Received: March 1, 2020. Revised: May 22, 2020.

International Journal of
Intelligent Engineering & Systems

http://www.inass.org/

Optimizing Machine Learning Parameters for Classifying the Sweetness of Pineapple Aroma Using Electronic Nose

Mhd Arief Hasan[1] **Riyanarto Sarno**[2]* **Shoffi Izza Sabilla**[2]

[1]*Informatics Study Program, Universitas Lancang Kuning, Pekanbaru 28265, Indonesia*
[2]*Department of Informatics, Faculty of Intelligent Electrical and Informatics Technology, Institut Teknologi Sepuluh Nopember (ITS), Surabaya 60111, Indonesia*
* Corresponding author's Email: riyanarto@if.its.ac.id

Abstract: Electronic nose (e-nose) has been widely used to distinguish various scents in food. The output of e-nose is a signal that can be identified, compared, and analyzed. However, many researchers use e-nose without using standardization tools, therefore e-nose is still often questioned for its validity. This paper proposes an electronic nose (e-nose) to classify the sweetness of pineapples. The standard sweetness levels are measured by using a Brix meter as a standardization tool. The e-nose consists of a series of gas sensors MQ Series which are connected to the Arduino micro-controller. The sweetness levels measured by the Brix meter are then ordered into low, medium, high sweetness groups. These sweetness groups are used as label ground-truth for the e-nose. Signal processing PCA and mother wavelet is employed to reduce noise from the e-nose signals. The signal processing methods obtain optimal parameters to find the characteristics of each signal. Machine learning methods were successfully carried out with optimized parameters for the classification of three levels of sweetness of pineapple. The best accuracy is 82% using KNN with 3 neighbors.

Keywords: E-nose, Classification, Pineapple, Statistical parameters, PCA, Threshold, Wavelet.

1. Introduction

An electronic nose (e-nose) has been widely used to distinguish various scents in food. The working principle of e-nose is to mimic the function of the human nose, in which there are several receptors to identify an aroma. These receptors will be replaced by sensors on the electronic nose. The electronic nose consists of several sets of gas sensors with different selectivity, a signal collection unit, and a series of pattern recognition software. E-nose has been widely used for research in the field of food [1] and fruits [2].

The output of e-nose is a signal that can be identified, compared, and analyzed. The e-nose, which consists of an array of non-selective chemical gas sensors, functions to capture and convert odors into electrical signals or sensor responses. The generated electrical signals are overlapping aromas forming compounds that are captured by the sensor array with each with a different sensitivity. Then, the electrical signals in the form of sensor responses form specific patterns for the type of scent that is captured. To be able to identify or classify the pattern of a sample object, it cannot be done just by looking directly at the pattern produced quantitatively, further analysis is needed using a pattern recognition machine. However, many researchers use e-nose without using standardization tools, therefore e-nose is still often questioned for its validity.

Pineapple, or ananas (Ananas comosus (L.) Merr.) It is a type of tropical plant originating from Brazil, Bolivia, and Paraguay. This plant belongs to the Bromeliaceae family. Smell the pineapple scent is one of the techniques to distinguish young and mature pineapple. A ripe pineapple is a pineapple whose bottom has a distinctive aroma of ripe, sweet, and fruity fruit. Whereas if a pineapple has an aroma like acid or something fermented like the smell of

Received: March 1, 2020. Revised: May 22, 2020.

vinegar, then this pineapple is classified as young pineapple. At present, to determine the level of fruit sweetness, people still rely on trained people. Therefore this study proposes to classify pineapples using E-Nose based on the sweetness level.

For the standard sweetness level is measured using a Brix meter as a standardization tool. Brix Meter is a tool used to determine sugar levels in food or fruit. This study uses a Brix meter to determine the sweetness level of each pineapple sample to be tested using E-nose.

The sweetness level measured by the Brix meter is then divided into 3 groups, namely low, medium, and high sweetness. These groups are used as ground-truth labels for e-nose. Signal processing using Principal Component Analysis (PCA) and wavelets aims to reduce the noise from signals generated by e-nose. PCA is a technique used to simplify data, by transforming data linearly to form a new coordinate system with maximum variance. Principal component analysis can be used to reduce the dimensions of data without significantly reducing the characteristics of the data. Meanwhile, Wavelet is one tool that can be used to analyze non-stationary signals. Wavelet analysis can be used to show temporary behavior on a signal. This Wavelet Transform Method can be used to filter data or improve the quality of data. The signal processing method is calculated to obtain the optimal parameters and find the characteristics of each signal.

The next process, the use of learning machines for the classification of three levels of sweetness pineapple. There are four methods used in this machine learning process, namely Support Vector Machine (SVM), k-nearest neighbor (KNN), Multilayer Perceptron (MLP), and Random Forest. Of the four Machine Learning methods, the highest accuracy value with optimal parameters will be used as the success level of this study.

2. E-Nose to determine the level of fruit sweetness.

E-Nose technology began to be created and developed in the early 1990s. A few years later, Benady and Simon in 1995 began to explore E-Nose research for the use of electronic noses to examine the use of E-Nose in the level of fruit maturity Then three years later Maul Et Early (1998) started to create a product for commercialization, also called the e-NOSE 4000 projects from Neotronics Scientific Inc., USA, which consists of 12 conductive polymers. This electronic nose has proven useful in predicting in a non-destructive way [2]. Many studies have used E-nose to measure fruit maturity using E-nose. In the research [3], E-nose is used to calculate the levels of OXIDE in ripe and young fruit. This study concluded that the resulting OXIDE levels could affect the young and ripe of fruit. Electronic noses can also distinguish odor differences between different fruit varieties such as apples [4]. The use of E-Nose allows complex aroma changes in the fruit to be monitored [5]. It is allowing a fresh pineapple to be very easily perishable (rotten) as it affects the results of the e-nose data that is produced. Another story in the research of [6] calculates the quality level of ripe olive oil. European regulations govern the parameters and methodology for determining the quality of olive oil, and this quality mainly depends on the condition of the olives after harvest. They succeeded in classifying with three methods of Naive Bayes (NB), Partial Least Squares Discriminant Analysis (PLS-DA) and Multilayer Perceptron (MLP) neural network. We determined the sweetness level of the pineapple with several sweetness levels measured using a Brix meter.

Previous studies have also carried out the classification process using E-nose for Java cocoa fruit [7]. The best predictive classification process is obtained by the E-nose - MLP- ANN procedure. However, this research only still discusses the odor that occurs. Not yet discussed the integration of the content of the fruit that affects the quality of the fruit itself. This research has not selected the best sensor in the feature selection process using E-Nose [8]. Instead, our research team tried to classify the pineapple we shared in the sweetness level process using E-nose and Brix Meter. Of course, this research is very useful for pineapple processing companies, where they require to choose the best pineapple in large quantities (container). We begin this basic research for broader development, namely the detection of pineapples in industrial capacity later.

3. Research methodology

3.1 Materials

In this research, the E-Nose used is an MQ Series. There are ten types of gas sensors used. These sensors are used to apprehend the difference from the aroma of pineapple. In communicating the data, a Universal Serial Bus (USB) port is connected to the Arduino microcontroller to transfer the data into a computer. The collection of gas sensors is inside a sample chamber of transparent glass. In Fig.

Received: March 1, 2020. Revised: May 22, 2020.

1, it is explained how the mechanism of the E-Nose tool system works. The pineapple was purchased from the same fruit store and was ripe at the same age as a sample. The pineapple was cut with the size of a dice and the weight of each sample of pineapple (dice) used was the same, specifically 100 grams.

The steps in sampling the data were as follows:

1. e-nose was turned on and warmed for 5 minutes;
2. before the sample was put into the sample room. The sample of pineapple was measured by sugar content using Brix meter;
3. the sample was put into the sample chamber;
4. the process of taking data took place where the USB port from the sample room was inserted into the computer. The data collection process lasted for approximately 10-15 minutes;
5. the sample room was cleaned of existing samples; this was managed to avoid the occurrence of residuals from gases that occur from previous data collection. This process took approximately 5 minutes. Therefore the sample chamber was released entirely from the smell of the existing gas.

3.2 E-nose system

The E-nose system developed in this study is shown in Fig. 1. Consisted of three components, which are data acquisition and transmission unit; sensor arrays and space units; power and gas supply unit. E-nose systems were always equipped with sensors that used gas sensors that were sensitive to aromatic compounds, nitrogen oxides, hydrogen, alkanes, methane, sulfur compounds, alcohol [9]–[11]. According to volatile compounds emitted by peaches in ripening procedures [12], volatile metabolites were released during the growth of microorganisms [13], [14], and sensor array selection in related literature [9]–[11], our E-nose sensor arrays were built with various sensors of eight semiconductor metal oxide sensors.

Table 1 explains the use of the number of sensors used on this e-nose. There were nine sensors that we use as described. The Gas Type column is the type of gas produced by the sensor. However, in this study, we did not discuss what type of sensors and what gases would dominate the data collection of pineapple using E-nose, as previous studies have done [15, 16].

Table 1. Sensors used and their gas compounds

Sensors	Gas Types
MQ 2	CH_4, Hydrogen (H2), Carbon Monoxide (CO), LPG, Alcohol, Propane, Smoke
MQ 4	CH_4, Natural Gas, Alcohol, Smoke
MQ 6	H2, Alcohol, LPG, Cooking Fumes
MQ 9	Coal Gas, CO, Liquefied Gas
MQ 135	NH_3, alcohol, NO_x, smoke, benzene, and CO_2
MQ 136	H_2S
MQ 137	CO, Ammonia
MQ 138	phenyl methane, ethanol, acetone, and methanol
DHT – 22	Temperature and Humidity

3.3 Brix meter

Brix Refractometer is used to measure sugar levels in fruits, fruit juices, coffee, soft drinks, and others. For fruit plants, refractometers help to determine when is the right time to harvest and classifying fruit based on the level of sweetness. This refractometer has a measurement range of: 0-32% Brix, Resolution: 0.2%, Working Temperature average in 10 ~ 30º C, Dimensions: 20.5 x 8 x 5.5 cm or 8 x 3.5 x 2.3 inch. How to use this tool is by using a pipette to take the liquid to be measured. Open the prism cover, and place 2-3 drops on the surface of the prism, then close it again. Point the prism to the light source, and observe from the side of the glass binoculars. Then Adjust the knob; therefore, the size is visible, its value is read based on the boundary between the blue and white areas. We use this to measure the sugar content in pineapples.

3.4 Stages of research

This research used several experimental techniques to produce the best quality data. Following Fig. 2 which explains the research flow are explained as follows:

1. pineapple data samples are measured using a Brix meter. Then proceed with the collection of E-Nose data for 15 minutes through Arduino with the output format in the form of Comma Separated Value (CSV). Resulting in E-Nose data of 518 Data Lines;
2. then a threshold is conducted to measure the sweetness of the existing pineapple. So that there are 3 classes of pineapple (Low, Medium, and High);
3. because the data generated in the calculation of pineapple time is not the same. So we decided to take lines 113 to 300 with the aim that data processing can be optimized;
4. then we reduce the noise that occurs in the data using the Mother wavelet method by using

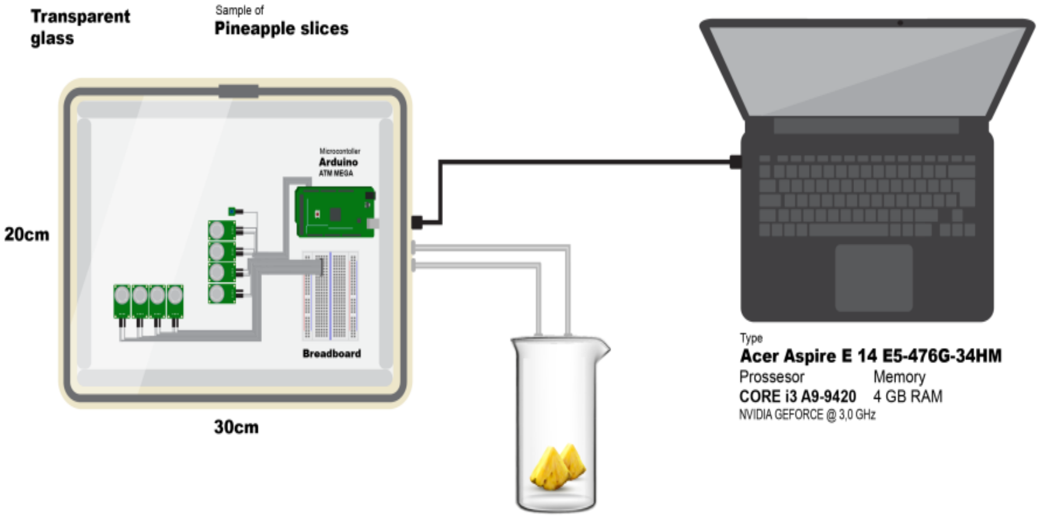

Figure. 1 E-nose system

several functions namely Sym5, db1, rbior1,1 using levels 1-10;

5. then we calculate the statistical parameters from the previous step (step 4). So we get 5400 features from these statistical parameters;
6. Then we did a feature selection using Pearson Correlation and Chi-Square;
7. After obtaining the features then enter the Principal Component Analysis (PCA) process. To simplify data, by transforming data linearly so that a new coordinate system with a maximum variant is formed. The PCA process is intended to determine the best components. The experiment continued with classification. This classification process uses four methods namely SVM, KNN, MLP, and Random Forest;
8. So from the 4 experimental methods obtained the best accuracy value. Obtained accuracy with optimal parameters.

3.4.1. Determination of the brix meter threshold value

In the initial stage, the value of the limit range of the sweetness of the pineapple is made. This calculation is done by determining the threshold value contained in the Brix meter. Brix meter value with a maximum value of 32%. This range produces three classes, mainly sweet, half sweet, and less sweet. The calculation uses the following formula. If based on the interval formula, the interval values are calculated wherein:

$$Reach\ (i) = largest\ \text{datum} - \text{datum} \quad (1)$$
$$= 22 - 5$$
$$= 7$$

Thus, the class interval length (C) is obtained.

$$C = \frac{Reach}{number\ of\ interval\ classess} \quad (2)$$

So as the following intervals and number of data are obtained, the result of the interval level shown in Table 2.

Table 2 explains the range of sweetness levels obtained after being calculated using a Brix meter. Then proceed with the threshold process. The interval is calculated using the manual Brix meter tool. In this range, there is no comma value, because the needle in the Brix meter is right in each row of numbers shown. Therefore, this study does not take into account the fraction of the range of sweetness of pineapple.

The data generated is in the form of CSV format with the amount following the explanation in Table 1. Each CSV produces 518 lines. Because during the data collection process there is a difference in the time difference between each turn on e-nose. We decided to retrieve data from row 113 to row 300. This is intended to get maximum results. Because these lines are lines of data retrieval directly from sensors that are on the E-Nose System tool described in Fig. 1.

Table 2. Pineapple sweetness interval level

Class	Sweetness Interval	Value	Number of data
1	0<X< 11	Low	19 Data
2	11-15	Medium	14 Data
3	> 15	High	28 Data

Received: March 1, 2020. Revised: May 22, 2020.

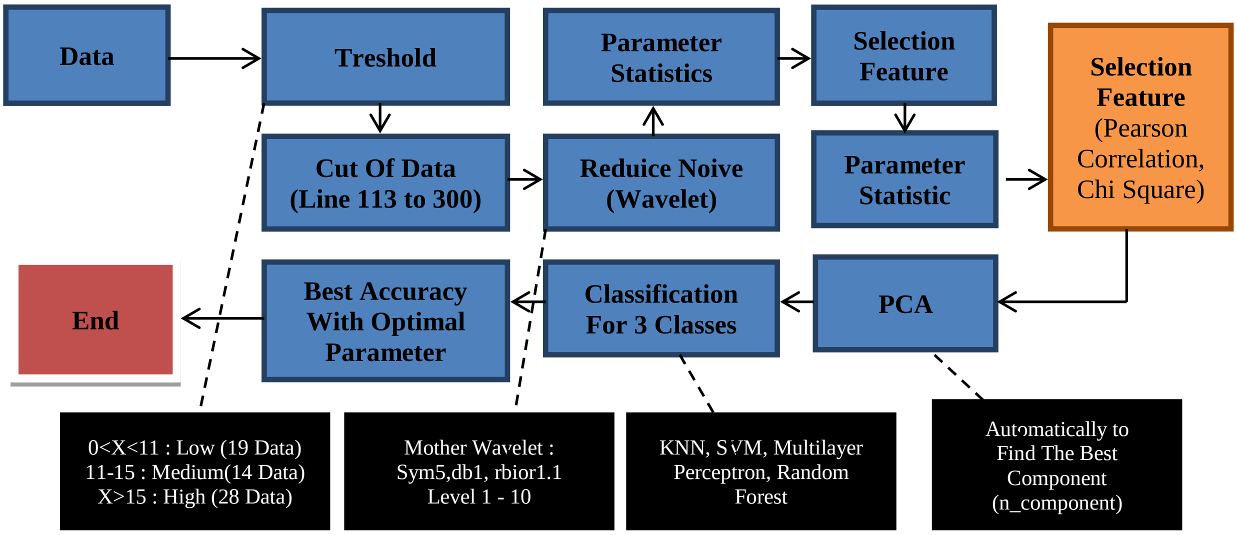

Figure. 2 Stages of research flow stages

3.4.2. Reduce noise (Wavelet) using mother wavelet

In general, the signal is susceptible to disturbances, such as interference from other gases entering the sensor, the stability of the electric current, which makes the sensor value unstable. These disturbances must need to be removed to proceed with further data analysis [17]. In the literature, many methods are mentioned to eliminate noise. In this study using Mother Wavelet to eliminate signal interference.

Therefore, the signal must be reconstructed using data preprocessing; this helps improve data quality, which will have an impact on the final result [18]. Pineapple data signal processing is carried out in two stages, particularly signal removal and data normalization. In experiments, pineapple signal obtains optimal parameters from discrete wavelet transforms (DWT) to improve signal reconstruction. The coefficient of DWT in the formula is initialized as dwt. The DWT of the original signal × (t) can be expressed with the following Eq. (3).

$$dwt(j.k) = \left(x(t), \psi j, k(t)\right) \tag{3}$$

$$\psi_{j,k(t)=\frac{1}{\sqrt{2^j}}\psi\left(\frac{t-k2^j}{2^j}\right)} \tag{4}$$

In this study using a mother wavelet with several types precisely: Sym5, db1, rbior1.1. This is intended to make a comparison of the resulting mother wavelet for comparison [19]. So that the maximum results are obtained.

3.4.3. Statistics parameters to determine the number of features

A statistical parameter is the characteristics of the measurement results of an object. The size of the statistical parameters is calculated from the sample data or population. Statistics parameters here are used to count the number of features available. The parameters used to use nine functions, specifically h5, n25, n75, n95, median, mean, std, var, and rms. From this parameter, it produces 5400 features. As for the explanation of the use of h5, n25, n75, n95, median, the mean is as follows in Table 3.

Table 3 explains the statistical parameter functions used. These statistical parameter functions are used in python using percentiles. Percentile is also called a point or value that divides the distribution of data into one hundred equal parts, therefore percentiles are often called "hundredths of measure". The point that divides data into one hundred equal parts.

Further explanation regarding N25, N75, and N95 and others.

```
n5    = np.nanpercentile(list_values, 5)
n25   = np.nanpercentile(list_values, 25)
n75   = np.nanpercentile(list_values, 75)
n95   = np.nanpercentile(list_values, 95)
```

N-Percentile is the lowest value that is equal to or greater than the N% of existing values.

Received: March 1, 2020. Revised: May 22, 2020.

Table 3. The Statistical Parameter function used

Function	Description
H5	The h5py package is a Pythonic interface to the HDF5 binary data format. HDF5 allows users to store large amounts of numerical data, and easily manipulate that data from NumPy.
N25	The binomial opportunity data distribution for N equals 25
N75	The binomial opportunity data distribution for N equals 75
N95	The binomial opportunity data distribution for N equals 95
Median	Used to find the middle value of data
Mean	The average value obtained from the sum of all values from each data, then divided by the amount of data available

3.4.4. Feature selection using pearson correlation and chi-square

Feature selection is an essential part of optimizing the performance of classification methods. The primary purpose of feature selection is to reduce complexity, increase accuracy, and choose optimal features from a data feature set [20]. Research conducted by Aytug Onan and Serdar Korukoglu discusses the comparison of several feature selection methods, the feature selection method using Pearson Correlation Coefficient can improve accuracy and has the advantage of being practical and fast for a massive number of features [21]. Following is the Pearson Correlation Coefficient equation shown in Eq. (5).

$$r_{xy} = \frac{n(\sum xy) - (\sum x)(y)}{\sqrt{[n\sum x^2 - (\sum x)2]}\ \sqrt{[n\sum y^2 - (\sum y)2]}} \tag{5}$$

where:

r_{xy} : Pearson's Product Moment Correlation Coefficient
n : Number of data pairs of x and y
$\sum x$: Total number of variables x
$\sum y$: Total number of variables y
$\sum x^2$: The square of the total number of variables x
$\sum y^2$: The square of the total number of variables y

Chi-Square feature selection uses statistical theory to test the independence of a term with its category [20]. One of the purposes of using this feature selection is to eliminate disruptive features in the classification. In the Chi-Square feature selection based on statistical theory, two of them are the appearance of the feature and the appearance of the category, which then each term value is ranked from the highest. Chi-Square test in statistics is applied to test the independence of two events.

$$X^2 = \sum \frac{(O-E)^2}{E} \tag{6}$$

where,

O : Observation value
E : Expected value (desire)

$$Df = (b-1)(k-1) \tag{7}$$

where :

b : number of rows
k : number of columns

3.4.5. Machine learning

One of the main topics in data mining or machine learning is classification. Classification is an activity to group data consisting of classes then labeled and targeted. The purpose of this algorithm is to solve the problem then create a classification category that is included in the process of supervised learning. The purpose of supervised learning is that label data or targets play a role as a 'supervisor' or 'teacher' who oversees the learning process in achieving a certain level of accuracy or precision. This classification process uses python because it has full features and has a platform that can be used both for research and for building production systems. This study uses four classification methods as follows:

a. **SVM (Support Vector Machine)** SVM is a machine learning method that carries out activities based on the rules of Structural Risk Minimization (SRM). SVM has a target to find the best hyperplane. The goal is to separate the two classes in the input space. The basic principle of SVM is a linear classifier and subsequently developed to work on non-linear problems. SVM tries to find the best hyperplane in the input space. The basic principle of SVM is a linear classifier and subsequently developed to work on non-linear problems. SVM has two parameters, which are C and gamma. This research creates an array for these parameters, as follows:

```
'model_C' : [0.0001; 0.001; 0.01; 0.1; 1; 10; 100]
'model_gamma' : [0.0001; 0.001; 0.01; 0.1; 1; 10; 100]
```

b. **KNN (*k-nearest neighbor*).** KNN classifies objects based on learning data that is the closest distance to the object. KNN classifies the projected learning data in multiple dimensions. This space is divided into sections that represent learning data criteria. Each learning data is represented as points c in many-dimensional spaces. The functions of KNN in Python are used as follows:

```
'model_n_neighbor' : [1; 2; 3; 4; 5; 6; 7; 8; 9; 10]
'model_weights' : [distance; uniform]
```

c. MLP Multilayer Perceptron (MLP) MLP is one type of classification method with an artificial neural network algorithm that adopts the workings of neural networks in living things. This algorithm is known to be reliable because of the learning process that can be carried out in a directed direction. Learning this algorithm is done by updating the back weight (backpropagation). Determination of the optimal weight will produce the right classification results. MLP consists of a simple system of interconnecting networks. The MLP functions that we use in Python are as follows :

```
'model_activation':[relu; logistic; tanh; relu]
 'model_hidden_layer_sizes':[500]
'model_max_iter':[10000]
```

d. Random Forest. This technique generates many decision tree classifications, then make decisions based on several decision trees that have been made. The advantage of using random forest is being able to classify data that has incomplete attributes and can be used for classification and regression. The Random Forest function used in python is as follows :

```
'model_n_estimator':[50; 100; 500; 1000]
```

3.4.6. Principal component analysis

Principal Component Analysis (PCA) is a method used to reduce the amount of data when a correlation occurs [22–24]. The aim is to find the essential part of which is a linear combination with an origin variable that explains each sensor. PCA projects a data matrix that initially has a high dimension to the lowest dimension (3 dimensions or two dimensions) without losing the necessary information. The correlation among samples can be visualized by a plot of each significant component [1, 25–30].

In this research, researchers want to find out the response pattern of a series of 10 gas sensors (electronic nose) toward the odor pattern on pineapple. Data processing methods are used to determine its classification using the PCA method. This technique is widely used to reduce the number of dimensions in a data set as it only uses the components that most contribute to tasks such as classification or regression in Machine Learning. The introduction of this pattern is expected to simplify testing the difference in the sweetness level of pineapple based on the aroma of pineapple.

3.4.7. Accuracy

Accuracy is one of the metrics for evaluating classification models. Informally, accuracy is the prediction fraction of our correct model. Formally, accuracy has the following definition :.

$$\text{Accuracy} = \frac{The\ Number\ of\ Predictions\ is\ Correct}{The\ total\ amount\ is\ correct}$$

For binary classification, accuracy can also be calculated in terms of positive and negative as follows:

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN}$$

Where,
TP = True Positive
TN = True Negative
FP = False Positive
FN = False Negative

4. Result and discussion

As described in Fig. 2, researchers conducted several models of machine learning testing experiments to determine the best results for data accuracy. In discussing these results we present the experiments that we have carried out and what the accuracy values obtained from these experiments. The experiments we use include processing data by classification, wavelets, PCA, Selection Feature. The following is a complete explanation of the results of the experiment and the results obtained for accuracy.

4.1 Data-classifier experiments

The first experiment conducted by us was to process pineapple data from E-Nose by using the SVM classification method. By using python we do SVM Classification so that the result of Parameter C is 100, The Gamma model is 0.001 with an accuracy of 0.770. This accuracy value does not meet the ideal classification value above 0.8. Therefore we do other clarification methods.

4.2 Data-wavelet-classifier experiments

In the second experiment is to process E-Nose data from pineapple earlier by denoising the signal obtained using wavelet. As explained in the previous experimental method in Fig. 2. Wavelet used is the mother wavelet Sym 5, db1, record 1.1 with levels 1-10. After Denoising, classification was continued with 4 methods namely SVM, KNN, MLP, Random Forest. In this experiment obtained better classification parameters than previous experiments. The parameter results obtained are as follows:

a. Classification using SVM obtained values obtained parameter C is 10, Gamma Model is 0.0001 with an accuracy of 0.787.
b. Classification using KNN obtained value obtained results 'model n_neighbors' is 6, 'model__weights' is ‘distance' with an accuracy of 0.656.
c. Classification using MLP obtained value obtained results ‘activation model’ is ‘relu’, ‘model__hidden_layer_sizes’ is 500, max_iter model is 1000 with accuracy is 0.705.
d. Classification using Random Forest obtained the value of the 'n_estimators model is 1000 with an accuracy of 0.738.

From this second step of experimentation, we have not to get maximum results. The accuracy value obtained is still far lower than the results of the previous stage.

4.3 Data-wavelet-PCA-classifier experiments

Then we proceed with the next experiment. Before doing the classification, we performed the PCA process, after the wavelet was obtained from the pineapple data E-Nose signal. The PCA process aims to carry out statistical processes on data to simplify existing data, By way of transforming data linearly so that a new coordinate system is formed with maximum variance. After conducting this experiment which ended with the classification process, obtained better results with the following details,

a. Classification using SVM model C is equal to 1, the gamma model is equal to 0.001, reduce_dim n_components is equal to 20 with an accuracy value of 0.803.
b. Classification using KNN obtained Parameter model ‘n_neighbors’ equals 1, ‘model__weights’ equals distance, reduce_dim n_components equal to 10 with an accuracy value of 0.689.
c. Classification using MLP obtained parameter activation model is equal to relu, model__hidden_layer_sizes is equal to 500, max_iter model is equal to 10000, reduce_dim n_components is equal to 20 with an accuracy value of 0.672.
d. Classification using Random Forest obtained parameter model n_estimators equal to 500 with an accuracy value of 0.770.

From the results of this test, it was proven that the PCA process was able to improve the accuracy of classification testing. The best value obtained from the results of the classification using this SVM method is 0.803.

4.4 Data-wavelet-selection feature-classifier experiments

In the fourth test, we replaced the third test, replacing the PCA process with feature selection. Feature selection is useful for removing data that does not correlate So that it will improve the accuracy of the data in making predictions. After the wavelet data is obtained then a Feature selection is done using Python. After the wavelet data is obtained then a Feature selection is done using Python:

a. The classification using the SVM method shows that the parameter value of model C is equal to 10, the gamma model is equal to 0.0001 with an accuracy value of 0.803.
b. Classification using the KNN Method obtained parameters with the n_neighbors model is 6, model__weights is 'distance' with an accuracy value equal to 0.770
c. Classification with MLP is obtained The activation model parameter is relu model__hidden_layer_sizes is equal to 500, max_iter model is equal to 10000 with an accuracy value of 0.705.
d. Random Forest classification obtained the results with the parameter model n_estimators is 500 with an accuracy value of 0.770.

4.5 Data-wavelet-selection feature-PCA-classifier experiments

In this final test, we performed all stages from steps 1 to four. Where after the data is processed with wavelet then performed Feature Selection the PCA process. Finally, the classification process is carried out. The results of this fifth test turned out to produce the best results from the previous test, the results of the test classification are as follows:

Received: March 1, 2020. Revised: May 22, 2020.

Table 3. Experimental results with a few steps classification experiments

Experiment	Classifier	Optimal parameter	Accuracy
Data – **Classifier**	SVM	With Prarameter {'model_C': 100, 'model_gamma': 0.001}	0.770
Data – **Wavelet** – Classifier	SVM	With Prarameter {'model__C': 10, 'model__gamma': 0.0001}	0.787
	KNN	With Prarameter {'model__n_neighbors': 6, 'model__weights': 'distance'}	0.656
	MLP	With Prarameter {'model__activation': 'relu', 'model__hidden_layer_sizes': 500, 'model max_iter': 10000}	0.705
	RandomForest	With Prarameter {'model__n_estimators': 1000}	0.738
Data – Wavelet– **PCA** – Classifier	SVM	With Prarameter {'model__C': 1, 'model__gamma': 0.001, 'reduce_dim__n_components': 20}	0.803
	KNN	With Prarameter {'model__n_neighbors': 1, 'model__weights': 'distance', 'reduce_dim__n_components': 10}	0.688
	MLP	With Prarameter {'model__activation': 'relu', 'model__hidden_layer_sizes': 500, 'model__max_iter': 10000, 'reduce_dim__n_components': 20}	0.672
	RandomForest	With Prarameter {'model__n_estimators': 50, 'reduce_dim n_components': 10}	0.787
Data – Wavelet– **Selection Feature** – Classifier	SVM	With Prarameter {'model__C': 10, 'model__gamma': 0.0001}	0.803
	KNN	With Prarameter {'model__n_neighbors': 6, 'model__weights': 'distance'}	0.770
	MLP	With Prarameter {'model activation': 'relu', 'model__hidden_layer_sizes': 500, 'model max_iter': 10000}	0.705
	RandomForest	With Prarameter {'model__n_estimators': 500}	0.770
Data - Wavelet – Selection Feature – **PCA** – Classifier	MLP	With Prarameter {'model_ _activation': 'relu', 'model hidden_layer_sizes': 500, 'model__max_iter': 10000, 'reduce_dim__n_components': 10}	0.738
	KNN	With Prarameter {'model__n_neighbors': 3, 'model__weights': 'uniform', 'reduce_dim__n_components': 10}	0.820
	SVM	With Prarameter {'model__C': 1, 'model__gamma': 0.001, 'reduce_dim__n_components': 10}	0.803
	RandomForest	With Prarameter {'model__n_estimators': 100, 'reduce_dim n_components': 30}	0.787

a. KNN classification with Parameter n_neighbors model equal to 3, model _weights equal to uniform, reduce_dim n_components equal to 10 with an accuracy of 0.820.

b. Classification SVM resulted for C and gamma parameters are 1 and 0.001 respectively. The reduction dimension using 10 n_component with accuracy is 0.803.

c. Random Forest Classification with Parameter model n_estimators is 100, reduce_sum n_components is 30 with accuracy results is 0.787.

The best accuracy is obtained by KNN Classification with an accuracy value of 0.820, It also provides the same thing as previous research that makes KNN as the best Machine Learning in the classification process [31]. Overall the results of this test are conveyed in Table 3. Experimental results with several steps of experimentation and classification.

5. Conclusion

From the results of several experiments in this study, we have managed to find the best accuracy in processing pineapple data using E-nose and Brix meters.We have processed the E-Nose data into three classes which are divided based on the threshold of the sweetness of pineapple. The best accuracy results we got with a value of 0.820. The accuracy results obtained by using the classification using K-Nearest neighbor (KNN). In this study, KNN is the best method for obtaining accuracy results from machine learning. These results are preceded by a wavelet process and then a selection feature.

In the next research, we will do a regression process from the E-Nose and Brixmeter data that we have found. This is intended to be able to see the effect between two or more variables. The variable relationship in question is functional which is realized in the form of a mathematical model. Then later from the results of this regression we will get a machine learning pattern linking historical data and labels or outputs that are interrelated, not alone.

Conflicts of Interest

The authors declare that there is no conflict of interest.

Author Contributions

RS is supervision that has the idea to classify the sweetness of pineapple using E-nose. He gave an idea to take Pineapple sample data using E-Nose and provide a threshold on the sweetness level of pineapple using Brix Meter. SIS gives a role to E-Nose development, she also calculating the threshold and classification process E-Nose and Brixmeter data using Python. MAH gives a role to the completion of this article so that this article can be written accordingly with the problem of this research.

Acknowledgments

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

References

[1] A. Loutfi, S. Coradeschi, G. K. Mani, P. Shankar, and J. B. B. Rayappan, “Electronic noses for food quality: A review”, *J. Food Eng.*, Vol. 144, pp. 103–111, 2015.

[2] J. Brezmes and E. Llobet, *Electronic Noses for Monitoring the Quality of Fruit*. Elsevier Inc., 2016.

[3] M. Mallick, S. Minhaz Hossain, and J. Das, “Graphene Oxide Based Fruit Ripeness Sensing e-Nose”, *Mater. Today Proc.*, Vol. 5, No. 3, pp. 9866–9870, 2018.

[4] D. Zhu, X. Ren, L. Wei, X. Cao, Y. Ge, H. Liu, and J. Li “Collaborative analysis on difference of apple fruits flavour using electronic nose and electronic tongue”, *Sci. Hortic. (Amsterdam).*, Vol. 260, No. August 2019, p. 108879, 2020.

[5] L. Torri, N. Sinelli, and S. Limbo, “Shelf life evaluation of fresh-cut pineapple by using an electronic nose”, *Postharvest Biol. Technol.*, Vol. 56, No. 3, pp. 239–245, 2010.

[6] D. M. Martínez Gila, J. Gámez García, A. Bellincontro, F. Mencarelli, and J. Gómez Ortega, “Fast tool based on electronic nose to predict olive fruit quality after harvest,” *Postharvest Biol. Technol.*, Vol. 160, Vo. November 2019, p. 111058, 2020.

[7] S. N. Hidayat, A. Rusman, T. Julian, K. Triyana, A. C. A. Veloso, and A. M. Peres, “Electronic nose coupled with linear and nonlinear supervised learning methods for rapid discriminating quality grades of superior java cocoa beans”, *International Journal of Intelligent Engineering and Systems*, Vol. 12, No. 6, pp. 167–176, 2019.

[8] S. Sabilla, R. Sarno, and K. Triyana, “Optimizing Threshold using Pearson Correlation for Selecting Features of Electronic Nose Signals”, *International Journal of Intelligent Engineering and Systems*, Vol. 12, No. 6, pp. 81–90, 2019.

[9] A. Hernández Gómez, J. Wang, G. Hu, and A. García Pereira, “Discrimination of storage shelf-life for mandarin by electronic nose technique”, *LWT - Food Sci. Technol.*, Vol. 40, No. 4, pp. 681–689, 2007.

[10] H. Zhang and J. Wang, “Evaluation of peach quality attribute using an electronic nose”, *Sensors Mater.*, Vol. 21, No. 8, pp. 419–431, 2009.

[11] S. Benedetti, S. Buratti, A. Spinardi, S. Mannino, and I. Mignani, “Electronic nose as a non-destructive tool to characterise peach cultivars and to monitor their ripening stage during shelf-life”, *Postharvest Biol. Technol.*, Vol. 47, No. 2, pp. 181–188, 2008.

[12] N. Narain, T. C. -. Hseih, and C. E. Johnson, “Dynamic Headspace Concentration and Gas Chromatography of Volatile Flavor Components in Peach”, *J. Food Sci.*, Vol. 55, No. 5, pp. 1303–1307, 1990.

[13] R. Article, A. Sani, A. A. Aliero, and S. E. Yakubu, “Volatile metabolites profiling to discriminate diseases of tomato fruits inoculated with three toxigenic fungal

pathogens", *Mol. Plant Pathol.*, Vol. 2, No. 3, pp. 14–22, 2011.

[14] A. D. Ibrahim, A. Sani, S. B. Manga, A. A. Aliero, R. U. Joseph, S. E. Yakubu, and H. Ibafidon, "Microorganisms Associated with Volatile Metabolites Production in Soft Rot Disease of Sweet Pepper Fruits (Tattase)", *Int. J. Biotechnol. Biochem.*, Vol. 7, No. 6, pp. 25–35, 2011.

[15] S. I. Sabilla, R. Sarno, and J. Siswantoro, "Estimating Gas Concentration using Artificial Neural Network for Electronic Nose", *Procedia Comput. Sci.*, Vol. 124, pp. 181–188, 2017.

[16] 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, pp. 659–671, 2016.

[17] R. Patil, "Noise Reduction using Wavelet Transform and Singular Vector Decomposition", *Procedia Comput. Sci.*, Vol. 54, pp. 849–853, 2015.

[18] L. Sharma and N. Mehta, "Data Mining Techniques: A Tool For Knowledge Management System In Agriculture", *Int. J. Sci. Technol. Res.*, Vol. 1, No. 5, pp. 67–73, 2012.

[19] M. Sushama, G. Tulasi Ram Das, and A. Jaya Laxmi, "Detection of high-impedance faults in transmission lines using Wavelet Transform", *J. Eng. Appl. Sci.*, Vol. 4, No. 3, pp. 6–12, 2009.

[20] A. Sharma and S. Dey, "Performance Investigation of Feature Selection Methods and Sentiment Lexicons for Sentiment Analysis", *Int. J. Comput. Appl.*, No. June, pp. 15–20, 2012.

[21] M. Shardlow, "An Analysis of Feature Selection Techniques", *Univ. Manchester*, No. 1, pp. 1–7, 2016.

[22] I. Tazi, K. Triyana, and D. Siswanta, "A novel Arduino Mega 2560 microcontroller-based electronic tongue for dairy product classification", *AIP Conf. Proc.*, Vol. 1755, No. July, 2016.

[23] I. Tazi, A. Choiriyah, D. Siswanta, and K. Triyana, "Detection of taste change of bovine and goat milk in room ambient using electronic tongue", *Indones. J. Chem.*, Vol. 17, No. 3, pp. 422–430, 2017.

[24] I. Tazi, K. Triyana, D. Siswanta, A. C. A. Veloso, A. M. Peres, and L. G. Dias, "Dairy products discrimination according to the milk type using an electrochemical multisensor device coupled with chemometric tools", *J. Food Meas. Charact.*, Vol. 12, No. 4, pp. 2385–2393, 2018.

[25] C. Li, P. Heinemann, and R. Sherry, "Neural network and Bayesian network fusion models to fuse electronic nose and surface acoustic wave sensor data for apple defect detection", *Sensors Actuators, B Chem.*, Vol. 125, No. 1, pp. 301–310, 2007.

[26] B. Wang, S. Xu, and D. W. Sun, "Application of the electronic nose to the identification of different milk flavorings", *Food Res. Int.*, Vol. 43, No. 1, pp. 255–262, 2010.

[27] Y. B. Che Man, A. Rohman, and T. S. T. Mansor, "Differentiation of lard from other edible fats and oils by means of Fourier transform infrared spectroscopy and chemometrics", *JAOCS, J. Am. Oil Chem. Soc.*, Vol. 88, No. 2, pp. 187–192, 2011.

[28] Z. Haddia, A. Amaria,b, A. Ould. Alia, N. El Baric, H. Barhoumid, A. Maarefd, N. Jaffrezic-Renaulte, and B. Bouchikhia, "Discrimination and identification of geographical origin virgin olive oil by an e-nose based on MOS sensors and pattern recognition techniques", *Procedia Eng.*, Vol. 25, pp. 1137–1140, 2011.

[29] M. Nurjuliana, Y. B. Che Man, and D. Mat Hashim, "Analysis of lard's aroma by an electronic nose for rapid Halal authentication", *JAOCS, J. Am. Oil Chem. Soc.*, Vol. 88, No. 1, pp. 75–82, 2011.

[30] R. Upadhyay, S. Sehwag, and H. N. Mishra, *Electronic nose guided determination of frying disposal time of sunflower oil using fuzzy logic analysis*, Vol. 221. 2017.

[31] D. R. Wijaya, R. Sarno, E. Zulaika, and S. I. Sabila, "Development of mobile electronic nose for beef quality monitoring", *Procedia Comput. Sci.*, Vol. 124, pp. 728–735, 2017.