092-8674(91)90418-X

[2] B. Malnic, J. Hirono, T. Sato, and L. B. Buck, "Combinatorial receptor codes for odors," *Cell*, 1999, doi: 10.1016/S0092-8674(00)80581-4

[3] Y. Soudry, C. Lemogne, D. Malinvaud, S. M. Consoli, and P. Bonfils, "Olfactory system and emotion: Common substrates," *European Annals of Otorhinolaryngology, Head and Neck Diseases*. 2011. doi: 10.1016/j.anorl.2010.09.007

[4] R. S. Herz, "Aromatherapy facts and fictions: A scientific analysis of olfactory effects on mood, physiology and behavior," *International Journal of Neuroscience*. 2009. doi: 10.1080/00207450802333953

[5] D. H. Zald and J. V. Pardo, "Emotion, olfaction, and the human amygdala: Amygdala activation during aversive olfactory stimulation," *Proc. Natl. Acad. Sci. U. S. A.*, 1997, doi: 10.1073/pnas.94.8.4119

[6] A. Keller, H. Zhuang, Q. Chi, L. B. Vosshall, and H. Matsunami, "Genetic variation in a human odorant receptor alters odour perception," *Nature*, 2007, doi: 10.1038/nature06162

[7] E. Kim *et al.*, "Pattern recognition for selective odor detection with gas sensor arrays," *Sensors (Switzerland)*, 2012, doi: 10.3390/s121216262

[8] Z. Yang, F. Sassa, and K. Hayashi, "A robot equipped with a high-speed LSPR gas sensor module for collecting spatial odor information from on-ground invisible odor sources," *ACS Sensors*, 2018, doi: 10.1021/acssensors.8b00214

[9] R. Chanonsirivorakul and N. Nimsuk, "Analysis of Relationship between the Response of Ammonia Gas Sensor and Odor Perception in Human," in *2020 8th International Electrical Engineering Congress, iEECON 2020*, 2020. doi: 10.1109/iEECON48109.2020.229522

[10] R. Yatabe *et al.*, "Odor Sensor System Using Chemosensitive Resistor Array and Machine Learning," *IEEE Sens. J.*, 2021, doi: 10.1109/JSEN.2020.3016678

[11] W. Zhang *et al.*, "A Novel Gas Recognition and Concentration Estimation Model for an Artificial Olfactory System with a Gas Sensor Array," *IEEE Sens. J.*, 2021, doi: 10.1109/JSEN.2021.3091582

[12] A. I. F. Al Isyrofie *et al.*, "Odor clustering using a gas sensor array system of chicken meat based on temperature variations and storage time," *Sens. Bio-Sensing Res.*, 2022, doi: 10.1016/j.sbsr.2022.100508

[13] D. Dobrzyniewski, B. Szulczyński, and J. Gębicki, "Determination of Odor Air Quality Index (OAQII) Using Gas Sensor Matrix," *Molecules*, 2022, doi: 10.3390/molecules27134180

[14] M. Aleixandre and T. Nakamoto, "Online Learning for Active Odor Sensing Based on a QCM Gas Sensor Array and an Odor Blender," *IEEE Sens. J.*, 2022, doi: 10.1109/JSEN.2022.3215127

[15] J. Wen, Y. Zhao, Q. Rong, Z. Yang, J. Yin, and Z. Peng, "Rapid odor recognition based on reliefF algorithm using electronic nose and its application in fruit identification and classification," *J. Food Meas. Charact.*, 2022, doi: 10.1007/s11694-022-01351-z

[16] J. Qian, A. Zhang, Y. Lu, J. Zhang, and P. Xu, "A Novel Multisensor Detection System Design for Odor Classification," *IEEE Sens. J.*, 2023, doi: 10.1109/JSEN.2023.3292310

[17] F. N. Abbas, I. M. Saadoon, Z. K. Abdalrdha, and E. N. Abud, "Capable of gas sensor MQ-135 to monitor the air quality with arduino uno," *Int. J. Eng. Res. Technol.*, 2020, doi: 10.37624/IJERT/13.10.2020.2955-2959

[18] M. U. Temel, "İŞ SAĞLIĞI VE GÜVENLİĞİ İÇİN ÇALIŞMA ORTAMLARINDA HAVA KALİTESİNİN ÖNEMİ VE TEKNOLOJİK ÖLÇÜMLERİ," ÇANAKKALE ONSEKİZ MART ÜNİVERSİTESİ, 2020.

[19] "Semiconductor Gas Sensor," 2024. Available: https://www.winsen-sensor.com/semiconductor-gas-sensor/. [Accessed: Oct. 28, 2024]

[20] I. Kilic and O. Yaman, "Classification of Spyware from Network Packets with Decision Trees Using Recursive Feature Elimination (RFE)," in *2024 32nd Signal Processing and Communications Applications Conference (SIU)*, 2024, pp. 1–4. doi: 10.1109/SIU61531.2024.10600885

[21] İ. Kılıç, N. N. Apaydın, M. Apaydın, and O. Yaman, "Decision Tree-Based Direction Detection Using IMU Data in Autonomous Robots TT - Otonom Robotlarda IMU Verilerini Kullanan Karar Ağacı Tabanlı Yön Tespiti," *Batman Üniversitesi Yaşam Bilim. Derg.*, vol. 14, no. 1, pp. 57–68, 2024, doi: 10.55024/buyasambid.1501521. Available: https://doi.org/10.55024/buyasambid.1501521