



Classification of Alcohol Type Using Gas Sensor and K-Nearest Neighbor

Munaf Ismail and Sri Arttini Dwi Prasetyowati

Universitas Islam Sultan Agung, Jl. Raya Kaligawe Km4, Semarang, 50112, Indonesia

ARTICLE INFORMATION

Received: December 15, 2021
Revised: March 05, 2022
Available online: March 29, 2022

KEYWORDS

Classification, Gas Sensor, K-Nearest Neighbor

CORRESPONDENCE

Phone: +62 (024) 6583584
E-mail: munaf@unissula.ac.id

A B S T R A C T

Ethanol, isopropyl and methanol belong to the same alcohol group. The latter is commonly used as an industrial solvent, not for personal consumption. Many traditional alcoholic drink sellers often mix alcoholic beverages, which are commonly called as "oplosan", this mixed drink is very dangerous for human if it contains methanol. Based on this problem, it is necessary to make a measuring device for the alcohol content in the liquid to classify the alcohol type. The design of this gas sensor-based alcohol classification system and method consists of a series of hardware and software applications. The block diagram of the alcohol classification system measures the ethanol and methanol substances in each alcoholic drink using the MQ3 gas sensor and WeMos as a data acquisition device and microcontroller. The computer was used to process the acquisition data from the gas sensor being used then calculates the K-Nearest Neighbor (K-NN) to obtain the prediction results. The K-NN system testing consists of testing the effect of the K value and testing its accuracy. The result of testing the effect of the K value produces 100% optimum accuracy at the values namely K=1, K=3, K=5, K=10 and 55% on K=20.

INTRODUCTION

Alcoholic beverages are a problem in Indonesia. The problem is often due to the emergence of illegal producers who make alcoholic beverages with alcohol content of more than 55%. Alcoholic beverages according to presidential regulation No. 74 of 2013 are defined as beverages that contain ethyl alcohol or flammable ethanol which is processed from agricultural materials containing carbohydrates by fermentation and distillation, or fermentation without distillation [1].

Ethanol, methanol and isopropyl belong to the same alcohol group. The naming of alcohol type compound will be determined based on the length of the carbon chain that composes the alcohol compound. Ethanol, also called ethyl alcohol, has chemical formula namely C_2H_5OH , while methanol or methyl alcohol has chemical formula namely CH_3OH . The latter is commonly used as an industrial solvent, not for personal consumption [2].

Table 1. The Property Comparison of Methanol and Ethanol

Properties	Methanol	Ethanol
Chemical formula	CH_3OH	C_2H_5OH
Molar mass, kg/kmol	32	46
Oxygen content, wt %	50	34.73
Carbon content, wt %	38	52
Hydrogen content, wt %	12	13.1

Methanol and ethanol have the same physical appearance, which is clear and colorless. So, it is very important to make sure the label is listed on the packaging before using it as shown in table 1, knowing the composition of the ingredients according to the label is the easiest way to tell the difference. It becomes a problem if there is no information label for the alcoholic liquid. Many traditional alcoholic drink sellers often mix alcoholic beverages, which are commonly called as "oplosan", this mixed liquor is very dangerous for human if it contains methanol. Until 2018, the number of victims who died due to "oplosan" liquors reached 18,000 people [3].

Based on this problem, it is necessary to make a measuring device for the alcohol content in the liquid according to its classification of alcohol type. It is hoped that the existence of this tool can help the public ensure that the beverages they drink do not contain methanol because they are harmful to human body and the result is close to accurate.

Similar studies have been carried out by several researchers using TGS 2602 sensor as the classification testing of a single gas of acetone or methanol [4] and the classification of tea aroma using Arduino Uno-Based Gas Sensor [5]. Another study examined the measurement of alcohol content using the most appropriate preparation method and the accuracy of gas chromatography to determine the levels of ethanol and methanol in "oplosan" liquors [6].

In this study, the MQ3 alcohol gas sensor was used. Besides trying variations in the use of this sensor type, this sensor was chosen because it was cheaper than the existing alcohol sensors, which has almost the same sensor sensitivity. However, this MQ3 gas sensor consumes quite a lot of power compared to other sensors, which is around 750 mW [7].

To find out the compounds contained in each type of liquid, this test will be calculated using the K-Nearest Neighbor (K-NN). K-NN is a method to classify the objects based on the closest data learning to the object [8], [9], [10]. Data learning is projected into a multidimensional space, where each dimension represents a feature of the data. Using K-Nearest Neighbor (K-NN) method, the accuracy of each contained compound in the type of alcohol can be known.

In this research, the K-NN method will be applied and processed in MS Excel based on the result of gas sensor reading, then the data was sent via the internet network. The gas sensor used is MQ3 to classify the type of alcohol gas, namely the types of ethanol and methanol scattered in the room. In this study, the accuracy of each sensor reading on the distribution of gas in a closed room will be tested.

METHOD

The design of this gas sensor-based alcohol classification system and method in this paper consists of a series of hardware and software applications. The block diagram of the alcohol classification system measures ethanol and methanol substances using the MQ3 gas sensor and WeMos as a data acquisition device, microcontroller, as well as a Wi-Fi network connection. The data is sent via Wi-Fi and received by the computer to be analyzed using the K-NN method as shown in Figure 1.



Figure 1 The Block Diagram of Alcohol Classification

Gas Sensor

The MQ3 sensor is a gas sensor which is suitable to detect alcohol level directly. The MQ3 alcohol sensor has high sensitivity and fast response time. The MQ3 alcohol sensor element consists of an SnO₂ layer with low conductivity in clean air [11]. The sensor resistance will change along with the detection of the ethanol gas presented by the sensor element. If the ethanol concentration is high, the sensor resistance will decrease so that the voltage output will increase.

When metal oxide (SnO₂) crystals are under normal conditions, namely at room temperature, the surface of the metal oxide (SnO₂) material interacts with oxygen molecules in the air. The oxygen atoms will be adsorbed and bind free electrons on the surface of the metal oxide (SnO₂) [12]. Inside the gas sensor, an electric current flows through the grain boundaries of the SnO₂

crystal. In the joint area, oxygen absorption prevents the load from moving freely. If the gas concentration is reduced, the oxidation process will occur. The surface density of the negative oxygen charge will decrease and will result in a decrease in the height of the barrier from the joint area. By lowering the barrier's height, that barrier sensor resistance will also decrease.

As shown in figure 2, the minimum circuit for the MQ3 sensor is very simple. The circuit consists of 1 variable resistor and H pin which is connected to a voltage of 5 V.

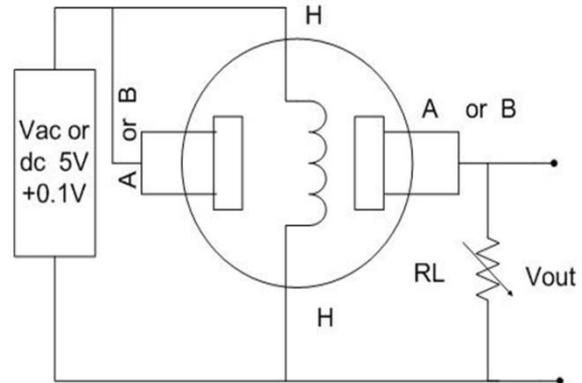


Figure 2 The MQ3 Sensor Driver Circuit [13]

The selected alcohol gas sensor on this system is the MQ3 type. The sensor is manufactured by Hanwei Electronic Co. Ltd. The MQ3 sensor has high sensitivity to alcohol gas and low sensitivity to benzene gas [14]. Table 2 shows the sensitivity level of MQ3 to some detectable gases.

Table 2 The Property Comparison of Methanol and Ethanol

Sensor	Sensitive Substance	Detection Range
MQ2	LPG, Propane	200 ppm-5000 ppm
MQ3	Alcohol	0.05mg/L 10mg/L (Alcohol)
MQ4	Methane	300 ppm-1000 ppm
MQ5	H ₂ , LPG, CH ₄ , CO	200-10000 ppm

The following solvents were used for the gas sensor evaluation, the alcohols tested were ethanol and methanol. To identify the type of gas used in this study, the researcher used a gas sensor, namely the MQ3 type gas sensor as an alcohol gas detector to detect the type of ethanol and methanol.

Data Acquisition

Data acquisition is part of the stages of the Knowledge Discovery in Database (KDD) process to record gas sensor data when testing the samples [15]. With data acquisition, we can classify, predict, estimate and get other useful information from the large data sets.

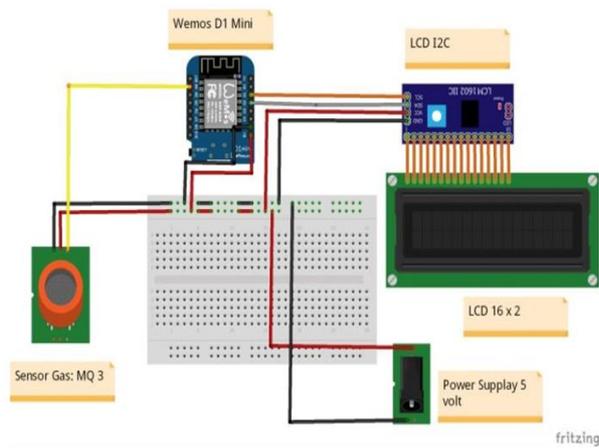


Figure 3 The Design Circuit of Alcohol Classification

The hardware of this research begins the blocking of gas sensor circuit consisting of the commercially available MQ3 gas sensor as shown in figure 3. The gas sensor to measure the volatile alcohol content in a holding container was simulated by placing a liquid alcohol in a closed room to measure the volatile alcohol content with the MQ3 sensor. The output of the MQ3 sensor is the conversion result of the alcohol content detection. The voltage was then forwarded to WeMos A0 analog input pin and the reading of alcohol content was displayed to a 16x2 LCD screen via D1 and D2 pins. Measurements were made by pouring alcohol into the reservoir. The MQ3 sensor was drilled into the alcohol reservoir cap to find out the compounds which were contained in each type of liquid alcohol.

From the sensor, the voltage and the reading of alcohol content were forwarded to WeMos A0 analog input pin, then the formula was processed. The unit was controlled by the ESP8266 IoT embedded system and then the information was sent to the database server with software via Wi-Fi internet network.

K-Nearest Neighbor

The computer was used to process the acquisition data from the gas sensor using the K-Nearest Neighbor (K-NN) method. It is a method that uses a supervised algorithm where the result of the new query instance are classified based on the majority of the categories in the K-NN. The purpose of this algorithm is to classify new objects based on attributes and training samples. The K-NN algorithm method is very simple. The algorithm works based on the shortest distance from the query instance to the training sample to determine its K-NN [16]. The training sample is projected onto a multidimensional space, where each dimension represents a feature of the data. This space is divided into sections based on training sample classification. A point in this space is marked by class *c*, if class *c* is the most suitable classification found in the K-Nearest Neighbors of that point [17]. The closest or furthest neighbor is usually measured based on Euclidean Distance which is presented as follows:

$$d = \sqrt{(a_1 - b_1)^2 + (a_2 - b_2)^2 + \dots + (a_n - b_n)^2} \quad (1)$$

$$d = \sqrt{\sum_{i=1}^n (a_i - b_i)^2} \quad (2)$$

Where:

a = sample data / training data

b = testing data

i = data variable
n = data dimension

The pattern of each gas type will be known through the K-NN classification.

RESULTS AND DISCUSSIONS

The MQ3 gas sensor was used to detect the gas vapor in each type of alcohol directly using methanol and ethanol in the driver circuit with 1 variable resistor. The output of the MQ3 sensor was an analog voltage proportional to the smell of alcohol gas received using the ADC, which was functioned to communicate with microcontroller.

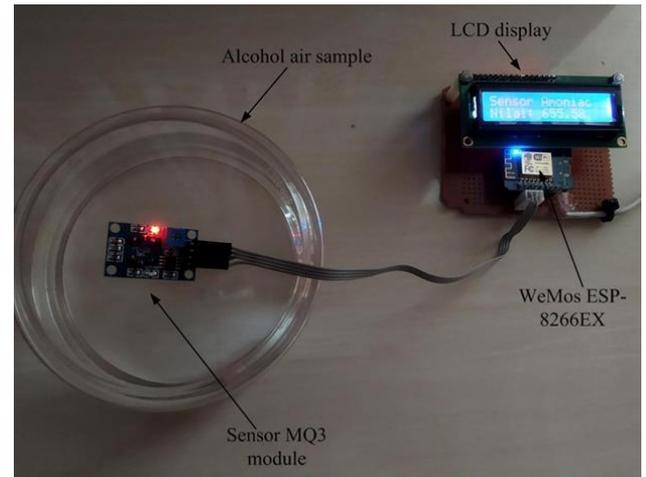


Figure 4 The Prototype of Alcohol Classification

The MQ3 gas sensor output was the conversion of yield voltage alcohol content detection. The voltage was forwarded to the WeMos A0 analog input pin and the alcohol content reading was displayed in a 16x2 LCD screen via D1 and D2 pins. The measurements were made by pouring alcohol into a round container with a diameter of 15 cm and a height of 10 cm as shown in Figure 4. The MQ3 sensor was drilled into the lid of the alcohol reservoir to read the type of gas which appeared from the ethanol or methanol alcohol being tested.

The first stage was to prepare the training data in advance. A training data is the data or information taken in the previous time and the class or label has already been known. In this study, the training data used the gas sensor reading for ethanol and methanol.

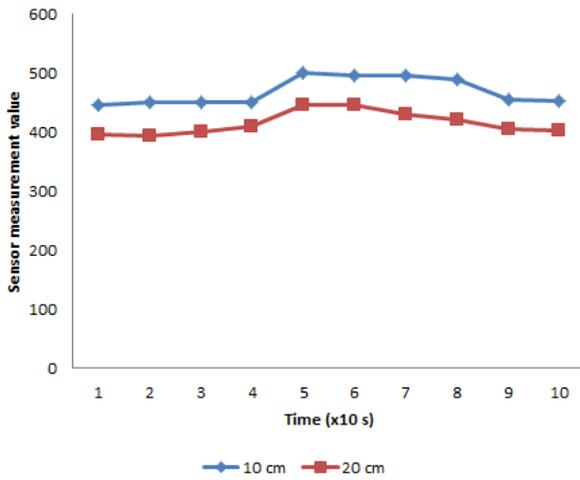


Figure 5 The Response of MQ3 Gas Sensor to Ethanol at Different Distances

Figure 5 shows the measurement of ethanol placed in a holding container using a gas sensor in which the distance varies from 10 cm and 20 cm.

The further the device was from the source of ethanol gas, the lower the concentration of ethanol vapor detected. This can be seen in Figure 5 in which the ethanol gas was read at a distance of 10 cm from the ethanol source, which produced an ADC value of 380-500 and 330-440 at a distance of 20 cm on the MQ3 gas sensor.

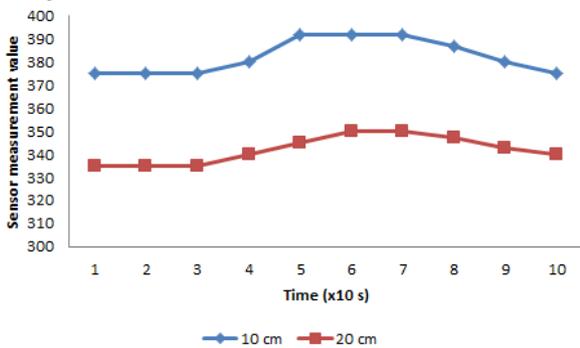


Figure 6 The Response of MQ3 Gas Sensor to Methanol at Different Distances

Figure 6 shows the measurement of methanol placed in a holding container using a gas sensor in which the distance varies from 10 cm and 20 cm. Figure 5 and 6 consists of the x-axis which contains the measurement time, and the y-axis which contains the ADC value of the gas sensor voltage measurement. The second stage was to create a table for the normalization calculation. The normalization formula uses the following equation:

$$Normalization = \frac{data_x - data_{min}}{data_{max} - data_{min}}$$

Then calculate each attribute and make sure the result obtained are the interval between the numbers 0 to 1.

The third step was to calculate the Euclidean distance. Euclidean Distance measures proximity between two objects described as straight line or direct measurement. This measurement method is

suitable when implemented on 14 data that the attributed values are numeric in particular with continuous attributes.

Table 2 The Normalization Data

No	10 cm	20 cm	Class
1	0.56	0.55	1
2	0.60	0.54	1
3	0.60	0.59	1
4	0.60	0.68	1
5	1	1	1
6	0.96	1	1
7	0.96	0.86	1
8	0.92	0.77	1
9	0.64	0.64	1
10	0.62	0.62	1
11	0	0	0
12	0	0	0
13	0	0	0
14	0.04	0.05	0
15	0.14	0.09	0
16	0.14	0.14	0
17	0.14	0.14	0
18	0.10	0.11	0
19	0.04	0.07	0
20	0.00	0.05	0

Table 2 shows the normalization calculation with the description of class 1 = ethanol and class 0 = methanol.

The most common distance calculation used in the calculation of K-NN algorithm is using Euclidean distance calculation. The closer the data similarity, the smaller the distance between the two points. Euclidean distance is said to be good if the new data has a minimum distance and has a high similarity.

Table 3 The Result of Euclidean Distance Calculation

No	Class	ED
1	1	612.10
2	1	612.08
3	1	612.04
4	1	611.98
5	1	611.47
6	1	611.50
7	1	611.59
8	1	611.68
9	1	611.98
10	1	612.01
11	0	612.88
12	0	612.88
13	0	612.88
14	0	612.82
15	0	612.72
16	0	612.69
17	0	612.69
18	0	612.74
19	0	612.80
20	0	612.85

Table 3 shows the Euclidean distance calculation, with the description of class 1 = ethanol and class 0 = methanol.

The value of Euclidean distance which had been sorted was limited and the majority values were taken based on the K value as the prediction result. The data testing used the closest value to the ethanol in the gas sensor reading. Then determine the closest distance to the K sequence and pair the corresponding classes. The K values used are 1, 3, 5 and 10. The sorting result for the K value is shown in Table 3.

Table 3 The Prediction Result

No	K	Predictive value	Name
1	K=1	1	ethanol
2	K=3	1	ethanol
3	K=5	1	ethanol
4	K=10	1	ethanol
5	K=20	1	ethanol

These data indicated that the prediction result obtained using K-NN with Euclidean distance was 100% similar to the test material. It can also be seen that there was no difference in the result obtained by applying variations in the K value from k=1 to k=20.

Furthermore, the measurement of prediction accuracy was obtained by comparing the number of correct predictions with the amount of new data (testing). The accuracy calculation to determine the level of prediction accuracy is obtained by using the equation:

$$\text{accuracy} = \frac{\text{true value}}{\text{number of data}} \times 100\% \quad (3)$$

Description:

True value = The number of data tested with the right result

Number of data = The number of data tested

An example of calculating accuracy with data from table 3 using K=5 (true value = 5, number of data =5) can be seen from the following calculations:

$$\text{accuracy} = \frac{5}{5} \times 100\% = 100\%$$

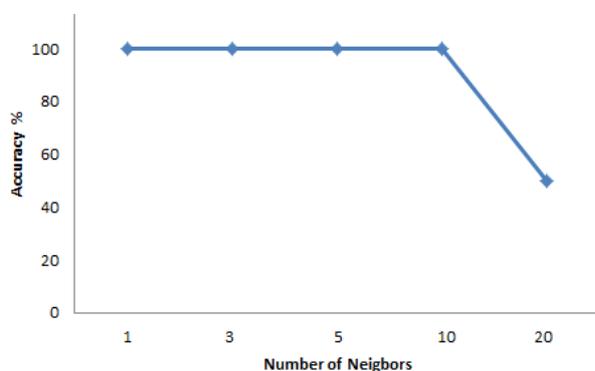


Figure 7. The Testing Accuracy

Figure 7 shows the result of testing the effect of K value that produced an optimum accuracy of 100% at the value of K=1, K=3, K=5, K=10 and the accuracy decreases at 55% when K

value = 20. The test results show that the parameter value of K is very influential on the classification results and the resulting accuracy. The average accuracy tends to decrease with the addition of the value of K [18] [19].

Based on Table 3 and the accuracy calculation of the data being tested, there are five results obtained in the field that are appropriate or correct as predicted using the prediction of K-NN calculation. Thus, based on these equations, the accuracy of the gas sensor reading is 74%.

Based on the comparison data above, the result of the application provided a minimum limit in obtaining the predicted result. In real condition, it was possible to get the more amounts of the cutting result so that the implications can be used for greater profits. This is another advantage obtained from the use of K-NN algorithm with Euclidean distance applied to the alcohol type classification application using the gas sensor that has been studied.

In the results of the study, it was concluded that if the K-NN method can classify the types of methanol and ethanol that are detected very well in a closed room. Other studies have succeeded in classifying alcohol gas using the SVM method [20]. K-NN and SVM methods are classification algorithms from Machine Learning.

CONCLUSIONS

Commercially available gas sensors showed that selectivity increased when the detection was combined from the response of MQ3 gas sensor and the calculation of K-Nearest Neighbor (K-NN) method so as to produce a classification of 2 types of alcohol, namely ethanol and methanol.

The gas sensor detected volatile organic compounds as the closest training data (similar) to the object in the new data or K-NN training and testing data.

The measurement of this system consisted of testing the effect of K value and testing its accuracy. The result of testing the effect of K value produced an optimum accuracy of 100% at the value of K=1, K=3, K=5, K=10 and 55% on K value = 20.

ACKNOWLEDGMENT

This work was funded by Lembaga Penelitian dan Pengabdian Masyarakat (LPPM) Universitas Islam Sultan Agung through Internal Research Fiscal Year 2021/2022.

REFERENCES

- [1] A. Yanti, S. Mursiti, N. Widiarti, B. Nurcahyo, and M. Alauhdin, "Indonesian Journal of Chemical Science Optimalisasi Metode Penentuan Kadar Etanol dan Metanol pada Minuman Keras Oplosan Menggunakan Kromatografi Gas (KG)," *Indones. J. Chem. Sci.*, vol. 8, no. 1, pp. 54–59, 2019, [Online]. Available: <http://journal.unnes.ac.id/sju/index.php/ijcs%0AOptimalisasi>.

- [2] S. Iliev, "A comparison of ethanol, methanol, and butanol blending with gasoline and its effect on engine performance and emissions using engine simulation," *Processes*, vol. 9, no. 8, 2021, doi: 10.3390/pr9081322.
- [3] R. H. Belvage, "Mabuk (di) Indonesia : Membongkar Struktur di Balik Fenomena Korban Tewas Karena Oplosan," *Umbara*, vol. 2, no. 2, pp. 97–112, 2019, doi: 10.24198/umbara.v2i2.20450.
- [4] J. Huang and J. Wu, "Robust and rapid detection of mixed volatile organic compounds in flow through air by a low cost electronic nose," *Chemosensors*, vol. 8, no. 3, 2020, doi: 10.3390/CHEMOSENSORS8030073.
- [5] A. Arisudin, M. Yahya, and D. Erwanto, "Klasifikasi Aroma Teh Dengan Menggunakan Sensor Gas Berbasis Arduino Uno," *JASEE J. Appl. Sci. Electr. Eng.*, vol. 2, no. 02, pp. 115–127, 2021, doi: 10.31328/jasee.v2i02.198.
- [6] A. Yanti, S. Mursiti, N. Wiardi, B. Nurcahyo, and M. Alauhdin, "Optimalisasi Metode Penentuan Kadar Etanol dan Metanol pada Minuman Keras Oplosan Menggunakan Kromatografi Gas (KG)," *Indones. J. Chem. Sci.*, vol. 8, no. 1, pp. 53–59, 2019.
- [7] M. Ismail, A. Marwanto, J. P. Hapsari, and M. Haddin, "Embedded Alcohol Sensing Design And Analysis For Air Samples," in *2021 8th International Conference on Electrical Engineering, Computer Science and Informatics (EECSI)*, 2021, pp. 85–89.
- [8] H. R. Estakhrouei and E. Rashedi, "Detecting moldy bread using an e-nose and the KNN classifier," in *2015 5th International Conference on Computer and Knowledge Engineering (ICCKE)*, 2015, pp. 251–255.
- [9] C. Deng *et al.*, "Enhancing the discrimination ability of a gas sensor array based on a novel feature selection and fusion framework," *Sensors (Switzerland)*, vol. 18, no. 6, 2018, doi: 10.3390/s18061909.
- [10] M. Abbatangelo, E. Núñez-Carmona, V. Sberveglieri, E. Comini, and G. Sberveglieri, "k-NN and k-NN-ANN combined classifier to assess mox gas sensors performances affected by drift caused by early life aging," *Chemosensors*, vol. 8, no. 1, 2020, doi: 10.3390/chemosensors8010006.
- [11] D. E. Williams, "Semiconducting oxides as gas-sensitive resistors," *Sensors Actuators B Chem.*, vol. 57, no. 1, pp. 1–16, 1999, doi: [https://doi.org/10.1016/S0925-4005\(99\)00133-1](https://doi.org/10.1016/S0925-4005(99)00133-1).
- [12] J. Kong *et al.*, "Nanotube Molecular Wires as Chemical Sensors," *Science (80-.)*, vol. 287, no. 5453, pp. 622 LP – 625, Jan. 2000, doi: 10.1126/science.287.5453.622.
- [13] H. Nazemi, A. Joseph, J. Park, and A. Emadi, "Advanced micro-and nano-gas sensor technology: A review," *Sensors (Switzerland)*, vol. 19, no. 6, 2019, doi: 10.3390/s19061285.
- [14] N. M. Tulung, "Rancang Bangun Alat Pendeteksi Kadar Alkohol Melalui Ekhalasi Menggunakan Sensor Tgs2620 Berbasis Mikrokontroler Arduino Uno," *J. Tek. Elektro dan Komput.*, vol. 4, no. 7, pp. 15–24, 2015.
- [15] M. Ismail, A. D. Prasetyowati, and J. P. Hapsari, "Desain dan Implementasi Akuisisi Data Suhu Murid Sekolah Berbasis Arduino Untuk Monitoring Kesehatan Komunal," *J. Nas. Tek. Elektro*, vol. 8, no. 2, pp. 58–64, 2019.
- [16] J. A. Firdaus *et al.*, "Sistem Pengukur Kesegaran Daging Sapi menggunakan Metode K-Nearest Neighbor (K-NN) dengan Fitur Penambahan Data Latih berbasis," vol. 4, no. 5, pp. 1555–1562, 2020.
- [17] K. Amiroh, D. Rahmawati, and A. Y. Wicaksono, "Intelligent System for Fall Prediction Based on Accelerometer and Gyroscope of Fatal Injury in Geriatric," *J. Nas. Tek. ELEKTRO*, vol. 10, no. 3, 2021.
- [18] M. I. P. Putra, D. T. Murdiansyah, and A. Aditsania, "Implementasi Algoritma Modified K-Nearest Neighbor (MKNN) untuk Klasifikasi Penyakit Kanker Payudara Tugas Akhir diajukan untuk memenuhi salah satu syarat memperoleh gelar sarjana dari Program Studi S1 Ilmu Komputasi M Ikhsan Perdana Putra Program Studi," *E-proceeding Eng.*, vol. 6, no. 1, pp. 2431–2441, 2019.
- [19] T. H. Simanjuntak, W. F. Mahmudy, and Sutrisno Sutrisno, "Implementasi Modified K-Nearest Neighbor Dengan Otomatisasi Nilai K Pada Pengklasifikasian Penyakit Tanaman Kedelai," *Pengemb. Teknol. Inf. dan Ilmu Komput.*, vol. 1, No.2, no. 2, pp. 75–79, 2017, [Online]. Available: <http://j-ptiik.ub.ac.id/index.php/j-ptiik/article/view/15/21>.
- [20] Husnawati and R. R. Putra, "Sistem Klasifikasi Jenis Gas Alkohol dan Butana Menggunakan Metode Suport Vector Machine," *Pros. Annu. Res. Semin.*, vol. 4, no. 1, pp. 305–310, 2018.

AUTHORS BIOGRAPHY

Munaf Ismail

Munaf Ismail currently is a lecturer at Electrical Engineering Department of Universitas Islam Sultan Agung. He is received a Bachelor's degree (2003) from Electrical Engineering Universitas Diponegoro and a Master's degree (2015) from Electrical Engineering Universitas Islam Sultan Agung. Research interesting : Sensors, Classification, Optimization.

Sri Arttini Dwi Prasetyowati

Sri Arttini Dwi Prasetyowati received a Bachelor's degree (1989) and a Master's degree (1998) from the Department of Mathematics at Gadjah Mada University, Indonesia. Doctoral Degree (2010) was received from Electrical Engineering Universitas Gadjah Mada Indonesia. Research Interest: Signal Processing and data mining.