Ripeness Assessment and Quality Control of Mango Gold *Susu* using an E-Nose System

Nur Irdina Fakhrul Anwar, Nor F. Za'bah*, and Aliza Aini Md Ralib

Department of Electrical and Computer Engineering, Kulliyyah of Engineering, International Islamic University Malaysia, Kuala Lumpur

*Corresponding author: adah510@iium.edu.my

Abstract—In this paper, the development and implementation of an electronic nose (e-nose) system utilizing the MQ sensor series from MOS-type gas sensors to classify mango gold *susu* ripeness is presented. The system's performance was enhanced through machine learning techniques, including Principal Component Analysis (PCA) for data dimensionality reduction and Support Vector Machine (SVM) for classification. The SVM classifier demonstrated high accuracy, particularly in identifying unripe and overripe mangoes, with accuracy scores of 1.00 and 0.99, respectively. A comprehensive database of volatile organic compound (VOC) profiles was established, leading to a precise prediction model for assessing the different stages of ripeness based on the mango's VOC profile.

1. INTRODUCTION

The fundamental solution to food waste is precise ripeness evaluation, which can prevent premature disposal and delayed consumption. Fruit ripeness has been determined manually by experienced farmers using fruit external characteristics such color, shape, firmness, and defect after post-harvest. As a result, only skilled and experienced farmers possess the ability to discern between different phases of fruit maturity using their own judgment. Diverse classifications for various fruits arise from the fact that professionals and farmers often evaluate ripeness classification in various manners based on their personal knowledge and experiences [1]. Consequently, human error may lead to inconsistencies in fruit maturity classification, impacting the quality control process. Therefore, there is a need for sophisticated, non-destructive alternatives since conventional quality control methods have limitations. These alternative methods offer less damage to fruits by utilizing characteristics such as aroma, color, and firmness for assessment. Various non-destructive techniques, such as NIR spectroscopy, electronic nose (e-nose), and RGB color sensors, have been explored to classify fruit ripeness stages accurately. Studies by Aghilinategh et al. [2] and Sabzi et al. [3] highlight the effectiveness of non-destructive methods like NIR spectroscopy and aerial video for fruit ripeness classification based on color. High accuracy levels, up to 97.88% for apple maturity classification, demonstrate the reliability of these techniques. Additionally, research by Baeitto and Wilson [4] emphasizes the role of fruit aroma in determining fruit quality and ripeness. E-nose instruments have shown promising results in accurately classifying ripeness stages for various fruits like blueberries, bananas, and apricots. The use of non-destructive methods, particularly e-nose technology, proves to be more effective, rapid, and non-damaging compared to traditional methods. This work will delve further into the usage of e-nose technology in monitoring the ripeness of mango gold susu, one of the common mangos in Malaysia. The main structure of an e-nose is illustrated in Fig. 1. Primarily, the e-nose systems are composed of a sensor array, a signal transducer, and a pattern recognition engine.

Fig. 1 shows that the sensor array is used to detect odor or in this work, the VOCs, generating and identifying the odor "fingerprint". However, unlike an olfactory system of mammals that has multiple receptor cells, the number of sensors for most electronic systems is limited.
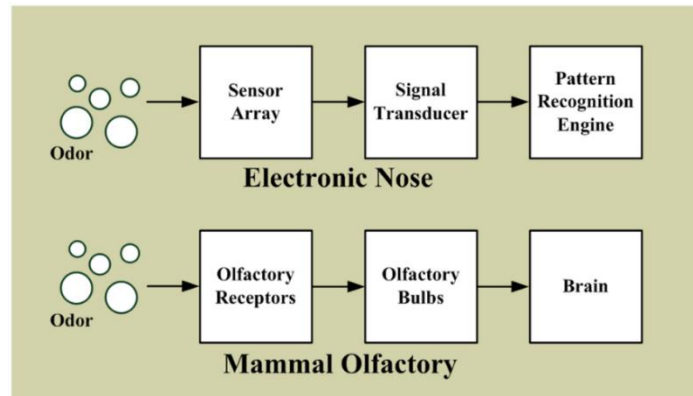
Fig. 1. An e-nose system [5]

2. METHODOLOGY

The methodology of this work involves a systematic four-step process, as shown in Fig. 2. In the first step, which is the data acquisition, the 6-MQ sensor array collects the VOC data within a controlled chamber to capture the volatile compounds emitted by the mangoes. The sensor, known for its versatility, cost-effectiveness, and sensitivity to various gases, is often used for gas detection in environmental monitoring and industrial settings. As mentioned before, the VOCs will go through the MQ series of sensors, which will detect the targeted gas and produce changes in its resistance. Table 1 shows the list of the 6 MQ sensors used in this work with the targeted gases.
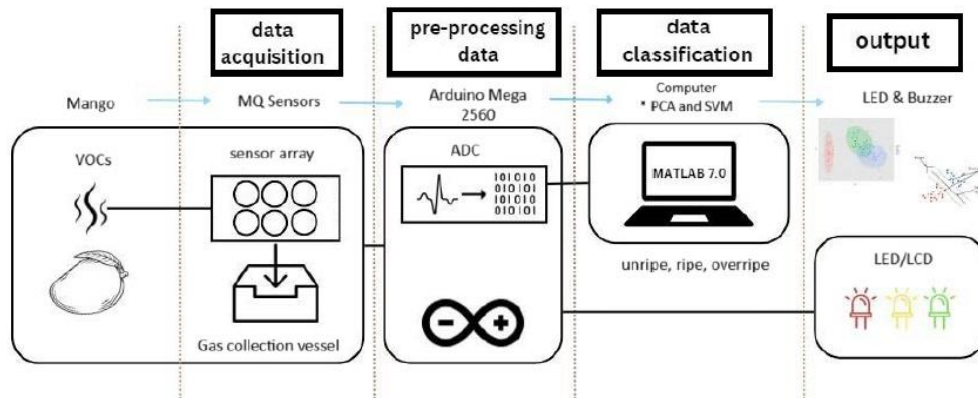

Fig. 2. The block diagram for the e-nose system for this work.

Table 1 - List of 6 MQ Sensors used for the e-nose system

Label	Sensor	Targeted Gases
S_1	MQ-2	Methane
		Butane
		LP Gas
		Smoke
S_2	MQ-3	Ethanol
		Alcohol
S_3	MQ-4	Methane
S_4	MQ-7	Carbon Monoxide
S_5	MQ-8	Hydrogen Gas
S_6	MQ-135	Ammonia
		Benzene
		Carbon Dioxide

This raw sensor data is then subjected to the pre-processing step, where an Arduino is employed for Analog-to-Digital Conversion (ADC). Additionally, baseline manipulation is performed on the raw e-nose data to normalize it, minimizing the impact of factors like temperature, humidity, and temporal drifts. Following pre-processing, the third step involves data classification utilizing Principal Component Analysis (PCA) and Support Vector Machines (SVM). PCA is employed to analyze data behavior and groupings, while SVM is used for training and classification. Finally, in the output step, this work employs three different colored LEDs, and an LCD connected to the Arduino, to produce the visual output where each color of the LED represents a specific ripeness category.

Fig. 3 shows the ripening stage of the mango fruit. Sample A's first batch of mangoes was in its early ripeness stage. After approximately seven days of storage, sample A became ripe, with a bright yellow color and a soft texture i.e. sample B. Then, overripe phases which took around five days, caused the mangoes to have a dark yellow color and a softer texture (sample C). Hence, we can safely say that the skin color and texture of the mangoes contributed to the assumption that they were at different stages of maturity. The machine learning model will categorize these samples by testing them with the e-nose setup.
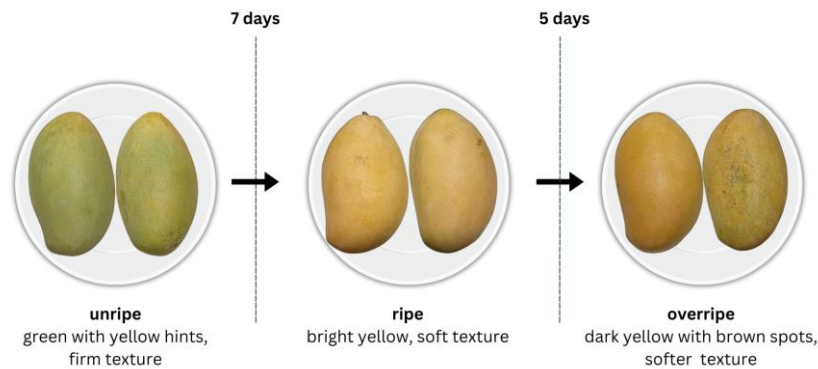

Fig. 3. The mango gold *susu* ripeness stages – unripe (sample A), ripe (sample B) and overripe (sample C)

2.1 Data Acquisition

On top of the array of sensors, the hardware configuration also consists of a smart digital thermometer and hygrometer for the recording of temperature and humidity. In order to calibrate the sensors for the first time, the sensors were preheated through electricity for 48 hours. After the calibration process, the MQ gas sensors must be preheated again for at least 30 minutes for the data collection to ensure the stability of the readings. This is based on the preliminary study that was carried out to guarantee that sensors' responses were stable during the data collection. This study is performed by comparing the sensor's reading obtained after 15 minutes against after 30 minutes. As shown in Fig. 4, it can be seen that notably, the sensors' readings were substantially more constant after 30 minutes.

Next, different types of VOCs produced by the mangoes during the three stages of ripening are collected. An airtight acrylic container is used in order to confine all the gas emitted from the mangoes. Three mangoes from each category were prepared for the data acquisition phase. Each mango was placed inside a container, and the sensors collected the information for 15 minutes, generating approximately 120 datasets per category.
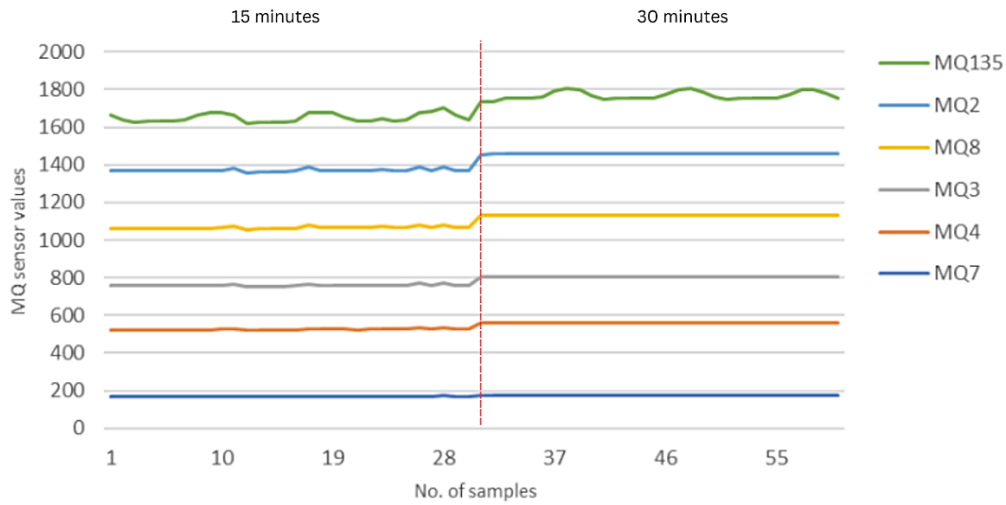
Fig. 4. MQ gas sensors' readings at the interval of 15 minutes vs 30 minutes

2.2 Data Pre-processing and Data Classification

The primary function of the Arduino microcontroller is to perform the Analog-to-Digital Conversion (ADC) procedure, which converts analog signals from the gas sensors into a digital format that can be classified. The constant analog signals generated by the gas sensors indicate the concentration of the gases they have detected. The Arduino code uses digital representations ranging from 0 to 1023 in a 10-bit ADC as the input data for the data classification stage.

A sample dataset's data analysis and classification are performed using PCA and SVM algorithms using PyCharm. PCA handles high-dimensional datasets and separates classes, especially in non-linear relationships. By reducing the dimensionality of the feature space while retaining essential information, PCA aids in uncovering the underlying patterns and structures in the MQ sensor readings. Fig. 5 illustrates the block diagram for the data analysis and classification. This work aims to use PCA for data grouping, while SVM is used for training and classification. PCA reduces the computational load on SVM during pre-processing, ensuring an efficient framework for precise ripeness evaluation based on the MQ sensor readings. The classification is based on three levels of mango ripeness: unripe, ripe, and overripe.
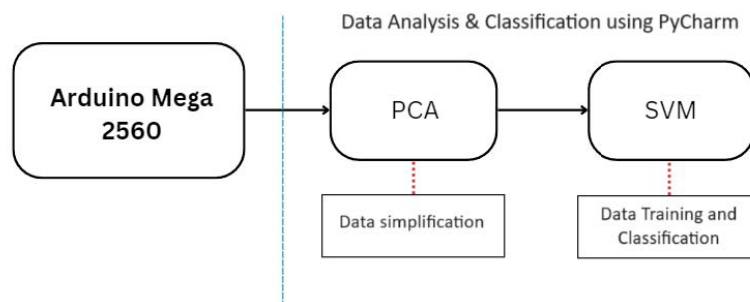

Fig. 5. The block diagram for data analysis and classification

Several measures such as accuracy, precision, recall, and F1 score may be used to assess the SVM model's performance. However, in this work, the accuracy parameter is employed. The formula for the accuracy is as shown in Eq. (1) where:

$$Accuracy = \frac{TP + TN}{TP + TN + FN + FP} \tag{1}$$

In this work, the definition of the accuracies are as follows:

- True Positives (TP): SVM correctly predicts positive classes, identifying some fruit as ripe when it is indeed ripe in ripeness assessment.

- True Negatives (TN): SVM correctly predicts negative classes, identifying some fruit as unripe or overripe when it is indeed unripe or overripe in ripeness assessment.

- False Positives (FP): SVM incorrectly predicts positive classes, mistakenly identifying an unripe or overripe fruit as ripe in ripeness assessment.

- False Negatives (FN): SVM incorrectly predicts negative classes, mistakenly identifying a ripe fruit as unripe or overripe in ripeness assessment.

3. RESULT AND DATA ANALYSIS

The illustration in Fig. 6 depicts significant distinctions in the sensors' responses corresponding to different ripeness stages. The data trends provide a solid foundation for the SVM model that allows it to classify the stages of ripeness.
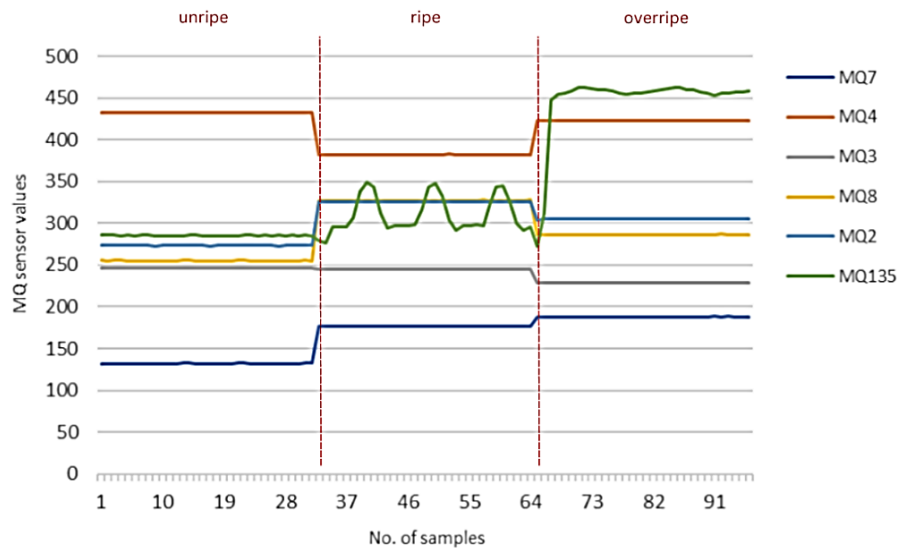

Fig. 6. Signal responses of the MQ gas sensors based on three stages of ripeness

3.1 Principal Component Analysis (PCA)

The output of the PCA analysis was promising, as shown in Fig. 7. The first principal component (PC1) explained a significant 59.54% of the variation in the sensor data. This indicates that the data has a dominant pattern or structure that PC1 is able to capture well in accordance with the three stages of ripeness. The second main component (PC2) accounted for an extra 23.19% of the variation, providing further information about the data's underlying structure. Significantly, by focusing simply on the first two main components, we were able to capture 82.73% of the overall data variability.

```
Training Data Variance Ratio


Principal Component 1 explains 59.54% of the variance
Principal Component 2 explains 23.19% of the variance
Total explained variance by the first 2 components: 82.73%
```
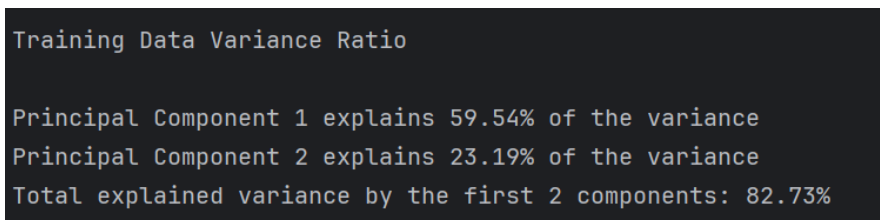
Fig. 7. Training Data Variance Ratio via PCA

The clustering patterns of the PCA model for mango ripeness classification are represented by a two-dimensional (2-D) scatter plot that shows individual mango samples as data points, with each color designated by numbers 0,1, and 2 representing different state of ripeness which are unripe, ripe, and overripe, respectively. This 2D plot reveals how the PCA model manages data based on the combined input gathered from particular MQ gas sensors. Data from the sensors are divided into three feature sets, namely:

Feature 1 which is the combination of MQ3, MQ4, and MQ7 (Fig. 8a)

Feature 2 which is the combination of MQ2, MQ8, and MQ135 (Fig. 8b)

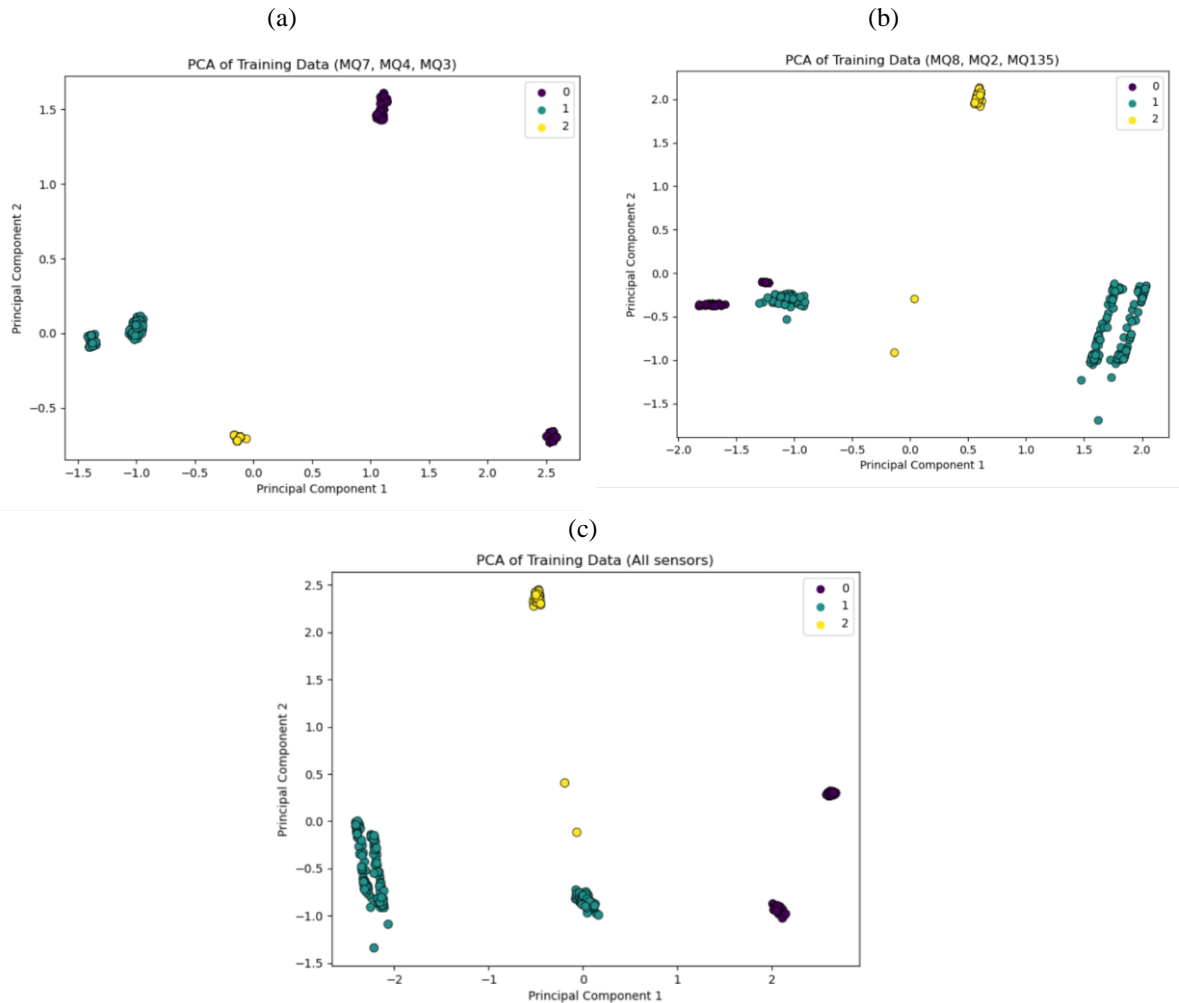Feature 3: which is the combination of all sensors. (Fig. 8c)

(a)   (b)

(c)

Fig. 8. 2-D graph plot of PCA clustering (a) Feature 1 (b) Feature 2 (c) Feature 3

3.2 Support Vector Machine (SVM)

Similar to the PCA clustering visualizations, the SVM model's decision boundaries for the ripeness classification are plotted in three independent 2-D graphs, as shown in Fig. 9. The decision boundaries created by the SVM model are reflected by the hyperplanes that divide the different ripeness classes based on the data collected from sensors.
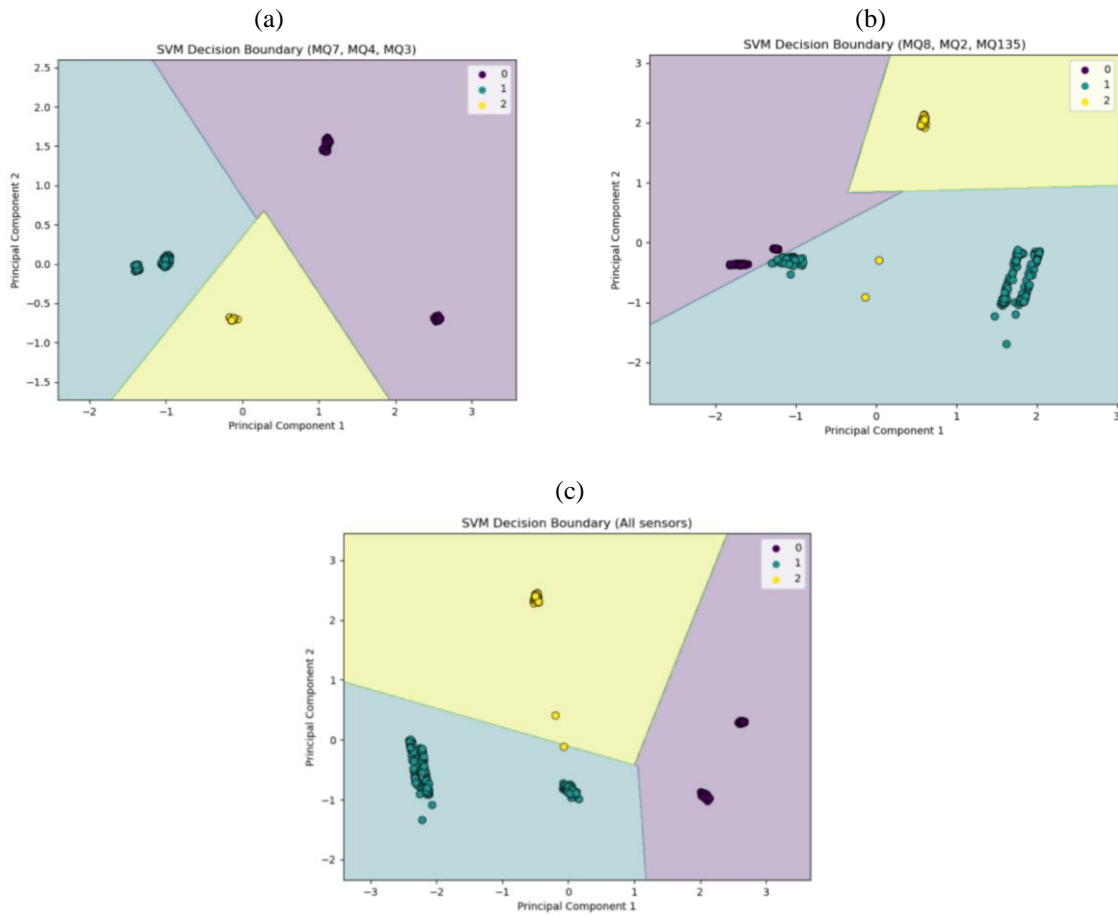
(a)   (b)

(c)

Fig. 9. 2-D graph plot of SVM Decision Boundaries for (a) Feature 1, (b) Feature 2, (c) Feature 3
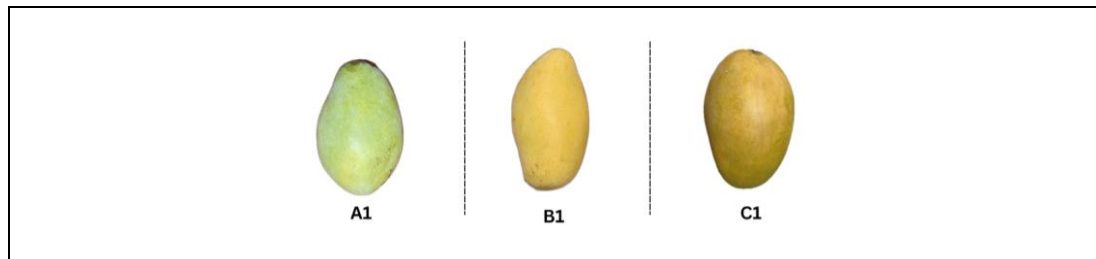
3.3 Classification

Using the testing data sets, the SVM model achieves a perfect accuracy of 100% in classifying the unripe category when utilizing all three features. This demonstrates that the model successfully learnt and incorporated the sensor data patterns to distinguish the unripe mangoes. However, both Feature 1 and Feature 2 failed to classify the ripe mango samples. This suggests that the sensor combinations did not offer enough information for the algorithm to correctly detect ripe mangoes. Nonetheless, the usage of Feature 3 resulted in an accuracy of 73% in detecting the ripe samples. Overall, as shown in Table 2, the accuracy of the e-nose system using the SVM model with Feature 3, which incorporates all six sensors, varies depending on the ripeness class. The system achieved accuracy rates of 100% for unripe, 73% for ripe, and 99% for overripe samples.

Table 2. The accuracy of SVM based on the three different sets

Ripeness	Feature 1	Feature 2	Feature 3
	MQ7, MQ4, MQ3	MQ2, MQ8, and MQ135	All sensors
Unripe	1.00	1.00	1.00
Ripe	0.00	0.00	0.73
Overripe	1.00	0.00	0.99

Three mango samples were used to test the mango ripeness classification using the Feature 3 model. As shown in Table 3, all three test samples—A1, B1, and C1—were accurately identified according to their respective ripeness stages. During the testing process, the sensor readings were recorded alongside the humidity and temperature values. The test was then repeated with another different set of three mango samples, yielding similar results.

Table 3. Datasets of the test samples

Sample	MQ2	MQ3	MQ4	MQ7	MQ8	MQ135	Humidity (%)	Temp (°C)	Outcome
A1	270	215	408	143	217	307	92	30.0	Unripe
B1	267	205	362	170	241	319	89	29.2	Ripe
C1	266	194	366	161	258	316	89	29.4	Overripe

4. CONCLUSION

In this work, 6 MQ sensor series from the MOS-type gas sensor was utilized to build an e-nose system, which has been successfully implemented in classifying the mango gold *susu* ripeness stages. In addition, the system's performance in the classification phase using machine learning techniques by identify the VOCs combination for each stage has been demonstrated using PCA and SVM. In conclusion, a comprehensive database of VOC profiles for mangoes at different ripening stages has been compiled, and a predictive model to evaluate the mango ripeness based on these VOC profiles has been developed using statistical learning approaches. Based on the Feature 3 model, the e-nose system has demonstrated that it can accurately classify various stages of mango ripeness, achieving precision comparable to that reported in previous studies.

REFERENCES

[1] "Tackling food loss and waste: A triple win opportunity," Newsroom. https://www.fao.org/newsroom/detail/FAO-UNEP-agriculture-environment-food-loss-wasteday-2022/en

[2] N. Aghilinategh, M. J. Dalvand, and A. Anvar, "Detection of ripeness grades of berries using an electronic nose," Food Sci Nutr, vol. 8, no. 9, pp. 4919-4928, Sep. 2020, https://doi.org/10.1002/fsn3.1788

[3] M. F. Mavi, Z. Husin, R. Badlishah Ahmad, Y. M. Yacob, R. S. M. Farook, and W. K. Tan, "Mango ripeness classification system using hybrid technique," Indonesian Journal of Electrical Engineering and Computer Science, vol. 14, no. 2, pp. 859-868, May 2019, https://doi.org/10.11591/ijeecs.v14.i2.pp859-868

[4] M. Baietto and A. D. Wilson, "Electronic-nose applications for fruit identification, ripeness and quality grading," Sensors (Switzerland), vol. 15, no. 1. MDPI AG, pp. 899-931, Jan. 06, 2015. https://doi.org/10.3390/s150100899

[5] Chiu, Shih-Wen, and Kea-Tiong Tang. 2013. "Towards a Chemiresistive Sensor-Integrated Electronic Nose: A Review" Sensors 13, no. 10: 14214-14247. https://doi.org/10.3390/s131014214