Received: June 25, 2022. Revised: July 20, 2022.

International Journal of
Intelligent Engineering & Systems

http://www.inass.org/

Electronic Nose Signals for Analysing Similarity of Male and Female Axillary Odour to Food Material Aroma

M. Syauqi Hanif Ardani[1] Riyanarto Sarno[1]* Mikhael Ming Khosasih[1] Malikhah Malikhah[2] Doni Putra Purbawa[1] Chastine Fatichah[1] Dwi Sunaryono[1] Rahadian Indarto Susilo[3] Shoffi Izza Sabilla[1] Kelly Rossa Sungkono[1]

[1] *Department of Informatics, Faculty of Intelligent Electrical and Informatics Technology, Institut Teknologi Sepuluh Nopember (ITS) Sukolilo, Surabaya, Indonesia*
[2] *Department of Engineering, Data Science Technology Study Program, Faculty of Advanced Technology and Multidiscipline, Airlangga University, Surabaya, Indonesia*
[3] *Department of Neurosurgery, Faculty of Medicine, Airlangga University, Surabaya, Indonesia.*
* Corresponding author's Email: riyanarto@if.its.ac.id

Abstract: Human axillary odour is a unique discussion because it stores various useful information, including health, food quality analysis, and the similarity to the smell of the food eaten which can be done using an electronic nose (e-nose). However, rarely for research use an e-nose to determine the similarity of human axillary odour to food material odours. This research aimed to determine the similarity between male and female odours based on food material odours, such as garlic, shallot, and red chilli using an e-nose. We propose a method using an electronic nose with 8 Taguchi gas sensor (TGS) and 1 digital & humidity (DHT) sensor, then processing the resulting data using the fast fourier transform (FFT) smoothing method. Sensor data that have anomalies are removed by the quantile method. The standardization process is carried out to the signal data and feature extraction is processed with mean, standard deviation, and minimum value. This study found that the proposed method can produce a fairly good cluster of aroma food materials, faithfully getting the highest accuracy of 96.67 % using support vector classifier (SVC) and k-nearest neighbour (KNN) with each best parameter. We also found that the smell of human armpit sweat generally has a lot of similarity with food materials, specifically that male has the highest similarity to the shallot, while female has the highest level of similarity to garlic. In contrast, red chilli has the lowest level of similarity with human armpit odour.

Keywords: Electronic nose, Similarity, Axillary odour, Food material aroma, Classification.

1. Introduction

The electronic nose is an artificial nose built to help smell an odour that contains many gas sensors to detect smells and transform them into data signals. Sometimes, a human nose fails to perceive smell because of illness or other conditions, so we can use an electronic nose to detect smell more easily. Many researchers have discussed e-nose for various problems, from detecting food quality [1], health interests like diabetes detection [2], detecting adulteration and quality control of olive oil [3], to detecting halal food by utilising pork and beef data [4]. The electronic nose has been used widely for many purposes because of its neutrality, stability, and durability than traditional odour analysis [5]. Overall, the electronic nose is a well-built system to perceive odour because of its sensitivity and quick data process. The electronic nose can obtain all information and data signal components about gases from odour smells.

Recently, some weaknesses have been found in the electronic nose processing data. It has cross-sensitivity [6] and low selectivity [7]. Its weakness makes the electronic nose have a limitation to discrimination [5]. Besides discrimination objects, the electronic nose has a limitation to distinguish

Received: June 25, 2022. Revised: July 20, 2022.

similar odours. Research rarely uses the electronic nose to discriminate between male and female body odours with food material odours. These food material odours are kitchen spices, such as garlic, red chilli, shallot, etc. One of the research in swiss states that male odour is similar to cheese and female odour is similar to onion [8]. The previous study explains that the human body odour is similar to what they consumed [9]. The similarity method uses to determine the similarity among texts [10]. The recent studies used e-nose to predict disease and outlier detection from human axillary odour [11], [12]. From these previous studies, no one discussed method to determine the similarity between human body odour and food material using electronic nose. Therefore, this research used the electronic nose to distinguish the food material odours, such as garlic, shallot, and red chilli then analyse the similarity to the human axillary odour. These food materials in this study were chosen based on the popularity of spices that Indonesian people used in their foods.

This research aims to determine similarities between male and female odours based on food material odour using an electronic nose. Gender (Male and female) and food material odour will be classified with classification methods, such as the k-nearest neighbour (KNN) method, support vector classifier (SVC), linear discriminant analysis (LDA), multi-layer perceptron (MLP), logistic regression (LR). Afterwards, the similarity method is used to determine similarities between males and females based on the food material.

This research has three main contributions, such as (i) making e-nose architecture with sensors that can detect correlations in food material aroma and human axillary odour; (ii) obtaining the best method for processing data on food material aroma and human axillary odour produced from e-nose; (iii) Finding the similarity correlation produced by food material aroma and human axillary odour. This paper is structured as follows: Section 1 provides the background of our research. Section 2 provides previous topics of other studies. Section 3 covers the proposed method. Section 4 explains the results of experiments and discussion. Section 5 is the conclusions and summarizes the main point of our research.

2. Related works

To this day, the electronic nose is widely used for various needs and in various fields, including health, aroma detection, disease detection, and so on. Each has its way of data collection, data processing, and determining the best results. Several related works that have been done by previous researchers are summarised in Table 1.

In [3], e-nose uses 12 different olive oil and contains 32 input sensors, than reduced into eight inputs to control the olive oil quality. The study uses several algorithms for classification, such as Naive Bayesian, KNN, LDA, Decision Tree, ANN, and SVM, and has two approaches, both are used with the same algorithms. The only one difference is that the second one is used with a dimension reduction algorithm named principal component analysis (PCA).

Human body odour was used to track improvement of the health status of athletes [13]. The

Table 1. Related works

Years	Author(s)	Dataset	Main Contribution	Methods
2017	Emre Ordukaya, et al. [3]	Olive oils	Olive oil quality control	Use several classification methods and data reduction
2018	Tanthip Eamsa-ard, et al. [13]	Axillary odour	Athlete health status monitoring	Use 3D plot from reduced data by PCA and analyse the clusters
2018	Sai Xu, et al. [14]	Rice stem	Brown planthopper infection detection	Recognition and similarity analysis by PCA
2018	Danang Lelono, et al. [15]	Chilli	Chilli quality classification	Use max value for feature extraction, PCA for see clusters.
2019	Madeshwari Ezhilan, et al. [16]	Broccoli	Freshness detection of broccoli	GC-MS, FTIR, PCA, centroid and completely link analysis
2020	Mayumi Nomura, et al. [17]	Banana juice	Find the differences between banana juice	Use e-nose called FF-2A and GC-MS for analysing
2022	Malikhah, et al. [11]	Axillary odour	Detection of Infectious Respiratory Disease (SARS)	Used e-nose to SARS detection using Stacked Deep Neural Network (DNN)
2022	Purbawa, et al. [12]	Axillary odour	Outlier Detection	Implemented adaptive filter with multiple feature extraction

Received: June 25, 2022. Revised: July 20, 2022.

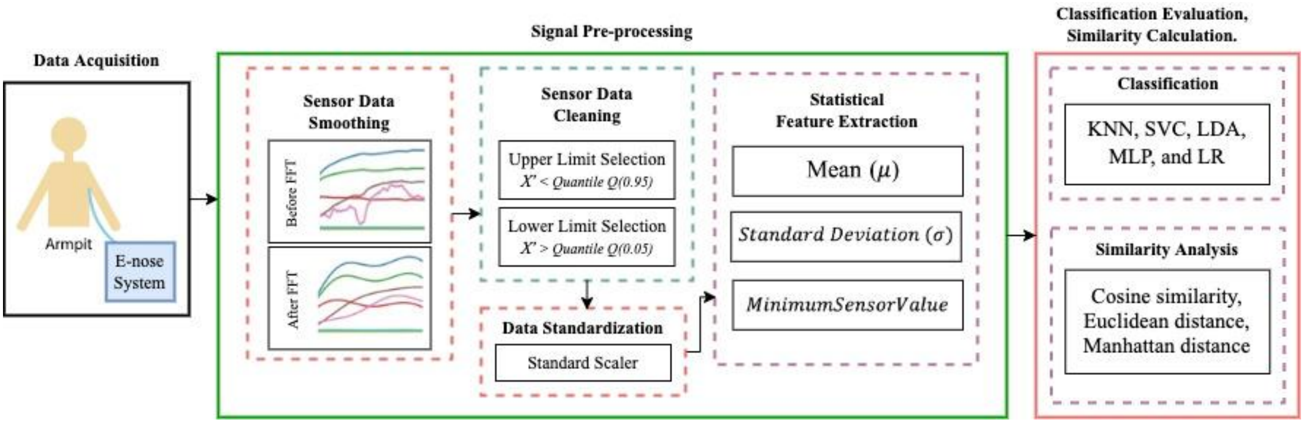

Figure. 1 Proposed methode

study implemented an electronic nose containing four calibrated gas sensors, namely PSE-COOH, PVP-COOH, Poly 4-styrenesulfonicCOOH, and PVA-COOH to obtain axillary odour from the athlete tested and analysed the volatile organic compounds (VOCs) by observing the 3D plot from clustered data by PCA. The data were collected from 3 volunteers and collected by 0, 3, 6, and 9 hours sequentially, then the data was finally gathered together with the time when it was taken.

In [18], the recognition and similarity between volatiles of brown rice plant hoppers and rice stem are discussed using an e-nose containing ten sensors, namely W1C, W5S, W3C, W6S, W5C, W1S, W1W, W2S, W2W, and W3S. The pre-processing proposed is by subtracting the original sensor value from each sample by the average value of the empty sensor without a sample. The study used hierarchical clustering analysis (HCA) to analyse the volatile similarity of all samples and used Euclidean distance to measure the interval of samples. The similarity technique used is by analysing the clusters of the plot results manually generated after PCA and the classifications used are k-nearest neighbour (KNN), probabilistic neural network (PNN), and support vector machine (SVM).

E-nose contains six Taguchi gas sensor (TGS) sensors, namely TGS 880, TGS 822, TGS 826, TGS 2602, TGS 2620, and TGS 2600, which also have been used for analysing the freshness of broccoli [19]. The moisture content of broccoli samples was examined by bacterial separation, gas chromatography-mass spectrometry (GC-MS) analysis, and fourier transform infrared (FTIR). To identify commonalities and differentiate between fresh broccoli, half-contaminated broccoli, and completely contaminated broccoli, the PCA, centroid-link cluster, and completely-link cluster analyses were used.

In [20], the quality of chilli sauce is measured by the e-nose containing TGS 2620, TGS 813, TGS 822, TGS 2600, and TGS 2620. The study uses baseline manipulation for pre-processing purposes because it can reduce multiplicative sensor drift. After pre-processing, only the max value is used as feature extraction for each sensor using multivariate pattern recognition to determine similarity and dissimilarity. Multivariate pattern recognition based on PCA is used to explore the similarity and dissimilarity among sensor responses.

In [21], an e-nose called FF-2A is used to detect the differences between several banana juice brands. The details of the sensor used are not stated but have been calibrated and standardised with nine ingredients, namely hydrogen sulfide, methyl mercaptan, ammonia, trimethylamine, propionic acid, butyraldehyde, butyl acetate, toluene, and heptane. The similarity and dissimilarity are measured by analysing the GC-MS of each brand on different days, and it is found that the most similar gases are hydrogen sulfide, ammonia, amine, ester, aldehyde, and carbon hydrate.

The recent studies also implement e-nose with human axillary odour. In [11] electronic nose was used to detect the infectious respiratory disease (SARS) which utilizes 5 metal oxide semiconductor (MOS) gas sensors. Using 29 statistical parameters and 2 hidden layers, the system generated the highest accuracy of 0.940 for the detection of 2 classes, where the best FCDN model has a total of 90,561 parameters. While on [12], the outlier detection was built to detect corrupted data in the human axillary odour.

Based on several previous studies, none of them has included e-nose used as an analysis of the classification and similarity of human axillary odour with the aroma of food ingredients. We propose this study to contribute to the development of e-nose implementation so that it can become a discussion and add references for further research.

3. Proposed method

This study aims to find the right sensor and the best method in detecting the similarity between the smell of human axillary odour and the food material aroma.

Fig. 1 shows the proposed system scheme to obtain the best classification and similarity result. The incoming signal would be raw data from the e-nose system that needed to be processed further. Then, the signal needs to be refined using fast fourier transform (FFT) to get a better signal. Next, the signal cleaning process is carried out using quantile method to remove the anomaly data that may exist by using the quantile upper limit and under limit selection. After that, the statistical feature extraction process is carried out to obtain the required features. After that, the classification and evaluation process are carried out, and then the similarity value is calculated and be evaluated.

3.1 Electronic nose

A gadget called the e-nose system is built around a condensed model of the biological olfactory system. It is made up of a number of electronic chemical partial specificity sensors and a pattern recognition system that can distinguish between simple and complicated odours. The two main hardware elements of the e-nose are the gas sensors and the data collection system.

Fig. 2 describes the e-nose architecture we used in this study. The vacuum pump hose is placed in the inner axilla. The scent of the axilla produced by the sweat is inhaled by the pump and sprayed into the sensor tube. Each sensor detects the gas levels in the aroma of the axilla and sends analog signal data to the microcomputer. A microcomputer converts the analog signal into a digital signal and then sent to the computer to be classified using machine learning.

The mechanism settings, essential parameters, and data collected from the e-nose can be stored on a computer and then processed using a designated program. E-nose contains a microcontroller, an electronic pump to draw the air sample into the sensor chamber, a sensor chamber which contains gas sensors, and a micro valve to select the odour samples.

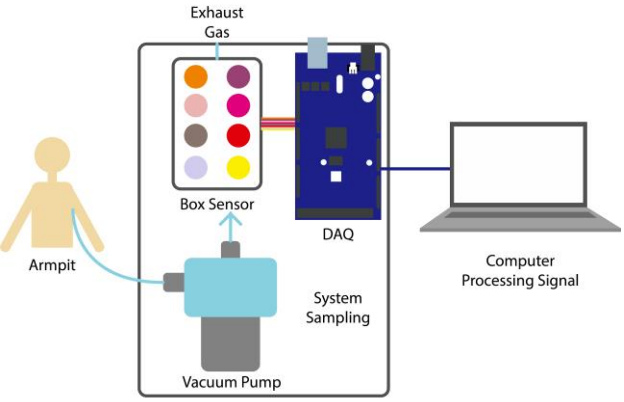

Figure. 2 E-nose architecture

Table 2. The sensor array explanations

Sensor	Gases Detected
TGS822	Methane, CO, Isobutane, n-Hexane, Benzene, Ethanol, Acetone
TGS2612	Ethanol, Methane, Isobutane, Propane
TGS2620	Methane, CO, Isobutane, Hydrogen, Ethanol
TGS832	R-12, R-134a, Ethanol, R-22
TGS826	Isobutane, Hydrogen, Ammonia, Ethanol
TGS2603	Hydrogen, H2S, Methyl Mercaptan, Trimethyl Amine, Ethanol
TGS2600	Methane, CO, Isobutane, Ethanol, Hydrogen
TGS813	CO, Methane, Ethanol, Propane, Isobutane, Hydrogen
DHT21	Temperature, Humidity

The selection of the sensors is based on the analysis of gases contained in human axillary odour, where human odour contains active compounds called volatile organic compounds (VOCs) [22] then we look for correlations of sensors that can detect each of these VOCs. Then, based on the analysis of previous studies, especially in [3, 14–17] which discussed the use of sensors for the detection of some plant products, we chose 8 MOS sensors that can detect the content between human axillary odour and food material aroma. The 8 sensors, each has its function to retrieve data on specific gases, as shown in Table 2.

The e-nose is connected to the computer using a USB interface for data acquisition [23]. The data taken were related to male and female axillary odour. The food material data taken uses raw ingredients, such as shallot, garlic, and red chilli as they are widely and oftentimes consumed by Indonesian people.

3.2 Sensor data smoothing

In the implementation of data collection, there are several challenges that cause the data obtained to be not smooth. To overcome this, we use fast fourier transform (FFT) so that the damaged data can be smoothed and produce a better signal. The FFT signal calculation can be seen in Eq. (1).

$$x[k] = \sum_{n=0}^{N-1} x[n] e^{\frac{-j2\pi kn}{N}} \quad (1)$$

where N is the total row of signal frequency data distribution from all the sample data to be determined.

Received: June 25, 2022. Revised: July 20, 2022.

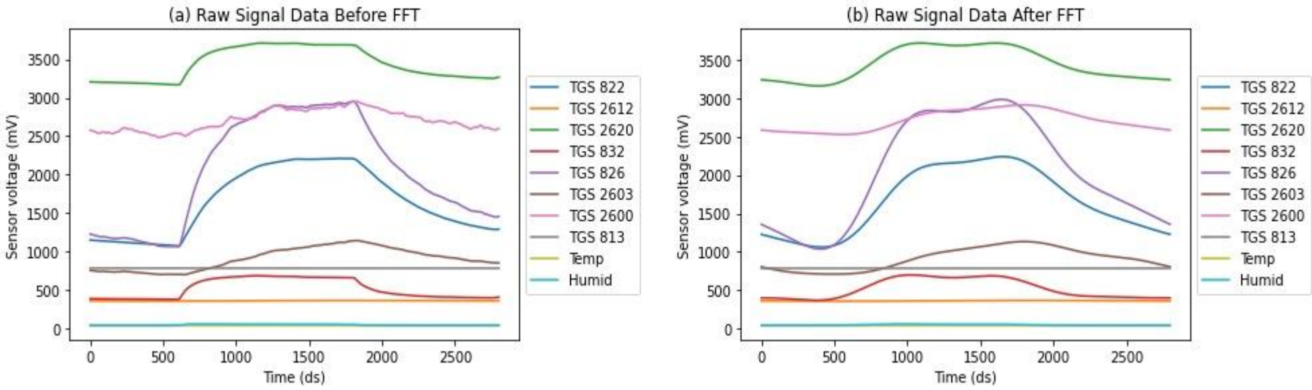

Figure. 3 Visual comparison of signal data before and after smoothing process

Table 3. Implemented quantile method

Parameter	Quantile Threshold
Upper_limit	0.95
Under_limit	0.05

The data generated before and after the FFT process can be seen in Fig. 3.

3.3 Sensor data cleaning

To overcome irregular sensor data, we propose cleaning using a quantile method. Quantile method (Q) allows for selecting unwanted data, such as data that rises too high or drops too low significantly and causes data anomalies. Each sensor has an upper and lower limit with the same quantile values with calculations in Eq. (2), and detailed threshold values in Table 3.

$$\boldsymbol{x}[n] = Q_{\text{Under\_limit}} \leq \boldsymbol{x} \leq Q_{\text{Upper\_limit}} \tag{2}$$

3.4 Statistical feature extraction

Feature extraction in this research used statistical parameters combination that we found the best result for this study. The previous study used this statistic feature extraction to obtain statistic parameters [2]. The process began with raw signal data containing a concentration of gases that captured stationary signals [24]. The stationary signal would be processed to obtain the statical parameter. This paper used three statistical parameters: mean, standard deviation, and minimum value.

Mean (μ) is the total signal frequency to be divided by the number of sample data. The average Mean (μ) equation can be seen in Eq. (3).

$$\mu = \frac{1}{n} \left(\sum_{i=1}^{N} x_i \right) \tag{3}$$

Standard deviation (σ) is a statistical value used to determine the closeness of a statistical sample to the mean of a data. The standard deviation (σ) can be calculated by Eq. (4).

$$\sigma = \sqrt{\frac{\sum_{i=1}^{N}(xi-\mu)^2}{n-1}} \tag{4}$$

The minimum sensor value is the lowest value in the raw sensor data. The minimum sensor value can be seen in Eq. (5).

$$MinimumValue = min(x_i) \tag{5}$$

3.5 Data standardization

To standardise the scale of the processed data with very large distance differences, we propose to use a standard scaler as a standardisation method. $Standar\ Scaler$ (z) can be calculated by Eq. (6).

$$z = \frac{x-\mu}{\sigma} \tag{6}$$

Algorithm 1 is a pseudocode for processing e-nose signal data taken from food material aroma and human axillary odour.

Algorithm 1 Pre-procesing Method

Input: E-nose Data of Food Material Aroma, E-nose Data of Huma Axillary Odour
Output: data features

FMA ← load(FoodMaterialAroma.csv)
HAO ← load(HumanAxillaryOdour.csv)
Sensor_columns ← all sensors in e-nose system

All_data = [FMA, HAO]

for sampling_id in All_data.sampling_id **do**
    Data_sampling ← All_data["sampling_id"]
        == sampling_id

Step 1: obtain smoothed sensor value $x[k]$
x ← FFT(Data_sampling)

Step 2: implement quantile method $\boldsymbol{x}[n]$
$x \leftarrow \boldsymbol{x} \leq Q_{\text{Upper\_limit}}$
$x \leftarrow \boldsymbol{x} \geq Q_{\text{Under\_limit}}$

Step 3: calculate statistical features
for sensor in Sensor_columns **do**
Feature1 ← $\mu(x_{sensor})$
Feature2 ← $\sigma(x_{sensor})$
Feature3 ← $MinimumValue(x_{sensor})$
Feature_sensor ← append(Feature1, Feature2, Feature3)
end for
Feature_sampling ← append(Feature_sensor)
end for

Step 4: standardize data with standard scaler (z)
Feature_sampling ← Scaler(Feature_sampling)

Feature_FMA ← Feature_sampling["sampling_id"] == FMA.sampling_id

Feature_HAO ← Feature_sampling["sampling_id"] == HAO.sampling_id

return Feature_FoodAroma, Feature_HumanOdour

3.6 Feature reduction and clustering

In this study, to plot and analyse the data distribution, we used principal component analysis (PCA) algorithm to perform feature reduction. PCA is a popular statistical approach for visualising data in two or three uncorrelated dimensions after being transformed from all correlated data [25]. PCA is a mathematical method that attempts to capture variance in datasets using a small number of factors [26]. Besides being useful for helping data visualisation, PCA can also find new information that can help the learning process. PCA is also useful as an analysis of data distribution by utilizing 2d or 3d visuals from dimensionally reduced data which represents all the features in the dataset.

The next step was to cluster the gender and food material data to ensure that each class is in the correct cluster, so that it can be confirmed that the proposed feature extraction is quite good and the classification process and similarity calculation can be more accurate. The clustering method uses k-means clustering (c), which can be obtained by Eq. (7).

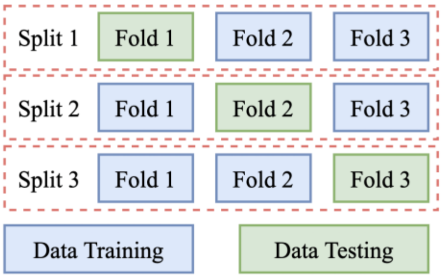

Figure. 4 K-fold validation mechanism

$$J = \sum_{j=1}^{k} \sum_{i=1}^{N} d(x_i^{(j)} - c_j)^2 \tag{7}$$

where k is number of clusters, d is the distance between data x and centroid c.

3.7 Classification

The data were then classified to determine the accuracy of each data and to confirm that the proposed method is well executed. Classification methods for testing the data are KNN, SVC, LDA, MLP, and LR. We use these methods because e-nose classification often uses these classification methods.

The stratified k-fold validation method is applied to evaluate the results of each mechanism and classification to measure the performance of each classifier [27] and to avoid overfitting or underfitting that occurs in the classification process. In this study, we use $K = 3$ for validation process. The k-fold mechanism is sown in Fig. 4.

3.8 Similarity

After classification, gender and food material odour data would go through similarity. Similarity analysis is useful for determining the acidity level of an object with another object. Many studies use this technique, including to find out if the text contains plagiarism with previous works or is an original work and has never been published [28], news analysis on social media [29], sentiment analysis, decision making on the social networks [30], and so on. There are several similarity analyses, and this study used cosine similarity, Euclidean distance and Manhattan distance analysis and discussed both to determine the similarity between human axillary odour and food aroma. First, we used the cosine similarity. The previous study [31] used cosine similarity to detect similarities between book data content. This method is a similarity between two vectors and measuring the cosine between its angle [31]. The greater the value of cosine similarity, the closer the similarity of the objects. The cosine similarity $\cos(a, b)$ equation can be seen in Eq. (8), where (a, b) are two objects whose the cosine similarity is calculated.

$$\cos(a, b) = \frac{a_1 b_{1+\cdots+} a_n b_n}{\sqrt{a_1^2+\cdots+a_n^2}\sqrt{b_1^2+\cdots+b_n^2}} \tag{8}$$

The second method is distance analysis using Euclidean distance. This method is also called L2 [32], aims to find the nearest distance between two-dimension data. The smaller the value of the Euclidean distance, the closer the similarity of the objects. Euclidean distance (d) equation can be seen in Eq. (9), where (a, b) are two objects that are calculated the Euclidean distance between them.

$$d(a, b) = \sqrt{\sum_{i=1}^{N}(a_i - b_i)^2} \tag{9}$$

The third method is distance analysis using Manhattan distance. Manhattan distance calculates the separation between two real-valued vectors, often known as the Taxicab distance or the city block distance. The smaller the value of the Manhattan distance, the closer the similarity of the objects. The Manhattan distance (d) calculation can be seen in Eq. (10), where (a, b) are two objects that are calculated the Manhattan distance between them.

$$d(a, b) = \sum_{i=1}^{N}|a_i - b_i| \tag{10}$$

4. Results and discussion

4.1 Data acquisition

Fig. 5 shows the sensor movement on the e-nose used. There are 8 TGS sensors that take data directly from the object being studied and two sensors (Temp, Humid) that provide information that the temperature and humidity conditions around the tool remain stable or not, but the sensor value are not use in the next calculation. The e-nose time required to collect each raw data is 280 seconds, where the first 60 seconds of the e-nose is flushing (P1), then at 60 to 180 seconds or in 120 seconds, the e-nose starts sensing time (P2). This e-nose picked up signals from the sensors on the axilla of men and women and the food material. At the last part, the e-nose starts to the cleaning process at 180 to 280 seconds or about 100 seconds (P3). The data that was successfully retrieved by the team consisted of 36 .csv data, 18 data for women and 18 data for men.

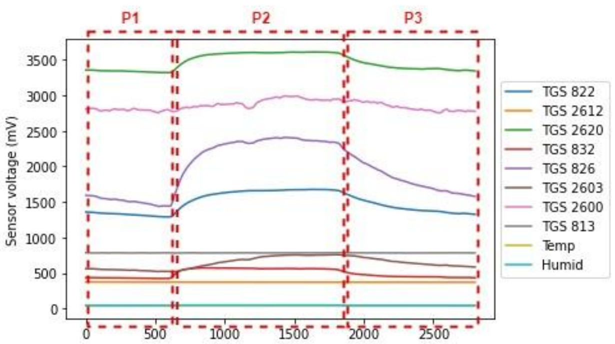

Figure. 5 Raw data of e-nose sensor array

According to the declaration of the Helsinki by world medical association (WHO), all studies involving humans as research subjects are required to obtain the consent of the subject. In this study, all the armpit sweat data that we collected from the people were with their consent and willingness to have their armpit data taken for use in this study. We also ensure that there is no improper use of data.

Data were collected two times, first using data from men and women, and secondly using food material data. We took male and female data using the e-nose, and then the e-nose tube was placed on the male and female axilla. The axilla had been chosen because this was the best sampling to obtain chemicals from the human body [25]. After taking male and female data, we took food material data by initially taking data from three food materials consisting of garlic, shallots, and red chillies. The data collection process was almost the same as collecting the human axillary odour for men and women. The human odour was collected by placing the e-nose tube directly on the human axilla, but the food aroma samples were placed into a glass container, of which the e-nose would do the suction to obtain the reacting signals, with the same dose of 0.1 grams for each sample, so the data collecting is fair for each sample. The results of the data collection for garlic, shallots and red chillies are 10 data each.

First, the male and female gender data were labelled to describe their data. Afterwards, the food material data were also labelled. Data labelling uses the acronym of food materials in Indonesian. Garlic is *Bawang Putih* (BP), shallot is *Bawang Merah* (BM), and red chilli is *Cabai Merah* (CM). All the data drop column time, temperature, and humidity. Data row will be used from 60 seconds to 180 seconds or from comma separated value (CSV) file row 600 until 1800 because this is the process to convert smell to the data signal.

4.2 Clustering result

Fig. 6 is the clustering result of food material aroma using PCA for feature reduction. Cluster 1, cluster 2, and cluster 3 represent the CM, BM, and BP class respectively. The proposed method produces a

Received: June 25, 2022. Revised: July 20, 2022.

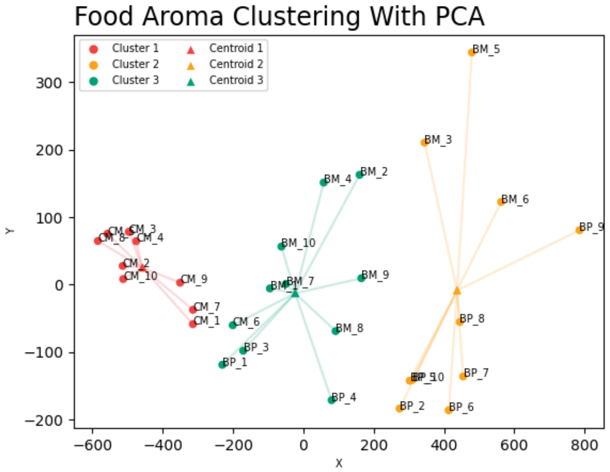

Figure. 6 Clustering result of food material aroma

fairly good clustering, although there are several classes that are in different clusters from the original. It can be seen that the CM class can be perfectly clustered, but in the BP and BM classes there are some cluster errors. This can happen because some of the BP and BM data are closed to each other, and the clustering uses 2 reduced dimensions by PCA method so there is some information that is not perfect.

4.3 Classification result

This study used n-nearest neighbour (KNN), support vector classifier (SVC), linear discriminant analysis (LDA), multi-layer perception (MLP), and logistic regression (LR) to analyse the classification results. The goal of this classification is to obtain accuracy from several classification methods. To find the best combination of parameters, we use the grid search method. This method allows the search for the most optimal parameters of each model to be built.

The method that we propose manages to get the best accuracy up to 96.66 % by using SVC and KNN with their best parameter. Then MLP and LR each get 93.33 % accuracy and the LDA method get 90.00 % accuracy. This accuracy is obtained by using k-fold validation to avoid the potential for overfitting and underfitting. Each classifier can carry out the classification process well so that it can show that the proposed method has produced quite good features and can be continued for the next process, similarity calculation. Details of the parameters used in each classifier are shown in Table 4.

4.3 Similarity result

Similarity analysis was carried out with 3 methods. Analysis of cosine similarity, it can be seen that the male gender is closed to the BM with the highest value of 0.9646, followed by BP and the

Table 4. Detail best parameters of each classifier

Methods	Best Parameters	Accuracy
Support Vector Classifier	{'C': 0.1, 'gamma': 0.01, 'kernel': 'linear'}	96.67%
K-Nearest Neighbor	{'metric': 'euclidean', 'n_neighbors': 1, 'weights': 'uniform'}	96.67%
Logistic Regression	{'verbose': 1}	93.33%
Multi-Layer Perceptron	{'activation': 'tanh', 'alpha': 0.0001, 'hidden_layer_sizes': (10, 30, 10), 'learning_rate': 'adaptive', 'max_iter': 100, 'solver': 'adam'}	90.00%
Linear Discriminant Analysis	{'solver': 'svd'}	90.00%

smallest is CM with values of 0.9610 and 0.4855, respectively. Meanwhile, female got the highest similarity value up to 0.9526 when compared to BP, followed by BM and CM with values of 0.9351 and 0.4664, respectively. The Euclidean distance analysis found that male also has the closest level of similarity with BM with a value of 1.8628, followed by BP and furthest with CM with values of 1.9988 and 5.4878, respectively. As for female, BP also obtained the closest similarity score with a value of 2.2771, followed by BM and CM with a value of 2.5115 and 5.2199, respectively. The Manhattan distance analysis found there are different results compared to the previous similarity method in male. The closest similarity value obtained by the male is with BP with a value of 6.6336, followed by BM and the farthest is still with CM with values of 6.7730 and 20.1891, respectively. Meanwhile for female, it is still the same as other similarity methods, namely BP has the closest value with a value of 6.9790, followed by BM and CM with a value of 8.4614 and 18.5678. The detailed result of the similarity of the human gender and the food material aroma using 3 different methods such Cosine similarity, Euclidean distance, and Manhattan distance is shown in Table 5.

5. Conclusions

The e-nose system we proposed can produce clusters of food material aromas quite well and produce quite good accuracy with several classifiers, which is the best method for classifying food material aroma using SVC with an accuracy of 96.67 %. This study also found a correlation between the food ingredients eaten by humans and the smell of sweat

Table 5. Similarity results using cosine similarity, Euclidean distance, and Manhattan distance

Cosine Similarity			Euclidean Distance			Manhattan Distance		
BM	BP	CM	BM	BP	CM	BM	BP	CM
0.9646	0.9610	0.4855	**1.8628**	1.9988	5.4878	6.7730	**6.6336**	20.1891
0.9351	**0.9526**	0.4664	2.5115	**2.2771**	5.2199	8.4614	**6.9790**	18.5678

produced by the body. The result of the calculation using cosine similarity in this research indicated that male odour has a very close similarity with shallot (BM) of 0.9646, while the lowest similarity is with red chilli (CM) of 0.4855. On the other hand, the female odour is also close to garlic (BP) of 0.9526, while having the lowest similarity with red chilli of 0.4664. The result of the calculation using Euclidean distance in this research indicated that male odour has the shortest distance also with shallot of 1.8628, while the longest distance also with red chilli of 5.4878 and the female odour is also close to garlic with the shortest distance of 2.2771, while having the longest distance with red chilli of 5.2199. The result of the calculation using Manhattan indicated that male odour has the shortest distance also with garlic of 6.6336, while the longest distance also with red chilli of 20.1891 and the female odour is also close to garlic with the shortest distance of 6.9790, while having the longest distance also with red chilli of 18.5678. Overall, we can conclude that human axillary odour generally have a lot of similarity with food materials, specifically that male has the highest level of similarity to shallot, while the female has the highest level of similarity to garlic. In contrast, red chilli has the lowest level of similarity with human armpit odour.

Conflicts of interest

The authors declare that there is no conflict of interest regarding the publication of this paper.

Author contributions

Conceptualisation, M. Syauqi Hanif Ardani; methodology, M. Syauqi Hanif Ardani, Mikhael Ming Khosasih; software, M. Syauqi Hanif Ardani, Mikhael Ming Khosasih; validation, M. Syauqi Hanif Ardani; writing—original draft preparation, M. Syauqi Hanif Ardani, Mikhael Ming Khosasih; writing—editing, M. Syauqi Hanif Ardani; assisted in methodology, Malikhah and Doni Putra Purbawa; writing—review, Shoffi Izza Sabilla, Kelly Rossa Sungkono, Riyanarto Sarno, Chastine Fatichah, Dwi Sunaryono; supervision, Riyanarto Sarno, Chastine Fatichah, Dwi Sunaryono, Rahadian Indarto Susilo; proposes problem ideas, Riyanarto Sarno.

Acknowledgments

This research was funded by Lembaga Pengelola Dana Pendidikan (LPDP) under Riset Inovatif-Produktif (RISPRO) Invitation Program managed, the Indonesian Ministry of Education and Culture under Penelitian Terapan Unggulan Perguruan Tinggi (PTUPT) Program, and Institut Teknologi Sepuluh Nopember (ITS) under project scheme of the Publication Writing and IPR Incentive Program (PPHKI).

References

[1] M. Ali, N. Hashim, S. A. Aziz, and O. Lasekan, "Principles and recent advances in electronic nose for quality inspection of agricultural and food products", *Trends in Food Science and Technology*, Vol. 99, pp. 1–10, 2020.

[2] Hariyanto, R. Sarno, and D. R. Wijaya, "Detection of diabetes from gas analysis of human breath using e-Nose", In: *Proc. of 2017 11th International Conference on Information & Communication Technology and System (ICTS)*, pp. 241–246, 2017.

[3] E. Ordukaya and B. Karlik, "Quality Control of Olive Oils Using Machine Learning and Electronic Nose", *Journal of Food Quality*, Vol. 2017, pp. 1–7, 2017.

[4] R. Sarno, K. Triyana, S. I. Sabilla, D. R. Wijaya, D. Sunaryono, and C. Fatichah, "Detecting Pork Adulteration in Beef for Halal Authentication Using an Optimized Electronic Nose System", *IEEE Access*, Vol. 8, pp. 221700–221711, 2020.

[5] W. Guo, H. Kong, J. Wu, and F. Gan, "Odor discrimination by similarity measures of abstract odor factor maps from electronic noses", *Sensors (Switzerland)*, Vol. 18, No. 8, 2018.

[6] Z. Liang, F. Tian, S. X. Yang, C. Zhang, H. Sun, and T. Liu, "Study on interference suppression algorithms for electronic noses: A review", *Sensors (Switzerland)*, Vol. 18, No. 4, 2018.

[7] T. Skov and R. Bro, "A new approach for modelling sensor based data", *Sensors and Actuators*, B: Chemical, Vol. 106, No. 2, pp. 719–729, 2005.

[8] J. Schwarcz, "Do men's and women's armpits

smell differently?", Mar, 2017, Accessed: Mar. 11, 2022. [Online]. Available: https://www.mcgill.ca/oss/article/general-science-you-asked/do-mens-armpits-smell-differently-womens

[9] S. L. Tan, H. S. Teo, and J. G. Guzmán, "E-nose screening of pesticide residue on chilli and double-checked analysis through different data-recognition algorithms", In: *Proc. of 2010 IEEE Electronics, Robotics and Automotive Mechanics Conference*, CERMA 2010, 2010, pp. 592–596.

[10] V. Cross, V. Mokrenko, K. Crockett, and N. Adel, "Using Fuzzy Set Similarity in Sentence Similarity Measures", In: *Proc. of 2020 IEEE International Conference on Fuzzy Systems (FUZZ-IEEE)*, Jul, 2020, pp. 1–8.

[11] Malikhah et al., "Detection of Infectious Respiratory Disease Through Sweat From Axillary Using an E-Nose With Stacked Deep Neural Network", *IEEE Access*, Vol. 10, pp. 51285–51298, 2022.

[12] D. P. Purbawa et al., "Adaptive filter for detection outlier data on electronic nose signal", *Sens Biosensing Res*, Vol. 36, 2022.

[13] T. E. Ard, T. Srikhirin, and T. Kerdcharoen, "Online Monitoring of Human Body Odor for the Improvement of Athlete's Health Status", In: *Proc. of 2018 IEEE International Conference on Consumer Electronics-Taiwan (ICCE-TW)*, 2018, pp. 1–2.

[14] S. Xu, Z. Zhou, L. Tian, H. Lu, X. Luo, and Y. Lan, "Study of the similarity and recognition between volatiles of brown rice plant hoppers and rice stem based on the electronic nose", *Computers and Electronics in Agriculture*, Vol. 152, pp. 19–25, 2018.

[15] D. Lelono, D. Permana, F. Achmad, T. W. Widodo, M. A. Bramantya, and K. Triyana, "Quality Classification of Chili Sauce Using Electronic Nose with Principal Component Analysis", In: *Proc. of 2018 4th International Conference on Science and Technology (ICST)*, pp. 1–5, 2018.

[16] M. Ezhilan, N. Nesakumar, K. J. Babu, C. S. Srinandan, and J. B. B. Rayappan, "Freshness Assessment of Broccoli using Electronic Nose", *Measurement: Journal of the International Measurement Confederation*, Vol. 145, pp. 735–743, 2019.

[17] M. Nomura, E. Osada, T. Tokita, T. Iwamoto, and Y. Manome, "Measurement and differentiation of banana juice scent using an electronic nose FF-2A", *PeerJ*, Vol. 9, p. e10638, 2021.

[18] S. Xu, Z. Zhou, L. Tian, H. Lu, X. Luo, and Y. Lan, "Study of the similarity and recognition between volatiles of brown rice plant hoppers and rice stem based on the electronic nose", *Computers and Electronics in Agriculture*, Vol. 152, pp. 19–25, 2018.

[19] M. Ezhilan, N. Nesakumar, K. J. Babu, C. S. Srinandan, and J. B. B. Rayappan, "Freshness Assessment of Broccoli using Electronic Nose", *Measurement: Journal of the International Measurement Confederation*, Vol. 145, pp. 735–743, 2019.

[20] D. Lelono, D. Permana, F. Achmad, T. W. Widodo, M. A. Bramantya, and K. Triyana, "Quality Classification of Chili Sauce Using Electronic Nose with Principal Component Analysis", In: *Proc. of 2018 4th International Conference on Science and Technology (ICST)*, pp. 1–5, 2018.

[21] M. Nomura, E. Osada, T. Tokita, T. Iwamoto, and Y. Manome, "Measurement and differentiation of banana juice scent using an electronic nose FF-2A", *PeerJ*, Vol. 9, 2021.

[22] R. Mebazaa, B. Rega, and V. Camel, "Analysis of human male armpit sweat after fenugreek ingestion: Characterisation of odour active compounds by gas chromatography coupled to mass spectrometry and olfactometry", *Food Chemistry*, Vol. 128, No. 1, pp. 227–235, 2011.

[23] R. Sarno, S. I. Sabilla, D. R. Wijaya, D. Sunaryono, and C. Fatichah, "Electronic nose dataset for pork adulteration in beef", *Data in Brief*, Vol. 32, 2020.

[24] 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, 2020, doi: 10.22266/ijies2020.0831.06.

[25] C. Wongchoosuk, M. Lutz, and T. Kerdcharoen, "Detection and classification of human body odor using an electronic nose", *Sensors*, Vol. 9, No. 9, pp. 7234–7249, Sep, 2009.

[26] D. Granato, J. S. Santos, G. B. Escher, B. L. Ferreira, and R. M. Maggio, "Use of principal component analysis (PCA) and hierarchical cluster analysis (HCA) for multivariate association between bioactive compounds and functional properties in foods: A critical perspective", *Trends in Food Science and Technology*, Vol. 72, pp. 83–90, Feb, 2018.

[27] S. Purushotham and B. K. Tripathy, "Evaluation of Classifier Models Using Stratified Tenfold Cross Validation Techniques", *Communications*

in Computer and Information Science, Vol. 2, pp. 680–690, 2012.

[28] R. Saptono, H. Prasetyo, and A. Irawan, “Combination of Cosine Similarity Method and Conditional Probability for Plagiarism Detection in the Thesis Documents Vector Space Model”, *Journal of Telecommunication, Electronic and Computer Engineering (JTEC)*, Vol. 10 (2–4), pp. 139–143, 2018. Accessed: Mar. 09, 2022. [Online]. Available: https://jtec.utem.edu.my/jtec/article/view/4332

[29] M. A. Smadi, Z. Jaradat, M. A. Ayyoub, and Y. Jararweh, “Paraphrase identification and semantic text similarity analysis in Arabic news tweets using lexical, syntactic, and semantic features”, *Information Processing and Management*, Vol. 53, No. 3, pp. 640–652, 2017.

[30] Z. J. Du, H. Y. Luo, X. D. Lin, and S. M. Yu, “A trust-similarity analysis-based clustering method for large-scale group decision-making under a social network”, *Information Fusion*, Vol. 63, pp. 13–29, 2020.

[31] D. Soyusiawaty and Y. Zakaria, “Book Data Content Similarity Detector With Cosine Similarity (Case study on digilib.uad.ac.id)”, In: *Proc. of 2018 12th International Conference on Telecommunication Systems, Services, and Applications (TSSA)*, pp. 1–6, 2018.

[32] M. D. Malkauthekar, “Analysis of euclidean distance and manhattan distance measure in face recognition”, In: *Proc. of Third International Conference on Computational Intelligence and Information Technology (CIIT 2013)*, pp. 503–507, 2013.