



## A Low Cost Electronic Nose System for Classification of Gayo Arabica Coffee Roasting Levels Using Stepwise Linear Discriminant and K-Nearest Neighbor

Indera Sakti Nasution<sup>1\*</sup>, Dian Putri Delima<sup>1</sup>, Zaidiyah Zaidiyah<sup>2</sup>, Rahmat Fadhil<sup>1</sup>

<sup>1</sup>Department of Agricultural Engineering, Faculty of Agriculture, Universitas Syiah Kuala, Banda Aceh 23123, Indonesia

<sup>2</sup>Department of Agricultural Product Technology, Faculty of Agriculture, Universitas Syiah Kuala, Banda Aceh 23123, Indonesia

Corresponding Author Email: [i.nasution@unsyiah.ac.id](mailto:i.nasution@unsyiah.ac.id)

<https://doi.org/10.18280/mmep.090514>

### ABSTRACT

**Received:** 14 June 2022

**Accepted:** 27 September 2022

#### Keywords:

*coffee roasting degree, Gayo arabica coffee, MOS sensors, stepwise linear discriminant, k-nearest neighbor*

A low-cost electronic nose (E-Nose) system using metal oxide sensor (MOS) was developed for Gayo arabica coffee roasting level. The developed electronic nose was designed to have a simple, rapid detection, as well as provides reliable results. The E-Nose system is equipped with MOS sensors, sensor chamber, microcontroller, computer, and data acquisition system. The level of coffee roasting was monitored by read the data from the sensors continuously in real time every second. The sensor signals were recorded in Excel file using data acquisition system and analysed using both stepwise linear discrimination and k-nearest neighbor classifiers. A high percentage (91.67%) of accuracy was obtained using stepwise linear discrimination method. Furthermore, k-nearest neighbor classifier using city block distance demonstrated higher accuracy than stepwise linear discrimination classifier. The results showed that the electronic nose system has a potential for assessing Gayo arabica coffee roasting level. The study confirmed that the proposed electronic nose equipped with at least two MOS sensors was suitable for monitoring the level of coffee roasting level. The result could be used for evaluating other varieties of roasted coffee.

## 1. INTRODUCTION

The chemical composition of green coffee is very complex, depending on the type of species, cultivar, climate, soil type, altitude, and cultivation technique [1, 2]. The aroma of green coffee tends to be weak and not easily detected [3], however roasted coffee tends to enhance the characteristic of the aroma. In general, some roasted coffee aromatics are: aldehyde, ketone, alcohol, ester, pyrazine, furan, acids, nitrogen compounds, and phenol compounds [4]. The measurement of aroma of each roasted coffee can vary according to the type of coffee used. The concentration of coffee aroma composition is not only caused by temperature and roasting time (roasting degree) but also influenced by the composition of green coffee aroma, where the composition of green coffee is influenced by the species, origin of coffee, maturity level, pre-harvest and post-harvest of green coffee [4].

Although the methods for detecting aroma are complex, aroma can be determined by conventional techniques such as: gas chromatography (GC) and gas chromatography-mass spectrometers (GC-MS) [4, 5]. GC or GC-MS is currently very popular to identify volatile compounds in coffee, but this method is very expensive, complex, and tedious [1]. Electronic nose (E-Nose) has the potential to substitute the conventional method, because it is easier to build and cheaper than GC or GC-MS.

Electronic nose is an artificial sense of smell to detect the aroma of certain products that includes sensor array, signal

conditioning and initial data processing, and pattern recognition. E-Nose technology has been widely used for fresh food analysis recently [6]. However, due to the complexity of the aroma in coffee, the type of E-Nose developed was varying. Several studies have been used E-Nose such as, E-Nose was used to detect changes in the aroma profile of Espresso coffee after grinding using the sensor array  $\alpha$ FOX system [7]. Identification of Arabica and Robusta coffee powder using 5 MOS sensors using the back propagation learning method [8].

E-Nose able to classify and determine aroma through statistical analysis based on sensor output [9], such as principal component analysis (PCA), linear discriminant analysis [10], support vector machine (SVM), and partial least squares (PLS) [11]. A considerable amount of literature has been published on stepwise linear discriminant analysis [12, 13], these studies provide important insight into E-Nose classification. Nonconformity measure based on k-nearest neighbors (K-NN) implemented separately as underlying algorithm of conformal prediction for discrimination of Ginsengs by a home-made electronic nose [14]. This study will focus on the use of gas array sensor on the aroma of Gayo specialty arabica coffee. Moreover, research on electronic nose for Gayo specialty arabica roasted coffee based on roasting level rarely conducted. The aim of this study was to classify the roasting level of coffee beans using electronic nose detection system such as stepwise linear discriminant analysis (SLDA) and k-nearest neighbor (K-NN) method.

## 2. MATERIAL AND METHODS

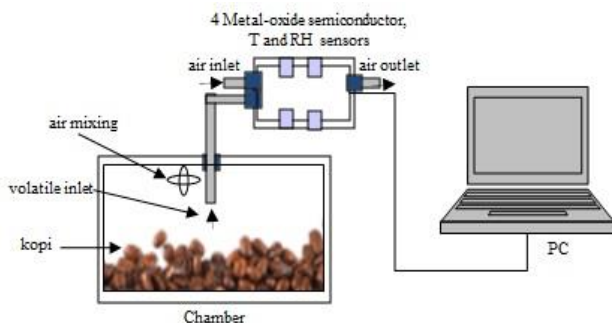
### 2.1. Sample preparation

Green coffees samples, (Gayo 1 fullwash specialty arabica) belong to Pondok Gajah Village, Bandar District, Bener Meriah Regency-Indonesia were employed. The samples were subjected to a roasting process using Didacto Italia TA413D roaster machine for 3 different times (7, 9, and 13 min) in order to obtain different roasting levels as confirming using visual evaluation (Table 1). Visual evaluation was measured according to color using SCAA (Specialty Coffee Association of America) roast coffee standard kit range from lightest (95) to darkest (25). The T1, T2, and T3 represented SCAA (Agrtron) number of 85, 48, and 35, respectively. Table 1 reported the temperatures reached of coffee samples roasted at different times. After the roasting process, the coffee beans placed in a room temperature for 3 days for degassing process. This degassing process, gases (mostly carbon dioxide) form inside the bean begin seeping out, otherwise gasses escapes so quickly that negatively affects the flavor or aroma of coffee [15].

**Table 1.** Roasting of arabica coffee for each roasting level

Treatment (Classification)	Roasting time (min)	Coffee beans temperature (°C)
T1	7	175
T2	9	180
T3	13	190

In order to obtain a value in the baseline position, the E-nose system without coffee (blank sample) tested using the proposed E-Nose. The temperature and humidity in the sensor chamber maintained uniformly. Each sample (150 g) was introduced into the closed sensor chamber (Figure 1), the chamber was sealed with a lid and silicon disk in a cap. All roasting level set were replicated by 6 times each for training (3 treatment x 4 MOS sensors x 6 repetitions) and 4 times each for testing (3 treatment x 4 MOS sensors x 4 replications), therefore producing a total of 120 roasted coffee samples. The aroma's sample was pumped into the sensor chamber, and sensor's response during 1,500 seconds was collected. The generated aroma in the sensors chamber recorded until reached a steady state signals. After the evaluation, the sensor chamber was exposed to pure air to purge the observation chamber.



**Figure 1.** E-Nose system design

### 2.2 Electronic nose (E-Nose) system

The E-Nose system constructed using four MOS sensors (MQ135, MQ9, MQ3 and MQ2). DHT22 sensor used to

measure temperature and humidity. The length, width and height of the sensor chamber were 13 cm x 13 cm x 7 cm, respectively. All gas sensors and temperature and humidity sensors placed in the room equipped with with air inlet and outlet. The E-Nose was equipped with MOS sensors, sensor chamber, microcontroller, computer, air mixing, and data acquisition system [16]. The design of the E-Nose system shown in the Figure 1.

The Arduino Uno microcontroller board pin input connected to each MOS sensors, and converted the analog signal into digital. The pin device connected to the sensor shown in the Table 2. The MOS sensors have an electrochemical sensor that changes its resistance for different levels of coffee beans. The sensor attached with a variable resistor to form a voltage divider circuit, and the variable resistor will change its sensitivity once coffee beans contact with the sensor after heating. The change in the resistance changes the voltage across the sensor that read by the microcontroller. In order to provide an accurate data response, a calibration step performed using fresh air. The E-Nose system read the data from the sensors continuously in real time every one second. Sensor signals that recorded in Excel file analyzed using the proposed analysis.

**Table 2.** Gas sensors connected to Arduino Uno pins

Sensors	Target pin
MQ-3	A0 (Analog)
MQ-9	A1 (Analog)
MQ-135	A2 (Analog)
MQ-2	A3 (Analog)
DHT22	2 (Digital)

### 2.3 Stepwise linear discriminant analysis (SLDA)

The SLDA evaluates a group of basis vectors that has the smallest within class scatter and the largest between class scatters. This method used as a tool for classification, dimensionality reduction, and maximum discrimination of the classes [12]. The size reduction utilized to select the most dominant input variables (MOS sensors) for reducing the computational load and improving the classification performance. Moreover, as a supervised methods the dimensionality reduction takes the class labels into consideration [17]. The stepwise method begins with the selection of variables carried out by Wilk's Lambda method [18]. A smaller value of Wilks' lambda indicates a greater discriminant power of the function. In order to perform LDA method, three steps are required [19]. Cutting score utilized to classify between two groups uniquely, this score used for constructing the classification matrix shown on Eq. (1) below:

$$C_s = \frac{(N_A * C_A) + (N_B * C_B)}{N_A + N_B} \quad (1)$$

where,  $C_s$  is an optimal cutting score between group A and B.  $N_A$  and  $N_B$  are number of observations of group A and B, respectively.  $C_A$  and  $C_B$  are centroid for group A and B, respectively.

### 2.4 K-nearest neighbor (K-NN)

The advantages of using K-NN are better performance in terms of prediction accuracy and computation time [20]. The K-NN is a supervised machine learning technique used for

classification, regression, pattern recognition and predictive analysis. The K-NN algorithm gathers data points that are close to it, and then sort those closest data points with regard to distance from the arrival data point. The distance can be measure in various ways. Various metric distances equations such as of Euclidean, city block, Hamming, and Mahalanobis distance have studied [21-23]. The equation of these metric distances as follows:

$$D_{euc} = \left[ \sum_{i=1}^k (x_i - y_i)^2 \right]^{\frac{1}{2}} \quad (2)$$

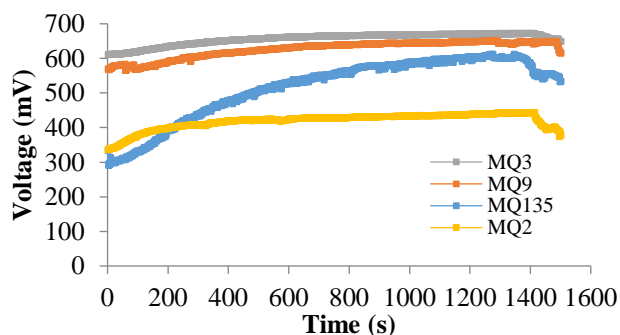
$$D_{cb} = \sum_{i=1}^k (x_i - y_i) \quad (3)$$

$$D_{ham} = \sum_{i=1}^k |x_i - y_i| \quad (4)$$

$$D_{mah} = \sqrt{(x - y)S^{-1}(x - y)^T} \quad (5)$$

### 3. RESULTS AND DISCUSSION

The signal output from the treatment 1 (see Table 1) was taken at one second intervals for 1,500 seconds are shown in Figure 2. Response curves with different roasting levels have different variations, the responses of sensors firstly increase and at the end will reach a dynamic balance.



**Figure 2.** Typical response of E-Nose sensors roasted coffee aroma (case: Treatment classification 1, roasting time 7 minutes at 175°C)

The typical response also occurs in treatments 2 and 3, respectively. In addition, all MOS sensors were sensitive to all the aromatic profile developed at different roasting level and for this reason they were considered for the statistical analysis. All signals at different level of roasting recorded were provided in the Table 3.

Stepwise linear discriminant analysis (SLDA) was applied to determine the best predictors. Linear discriminant analysis (LDA) able to reduce the number of features to a more manageable number before the classification process. Despite its simplicity, LDA often provides robust, sane, and

interpretable classification results. The best model obtained by the SLDA selected the 2 predictors that most contributed to the differentiation of Gayo Arabica coffee from different level of roasting. The variables selected were MQ2 and MQ135. These results are similar to that found in Radi et al. [24] who stated the highest responsive E-Nose sensor for coffee under roasting condition using MQ series sensors based linear regression parameter was MQ135. The MQ135 have been used for classifying Arabica and Robusta coffee using support vector machine and perceptron [25]. The MQ135 sensor has sensitivity on several gases, such as: CO<sub>2</sub>, CO, alcohol, NH<sub>4</sub>, toluene, and acetone. Moreover, roasted coffee generates many chemical compounds containing alcohols, acids, aldehydes, azines, hydrazides, and ketones [1].

**Table 3.** The mean of E-Nose response aroma at different roasting levels

MOS sensors	E-Nose signal (mV)	SD	Treatment classification
MQ135	570.50	22.845	T1
	647.33	10.930	T2
	660.17	17.394	T3
MQ9	641.83	6.997	T1
	648.67	11.944	T2
	667.67	9.1797	T3
MQ3	652.33	9.395	T1
	657.00	13.638	T2
	685.00	4.690	T3
MQ2	402.50	21.751	T1
	425.50	23.637	T2
	645.33	25.866	T3

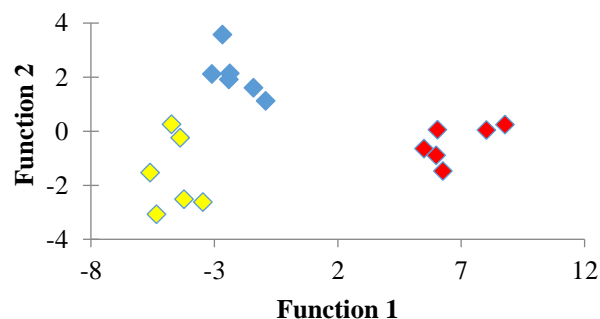
SD: Standard Deviation; N=6 replications

The selected sensors were used for the prediction studies by means of LDA methods. Discriminant scores (DS) were calculated according to the discriminant functions (Eqns. (6) and (7)).

$$DS_1 = -32.079 + 0.020339 X + 0.039394 Y \quad (6)$$

$$DS_2 = -25.763 + 0.052585 X - 0.014570 Y \quad (7)$$

where,  $X$  is E-Nose signal of MQ135, and  $Y$  is E-Nose signal of MQ2,  $DS_1$  is discriminant function 1, and  $DS_2$  is discriminant function 2.



**Figure 3.** Discriminant function plot of the first two functions obtained from data training of LDA analysis of roasted coffee aroma using E-Nose. Numbering in the graph represent treatment (group centroid) T1=175°C for 7 minutes, T2=180°C for 9 minutes, and T3=190°C for 13 minutes, respectively

Figure 3 shows the discriminant function plot of the first two functions obtained from LDA method, which able to classify 100% of accuracy for data training with a similar value after cross validation. Therefore, the results showed that this statistical approach was useful for evaluating the differences between coffee samples. Table 4 shows the successful discrimination of the data testing by using the Eq. (6). A high percentage (91.67%) of successful classification was obtained. The classification results significantly decrease (67.67%) when the Eq. (7) was used (Table 5).

The accuracies of four MOS sensors use *K-NN* analysis with varied factor of distance metrics and number of neighbors from one to eight shown in Table 6. The result revealed that at  $k=1$  until  $k=5$  for Euclidean distance metric, the accuracy of training and testing is 100% and 83.3%, respectively. City block distance at  $k=3$  until  $k=8$ , achieved the best result with the accuracy is 100% for both training and testing dataset. For Hamming distance metric, the result of accuracy ranged within 33.3% to 72.2% for training and 23.3% to 41.7% for testing.

**Table 4.** Successful discrimination of E-Nose testing set of two MOS sensors selected using discriminant function 1

MOS Sensors (mV)		Discriminant Scores	Class	Successful (yes/no)
MQ135	MQ2			
565	377	-5.736	T1	yes
540	390	-5.732	T1	yes
559	435	-3.573	T1	yes
560	420	-4.144	T1	yes
645	418	-2.494	T2	yes
640	442	-1.649	T2	yes
640	397	-3.423	T1	no
640	418	-2.595	T2	yes
673	666	7.845	T3	yes
600	617	4.430	T3	yes
680	680	8.539	T3	yes
644	621	5.483	T3	yes

Cutting score between group T<sub>1</sub> and T<sub>2</sub> is -3.385; cutting score between group T<sub>2</sub> and T<sub>3</sub> is 2.309.

**Table 5.** Successful discrimination of E-Nose testing set of two MOS sensors selected using discriminant function 2

MOS Sensors (mV)		Discriminant Scores	Class	Successful (yes/no)
MQ135	MQ2			
565	377	-5.855	T1	yes
540	390	-5.171	T1	yes
559	435	-2.836	T1	no
560	420	-3.617	T1	yes
645	418	-3.746	T1	no
640	442	-2.496	T2	yes
640	397	-4.837	T1	no
640	418	-3.744	T1	no
673	666	9.144	T3	yes
600	617	6.618	T3	yes
680	680	9.871	T3	yes
644	621	6.813	T3	yes

Moreover, Mahalanobis distance metric yielded accuracy ranged within 72.2% to 88.9% for training and 58.3% to 83.3% for testing. The performance of *K-NN* classifier is affected by the choice of  $k$  value. Choosing the optimal  $k$  becomes challenge for different experiments. The accuracy of data sets could be measured using  $k$  value from 1 to

$\sqrt{\text{number of sample in a data set}}$ , sometimes the majority of the best  $k$  value is 1 or in other experiment a number of data sets are not sensitive to the  $k$  value [26]. Small  $k$  value is sensitive to noisy data, whereas high  $k$  value (exceed its convenient level) results unreliable prediction [27].

**Table 6.** Accuracies of four MOS sensors for E-Nose system using K-Nearest Neighbor

k	Distance metrics							
	Euclidean		City block		Hamming		Mahalanobis	
	T	t	T	t	T	t	T	t
1	100	83.3	100	92	72.2	41.7	83.3	58.3
2	100	83.3	100	92	50	41.7	83.3	83.3
3	100	83.3	100	100	50	33.3	88.9	75
4	100	83.3	100	100	50	25	72.2	58.3
5	100	83.3	100	100	33.3	23	83.3	83.3
6	94.4	83.3	100	100	33.3	33.3	83.3	83.3
7	94.4	91.7	100	100	33.3	33.3	88.9	83.3
8	88.9	83.3	100	100	33.3	33.3	83.3	83.3

$k$ =number of neighbors; T=training accuracy; t=testing accuracy

According to Table 6, it shows that the city block distance metric (also called Manhattan distance) performed the best distance metrics compared to all distance metrics. City block distance is usually preferred over the more common Euclidean distance when there is high dimensionality in the data and the number of features is large [28, 29]. Moreover, city block distances metrics it works well with datasets with compact or isolated clusters, however, it sensitive to outliers [30]. In spite of Euclidean distance is very common in *K-NN* clustering, however, in the event that two data vectors have no property values in common, they may have a smaller distance than the other pair of data vectors containing the same trait values [31]. Hamming distance used to measure the distance between two vectors, and its low accuracy affected by the ratio of members in each class, while the other distance metrics not affected by such phenomenon [32]. Mahalanobis distance metrics proposed to represent the distance between particular point and the mean of the normal data [33]. Mahalanobis has higher result for low-dimensional datasets [34], however lack of robustness due to outliers [35].

#### 4. CONCLUSIONS

This study proposed customized electronic nose based on 4 metal oxide semiconductor sensors to classify Gayo arabica coffee roasting level. The potential of the electronic nose to monitor changes in aroma of different roasting level was studied. The application demonstrated its reliability in documenting significant changes among roasting levels as it was also noticed during data analysis and interpretation. It did not require any sophisticated or expensive laboratory equipment. The suggested e-nose provides a low-cost, non-destructive tool that non-specialists can use. In the process of classification by SLDA and *K-NN*, the *KNN* demonstrated higher accuracy than the SLDA. The study confirmed that the proposed electronic nose equipped with at least MQ2 and MQ135 sensors was suitable for monitoring the level of coffee roasting level. The obtained results could represent the capability of the sensor array for evaluating Gayo arabica coffee roasting level. These results could be used in further studies by using other agricultural commodities.

## REFERENCES

- [1] Marek, G., Dobrzański Jr, B., Oniszczuk, T., Combrzyński, M., Ćwikła, D., Rusinek, R. (2020). Detection and differentiation of volatile compound profiles in roasted coffee arabica beans from different countries using an electronic nose and GC-MS. *Sensors*, 20(7): 2124. <https://doi.org/10.3390/s20072124>
- [2] Seninde, D.R., Chambers, E. (2020). Coffee flavor: A review. *Beverages*, 1-25. <https://doi.org/10.3390/beverages6030044>
- [3] Liu, C., Yang, N., Yang, Q., Ayed, C., Linforth, R., Fisk, I.D. (2019). Enhancing robusta coffee aroma by modifying flavour precursors in the green coffee bean. *Food Chemistry*, 281: 8-17. <https://doi.org/10.1016/j.foodchem.2018.12.080>
- [4] Caporaso, N., Whitworth, M.B., Cui, C., Fisk, I.D. (2018). Variability of single bean coffee volatile compounds of Arabica and robusta roasted coffees analysed by SPME-GC-MS. *Food Research International*, 108: 628-640. <https://doi.org/10.1016/J.FOODRES.2018.03.077>
- [5] Yuwono, S.S., Hanasasmita, N., Sunarharum, W.B., Harijono. (2019). Effect of different aroma extraction methods combined with GC-MS on the aroma profiles of coffee. *IOP Conference Series: Earth and Environmental Science*, Institute of Physics Publishing, p. 12044.
- [6] Shi, H., Zhang, M., Adhikari, B. (2018). Advances of electronic nose and its application in fresh foods: A review. *Critical Reviews in Food Science and Nutrition*, 58(16): 2700-2710. <https://doi.org/10.1080/10408398.2017.1327419>
- [7] Severini, C., Ricci, I., Marone, M., Derossi, A., De Pilli, T. (2015). Changes in the aromatic profile of espresso coffee as a function of the grinding grade and extraction time: A study by the electronic nose system. *Journal of Agricultural and Food Chemistry*, 63(8): 2321-2327. <https://doi.org/10.1021/jf505691u>
- [8] Rabersyah, D. (2016). Identifikasi jenis bubuk kopi menggunakan electronic nose dengan metode pembelajaran backpropagation. *Jurnal Nasional Teknik Elektro*, 5(3): 332-338. <https://doi.org/10.25077/jnte.v5n3.305.2016>
- [9] Hidayat, S.N., Triyana, K., Fauzan, I., Julian, T., Lelono, D., Yusuf, Y., Ngadiman, N., Veloso, A.C.A., Peres, A.M. (2019). The electronic nose coupled with chemometric tools for discriminating the quality of black tea samples in situ. *Chemosensors*, 7(3): 29. <https://doi.org/10.3390/chemosensors7030029>
- [10] Edita, R., Darius, G., Vinauskienė, R., Eisinaite, V., Balčiūnas, G., Dobilienė, J., Tamkutė, L. (2018). Rapid evaluation of fresh chicken meat quality by electronic nose. *Czech Journal of Food Sciences*, 36(5): 420-426. <https://doi.org/10.17221/419/2017-cjfs>
- [11] Mohammad-Razdari, A., Ghasemi-Varnamkhasti, M., Yoosefian, S.H., Izadi, Z., Siadat, M. (2019). Potential application of electronic nose coupled with chemometric tools for authentication assessment in tomato paste. *Journal of Food Process Engineering*, 42(5): e13119. <https://doi.org/10.1111/jfpe.13119>
- [12] Galán, R.D.B., Herrera, M.B., de la Cuesta, P.L., Pérez-Magariño, S. (2020). Stepwise linear discriminant analysis to differentiate Spanish red wines by their Protected Designation of Origin or category using physico-chemical parameters. *Oeno One*, 54(1): 86-99. <https://doi.org/10.20870/oeno-one.2020.54.1.2588>
- [13] Tian, X., Wang, J., Ma, Z., Li, M., Wei, Z. (2019). Combination of an E-nose and an E-tongue for adulteration detection of minced mutton mixed with pork. *Journal of Food Quality*, 2019: 4342509. <https://doi.org/10.1155/2019/4342509>
- [14] Wang, Z., Sun, X., Miao, J., Wang, Y., Luo, Z., Li, G., (2017). Conformal prediction based on K-nearest neighbors for discrimination of ginsengs by a home-made electronic nose. *Sensors (Switzerland)*, 17(8): 1869. <https://doi.org/10.3390/s17081869>
- [15] Wang, X., Lim, L.T. (2014). Effect of roasting conditions on carbon dioxide degassing behavior in coffee. *Food Research International*, 61: 144-151. <https://doi.org/10.1016/j.foodres.2014.01.027>
- [16] Nasution, I.S., Sundari, S., Rifky, N. (2020). Data acquisition of multiple sensors in greenhouse using arduino platform. In *IOP Conference Series: Earth and Environmental Science*, 515(1): 012011. <https://doi.org/10.1088/1755-1315/515/1/012011>
- [17] Duda, R.O., Hart, P.E., Stork, D.G. (2000). *Pattern Classification*. 2nd Ed., California: John Wiley and Sons Inc.
- [18] Iduseri, A., Osemwenkhae, J.E. (2015). An efficient variable selection method for predictive discriminant analysis. *Annals of Data Science*, 2(4): 489-504. <https://doi.org/10.1007/s40745-015-0061-9>
- [19] Tharwat, A., Gaber, T., Ibrahim, A., Hassanien, A.E. (2017). Linear discriminant analysis: A detailed tutorial. *AI Communications*, 30(2): 169-190. <https://doi.org/10.3233/AIC-170729>
- [20] Sathish, T., Rangarajan, S., Muthuram, A., Praveen Kumar, R. (2020). Analysis and modelling of dissimilar materials welding based on k-nearest neighbour predictor. *Materials Today: Proceedings*, 21: 108-112. <https://doi.org/10.1016/j.matpr.2019.05.371>
- [21] Ehsani, R., Drabløs, F. (2020). Robust distance measures for kNN classification of cancer data. *Cancer informatics*, 19: 1176935120965542. <https://doi.org/10.1177/1176935120965542>
- [22] Bonet, I., Rodríguez, A., Grau, R., García, M.M., Saez, Y., Nowé, A. (2008). Comparing distance measures with visual methods. *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 5317: 90-99. <https://doi.org/10.1007/978-3-540-88636-58>
- [23] Chomboon, K., Chujai, P., Teerarassamee, P., Kerdprasop, K., Kerdprasop, N. (2015). An empirical study of distance metrics for k-nearest neighbor algorithm. In *Proceedings of the 3rd international conference on industrial application engineering*, pp. 280-285. <https://doi.org/10.12792/iciae2015.051>
- [24] Radi, Rivai, M., Purnomo, M.H. (2016). Study on electronic-nose-based quality monitoring system for coffee under roasting. *Journal of Circuits, Systems and Computers*, 25(10): 1650116. <https://doi.org/10.1142/S0218126616501164>
- [25] Magfira, D.B., Sarno, R. (2018). Classification of Arabica and Robusta coffee using electronic nose. In *2018 International Conference on Information and Communications Technology (ICOIACT)*, pp. 645-650. <https://doi.org/10.1109/ICOIACT.2018.8350725>
- [26] Yang, Z.G., Li, H.Q., Zhu, L.P., Liu, Q., Sikandar, A.

- (2017). A case based method to predict optimal k value for k-NN algorithm. *Journal of Intelligent and Fuzzy Systems*, 33(1): 55-65. <https://doi.org/10.3233/JIFS-161062>
- [27] Bulut, F., Amasyali, M.F. (2017). Locally adaptive k parameter selection for nearest neighbor classifier: One nearest cluster. *Pattern Analysis and Applications*, 20(2): 415-425. <https://doi.org/10.1007/s10044-015-0504-0>
- [28] Suwanda, R., Syahputra, Z., Zamzami, E.M. (2020). Analysis of euclidean distance and manhattan distance in the K-means algorithm for variations number of centroid K. *Journal of Physics: Conference Series*, 1566(1): 012058. <https://doi.org/10.1088/1742-6596/1566/1/012058>
- [29] Guzun, G., Canahuate, G. (2017). Supporting dynamic quantization for high-dimensional data analytics. *Proceedings of the ExploreDB'17. International Workshop on Exploratory Search in Databases and the Web (4th: 2017: Chicago, Ill.)*, pp. 1-6. <https://doi.org/10.1145/3077331.3077336>
- [30] Gan, G., Ma, C., Wu, J. (2020). *Data clustering: Theory, algorithms, and applications*. Society for Industrial and Applied Mathematics. <https://doi.org/10.1137/1.9780898718348>
- [31] Basaran, B., Günes, F. (2016). *Data clustering. Intelligent Multidimensional Data Clustering and Analysis*, 31(3): 28-72. <https://doi.org/10.4018/978-1-5225-1776-4.ch002>
- [32] Chomboon, K., Chujai, P., Teerarassamee, P., Kerdprasop, K., Kerdprasop, N. (2015). An empirical study of distance metrics for k-nearest neighbor algorithm. In *Proceedings of the 3rd international conference on industrial application engineering*, pp. 280-285. <https://doi.org/10.12792/iciae2015.051>
- [33] Dokas, P., Ertoz, L., Kumar, V., Lazarevic, A., Srivastava, J., Tan, P.N. (2002). *Data mining for network intrusion detection*. National Science Foundation Workshop on Next Generation Data Mining, 38(7): 21-30.
- [34] Shirkorshidi, A.S., Aghabozorgi, S., Ying Wah, T. (2015). A comparison study on similarity and dissimilarity measures in clustering continuous data. *PLoS ONE*, 10(12): 1-20. <https://doi.org/10.1371/journal.pone.0144059>
- [35] Leys, C., Klein, O., Dominicy, Y., Ley, C. (2018). Detecting multivariate outliers: Use a robust variant of the Mahalanobis distance. *Journal of Experimental Social Psychology*, 74: 150-156. <https://doi.org/10.1016/j.jesp.2017.09.011>