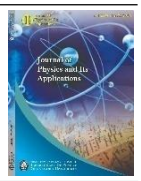
Journal of Physics and Its Applications

Journal homepage : https://ejournal2.undip.ac.id/index.php/jpa/index



Environmental Compensation for Robust Tea Aroma Classification Using an Electronic Nose and Machine Learning

Budi Sumanto[1*], Ummi Kaltsum[2], Galih Setyawan[1], Ganjar Alfian[1], Dzulkifli Daeng Syauqi[1], Ariesta Martiningtyas Handayani[1], and Kombo Othman Kombo[3]

[1]*Department of Electrical Engineering and Informatics, Vocational College, Universitas Gadjah Mada, Yogyakarta, Indonesia*
[2]*Department of Physics Education, Faculty of Mathematics, Natural Sciences, and Information Technology Education, Universitas Persatuan Guru Republik Indonesia, Semarang, Indonesia*
[3]*Department of Natural Sciences, College of Science and Technical Education, Mbeya University of Science and Technology, Mbeya, Tanzania*

**Corresponding author: budi.sumanto@ugm.ac.id*

ARTICLE INFO

Article history:
Received: 20 July 2026
Accepted: 26 August 2026
Available online: 31 August 2026

Keywords:
Electronic nose
MOS sensor
Tea aroma classification
Environmental compensation
Machine learning

ABSTRACT

Classifying tea aromas using an electronic nose (e-nose) system offers rapid, non-destructive quality assessment. However, metal oxide semiconductor (MOS)-based gas sensors are often affected by temperature and humidity, reducing classifier robustness. To address this, we propose an environmental compensation approach to boost robustness in machine learning-based tea aroma classification. Specifically, we analyzed 400 tea samples (100 per class: black, green, red, yellow) using an e-nose with 10 MOS sensors under three scenarios: (A) sensor features only, (B) integration of temperature–humidity features, and (C) temperature–humidity-based signal compensation before feature extraction. For classification, we used SVM, Random Forest, KNN, and a soft voting ensemble. Notably, Scenario C performed best, achieving 85.00% accuracy with SVM. Furthermore, robustness analysis revealed that KNN led on test data (RI=0.9851), while SVM was perfect in cross-validation (RI=1.000). These results confirm that environmental compensation effectively improves the MOS e-nose system stability.

1. Introduction

Tea is enjoyed around the world for its aroma, taste, and health benefits. Depending on how much the leaves are oxidized or fermented, tea is usually grouped into green, yellow, red, and black varieties. Each type has its own unique flavor because of the way it is processed and the changes that happen during fermentation [1],[2]. The aroma of tea is an important factor in judging its quality and value.

Conventional tea quality evaluation relies on laboratory analysis or expert sensory testing [3]. While accurate, these methods are costly, time-consuming, and subjective [4]. Thus, faster, more objective, and non-destructive alternatives are needed for tea quality assessment.

The electronic nose (e-nose), which mimics human smell using gas sensor arrays and pattern recognition algorithms [5],[6], is a rapidly evolving technology for aroma analysis. It has been widely used to assess the quality of food and beverages such as coffee, milk, meat, and tea [7]. Using machine learning methods, the e-nose can automatically classify odors based on multidimensional response patterns from gas sensors [8]-[10].

Various classification algorithms, such as Support Vector Machine (SVM), Random Forest (RF), K-Nearest Neighbor (KNN), and ensemble methods, have been widely used to improve the performance of e-nose-based aroma classification [11],[12]. However, improvements in classification accuracy are not only influenced by the choice of classifier but also depend heavily on the quality of the sensor signals used as model input.

Recent research into the use of e-noses for classifying tea has also shown how important it is to represent signal features and to carry out feature optimization. It has been found that an optimized combination of statistical and dynamic sensor response features helps to improve the ability to distinguish between fermented and non-fermented teas, and that the use of mean, median, and gradient features together with Linear Discriminant Analysis (LDA) results in an accuracy of 85% and an AUC of 0.97 [13]. These results show that proper feature representation and selection can greatly assist in the discrimination between different tea aroma categories. Yet this method mainly concentrated on feature optimization without looking into or compensating for the effect of environmental temperature and humidity on MOS sensor responses.

Sensor responses in metal-oxide-semiconductor (MOS)-based electronic nose tea classification systems are highly sensitive to environmental conditions, especially temperature and relative humidity [14]. Fluctuations in these parameters during measurement may result in baseline shifts, response drift, and signal pattern distortion, which compromise the stability and robustness of classification models [15],[16]. Therefore, although previous tea classification studies have demonstrated the effectiveness of feature optimization [13], the environmental dependence of MOS sensor responses remains an important issue when classification models are expected to operate robustly under

varying measurement conditions. Despite this, the majority of existing studies emphasize feature extraction and classifier optimization, with limited attention given to environmental compensation. Consequently, the influence of environmental compensation on classification robustness remains largely unexplored [17].

In particular, there is still limited research comparing the effectiveness of integrating temperature–humidity features with direct compensation at the sensor signal level to improve the robustness of tea aroma classification. However, understanding the most effective environmental compensation strategies is crucial for developing e-nose systems that are more stable under real-world operating conditions. Accordingly, rather than further optimizing feature combinations alone, this study shifts the analytical focus toward environmental robustness by explicitly evaluating how temperature and humidity information can be incorporated at different stages of the classification pipeline.

To address these challenges, this study seeks to enhance the robustness of tea aroma classification by employing a metal-oxide semiconductor (MOS)-based electronic nose (e-nose) system, utilizing environmental compensation and machine learning techniques. Three experimental scenarios are evaluated. First, classification is performed based solely on sensor features. Next, temperature and humidity features are integrated at the feature level. Finally, temperature and humidity compensation are applied at the signal level before feature extraction. Through this comparative framework, the main contribution of this study is to determine whether environmental information is more effective when introduced as additional classification features or when used directly to compensate sensor responses before feature extraction, while also evaluating the resulting classification stability across different machine learning models.

2. Method

2.1. Tea Samples

Four types of tea, each representing a different level of fermentation green tea, yellow tea, red tea, and black tea were selected for this study. Variations in processing produce distinct volatile compound compositions and aroma profiles, which are appropriate for assessing the electronic nose system's classification capabilities. The characteristics of each sample are summarized in Table 1 and illustrated in Fig 1.

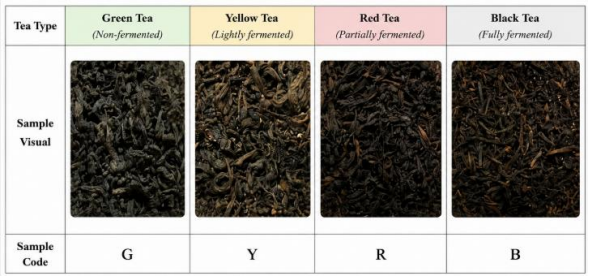


Fig. 1: Visual appearance and sample coding of the four tea categories used in this study

For each type of tea, 100 samples were collected, resulting in a total dataset of 400 samples. All samples were obtained from commercial products in good condition and stored in sealed containers at room temperature to maintain aroma stability during the measurement process.

Table 1: Description of the Tea Samples Used in the Study

Tea Type	Fermentation Level	General Aroma Characteristics
Green Tea	Non-Fermented	Fresh, vegetal
Yellow Tea	Lightly Fermented	Mild, Floral
Red Tea	Semi Fermented	Fruity, Sweet
Black Tea	Fully Fermented	Strong, Earthy

2.2. Electronic nose system

The electronic nose (e-nose) system used consists of a sample acquisition chamber, a metal-oxide-semiconductor (MOS)-based gas sensor array, a microcontroller-based data acquisition unit, and a computer for data storage and processing, as shown in Fig 2.

Table 2: Sensor Configuration in an Electronic Nose System

Sensor	Description
S1	Carbon monoxide, ethanol, hydrogen, isobutane, and methane
S2	Ammonia, ethanol, hydrogen, hydrogen sulfide, and toluene
S3	Ethanol, hydrogen, isobutane, and methane
S4	Carbon monoxide, ethanol, hydrogen, isobutane, and methane
S5	Carbon monoxide, ethanol, hydrogen, isobutane, methane, and propane
S6	Carbon monoxide, ethanol, hydrogen, isobutane, methane, and propane
S7	Carbon monoxide, ethanol, hydrogen, and methane
S8	Acetone, benzene, carbon monoxide, ethanol, isobutane, methane, and n-hexan
S9	Ammonia, ethanol, hydrogen, and isobutane
S10	Chlorofluorocarbons, ethanol, and hydrofluorocarbons
Temperature, Humidity	

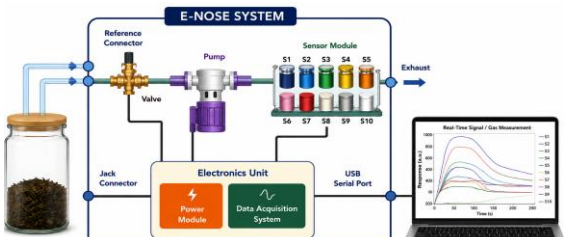


Fig. 2: Schematic diagram of the electronic nose system used for tea aroma measurement.

The sensor array comprises ten gas sensors with distinct sensitivities to volatile organic compounds (VOCs), generating a unique multidimensional response pattern and an aroma fingerprint for each tea type, as presented in Table 2. Additionally, the system incorporates temperature and humidity sensors to track environmental conditions, given that metal-oxide semiconductor (MOS) sensor responses

are significantly affected by variations among these parameters.

2.3. Data Acquisition Procedure

Data acquisition is initiated by placing a tea sample in the electronic nose measurement chamber. Each measurement cycle consists of baseline, sampling, and recovery phases.

During the baseline phase, sensors are exposed to clean air for 5 seconds to record the initial response. In the sampling phase, tea aroma is introduced for 60 seconds, resulting in changes to each gas sensor's response. The recovery phase then applies clean air for 240 seconds to restore the sensors prior to the subsequent measurement.

All gas, temperature, and humidity sensor responses were continuously recorded using data acquisition software with a sampling interval of 100 ms. With this configuration, each sample produced a time-series dataset of dimensions 3050 × 10, representing 3050 measurement points for 10 gas sensor outputs. For further analysis, only the signals from the active sensing phase (60 seconds) are used as input for the preprocessing and feature extraction stages.

2.4. Signal Preprocessing and Feature Extraction

The output from the electronic nose consists of time-series signals from 10 gas sensors, which include baseline offset, noise, and response variations. Before classification, these signals undergo preprocessing steps: baseline correction, windowing, and normalization.

In the baseline correction stage, each sensor signal is adjusted against the average response during the baseline phase to minimize offset and noise. This correction follows the equation 1.

$$x_c(t) = x(t) - \bar{x}_{\text{baseline}} \quad (1)$$

Where $x(t)$ represents the original sensor signal, $x_c(t)$ represents the corrected signal, and $\bar{x}_{\text{baseline}}$ represents the average sensor response during the baseline phase, calculated as equation 2:

$$\bar{x}_{\text{baseline}} = \frac{1}{n}\sum_{i=1}^{N} x_i \quad (2)$$

where N is the number of data points in the baseline phase. In this study, the baseline phase lasted 5 seconds with a sampling interval of 100 ms, resulting in 50 baseline data points for each sensor.

After baseline correction, a windowing process was performed to extract the most informative portion of the signal. This selection focused on the active sensing phase, specifically from 5 to 65 seconds. During this period, the tea aroma interacts directly with the sensor array. Window selection aims to focus the analysis on sensor responses most relevant to the sample's aroma characteristics.

The windowed signal was then normalized using the z-score normalization method. This step reduced scale differences between sensors, as expressed equation 3:

$$x_n = \frac{x - \mu}{\sigma - \epsilon} \quad (3)$$

where μ is the mean of the signal, σ is the standard deviation, and ϵ is a small constant used to avoid division by zero.

After the preprocessing stage was completed, the sensor signals were extracted into statistical and dynamic features that represent the aroma characteristics of the tea. The features used include the mean, maximum value, standard deviation, skewness, kurtosis, slope, and area under the curve (AUC), as summarized in Table 3. With 10 sensors and 7 features for each sensor, each sample yielded a total of 70 features.

Table 3: Statistical and Dynamic Features Used in the Study

Feature	Formula	Description
Mean	$\frac{1}{N}\sum x_i$	Average sensor response
Maximum	Max(x)	Peak response
Std	$\sqrt{\frac{1}{N-1}\sum(x_i-\mu)^2}$	Signal dispersion
Skewness	$\frac{1}{N}\sum_{i=1}^{N}(\frac{x_i-\mu}{\sigma})^3$	Distribution asymmetry
Kurtosis	$\frac{1}{N}\sum_{i=1}^{N}(\frac{x_i-\mu}{\sigma})^4-3$	Excess kurtosis
Slope	$\frac{x_{end}-x_{start}}{\Delta t}$	Response rate
AUC	$\int x(t)dt$	Total response energy

To reduce redundancy and retain the most discriminatory features, feature selection was performed using the ANOVA F-test (SelectKBest) method. Based on the selection results, the top 20 features were chosen as input for the machine learning classification stage.

2.5. Experimental Scenario

To analyze the effect of environmental parameters on tea aroma classification performance, this study designed three different experimental scenarios. These three scenarios aim to evaluate the extent to which temperature and humidity data can improve the robustness of the electronic nose system against variations in environmental conditions. A summary of the experimental scenarios is shown in Table 4.

Table 4: Experimental Scenarios

Scenario	Input	Objective
A	20 sensor features	Baseline classification
B	20 sensor features + T + H	Environmental feat integration
C	20 compensated sensor features	Environmental compensation

Scenario A: Sensor Features Only

Scenario A is used as the baseline model, in which the classification process uses only features extracted from the signals of 10 gas sensors, without taking environmental parameters into account. Thus, the feature vector in this scenario consists solely of sensor features.

$$F_A = F_{sensor} \quad (4)$$

This scenario is used to evaluate the basic ability of an electronic nose to distinguish tea aromas based solely on gas sensor responses.

Scenario B: Integration of Temperature and Humidity Sensor Features

In this scenario, the temperature (T) and relative humidity (H) parameters are explicitly added to the feature vector extracted from the sensors. The feature vector in this scenario is expressed as:

$$F_B = [F_{sensor}, T, H] \quad (5)$$

This approach is used to test the hypothesis that environmental information can help machine learning models understand changes in sensor responses influenced by temperature and humidity, thereby improving classification stability.

Scenario C: Temperature and Humidity-Based Compensation

Scenario C applies a direct compensation approach to the sensor signal before feature extraction is performed. Compensation is performed by normalizing the sensor response using temperature and humidity parameters.

$$x_{comp}(t) = \frac{x(t)}{(T+\epsilon)(H+\epsilon)} \quad (6)$$

where *x(t)* represents the original sensor response, and ϵ is a small constant used to avoid division by zero. This scenario aims to evaluate whether direct signal compensation is more effective at reducing environmental effects than the integration of explicit features in Scenario B.

2.6. Machine Learning Models and Ensemble Strategies

We used Support Vector Machine (SVM), Random Forest (RF), and K-Nearest Neighbor (KNN) as base classifiers due to their complementary classification characteristics.

SVM offers good generalization and handles nonlinear problems with kernels. RF is robust to noise and captures feature relationships using multiple trees. KNN classifies samples based on feature proximity.

To obtain the optimal model configuration, hyperparameter tuning was performed using the Grid Search method with 10-fold cross-validation for each classifier. The optimized parameters included the C and gamma values for SVM, the number of estimators and maximum depth for RF, and the number of neighbors for KNN.

In addition to evaluating individual models, this study also applied an ensemble learning approach using the soft voting method to improve the stability and robustness of the classification. In the soft voting ensemble method, each classifier predicts probabilities for each class. These probabilities are averaged across all classifiers, and the class with the highest average probability is selected as the final decision.

$$\hat{y} = \arg max \sum_{m=1}^{M} w_m P_m(y|x) \quad (7)$$

where P_m (y|x) represents the class prediction probability of the *mth* classifier, and w_m is the weight of each classifier. In this study, all classifiers were assigned the same weight, so that each model contributed equally to the final decision.

2.7. Model Performance Evaluation and Robustness Analysis

Model performance evaluation was conducted using the hold-out validation method with a stratified 80:20 data split, in which 80% of the data was used as training data and the remaining 20% as test data. The stratified split approach was used to ensure that the proportion of samples in each class remained balanced between the training and test data.

Model performance was assessed using several classification metrics: accuracy, precision, recall, and F1-score. Accuracy measures the proportion of correct predictions relative to the total number of samples and is calculated as follows:

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (8)$$

where TP, TN, FP, and FN represent true positive, true negative, false positive, and false negative, respectively.

In addition, a confusion matrix was used to analyze classification error patterns among tea classes. The evaluation was further supported by the Receiver Operating Characteristic (ROC) curve and the multiclass Area Under the Curve (AUC) to assess the model's ability to distinguish between each class.

To evaluate the model's generalization ability, 10-fold cross-validation was performed, in which the dataset was divided into 10 subsets, and the model was trained and tested alternately on each fold. For each fold, feature standardization and ANOVA feature selection were performed within the evaluation pipeline to prevent data leakage. If an ensemble method was used, its procedure, including model selection, aggregation, and prediction combination, was also integrated within each cross-validation iteration to ensure unbiased evaluation.

In addition to conventional performance evaluations, this study also proposes an additional metric, the Robustness Index (RI), to measure the stability of the model's performance across the three experimental scenarios. The Robustness Index is calculated using the following equation:

$$RI = 1 - \frac{\sigma_{acc}}{\mu_{acc}} \quad (9)$$

where σ_{acc} and μ_{acc} represent the standard deviation and mean of the model's performance, respectively, across the three scenarios.

3. Result and Discussion

3.1. Sensor Response Characteristics of Tea Aroma

The sensory response characteristics of the gas sensor to tea aroma were analyzed to evaluate the electronic nose system's ability to detect the aroma patterns produced by the tea samples. Examples of the electronic nose output signals before and after the

preprocessing stage are shown in Fig. 3(a) and Fig. 3(b).

Fig. 3(a) shows that the raw signals from the 10 gas sensors exhibit three distinct measurement phases: baseline, sampling, and recovery. The transition from baseline to sampling begins precisely at the moment the sensors are first exposed to the analyte, initiating a noticeable change in sensor response. The start of the recovery phase occurs exactly when the analyte is removed from the environment, leading the sensor responses to return toward baseline values. During the baseline phase, sensors display stable responses despite significant initial offset differences, with initial values ranging from approximately 600 to 3700, which reflect the intrinsic sensitivity of each MOS sensor.

When the tea aroma began to be introduced into the sensor chamber during the sampling phase, most sensors exhibited a significant increase in response due to the interaction between the tea's volatile compounds and the sensor surface. However, the response amplitudes varied among the sensors. Sensors S2, S8, and S10 exhibited the most dominant responses with the highest response amplitudes, while sensors S5, S7, and S9 showed relatively small changes in response. These differences indicate that each sensor has a different sensitivity to volatile organic compounds (VOCs), so the combined responses of all sensors form a distinctive multidimensional pattern (aroma fingerprint).

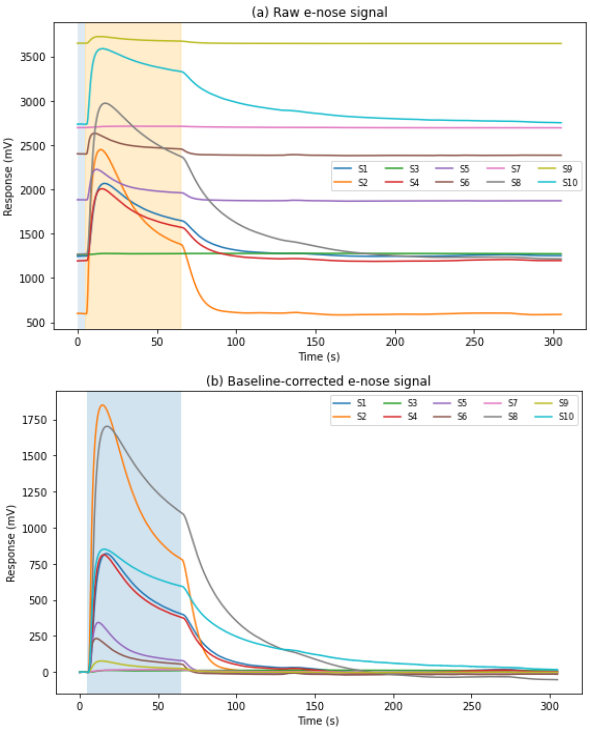


Fig. 3: Sensor response signals: (a) raw signal and (b) baseline-corrected signal

After baseline correction, as shown in Fig. 3(b), the initial offset across all sensors was successfully reduced, so that all signals were normalized relative to the sensors' initial conditions. This result makes the sensor response patterns easier to compare and more representative for the feature extraction process. Furthermore, it is evident that some sensors exhibit different recovery characteristics; sensors such as S1, S3, and S4 show relatively rapid recovery, whereas S8 and S10 experience a slower decline in response. These differences in adsorption–desorption characteristics indicate variations in response dynamics among sensors, which have the potential to enhance the system's discriminatory capability.

Overall, these results show that the electronic nose system consistently captures changes in tea aroma and generates diverse sensor responses, supporting classification in the next stage.

3.2. Feature Space Visualization Using Principal Component Analysis

Principal Component Analysis (PCA) was used to analyze feature separability in the projected feature space. We performed a visualization using Principal Component Analysis (PCA) for all three experimental scenarios. The results of the two-component projection are shown in Fig.s 4(a)–4(c).

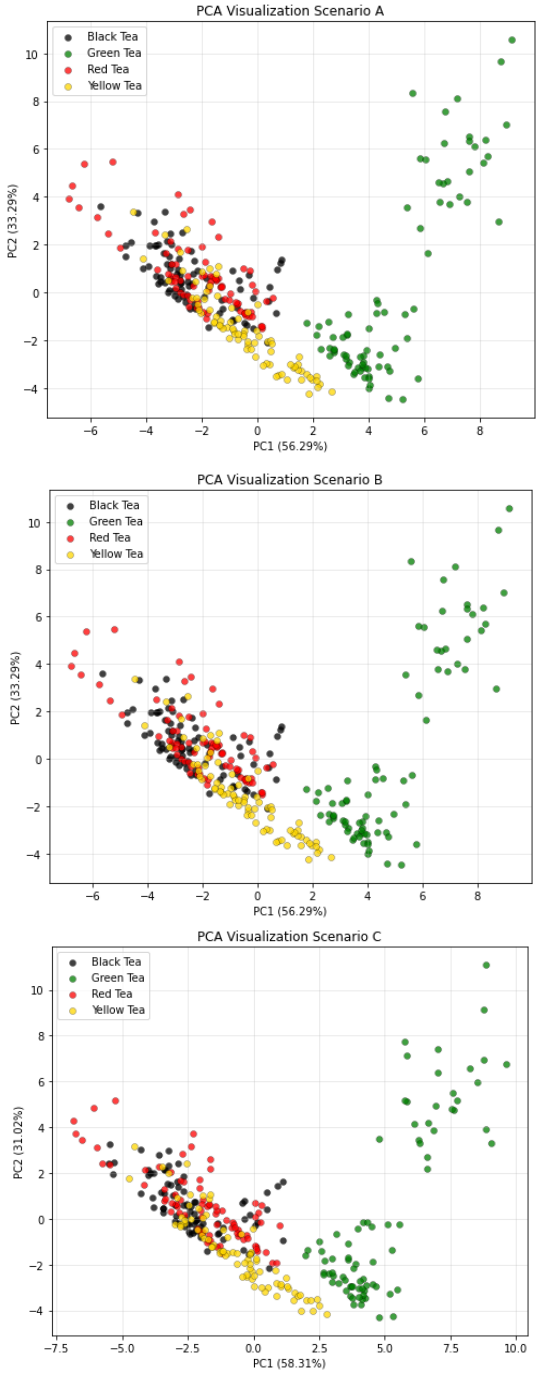


Fig. 4: PCA visualization of feature distribution for (a) Scenario A, (b) Scenario B, and (c) Scenario C.

In Scenario A, the first two principal components accounted for 89.58% of the total data variation, with PC1 contributing 56.29% and PC2 contributing 33.29%. The projection results showed that green tea formed a cluster clearly distinct from the other tea categories, indicating that this category had the most distinct aroma fingerprint. In contrast, the black tea and red tea categories show a fairly significant overlap, suggesting that these two tea categories have similar volatile compound compositions, which may make distinguishing between them more challenging.

In Scenario B, the overall PCA distribution pattern appears relatively similar to that in Scenario A. However, the integration of temperature and humidity features provides additional information in the high-dimensional feature space that helps the machine learning model improve its classification decision boundaries. This indicates that the influence of environmental parameters is not always explicitly visible in the two-dimensional PCA projection, yet it still contributes to improved classification performance.

In Scenario C, two principal components explain 89.33% of the total variation, with PC1 contributing 58.31% and PC2 contributing 31.02%. Although the separation of the green tea group remains clearly defined, the distribution of some other groups shows greater dispersion compared to the previous two scenarios. This suggests that a direct compensation approach to sensor signals may reduce some of the discriminative information crucial for classification.

Overall, the PCA analysis demonstrates that the electronic nose system creates a feature space sufficiently discriminative to distinguish among the four tea groups. However, the overlap that remains between the black tea group and the red tea group indicates that distinguishing these two groups continues to pose a significant challenge in this study.

3.3. Classification Performance Across Various Experimental Scenarios

The classification performance of each machine learning model across the three experimental scenarios is summarized in Table 5. The evaluation was conducted using test set accuracy to compare the classification performance of both individual models and ensemble models.

Table 5: Classification accuracy (%) under different experimental scenarios

Model	Scenario A	Scenario B	Scenario C
SVM	78.75	78.75	85.00
Random Forest	76.25	76.25	81.25
KNN	80.00	80.00	77.50
Ensemble	78.75	80.00	83.75

Table 5 shows that classification performance depends on the experimental scenario used. In Scenario A, performance was evaluated using only features from gas sensors, where the KNN model achieved the highest accuracy at 80.00%, while SVM and the ensemble model each reached 78.75%.

In Scenario B, adding temperature and humidity as new features did not markedly improve the models' performance. The SVM and Random Forest models' accuracies remained unchanged from Scenario A, while the ensemble model's accuracy increased slightly from 78.75% to 80.00%.

The most interesting results were observed in Scenario C, where temperature and humidity-based signal compensation was applied prior to the feature extraction process. In this scenario, all models except KNN showed a significant improvement in performance. The SVM model achieved the highest accuracy of 85.00%, followed by the ensemble model at 83.75%, and the Random Forest model at 81.25%.

These results show that direct compensation at the signal level is more effective than integrating environmental parameters at the feature level. This finding indicates that the effects of temperature and humidity on the MOS sensor are more dominant during the raw signal formation stage than in the statistical feature representations resulting from feature extraction.

3.4. Detailed Analysis of the Best Classification Model

To better understand classification performance, we further analyzed the best model in Scenario C, which had the highest accuracy among all scenarios. We used a confusion matrix, a classification report, and a multiclass Receiver Operating Characteristic (ROC) curve for this analysis.

Fig. 5's confusion matrix shows that classification was most successful for the green tea class: all samples were correctly classified with no misclassifications. This aligns with the previous PCA analysis, where green tea formed a distinct cluster. In the SVM Scenario C model, green tea achieved precision, recall, and F1-scores of 1.00, indicating clear separability.

The highest classification error occurred between black tea and red tea, with frequent misclassification between these categories. In the optimal model, black tea had a precision of 0.67, a recall of 0.80, and an F1-score of 0.73, while red tea achieved a precision of 0.81, a recall of 0.85, and an F1-score of 0.83. These results show that, although the classes can be distinguished, overlapping features still cause misclassifications.

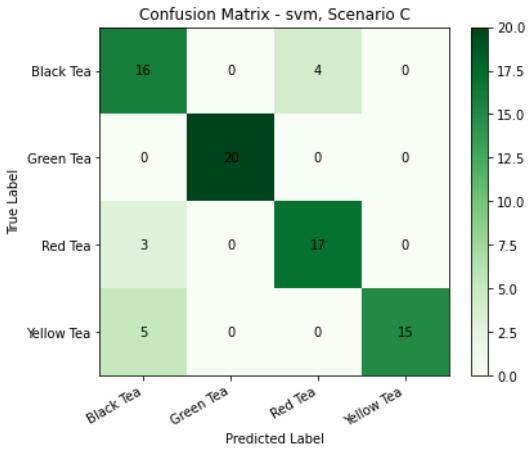


Fig. 5: Confusion matrix of the SVM classifier for Scenario C

This aligns with the principal component analysis (PCA) visualization in Subsection 3.2, where black and red tea clusters appear close together. Their proximity likely stems from similar volatile compound profiles, as both teas undergo extensive

oxidation and share overlapping aroma characteristics.

Evaluation using the multiclass ROC curve in Fig. 6 further reinforces the classification results. The AUC values obtained for each class indicate excellent discriminative performance, specifically around 0.91 for black tea, 1.00 for green tea, 0.94 for red tea, and 0.97 for yellow tea. These high AUC values indicate that the model has excellent probabilistic separation capabilities, although overlap between some classes still influences the final classification decision.

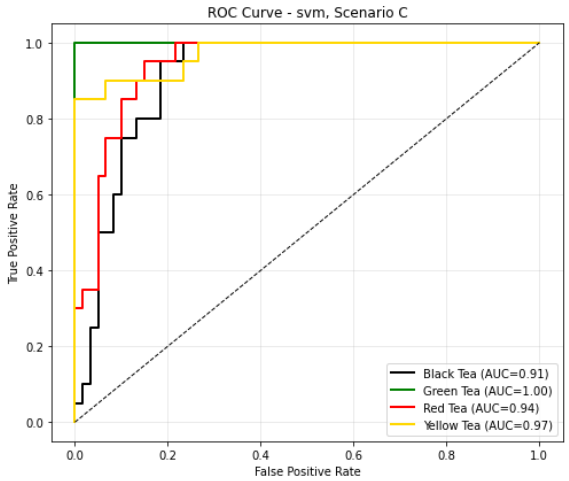


Fig. 6: Multiclass ROC curves of the SVM classifier under Scenario C

Table 6: Test-set robustness index of different classification models

Model	Mean Accuracy	Std Accuracy	RI
SVM	0.8083	0.0295	0.9636
Random Forest	0.7792	0.0236	0.9697
KNN	0.7917	0.0118	0.9851
Ensemble	0.8083	0.0212	0.9737

Overall, these results show that the temperature- and humidity-based compensation approach in Scenario C not only improves classification accuracy but also enhances the separability between classes, particularly for classes with closely related aroma characteristics.

3.5. Robustness Analysis

In addition to evaluating classification accuracy, this study also analyzed the model's robustness to changes in experimental scenarios using the Robustness Index (RI). This metric is used to measure the stability of the model's performance based on variations in accuracy across scenarios. An RI value closer to 1 indicates that the model's performance is more stable in the face of changes in experimental conditions. The results of the robustness index calculations are shown in Table 6.

Based on Table 6, the KNN model had the highest robustness in test data (RI: 0.9851), even though it was not the most accurate. KNN performed most consistently across experimental scenarios, meaning different test conditions or data subsets during evaluation. In contrast, although SVM had the highest classification accuracy in Scenario C, its performance varied more, resulting in a lower RI. To further assess stability, a 10-fold cross-validation robustness index was also calculated, as shown in Table 7.

The results in Table 7 show that the SVM achieved a perfect RI value of 1.000, indicating no variation in performance across scenarios in the cross-validation evaluation. This finding suggests that the SVM exhibits highly stable generalization even as experimental scenarios change. Meanwhile, the ensemble model produced the highest average cross-validation accuracy of 81.75% with an RI value of 0.9934, indicating a good balance between classification performance and model stability.

Table 7: Cross-validation robustness index of different classification models

Model	Mean CV Accuracy	Std Accuracy	RI
SVM	0.8050	0.0000	1.0000
Random Forest	0.7975	0.0035	0.9956
KNN	0.7783	0.0094	0.9879
Ensemble	0.8175	0.0054	0.9934

Overall, the robustness analysis results show a trade-off between maximum accuracy and model stability. SVM in Scenario C provided the highest classification accuracy, while KNN demonstrated the best robustness on the test data, and SVM showed the best stability in the cross-validation evaluation. These findings indicate that selecting the best model requires considering not only accuracy but also performance stability in the face of variations in environmental conditions and experimental scenarios.

5. Conclusions

This study developed an electronic nose-based approach to classify tea aromas. It evaluates how integrating and compensating for temperature and humidity affect machine learning model performance. The electronic nose system successfully captured aroma response patterns for four types of tea: black, green, red, and yellow. These patterns were represented through feature extraction and selection.

Building on these findings, the experimental results show that the temperature–humidity-based signal compensation approach (Scenario C) delivers the best performance compared to the other scenarios. The Support Vector Machine (SVM) model achieved the highest classification accuracy of 85.00%, outperforming Random Forest, KNN, and ensemble models. PCA analysis indicates that green tea exhibits the highest separability, while the greatest overlap occurs between black tea and red tea. Additional evaluation using the Robustness Index (RI) shows that KNN exhibits the best stability in test data evaluation (RI = 0.9851), whereas SVM demonstrates perfect robustness in cross-validation evaluation (RI = 1.000).

Taken together, these results indicate that compensating for environmental parameters at the signal level is more effective than integrating environmental features at the statistical feature level in improving the robustness of electronic nose-based tea aroma classification. These findings underscore the importance of mitigating environmental influences in MOS-based gas sensor systems. Looking

ahead, future research could focus on exploring domain adaptation methods, deep learning, or more adaptive drift compensation techniques to improve the system's generalization in more dynamic environmental conditions.

The proposed environmental compensation framework provides a practical solution for improving the robustness of MOS-based electronic nose systems under varying environmental conditions.

References

[1] Y. Ye, Z. Ai, R. Li, T. Chen, X. Wang, and S. Yu, "Quality analysis and antioxidant activity of different types of tea powder," *Food Prod. Process. Nutr*., vol. 6, Art. no. 36, Dec. 2024, doi: 10.1186/s43014-023-00198-1.

[2] M. Assad, T. J. Ashaolu, I. Khalifa, A. A. Sary, I. M. AlSalami, and S. G. El-Deeb, "Dissecting the role of microorganisms in tea production of different fermentation levels: a multifaceted review of their action mechanisms, quality attributes and future perspectives," *World J. Microbiol. Biotechnol*., vol. 39, no. 10, Art. no. 265, Oct. 2023, doi: 10.1007/s11274-023-03701-5.

[3] J. Moreira, J. Aryal, L. Guidry, A. Adhikari, Y. Chen, S. Sriwattana, and W. Prinyawiwatkul, "Tea Quality: An Overview of the Analytical Methods and Sensory Analyses Used in the Most Recent Studies," *Foods*, vol. 13, no. 22, Art. no. 3580, Nov. 2024, doi: 10.3390/foods13223580.

[4] J. Gehlken, M. Pour Nikfardjam, and C. Zörb, "Prediction of sensory attributes in winemaking grapes by on-line near-infrared spectroscopy based on selected volatile aroma compounds," *Anal. Bioanal. Chem*., vol. 415, no. 8, pp. 1515–1527, Mar. 2023, doi: 10.1007/s00216-023-04549-2.

[5] C. Fosalau, C. Zet and A. Hariton, "Evaluation of Performances and Capabilities of an Electronic Nose," 2025 *International Conference on Electromechanical and Energy Systems (SIELMEN)*, Iasi, Romania, 2025, pp. 1-6, doi: 10.1109/SIELMEN67352.2025.11260833.

[6] L. Cheng, Q.-H. Meng, A. J. Lilienthal, and P.-F. Qi, "Development of compact electronic noses: a review," *Meas. Sci. Technol*., vol. 32, no. 6, Art. no. 062002, Jun. 2021, doi: 10.1088/1361-6501/abef3b.

[7] T. Ren, Y. Lin, Y. Su, S. Ye, and C. Zheng, "Machine Learning-Assisted Portable Microplasma Optical Emission Spectrometer for Food Safety Monitoring," *Anal. Chem*., vol. 96, no. 13, pp. 5170–5177, Apr. 2024, doi: 10.1021/acs.analchem.3c05332.

[8] N. Dhariwal, S. Bisht, P. Yadav, T. Ueda, T. Hyodo and V. Kumar, "Machine Learning in Gas Sensing: A Paradigm Shift for Enhanced Detection and Selectivity—A Review," in *IEEE Sensors Journal*, vol. 26, no. 6, pp. 7897-7921, 15 March15, 2026, doi: 10.1109/JSEN.2026.3655422.

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

[10] H. Mei, J. Peng, T. Wang, M. Zhang, X. Li, and J. Wang, "Overcoming the Limits of Cross-Sensitivity: Pattern Recognition Methods for Chemiresistive Gas Sensor Array," *Nano-Micro Lett*., vol. 16, no. 1, Art. no. 269, Dec. 2024, doi: 10.1007/s40820-024-01489-z.

[11] P. Jia, X. Li, M. Xu, and L. Zhang, "Classification techniques of electronic nose: a review," *Int. J. Bio-Inspired Comput*., vol. 23, no. 1, pp. 16–27, Jan. 2024, doi: 10.1504/IJBIC.2024.136224.

[12] L. A. Putri, I. Rahman, M. Puspita, S. N. Hidayat, A. B. Dharmawan, A. Rianjanu, S. Wibirama, R. Roto, K. Triyana, and H. S. Wasisto, "Rapid analysis of meat floss origin using a supervised machine learning-based electronic nose towards food authentication," *npj Sci. Food*, vol. 7, Art. no. 31, Jun. 2023, doi: 10.1038/s41538-023-00205-2.

[13] U. Kaltsum, K. . Othman Kombo, R. Roto, S. Sholihun, and K. Triyana, "Feature-Optimized Electronic-Nose Classification of Fermented and Non-Fermented Teas", Trends Sci, vol. 23, no. 2, p. 11396, Nov. 2025, doi: 10.48048/tis.2026.11396

[14] S. Robbiani, B. J. Lotesoriere, R. L. Dellacà, and L. Capelli, "Physical Confounding Factors Affecting Gas Sensors Response: A Review on Effects and Compensation Strategies for Electronic Nose Applications," *Chemosensors*, vol. 11, no. 10, Art. no. 514, Sep. 2023, doi: 10.3390/chemosensors11100514.

[15] G. Ansari, R. Singh, S. Kumar, and S. C. Mukhopadhyay, "Fast and robust mixed gas identification and recognition using tree-based machine learning and sensor array response," *Sci. Rep*., vol. 15, no. 1, Art. no. 34472, Dec. 2025, doi: 10.1038/s41598-025-19063-x.

[16] B. Sumanto, K. Triyana, and A. Kusumaatmaja, "Optimized calibration for metal oxide gas sensor arrays to overcome tool variability," *Instrum. Sci. Technol*., vol. 54, no. 1, pp. 118–135, Jan. 2026, doi: 10.1080/10739149.2025.2504478.

[17] A. D. Guritno, H. Suryo, T. W. Wibowo, and M. N. Cahyanti, "Development of a portable electronic nose for the classification of tea quality based on tea dregs aroma," *Int. J. Smart Sens. Intell. Syst*., vol. 17, no. 1, Art. no. 0019, Jan. 2024, doi: 10.2478/ijssis-2024-0019.