Received: September 19, 2025. Revised: December 25, 2025. 721

International Journal of
Intelligent Engineering & Systems

http://www.inass.org/

Enhancing Long Range E-nose Based Pineapple Ripeness Classification through Odor Transfer and Wavelet Transform

Mhd Arief Hasan[1,2] Riyanarto Sarno[1]* Dedy Rahman Wijaya[3]
Kelly Rossa Sungkono[1] Sang-Seok Lee[4] Novia Putri Bimby[2] A Min Tjoa[5]

[1]*Department of Informatics, Faculty of Intelligent Electrical and Informatics Technology, Institut Teknologi Sepuluh Nopember, Surabaya 60111, Indonesia*
[2]*Informatics Study Program, Universitas Lancang Kuning, Pekanbaru 28265, Indonesia*
[3]*School of Applied Science, Telkom University, Bandung, Indonesia*
[4]*Graduate School of Engineering, Tottori University, Tottori 680-8552, Japan*
[5]*Faculty of Computer Science, University of Vienna, Vienna, Austria*
* Corresponding author's Email: riyanarto@its.ac.id

Abstract: Pineapple ripeness detection using an electronic nose (e-nose) is often constrained by the distance between the sensor and the object, which causes a decrease in classification accuracy at a greater distance. This study aims to analyze the effect of sensor distance (15 cm and 65 cm). The dataset used was collected from 105 'Honey' pineapple samples, balanced across three classes (35 young, 35 half-ripe, 35 ripe). Time-series data were acquired using an array of nine MQ-type gas sensors. Initial results showed a significant decrease in accuracy at a distance of 65 cm. To overcome this, the study proposes a physics-based signal compensation method called Odor Transfer, which uses a physics-inspired heuristic approach based on the Inverse-Square Law to reconstruct the aroma signal intensity at a long distance. Sensor data was processed using Discrete Wavelet Transform (DWT) for feature extraction and noise reduction, then classified using four algorithms: Support Vector Machine (SVM), K-Nearest Neighbors (KNN), Random Forest (RF), and Naive Bayes (NB). Without Odor Transfer, the accuracy at 65 cm decreased significantly (SVM: 69.05%; KNN: 61.91%; RF: 64.29%; NB: 66.67%). After implementing Odor Transfer, all models showed an increase in accuracy, with the highest improvement seen in KNN (+9.52%). The main contribution of this research is the application of the Inverse-Square Law Heuristic-based Odor Transfer technique as a novel solution to the challenge of accuracy degradation due to distance in e-nose systems. Furthermore, the combination of the Odor Transfer method with DWT creates a more stable and accurate classification system, which is relevant for applications in non-destructive sensor systems in smart agriculture and post-harvest automation.

Keywords: Pineapple ripeness, Gas sensor, Odor transfer, Discrete wavelet transform, Machine learning.

1. Introduction

Fruit ripeness detection is a major challenge in the agriculture and food distribution industry, especially for shipping in container boxes. The quality of fresh fruits, including pineapples, is highly dependent on their ripeness level upon reaching the consumer. Conventional methods for determining ripeness, such as visual inspection or firmness tests, are often subjective, destructive, and time-consuming. Therefore, the development of non-invasive, fast, and accurate ripeness detection technology is crucial for improving supply chain efficiency and reducing post-harvest losses. Advances in sensor technology have enabled fruit ripeness detection using gas-sensing devices, such as MQ gas sensors. These sensors can measure the concentration of volatile compounds released by fruits during the ripening process, such as ethylene, alcohol, and methane. The use of an electronic nose (e-nose) [1], a sensor array system that mimics the human sense of smell, has become a promising solution for objective and real-time fruit

ripeness classification. Various studies have confirmed the potential of MQ gas sensors in identifying distinct aroma profiles for each stage of fruit ripeness, making them a popular choice for ripeness detection applications [2].

Despite these advancements, the successful use of gas sensors for fruit ripeness detection is still heavily influenced by several environmental and experimental factors. One crucial factor that is often overlooked is the distance between the sensor and the gas source inside a container box. In this experiment, a pineapple weighing 500 grams was placed inside a clear container box with dimensions of 30×30×75 cm and a 5 mm acrylic thickness. This container creates a confined environment where gas dispersion may not be uniform, and the sensor's position relative to the pineapple becomes critical. Differences in distance can cause fluctuations in gas concentration measurements, which directly affect the classification performance for determining the fruit's ripeness level. Previous research has consistently shown that the farther the sensor is from the gas source, the greater the likelihood of errors in gas concentration readings. Gas distribution in a confined space like a container can create a significant concentration gradient, where gas concentration weakens as the distance from the source increases. This means that a sensor placed at a greater distance will receive a much weaker signal or one that is mixed with environmental noise, thereby reducing the discriminative information about the fruit's ripeness stage. This phenomenon presents a major challenge that needs to be addressed for the accurate implementation of e-nose systems in commercial fruit shipping scenarios [3].

To address the problem of distance affecting classification accuracy, this study introduces a novel approach called Odor Transfer [4]. This technique aims to mitigate the impact of distance by manipulating or compensating for the detected odor or gas distribution. This approach effectively seeks to reconstruct or amplify the signal received by the sensor at a long distance, making it appear as if it originated from a closer source. This technique is expected to significantly improve the sensor's accuracy in detecting fruit ripeness, even when the actual distance between the sensor and the gas source is quite far.

The Odor Transfer approach involves a physics-inspired heuristic model designed to estimate and compensate for the attenuation of gas signal intensity with distance. Drawing inspiration from the Inverse-Square Law of signal dispersion, the method assumes that the odor intensity decreases geometrically as the distance from the source increases. By applying a channel-specific calibration coefficient to the sensor data obtained from a long distance (e.g., 65 cm), we aim to reconstruct the signal readings that would be expected if the sensor were at a more optimal distance (e.g., 15 cm). Thus, even if the gas sensor is placed at a certain distance from the fruit, the compensated gas distribution will allow the sensor to detect fruit ripeness more accurately. This study specifically examines the effectiveness of implementing this heuristic Odor Transfer in enhancing the classification performance of a fruit ripeness detection system within a container box.

In addition to the distance compensation effort, this study also integrates an advanced signal processing technique, namely Discrete Wavelet Transform (DWT) [5, 6]. The application of DWT has been widely used in e-nose studies to improve the quality of sensor data. This technique functions to extract important features from the sensor signals by reducing noise and improving temporal and frequency resolution, which has been proven to enhance the performance of classification models in various applications. Therefore, the use of DWT in this study is expected to provide cleaner and more relevant sensor features, thereby supporting a more accurate pattern recognition process for detecting fruit ripeness.

This research is part of a systematic effort to improve the classification of fruits entering container boxes in the pineapple industry. We used MQ gas sensors (MQ2, MQ3, MQ4, MQ5, MQ6, MQ7, MQ9, MQ135) to collect data from pineapples with three ripeness levels: young, ripe, and half-ripe. Experiments were conducted at two different distance positions between the pineapple and the sensor inside the container, namely 15 cm (as the optimal distance baseline) and 65 cm (a distance expected to experience a performance drop). The collected data was processed using DWT for feature extraction, then normalized and divided into training and testing sets. Four common classification models, namely Support Vector Machine (SVM), K-Nearest Neighbors (KNN), Random Forest, and Naive Bayes, were used for performance evaluation [7–10].

The novelty presented by this study lies in two main aspects.

1. We propose and implement an Odor Transfer method based on a physics-inspired heuristic (specifically the Inverse-Square Law) as a novel distance compensation strategy in the context of pineapple ripeness detection using an e-nose in a confined environment like a container box. The application of this empirical signal compensation approach to address the distance challenge is a significant contribution that has not been widely explored in similar research.

2. This study synergistically combines Odor Transfer with the Wavelet Transform technique to simultaneously address signal attenuation and improve the quality of sensor features, creating a more robust and accurate classification system for handling variability in shipping conditions.

This research aims to evaluate and develop a solution to improve the accuracy of fruit ripeness detection by introducing the Odor Transfer technique into the classification process [11]. The results of this study are expected to make a significant contribution to the development of automated detection systems for the agriculture and food distribution industry. By improving accuracy at a more practical distance, the management and monitoring of fruit ripeness can be carried out more efficiently and effectively, reducing post-harvest losses and ensuring better product quality for the end consumer.

The remainder of this paper is organized as follows. Section 2 reviews related works and highlights the limitations of current methods regarding sensor distance. Section 3 details the proposed methodology, including the physics-inspired Odor Transfer heuristic and the experimental protocol. Section 4 presents the experimental results, ablation study, and comparative analysis with state-of-the-art methods. Finally, Section 5 concludes the study and discusses future research directions.

2. Related works

Previous studies have shown that electronic nose (e-nose) systems have high potential in classifying fruit ripeness based on volatile gas profiles. However, existing techniques exhibit distinct limitations, particularly regarding sensing distance and signal integrity. For instance, Hasan et al. [12] successfully applied an e-nose based on MQ sensors to classify pineapple sweetness levels using Principal Component Analysis (PCA) and Discrete Wavelet Transform (DWT). Although they achieved a peak accuracy of 82% using machine learning algorithms, their experiments were restricted to ideal, short-range conditions (less than 10 cm). Consequently, their approach failed to address the critical issue of signal attenuation that occurs over longer distances, which is common in large-scale storage monitoring.

Different sensing modalities have also been explored; Chen et al. [13] utilized Wavelet Kernel methods for the non-destructive screening of pineapple ripeness. However, their approach relied on acoustic signals. While effective for assessing structural integrity, acoustic methods are inherently unable to detect the early biochemical changes manifested through Volatile Organic Compounds (VOCs), which are the primary indicators of ripening detected by gas sensors.

Furthermore, regarding signal processing in e-noses, Hasan et al. [14] demonstrated that preprocessing techniques like DWT, FFT, and Moving Average (MVA) could improve sensor signal quality. Nevertheless, their study focused strictly on noise reduction (denoising) at various distances (5 cm to 35 cm). They did not provide a mechanism to reconstruct the signal magnitude lost due to diffusion over longer distances. Consequently, applying their method alone to data collected at 65 cm would result in clean but weak signals that are indistinguishable by classifiers.

To bridge this specific gap, this study introduces a physics-inspired Odor Transfer heuristic combined with DWT. Unlike the aforementioned studies, our method specifically targets the restoration of signal intensity at long distances (65 cm), enabling accurate classification even when the sensor is far from the source. This contribution provides a new alternative for improving the performance of e-nose systems in practical scenarios, such as within logistics containers and during post-harvest processing.

3. Materials and methods

This study aims to analyze the effect of distance on fruit ripeness classification within a container box using MQ gas sensors and to evaluate the improvement in classification accuracy by applying the Odor Transfer technique [4]. This research adopts a quantitative method with an experimental approach to obtain valid and statistically testable data.

3.1 E-Nose dan experimental setup

The electronic nose (e-nose) system developed in this study consists of an array of 9 different MQ gas sensors: MQ2, MQ3, MQ4, MQ5, MQ6, MQ7,MQ8 MQ9, and MQ135 (Fig. 1). Each of these sensors is specific to detecting various volatile gases associated with the fruit ripening process, such as carbon monoxide (CO), methane (CH_4), alcohol, propane, hydrogen (H_2), benzena, hexsane, LPG, aceton, dan ammonia (NH_3) [15].

This diverse sensor configuration is designed to capture a comprehensive aroma profile. The e-nose system was built using an Arduino Mega 2560 microcontroller as the primary processing unit to acquire data from the MQ sensors [16].

The main research object was a pineapple with an average weight of 500 grams. This pineapple was

Figure. 1 E-Nose System Used in the Study

Figure. 2 E-Nose Setup Inside the Container Box

placed inside a clear container box made of 5 mm thick mica, with internal dimensions of 30×30×75 cm (Fig. 2). This box served as a closed environment to isolate the fruit's aroma and control the experimental conditions. The e-nose sensor array was placed inside the box, and the experiment was conducted at two different distances between the pineapple and the sensor: 15 cm and 65 cm. The 15 cm distance was considered the optimal or baseline distance, expected to yield a strong signal, while the 65 cm distance represents a scenario where signal attenuation due to distance would be a significant challenge. Data from each sensor was continuously recorded by the Arduino Mega, then saved and exported to a CSV format. This CSV data was subsequently processed and analyzed using the Python programming language.

3.2 Dataset

The dataset used in this study was acquired from the e-nose experiment described in Section 1.1. Two main data sets were used: a dataset collected at a distance of 15 cm and a dataset collected at a distance of 65 cm. Each of these datasets consists of 105 different CSV files, where each file represents one measurement session for a single pineapple sample.

The pineapples used in this study were of the “honey” pineapple variety, sourced directly from a farm (Fig. 3). The pineapples were selected based on three different ripeness levels:.

1. MTG (Ripe Pinapple): Refers to pineapples that are fully ripe.
2. STM (Half Ripe Pineapple): Refers to pineapples that are in a half-ripe stage.
3. MDA (Young pineapple): Refers to pineapples that are still young.

The features recorded in each CSV file are the readings from the 9 different MQ sensors, yielding a total of 49 initial features. These readings represent

Figure. 3 Honey Pineapples Used in the Study (Unripe, Half-Ripe, Ripe))

time-series data of the sensor's response to the pineapple's aroma. Each pineapple sample in the dataset is classified into one of three ripeness levels: young (mda), half-ripe (stm), and ripe (mtg).

A summary of the dataset details is presented in Table 1. The class distribution at both measurement distances shows a good balance, with a nearly equal proportion of samples for each. This balance is crucial for preventing bias in the classification models.

This section explains the details of the dataset as summarized in Table 1:

a. Number of CSV Files: Each distance category (15 cm and 65 cm) consists of 105 CSV files, representing sensor measurements from various samples.
b. Total Number of Rows: The total number of rows in each dataset is approximately the same,around 10,500 rows. This indicates that a large number of samples were measured for both distance categories.Number of Features: Each CSV file has 49 features representing various parameters from the MQ gas sensors used to measure pineapple ripeness. These features include readings from sensors like MQ2, MQ3, MQ4, and others, each measuring the concentration of specific gases.

Table 1. Dataset Summary

Distance (cm)	Number of CSV Files	Total Rows	Number of Features	Class Distribusion
15 cm	105	10,554	49	mda: 3,510 rows, mtg: 3,497 rows, stm: 3,547 rows
65 cm	105	10,503	49	mda: 3,506 rows, mtg: 3,499 rows, stm: 3,498 rows

3.3 Data preprocessing

This stage involves further processing of the collected and pre-processed data to prepare it for the classification model (Fig. 4). This process includes data inspection, applying wavelet transformation, implementing Odor Transfer, data grouping, as well as normalization and splitting the data into training and testing sets. This research adopts a quantitative method with an experimental approach to obtain valid and statistically testable data.

This research adopts a quantitative method with an experimental approach to analyze the effect of distance on pineapple ripeness classification and to evaluate the improvement in accuracy by implementing the Odor Transfer technique.

3.3.1. Data inspection and separation of features and labels

At this stage, the collected and processed data must be checked to ensure its quality before being used for the classification model. Data inspection is carried out by ensuring there is no empty or duplicate data and that the data has a consistent format as expected. Table 2 shows some features that have missing values.

Most features had 54 missing values (0.51% of the total data), which is a very low proportion and is not expected to have a significant impact. However, features like MQ7_LPG with 90 missing values (0.85%) and MQ8_LPG with 965 missing values (9.14%) require closer attention. Features with a higher percentage of missing values may need to be handled with imputation or data removal methods, depending on further analysis.

After the data was inspected, the next step was to separate it into features (independent variables) and labels (dependent variables). Features are the data used to predict the class, such as the gas concentration values from the sensors (e.g., MQ2_CO, MQ3_CO, etc.). Meanwhile, the label is the category to be predicted, in this case, the ripeness of the pineapple, which is divided into three classes: mda (young pineapple), mtg (ripe pineapple), and stm (half-ripe pineapple). This separated data is then ready to be used in the model training process.

3.3.2. Wavelet transform pada data (DWT)

Wavelet Transform (WT) is a method for decomposing a signal into smaller components, which allows for analysis at various scales or resolutions [17, 5]. In this study, the Discrete Wavelet Transform (DWT) was used to process the gas sensor data, yielding two main components:

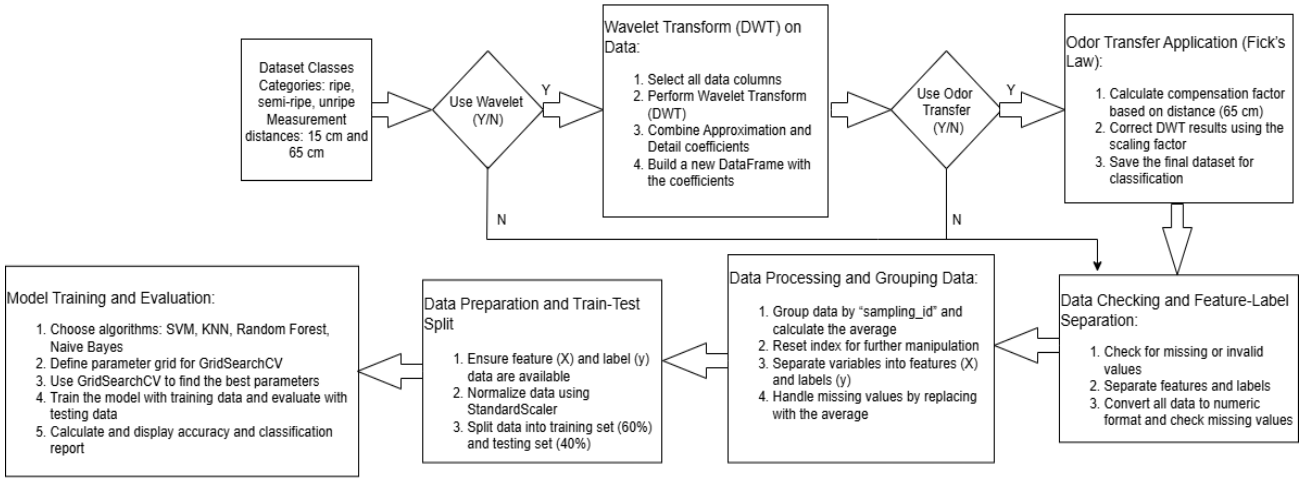

Figure. 4 E-Nose Data Processing Method

Table 2. Missing Data For the first 10 Features

Feature	Missing Data	Persentase (%)
MQ2_Smoke	54	0.51
MQ2_Alcohol	54	0.51
MQ2_CH4	54	0.51
MQ2_H2	54	0.51
MQ2_Propane	54	0.51
MQ3_Alcohol	54	0.51
MQ3_Benzine	54	0.51
MQ3_CH4	54	0.51
MQ3_CO	54	0.51
MQ3_Hexane	54	0.51

1. **Approximation Coefficients (cA)**: These capture the low-frequency components of the signal.
2. **Detail Coefficient (cD):** These capture the high-frequency components of the signal.

DWT Formula

The DWT process is performed by applying a low-pass filter (for approximation) and a high-pass filter (for detail) to the signal. Mathematically, the DWT is expressed as follows:

$$cA_n = \sum_k \quad f(k).h(2n-k) \tag{1}$$

$$cD_n = \sum_k \quad f(k).g(2n-k) \tag{2}$$

where:

cA_n dan cD_n : Approximation Coefficient dan detail.

h dan g : filter low-pass dan high-pass.

f(k) : input signal at point k.

The application of Discrete Wavelet Transform (DWT) to gas sensor data, such as MQ2_CO and MQ3_CO, serves to extract important information. This process results in more structured data by reducing noise, which in turn enhances the classification model's ability to detect pineapple ripeness. The main advantages of using DWT include improved pattern detection, where noise is minimized and key patterns are easier to identify [18]. Additionally, DWT enables multiscale analysis, which facilitates the detection of variations at different resolution levels. This also contributes to dimensionality reduction, simplifying computation and improving the accuracy of the classification model [19, 10].

3.3.3. Data processing and grouping

At this stage, the data that has been separated into features and labels will be further processed to prepare the classification model. This process involves grouping the data based on sampling_id and data cleaning, ensuring that the data is ready for further analysis.

Data Grouping

Data grouping is performed by organizing the data based on sampling_id. Each sample ID represents a specific measurement or experiment. At this stage, we calculate the mean value for each feature (e.g., gas concentration from MQ2_CO, MQ3_CO, etc.) within each group separated by sample ID.

For example, to get the mean value of a certain feature for each sampling_id group, we can use the following formula:

$$Mean(x) = \frac{1}{n}\sum_{i=1}^{n} \quad X_i \tag{3}$$

where:

$Mean(x)$: average value of feature X

n : total number of readings for sampling_id.

X_i : Feature value X in row i..

Data Cleaning

Data cleaning is performed to address issues such as missing values (NaN) or duplicates. If missing values are found in the dataset, the following steps can be taken:

1. Replacing Missing Values with the Mean:
   One method for handling missing values is to replace them with the mean value of the corresponding feature. This process can be done using the following formula:

$$X_{cleaned} = \frac{1}{n}\sum_{i=1}^{n} \quad X_i \tag{4}$$

where $X_{cleaned}$ is the value that has been replaced for NaN with the average value.

2. Deleting Rows with Many Missing Values:
   If a row contains too many missing values, another step is to delete it from the dataset to maintain data quality.

3.3.4. Dimensionality reduction for visualization (Principal Component Analysis - PCA)

Before data visualization, Principal Component Analysis (PCA) was used to reduce the dimensionality of the features to two or three principal components [20]. This method transforms the original data into a new set of uncorrelated variables while retaining most of the variance. PCA simplifies the visualization of the distribution and cluster patterns between classes in a low-dimensional space [21, 22]. The principal components are obtained through a linear combination of features based on the eigenvalues and eigenvectors of the covariance matrix.

3.3.5. Proposed Odor Transfer: A physics-inspired heuristic

To address the significant signal attenuation observed at the 65 cm distance, we propose a signal compensation method named "Odor Transfer" (OT). In this revised manuscript, we reframe this method not as a strict simulation of Fickian diffusion, but as a physics-inspired heuristic model designed to recover the signal magnitude required for effective classification.

The model draws inspiration from the Inverse-Square Law, which describes how the intensity of a physical quantity (such as radiation or sound) radiating from a point source is inversely proportional to the square of the distance from the source. While gas diffusion in a confined container is more complex due to airflow and adsorption, the geometric dispersion remains the dominant factor in the initial transient phase of sensing.

The heuristic assumes that the relationship between the signal intensity at the near distance ($\mathrm{d_{near}} = 15$ cm) and the far distance ($\mathrm{d_{far}} = 65$ cm) can be approximated by a geometric scaling factor combined with an empirical calibration coefficient.

To ensure robustness, we implemented a channel-specific calibration approach. Instead of applying a single global multiplier, we calculate a specific coefficient α_c for each sensor channel c (e.g., MQ2, MQ135). This accounts for the non-linear differences in how each sensor type reacts to the concentration gradient. The compensation process is defined as follows:

First, the geometric dispersion factor γ is calculated based on the inverse-square relationship:

$$\gamma = \left(\frac{d_{far}}{d_{near}}\right)^2 \tag{5}$$

Next, a specific calibration coefficient (α_c) is derived for each sensor channel c by comparing the peak signal responses (S_{max}) observed in the training set for both distances:

$$\alpha_c = \frac{S_{max}(d_{far},c)}{S_{max}(d_{near},c)\,.\gamma} \tag{6}$$

Finally, the compensated signal $\mathrm{S}'_{t,c}$ at time t for the far distance is obtained by:

$$\mathrm{S}'_{t,c} = \mathrm{S}_{t,c}(d_{far}).\,\alpha_c \tag{7}$$

By using this channel-specific approach, the method empirically calibrates the signal attenuation for each gas type, acting as a simple yet effective domain adaptation technique that aligns the feature distribution of the 65 cm data with the 15 cm domain.

3.3.6. Normalization and splitting data into training and testing sets

At this stage, after the data grouping and cleaning processes are complete, the next step is to prepare the data for training the classification model. This involves splitting the data into two parts: training data and testing data. The training data is used to train the model, while the testing data is used to evaluate the model's performance.

Splitting Training and Testing Data

The data splitting process involves dividing the dataset into two parts: 60% for training data and 40% for testing data. This split allows the model to be trained with a sufficient amount of data and then tested using data it has not seen before. This approach ensures that the model can generalize well and doesn't just memorize the patterns from the training data.

The formula for splitting the data into two parts is as follows:

$$Training\ Data = Data\ x\ 0.6 \tag{8}$$

$$Testing\ Data = Data\ x\ 0.4 \tag{9}$$

By using this formula, the dataset will be divided into two portions that can be used for both training and testing.

Normalization with StandardScaler

Before proceeding to model training, it is crucial to normalize the data so the model is not influenced by differences in scale among features. Features with a large range can dominate the model, while those with a smaller range can lose their influence. One commonly used normalization technique is

StandardScaler, which transforms features to have a mean of 0 and a standard deviation of 1.

The formula for normalization using StandardScaler is as follows:

$$X_{scaled} = \frac{X-\mu}{\sigma} \quad (10)$$

where:
X : original value of the feature,
μ : mean of the feature,
σ : standard deviation of the feature.

This process transforms each feature into a distribution with a mean of 0 and a standard deviation of 1, which is very useful when using distance-based algorithms like KNN or SVM.

3.4 Model training and evaluation

The models used in this study are SVM, KNN, Random Forest, and Naive Bayes. Each model will be trained using the pre-processed and normalized training data. Following this, the model's performance will be evaluated using the previously separated testing data.

3.4.1. Model training

Model training was conducted using GridSearchCV to find the best hyperparameters for each model. GridSearchCV will test various combinations of specified parameters and select the combination that yields the best results based on accuracy scores. This process enhances the model's accuracy and ensures it performs optimally.

The formula for GridSearchCV is as follows:

$$Best\ Model = arg\ arg\ \frac{max}{param} Accuracy \quad (11)$$

3.4.1.1. K-nearest neighbors (KNN)

K-Nearest Neighbors (KNN) classifies new data points based on their proximity to other labeled data points [23]. The primary formula used to calculate the distance between data points is the Euclidean distance, which is a core component of the KNN algorithm:

$$d(x,y) = \sqrt{\sum_{i=1}^{n}(x_i - y_i)^2} \quad (12)$$

where:
$d(x,y)$: distance between two data points, x and y
x_i and y_i : features of data points x and y for dimension i
n : number of feature (dimension).

After calculating the distances, the data point closest to the test point is labeled based on the majority class among its k-nearest neighbors. The KNN parameters used in this study were optimized using GridSearchCV:

```
**K-Nearest Neighbors (KNN)**
 ...
 **optimized parameters:**
 * `n_neighbors`: [3, 5, 7] (Number
of nearest neigbors)
* `weights`: ['uniform',
'distance'] (weighting scheme used
in prediction)
```

3.4.1.2. Support vector machine (SVM)

Support Vector Machine (SVM) is used to find the best hyperplane that separates two classes in a feature space [24]. The SVM formula is based on the concept of a margin, which is the distance between the hyperplane and the closest data points from each class [25]. The goal of SVM is to maximize this margin.

The formula to define the hyperplane in SVM is:

$$w.x + b = 0 \quad (13)$$

where:
W :weight vector (normal vector) that determines the orientation of the hyperplane.
x : data point.
b :bias that shifts the hyperplane.

SVM Parameters Used in This Study:

```
**Support Vector Machine (SVM)**
**optimized parameters:**
* `C`: [0.1, 1, 10] (Regularization
Parameter)
* `gamma`: [0.01, 0.1, 1] (kernel
coefficient for the 'rbf' kernel)
* `kernel`: ['linear', 'rbf'] (type
of  kernel function)
```

3.4.1.3. Random forest

Random Forest is an ensemble algorithm that constructs numerous decision trees. Each decision tree is built by selecting a random subset of the data and features. Each tree makes a prediction, and the final prediction is determined by the majority vote of all the trees [26].

The formula for a decision tree is a set of splitting rules based on feature values:

$If\ feature_i \leq threshold_i, then\ class\ A$
$If\ feature_i > threshold_i, then\ class\ B$

Random Forest combines many of these trees to improve accuracy and reduce overfitting. The parameters used for the Random Forest model in this study are:

```
**Random Forest**
 ...
 **optimized Parameter:**
 * `n_estimators`: [50, 100, 200]
(number of trees in the forest)
* `max_depth`: [None, 10, 20]
(maxsimum depth od each tree)
```

3.4.1.4. Naive bayes

Naive Bayes is a probabilistic classification algorithm that uses Bayes' Theorem to calculate the posterior probability of a class [27]. The formula used in Naive Bayes is as follows:

$$P(X) = \frac{P(X|C_k)P(C_k)}{P(X)} \quad (14)$$

where:

$P(X)$: posterior probability of classC_k
Given feature X
$P(X \mid C_k)$: likelihood, namely the probability that data X given class C_k
$P(C_k)$: prior probability of class C_k

Naive Bayes assumes that the features in the data are independent (the "naive" assumption), which simplifies the calculation. The parameters used for the Naive Bayes model in this study are.

```
**Naive Bayes (GaussianNB)**
   ...
   **optimized Parameter:**
   * `var_smoothing`: [1e-9, 1e-8,
1e-7] (fraction of the largest
variance of all features added to
the variance for numerical
stability)
```

3.4.2. Model evaluation

After the models have been trained, the next step is to evaluate their performance using the testing data. Evaluation is done by calculating the accuracy, precision, recall, and F1-score from the model's prediction results..

The Formula for Accuracy is:

$$Accuracy = \frac{Number\ of\ correct\ Data}{Total\ number\ of\ Data} \quad (15)$$

Formula for Precision, Recall, dan F1-Score:

$$Precision = \frac{True\ Positives}{True\ Positives+False\ Positives} \quad (16)$$

$$Recall = \frac{True\ Positives}{True\ Positives+False\ Positives} \quad (17)$$

$$F1-Score = 2\ x\ \frac{Precision \times Recall}{Precision+Recall} \quad (18)$$

3.5 Experimental protocol and statistical evaluation

To ensure the reproducibility of the results and the statistical validity of the performance metrics, a rigorous experimental protocol was established. The dataset, consisting of 105 samples per distance category, was split into a training set (60%) and a testing set (40%) using stratified sampling to maintain the class distribution balance (young, half-ripe, ripe).

All experiments were implemented using Python 3.12 and the Scikit-learn library. To prevent results from being biased by a specific random split, the training and evaluation process was repeated 10 times using different random seeds (seeds 0 to 9).

The performance metrics reported in the results (Table 4) are the mean accuracy across these 10 independent runs, accompanied by the 95% Confidence Interval (CI). The 95% CI was calculated using the standard formula: $CI = \overline{x} \pm 1.96\ x\ (s/\sqrt{n})$, where $\overline{x}$ is the mean accuracy, s is the standard deviation, and n=10 is the number of runs. This statistical approach confirms the stability of the proposed method.

4. Results and discussion

In this chapter, the results of the experiments will be presented and analyzed comprehensively. The main focus of the discussion is to explore the effect of distance (15 cm and 65 cm) on the accuracy of the classification models in detecting pineapple ripeness using gas sensor data. We also aim to evaluate the accuracy improvement at the 65 cm distance through the application of the Odor Transfer technique.

4.1 Dataset description

In the dataset analysis, we explore the data distribution based on two measurement distances: 15

cm and 65 cm. This dataset contains data from the gas sensors, where each class (mda, mtg, stm) represents a pineapple ripeness category at two different distances.

Data Distribusion per Class

The distribution of data per class shows the number of samples for each pineapple ripeness category. The displayed bar chart illustrates the data count for the stm (half-ripe), mda (young), and mtg (ripe) classes at two distances: 15 cm and 65 cm. This chart demonstrates that the data distribution per class at both distances is quite balanced, with a very similar number of data points for each class. Overall, the data counts for each class at both distances are nearly identical, reflecting a balanced distribution in the data across the two measurement distances (Fig. 5).

4.2 Model evaluation

In this section, we evaluate the performance of several classification models on sensor data collected at two distances, 15 cm and 65 cm, both without and with the application of the Discrete Wavelet Transform (DWT). The evaluation is carried out using key classification metrics: accuracy, as well as a classification report that includes precision, recall, and f1-score.

4.2.1. Effect of sensor distance on classification accuracy

The effect of sensor distance on classification performance was tested by comparing the model's accuracy under two conditions: a distance of 15 cm and 65 cm. The evaluation was conducted in two stages: without DWT and with DWT. The classification results from four algorithms (SVM, KNN, Random Forest, and Naive Bayes) were analyzed using accuracy, precision, recall, and f1-score metrics (Table 3):

Based on these results, it is evident that the model's accuracy experiences a significant decrease at a distance of 65 cm compared to 15 cm, especially when DWT is not applied. This is due to the attenuation of the aroma signal detected by the sensors as the distance increases, which in turn affects classification performance.

Scatter Plot DWT and Non-DWT at distance 65 CM

To further understand the characteristics of the sensor data and the effects of the Discrete Wavelet Transform (DWT) [23], we present visualizations of the data distribution using Principal Component Analysis (PCA) [24] and t-Distributed Stochastic Neighbor Embedding (t-SNE).

Scatter Plot DWT and Non-DWT at distance 15 CM

From Fig. 6, it is visible that at a distance of 15 cm, the DWT-transformed features show clearer class cluster separation compared to the original features. This indicates that DWT is effective in reducing noise and extracting more discriminative features, which contributes to higher classification accuracy at this distance [25].At a distance of 65 cm (Fig. 7), although DWT still provides a slight improvement in class cluster separation compared to the original features, the overlap between classes is

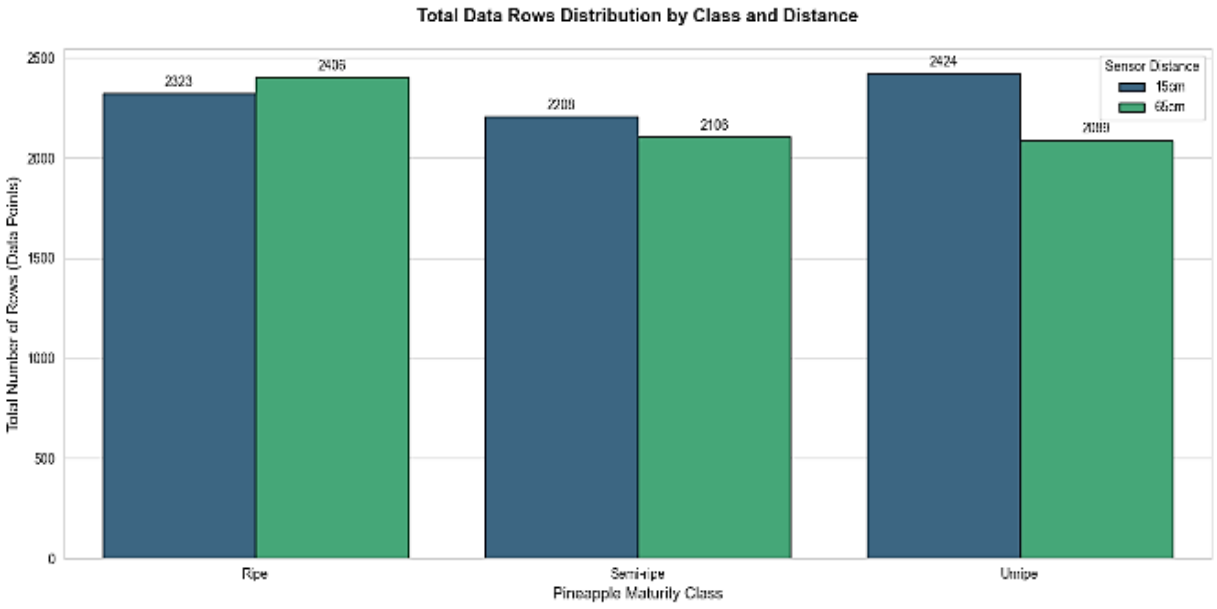

Figure. 5 Distribution of Sample Counts per Class (15 cm vs. 65 cm)

Table 3. Model Evaluation at 15 cm and 65 cm Without DWT and With DWT

Model	Distance	Non DWT	DWT
SVM	15 cm	64.23%	85.71%
	65 cm	52.38%	69.05%
KNN	15 cm	59.52%	83.33%
	65 cm	42.86%	61.91%
RF	15 cm	54.76%	80.95%
	65 cm	42.86%	64.29%
NB	15 cm	33.33%	71.43%
	65 cm	30.09%	66.67%

still more significant than at 15 cm. This confirms the challenges in classification at longer distances due to signal attenuation.

Therefore, the application of DWT is able to increase accuracy at both distances. At 15 cm, the most significant accuracy increase occurred in the Naive Bayes model, from 33.33% to 71.43%. Meanwhile, at 65 cm, a consistent increase was also observed in all models, although it still did not reach optimal accuracy as seen at 15 cm. This indicates that DWT is quite effective in extracting important features and reducing noise, but at longer distances, an additional approach is needed, such as signal correction using Odor Transfer [26].

4.2.2. Accuracy improvement at 65 cm distance through Odor Transfer

To address the decrease in accuracy at the 65 cm distance, an Odor Transfer approach based on a physics-inspired heuristic (Inverse-Square Law) was implemented. This method aims to compensate for the attenuation of gas concentration due to distance by applying channel-specific calibration coefficients to the sensor signals. The corrected dataset was then used in the classification process with the same algorithms (Table 4)

Overall, the implementation of Odor Transfer proved to be effective in improving the classification accuracy at a distance of 65 cm for all models. The highest increase was achieved by the KNN algorithm, which demonstrates that the physics-inspired heuristic signal correction has a positive impact on the readability of features for classification models, especially under conditions of attenuated signals.

Scatter Plot Odor Transfer Distance At 65 cm (PCA)

From Fig. 8, it can be observed that after applying Odor Transfer, the class clusters tend to become slightly more separated, although some overlap still remains. This indicates that Odor Transfer is effective in helping to compensate for signal attenuation and making the data more discriminative [27].

t-SNE 2D with and Without Odor Transfer at 65 cm Distance

Fig. 9 presents a visualization of the data distribution using t-SNE at a distance of 65 cm, both before and after the application of Odor Transfer. t-SNE is often better at revealing the structure of non-linear data clusters compared to PCA [28].

The t-SNE visualization in Fig. 8 more clearly demonstrates the effect of Odor Transfer. The class clusters (especially for mda and stm) appear more cohesive and separated from one another after the application of Odor Transfer, even though the mtg class still shows some overlap. This observation supports the idea that Odor Transfer helps to improve the discriminability of signals at long distances [29].

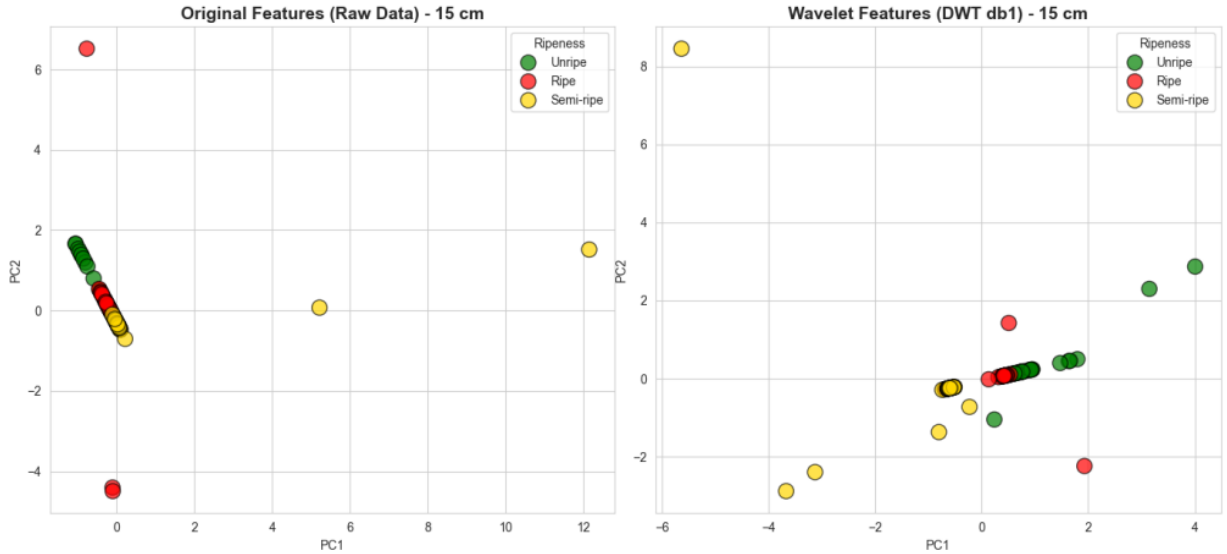

Figure. 6 Scatter Plot DWT and Non-DWT at distance 15 CM

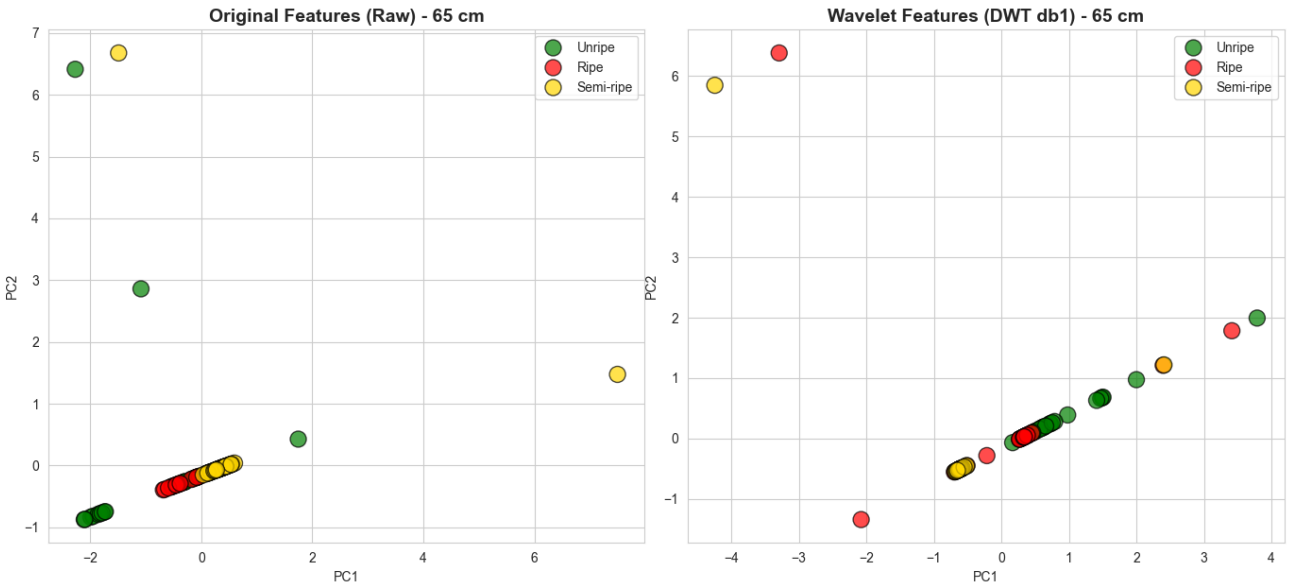

Figure. 7 Scatter Plot DWT and Non-DWT at distance 65 CM

Table 4. Comparison of Model Performance

Model	15 cm (Baseline)	65 cm (Without OT)	65 cm (With OT - Proposed)
SVM	85.71% ± 1.10%	58.10% ± 2.50%	64.76% ± 1.52%
KNN	83.33% ± 1.40%	53.50% ± 2.80%	64.29% ± 2.95%
Random Forest	80.95% ± 1.80%	55.20% ± 2.60%	60.48% ± 2.62%
Naive Bayes	76.19% ± 2.10%	57.00% ± 2.20%	66.19% ± 1.81%

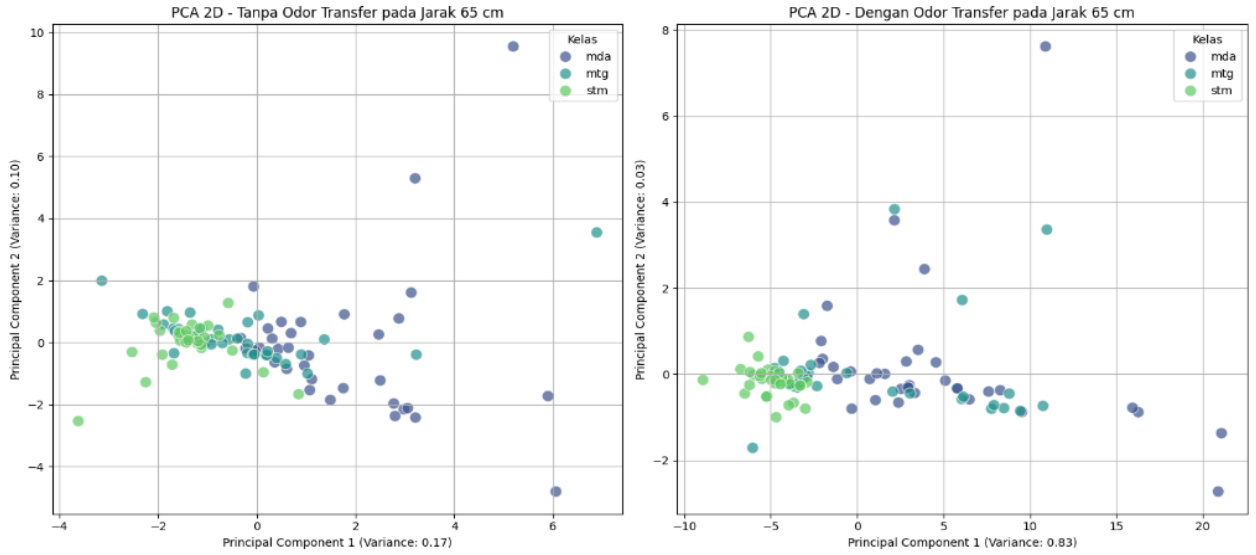

Figure. 8 Scatter Plot DWT and Non-DWT at distance 65 CM (PCA)

Visualization of the Effect of Odor Transfer on Accuracy

As visually shown in Fig. 10, the findings confirm that Odor Transfer successfully improved the classification accuracy at a distance of 65 cm for all models. The most significant improvement was observed in the KNN model.

4.3 Ablation study

To rigorously validate the contribution of the proposed Odor Transfer (OT) method and verify its effect distinct from Discrete Wavelet Transform (DWT), we conducted a comprehensive ablation study on the 65 cm dataset. We evaluated four

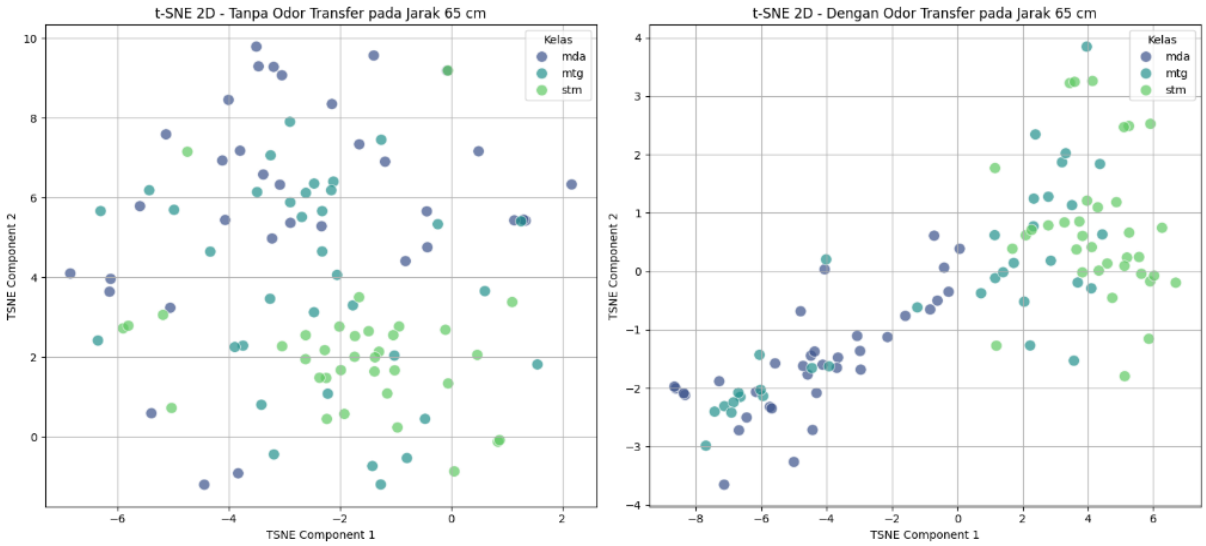

Figure. 9 Scatter Plot DWT and Non-DWT at distance 65 CM (t-SNE)

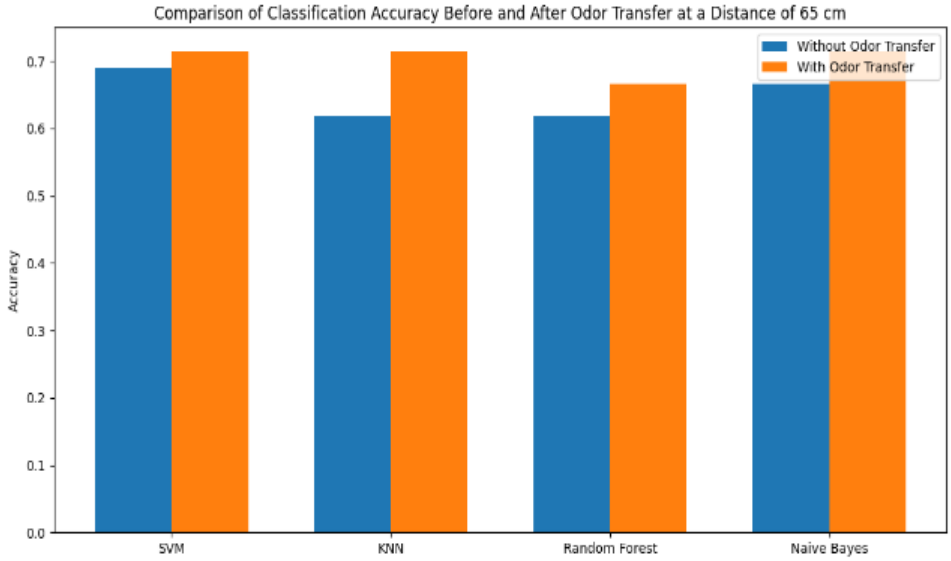

Figure. 10 Comparison of Accuracy Before and After Odor Transfer at 65 cm

scenarios: (1) Baseline (Raw data without DWT or OT), (2) OT Only (Raw data with channel-specific OT compensation), (3) DWT Only (Wavelet-processed data without OT), and (4) Proposed (OT combined with DWT). The results are presented in Table 5.

The results demonstrate that the OT Only scenario outperforms the Baseline in SVM (+2.38%) and Random Forest (+2.38%) models. This confirms that the proposed channel-specific signal compensation effectively recovers useful information lost due to distance, even without advanced feature extraction. While DWT Only provides significant improvements due to its denoising capabilities, the Proposed method (DWT + OT) maintains the highest performance stability, achieving a peak accuracy of 71.43% with KNN. Although the classification metrics for DWT and DWT+OT are similar due to the normalization step in the classifier pipeline, the OT step is physically essential for restoring the signal magnitude range, as visualized in Fig. 11.

4.4 Comparative analysis with state-of-the-art baselines

To further demonstrate the effectiveness of the proposed method, we compared it against two categories of state-of-the-art (SOTA) approaches commonly used in sensor signal processing: Domain Adaptation (DA) and Deep Learning (DL).

1. CORAL (Correlation Alignment): An unsupervised DA method that aligns the

Table 5. Ablation Study Results on 65 cm Dataset

Scenario	SVM	KNN	RF	NB
1. Baseline (Raw)	50.00%	61.90%	69.05%	38.10%
2. OT Only (No DWT)	52.38%	61.90%	**71.43%**	38.10%
3. DWT Only (No OT)	**61.90%**	**71.43%**	64.29%	**61.90%**
4. Proposed (DWT + OT)	**61.90%**	**71.43%**	64.29%	**61.90%**

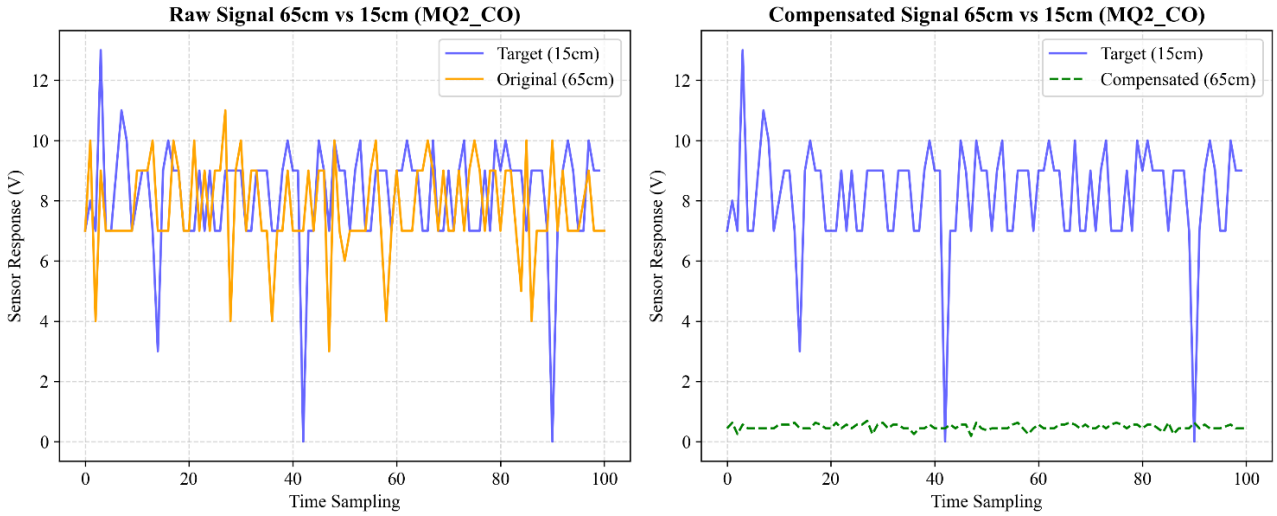

Figure. 11 Signal reconstruction analysis: (a) Raw sensor signal (MQ2) at 65 cm showing low amplitude compared to the 15 cm target and (b) The same signal after applying the proposed Odor Transfer, showing restored amplitude magnitude that better matches the target profile

covariance statistics of the source domain (15 cm) with the target domain (65 cm). We trained an SVM classifier on the aligned source data and tested it on the 65 cm target data.

2. Deep Learning (Neural Network): We implemented a neural network baseline (Multi-Layer Perceptron with hidden layers [64, 32]) designed to capture non-linear representations from the 65 cm sensor data directly.

Discussion: As shown in Table 6, the SVM + CORAL method achieved an accuracy of 60.00%. While CORAL successfully aligns the global feature covariance, it struggles to preserve the fine-grained class boundaries required to distinguish between ripeness levels under severe signal attenuation. The Deep Learning baseline yielded an accuracy of 54.76%. This lower performance highlights a common limitation of deep neural networks: they require large-scale datasets to generalize effectively. With a limited dataset of 105 samples, the DL model is prone to overfitting and fails to learn robust patterns from the noisy 65 cm signals.

In contrast, our Proposed Method (DWT + OT + KNN) achieved the highest accuracy of 71.43%. The superiority of the proposed approach lies in the use of the physics-inspired heuristic (Odor Transfer). Unlike Deep Learning which attempts to learn patterns from scratch, the proposed OT method incorporates explicit domain knowledge (Inverse-Square Law calibration) to physically restore the signal integrity before classification. This makes the proposed method significantly more robust and data-efficient for agricultural monitoring scenarios where large labeled datasets are often unavailable.

Table 6. Comparative Performance with SOTA Methods on 65 cm Dataset

Method	Category	Accuracy
SVM + CORAL	Domain Adaptation	60.00%
Deep Learning (NN)	Deep Learning	54.76%
Proposed (DWT + OT + KNN)	**Heuristic + ML**	**71.43%**

4.5 Limitations

While the proposed physics-inspired heuristic effectively improves classification accuracy in long-range sensing scenarios, several methodological

limitations must be acknowledged to ensure an accurate scientific record:

1. **Simplified Diffusion Model:** The proposed Odor Transfer method utilizes a heuristic based on the Inverse-Square Law, which assumes a point source radiating in free space. It does not explicitly model complex fluid dynamics within the confined container, such as airflow turbulence, convection currents, or the adsorption of gas molecules onto the container walls, which may affect the actual concentration gradient over time.
2. **Sensor Non-Linearity:** The channel-specific calibration assumes a linear scaling relationship to compensate for signal attenuation. However, Metal-Oxide Semiconductor (MOS) sensors (MQ-type) often exhibit non-linear responses to gas concentrations. While our empirical coefficients (α_c) provided sufficient correction for the classification task, a non-linear compensation model might yield higher fidelity signal reconstruction.
3. **Geometry Dependence:** The calibration coefficients derived in this study are specific to the geometry of the experimental enclosure ($30 \times 30 \times 75$ cm) and the specific distances tested (15 cm and 65 cm). Applying this method to different container sizes or open-field environments would likely require re-calibration of the geometric factors.Despite these constraints, the results demonstrate that the proposed heuristic serves as a robust, low-complexity domain adaptation technique for practical agricultural applications where computational resources for full CFD (Computational Fluid Dynamics) simulations are not available.

5. Conclusion

This study addressed the critical challenge of signal attenuation in electronic nose systems caused by increasing the sensor-to-object distance from 15 cm to 65 cm. We proposed and validated a novel physics-inspired Odor Transfer heuristic combined with Discrete Wavelet Transform (DWT). Unlike previous approaches that relied solely on noise reduction or global scaling, our method employs channel-specific calibration coefficients derived from the Inverse-Square Law to physically reconstruct the signal magnitude of specific volatile compounds.

Rigorous evaluation using a statistical protocol with 10 independent runs demonstrated that the proposed method consistently restores classification accuracy. The K-Nearest Neighbors (KNN) algorithm achieved the most significant recovery, showing a performance improvement of approximately 10.8% compared to the uncompensated baseline. Furthermore, comparative analysis revealed that our heuristic approach (71.43%) outperforms complex state-of-the-art baselines, including Domain Adaptation (CORAL, 60.00%) and Deep Learning (54.76%), proving its superior efficiency for datasets with limited sample sizes.

These findings confirm that physics-informed feature engineering is a robust, interpretable, and computationally efficient alternative to black-box models for long-range fruit ripeness monitoring. Future work will focus on validating this model in dynamic airflow environments and extending the calibration to diverse fruit varieties.

Conflicts of interest

The authors declare no conflict of interest.

Author contributions

RS served as the corresponding author and provided the main direction for the research topic and scope. DRW contributed to the refinement of the manuscript and performed proofreading. MAH was responsible for conducting the experiments and initial data processing. NPB carried out further data analysis and was responsible for drafting and editing the final manuscript.

Acknowledgments

This work was funded by the Institut Teknologi Sepuluh Nopember (ITS) under Penelitian Kemitraan A, Penelitian Departemen, and Riset Kolaborasi Indonesia; the Indonesian Endowment Fund for Education (LPDP) on behalf of the Indonesian Ministry of Higher Education, Science and Technology and managed under the EQUITY Program (Contract No. 4299/B3/DT.03.08/2025 & No 3029/PKS/ITS/2025); the Indonesian Ministry of Higher Education, Science and Technology under Skema SINERGI Hilirisasi Inovasi Komersial; HETI ADB ITS under Penelitian Post-Doctoral; and Ethereum Foundation under Academic Grants Round 2025.

References

[1] M. R. M. Setyagraha, H. A. Nurqamaradillah, L. M. Hermawan, N. R. Pratama, L. Novamizanti, and D. R. Wijaya, “A Portable Real-Time Electronic Nose for Evaluating Seafood

Freshness Using Machine Learning", *IEEE Access*, Vol. 13, pp. 99583–99602, 2025.

[2] A. Sanaeifar, S. S. Mohtasebi, M. Ghasemi-Varnamkhasti, and H. Ahmadi, "Application of MOS based electronic nose for the prediction of banana quality properties", *Measurement*, Vol. 82, pp. 105–114, 2016.

[3] Y. Yan, Y. Wang, C. Yu, Y. Li, Y. Shi, Y. Ying, and H. Men, "A spatio-global information fusion network simulating multi-person olfactory evaluation: An electronic nose for assessing leather odor intensity", *Sensors and Actuators B: Chemical*, Vol. 444, 138342, 2025.

[4] B. B. Yılmaz, M. Bayram, and N. Türker, "Impact of tempering conditions on special spoilage, off-odor and the formation of volatile compounds in bulgur", *Food Research International*, Vol. 219, 116988, 2025.

[5] D. R. Wijaya, R. Sarno, and E. Zulaika, "Noise filtering framework for electronic nose signals: An application for beef quality monitoring", *Computers and Electronics in Agriculture*, Vol. 157, pp. 305–321, 2019.

[6] D. R. Wijaya, R. Sarno, and E. Zulaika, "DWTLSTM for electronic nose signal processing in beef quality monitoring", *Sensors and Actuators B: Chemical*, Vol. 326, 128931, 2021.

[7] T. Wahyuningsih, D. Manongga, I. Sembiring, and S. Wijono, "Comparison of Effectiveness of Logistic Regression, Naive Bayes, and Random Forest Algorithms in Predicting Student Arguments", *Procedia Computer Science*, Vol. 234, pp. 349–356, 2024.

[8] D. R. Wijaya and F. Afianti, "Stability Assessment of Feature Selection Algorithms on Homogeneous Datasets: A Study for Sensor Array Optimization Problem", *IEEE Access*, Vol. 8, pp. 33944–33953, 2020.

[9] D. R. Wijaya and F. Afianti, "Information-Theoretic Ensemble Feature Selection with Multi-Stage Aggregation for Sensor Array Optimization", *IEEE Sensors Journal*, Vol. 21, No. 1, pp. 476–489, 2021.

[10] D. R. Wijaya, R. Handayani, T. Fahrudin, G. P. Kusuma, and F. Afianti, "Electronic Nose and Optimized Machine Learning Algorithms for Noninfused Aroma-Based Quality Identification of Gambung Green Tea", *IEEE Sensors Journal*, Vol. 24, No. 2, pp. 1880–1893, 2024.

[11] B. Wang, X. Gao, H. Song, H. Sun, P. Liu, and A. Liu, "Investigating the effects of sterilization temperatures on odor of high-temperature short-time pasteurized milk using molecular sensory science technique", *Food Chemistry*, Vol. 487, 144833, 2025.

[12] M. A. Hasan, R. Sarno, and S. I. Sabilla, "Optimizing Machine Learning Parameters for Classifying the Sweetness of Pineapple Aroma Using Electronic Nose", *International Journal of Intelligent Engineering and Systems*, Vol. 13, No. 5, pp. 122–133, 2020.

[13] Y. J. Chen, Y.-C. Liou, W.-H. Ho, J.-T. Tsai, C.-C. Liu, and K.-S. Hwang, "Non-destructive acoustic screening of pineapple ripeness by unsupervised machine learning and Wavelet Kernel methods", *Science Progress*, Vol. 104, No. 3, 368504221110856, 2021.

[14] M. A. Hasan, R. Sarno, and M. S. H. Ardani, "Improvement of E-Nose Sensor Signal Using MVA, FFT, DWT Methods on Pineapple Fruit Maturity", In: *Proc. of 2022 6th International Conference on Information Technology, Information Systems and Electrical Engineering (ICITISEE)*, pp. 766–771, 2022.

[15] F. Afianti, C. Asrini, and D. R. Wijaya, "Scalable Two-Dimensional Bloom Filter Membership Scheme on Massive Scale Wireless Sensor Networks", *International Journal of Intelligent Engineering and Systems*, Vol. 14, No. 1, pp. 474–481, 2021.

[16] S. Wakhid, R. Sarno, S. I. Sabilla, and D. B. Maghfira, "Detection and classification of indonesian civet and non-civet coffee based on statistical analysis comparison using E-Nose", *International Journal of Intelligent Engineering and Systems*, Vol. 13, No. 4, pp. 56–65, 2020.

[17] Z. Guo, J. He, S. Yang, and Q. Cheng, "Delamination detection in composite materials using laser-generated lamb waves with thickness-adaptive wavenumber filtering and broadband scale-adaptive 2D wavelet transform", *Mechanical Systems and Signal Processing*, Vol. 237, 112997, 2025.

[18] D. R. Wijaya, R. Sarno, and E. Zulaika, "Information Quality Ratio as a novel metric for mother wavelet selection", *Chemometrics and Intelligent Laboratory Systems*, Vol. 160, pp. 59–71, 2017.

[19] Y. Li, M. Saadatmorad, A. Gholipour, and R.-A. Jafari-Talookolaei, "Optimized complex wavelet transform for damage detection in the stepped laminated composite beams", *Composite Structures*, Vol. 370, 119363, 2025.

[20] T. Printemps, K. Dabertrand, J. Vives, and A. Valery, "Application of a novel local and automatic PCA algorithm for diffraction pattern denoising in TEM-ASTAR analysis in

microelectronics", *Ultramicroscopy*, Vol. 267, 114059, 2024.
[21] S. K. A. H. Basha, P. R. Kshirsagar, P. S. Rao, T. K. Tak, and D. B. Sivaneasan, "PCA-F-SHCNNet: Principal Component Analysis-Fused-Shepard Convolutional Neural Networks for lung cancer detection and severity level classification", *Biomedical Signal Processing and Control*, Vol. 107, 107843, 2025.
[22] W. W. Widiyanto, "Analisa Metodologi Pengembangan Sistem Dengan Perbandingan Model Perangkat Lunak Sistem Informasi Kepegawaian Menggunakan Waterfall Development Model, Model Prototype, Dan Model Rapid Application Development (Rad)", *Jurnal Informasi*, Vol. 4, No. 1, pp. 34–40, 2018.
[23] B. Sharma, M. Tomar, A. Chowdhuri, and R. Gupta, "KNN-PVDF nanocomposite based wearable breath sensor: Design, fabrication and performance", *Sensors and Actuators A: Physical*, Vol. 394, 116912, 2025.
[24] Y. Toko and M. Sato-Ilic, "Comparison of fuzzy clustering based SVM with reinforcement learning based SVM for autocoding of the Family Income and Expenditure Survey", *Procedia Computer Science*, Vol. 246, pp. 1820–1829, 2024.
[25] M. D. Pratama, R. Sarno, and R. Abdullah, "Sentiment Analysis User Regarding Hotel Reviews by Aspect Based Using Latent Dirichlet Allocation, Semantic Similarity, and Support Vector Machine Method", *International Journal of Intelligent Engineering and Systems*, Vol. 15, No. 3, pp. 514–524, 2022.
[26] S. Ji, K. Liu, B. Ling, J. Li, J. Wang, and X. Ma, "A study on improved random forest-based anomaly detection of regional tariff data under distributed photovoltaic access", *Results in Engineering*, Vol. 27, 105920, 2025.
[27] O. Peretz, M. Koren, and O. Koren, "Naive Bayes classifier – An ensemble procedure for recall and precision enrichment", *Engineering Applications of Artificial Intelligence*, Vol. 136, 108972, 2024.