**Matrik: Jurnal Manajemen, Teknik Informatika, dan Rekayasa Komputer**
Vol. 24, No. 3, July 2025, pp. 381∼394
ISSN: 2476-9843, accredited by Kemenristekdikti, Decree No: 10/C/C3/DT.05.00/2025
DOI: 10.30812/matrik.v24i3.4291 

# Comparison of Random Forest Support Vector Machine and Passive Aggressive Models on E-nose-Based Aromatic Rice Classification

**Salima Nurrahma, Budi Sumanto**
Universitas Gadjah Mada,Yogyakarta, Indonesia

**Article Info**

***Article history:***

Received July 23, 2024
Revised February 13, 2025
Accepted March 21, 2025

***Keywords:***

*Aromatic Rice;*
*Electronic Nose;*
*Passive Aggressive;*
*Random Forest;*
*Support Vector Machine.*

***ABSTRACT***

Accurate classification of aromatic rice types is crucial for maintaining quality and meeting consumer preferences. **The purpose of this study** is to classify Mentik Wangi, Pandan Wangi, and C4 rice based on their volatile content using e-nose. C4 rice, as a popular non-aromatic variety, was included as a comparison for sensor response analysis. **The research method** involved preprocessing the e-nose gas sensor readings, including feature extraction, baseline manipulation, and missing value checking, to ensure data quality. The classification was performed using Random Forest, Support Vector Machine, and Passive-Aggressive methods. **The results** showed that the Random Forest model achieved the highest accuracy of 97%, followed by the Support Vector Machine at 95% and Passive Aggressive at 90%. The model evaluation utilized a Confusion Matrix and Receiver Operating Characteristics, which confirmed that Random Forest was the best-performing model. **This study concludes** that e-nose-based classification effectively differentiates between aromatic rice types, providing significant potential for objective and efficient quality assessment and offering valuable insights for further research in areas such as food technology, agricultural science, and chemical analysis.



***Corresponding Author:***

Budi Sumanto,
Departement of Electrical Engineering and Informatics,
Universitas Gadjah Mada, Yogyakarta, Indonesia,
Email: budi.sumanto@ugm.ac.id.

How to Cite:
B. Sumanto and S. Nurrahma, ”Comparison of Random Forest Support Vector Machine and Passive Aggressive Models on E-nose-Based Aromatic Rice Classification”, *MATRIK : Jurnal Manajemen, Teknik Informatika dan Rekayasa Komputer*, vol. 24, no. 3, pp. 381–394, doi: 10.30812/matrik.v24i3.4291.

## 1. INTRODUCTION

Rice aroma is an important indicator of quality that affects consumer preferences and market prices. Aromatic rice, with its coveted distinctive aroma, is valued higher than non-aromatic rice [1]. This unique aroma comes from volatile compounds, mainly 2-acetyl-1-pyrroline (2AP), which are formed during rice growth and influenced by environmental factors such as soil composition and sun exposure [1, 2].

Some previous studies that are in line with this research are conventional methods for assessing rice aroma, such as human sensory panels, which are subject to subjectivity limitations and are less efficient for large-scale analysis [3]. Therefore, an objective and fast method for rice aroma classification is required. Electronic nose (e-nose) technology offers a promising solution for this purpose [4]. The e-nose, with its sensors sensitive to various volatile compounds, can ”smell” the aroma of rice and generate quantitative data that can be further analyzed. This research focuses on the classification of three types of aromatic rice-Mentik Wangi, Pandan Wangi (Sedap Wangi), and C4 rice-using e-nose. Mentik Wangi and Pandan Wangi are known as premium aromatic rice varieties in Indonesia. In contrast, C4 rice, although not an aromatic variety, is included as a comparison to evaluate the ability of e-nose to distinguish aromas. This study aims to (1) explore the potential of e-nose in distinguishing the aroma of the three types of rice [5], (2) apply differential feature extraction and feature selection methods to optimize e-nose data [6], and (3) compare the performance of three classification models (Random Forest, Support Vector Machine, and Passive Aggressive) to identify the best model for classifying aromatic rice.

Previous research has utilized e-noses for rice quality classification [7]. However, most of these studies are limited to raw data analysis or rely on simple feature extraction methods such as Principal Component Analysis (PCA) [8, 9], which are often suboptimal in capturing the complexity of aromatic rice volatile data. In contrast to these approaches, this research goes a step further by integrating more advanced machine-learning classification methods [10], such as Random Forest and Support Vector Machines, which are known to excel at handling non-linear and high-dimensional data. The novelty of this study lies in the comprehensive comparison of three different classification models, Forest, Support Vector Machine, and Passive-Aggressive-specifically, to identify the most accurate and efficient model in aromatic rice classification. This approach allows us to evaluate the performance of each model in depth and determine the best model that can be applied objectively and efficiently in aromatic rice quality assessment.

This research aims to classify the aroma of aromatic rice using e-nose and machine learning. Adopting data acquisition techniques from previous research [11], the results showed that Random Forest achieved the best accuracy, outperforming other models. In conclusion, e-nose and Random Forest are effective in distinguishing aromatic rice types. This research contributes to agricultural technology with a more accurate and efficient classification of rice aroma. The implication is that this method can improve the quality of aromatic rice for farmers, producers, and certification bodies and provide insights for the development of e-nose in other agricultural fields.

## 2. RESEARCH METHOD

### 2.1. Research Materials

The materials used in this study consisted of three types of rice, namely Mentik Wangi, Pandan Wangi (Sedap Wangi), and C4, as shown in Table 1. Raw data from the electronic nose (e-nose) was obtained for each type of rice. An e-nose with 9 sensors was used to capture rice aroma readings in mV for 70 seconds, with an acquisition frequency of 10 Hz (every 0.1 seconds). Each data point is labeled according to the type of rice, i.e., Mentik Wangi (M), Pandan Wangi (Sedap Wangi) (S), or C4 (C). The dataset consists of 300 CSV files, with each file representing one measurement. The data then underwent preprocessing, including baseline manipulation using the differential method and feature extraction using the average value of the sensor readings. The preprocessed data is split into training data (80%) and testing data (20%). The training data is used to train the classification models Random Forest, Support Vector Machine, and Passive Aggressive, enabling the models to classify the test data accurately. The classification results, including accuracy, precision, and recall, are then evaluated to determine the effectiveness of each model in distinguishing the types of aromatic rice.

Table 1. Proportion and Type of Research Sample

| No | Types of Aromatic Rice | Symbol | Quantity |
|---|---|---|---|
| 1 | Mentik Wangi Rice | M | 20 sample @25 gram |
| 2 | Pandan Wangi Rice (Sedap Wangi) | S | 20 sample @25 gram |
| 3 | C4 Rice | C | 20 sample @25 gram |

### 2.2. Research Equipment

In this study, we used the following components. The sample container, made of sealed aluminum glass, serves as a container for the object to prevent the object's scent from spreading. The e-nose system consists of 9 types of gas sensors to detect the object's scent, along with the air suction and exhaust system inside. The laptop is used to store the data from the sensor readings and process the data to classify the types of rice. All these components are connected, as shown in Figure 1. A tube is attached to the lid of the aluminum container, which is connected to the airflow channel of the e-nose, with two other channels on the e-nose for exhaust and clean air as references. The e-nose sends data to the laptop through a connected USB cable.

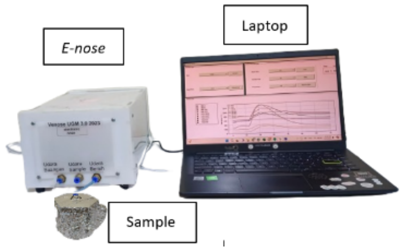


Figure 1. Research equipment

The electronic nose used in this study is arranged with a sensor array containing several sensors. The sensors used include MQ3, MQ9, MQ135, TGS822, TGS2600, TGS2611. The reading data of rice aroma by these sensors, after being obtained, will then be processed using a feature from Google, namely Google Colab. Metal oxide semiconductor (MOS) sensors have different sensitivities to certain gases [12] as described in Table 2.

Table 2. Sensors Spesification

| No | Sensors | Detected object |
|---|---|---|
| 1 | MQ-3 | Alcohol and benzena |
| 2 | MQ-9 | Carbon monoxide $(CO)$, LPG (propane and butane) methane |
| 3 | MQ-135 | Ammonia $(NH_3)$, nitrous oxide $(NO_x)$, benzena, alcohol |
| 4 | MQ-137 | Ammonia |
| 5 | TGS-813 | Carbon monoxide, ethanol, propane, butane, methane, and other combustible gases |
| 6 | TGS-822 | Ethanol, benzene, methane, isobutane, carbon monoxide |
| 7 | TGS-2600 | Carbon monoxide, cigarette smoke, ethanol, hydrogen, isobutane |
| 8 | TGS-2602 | Volatile Organic Compounds $(VOCs)$, Ammonia $(NH_3)$, thanol, and toluene |
| 9 | TGS-2611 | Methane, ethanol, isobutane, hydrogen |

### 2.3. Research stages

This research consists of several main stages, which include: (1) Sample Preparation, (2) Data acquisition, (3) Data preprocessing, (4) Feature extraction, (5) Data splitting, (6) Classification, and (7) Evaluation. In the sample preparation stage, 25 grams of rice from each type (Mentik Wangi, Pandan Wangi (Sedap Wangi), and C4) were placed in a glass sample container and heated for 3 minutes using an aluminum heater. This heating aims to increase the release of volatile compounds from the rice. Furthermore, at the data acquisition stage, the heated samples were tested using e-nose. The resulting data was stored in Comma Separated Value (.csv) format. The data preprocessing stage is performed with baseline correction using the differential method to remove noise and improve the signal. The feature extraction stage is performed to obtain the important features of the rice aroma data. In this study, the average value of each sensor reading is used as the feature. Then, the data is separated into training data (80%) and testing data (20%). The training data is used to train classification models (Random Forest, SVM, and Passive Aggressive). The trained model is then tested using test data to evaluate classification performance. The evaluation is based on accuracy, precision, and recall metrics.

### 2.4. Data Acquisition

The data acquisition process on the e-nose has three stages. The first stage of flushing serves to get a baseline signal by flowing clean air to the sensor. Baseline is used as a basic reference signal for taking data. The flushing stage runs in the first 10 seconds.

Furthermore, the sampling stage directs the sample gas into the chamber or sensor room, allowing the sensor to detect the volatile object. The sampling stage lasts 15 seconds after the flushing stage. The purging stage stops the gas flow from the sensor chamber to clean the aroma that is still left in the sensor chamber. This stage is carried out to obtain optimal sampling and recognizable object aroma patterns. The purging stage is carried out for 45 seconds. The results of the data acquisition will be stored in the system directory in a file with the type Comma Separated Value (.csv).

### 2.5. Baseline Manipulation

The sensor's initial results show a different reading of the detected aroma-this is caused by differences in the sensitivity of each sensor in reading the sample aroma [12]. The baseline manipulation process can manipulate the sensor's baseline value to all zero values. This research uses differential baseline manipulation by reducing the maximum voltage value obtained by the sensor with the minimum voltage value of the sensor, as in Equation 1.

$$V_{diff} = V_{max} - V_{min} \tag{1}$$

### 2.6. Feature Extraction

Feature extraction is the process of mapping original features to new features to produce better classification. The characteristics of the sensor response used in e-nose can be obtained by extracting sensor response information in the form of different patterns [13]. The statistical feature mean or average is used in this research. This feature refers to the average value of each column or each feature of the dataset. The mean value can estimate the value of the whole data [14].

### 2.7. Classification Models

The classification methods used are random forest, passive-aggressive, and support vector classifier. The classification process is carried out using all the collected data. It begins with an input dataset of 300 CSV files, which will be in the process of differential baseline manipulation and mean feature extraction. The result of feature extraction data is one CSV file. Will be split into training data (80%) and testing data (20%) [15]. Train data is used to train the classification method model so that it can hit test data and classify correctly. After that, testing the classification model with test data (test data) [16]. Based on this process, the classification results will be obtained in the form of accuracy, precision, and recall values for further analysis and evaluation. The first classification model is a Random Forest with a flowchart, as in Figure 2.

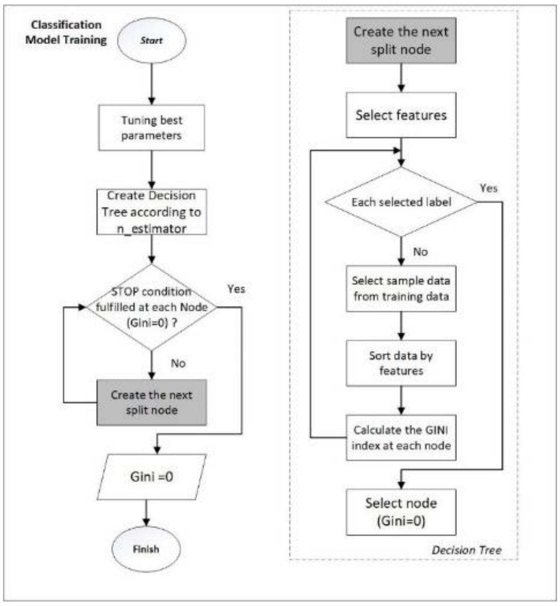


Figure 2. Flowchart random forest

The Random Forest model starts with the process of tuning the best Random Forest parameters against the research data and then creating a Decision Tree model and selecting a random subset of the training data. The Decision Tree continues the next process; the results of the process will be selected whether the node has a tree depth that has been fulfilled and the node has produced the smallest gini index value [17]. If not, it will be processed on the Decision Tree to find the next separator node, and if it has been fulfilled, the model prediction error value will be calculated, and the Random Forest process is complete.

Random Forest uses the Gini index as a separation criterion to create a decision tree in a random forest. This Gini value represents the value of data impurity at each node on the decision tree. If the Gini value is low, it means that the node is mutni or most of the data at the node belongs to one class. The $p_i$ value is the probability of a row in the dataset belonging to class $c_i$ as in Equation 2. Each data will go through the split data process so that the data will be D1 and D2.

$$Gini(D) = 1 - \sum_{i=1}^{m} pi^2 \tag{2}$$

The optimal node separation criterion is obtained by solving the minimum Gini value in $index_A$ (D) at each node. Then, the classification row of k decision trees will be obtained, and the predicted value of random forest using the most votes [18]. The second classification model used in this research is the Support Vector Machine (SVM). This classification model is widely used for regression and pattern classification problems [19]. Flowchart Figure 3. describes the classification workflow using a support vector machine, which starts with the process of tuning the best parameters of the support vector machine classification model then the raw data set or dataset is inputted to the system 19. Then based on the results of the best tuning parameters obtained, it will be processed to find the location of the data closest to the hyperplane and will calculate the margin between one sample data and other types of sample data. Training data is used for the classification process using a support vector machine.

$$K(x_i, x) = (\gamma(x_i^T x) + r)^p \tag{3}$$

In the parameter tuning process carried out, the best kernel type for this research data is the polynomial kernel type. The formula for the polynomial kernel type is shown in Equation 3. Kernel is a mathematical equation that projects data into a new feature space, and the data can be linearly separated using a hyperline. Polynomial Kernel works for cases that have non-linear data but still have certain patterns, such as circular patterns [20]. If the polynomial kernel has a polynomial level that is too low, it will cause the kernel to be unable to map the data to higher features. The classification model supports vector Machines with a flowchart, as shown in Figure 3.

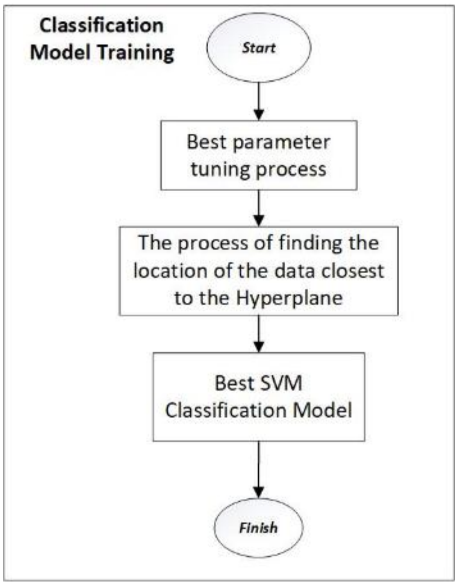


Figure 3. Flowchart support vector machine

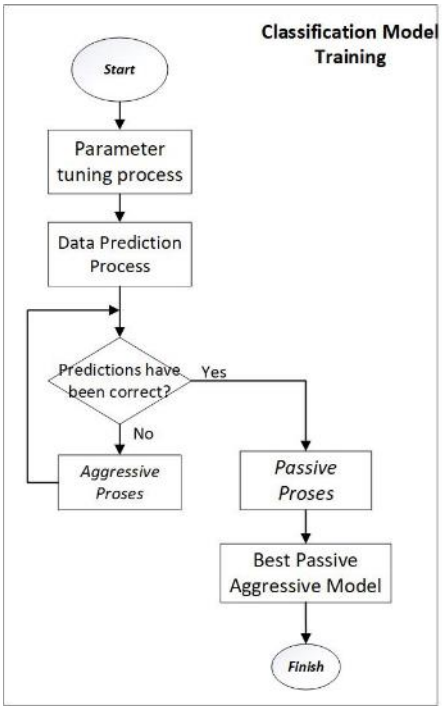


Figure 4. Flowchart passive aggressive

Passive Aggressive is the last classification model used in this research. Passive aggressive is a machine learning algorithm with an algorithm model that will be passive when the prediction is correct and will maintain the model[21]. Conversely, this model will be aggressive when the prediction is incorrect and will continue to make adjustments to produce the correct prediction. This type of model does not converge in contrast to other models. The passive aggressive model will increase its accuracy when there is a prediction error that continues to be corrected [22].

$$\bar{w}_{t+1} = \bar{w}_t + \frac{max(0.1 - y_t(\tilde{x}_t))}{||x_t||^2 + \dfrac{1}{2C}} y_t \bar{x}_t \tag{4}$$

Equation 4 describes how the model weights are updated every iteration based on the prediction results and the actual labels [23]. The components required in the calculation process are $\frac{w}{w}(t+1)$, the model weight vector value at iteration $t+1$, $x$, the input feature vector value at iteration $t$, $C$, the regulation parameter value, and $y$, the class label.

Based on Figure 4, the process starts with finding the best parameters in PA and then predicting data using dataset reading as raw data and splitting data to divide test data and training data. For training data, the $ypred_{PA}$ data prediction will be read, and then the prediction will be selected to determine whether it is correct. If not, the data will enter the aggressive process, where model changes will be made that will improve the model and will be checked again to determine whether the prediction is correct. If the prediction is correct, the model will enter the Passive process using the testing data, which will maintain the model and not make any changes. Next, the model accuracy value will be displayed and evaluated whether the model has worked well.

### 2.8. Evaluation

The process of evaluating the performance of the classification model in this study was carried out using two methods, namely, confusion matrix and receiver operating characteristic. Confusion matrix (CM) is often used in machine learning projects to see the performance of classification models in recognizing test data and sample data. Confusion Matrix provides detailed information about the errors made by the model and simplifies the process of analyzing model performance.

Another performance evaluation model is the receiver operating characteristic. Receiver Operating Characteristic (ROC) is a machine learning operation that serves to evaluate the performance of classification models in binary and multiclass cases [22]. The curve generated in the receiver operating characteristic process represents the performance of the classification model in distinguishing between object classes. The receiver operating characteristic curve connects coordinate points using ”1-specificity” (false positive rate) as the x-axis and ”sensitivity” as the y-axis for all boundary values measured from the test results.

## 3. RESULT AND ANALYSIS

### 3.1. Baseline Manipulation

The sensor readings in the previous process will be processed using baseline manipulation to correct the initial baseline of the sensor response [12], where the entire sensor's initial reading value is in a different range. The differential baseline manipulation used in this study restores the initial signal of the sensor response or equalizes the base point of each metal oxide sensor [24]. The differential method used in this study uses Equation 1.

Based on Figure 5, the baseline manipulation results represent a large sample of Mentik Wangi. The results of baseline manipulation show that the initial values of the nine sensors are the same relative to zero, showing a response pattern that can be identified. Each sample will be subjected to a baseline manipulation process so that the reading scale on the sensor response graph matches the reading value of each type of rice.

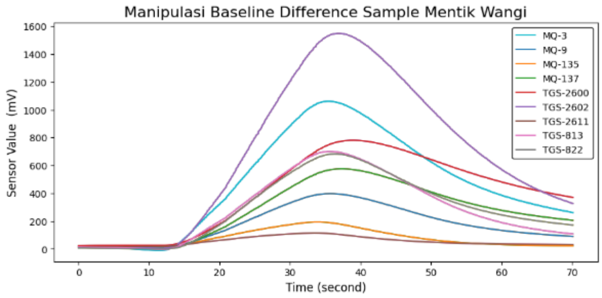


Figure 5. Baseline manipulation mentik wangi

### 3.2. Feature Extraction Result

Based on Figure 6, it can be seen that C4 rice has the highest value on the TGS-2602 sensor of 1500mV where this sensor has high sensitivity to detect Volatile Organic Compound (VOC) which is a volatile organic compound [25]. Mentik rice has a high value on the TGS-2602 sensor because this rice contains 2-acetyl-1-pyrroline (2AP), which forms a fragrant aroma typical of Mentik Wangi rice; this content is included in VOC. In addition, Mentik Wangi is superior to the MQ-3 sensor, which functions to detect alcohol, which means that in the C4 used as this object, there is contamination or undergoes a fermentation process. Finally, C4 rice excels at the TGS-813 sensor which is a sensor for organic compounds containing nitrogen in the form of proteins and amino acids which can be detected on this sensor.

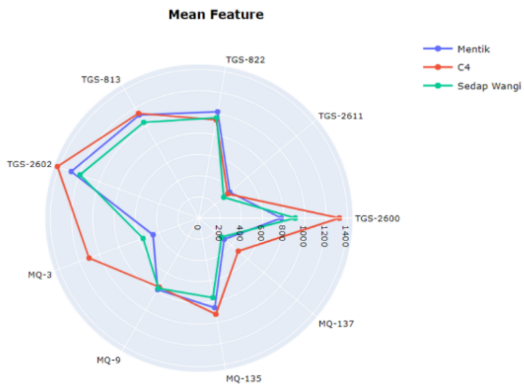


Figure 6. Baseline manipulation mentik wangi

### 3.3. Classification Model Results

The results of feature selection obtained the best mean feature for the classification process of this research. The data extracted from the mean feature is used as classification data in the classification model. This mean data is processed with each classification model used, resulting in individual readings from each model. Each model aims to classify the types of rice according to the sensor readings received, with each method pattern. The results of each classification model include accuracy, precision, recall, and F1-Score values, as shown in Table 3.

Table 3. Comparison of Three Model Results

| Model | Acuration | Presisi | Recall | F1-Score |
|---|---|---|---|---|
| RF | 96.67% | 96.75% | 96.67% | 96.67% |
| SVM | 95.00% | 94.99% | 95.00% | 94.96% |
| PA | 90.00% | 90.32% | 90.00% | 89.88% |

Based on the results of the three classification models above, the random forest model is the most superior model with the highest overall value. The high accuracy value indicates that the model can accurately predict both positive and negative values. A high precision value means that the classification model correctly predicts many positive values from the total positive values available. Meanwhile, a high recall or sensitivity means that the model can classify new data well.

The next process is the evaluation of the classification model to determine the performance of the model [26]. The first model evaluation uses the following confusion matrix results from three classification models. The selection of classification methods used, including Random Forest, Support Vector Machine, and Passive Aggressive, does not have a comparison with previous studies involving aromatic rice as the subject, as there are no prior studies related to this topic. Previous research using rice as the subject employed the PCA method to classify aromatic and non-aromatic rice, which is less relevant to this study. The researchers chose Random Forest, Support Vector Machine, and Passive Aggressive methods because these methods feature robust algorithms and are better suited for classifying aromatic rice, which exhibits significant similarities between types. These methods have proven to yield high accuracy in classifying aromatic rice types.

The results of Figure 7 show that the Y-axis column represents the actual data that the random forest model has carried out, while the X-axis represents the predicted data or original data that is input to the classification process. Based on the confusion matrix, the Random Forest model classified 60 test samples of aromatic rice into three categories: C4, Mentik Wangi, and Pandan Wangi. The model predicted 58 out of 60 samples correctly, resulting in a total accuracy of 96.67%. The number of samples that are true positive includes C4 rice. There are 19 data, Mentik Wangi as much as 19 data, and Pandan Wangi as much as 20 data. The total false positive value is zero from C and S (0 + 0), the false negative value is 1 from the result of C(1 + 0), and the true negative is 40 from M and S (19 + 1 + 0 + 20).

The last evaluation model used, the Receiver Operating Characteristic of random forest, is shown in Figure 8, which is a probability curve that shows how well the model classifies samples. The Area Under the Curve (AUC) value shows the performance value of the model in the classification of each class [25]. The best area under the curve value in class S shows that the random forest model has no classification prediction error in class S (fragrant savory). As for the other two classes, there is a slight error shown by the area under the curve value below class S.

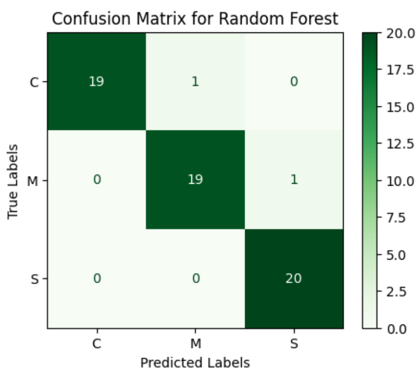


Figure 7. Confusion matrix random forest

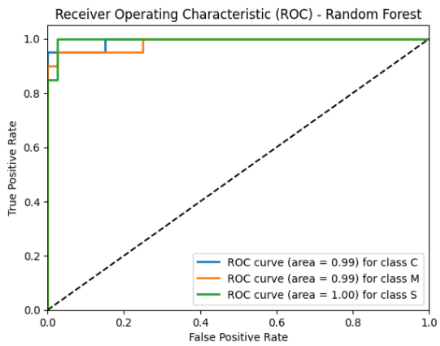


Figure 8. Receiver operating characteristic random forest

The confusion matrix of the support vector machine model is shown in Figure 9. The total data in the confusion matrix is 60, which is the value of the entire sample testing data. There are a total of 57 data in the true positive and true negative categories, namely 39 amounts of data predicted by the support vector machine. The number of each sample that is true positive includes C4 rice; there are 18 data, while the true negative is Mentik Wangi, as much as 19 data, and Pandan Wangi, as much as 20 data. The total false negative value is one from C (1 + 0), and the false positive value is one from the results of M and S (1 + 0).

From Figure 10, the result is obtained. Class S has the highest area under the curve value so that the support vector machine model can classify it very well. Class C has the second highest value; there is a classification error in that class. While class M has the smallest area under the curve value in this model, the support vector machine model produces the most misclassification of positive and negative samples in class M, characterized by the M graph having the largest number of serrations.

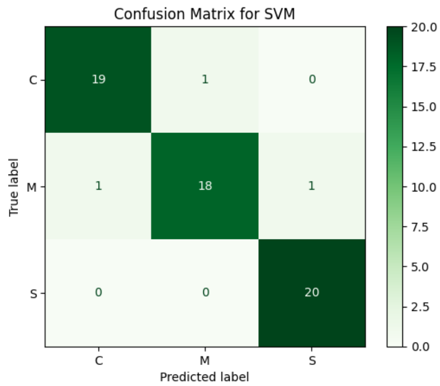


Figure 9. Confusion matrix support vector machine

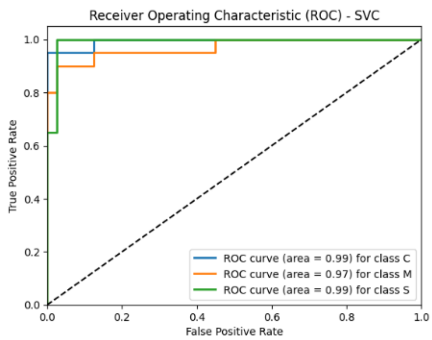


Figure 10. Receiver Operating Characteristic SVM

The last classification model, passive aggression, evaluated its performance in the classification process of each class. The confusion matrix of the passive-aggressive model is shown in Figure 11. The passive-aggressive model has a total true positive value of 17 data in class C, true negative, with 19+1 data in class M and 19 data in class S. In class, there are four False Positive values. Further evaluation using receiver operating characteristics was also carried out using the passive-aggressive model.

Figure 12 shows all the serrations, which means that the entire sample has misclassified the positive class or negative class. The largest area under the curve value of 0.98 is owned by class S, which is successfully classified best by the passive-aggressive model. Even though it has a marked misclassification, there are serrations on the S curve. The smallest area under the curve value is owned by class C, characterized by many serrations on the curve and a value of 0.92, where there is the most misclassification in this class. The smaller the area under the curve, the smaller the area under the curve value each class has and the lower the performance of the Passive Aggressive model in classification.

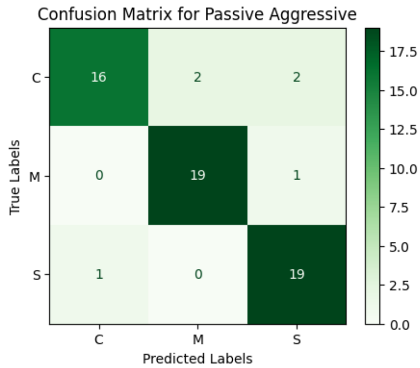


Figure 11. Confusion matrix passive aggressive

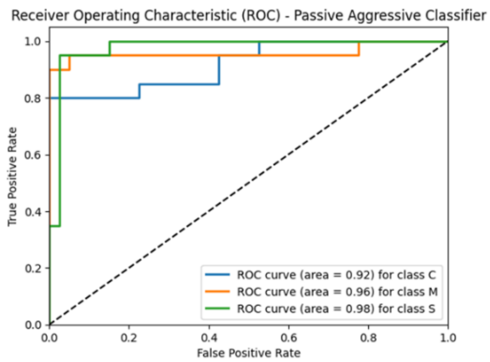


Figure 12. Receiver operating characteristic passive aggressive

The accuracy results of the Random Forest model reached 97%, Support Vector Machine achieved 95%, and Passive Aggressive reached 90%. Thus, it can be concluded that the Random Forest model is the best model. Additionally, validation of the comparison results was performed using the Receiver Operating Characteristic (ROC) to display the AUC probability curve, which represents the measure of separability and how well the model distinguishes between classes. The research on the classification of aromatic rice types using an electronic nose instrument and the application of machine learning has been optimally utilized in various research fields and rice production processes.

In the previous study by Aimi Aznan et al., 2022, [8] which aimed to recognize the quality of 17 types of rice using an electronic nose and machine learning with Artificial Neural Network (ANN) pattern recognition, the research resulted in the classification of rice types based on aroma, color, and texture as classification targets. Seven types of gas sensors were used. Based on two ANN models, the Bayesian regularization with 7 neurons achieved an accuracy of 95%, and the other model, Bayesian regularization with 10 neurons, also achieved an accuracy of 95%. Building on the research above, I have developed a new study for classifying aromatic rice types based solely on aroma, not related to other factors such as texture or color, using 9 types of gas sensors to capture more gas types and further improve the accuracy. The classification process in this study used three machine learning methods: Random Forest, Support Vector Machine, and Passive Aggressive. The accuracy results obtained from these classification models are superior at 97%, compared to the previous study, and they generate a Receiver Operating Characteristic (ROC) graph that is closer to the true positive value, indicating a higher sensitivity of the model in the classification process.

## 4. CONCLUSION

This study evaluated the performance of several classification methods in identifying aromatic rice types and found that Random Forest significantly outperformed the other methods. This superiority is shown by higher accuracy, precision, and recall values, as well as the lowest prediction error based on confusion matrix analysis. Validation using Receiver Operating Characteristic (ROC) curves resulted in near-ideal curves (1:1), confirming the ability of the Random Forest model to predict sample classes accurately. These findings recommend Random Forest as an effective method for aromatic rice type classification.

## 5. ACKNOWLEDGEMENTS

Thanks to the Department of Electrical Engineering and Informatics, Instrumentation and Control Engineering Technology Study Program and Instrumentation and Control Engineering Technology friends.

## 6. DECLARATIONS

AUTHOR CONTIBUTION

S.N. and B.S. shared contributions in conceptualization, methodology, validation, and interpretation of data. S.N.'s specific contributions include initial manuscript writing, machine learning computation, visualization, and laboratory experiments. B.S. was responsible for hardware and electronic development, manuscript revision, supervision, project leadership, and securing funding. All authors approved the final version of the manuscript.

FUNDING STATEMENT

The Department of Electrical Engineering and Informatics at Universitas Gadjah Mada funds this research.

COMPETING INTEREST

The authors declare that there are no conflicts of interest.

## REFERENCES


[1] Y. Xiong, X. Zheng, X. Tian, C. Wang, J. Chen, L. Zhou, D. Xu, J. Wang, V. Gilard, M. Wu, and A. You, "Comparative study of volatile organic compound profiles in aromatic and non-aromatic rice cultivars using HS-GC–IMS and their correlation with sensory evaluation," *LWT*, vol. 203, july, p. 116321, 2024, https://doi.org/10.1016/j.lwt.2024.116321.

[2] A. Akwero, D. Ocan, W. Akech, J. Lamo, M. Ochwo-Ssemakula, and P. Rubaihayo, "Allelic variations in aroma gene in cultivated rice varieties," *African Crop Science Journal*, vol. 28, no. 2, pp. 241–254, Jul. 2020, https://doi.org/10.4314/acsj.v28i2.9.

[3] X. Lu, X. Cai, and S. Liu, "Application of practical bionic electronic nose in rice wine detection," in *2ND International Conference on Frontiers Of Biological Sciences and Engineering (FSBE 2019)*, Jinan City, China, 2020, p. 020035, https://doi.org/10.1063/5.0000313.

[4] A. J. Moshayedi, A. Sohail Khan, J. Hu, A. Nawaz, and J. Zhu, "E-Nose-Driven Advancements in Ammonia Gas Detection: A Comprehensive Review from Traditional to Cutting-Edge Systems in Indoor to Outdoor Agriculture," *Sustainability*, vol. 15, no. 15, p. 11601, july 2023, https://doi.org/10.3390/su151511601.

[5] E. Kartikadarma, S. Wijayanti, S. A. Wulandari, and F. A. Rafrastara, "Principle Component Analysis for Classification of the Quality of Aromatic Rice," september 2020, https://doi.org/10.48550/arXiv.2009.06496.

[6] Z. Ye, Y. Liu, and Q. Li, "Recent Progress in Smart Electronic Nose Technologies Enabled with Machine Learning Methods," *Sensors*, vol. 21, no. 22, p. 7620, november 2021, https://doi.org/10.3390/s21227620.

[7] R. Ruslan, S. Khairunniza-Bejo, I. Rukunudin, M. Jahari, and M. Ibrahim, "Development of a machine vision system for rice seed inspection system," *Food Research*, vol. 4, no. S6, pp. 150–156, Dec. 2020, https://doi.org/10.26656/fr.2017.4(S6).049.

[8] A. Aznan, C. Gonzalez Viejo, A. Pang, and S. Fuentes, "Rapid Assessment of Rice Quality Traits Using Low-Cost Digital Technologies," *Foods*, vol. 11, no. 9, p. 1181, Apr. 2022, https://doi.org/10.3390/foods11091181.

[9] A. A. Munawar and Z. Sabaruddin, "Fast classification of rice (Oryza sativa) cultivars based on fragrance and environmental origins by means of near infrared spectroscopy," *IOP Conference Series: Earth and Environmental Science*, vol. 644, no. 1, p. 012003, january 2021, https://doi.org/10.1088/1755-1315/644/1/012003.

[10] M. H. Alsharif, A. H. Kelechi, K. Yahya, and S. A. Chaudhry, "Machine Learning Algorithms for Smart Data Analysis in Internet of Things Environment: Taxonomies and Research Trends," *Symmetry*, vol. 12, no. 1, p. 88, Jan. 2020, https://doi.org/10.3390/sym12010088.

[11] N. A. Isaac, I. Pikaar, and G. Biskos, "Metal oxide semiconducting nanomaterials for air quality gas sensors: operating principles, performance, and synthesis techniques," *Microchimica Acta*, vol. 189, no. 5, p. 196, 2022, https://doi.org/10.1007/s00604-022-05254-0.

[12] B. Sumanto, D. R. Java, W. Wijaya, and J. Hendry, "Seleksi Fitur Terhadap Performa Kinerja Sistem E-Nose untuk Klasifikasi Aroma Kopi Gayo," *MATRIK : Jurnal Manajemen, Teknik Informatika dan Rekayasa Komputer*, vol. 21, no. 2, pp. 429–438, march 2022, https://doi.org/10.30812/matrik.v21i2.1495.

[13] S. Dhanekar, "Smart and Intelligent E-nose for Sensitive and Selective Chemical Sensing Applications," in *Smart Sensors for Environmental and Medical Applications*, 1st ed., H. Hallil and H. Heidari, Eds. Wiley, april 2020, pp. 149–171, https://doi.org/10.1002/9781119587422.ch8.

[14] E. S. A. Nugraha, N. D. Resty, Y. I. T. Wijaya, and B. Sumanto, "Electronic Nose to Classify Tobacco Origin using a Naive Bayes Method based on Feature Selection," in *2023 International Conference on Computer Science, Information Technology and Engineering (ICCoSITE)*. Jakarta, Indonesia: IEEE, february 2023, pp. 267–272, https://doi.org/10.1109/ICCoSITE57641.2023.10127800.

[15] D. R. Prehanto, A. D. Indriyanti, I. K. D. Nuryana, and G. S. Permadi, "Classification based on K-Nearest Neighbor and Logistic Regression method of coffee using Electronic Nose," *IOP Conference Series: Materials Science and Engineering*, vol. 1098, no. 3, p. 032080, march 2021, https://doi.org/10.1088/1757-899X/1098/3/032080.

[16] 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," *International Journal of Intelligent Engineering and Systems*, vol. 13, no. 4, pp. 56–65, augustus 2020, https://doi.org/10.22266/ijies2020.0831.06.

[17] S. Tangirala, "Evaluating the Impact of GINI Index and Information Gain on Classification using Decision Tree Classifier Algorithm*," *International Journal of Advanced Computer Science and Applications*, vol. 11, no. 2, pp. 612–619, 2020, https://doi.org/10.14569/IJACSA.2020.0110277.

[18] N. Kunhare, R. Tiwari, and J. Dhar, "Particle swarm optimization and feature selection for intrusion detection system," *Sādhanā*, vol. 45, no. 1, p. 109, december 2020, https://doi.org/10.1007/s12046-020-1308-5.

[19] X. Kan and T. Li, "Slant Split Criterion Random Forests Classification Algorithm based on Soft Margin Hyperplane," in *2020 IEEE 9th Joint International Information Technology and Artificial Intelligence Conference (ITAIC)*. Chongqing, China: IEEE, december 2020, pp. 1332–1337, https://doi.org/10.1109/ITAIC49862.2020.9339073.

[20] B. Sumanto, I. Nathasya, W. Wijaya, and J. Hendry, "Aplikasi Electronic Nose Untuk Klasifikasi Kopi Aceh Gayo Berdasarkan Proses Pasca Panen," *Jurnal Teknologi dan Sistem Komputer*, vol. 11, no. 1, 2024, https://doi.org/10.14710/jtsiskom.2023.14421.

[21] P. Lemberger and I. Panico, “A primer on domain adaptation theory and applications,” *ArXiv*, no. M1, 2020.

[22] V. Shabani, A. Havolli, A. Maraj, and L. Fetahu, “Fake News Detection using Naive Bayes Classifier and Passive Aggressive Classifier,” in *2023 12th Mediterranean Conference on Embedded Computing (MECO)*. Budva, Montenegro: IEEE, june 2023, pp. 1–6, https://doi.org/10.1109/MECO58584.2023.10155036.

[23] R. Muzzammel and A. Raza, “A Support Vector Machine Learning-Based Protection Technique for MT-HVDC Systems,” *Energies*, vol. 13, no. 24, p. 6668, december 2020, https://doi.org/10.3390/en13246668.

[24] A. A. H. De Hond, E. W. Steyerberg, and B. Van Calster, “Interpreting area under the receiver operating characteristic curve,” *The Lancet Digital Health*, vol. 4, no. 12, pp. e853–e855, december 2022, https://doi.org/10.1016/S2589-7500(22)00188-1.

[25] B. Sumanto, Y. Idianingrum Tw, S. Humaira, R. Lestari Budiani, and M. Arrofiq, “E-nose Application With Chemometrics for Monitoring Kombucha Tea Fermentation Process,” *JST (Jurnal Sains dan Teknologi)*, vol. 12, no. 1, march 2023, https://doi.org/10.23887/jstundiksha.v12i1.50994.

[26] M. Riadi, A. M. Okasa, R. Sjahril, and M. Mahendradatta, “Volatile Compound Analysis of Aromatic Rice Mutant Lines Using HS-SPME/GC-MS,” *Jurnal Bioteknologi & Biosains Indonesia*, vol. 9, no. 2, pp. 257–267, 2022.

**[This page intentionally left blank.]**