Optimizing Electronic Nose Performance for Detecting Coconut Sap Preservatives: A Comparative Analysis of Feature Extraction and Machine Learning Techniques

Yahya Efendi, Agus Naba* and Arinto Yudi Ponco Wardoyo

Department of Physics, Faculty of Mathematics and Natural Sciences, Universitas Brawijaya, Malang, Indonesia

(*Corresponding author's e-mail: anaba@ub.ac.id)

Abstract

Coconut sap, the raw material for coconut sugar, is highly susceptible to rapid fermentation, prompting farmers to use natural or chemical preservatives. Detecting these preservatives is challenging, as conventional techniques like Gas Chromatography-Mass Spectrometry (GC-MS) are costly and impractical for field applications. This study evaluated 3 feature extraction methods—maximum, difference, and integral—using an Electronic Nose (e-nose) to classify sap samples: Without preservatives (S-O (Original Sap)), with natural preservatives (S-NP (Sap with Natural Preservative)), and with chemical preservatives (S-CP (Sap with Chemical Preservative)). Data from ten Metal-Oxide Semiconductor sensors were analyzed using Principal Component Analysis (PCA) and 4 machine learning models: Random Forest (RF), Gradient Boosting (GB), Quadratic Discriminant Analysis, and k-Nearest Neighbor (k-NN). A stratified 5-fold cross-validation protocol was employed to ensure model robustness. The results demonstrated that the integral feature consistently outperformed the other methods, yielding superior PCA cluster separation and classification accuracy. The k-NN model with raw integral features achieved the highest test accuracy of 93.33% and a mean validation accuracy of 86.11%. Although GB with 2PC input also performed well (91.11% test accuracy, 87.78% validation), the k-NN model's misclassification pattern was safer for food safety, as it avoided labeling S-CP samples as S-O—a high-risk error. Sensors sensitive to organic solvent vapors and alcohols were the most significant contributors to detection accuracy. These findings confirm that integral feature extraction provides a reliable, rapid, and non-destructive method for preservative detection in coconut sap, offering a cost-effective alternative to GC-MS for quality control.

Keywords: Coconut sap, E-nose, Feature extraction, Machine learning, Preservative detection, Food safety

Introduction

Sugar, a fundamental commodity globally, plays an indispensable role in daily life, serving as a primary energy source and a versatile ingredient in countless food products. Beyond refined white sugar, alternative sweeteners such as coconut sugar have gained considerable prominence, particularly in rural communities and increasingly in urban markets due to growing consumer demand for natural and minimally processed foods. Coconut sugar is derived from coconut sap, a sweet, translucent liquid collected from the unopened inflorescence of the coconut palm (*Cocos nucifera*). This natural sap is highly valued not only for its distinct caramel-like flavor but also for its rich nutritional profile, which includes approximately 85.62% water, 13.64% sucrose, 0.04% reducing sugars, 0.17% amino acids, and 0.3% vitamin C [1-3]. Its inherent nutritional richness, however, also renders it highly susceptible to rapid microbial degradation, posing significant challenges to its quality and shelf-life.

Coconut sap exhibits physical and chemical characteristics that are strongly influenced by various factors, including environmental conditions, the cleanliness of tapping tools, and the time of collection [4]. These factors determine the sugar composition and

the content of Volatile Organic Compounds (VOCs), which play a crucial role in the aroma and flavor profile of the sap [5]. VOCs may undergo changes due to interactions with the surrounding environment or during storage, thereby accelerating quality degradation. Microbial activity is also a key factor: yeasts such as *Saccharomyces cerevisiae* convert sugars into ethanol, lactic acid bacteria produce lactic acid, and acetic acid bacteria oxidize ethanol into acetic acid [6-9]. The accumulation of these compounds leads to noticeable alterations in the taste, aroma, and color of coconut sap [10].

The inherent susceptibility of coconut sap to fermentation presents a critical challenge for its processing and commercialization. The sap must be processed immediately after collection. Otherwise, it undergoes rapid fermentation driven by naturally occurring microorganisms [11]. This process converts sugars into alcohol and organic acids, causing undesirable changes in taste, aroma, and color, which in turn reduces the sap's nutritional properties and safety [12-14]. As a result, the quality of the final coconut sugar is diminished, manifesting as less sweetness, a coarser texture, and reduced nutritional value [15].

To combat the rapid fermentation of coconut sap, farmers often employ various preservation methods. While traditional natural preservatives like jackfruit wood and mangosteen peel are commonly used and generally considered safe [8], there is a pervasive issue of farmers resorting to synthetic alternatives. These chemical additives, though effective in extending shelf life, pose significant health risks to consumers. Consequently, both preventive strategies and innovative preservation methods are crucial for sustaining coconut sap quality in industry [16,17]. For instance, exploring antimicrobial substances like extracts from plants such as the bark of *Hopea beccariana* has shown promise in maintaining sap integrity [18]. Additionally, optimizing harvesting time and techniques can improve sap quality by minimizing exposure to contaminants that accelerate fermentation [19]. Despite ongoing educational efforts to discourage the use of harmful chemicals, direct quality control of coconut sap remains limited, especially at the farm level, highlighting the need for more effective solutions to address the health concerns associated with synthetic preservatives.

Conventional methods like Gas Chromatography-Mass Spectrometry (GC-MS) are ill-suited for on-site coconut sap quality control, being costly, slow [20,21], and technically limited [1,22]. This necessitates rapid, affordable, non-destructive alternatives [2,12]. Electronic Nose (e-nose) technology offers a solution [19], utilizing an array of Metal-Oxide Semiconductor (MOS) sensors to detect VOCs in real-time [20,21]. Although susceptible to environmental factors [23,24], the e-nose process digitizes sensor signals [25,26], extracts features, and uses pattern recognition algorithms like Principal Component Analysis (PCA) for classification, providing a portable, cost-effective food safety tool.

The effectiveness of an e-nose system heavily relies on appropriate feature extraction techniques, which significantly enhance classification accuracy in complex matrices [27,28]. Robust methods are crucial for improving performance in diverse applications like quality control and disease detection [29-31]. While e-nose technology is well-established for food quality assessment, a research gap exists for identifying the optimal feature extraction method for detecting preservatives in coconut sap. Previous studies have not offered a comprehensive comparison of "maximum response," "difference," and "integral" techniques combined with robust machine learning validation in this specific context. This study aims to fill this gap by evaluating these 3 feature extraction methods in an e-nose system for coconut sap analysis. By combining these techniques with pattern recognition algorithms like PCA and machine learning, this research provides a focused comparative analysis to identify the most effective approach for enhancing detection accuracy.

This paper is structured as follows: "Related work" reviews prior studies. "Material and methods" details the sample preparation, e-nose setup, and algorithms for feature extraction and pattern recognition. The "Results" section presents sensor analysis, PCA, and classification performance. Finally, the "Discussion" and "Conclusions" interpret the findings and summarize the study's key contributions and implications.

Related work

The efficacy of e-nose systems in food quality detection hinges on feature extraction techniques. Prior

studies demonstrate this (**Table 1**), with Kombo *et al.* [32] achieving high accuracy in tea classification using piecewise features, and Chen *et al.* [33] classifying Chinese vinegar flavors with LSTM (Long Short-Term Memory). Advanced CNN (Convolutional Neural Network)-based methods by Zhai *et al*. [34] and Chen *et al*. [35] show impressive results in gas and chili classification. However, a gap exists in optimizing feature extraction for detecting preservative mixtures in coconut sap, which this study addresses.

Table 1 Related works on e-nose systems and feature extraction.

Topic	Method	Subject
High-quality tea classification [32]	Piecewise feature + line fitting	Testing accuracy 96.50%, sensitivity 98.60%
Flavor classification of Chinese vinegar [33]	Parallel LSTM (time-series)	Accuracy 95.8% using softmax; higher accuracy achieved using SVM
Industrial gas classification using MIGACN [34]	Multilevel group and temporal attention-based CNN	Achieved 98.19% accuracy with data augmentation
Chili variety and origin classification [35]	Sensor-aware CNN (SACNet)	Achieved 98.56, 97.43, and 99.31% accuracy on different datasets

Materials and methods

This section details the experimental design, sample preparation, e-nose device setup, data acquisition protocols, feature extraction algorithms, and pattern recognition techniques employed in this study. The methodological framework was designed to ensure reliability and reproducibility, aligning with academic research standards in sensor technology and food quality analysis.

Sample preparation

Coconut sap samples were meticulously collected from the Manggar Sari farmer group area in Wonoanti Village, Pacitan Regency, Indonesia (–8.200871553846548, 111.22681225261131). To ensure consistency and minimize variability in initial raw material quality, all sap samples were specifically tapped during the early morning hours between 05:00 and 06:00 AM. This standardized collection time is crucial as the composition of coconut sap can vary significantly throughout the day due to factors such as environmental temperature, humidity, and the palm's metabolic activity [36]. Such factors, along with the microflora present at different locations, can alter the sap's chemical profile, affecting aroma characteristics critical for accurate e-nose detection [37].

.

Following collection, the fresh coconut sap was systematically prepared under 3 distinct conditions to represent various preservation scenarios. These conditions were: Original sap without any added preservatives (S-O), sap mixed with natural preservatives (S-NP), and sap mixed with chemical preservatives (S-CP). The natural preservatives, widely utilized by local farmers, included extracts from jackfruit wood (*Artocarpus heterophyllus*) and mangosteen peel (*Garcinia mangostana*), which were traditionally prepared by local farmers. For the S-NP samples, 10 mL of the natural preservative extract was added to every 1 L of coconut sap. For the S-CP samples, an unbranded bulk soap made locally, commonly used by farmers in Pacitan Regency, was applied at a concentration of 5 g per 1 L of sap. The prepared samples were allowed to stand for 4 h to observe the effects of the preservatives. Prior to each measurement, 300 mL of the sample was transferred into a 600 mL bottle and shaken for 1 min to ensure homogeneity and facilitate the release of volatile compounds. For each testing session, 25 individual samples of 1 type were analyzed, with each sample measured 3 times, resulting in 75 data points per sensor for each sample type, as summarized in **Table 2**. This protocol yielded a total of 225 data points per sensor across all conditions.

Table 2 Three coconut sap samples used and their quantities

No.	Sample	Number of samples
1	S-O	25
2	S-NP	25
3	S-CP	25

The e-nose device setup

The e-nose device utilized in this study, shown in **Figure 1**, was custom-built and comprised an array of ten MOS sensors from the TGS series (Figaro Engineering Inc., Japan) and MQ series (Hanwei Electronics Co., China). These sensors were integrated into a custom-built chamber. The system was controlled using a microcontroller (Arduino Mega 2560, Arduino, Italy) and data acquisition was managed through UB-Nose software, developed by our laboratory. Environmental parameters were monitored with a DHT22 temperature and humidity sensor (Aosong Electronics Co., China). These sensors were selected for their specific sensitivities to various chemical compounds, as detailed in **Table 3**. MOS sensors operate on the principle of changing electrical resistance upon exposure to specific gases [38]. When target gas molecules interact with the semiconductor surface, a redox reaction with adsorbed oxygen ions alters the sensor's electrical resistance [39]. The magnitude and direction of this change depend on the gas type and sensor material characteristics [40].

Figure 1 The custom-built e-nose device used in this study, showcasing its physical components.

Table 3 MOS gas sensors used in e-nose with their specifications.

Sensor	Gas detection	Detection range
S1	Organic solvent vapors, flammable gases such as CO, LPG (Liquefied Petroleum Gas), alcohol	50 - 5000 ppm
S2	LPG, propane, hydrogen	200 - 10000 ppm
S3	Alcohol (ethanol), Benzine, Methane	0.05 - 10 mg/L alcohol
S4	Methane (CH_4), LPG	300 - 10000 ppm
S5	Carbon monoxide, methane, isobutane, ethanol	1 - 100 ppm
S6	LPG, isobutane, propane	300 - 10000 ppm
S7	Carbon monoxide; also responds to hydrogen, LPG, methane, alcohol	20 - 2000 ppm (CO)

Sensor	Gas detection	Detection range
S8	Ammonia (NH$_3$), organic amines like trimethylamine, monoethanolamine	5 - 500 ppm
S9	Carbon monoxide, combustible gases like methane and LPG	10 - 1000 ppm (CO), 200 - 10000 ppm (CH$_4$)
S10	Ammonia, sulfide, benzene, alcohol vapors, smoke, carbon dioxide (CO$_2$)	10 - 1000 ppm

Prior to data acquisition, the e-nose device was powered on for a 60-min stabilization period to ensure the sensor array reached its optimal operating temperature for consistent responses [41]. The data acquisition process was controlled by the UB-Nose software, which managed 2 main stages: Sample injection (5 s) and purging (35 s), as depicted in **Figure 2**. During injection (Part A), an inlet pump drew VOCs from the sample container into the sensor chamber. The gases interacted with the sensor array (Part B), causing resistance changes that were converted into digital signals and transmitted to a computer (Part C). Following this, the purging phase cleared residual VOCs from the chamber to return the sensors to their baseline state (Part D). An example of the resulting sensor response curve for an S-O sample is provided in **Figure 3**.
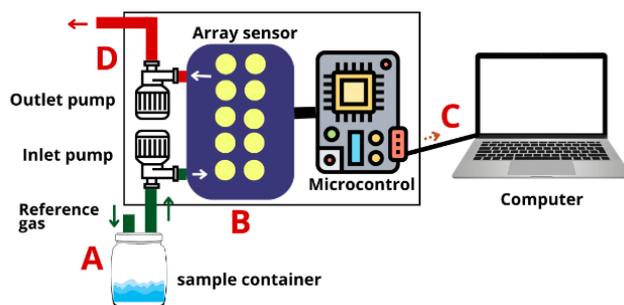

Figure 2 Diagram of the e-nose system workflow, illustrating the sequential stages of sample introduction, sensor interaction, signal processing, and chamber purging.
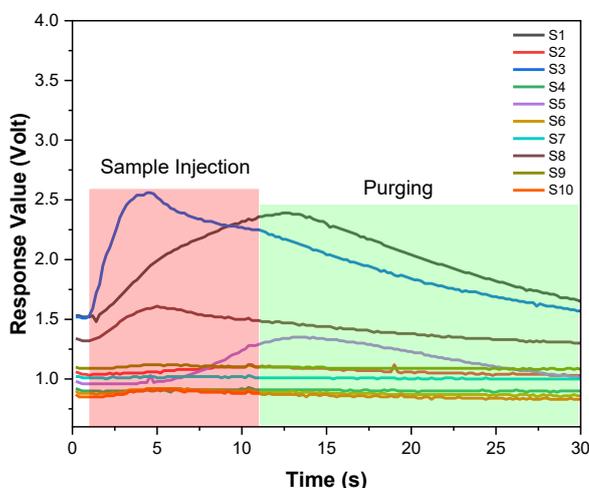

Figure 3 Example of e-nose sensor response curve from an S-O sample, showing the change in sensor resistance over time during the gas exposure and recovery phases, typically used for feature extraction.

Feature extraction algorithms

Raw sensor signals were processed using 3 distinct feature extraction techniques to derive meaningful information for pattern recognition. The maximum response method extracts the peak resistance value:

$$f_{\max} = \max\big(S(t)\big), \tag{1}$$

where $S(t)$ is the signal value at time t, capturing the strongest reaction to the sample's VOCs. The difference method calculates the change between the maximum response and the baseline:

$$f_{\text{diff}} = S_{\max} - S_{\text{base}}, \tag{2}$$

where $S_{\max}$ is the maximum signal value during exposure and S_{base} is the baseline signal value before gas exposure, quantifying the magnitude of the sensor's reaction relative to its resting state. Finally, the integral method computes the area under the sensor response curve over a defined period:

$$f_{\text{int}} = \int_{t_0}^{t_n} S(t)\ dt, \tag{3}$$

where t_0 is the start time and t_n is the end time of the measurement, representing the total sensor exposure to the volatile compounds.

Pattern recognition

The extracted features were used for pattern recognition to classify the coconut sap samples. PCA was employed for dimensionality reduction and visualization. Subsequently, 4 machine learning algorithms were applied for classification: Random Forest (RF), Gradient Boosting (GB), Quadratic Discriminant Analysis (QDA), and k-Nearest Neighbor (k-NN). The dataset was randomly divided into 2 subsets with a ratio of 80:20, corresponding to training and testing, respectively. To further ensure the robustness of model evaluation, a stratified 5-fold cross-validation procedure was applied to the training set (80%) for each model. Stratified cross-validation is a validation technique specifically designed to maintain the balance of class distribution in each fold, ensuring that the sample proportions in every subset accurately

represent the overall dataset [42]. This approach minimizes the risk of overfitting, reduces the influence of sampling bias, and provides a more reliable estimate of classification accuracy. All machine learning analyses were performed using Python (v.3.12) with the scikit-learn library (v.1.5.1) in the Google Colab environment.

Statistical analysis was conducted to evaluate the significance and reliability of the extracted features. Data normality was tested using the Shapiro-Wilk test. One-way ANOVA (Analysis of Variance) was applied for normally distributed data, while the Kruskal-Wallis test was used for non-parametric cases. Dimensionality reduction was performed with PCA to visualize sample clustering. Classification performance was evaluated through training, validation, and test datasets, with accuracy and confusion matrices used as key performance indicators. All statistical graphs were generated using Origin 2024 (OriginLab Corp., Massachusetts, USA).

Results and discussion

This chapter presents the findings of the study, beginning with an analysis of the individual sensor responses to different coconut sap samples. It then examines the influence of environmental factors on sensor stability, followed by a detailed evaluation of dimensionality reduction using PCA. Finally, the performance of various machine learning models for classifying the samples is assessed, leading to the selection of the optimal analytical approach.

Sensor response analysis

The average output voltage from each MOS sensor, as depicted in **Figure 4**, reveals the differential responses among the sample groups. The data shows significant variations for several sensors, while others had uniform responses. S3 consistently yielded the highest average values and demonstrated the most pronounced differentiation between samples. This indicates a high sensitivity to specific compounds in coconut sap, particularly alcohols, which aligns with its specifications [43]. The higher concentration of volatile compounds like ethanol in S-O and S-NP samples, resulting from natural fermentation, led to stronger signal responses from S3 compared to the S-CP samples

where fermentation was inhibited. As shown in **Table 3**, S3 is specifically sensitive to alcohol vapors such as ethanol. This characteristic explains its higher responses in S-O and S-NP samples, where natural fermentation

produced ethanol as the dominant volatile compound [6-8]. In contrast, the lower responses observed in S-CP samples indicate that chemical preservatives effectively inhibited the fermentation process.
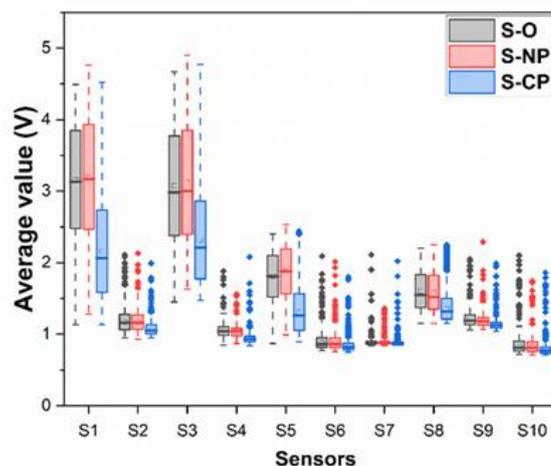

Figure 4 Average response values of each sensor, illustrating the distinct signal patterns across different coconut sap sample types. The figure highlights the varying sensitivities of individual MOS sensors to the volatile compounds emitted by each sample condition, providing a visual representation of their discriminative capabilities.

Sensors S1 and S5 also showed a clear signal distribution across the sample groups, demonstrating sensitivity to organic solvent vapors and ethanol. This explains their elevated responses in S-O and S-NP samples, where ongoing fermentation produced more organic compounds. Conversely, the lower signals in S-CP samples are attributed to the synthetic preservative inhibiting the formation of these volatiles. These chemical variations significantly impact the e-nose response patterns, as the production of alcohols and organic acids during fermentation alters the sap's aroma profile [6,10].

In contrast, sensors S4, S6, S7, S2, and S9 displayed uniform or overlapping distribution patterns, indicating lower discriminative capability for the key volatile compounds that differentiate the samples. These sensors are primarily designed for gases like propane and hydrogen, which were not dominant in the sap's volatile profile. S8 showed some response to amine compounds, potentially indicating protein degradation. This analysis underscores that a diverse sensor array is crucial for creating a holistic chemical fingerprint, with S1, S3, and S5 emerging as the key indicators for distinguishing between fresh, naturally fermented, and chemically preserved coconut sap [34].

Effect of temperature and humidity

Environmental factors like temperature and humidity are critical for the stability and accuracy of MOS sensors. As shown in **Figure 5(a)**, these conditions were carefully monitored within the measurement chamber. Stable conditions are paramount because fluctuations in temperature and humidity can alter a sensor's conductivity and the adsorption of volatile compounds, leading to signal drift and unreliable data.

Figure 5(b) depicts the average baseline shift, which is the gradual change in a sensor's signal over time. This drift can be caused by prolonged VOC exposure, temperature changes, or sensor aging, and can introduce significant errors if unaddressed. Standard mitigation practices include regular calibration and monitoring. The minimal baseline shift observed, along with the stable temperature and humidity, confirms that the experimental setup effectively controlled these variables, ensuring the reliability of the sensor data. This is consistent with previous studies that demonstrated fluctuations in ambient temperature and humidity significantly affect MOS sensor conductivity, often leading to signal drift if uncontrolled [25,39,44-48].

Consequently, regulating environmental factors and applying regular calibration have been regarded as standard practices in e-nose research to ensure reproducibility and sensor stability [32,49]. The minimal drift observed in this study supports these findings, confirming the effectiveness of the applied control measures.
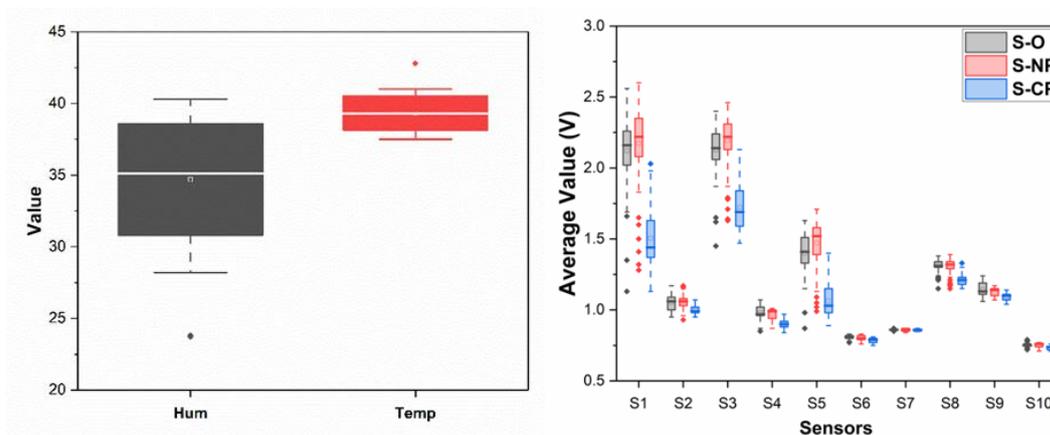

Figure 5 (a) Average temperature and humidity within the measurement chamber during the experimental process, highlighting the controlled environmental conditions maintained to ensure sensor stability and data consistency. (b) Average baseline shift observed during the sample measurement process, indicating the stability of the e-nose sensors over time and the effectiveness of environmental control in minimizing signal drift.

Dimensionality reduction with PCA

PCA was used to reduce the dimensionality of the e-nose data and visualize the clustering of the 3 coconut sap sample groups (S-O, S-NP, and S-CP), as shown in the 2D and 3D projections in **Figures 6** and **7**. Among the 3 feature extraction methods evaluated, the integral feature consistently provided the most distinct separation. In both visualizations, the clusters derived from the integral feature are compact and clearly separated, demonstrating their strong discriminative capability. This is because the integral method accumulates the entire sensor response over time, creating a comprehensive chemical fingerprint that enhances sensitivity. The effectiveness of this method is particularly evident in the 3D plot, where the additional dimension (PC3) further clarifies the spatial separation between the clusters.
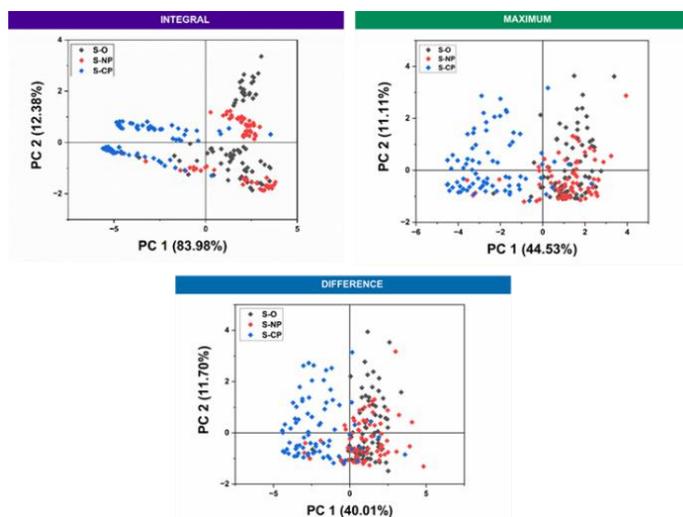

Figure 6 Two-dimensional PCA scatter plots showing the cluster separation of coconut sap samples based on 3 feature extraction methods.
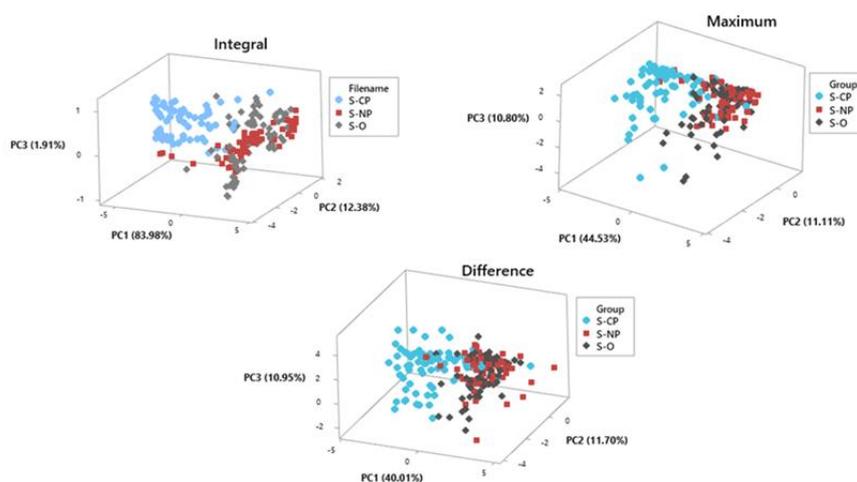
Figure 7 Three-dimensional PCA plots of coconut sap sample clusters using 3 different feature extraction methods.

In contrast, the maximum feature method resulted in a significant overlap between the S-O and S-NP clusters. This indicates that relying solely on the peak signal is less effective, as it overlooks crucial temporal information in the sensor's dynamic response. Even in the 3D projection, the overlap remains, confirming that this method lacks the detail needed for consistent class separation. Similarly, the difference feature performed poorly, yielding substantial overlap among all 3 sample groups. Capturing only the change from baseline to maximum response proved insufficient for distinguishing the complex volatile profiles of the sap samples, resulting in diminished discriminative power.

Classification performance of machine learning models

Prior to the classification stage using machine learning models, the extracted features were first subjected to statistical testing. Normality testing using the Shapiro-Wilk test indicated that none of the feature distributions met the assumption of normality ($p < 0.0001$ across all sensors). Therefore, the non-parametric Kruskal-Wallis test was employed to examine whether there were significant differences among the sample groups (S-O, S-NP, and S-CP) for each feature extraction method. The results demonstrated statistically significant differences across all sensors and feature extraction methods ($p < 0.0001$), confirming that the extracted features contain discriminative information. These findings validate the suitability of the features as input variables for the subsequent machine learning classification process.

The subsequent stage involved the classification of coconut sap samples based on the extracted features, using 4 different machine learning algorithms. The parameters for each machine learning model were configured as specified in **Table 4**. To evaluate model performance under various input scenarios, the machine learning inputs were organized and compared in 3 formats: 1) Raw data without dimensionality reduction, 2) Two principal components (2PC), and 3) Three principal components (3PC) obtained from the PCA process. This approach was intended to examine the impact of dimensionality reduction on classification performance. The results of the machine learning model evaluations are presented in **Table 5**.

Table 4 Parameters used for each machine learning model during the classification of coconut sap samples, detailing the specific configurations that optimized their performance.

No.	Model	Parameters
1	Random Forest	n_estimators = 50, max_depth = 5, min_samples_split = 2, min_samples_leaf = 1, bootstrap = True, random_state = 42
2	Gradient Boosting	n_estimators = 50, learning_rate = 0.1, max_depth = 2, subsample = 0.8, random_state = 42
3	QDA	reg_param = 0.01
4	k-NN	n_neighbors = 2, weights = 'uniform', metric = 'minkowski'

Table 5 Accuracy levels of machine learning models.

Feature extraction	Machine learning models	Accuracy (%)		
		Training	Validation (Mean ± Std)	Test
Integral	**GB (2PC)**	**97.22**	**87.78 ± 5.98**	**91.11**
	KNN (2PC)	**91.67**	**81.11 ± 5.39**	**91.11**
	QDA (2PC)	63.89	61.11 ± 9.78	62.22
	RF (2PC)	96.11	86.11 ± 5.83	88.89
	GB (3PC)	99.44	88.33 ± 4.08	88.89
	KNN (3PC)	**94.44**	**82.22 ± 4.51**	**91.11**
	QDA (3PC)	84.44	83.89 ± 4.08	86.67
	RF (3PC)	97.22	89.44 ± 4.44	88.89
	GB (RAW)	98.89	86.11 ± 3.93	88.89
	KNN (RAW)	**95.56**	**86.11 ± 3.93**	**93.33**
	QDA (RAW)	96.11	88.33 ± 4.78	84.44
	RF (RAW)	96.67	83.89 ± 5.93	88.89
Difference	GB (2PC)	86.67	67.22 ± 5.39	51.11
	KNN (2PC)	81.11	57.22 ± 5.15	53.33
	QDA (2PC)	61.67	57.78 ± 4.08	64.44
	RF (2PC)	85.00	63.33 ± 4.08	57.78
	GB (3PC)	90.56	67.22 ± 3.69	57.78
	KNN (3PC)	75.56	49.44 ± 4.08	55.56
	QDA (3PC)	62.22	58.89 ± 5.39	64.44
	RF (3PC)	88.33	62.78 ± 5.15	60.00
	GB (RAW)	95.56	72.22 ± 9.13	82.22
	KNN (RAW)	75.56	55.56 ± 4.3	57.78
	QDA (RAW)	72.22	58.33 ± 11.25	57.78
	RF (RAW)	93.33	72.78 ± 7.74	77.78

Feature extraction	Machine learning models	Accuracy (%)		
		Training	Validation (Mean ± Std)	Test
Maximum	GB (2PC)	88.89	73.33 ± 8.35	75.56
	KNN (2PC)	83.89	64.44 ± 3.69	73.33
	QDA (2PC)	61.67	59.44 ± 6.71	68.89
	RF (2PC)	86.67	70.0 ± 4.78	77.78
	GB (3PC)	92.78	70.56 ± 7.58	73.33
	KNN (3PC)	82.22	67.22 ± 4.44	62.22
	QDA (3PC)	65.56	58.89 ± 8.31	66.67
	RF (3PC)	90.00	65.56 ± 4.51	71.11
	GB (RAW)	97.22	82.22 ± 9.06	84.44
	KNN (RAW)	85.56	66.67 ± 6.33	71.11
	QDA (RAW)	77.78	67.78 ± 8.71	57.78
	RF (RAW)	95.56	82.22 ± 8.53	86.67

The classification results presented in **Table 5** indicate that the integral feature extraction method outperforms the maximum and difference approaches. The k-NN algorithm consistently emerged as the best-performing model when applied to integral features in raw, 2PC, and 3PC formats. In particular, the use of raw integral features enabled k-NN to achieve the highest test accuracy (93.33%), while its performance in the 2PC and 3PC cases also remained robust. This consistency suggests that the cumulative information captured by the integral method provides a more comprehensive representation of the volatile profiles in coconut sap compared to maximum or difference features, which tend to lose important temporal dynamics.

GB with integral 2PC input may also be considered, as it demonstrated relatively high accuracy and satisfactory validation performance. Nevertheless, its overall performance was inferior to that of k-NN, both in terms of accuracy and stability. This finding underscores that, although GB represents a strong candidate, the combination of integral features with k-NN remains the most optimal approach for preservative detection in coconut sap. By iteratively training and validating the models across 5 balanced folds, stable accuracy values were obtained while minimizing the effect of sampling bias. The k-NN model with raw integral features achieved a validation accuracy of 86.11% ± 3.93, reflecting low variability across folds. This outcome demonstrates that the model's performance is consistent and not overly dependent on any particular subset of the data.

Figure 8 presents the confusion matrices for the classification of coconut sap samples using GB (integral 2PC) and k-NN (integral raw data). Both models achieved high accuracy; however, differences were observed in the type of misclassification. GB incorrectly classified 1 S-CP sample as S-O, which can be considered a critical error in the context of food safety, as S-CP samples should never be identified as S-O. In contrast, the k-NN model misclassified 1 S-CP sample as S-NP, which is comparatively less critical since the sample was still recognized as containing preservatives, albeit of a different category. This distinction underscores an important practical implication: Although GB and k-NN achieved comparable accuracy, the misclassification pattern of k-NN is more conservative and therefore safer for preservative detection in coconut sap. Consequently, when both accuracy and food safety considerations are taken into account, the integral-kNN combination emerges as the most reliable model.
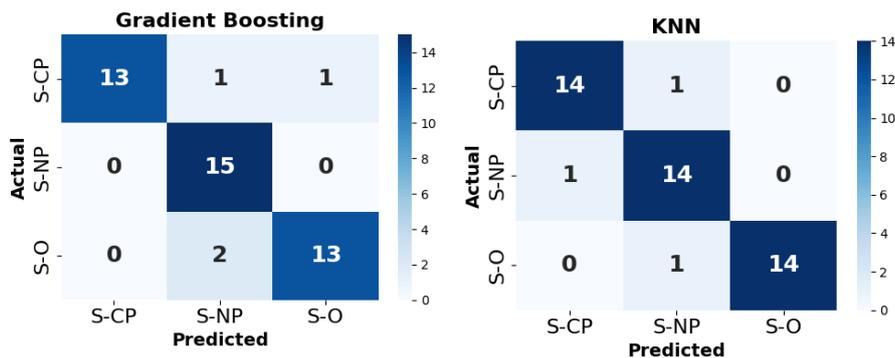
Figure 8 Confusion matrix for the classification of coconut sap samples using GB (integral 2PC) and k-NN (integral raw data). Both models are accurate, with differences in the location of misclassification within the S-CP class.

Discussion

The findings of this study confirm that the integral feature extraction method delivers the most effective performance in distinguishing the 3 types of coconut sap samples. PCA visualizations demonstrate that the integral feature produced the clearest cluster separation, thereby reflecting a more comprehensive chemical fingerprint compared to the maximum and difference features. This study further shows that the k-NN model with raw integral features consistently achieved the highest accuracy and stable validation performance, thereby confirming its reliability.

Interestingly, **Table 5** also reveals that using principal components as input features yielded mixed results. While some models, such as GB with 2PC input, showed an increase in test accuracy compared to their raw data counterparts, the best-performing model overall (k-NN) achieved its peak performance using the original raw features. This highlights a key consideration: PCA is an unsupervised dimensionality reduction technique that identifies directions of maximum variance, which may not align with the directions of maximum class separability and can discard valuable discriminative information [50]. This is critical in food safety applications, where the nature of misclassification is as important as overall accuracy. For instance, while the GB model with 2PC input had high accuracy, it critically misclassified S-CP as S-O. In contrast, the k-NN model's only error was classifying S-CP as S-NP, a more conservative and safer outcome.

Validation using stratified 5-fold cross-validation further supports these findings, as the small deviation indicates that the k-NN model has good generalization capability and is not overly dependent on specific data

partitions. Previous studies have also highlighted the effectiveness of stratified k-fold cross-validation in preventing overfitting and ensuring generalization, particularly in sensor-based classification tasks. For instance, stratified k-fold cross-validation was applied in the Gasformer model for lung cancer screening using e-nose breath analysis and reported a validation accuracy of $97.4 \pm 3.6\%$ [42]. In line with such findings, the relatively small standard deviation observed in our study further supports that the proposed integral-kNN model is both generalizable and stable across different data partitions.

It is important to note that this high performance was achieved under controlled laboratory conditions. A critical consideration for real-world deployment in uncontrolled farm or market settings is the impact of variable temperature and humidity, which are known to affect MOS sensor conductivity and cause signal drift [25,39,44]. While our setup successfully minimized this drift through environmental control, future applications must integrate robust compensation strategies. The inclusion of an environmental sensor (DHT22) in our system provides a foundation for developing such algorithms (e.g., humidity compensation models [41,44]) to correct readings in real-time, which is an essential next step towards field-ready deployment.

Overall, these results underscore that the integral-kNN combination represents the most reliable and optimal approach for preservative detection in coconut sap. These findings align with previous studies highlighting the strength of cumulative time-domain features such as integral over instantaneous or difference-based features [51]. While advanced methods—such as piecewise feature extraction [32],

feature discretization [52], or deep learning approaches like vision transformers [53] and ResNet-LSTM models [54]—offer high accuracy, they demand large datasets, high computational resources, and often reduce interpretability. In contrast, this study demonstrates that a simple and interpretable feature, such as the integral method, when combined with a robust classifier like k-NN, can achieve comparable accuracy while maintaining efficiency and transparency. This makes it a practical solution for rapid, non-destructive detection of preservatives in coconut sap, where reliability, interpretability, and food safety are paramount.

Conclusions

This study employed an e-nose system to evaluate the effectiveness of 3 feature extraction techniques—maximum, difference, and integral—in distinguishing preservative treatments in coconut sap. Among the tested approaches, the integral feature extraction method consistently outperformed maximum and difference methods, as demonstrated by clearer cluster separation in PCA visualizations and superior classification accuracy.

A key finding was that using raw features, rather than principal components from PCA, yielded the best results. The k-NN model with raw integral features achieved the highest test accuracy of 93.33%, supported by stable validation accuracy (86.11 ± 3.93%), confirming its robustness. While other models using PCA-derived features showed high performance, this study highlights that dimensionality reduction does not guarantee improved accuracy and can sometimes discard valuable discriminative information.

Furthermore, the choice of the k-NN model is reinforced by its safer misclassification profile. Unlike the GB model, which made the critical error of classifying S-CP as S-O, the k-NN model's only error was more conservative (S-CP as S-NP). This underscores the practical advantage of the integral-kNN combination for real-world food safety.

The proposed approach offers practical implications for food safety laboratories, small-scale producers, and regulatory bodies. Overall, this research establishes the integral-kNN framework with raw data as the most robust solution for rapid, non-destructive detection of chemical preservatives in coconut sap. While this study provides a strong proof-of-concept under controlled conditions, performance in variable field environments remains a key challenge. Future work should focus on expanding the dataset, exploring advanced feature extraction techniques like deep learning, and developing portable e-nose prototypes to enhance field-based monitoring.

Declaration of Generative AI in Scientific Writing

The authors acknowledge the use of generative AI tools (e.g., Grammarly and ChatGPT by OpenAI) in the preparation of this manuscript, specifically for language editing and grammar correction. No content generation or data interpretation was performed by AI. The authors take full responsibility for the content and conclusions of this work.

CRediT Author Statement

Yahya Efendi: Conceptualization, methodology, writing - original draft, investigation, visualization; **Agus Naba:** Writing - review and editing, data curation, supervision, validation; **Arinto Yudi Ponco Wardoyo:** Data curation, supervision, validation.

References

[1] MT Asghar, YA Yusof, MN Mokhtar, ME Ya'acob, HM Ghazali, LS Chang and YN Manaf. Coconut (*Cocos nucifera* L.) sap as a potential source of sugar: Antioxidant and nutritional properties. *Food Science and Nutrition* 2020; **8(4)**, 1777-1787.

[2] J Wiboonsirikul, P Ongkunaruk and P Poonpan. Determining key factors affecting coconut sap quality after harvesting. *Heliyon* 2024; **10(8)**, e29002.

[3] H Purnomo. Volatile components of coconut fresh sap, sap syrup and coconut sugar. *ASEAN Food*

Journal 2007; **14(1)**, 45-49.

[4]   KB Hebbar, M Arivalagan, KC Pavithra, TK Roy, M Gopal, KS Shivashankara and P Chowdappa. Nutritional profiling of coconut (*Cocos nucifera* L.) inflorescence sap collected using novel coco-sap chiller method and its value added products. *Journal of Food Measurement and Characterization* 2020; **14(5)**, 2703-2712.

[5]   Q Xia, R Li, S Zhao, W Chen, H Chen, B Xin, Y Huang and M Tang. Chemical composition changes of post-harvest coconut inflorescence sap during natural fermentation. *African Journal of Biotechnology* 2011; **10(66)**, 14999-15005.

[6]   N Udomsaksakul, K Kodama, S Tanasupawat and A Savarajara. Diversity of ethanol fermenting yeasts in coconut inflorescence sap and their application potential. *ScienceAsia* 2018; **44(5)**, 371-381.

[7]   N Udomsaksakul, K Kodama, S Tanasupawat and A Savarajara. Indigenous *Saccharomyces cerevisiae* strains from coconut inflorescence sap: Characterization and use in coconut wine fermentation. *Chiang Mai University Journal of Natural Sciences* 2018; **17(3)**, 219-230.

[8]   Karseno, Erminawati, T Yanto and I Handayani. The effect of coconut sap and skim milk concentration on physicochemical and sensory characteristic of coconut sap drink yogurt. *In*: Proceedings of the 2nd International Conference on Sustainable Agriculture for Rural Development, Purwokerto, Indonesia. 2021, p. 12045.

[9]   R Pandiselvam, MR Manikantan, SM Binu, SV Ramesh, S Beegum, M Gopal, KB Hebbar, AC Mathew, A Kothakota, R Kaavya and S Shil. Reaction kinetics of physico-chemical attributes in coconut inflorescence sap during fermentation. *Journal of Food Science and Technology* 2021; **58(9)**, 3589-3597.

[10]   BB Borse, LJM Rao, K Ramalakshmi and B Raghavan. Chemical composition of volatiles from coconut sap (*neera*) and effect of processing. *Food Chemistry* 2007; **101(3)**, 877-880.

[11]   JD Atputharajah, S Widanapathirana and U Samarajeewa. Microbiology and biochemistry of natural fermentation of coconut palm sap. *Food Microbiology* 1986; **3(4)**, 273-280.

[12]   Y Somawiharja, DM Wonohadidjojo, M Kartikawati, FRT Suniati and H Purnomo. Indigenous technology of tapping, collecting and processing of coconut (*Cocos Nucifera*) sap and its quality in Blitar Regency, East Java, Indonesia. *Food Research* 2018; **2(4)**, 398-403.

[13]   M Chinnamma, S Bhasker, MB Hari, D Sreekumar and H Madhav. Coconut Neera - A vital health beverage from coconut palms: Harvesting, processing and quality analysis. *Beverages* 2019; **5(1)**, 22.

[14]   R Somashekaraiah, B Shruthi, BV Deepthi and MY Sreenivasa. Probiotic properties of lactic acid bacteria isolated from neera: A naturally fermenting coconut palm nectar. *Frontiers in Microbiology* 2019; **10**, 1382.

[15]   J Wrage, S Burmester, J Kuballa and S Rohn. Coconut sugar (*Cocos nucifera* L.): Production process, chemical characterization, and sensory properties. *LWT* 2019; **112**, 108227.

[16]   A Saraiva, C Carrascosa, F Ramos, D Raheem, M Lopes and A Raposo. Coconut sugar: Chemical analysis and nutritional profile; health impacts; safety and quality control; food industry applications. *International Journal of Environmental Research and Public Health* 2023; **20(4)**, 3671.

[17]   L Sukumaran and M Radhakrishnan. Impact of nisin in combination with sodium benzoate and calcium carbonate on the bacterial and yeast population of coconut neera (*Coconut inflorescence* sap). *Journal of Pure and Applied Microbiology* 2021; **15(4)**, 2050-2058.

[18]   D Raharjo, MZ Zaman, D Praseptiangga and A Yunus. Physicochemical and microbiological characteristics of various stem bark extracts of *Hopea beccariana* Burck potential as natural preservatives of coconut sap. *Open Agriculture* 2023; **8(1)**, 20220175.

[19]   Mustaufik, L Sutiarso, S Rahayoe and KH Widodo. The effect of time and duration of tapping and the addition of laru as natural preservative in coconut sap quality. *In*: Proceedings of the 2nd International Conference on Sustainable Agriculture for Rural Development, Purwokerto, Indonesia. 2021, p. 12084.

[20]   SB Sulistyo and P Haryanti. Regression analysis for determination of antioxidant activity of

coconut sap under various heating temperature and concentration of lysine addition. *Food Research* 2020; **4(4)**, 976-981.

[21] P Haryanti, Karseno, I Handayani and SB Sulistyo. The chemical composition of coconut sap at different tapping condition. *AIP Conference Proceedings* 2023; **2586**, 60010.

[22] V Zulfia, M Ainuri, N Khuriyati, R Yusuf, AS Alim and U Pato. Optimizing of the parameters of coconut sugar production using taguchi design in Riau, Indonesia. *International Journal on Advanced Science, Engineering and Information Technology* 2022; **12(2)**, 752-758.

[23] Y Gan, T Yang, LP Guo, R Qiu, S Wang, Y Zhang, M Tang and Z Yang. Using HS-GC-MS and flash GC e-nose in combination with chemometric analysis and machine learning algorithms to identify the varieties, geographical origins and production modes of *Atractylodes lancea*. *Industrial Crops and Products* 2024; **209**, 117955.

[24] Z Gan, Q Zhou, C Zheng and J Wang. Challenges and applications of volatile organic compounds monitoring technology in plant disease diagnosis. *Biosensors and Bioelectronics* 2023; **237**, 115540.

[25] Y Huang, IJ Doh and E Bae. Design and validation of a portable machine learning-based electronic nose. *Sensors* 2021; **21(11)**, 3923.

[26] W Yao, H Wu, Y Cai, Y Chen, D Liu and M Zhang. Comprehensive analysis of geographical impact on flavor profiles of braised chicken across Eastern, Central, and Western China using GC-IMS, E-Nose techniques and sensory evaluation. *Food Chemistry Advances* 2024; **5**, 100808.

[27] W Xu, Y He, J Li, J Zhou, E Xu, W Wang and D Liu. Portable beef-freshness detection platform based on colorimetric sensor array technology and bionic algorithms for total volatile basic nitrogen (TVB-N) determination. *Food Control* 2023; **150**, 109741.

[28] C Avian, JS Leu, SW Prakosa and M Faisal. An improved classification of pork adulteration in beef based on electronic nose using modified deep extreme learning with principal component analysis as feature learning. *Food Analytical Methods* 2022; **15(11)**, 3020-3031.

[29] Z Li, T Wang, H Jiang, WT Wang, T Lan, L Xu, YH Yun and W Zhang. Comparative key aroma compounds and sensory correlations of aromatic coconut water varieties: Insights from GC×GC-O-TOF-MS, E-nose, and sensory analysis. *Food Chemistry: X* 2024; **21**, 101141.

[30] DR Wijaya, F Afianti, A Arifianto, D Rahmawati and VS Kodogiannis. Ensemble machine learning approach for electronic nose signal processing. *Sensing and Bio-Sensing Research* 2022; **36**, 100495.

[31] MA Khan, I Ashraf, M Alhaisoni, R Damaševičius, R Scherer, A Rehman and SAC Bukhari. Multimodal brain tumor classification using deep learning and robust feature selection: A machine learning application for radiologists. *Diagnostics* 2020; **10(8)**, 565.

[32] KO Kombo, N Ihsan, TS Syahputra, SN Hidayat, M Puspita, Wahyono, R Roto and K Triyana. Enhancing classification rate of electronic nose system and piecewise feature extraction method to classify black tea with superior quality. *Scientific African* 2024; **24(14)**, e02153.

[33] Y Chen, J Fu, X Weng, J Chen, R Hu and Y Zhu. A feature extractor for temporal data of electronic nose based on parallel long short-term memory network in flavor discrimination of Chinese vinegars. *Journal of Food Engineering* 2024; **379**, 112132.

[34] S Zhai, Z Li, H Zhang, L Wang, S Duan and J Yan. A multilevel interleaved group attention-based convolutional network for gas detection via an electronic nose system. *Engineering Applications of Artificial Intelligence* 2024; **133**, 108038.

[35] Y Chen, X Wang, W Yang, G Peng, J Chen, Y Yin and J Yan. An efficient method for chili pepper variety classification and origin tracing based on an electronic nose and deep learning. *Food Chemistry* 2025; **479**, 143850.

[36] Mustaufik, L Sutiarso, S Rahayu and KH Widodo. Technique engineering of tapping and shelter of coconut sap and its effect on the quality of crystal coconut sugar. *Food Research* 2022; **6(2)**, 248-254.

[37] W Chen, Q Zhu, Q Xia, W Cao, S Zhao and J Liu. Reactive oxygen species scavenging activity and DNA protecting effect of fresh and naturally fermented coconut sap. *Journal of Food Biochemistry* 2011; **35(5)**, 1381-1388.

[38] YF Sun, SB Liu, FL Meng, JY Liu, Z Jin, LT Kong and JH Liu. Metal oxide nanostructures and their gas sensing properties: A review. *Sensors* 2012; **12(3)**, 2610-2631.

[39] Z Jiang, P Xu, Y Du, F Yuan and K Song. Balanced distribution adaptation for metal oxide semiconductor gas sensor array drift compensation. *Sensors* 2021; **21(10)**, 3403.

[40] A Khorramifar, M Rasekh, H Karami, U Malaga-Toboła and M Gancarz. A machine learning method for classification and identification of potato cultivars based on the reaction of mos type sensor-array. *Sensors* 2021; **21(17)**, 5836.

[41] M Yan, Y Wu, Z Hua, N Lu, W Sun, J Zhang and S Fan. Humidity compensation based on power-law response for MOS sensors to VOCs. *Sensors and Actuators B: Chemical* 2021; **334**, 129601.

[42] Y Lin, J Jing, Y He, X Wang, A Zhong, W Ye, W Xu, X Zhao and X Pan. A fast, non-invasive auxiliary screening algorithm for lung cancer based on electronic nose system. *Sensors and Actuators A: Physical* 2025; **389**, 116490.

[43] M Gopal, S Shil, A Gupta, KB Hebbar and M Arivalagan. Metagenomic investigation uncovers presence of probiotic-type microbiome in Kalparasa® (Fresh Unfermented Coconut Inflorescence Sap). *Frontiers in Microbiology* 2021; **12**, 662783.

[44] F Lu and J Zhang. Drift compensation of the gas sensor based on self-training and semi-supervised learning. *In*: Proceedings of the 2nd International Conference on Artificial Intelligence, Big Data and Algorithms, Nanjing, China. 2022, p. 799-803.

[45] S Liu, X Chen, X Xia, Y jin, G Wang, H Jia and D Huang. Electronic sensing combined with machine learning models for predicting soil nutrient content. *Computers and Electronics in Agriculture* 2024; **221**, 108947.

[46] Y Sun and Y Zheng. A method of gas sensor drift compensation based on intrinsic characteristics of response curve. *Scientific Reports* 2023; **13(1)**, 11971.

[47] X Dong, H Duan, X Xu and S Han. A novel memory mechanism for postponing the drift of chemical gas sensors. *In*: Proceedings of the 2021 IEEE 11th International Conference on Electronics Information and Emergency Communication, Beijing, China. 2021, p. 1-4.

[48] H Se, K Song, H Liu, W Zhang, X Wang and J Liu. A dual drift compensation framework based on subspace learning and cross-domain adaptive extreme learning machine for gas sensors. *Knowledge-Based Systems* 2023; **259**, 110024.

[49] Chotimah, K Saifullah, FN Laily, M Puspita, KO Kombo, SN Hidayat, ET Sulistyani, Wahyono and K Triyana. Electronic nose-based monitoring of vacuum-packaged chicken meat freshness in room and refrigerated storage. *Journal of Food Measurement and Characterization* 2024; **18(10)**, 8825-8842.

[50] J Lever, M Krzywinski and N Altman. Principal component analysis. *Nature Methods* 2017; **14(7)**, 641-642.

[51] J Yan, X Guo, S Duan, P Jia, L Wang, C Peng and S Zhang. Electronic nose feature extraction methods: A review. *Sensors* 2015; **15(11)**, 27804-27831.

[52] Y Shi, H Lin, Y Yu, C Yin and Y Wang. A gas-spectral bimodal information fusion method combining electronic nose and hyperspectral system to identify the rice quality in different storage periods. *IEEE Transactions on Instrumentation and Measurement* 2024; **73**, 2526611.

[53] G Hu, B Du, X Wang and G Wei. An enhanced black widow optimization algorithm for feature selection. *Knowledge-Based Systems* 2022; **235**, 107638.

[54] G Wei, X Liu, A He, W Zhang, S Jiao and B Wang. Design and implementation of a ResNet-LSTM-Ghost architecture for gas concentration estimation of electronic noses. *IEEE Sensors Journal* 2024; **24(16)**, 26416-26428.