ISSN (Print) : 2337-9030 ISSN (Online) : 2685-2608
PILLAR OF PHYSICS
JURNAL BERKALA ILMIAH FISIKA
Vol. 18 (1), 2024, page. 41-49
DOI : 10.24036/16783171074

THE USE OF ELECTRONIC NOSE IN MACHINE LEARNING-BASED OF JENGKOL (*ARCHIDENDRON PAUCHIFLORUM*) AND KABAU SEEDS (*ARCHIDENDRON BUBALINUM*) AUTHENTICATION

Dian Putri Mustika[1], Yenni Darvina[1*], Yulkifli[1] and Kuwat Triyana[2]

[1] *Departement of Physics, Universitas Negeri Padang, Jl. Prof. Dr. Hamka Air Tawar Padang 25131, Indonesia*
[2] *Departement of Physics, Universitas Gadjah Mada, Jl. Farmako Sekip Utara Yogyakarta 55281, Indonesia*
Corresponding author. Email: ydarvina@fmipa.unp.ac.id

ABSTRACT

This research aims to develop an authentication method for jengkol (Archidendron pauciflorum) and kabau (Archidendron bubalinum) seeds using Electronic Nose (E-Nose) integrated with Machine Learning based on Support Vector Machine (SVM). The background of this research is the physical and aroma similarities between the two types of seeds, which make visual identification difficult, especially after cutting, as well as the potential for counterfeiting due to differences in economic value, although to date no concrete evidence of counterfeiting has been found. The research was conducted using an E-Nose consisting of 10 MOS-based gas sensors to detect the aroma profile of jengkol and kabau seeds. The aroma data obtained was extracted using several statistical features such as mean, median, variance, standard deviation, and Area Under the Curve (AUC), then analyzed with the SVM kernel RBF model. Tests were conducted on 200 sample data with the results of only 2 misclassified data, resulting in accuracy, Recall_0, and Recall_1 of 99.00. These results show that the developed model is very stable and effective in distinguishing between the two types of seeds, even with variations in data collection methods and different collection times. These findings show that the combination of E-Nose and Machine Learning can be an efficient and economical solution for aroma-based product authentication, especially for food commodities that are difficult to distinguish visually, and has the potential to be applied to other products that require assurance of authenticity based on volatile profiles.

Keywords : *Electronic Nose, Machine Learning, SVM, Jengkol, Kabau Seeds*

I. INTRODUCTION

Indonesia is one of the tropical countries that has a high biodiversity of both flora and fauna whose distribution is very wide in the world, including unique and distinctive food crops and spices [1]. Two of these plants are jengkol (Archidendron pauchiflorum) and jaling or kabau seed (Archidendron bubalinum). Both plants are often used in traditional Indonesian cuisine. However, kabau seeds are not considered to have a high economic value which is why they are less popular than jengkol [2].

Jengkol is a plant that grows in tropical areas such as in several Southeast Asian countries, namely Indonesia, Malaysia and Thailand. Based on a report from the Central Statistics Agency (BPS) in 2023, jengkol production in Indonesia reached 157.16 thousand tons. This figure increased by 1248 tons (0.8%) compared to the previous year and increased by 60.23 thousand tons (62.14%) in the last five years. In 2023, West Sumatra was ranked as the top jengkol producer nationally with 21.3 thousand tons. Followed by West Java with 19.76 thousand tons, Lampung with 18.84 thousand tons, and Central Java with 16.79 thousand tons [3]. Although jengkol has an unpleasant odor, there are many benefits of this plant, such as the young shoots of jengkol that we can eat as vegetables and the seeds that we can eat with rice before or after processing such as boiled, fried, or added with spices [4]. Although jengkol has many benefits, extracts from jengkol can cause hypertrophy and damage to the heart, kidneys, lungs and pancreas. In addition, jengkol may also reduce the risk of ethanol induction of gastric lesions and protect the gastrointestinal mucosa [5].

Jaling, also known as biji kabau, is a plant that grows naturally in the forest. Various types of jengkol (Archidendron pauciflorum) and petai (Parkia speciosa) are used as fresh vegetables by the community. Kabau seeds are a traditional medicine in Sumatra for diabetes and stomach pain. The fruit's active phytochemicals

Submitted: Nov 4, 2024 Accepted: April 29, 2025 Published: April 30, 2025

include alkaloids, flavonoids, phenolics, tannins, and steroids. These compounds can be used as natural insecticides, blood glucose lowering, antidiabetics, antifungals, and antioxidants. [6]. In terms of physical size, kabau seeds do look different from jengkol, but the distinctive aroma released by kabau seeds is almost the same. The content of compounds in kabau seeds is also the same as the content of jengkol compounds which are one clan with kabau seeds.

Jengkol generally has a higher economic value than kabau [2]. This is because kabau is less popular outside of its native region in Sumatra [7]. This difference in economic value may create a potential motive for counterfeiting, although to date there is no concrete evidence to suggest that counterfeiting between these two types of seeds occurs. However, jengkol and kabau are also physically quite similar, especially when chopped, making it difficult to distinguish them visually despite their different odors. Therefore, it is important to develop authentication methods that can ensure product authenticity, especially in the context of increasingly competitive trade.

In the process of product authentication related to gas or odor, Gas Chromatography-Mass Spectrometry (GC-MS) is commonly used. GCMS is a highly effective analytical technique for distinguishing samples based on their chemical composition. The method works by separating components in a gas sample and then identifying them based on their molecular mass. GCMS can be used to distinguish pure essential oils and essential oils mixed with alcohol [8].

As technology evolves, the Electronic Nose (E-Nose) comes as a more economical and efficient alternative. An E-Nose is a device that mimics the human olfactory function by using chemical sensors to detect and recognize odors [9]. The way the E-Nose works involves picking up gases, which are then measured by the sensors to generate data patterns that can be interpreted by Machine Learning algorithms. Machine Learning is the ability of computers to learn and evolve like humans, but at a much greater speed and scale. The main goal of Machine Learning is to recognize patterns in data [10]. Recently, many Machine Learning techniques have been developed, and integrated into feature extraction, modeling and drift compensation of gas sensors. The goal of feature extraction is to preserve strong pattern information in the raw signal while removing noise [11]. In addition, feature extraction also aims to reduce the dimensionality of the data while ensuring that key information that defines the structure or pattern in the data is not lost. But to validate the results of enose, we used GC-MS to see the chemical composition of the two plants.

From several previous studies, such as research [12] which differentiated beef, chicken, and pork, it was found that beef and chicken have hydrocarbon and alcohol compound groups, while pork is dominant with aldehyde compounds such as dodecanal and 9-octadecanal. The results of the LDA model used show validation and testing accuracy reaching >99% for all extracted features. Thus, this E-Nose system has great potential in food authenticity testing and food adulteration detection, especially in identifying pork content in meat products.In addition to animal meat content, research has also been conducted in [13] which differentiates processed pork, beef and goat skin products. It is evident that the trend line graphs that appear different from each other. Likewise, the total value of the discriminant function of the LDA plot graph is 100% with each skin grouping center, namely pork skin, cow skin and goat skin at the center (-27, (-1)), (10,8) and (15,(-8)). This is in accordance with the original characteristics of each leather which has a distinctive odor. As for the comparison of the original leather characteristics with the samples of leather products; tanned leather and rambak crackers, there is a deviation of 5.77% for tanned leather products and 2.06% for rambak crackers from the center of the original leather grouping. However, the respective analyses of the leather samples and leather products are well identified and still refer to the analysis of the genuine leather samples.

Meanwhile, research [14] has also investigated the differences between green tea and black tea from different regions using E-Nose with Machine Learning algorithms. The PCA method reduces the data matrix of the feature extraction results by mapping the aroma pattern of each sample using two PCA principal components. The PCA reduction results in the integral feature extraction method show the largest cumulative percentage of variation of 97% and 100% for green tea samples, which indicates that PCA can accurately distinguish between green tea and black tea samples. This states that the E-Nose allows scent detection to be done quickly and without damaging the sample, thus saving time and money compared to conventional methods that may require longer and costly laboratory testing.

II. METHOD

This research was conducted using laboratory experimental method. This research was conducted at the Material Physics and Instrumentation Laboratory, Department of Physics, FMIPA, Gadjah Mada University, Yogyakarta. The independent variable in this study is the type of sample used, namely jengkol and kabau seeds. The dependent variable in this study is the aroma of jengkol and kabau seeds. The control variable in this study is the mass of jengkol and kabau seeds. The first step before conducting this research is to prepare tools and materials. The sample preparation tools used are digital scales and knives.

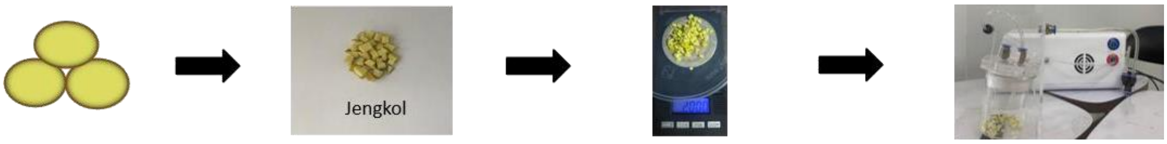

Fig. 1. Jengkol preparation

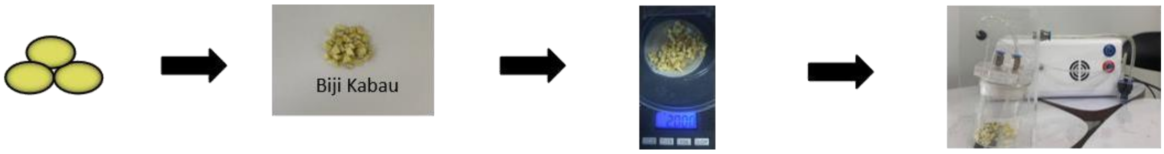

Fig. 2. Kabau seeds preparation

Fig. 3. Comparison of chopped Jengkol and Kabau seeds

While the data collection tools used are electronic nose, personal computer, data logger, usb, drain pump, teflon hose, 100 ml beaker, and acrylic box. The materials used are jengkol and kabau seeds. The following is a data collection scheme using enose.

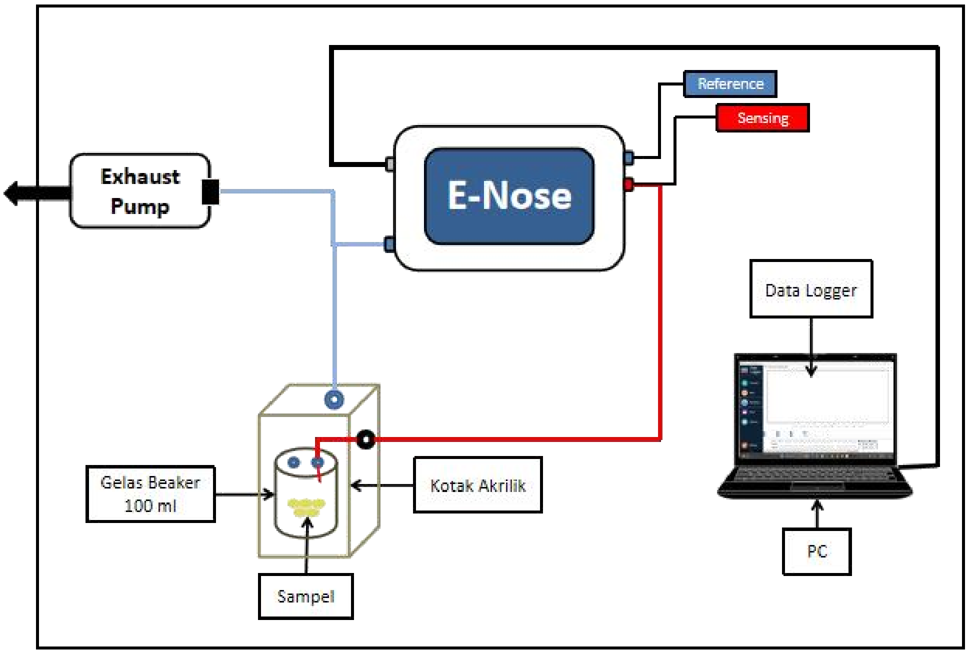

Fig. 4. Data retrieval scheme

An E-nose is a device that works like a sense of smell, with multiple gas sensors instead of olfactory receptors that can detect odors or scents. Scents are made up of molecules that each have a specific size and shape. The receptor captures a molecule and sends a signal to the brain, which identifies the scent associated with that molecule. The purpose of such identification is to recognize simple or complex scent patterns. Therefore, the output of the electronic nose can be the identity or characteristic properties of the sample, allowing the characteristic properties of the aroma [15].

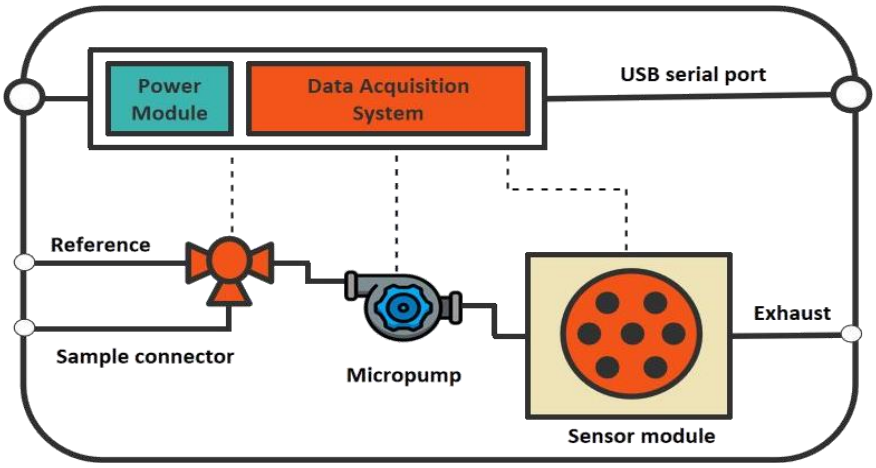

Fig. 5. Schematic of scent sampling on Genose C19

The E-nose consists of three main functional components that operate serially on aroma samples: the sample chamber, the sensor array and the data acquisition system. The sample to be tested is placed in the sample chamber. The aroma sample is flowed into the sensor array chamber located in the sensor chamber using an airflow system. Gas sensors are arranged into a sensor array to respond to certain chemical compounds in the sample. The response is in the form of an analog signal which is then captured by ADC (Analog to Digital Converter) so that it can be read by a computer.

The gas sensor that is often used in E-Nose is the MOS gas sensor. This sensor is in high demand due to its relative ease of use. MOS sensors have relatively low sensitivity, whereas new sensors can detect the presence of a substance when the substance has a concentration of several parts per million (ppm). The sensors used in Genose C19 are as follows:

Table 1. Gas Sensor Used

Gas Sensor	Selective Target Gas	Sensor Color Description on Data Logger
S1	Carbon monoxide, ethanol, hydrogen, isobutane, and methane	Dark Blue
S2	Ammonia, ethanol, hydrogen, hydrogen sulfide, and toluene	Orange
S3	Ethanol, hydrogen, isobutane, and methane	Green
S4	Carbon monoxide, ethanol, hydrogen, isobutane, and methane	Red
S5	Carbon monoxide, ethanol, hydrogen, isobutane, methane and propane	Purple
S6	Carbon monoxide, ethanol, hydrogen, isobutane, methane and propane	Brown
S7	Carbon monoxide, ethanol, hydrogen, methane	Lilac
S8	Acetone, benzene, carbon monoxide, ethanol, isobutane, methane, and n-hexane	Grey
S9	Ammonia, ethanol, hydrogen, and isobutane	Dark Yellow
S10	Chlorofluorocarbons, ethanol, and hydrofluorocarbons	Sky Blue

Of the 10 sensors used in the C-19 E-Nose, there are 2 sensors that have the lowest value compared to other sensors, namely S3 andS7 because of the little selective gas they detect.

A fundamental aspect of data analysis is feature extraction, where we extract important information from sensor response curves for further analysis. The goal is to capture sample characteristics from the raw data, then facilitate easy differentiation with pattern recognition algorithms and improve performance for classification. Typically, the output of an E-nose contains many signals to process. Therefore, the features of the data should be extracted first so that it can be easier to classify [16].

The feature extraction method applied in this research is :

$$Mean\,() = \frac{1}{} \sum_{=1} \boldsymbol{I} \tag{1}$$

$$Median = \text{median}\,() \tag{2}$$

$$Varians\,(K) = \frac{1}{} \sum_{=1} (\boldsymbol{i} - \bar{})^2 \tag{3}$$

$$Standard\ Deviation\,(K) = \sqrt{\frac{1}{} \sum_{=1} (\ -\ )^2} \tag{4}$$

$$Area\ Under\ the\ Curve\,(\text{AUC}) = \sum_{=1}^{-1} \left(\frac{++1}{2} \times (+1 -) \right) \tag{5}$$

For classification and regression tasks, an evaluation is performed. For multiclass classification, several performance metrics such as accuracy, precision, recall (sensitivity), specificity, and F score are used. The F score is calculated as a macro average for treating all classes in the same way. This matrix can be calculated by the following equation :

$$\text{Accuracy} = \frac{p+}{p+\ +p+} \tag{6}$$

$$\text{Recall\_0} = \frac{}{+p} \tag{7}$$

$$\text{Recall\_1} = \frac{p}{p+} \tag{8}$$

Where, tp, tn, fp, fn are true positive, true negative, false positive, and false negative values, respectively [17].

In E-Nose data analysis, the concepts of true positive, true negative, false positive, and false negative are crucial to evaluate the model's performance in classifying samples based on their volatile profiles. True positive represents a sample that is truly positive (i.e., contains a certain compound) and is successfully identified as positive by the model. Conversely, negative samples that are also correctly classified are called true negatives. A false positive occurs when the model incorrectly identifies a negative sample as positive, and a false negative occurs when the model incorrectly classifies a positive sample as negative. This parameter is often used to compare the performance of various classifications and data pre-processing in E-nose applications such as disease detection, product quality identification, and environmental analysis.

III. RESULTS AND DISCUSSION

Data collection of jengkol and kabau seeds aroma samples with E-Nose was carried out for 8 days. Each type of sample is done 1 time sampling data, so that 200 comma-separeted values (csv) are obtained.

Table 2. Total E-Nose data collected

Retrieval Method	**Description**	**Jengkol (0)**	**Kabau Seeds (1)**	**Total**
Change	Samples of jengkol and kabau seeds were taken alternately	100	100	200

Sampling data of jengkol (0) and kabau seeds (1) were collected using two methods as can be seen in Table 2. The first method was to separate the sampling of jengkol and kabau seeds into two sessions. In the morning, jengkol samples were collected and kabau seeds were collected in the afternoon. This separate retrieval method is used as initial data, which serves as training data. The second retrieval method is done by taking the sensor response of the jengkol and kabau seeds samples alternately. The data from this alternate retrieval method will be used as external test data. From both data collection methods, a total of 200 data were obtained.

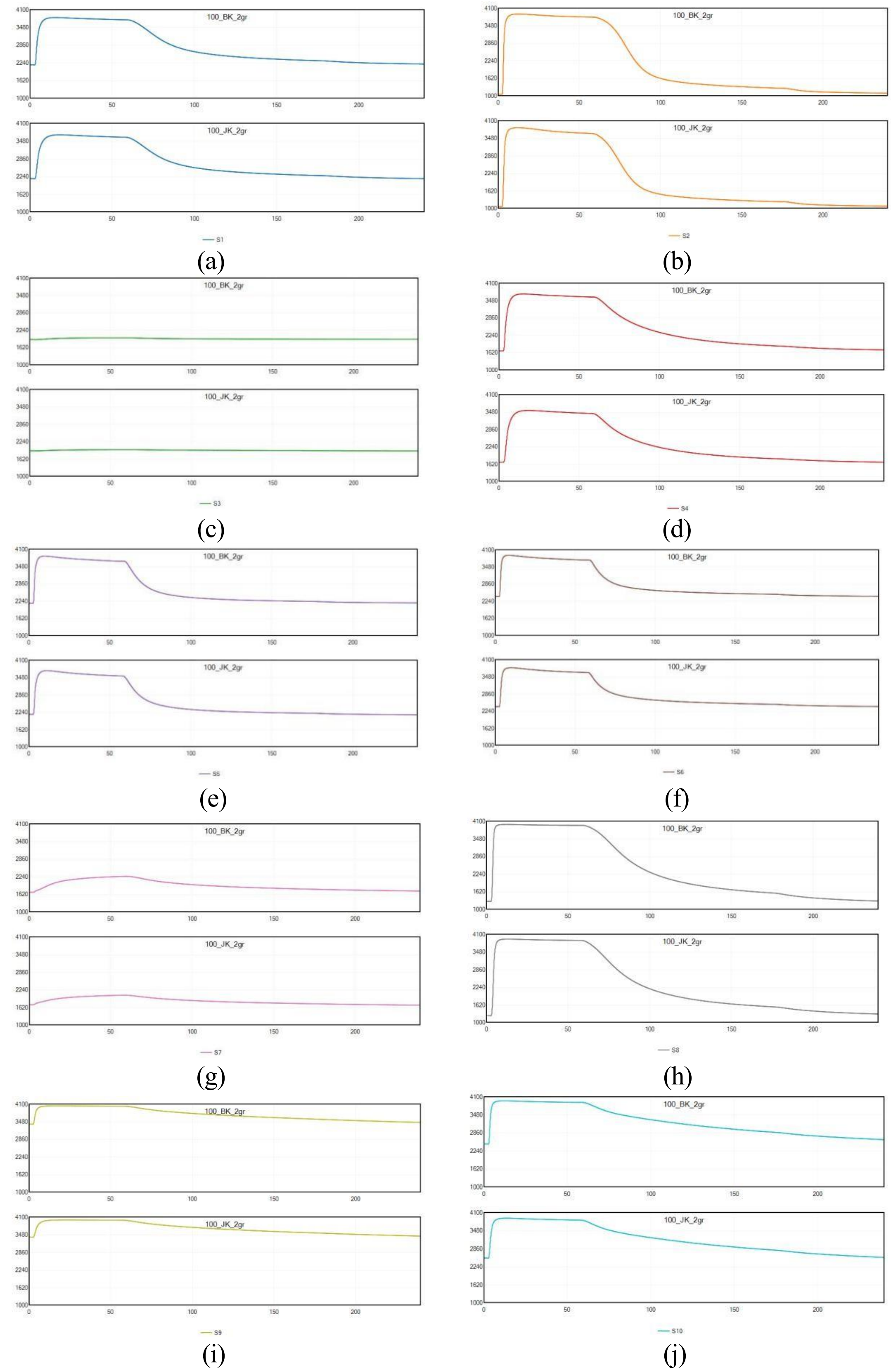

Fig 8. Response of each E-Nose sensor (a) S1 (b) S2 (c) S3 (d) S4 (e) S5 (f) S6 (g) S7 (h) S8 (i) S9 (j) S10

On each sensor, there is a maximum point reached, can be seen in Table 3 :

Table 3. Absolute value

Gas Sensor	Sensor Color Description on Data Logger	Absolut Value
S1	Dark Blue	3712.5
S2	Orange	3865.125
S3	Green	1959.25
S4	Red	3546.25
S5	Purple	3738.375
S6	Brown	3821
S7	Lilac	2048.875
S8	Grey	3943.875
S9	Dark Yellow	4013.625
S10	Sky Blue	3921.125

Electronic Nose (E-Nose) response data is taken using an E-Nose unit consisting of 7 gas sensors, temperature and humidity sensors. As for the sampling process, the response data is divided into 3 phases, namely delay, sampling, purging. The delay phase is set for 2 seconds, where in this phase the E-Nose system will try to clean the remaining environmental gas in the E-Nose system. In this delay phase it is also used as an environmental reference reference to shift the data response pattern, which aims to homogenize the initial value of the data response to 0 mV. The sampling phase is set for 173 seconds, where in this phase the response from the target sample will be measured. The purging phase is set for 65 seconds, which serves to clean the remaining gas from the target sample that was measured.

The Support Vector Machine (SVM) algorithm is a method in Machine Learning (ML) that is used in classifying between labels. Inter-label classification is performed using kernels to find the plane that can linearly separate the data distribution. The kernel serves to find the correlation between data points, where this information will be used by SVM in finding the best plane in separating the distribution between labels. In this research, SVM kernel variation was conducted to determine the best kernel that can separate jengkol (0) and kabau seeds (1).

Table 4. CV performance of SVM model with kernel variation using maximum value extraction feature

Kernel	Accuracy	Recall_0	Recall_1
Linear	98.50 ± 2.24	98.00 ± 4.47	99.00 ± 2.24
Polynomial	86.50 ± 3.79	100.00 ± 0.00	73.00 ± 7.58
RBF	99.00 ± 1.37	99.00 ± 2.24	99.00 ± 2.24
Sigmoid	99.00 ± 1.37	99.00 ± 2.24	99.00 ± 2.24

The performance of the model is measured using the cross-validation (CV) method which takes into account the accuracy and recall performance. The CV performance of the SVM model with kernel variations can be seen in Table 4. It can be seen that the average performance of the model can distinguish between jengkol and kabau seeds labels quite well. This is in line with the results of PCA, LDA and euclidean distance analysis. In addition, the deviation of the CV performance obtained also tends to be small. This small deviation indicates that the model performance tends to be stable. The best performance is obtained when using RBF and Sigmoid kernels.

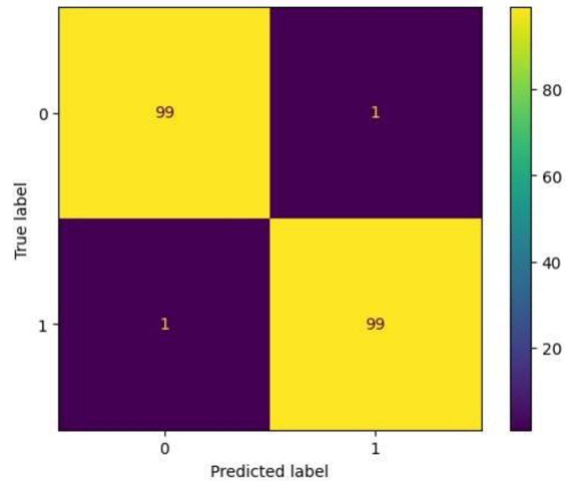

Fig 9. Confution matrices of SVM model with RBF kernel on external test data

External test data from alternate collection was used to test the developed model with separate collection data. This was done to simulate the conditions in the field, where jengkol and kabau seeds are not sampled separately. But it is done at one time. The performance of the external test data on the SVM model with RBF kernel can be seen in Figure 9. It can be seen that out of 200 data, there are 2 data that are misclassified. Where from the confution matrices, the accuracy is 99.00, Recall 0 is 99.00, and Recall 1 is 99.00. The research that has been conducted shows that the model that has been developed remains stable despite changes in the method and time of data collection.

The advantage of this method lies in its efficiency and ability to detect differences in volatile aromas that cannot be identified visually, especially after the seeds have been cut. This is particularly important given the physical similarities between jengkol and kabau and the potential for adulteration due to differences in the economic value of the two commodities, although to date no concrete evidence of adulteration has been found in the field. Compared to conventional methods such as Gas Chromatography-Mass Spectrometry (GC-MS), E-Nose offers a more economical and rapid solution without the need for complicated and expensive laboratory processes. In addition, the integration with machine learning enables automatic and accurate processing of aroma data, so it has the potential to be applied to the authentication of other food products that are difficult to distinguish visually but have distinctive volatile profiles.

Methodologically, this study also highlights the importance of extracting features from sensor data, such as mean, median, variance, standard deviation, and area under the curve (AUC), which are proven to improve the classification performance of the model. The use of SVM with RBF kernel was shown to be effective in managing complex volatile data and resulting in precise classification. Thus, this research not only provides a practical solution to the problem of authenticating jengkol and kabau, but also opens up opportunities for wider applications of machine learning-based E-Nose technology in the fields of food safety, quality control, and detection of counterfeit aroma-based products.

IV. CONCLUSION

Classification of jengkol and kabau seeds using E-Nose is done by varying several extraction features. Where the maximum value extraction feature shows the greatest separation value when compared to other feature extraction methods used. Kernels that function to find correlations between data are varied in developing SVM models. The kernel variations used are linear, polynomial, RBF, and Sigmoid kernels. Where SVM models with RBF and Sigmoid kernels provide the best CV performance when compared to linear and polynomial kernels. Meanwhile, to further test the stability of the model, the model that has been developed is then tested on external test data obtained from random data collection. The results of this external test show that the stability of the model performance can still be maintained, even though the retrieval method is changed and carried out in different weeks.

REFERENCES

[1] Rahman, I. N., Lelono, D., & Triyana, K. (2018). Klasifikasi Kakao Berbasis e-nose dengan Metode Neuro Fuzzy. IJEIS (Indonesian Journal of Electronics and Instrumentation Systems), 8(1), 49. https://doi.org/10.22146/ijeis.25512

[2] Harmoko, H., Samitra, D., & Rozi, Z. F. (2020). Pemberdayaan masyarakat Desa Taba Remanik Kecamatan Selangit Kabupaten Musi Rawas melalui pelatihan pembuatan kerupuk kabau. J-ABDIPAMAS (Jurnal Pengabdian Kepada Masyarakat), 4(1), 11. https://doi.org/10.30734/j-abdipamas.v4i1.896

[3] Fadhlurrahman, I. (n.d.). Indonesia Produksi 157,16 Ribu Ton Jengkol pada 2023. Retrieved from https://databoks.katadata.co.id/datapublish/2024/07/22/volume-produksi-jengkol-menurut-provinsi#:~:text=No.&text=Produksi jengkol Indonesia mencapai 157,%25) dalam 5 tahun terakhir.

[4] Siburian, R. (2018). Methyl gallate from jiringa (Archidendron jiringa) and antioxidant activity, (January), 11–16. https://doi.org/10.22159/ajpcr.2018.v11i1.21637

[5] Mainuddin, M., Kamal, M., Ahmed, A., Ferdous, G., Haque, H., & Khatun, A. (2021). Biocatalysis and Agricultural Biotechnology Chemical composition , bioactive compounds , antioxidants potential and mycotoxin of minor exotic archidendron pauciflorum fruit with the focus to Bangladesh. Biocatalysis and Agricultural Biotechnology, 34(June 2020), 102039. https://doi.org/10.1016/j.bcab.2021.102039

[6] Serunting, M. A., Zulfikar, M. A., Setyorini, D. A., Rizki, W. O. S., Kurniawan, R., & Setiyanto, H. (2024). Facile sunlight-irradiation mediated green synthesis of highly stable silver nanoparticles using Archidendron bubalinum pods extract for antibacterial activity application. Case Studies in Chemical and Environmental Engineering, 10(June), 100811. https://doi.org/10.1016/j.cscee.2024.100811

[7] Komariah, D., & Hartana, A. (2016). Variasi Morfologi Kabau (Archidendron Bubalinum) Dan Pe-Manfaatannya Di Sumatra. Floribunda, 5(5), 157–164

[8] Hardoyono, F. (2017). Penelitian Dan Pengembangan Sensor Aromatik Wewangian Untuk Autentikasi Produk Parfum Halal. Jurnal Penelitian Agama, 18(2), 302–322

[9] Zhang, E., Wang, F., Yin, C., Wang, H., Zhu, Y., Duan, H., … Hu, J. (2024). Food Bioscience Application of an electronic nose for the diagnosis of ketosis in dairy cows. Food Bioscience, 60(February), 104355. https://doi.org/10.1016/j.fbio.2024.104355

[10] Carleo, G., Cirac, I., Cranmer, K., Daudet, L., Schuld, M., Tishby, N., … Zdeborová, L. (2019). Machine learning and the physical sciences. Reviews of Modern Physics, 91(4), 45002. https://doi.org/10.1103/RevModPhys.91.045002

[11] Ye, Z., Liu, Y., & Li, Q. (2021). Recent progress in smart electronic nose technologies enabled with machine learning methods. Sensors. https://doi.org/10.3390/s21227620

[12] Putri, L. A., Rahman, I., Puspita, M., Hidayat, S. N., Dharmawan, A. B., Rianjanu, A., … Wasisto, H. S. (2023). Rapid analysis of meat floss origin using a supervised machine learning-based electronic nose towards food authentication. Npj Science of Food, 7(1), 1–15. https://doi.org/10.1038/s41538-023-00205-2

[13] Putra, Y. (2021). Perbandingan Karakteristik Kulit Babi Ternak dengan Kulit Sapi dan Kambing Menggunakan lnstrumen Electronic Nose Berbasis Chemometric untuk Jaminan Kualitas Halal (Doctoral dissertation, Universitas Negeri Padang)

[14] Inca, I., Widodo, T. W., & Lelono, D. (2018). Klasifikasi Teh Hijau dan Teh Hitam Tambi-Pagilaran dengan Metode Principal Component Analysis (PCA) Menggunakan E-Nose. IJEIS (Indonesian Journal of Electronics and Instrumentation Systems), 8(1), 61. https://doi.org/10.22146/ijeis.28718

[15] Aji Lintang, C., Wahyu Widodo, T., & Lelono, D. (2016). Rancang Bangun Electronic Nose untuk Mendeteksi Tingkat Kebusukan Ikan Air Tawar CORE View metadata, citation and similar papers at core.ac.uk provided by IJEIS (Indonesian Journal of Electronics and Instrumentation Systems). Ijeis, 6(2), 129–140.

[16] Kombo, K. O., Ihsan, N., Syahputra, T. S., Hidayat, S. N., Puspita, M., Wahyono, … Triyana, K. (2024). Enhancing classification rate of electronic nose system and piecewise feature extraction method to classify black tea with superior quality. Scientific African, 24, e02153. https://doi.org/10.1016/j.sciaf.2024.e02153

[17] Wijaya, D. R., Afianti, F., Arifianto, A., Rahmawati, D., & Kodogiannis, V. S. (2022). Ensemble machine learning approach for electronic nose signal processing. Sensing and Bio-Sensing Research, 36(March), 100495. https://doi.org/10.1016/j.sbsr.2022.100495