Recognition of Original Arabica Civet Coffee based on Odor using Electronic Nose and Machine Learning

Whilly Harsono[1], Riyanarto Sarno[1,2*], Shoffi Izza Sabilla[2]
[1]*Departement of Information Technology Management, Faculty of Creative Design and Digital Business,*
[2]*Department of Informatics, Faculty of Intelligent Electrical and Informatics Technology,*
Institut Teknologi Sepuluh Nopember (ITS) Sukolilo, Surabaya 60111, Indonesia
** Corresponding author's Email: riyanarto@if.its.ac.id*

***Abstract*—Many studies have used an electronic nose (E-nose) to detect several types of coffee. To the best of our knowledge, none of the studies have tried to detect odors from a mixture of several types of coffee. Therefore, this research proposes E-nose which can be used to recognize original Arabica civet coffee. The mixture of Arabica civet coffee and Robusta coffee (non-civet coffee) is used as the object of this research. Nine combinations of mixture are prepared in this study. Those combinations are referred to as classes. After collecting the data, a statistical calculation would be determined to obtain parameter statistics. Moreover, the classification method used in this study is to recognize original Arabica civet coffee and original Robusta coffee. Several classifications had been compared, namely Logistic Regression (LR), Linear Discriminant Analysis (LDA), and K-Nearest Neighbors (KNN). The best result is the KNN method with an accuracy value of 97.7% for nine classes.**

Keywords—E-nose, Classification, Sensors, Arabica Coffee, Robusta Coffee, Civet Coffee.

I. INTRODUCTION

Traditionally, the aroma of coffee has been used to differentiate the originality of coffee. The aroma of coffee contains gas which is obtained by determining the gas content. During the roasting, temperature increases, and the biological process occurs. Then, coffee releases a robust aroma [1]. New compounds formed by physical and chemical reactions evaporate. E-nose has the ability to simulate the work of the human sense of smell. An electronic nose is made to catch the gas and recognize odors by using sensors [2]. The database of aroma produced by coffee is a pattern of odor, one of which functions to develop the system that can recognize a pattern, so it can be classified and be inspected [3].

In 2016, a study was conducted to classify coffee using a backpropagation neural network. The result showed that backpropagation neural network is capable of determining the differences [4] between Arabica and Robusta with a success rate of 40%.

Another E-nose study attained an accuracy of 71% for the Support Vector Machine (SVM) method and 57% for the Perceptron method. The study tried to classify the aroma of Arabica coffee and the aroma of Robusta coffee. The SVM method could recognize Arabica coffee and Robusta coffee with better results than the Perceptron method [5]. However, the research had a weakness in the classification method. The result of the classification has a lesser percentage of accuracy. Moreover, there was not any statistical calculation that could be used for preprocessing before classifying the data.

Therefore, this study aims to improve the weaknesses of the previous studies. E-nose used in this study has different characteristics to identify the odor and aroma of the gas because it consists of various types of sensors [6]. Then, the preprocessing stage using statistical calculations can obtain the characteristics from each signal response. This study has three values from a combination of statistical calculations; there are the values of average and standard deviation, the values between the minimum and maximum, and the values between average, standard deviation, and minimum and maximum values. After preprocessing stage, the calculation continues with the classification phase. Confusion matrix [7] is used in this study to evaluate the classification method.

Besides, we try to use other classification methods, so the results can be compared. Other classification methods that we use in this study are Logistic Regression (LR), Linear Discriminant Analysis (LDA) [8], and K-Nearest Neighbors (KNN). Comparing the result generated from three classification methods use the confusion matrix, and the best accuracy is chosen for the purpose of this study.

II. RELATED RESEARCH

A. *E-nose using Backpropagation Neural Network*

This study uses a system that puts several sets of gas sensors and receives input signals from TGS 2610, TGS 2611, TGS 2602, TGS 2620, and TGS 822 [4]. The resistance of the sensor results in a change of voltage when the sensor detects the presence of gaseous elements from the aroma of coffee. This signal is operated by a signal conditioning circuit to be delivered to the analog-digital converter (ADC) circuit and to change over into digital form. The process continues when the digital signal transmitted to the Personal Computer (PC) and to be processed using backpropagation NN (Neural Network). Backpropagation NN used is built with architecture 1 input layer 5 nodes (x_1-x_5), 1 hidden layer 6 nodes, and 1 output layer 2 nodes as shown in Fig. 1.

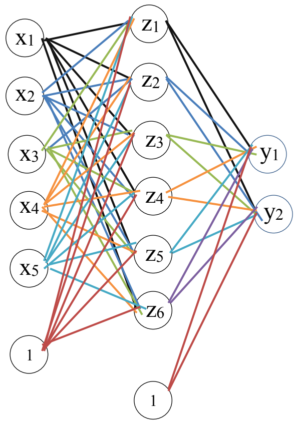

Figure 1. Backpropagation Design

The conclusion of this study is obtained after several tests and analyses. It can be concluded that the identification of Arabica coffee using Backpropagation NN is able to identify with a success rate of 40% and then for Robusta coffee of 100%. Besides, the system can also identify air or without coffee with an accuracy of 100%. Unfortunately, this study can only distinguish two classes of coffee, namely Arabica coffee and Robusta coffee.

B. *E-nose using SVM and Perceptron*

This study uses two methods for classifying the aroma of Arabica coffee and the aroma of Robusta coffee; they are the SVM and Perceptron methods. First, the study reduced data noise using discrete wavelet transforms. After finishing the process, the data would be passing through the feature extraction stage. The next step was using the SVM [5] and Perceptron methods for classification. The result of both methods showed the highest accuracy values and the lowest error for the classification of Arabica coffee and Robusta coffee. At the end of this study, the researcher would like to show that after classified using SVM and Perceptron methods [9], each result would be compared and the best result would be chosen. So, the conclusion is that the E-nose is able to identify between Arabica coffee and Robusta coffee. The best accuracy value is generated by the SVM method of which accuracy is 71%; however, this result still needs improvement.

C. *Classification using Radial Basis Function*

The gas sensors receive the aroma of coffee, change it into the transmitted signal, and analyze using pattern recognition. The E-nose used in this study is a sensor array polymer-coated for classification coffee variant, such as Arabica Bengkulu, Arabica Sidikalang, Arabica Papandayan, and Arabica Kerinci planted in different areas [10]. This study uses Artificial Neural Network Radial Basis Function to classify the coffees. The features comprise energy, contrast, correlation, and homogeneity. These features are trained using the Radial Basis Function neural network in order to classify coffee into four classes, namely Arabica Kerinci, Arabica Papandayan, Arabica Bengkulu, and Arabica Sidikalang.

D. *Statistical Data*

Statistics are a group of data in the system of numbers or not numbers relating to certain problems arranged in the form of tables, lists, diagrams, or others, so statistics are the result of data processing presented in tables, graphs, diagrams, etc. The purpose of statistics is to make it easier to interpret data used for a particular purpose. Statistics is scientific methods of how to collect, manage, analyze, interpret, and present data [11]. The purpose of statistics is to obtain a picture of a set of data that has been reviewed so that conclusions can be drawn from the data.

1) Random Data

Random or single data are data that have not yet been arranged or grouped into interval classes. A single data example in this study is when the MQ135 gas sensor receives gas in ppm units with the following results.

2) Group Data

Group data are data that have been arranged or grouped into interval classes. Group data are arranged in the form of frequency distributions or frequency tables using the formula:

$$K = 1 + 3{,}33 \log n$$
$$R = Biggest\ data - Smallest\ data$$
$$C = \frac{R}{K}$$

We can see the group of data in this study in Table 1. It consists of the values of the group data and the average of each range of values.

Table 1. Output data of five gas sensors

Data	Average
15-17	15.86
49-53	51.20
82-95	87.42
105-109	107.20
174-180	177.20

The data group in Table 1 is the output data from five gas sensors used in this study. The values obtained are different for each sensor. Therefore, we use the mean statistical calculation or so-called Average (Avg). The average value is very dependent on the magnitude of each data, including if there is an extreme value in the data, which is a very small or very large value and much different from the data group.

$$\bar{X} = \frac{\sum_{i=1}^{n} xi}{n} = \frac{x1+x2+x3+\cdots+xi}{n} \quad (1)$$

$$SD = \sqrt{\frac{\sum_{i=1}^{n}(xi-\mu)^2}{n-1}} \quad (2)$$

Eq. (1) means the standard deviation (SD) is the root of the middle of the square of the deviation of the mean or the square root of the mean squared. Eq. (2) is the standard deviation/sample deviation symbolized with s.

To determine the standard deviation, the method is to draw the root of the variance. For a set of data $x_1, x_2, x_3, \ldots, x_n$ (single data), the standard deviation can be determined i.e. the formula.

The smallest (minimum) and the largest (maximum) value is often used in calculating statistical data, including to find out how large the range or the difference is between the smallest data and the largest data. Minimum and maximum (Minmax) normalization is a normalization method by performing a linear transformation of the original data to produce a balance of comparative [12] values of data before and after the process.

III. Methodology

The implementation of this study began when E-nose received the signal [13] from the aroma of coffee. Then the signal was processed by Arduino into data and sent to the computer. The initial stage of data analysis [14] in this study was statistical calculation. The combined values of the statistical calculation with the classification value determined the best value of each method. The best method was evaluated by comparing the accuracy [15] values of the three methods in this study.

A. *Data Collection*

E-nose is a combination of several gas sensors to form an instrument that has the same function as the human sense of smell to detect odors. The gas sensor used is a gas sensor from the MQ family. It is composed of electrochemical sensors. Each of the sensors has a different level of selectivity combined to form a sensor array. The individual sensor patterns may not be selective, but the collective response of the whole array can be predicted. The sensor array characteristic pattern in the presence of a particular gas is tantamount to a signature which can be effectively learned with sufficient training data [16]. The list of sensors in this study aiming to recognize the aroma of coffee can be seen in Table 2.

Table 2. Gas Sensors in This Study

Sensor	Target
MQ 2	LPG, I-Butane, Propane, Methane, Alcohol, H2, Smoke
MQ 3	Alcohol, Methane, Benzene, Hexane, LPG, CO
MQ 4	Methane, Natural gas
MQ 7	CO
MQ 135	Carbon Dioxide

There are sensing elements, sensor base, and sensor cap that build the sensor. The detector of elements is divided into two common parts; they are sensing material and heater that function to heat the sensing element. Depending on the target of the gas, the detector reprocesses different gases, such as Alcohol, NH4, and CO_2. The gas contained in the aroma of coffee will be recognized and captured by each sensor used in this study. The sensors will send the data through Arduino to the computer. The E-nose design capturing the aroma of coffee can be seen in Figure 2.

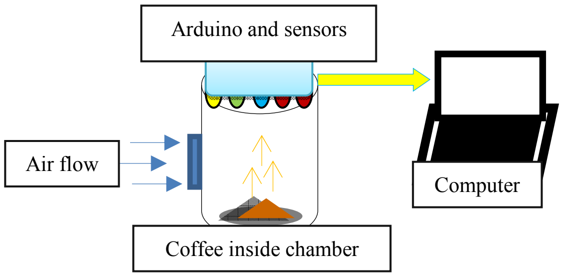

Figure 2. E-Nose Design for Recognition

The dataset used is the data in the form of a gas of coffee aroma. The aroma released by the coffee is the result of roasting process. The process begins with a process called the first crack; in which coffee explode due to gas pressure from the coffee until the flavour is formed from a coffee due to changes in chemical reactions within, this is called the second crack [17]. This study uses a comparison of mixed coffee. This study uses a comparison of mixed coffee (mixture) between Arabica Civet coffee and Non-civet coffee [18]. Nine mixes are used as data which are referred to as classes. The comparison of Arabica Coffee in Table 3 was carried out 50 times to nine classes.

Table 3. Arabica Civet Coffee and Non-Civet Coffee Classes

Proportion Arabica Civet Coffee and Non-Civet Coffee	Class
Civet 0% and Non-Civet 100%	L0-NL100
Civet 10% and Non-Civet 90%	L10-NL90
Civet 20% and Non-Civet 80%	L20-NL80
Civet 25% and Non-Civet 75%	L25-NL75
Civet 50% and Non-Civet 50%	L50-NL50
Civet 75% and Non-Civet 25%	L75-NL25
Civet 80% and Non-Civet 20%	L80-NL20
Civet 90% and Non-Civet 10%	L90-NL10
Civet 100% and Non-Civet 0%	L100-NL0

Each data was collected for 15 minutes at room temperature. During 15 minutes, 300 data were collected. Each class had 50 tests. Coffee used as the experiment material is ground coffee with an ideal grinder level, from coarse to medium size, with coffee weight of each class being 15 grams. The output of the detection of coffee aroma produced a digital value derived from each sensor.

B. *Preprocessing Data*

The next step is to process the data detected by E-nose on the coffee as seen in Figure 3. This process used machine learning to estimate the accuracy of the best model in the invisible data by evaluating the actual data that is not visible [19]. So, the accuracy was estimated using statistical methods for data validation purposes. The first process study is to calculate the Avg and SD value. The second is to calculate the Minmax value. The last is to calculate using the Avg, SD, and Minmax statistical methods.

The first calculation started by collecting data on Avg values and SD values of nine classes of the mixture Arabica civet coffee and Non-civet Arabica coffee into tables and stored in a Microsoft Excel file. Afterward, a new file was created in Microsoft Excel which contained the recapitulation result of the nine classes. The amount of data collected from the recapitulation was 450 data with a detail of 50 data from each class. The second data processing is to calculate the Minmax value of the nine classes of the mixture of Arabica civet coffee and Non-civet Arabica coffee [18].

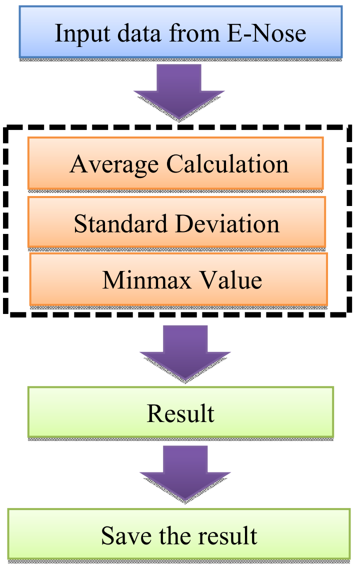

Figure 3. Preprocessing Data Flowchart

The result of data processing of this study was saved in a .csv (Comma Separated Values) format. The process of changing the file extension from (Microsoft Office Excel) .xls to .csv was done online by uploading the file, selecting the convert option, and downloading the result. The next stage is to run the .csv file using the Visual Studio Code. Eventually, the accuracy was obtained using formulation in a python programming language.

C. *Data Analysis*

The data used in this study were divided into two parts with the cross-validation method. The process was followed by the classification phase. Finding an effective data partition or sampling method is a method used to minimize errors in estimating accuracy, comparing methods, and finding the best method [20].

Calculations carried out at the data processing stage were classified using the Logistic Regression (LR), Linear Discriminant Analysis (LDA), and K-Nearest Neighbors (KNN) methods. The aim is to estimate the accuracy of the statistical calculation when processing data. On the classification stage, the dataset that had been stored was divided into two types of data, namely training data and testing data. The data stored were of nine classes of the mixture of Arabica civet coffee and Non-civet Arabica coffee. The data, amounting to 50 data from each class, were divided into 80% for training data and 20% for testing data from the total input data.

D. *Data Evaluation*

The evaluation of the data classified in this study used the CF (Confusion Matrix) method [21]. CF is one of the tools commonly used in evaluating machine learning [22] which contains two or more categories. This method divides data into 2 classes, namely data generated from the classifier (Predictive Class) and originally known data (Actual Class). The classification process using CF had four terms of results used to calculate the performance of classification, namely TP, TN, FP, and FN. TP is positive data that were discovered correctly. If the category generated by the classifier [23] is similar to the existing data class, the data are recorded in the TP. TN is the number of negative data that were discovered correctly, while FP is negative data but detected as positive. On the other hand, FN is the reverse of TP; it means that the data are positive, but detected as negative [24].

TP is the data from class 1 classified as class 1. Data from class 0 that is correctly classified as class 0 in TN. Then, the opposite of TN is FN meaning the amount of the data from class 1 incorrectly classified as class 0 [25]. Based on the value of TN, FP, FN, and TP [26], the value of precision, memory, and accuracy can be obtained. The precision value describes the amount of data that is categorized positively, then classified correctly divided by the total data results in positive classification. Precision can be seen in Eq. (3).

$$Prc = \frac{TP}{(TP+FP)} \times 100\% \tag{3}$$

Meanwhile, recall establishes the percentage of positive data correctly classified by the system. In binary classification, recall is also known as sensitivity. The calculation can be seen in Eq. (5).

$$Recall = \frac{TP}{(TP+FN)} \times 100\% \tag{4}$$

$$f1-score = \frac{2TP}{(2TP+FP+FN)} \times 100\% \tag{5}$$

$$Acc = \frac{(TP+TN)}{(TP+TN+FP+FN)} \times 100\% \tag{6}$$

The accuracy value describes how precise the system can classify data correctly [27]. In other words, the value of accuracy is the comparison between correctly classified data and all data. The accuracy value can be obtained by Eq. (6).

Table 4. Confusion Matrix Result in Python

Classes	Precision	Recall	F-1 Score	Support
L75-NL25	0,90	0,90	0,90	10
L0-NL100	1,00	1,00	1,00	9
L10-NL90	1,00	1,00	1,00	8
L100-NL0	1,00	1,00	1,00	15
L20-NL80	1,00	1,00	1,00	12
L25-NL75	1,00	1,00	1,00	5
L50-NL50	1,00	1,00	1,00	14
L80-NL20	1,00	1,00	1,00	7
L90-NL10	0,86	0,86	0,86	10

The result of calculating the Confusion Matrix using the LR classification method to form a classifier model is presented in Table 4 [28]. This model is a representation used to predict new data classes that have never existed. The logic is to let the machine learn from the training set and be tested using the testing set.

IV. Results and Discussion

This study was conducted aiming at utilizing E-nose to detect the aroma of coffee followed by calculating the statistics included in the initial process (pre-processing). The data were divided into nine classifications based on the value of percentages of coffee mixture as presented in Table 3. First, we attempted to obtain the accuracy of the average calculation and standard deviation. The accuracy of the LR method calculated statistically in terms of average and standard deviation is 91.38%, and of the LDA method is 91.38%. The highest result of accuracy is the KNN method of 96.11%. Upon the completion of calculating the accuracy from average and standard deviation, the next step is to determine the other accuracy value from other statistical calculations. The highest accuracy is obtained by the KNN method of 95.27%. The process is followed by the statistical calculation of average-standard deviation-min-max. So, three statistical calculations are essential to make to classify the comparison between one method to the others. From the result, the highest result is the KNN method of which accuracy is 96.38%. The result of the classification in this study was compared and selected by statistical methods in terms of the highest accuracy value.

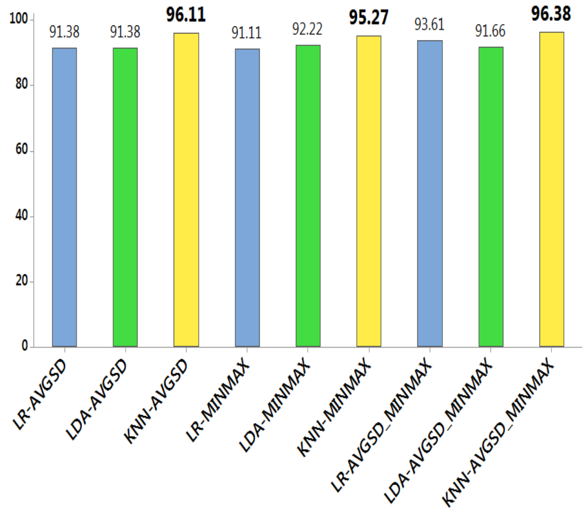

Figure 4. Classification Using Statistical Methods

Figure 4 shows that the KNN classification method obtains the highest accuracy result. The KNN classification method had passed three statistical calculations for data validation, namely the calculation of the minmax value of 95.27%, the calculation of the average and standard deviation value of 96.11%, and the calculation of the min, max, average and standard deviation value of 96.38% [29]. The process is followed by evaluating the accuracy of the classification using the confusion matrix in machine learning.

Classification of the three methods was combined using machine learning calculation and the result showed that the KNN method obtained the highest value among the other methods. The classification value with the initial process using the overall statistical calculation obtained the highest value of 96.38. The next step is to calculate classification accuracy using CF (Confusion Matrix). The calculation was also done using ML.

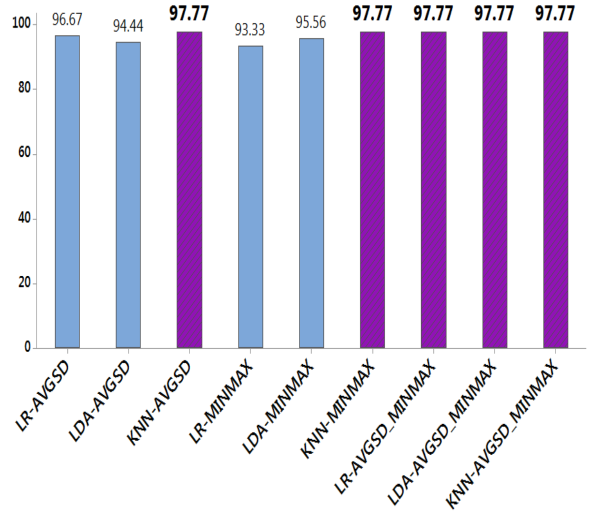

Figure 5. Accuracy of Classification Using Confusion Matrix

Once the best classification method was determined, the KNN method was applied to the existing data using a confusion matrix to prove its accuracy. The result of CF in finding the best classification method and finding the best validation method using the statistical method is KNN of which value is 97.77%. For the data validation method, the best result is obtained when using more statistical calculations. In Figure 5, for instance, the best result is obtained when all statistical calculations were applied (Avg, SD, minimum and maximum value).

Table 5. Confusion Matrix of KNN for Avg-SD Value

PREDICTION \ TARGET	L75-NL25	L0-NL100	L10-NL90	L100-NL0	L20-NL80	L25-NL75	L50-NL50	L80-NL20	L90-NL10
L75-NL25	**10**	0	0	0	0	0	0	0	0
L0-NL100	0	**9**	0	0	0	0	0	0	0
L10-NL90	0	0	**8**	0	0	0	0	0	0
L100-NL0	0	0	0	**15**	0	0	0	0	0
L20-NL80	0	0	0	0	**12**	0	0	0	0
L25-NL75	0	0	0	0	0	**5**	0	0	0
L50-NL50	0	0	0	0	0	0	**14**	0	0
L80-NL20	0	0	0	0	0	0	0	**7**	0
L90-NL10	0	0	0	0	0	0	0	0	**8**

The result of the classification accuracy, using the confusion matrix of the mixture of Arabica civet coffee and non-civet Arabica coffee obtained an accuracy of 97.77% from the KNN algorithm and Avg-SD statistical calculation. From Table 5, the coffee mixture data of the L75-NL25 prediction class is classified into 10 data in the target class. The L0NL100 prediction class classifies 9 data in its target class. The L10NL90 prediction class classifies 8 data in its target class. The L100NL0 prediction class classifies 15 data in its target class. The L20NL80 prediction class classifies 12 data in its target class. The L25NL75 prediction class classifies 5 data in its target class. The L50NL50 prediction class classifies

14 data in its target class. The L80NL20 prediction class classifies 7 data in its target class. And, the L90NL10 prediction class classifies 8 data in its target class.

Table 6. Confusion Matrix of KNN for Minmax Value

		TARGET								
		L75-NL25	L0-NL100	L10-NL90	L100-NL0	L20-NL80	L25-NL75	L50-NL50	L80-NL20	L90-NL10
PREDICTION	L75-NL25	**10**	0	0	0	0	0	0	0	0
	L0-NL100	0	**9**	0	0	0	0	0	0	0
	L10-NL90	0	0	**8**	0	0	0	0	0	0
	L100-NL0	0	0	0	**15**	0	0	0	0	0
	L20-NL80	0	0	0	0	**12**	0	0	0	0
	L25-NL75	0	0	0	0	0	**5**	0	0	0
	L50-NL50	0	0	0	0	0	0	**14**	0	0
	L80-NL20	0	0	0	0	0	0	0	**7**	0
	L90-NL10	0	0	0	0	0	0	0	2	**8**

The accuracy result using confusion matrix in Table 6 of the L90NL10 classification shows that there are 2 data included in the L80NL20 class target, meaning that the detection between the aroma of coffee from the 2 types of mixture indicates similarity when detecting the aroma of coffee. The similarity in the aroma detection data leads to the predicted [30] data to a class different from the target class. So, the result of the classification of the mixture of Arabica civet coffee and non-civet Arabica coffee obtains an accuracy of 97.77% from the KNN algorithm and Minmax statistical calculation.

Table 7. Confusion Matrix of KNN for Avg-SD-Minmax

		TARGET								
		L75-NL25	L0-NL100	L10-NL90	L100-NL0	L20-NL80	L25-NL75	L50-NL50	L80-NL20	L90-NL10
PREDICTION	L75-NL25	**10**	0	0	0	0	0	0	0	0
	L0-NL100	0	**9**	0	0	0	0	0	0	0
	L10-NL90	0	0	**8**	0	0	0	0	0	0
	L100-NL0	0	0	0	**15**	0	0	0	0	0
	L20-NL80	0	0	0	0	**12**	0	0	0	0
	L25-NL75	0	0	0	0	0	**5**	0	0	0
	L50-NL50	0	0	0	0	0	0	**14**	0	0
	L80-NL20	0	0	0	0	0	0	0	**7**	0
	L90-NL10	0	0	0	0	0	0	0	2	**8**

Similarly, the result of the classification accuracy using the confusion matrix in Table 7 of a coffee mixture between Arabica civet coffee and non-civet Arabica coffee using the KNN algorithm and Avg-SD-Minmax statistical calculation obtains an accuracy percentage of 97.77%. If seen from the average of all accuracy generated, the classification of Arabica civet coffee and non-civet Arabica coffee can be done using data from the aroma detection result using E-nose. Data classification highly affects the value of accuracy produced, the more the attributes used in the classification process, the higher the accuracy value.

Table 8. Algorithm Performance of Confusion Matrix

	Precision	Recall	F1-Score
Mix 75-NL25	1.00	1.00	1.00
Mix L0-NL100	1.00	1.00	1.00
Mix L10-NL90	1.00	1.00	1.00
Mix L100-NL0	1.00	1.00	1.00
Mix L20-NL80	1.00	1.00	1.00
Mix L25-NL75	1.00	1.00	1.00
Mix L50-NL50	0.78	1.00	0.88
Mix L80-NL20	1.00	1.00	1.00
Mix L90-NL10	1.00	1.00	1.00
Accuracy		**0.98**	

Table 8 shows that the accuracy result of the dataset classification is significant, as the amount of false negative data and false positive data produced is highly similar or of symmetric value.

V. Conclusion

This study aims to recognize original Arabica civet coffee using the Electronic Nose (E-nose) system with the MQ family of gas sensors, namely MQ2, MQ3, MQ4, MQ7, and MQ135. Statistical calculation in the preprocessing stage was used to obtain the characteristics of each signal which, in turn, indicated a good enhancement. The highest result of the classification accuracy using the confusion matrix of the mixture of Arabica civet coffee and non-civet Arabica coffee was from the KNN algorithm of which accuracy is 97.77%. The result of the confusion matrix also showed that the best method to validate data is to use all of the statistical calculations. Thus, it is imperative for future studies to achieve higher accuracy values by adding other combination statistical calculation methods in the preprocessing data stage.

ACKNOWLEDGMENTS

This study was partially funded by the Ministry of Research, Technology and Higher Education of the Republic of Indonesia under the PMDSU Program managed by Institut Teknologi Sepuluh Nopember (ITS).

REFERENCES

[1] U. Jumhawan, S. P. Putri, Yusianto, T. Bamba, and E. Fukusaki, "Quantification of coffee blends for authentication of Asian palm civet coffee (Kopi Luwak) via metabolomics: A proof of concept," *J. Biosci. Bioeng.*, vol. 122, no. 1, pp. 79–84, Jul. 2016, doi: 10.1016/j.jbiosc.2015.12.008.

[2] E. Ongo *et al.*, "Chemometric discrimination of philippine civet coffee using electronic nose and gas chromatography mass spectrometry," in *Procedia Engineering*, 2012, vol. 47, pp. 977–980, doi: 10.1016/j.proeng.2012.09.310.

[3] L. Eusebio, L. Capelli, and S. Sironi, "Electronic nose testing procedure for the definition of minimum performance requirements for environmental odor monitoring," *Sensors (Switzerland)*, vol. 16, no. 9, Aug. 2016, doi: 10.3390/s16091548.

[4] D. Rabersyah, . F., and . D., "Identifikasi Jenis Bubuk Kopi menggunakan Electronic Nose dengan Metode Pembelajaran Backpropagation," *J. Nas. Tek. ELEKTRO*, vol. 5, no. 3, p. 332, Oct. 2016, doi: 10.25077/jnte.v5n3.305.2016.

[5] D. B. Magfira and R. Sarno, "Classification of Arabica and Robusta coffee using electronic nose," in *2018 International Conference on Information and Communications Technology, ICOIACT 2018*, Apr. 2018, vol. 2018-January, pp. 645–650, doi: 10.1109/ICOIACT.2018.8350725.

[6] S. I. Sabilla, R. Sarno, and J. Siswantoro, "Estimating Gas

Concentration using Artificial Neural Network for Electronic Nose," in *Procedia Computer Science*, 2017, vol. 124, pp. 181–188, doi: 10.1016/j.procs.2017.12.145.

[7] S. Wakhid, R. Sarno, S. I. Sabilla, and D. B. Maghfira, "Detection and Classification of Indonesian Civet and Non-Civet Coffee Based on Statistical Analysis Comparison Using E-Nose," *Int. J. Intell. Eng. Syst.*, vol. 13, no. 4, 2020, doi: 10.22266/ijies2020.0831.06.

[8] S. I. Sabilla and R. Sarno, "Development of wavelet transforms to predict methane in chili using the electronic nose," in *Proceeding - ICAMIMIA 2017: International Conference on Advanced Mechatronics, Intelligent Manufacture, and Industrial Automation*, Jun. 2018, pp. 271–276, doi: 10.1109/ICAMIMIA.2017.8387600.

[9] A. Amrin, "Perbandingan Metode Neural Network Model Radial Basis Function Dan Multilayer Perceptron Untuk Analisa Risiko Kredit Mobil." .

[10] I. W. Ramadiansyah and S. S. . M. T. Dr. Danang Lelono, "Klasifikasi Kopi Dengan Jaringan Syaraf Tiruan Berbasis E-Nose Quartz Crystal Microbalance," 2019.

[11] S. M. Scott, D. James, and Z. Ali, "Data analysis for electronic nose systems," *Microchimica Acta*, vol. 156, no. 3–4. pp. 183–207, Dec. 2006, doi: 10.1007/s00604-006-0623-9.

[12] W. Dong *et al.*, "Comparative evaluation of the volatile profiles and taste properties of roasted coffee beans as affected by drying method and detected by electronic nose, electronic tongue, and HS-SPME-GC-MS," *Food Chem.*, vol. 272, pp. 723–731, Jan. 2019, doi: 10.1016/j.foodchem.2018.08.068.

[13] D. R. Wijaya, R. Sarno, E. Zulaika, and S. I. Sabila, "Development of mobile electronic nose for beef quality monitoring," in *Procedia Computer Science*, 2017, vol. 124, pp. 728–735, doi: 10.1016/j.procs.2017.12.211.

[14] D. R. Wijaya, R. Sarno, and E. Zulaika, "Sensor Array Optimization for Mobile Electronic Nose: Wavelet Transform and Filter Based Feature Selection Approach," *Int. Rev. Comput. Softw.*, vol. 11, no. 8, p. 659, Aug. 2016, doi: 10.15866/irecos.v11i8.9425.

[15] S. I. Sabilla, R. Sarno, and K. Triyana, "Optimizing Threshold Using Pearson Correlation for Selecting Features of Electronic Nose Signals," *Int. J. Intell. Eng. Syst.*, vol. 12, no. 6, 2019, doi: 10.22266/ijies2019.1231.08.

[16] A. Fahmi, E. Sugiarto, A. Winarno, S. Sumpeno, and M. H. Purnomo, "Waqf lands assets classification based on productive value for business development using Naïve bayes," in *2018 International Seminar on Research of Information Technology and Intelligent Systems, ISRITI 2018*, Nov. 2018, pp. 622–626, doi: 10.1109/ISRITI.2018.8864489.

[17] Radi, M. Rivai, and M. H. Purnomo, "Study on Electronic-Nose-Based Quality Monitoring System for Coffee Under Roasting," *J. Circuits, Syst. Comput.*, vol. 25, no. 10, p. 1650116, Oct. 2016, doi: 10.1142/S0218126616501164.

[18] E. R. Arboleda, "Discrimination of civet coffee using near infrared spectroscopy and artificial neural network," *Int. J. Adv. Comput. Res.*, vol. 8, no. 39, pp. 324–334, Nov. 2018, doi: 10.19101/IJACR.2018.839007.

[19] I. H. Sarker, A. S. M. Kayes, and P. Watters, "Effectiveness analysis of machine learning classification models for predicting personalized context-aware smartphone usage," *J. Big Data*, vol. 6, no. 1, p. 57, Dec. 2019, doi: 10.1186/s40537-019-0219-y.

[20] G. Jiang and W. Wang, "Error estimation based on variance analysis of k-fold cross-validation," *Pattern Recognit.*, vol. 69, pp. 94–106, Sep. 2017, doi: 10.1016/j.patcog.2017.03.025.

[21] D. Seka, B. S. Bonny, A. N. Yoboué, S. R. Sié, and B. A. Adopo-Gourène, "Identification of maize (Zea mays L.) progeny genotypes based on two probabilistic approaches: Logistic regression and naïve Bayes," *Artif. Intell. Agric.*, vol. 1, pp. 9–13, Mar. 2019, doi: 10.1016/j.aiia.2019.03.001.

[22] R. Kohavi, N. J. Rothleder, and E. Simoudis, "Emerging trends in business analytics," *Commun. ACM*, vol. 45, no. 8, pp. 45–48, 2002, doi: 10.1145/545151.545177.

[23] F. Kulapichitr, C. Borompichaichartkul, I. Suppavorasatit, and K. R. Cadwallader, "Impact of drying process on chemical composition and key aroma components of Arabica coffee," *Food Chem.*, vol. 291, pp. 49–58, Sep. 2019, doi: 10.1016/j.foodchem.2019.03.152.

[24] N. Shiri Harzevili and S. H. Alizadeh, "Mixture of latent multinomial naive Bayes classifier," *Appl. Soft Comput. J.*, vol. 69, pp. 516–527, Aug. 2018, doi: 10.1016/j.asoc.2018.04.020.

[25] A. Lawi and Y. Adhitya, "Classifying Physical Morphology of Cocoa Beans Digital Images using Multiclass Ensemble Least-Squares Support Vector Machine," in *Journal of Physics: Conference Series*, Mar. 2018, vol. 979, no. 1, doi: 10.1088/1742-6596/979/1/012029.

[26] M. Huljanah, Z. Rustam, S. Utama, and T. Siswantining, "Feature Selection using Random Forest Classifier for Predicting Prostate Cancer," in *IOP Conference Series: Materials Science and Engineering*, Jul. 2019, vol. 546, no. 5, doi: 10.1088/1757-899X/546/5/052031.

[27] M. Xu, J. Wang, and L. Zhu, "The qualitative and quantitative assessment of tea quality based on E-nose, E-tongue and E-eye combined with chemometrics," *Food Chem.*, vol. 289, pp. 482–489, Aug. 2019, doi: 10.1016/j.foodchem.2019.03.080.

[28] B. Farran, C. Saunders, and M. Niranjan, "Machine learning for intrusion detection: Modeling the distribution shift," in *Proceedings of the 2010 IEEE International Workshop on Machine Learning for Signal Processing, MLSP 2010*, 2010, pp. 232–237, doi: 10.1109/MLSP.2010.5589161.

[29] T. Badriyah, M. Tahrir, and I. Syarif, "Predicting the Risk of Preeclampsia with History of Hypertension Using Logistic Regression and Naive Bayes," in *Proceedings - 2018 International Conference on Applied Science and Technology, iCAST 2018*, Oct. 2018, pp. 399–403, doi: 10.1109/iCAST1.2018.8751588.

[30] S. I. Sabilla, R. Sarno, and K. Triyana, "Optimizing threshold using pearson correlation for selecting features of electronic nose signals," *Int. J. Intell. Eng. Syst.*, vol. 12, no. 6, pp. 81–90, 2019, doi: 10.22266/ijies2019.1231.08.