Sensors and Materials, Vol. 38, No. 10 (2026) 6205–6226
MYU Tokyo

S & M 4676

Nondestructive Evaluation of Pineapple Texture from Vibration Data

Chang-Sheng Lin,[1*] Chih-Chung Hsu,[2] Chin-Hui Yeh,[3]
Chi-Han Tsai,[2] and Hung-Tse Chiang[4]

[1]Department of Mechanical Engineering, National Kaohsiung University of Science and Technology, Kaohsiung, Taiwan
[2]Institute of Intelligent Systems, College of Artificial Intelligence, National Yang Ming Chiao Tung University, Tainan, Taiwan
[3]LIMO System Technology Co., Ltd., Kaohsiung, Taiwan
[4]Chroma ATE Inc., Taoyuan, Taiwan

(Received May 25, 2026; accepted August 5, 2026)

Keywords: pineapple texture evaluation, nondestructive testing, frequency response function, Matthews correlation coefficient, linear discriminant analysis

Previously, pineapple texture grading methods relied on farmers' experience as a reference, leading to differences in judgment criteria depending on the operator. In this study, we employ a nondestructive testing method to estimate the texture of pineapples. By using excitation signals from a vibration platform and measuring the response signals of pineapples, the frequency response function is calculated to analyze the dynamic characteristics of the fruit. A texture discrimination threshold is then established as a basis for classifying pineapples into "hollow-sounding" and "solid-sounding" categories. The developed algorithm was evaluated using a confusion matrix and Matthews correlation coefficient (*MCC*) to confirm the texture discrimination threshold. The discriminative power of the classification model was further validated using the receiver operating characteristic curve (ROC) and the area under the ROC curve, thereby establishing a technique for quantifying pineapple texture. Principal component analysis and linear discriminant analysis (LDA) were used to reduce the dimensionality of the data and extract relevant features for texture identification. The results indicated that the *MCC*-based classification showed a significant consistency with farmers' texture judgments. Moreover, LDA proved effective in distinguishing different pineapple textures. The proposed methodology provides a reliable technique for pineapple texture identification, reducing human errors and costs while significantly improving the accuracy of texture classification.

1. Introduction

Taiwan is geographically located in the subtropical region, providing natural environmental advantages for the cultivation of tropical fruits. Among them, pineapple is one of the most important tropical fruits in Taiwan, sold in both domestic and international markets. The texture

*Corresponding author: e-mail:changsheng@nkust.edu.tw
https://doi.org/10.18494/SAM6451

ISSN 0914-4935 © MYU K.K.
https://myukk.org/

and mechanical properties of pineapples directly affect their market value and shelf life. To reliably assess the internal quality of pineapples, it is necessary to evaluate their texture, which in turn reflects their level of ripeness.

Traditionally, farmers assess the external texture of pineapples by visual inspection, while the internal texture is judged by the tone produced when tapping the fruit using fingers or rubber sticks. On the basis of the tapping sound, pineapples are generally categorized into "hollow-sounding fruits" and "solid-sounding fruits." Hollow-sounding fruits produce a drum-like sound, have golden yellow flesh, higher storability, and moderate moisture and ripeness. In contrast, solid-sounding fruits emit a meat-like sound, have darker yellow flesh, lower storability, and higher moisture and ripeness levels. These judgments are made through the tapping and subsequent slicing of the fruit for cross-sectional observation. However, such assessments are subjective and can vary depending on individual experience, leading to inconsistent results. Therefore, converting the assessment of pineapples into hollow- or solid-sounding categories using objective and quantitative analysis is a major focus of this study. To effectively classify pineapple texture, machine learning algorithms are employed to analyze characteristic features of the fruit and determine classification parameters and threshold values. These methods enable efficient quality evaluation, threshold establishment, and the reduction of labor costs and time, thereby improving classification accuracy.

As pineapple is an important economic fruit in Taiwan, quality grading is essential. Research on the nondestructive testing (NDT) of fruits has been continuously developed and applied in various fields. These techniques enable the assessment of internal characteristics without damaging the specimen, thus providing effective tools for evaluating fruit quality. For instance, Pathaveerat *et al.*[1] utilized an acoustic impulse response method involving impact rods and microphones to nondestructively measure the main resonance frequency of pineapples and compared the results with those of destructive cross-sectional analysis. They assessed parameters such as flesh firmness and acidity to examine correlations between destructive measures and ripeness, achieving a classification accuracy of 82.8%. Ay[2,3] multiplied the first resonance frequency (f) by mass (m) to derive a classification index, successfully distinguishing between hollow- and solid-sounding pineapples, with accuracy rates of 70–80%. Abbaszadeh and Moosavian.[4] adopted a nondestructive method using a laser Doppler vibrometer (LDV) to analyze vibration signals in watermelons for internal quality evaluation. Taniwaki and Sakurai[5] explored the application of acoustic vibration techniques in agriculture to assess ripeness nondestructively, using them as tools for quality determination. Tiplica *et al.*[6] employed acoustic measurements to evaluate apple texture by striking the fruit with an impact hammer and using a microphone to capture spectral response signals. Eighteen key features were extracted and used to classify ten different apple varieties. Zhang *et al.*[7] evaluated the firmness of differently shaped pears using an NDT approach and validated the prediction capability through receiver operating characteristic (ROC) curves. Ballabio *et al.*[8] investigated classification performance by principal component analysis (PCA) and linear discriminant analysis (LDA) in both binary and multiclass datasets, considering balanced and imbalanced distributions. They compared different classification metrics using LDA-derived models.

Addressing the challenges of learning from imbalanced datasets, Bekkar *et al.*[9] proposed alternative evaluation methods using combined metrics such as accuracy, F_1 score, ROC curves, and area under the ROC curve (AUC). Brito *et al.*[10] classified cereal bars by NIR spectroscopy and supervised pattern recognition, developing classification models based on PCA and LDA. While the accuracy and F_1 score calculated from the confusion matrix are common evaluation metrics, these statistics may overestimate performance, especially in imbalanced datasets. Hence, Chicco and Jurman[11] suggested that Matthews correlation coefficient (*MCC*) outperforms both accuracy and F_1 score, as demonstrated through experiments on six datasets with varying class imbalance. Fawcett[12] and Tharwat[13] highlighted the ROC curve as an effective tool for visualizing and evaluating classification models, with AUC used to confirm model discriminability and assess stability under imbalanced data conditions. Huang *et al.*[14] generated acoustic spectrograms from the data collected using the hitting machine under various conditions, including farm noise, and used them to train a deep learning convolutional neural network (CNN). The model achieved high classification accuracy, and together with the hitting device, provides an efficient and reliable method for automatic pineapple sorting. Chen *et al.*[15] developed a nondestructive device that taps pineapple surfaces and analyzes the resulting sound using wavelet features and unsupervised machine learning. With optimized acoustic coupling and a custom thorn board for higher tapping performance, the system delivers 98.56% accuracy and 0.93 F_1 score in automatic classification. Arwatchananukul *et al.*[16] used acoustic response to mimic traditional fruit-knocking for pineapple quality assessment. Total soluble solids (TSS) and TSS/TA, which combines TSS and titratable acidity (TA), are effective for maturity, while stiffness coefficients help detect defects. Random forest achieved accuracies of 99.38% for maturity and 99.58% for defects. The method is low-cost and practical for farmers, though efficiency and reliability require further validation.

2. Research Method

In this study, commercially available Tainung No. 17 (commonly known as Golden Diamond) pineapples were used as experimental samples. These samples were obtained from local fruit markets and were representative of typical pineapples available for sale. The pineapples used in the experiments were free of visible defects and selected to ensure consistency in maturity and size, thereby minimizing variability during physical and vibration measurements. The physical properties of the pineapple samples were measured, and vibration analysis was conducted using the internal texture detection module provided by the Industrial Technology Research Institute (ITRI) of Taiwan to collect vibration signals from the pineapples. Subsequently, a raw dataset was constructed for further processing. PCA and LDA were applied as dimensionality reduction methods, followed by the calculation of relevant evaluation metrics. Finally, a classification threshold for pineapple texture was established, and the results of the discriminant analysis were discussed.

2.1 Experimental materials

Since the physical properties of pineapples begin to change after harvesting, all samples used in this study were tested on the day they were collected. Prior to the experiments, experienced farmers classified the pineapple samples on the basis of their texture by tapping and other related assessment methods. The pineapples were categorized as either hollow-sounding or solid-sounding fruits, and these classifications served as reference labels for the experimental dataset. To ensure the reliability of the farmers' classification, a portion of the pineapples was cut longitudinally after the experiments to inspect their internal characteristics. If the cut section revealed golden-yellow flesh with relatively low moisture content, the sample was confirmed as a hollow-sounding fruit, as shown in Fig. 1(a). Conversely, if the flesh appeared dark yellow and retained higher moisture content, the sample was identified as a solid-sounding fruit, as shown in Fig. 1(b).

2.2 Pineapple parameter combinations

Pineapple texture is often closely related to its physical properties. NDT methods are based on the measurement of these physical properties. In this study, three experimental measurements were conducted: mass measurement, volume estimation, and density estimation. Each pineapple sample was placed on a weighing scale three times, and the average of the three measurements was calculated to serve as the mass parameter for that sample. An electronic weighing scale was used for these measurements.

In previous studies, sample volume was also measured by the water displacement method. However, submerging pineapples in water accelerates fermentation and adversely affects subsequent analyses. Therefore, in this study, we estimated the volume of pineapples by measuring their maximum length, width, and height. The volume was then calculated as



(a) (b)

Fig. 1. (Color online) Cross-sectional views of (a) hollow-sounding and (b) solid-sounding pineapples.

$$V = \frac{4}{3}\pi HLW\,. \tag{1}$$

Here, H represents the maximum height of the pineapple measured, excluding the crown; only the fruit body is measured. L and W denote the maximum length and width of the pineapple, respectively, as illustrated in Fig. 2. The volume of the pineapple is estimated using the formula for an ellipsoid, which yields an approximate actual pineapple volume.

Because of the lack of equipment for directly measuring the density of the test samples, it was difficult to obtain the actual density parameters of the pineapples. Therefore, density parameters were estimated indirectly. By using the measured mass and estimated volume of each pineapple sample, the density ρ was calculated as

$$\rho = \frac{m}{V}, \tag{2}$$

where m represents the measured mass of the pineapple and V represents the estimated volume. The density of the pineapple is approximated using the density formula, thereby obtaining an estimate close to the actual pineapple density. By randomly sampling 20 pineapples, repeated five times, and comparing ellipsoid volume estimation with water displacement, the average error was found to be below 4%, and sensitivity analysis showed that this had a negligible effect on classification (<1% change).

2.3 Vibration signal analysis

The equipment used in this experiment consists of both hardware and software components. The hardware setup is designed to excite and measure the vibration signals of pineapples,

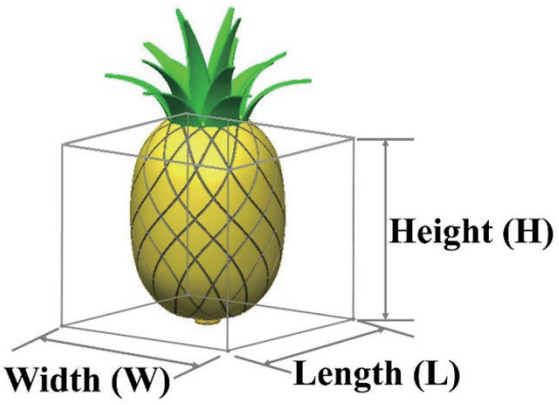


Fig. 2. (Color online) Schematic diagram illustrating pineapple dimensions: height (H), length (L), and width (W).

utilizing the internal texture detection module provided by ITRI, as shown in Fig. 3(a). This module comprises an exciter, an accelerometer, and a force sensor. Vibration signals from the pineapples are acquired through a signal acquisition device, enabling texture analysis based on dynamic characteristics. The software system includes signal acquisition and data post-processing. The entire system is controlled by a programmable logic controller (PLC), and the human–machine interface (HMI) is developed using LabVIEW. Through the operation of the detection module, vibration signals are collected and subsequently processed in MATLAB R2023a to generate spectral plots. These spectral analyses are then used for further discriminant analysis to distinguish pineapple texture on the basis of the vibration signals.

The pineapple was placed on the vibration platform of the internal texture detection module and secured by clamps on the platform to prevent movement, as shown in Fig. 3(b). In this study, the vibration exciter SA-JZ005, located beneath the vibration platform, was used to induce vibration on the pineapple's crown (also called the fruit stalk). The excitation signal was a sine sweep in the range 10–4000 Hz, with amplitude 25 g and duration 2 s. The excitation signal from the exciter was captured by a force sensor, while the pineapple's response signal was measured by an accelerometer, as illustrated in Fig. 3(c). Both the excitation and response signals were acquired using a signal acquisition device, Dytran 5860B. Power spectral density (PSD) calculations were performed using the analysis software provided by ITRI to obtain the pineapple's frequency response function (FRF). To ensure experimental consistency, each

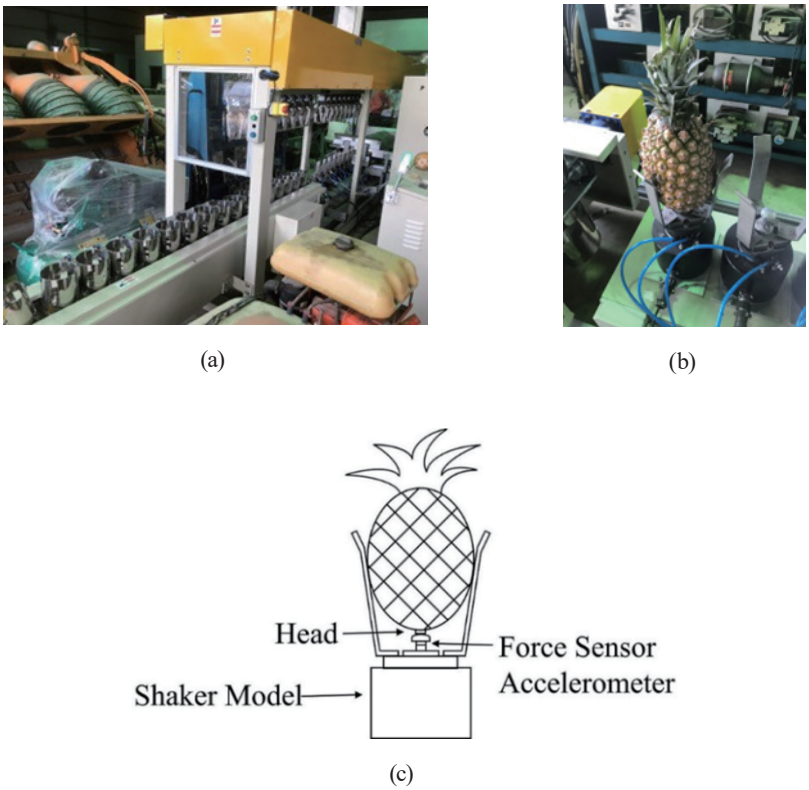


Fig. 3. (Color online) (a) Internal texture detection module. (b) Experimental setup of pineapple testing. (c) Schematic of pineapple measurement instruments.

pineapple was tested three times, with the fruit rotated and repositioned on the vibration platform for each trial. The pineapple was rotated 120° each time, and the clamping force was kept constant using a torque-controlled fixture. The repeatability of FRF across trials showed a variation coefficient < 3%, confirming good stability. Consequently, the clamping position was varied in each test, allowing for the evaluation of reproducibility across different measurement points and repeatability at the same measurement point on the same pineapple.

Using the force excitation signal $f(t)$ measured by the force sensor in the time domain, and the acceleration response signal $a(t)$ of the pineapple measured by the accelerometer in the time domain, the FRF of the pineapple can be obtained. By applying the fast Fourier transform (FFT) to the time-domain signals, the power spectra are calculated. FRF is defined as the ratio between the system's output and input in the frequency domain. The equation for FRF is

$$H(\omega) = \frac{A(\omega)}{F(\omega)}, \tag{3}$$

where $A(\omega)$ and $F(\omega)$ are respectively the frequency-domain response and excitation signals obtained through FFT. The FRF of the pineapple sample can thus be obtained.

In this study, the FRFs of pineapple samples were obtained using a texture detection module and then imported into MATLAB R2023a for further identification analysis. Figure 4 shows the overlaid FRFs of all pineapples, allowing for a more comprehensive observation of the measured FRFs. It was observed that the pineapple FRFs exhibited a prominent first peak at around 200 Hz, which is defined as the primary peak frequency of the pineapple. Low-frequency components typically contribute the most to the structural vibration response, and the amplitude corresponding to the primary peak frequency is examined to determine whether it can be used as an indicator of pineapple texture. This parameter is further explored as a basis for classification. To verify the repeatability and reproducibility of the pineapple FRFs, individual frequency spectra for each pineapple are plotted in Fig. 5. A custom MATLAB script was

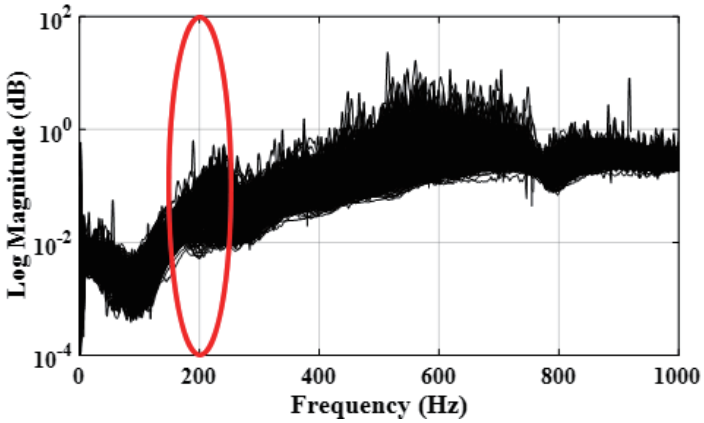


Fig. 4. (Color online) FRF of pineapple.

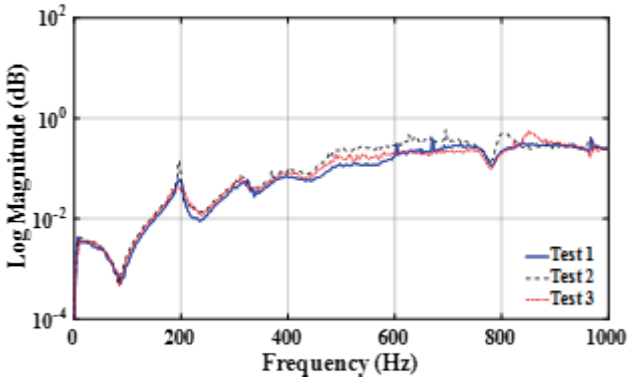


Fig. 5. (Color online) FRF of a single pineapple.

developed to extract the primary peak frequency and its corresponding amplitude for subsequent classification analysis, in order to assess the discriminative capability of these features.

2.4 PCA

PCA is commonly used to reduce the dimensionality of a dataset while retaining the features that contribute most to the variance within the data. Suppose the original dataset consists of n samples, each with d features:

$$\boldsymbol{X} = [\boldsymbol{X}_1\ \boldsymbol{X}_2\ \cdots \boldsymbol{X}_d] = \begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1d} \\ x_{21} & x_{22} & \cdots & x_{2d} \\ \vdots & \vdots & \ddots & \vdots \\ x_{n1} & x_{n2} & \cdots & x_{nd} \end{bmatrix}. \tag{4}$$

The original data are reduced to k dimensions (i.e., dimensionality reduction) by the following process:

$$\boldsymbol{Y} = \begin{bmatrix} y_{11} & y_{12} & \cdots & y_{1k} \\ y_{21} & y_{22} & \cdots & y_{2k} \\ \vdots & \vdots & \ddots & \vdots \\ y_{n1} & y_{n2} & \cdots & y_{nk} \end{bmatrix}. \tag{5}$$

Therefore, it is necessary to find a basis vector matrix $\boldsymbol{W}_{d\times k}$ such that $\boldsymbol{Y}_{n\times k} = \boldsymbol{X}_{n\times d}\boldsymbol{W}_{d\times k}$ holds. The obtained data are then used to calculate their mean value and variance. Subsequently, data standardization is performed with the aim of unifying the scale of features and reducing identification difficulties caused by scale differences. The standardization is defined as

$$Z = \frac{X - \mu}{\sigma}, \tag{6}$$

where μ represents the sample mean and σ denotes the sample standard deviation. The covariance matrix is subsequently calculated as

$$\boldsymbol{C} = \begin{bmatrix} \operatorname{cov}(\boldsymbol{X}_1, \boldsymbol{X}_1) & \operatorname{cov}(\boldsymbol{X}_1, \boldsymbol{X}_2) & \cdots & \operatorname{cov}(\boldsymbol{X}_1, \boldsymbol{X}_d) \\ \operatorname{cov}(\boldsymbol{X}_2, \boldsymbol{X}_1) & \operatorname{cov}(\boldsymbol{X}_2, \boldsymbol{X}_2) & \cdots & \operatorname{cov}(\boldsymbol{X}_2, \boldsymbol{X}_d) \\ \vdots & \vdots & \ddots & \vdots \\ \operatorname{cov}(\boldsymbol{X}_d, \boldsymbol{X}_1) & \operatorname{cov}(\boldsymbol{X}_d, \boldsymbol{X}_2) & \cdots & \operatorname{cov}(\boldsymbol{X}_d, \boldsymbol{X}_d) \end{bmatrix}. \tag{7}$$

By singular value decomposition (SVD), the eigenvectors and the corresponding eigenvalues of the covariance matrix $\boldsymbol{C}$ can be obtained. The eigenvalues are arranged in descending order, and the eigenvectors are rearranged accordingly. The top k eigenvectors corresponding to the largest eigenvalues are selected to form the basis matrix $\boldsymbol{W}_{d\times k}$. This allows the projection of the original data into a new reduced-dimensional space, resulting in the data projection matrix $\boldsymbol{Y}_{n\times k}$. In this study, there are a total of five eigenvalues and 240 data samples.

2.5 LDA

In terms of dimensionality reduction methods, LDA is not only an extension of PCA, but also a supervised learning technique that utilizes information from labeled data. Since the dataset is distinguished by labels, LDA offers higher discrimination ability for classification than PCA during the dimensionality reduction and feature separation processes. The aim of PCA is to find a projection axis that maximizes the variance of the projected data, whereas LDA seeks to maximize the scatter between different classes. Assuming there are two classes represented by $\boldsymbol{X}_{n_1\times d}$ and $\boldsymbol{Y}_{n_2\times d}$, the within-class scatter is defined as the sum of the covariance matrices of the two classes:

$$\boldsymbol{S}_w = \operatorname{cov} \boldsymbol{X} + \operatorname{cov} \boldsymbol{Y}, \tag{8}$$

where cov $\boldsymbol{X}$ denotes the covariance matrix of the hollow-sounding fruit class and cov $\boldsymbol{Y}$ denotes the covariance matrix of the solid-sounding fruit class. After obtaining the within-class scatter, the between-class scatter is then calculated. The between-class scatter is defined as the product of the difference between the mean vectors of the two classes:

$$\boldsymbol{\mu}_x = \begin{bmatrix} E(\boldsymbol{X}_1) \\ \vdots \\ E(\boldsymbol{X}_d) \end{bmatrix}, \tag{9}$$

$$\boldsymbol{\mu}_y = \begin{bmatrix} E(\boldsymbol{Y}_1) \\ \vdots \\ E(\boldsymbol{Y}_d) \end{bmatrix}, \tag{10}$$

$$\boldsymbol{S}_b = \left[\left(\boldsymbol{\mu}_x - \boldsymbol{\mu}_y \right) \left(\boldsymbol{\mu}_x - \boldsymbol{\mu}_y \right)^T \right], \tag{11}$$

where $\boldsymbol{\mu}_x$ represents the mean vector of the five features of the hollow-sounding fruit and $\boldsymbol{\mu}_y$ represents the mean vector of the five features of the solid-sounding fruit. $\mathbf{S}_b$ denotes the difference between the mean vectors of the hollow-sounding and solid-sounding fruits. Using these equations, the within-class scatter $\boldsymbol{S}_w$ and between-class scatter $\boldsymbol{S}_b$ are obtained. To maximize the separation between different classes, the goal is to minimize the within-class scatter $\boldsymbol{S}_w$ while maximizing the between-class scatter $\boldsymbol{S}_b$.

$$\boldsymbol{S}_w^{-1} \boldsymbol{S}_b w = \lambda w \tag{12}$$

The ratio between the within-class scatter and the between-class scatter, similar to the approach used in PCA, is analyzed by SVD to obtain the eigenvectors and eigenvalues of the data, which are then used for subsequent discriminant analysis.

2.6 Evaluation metrics

To evaluate the performance of the classification model for distinguishing pineapple textures, several evaluation metrics were selected as the basis for assessing classification effectiveness. The metrics include the confusion matrix, accuracy, F_1 score, *MCC*, ROC curve, and AUC. To effectively distinguish between different pineapple textures, it is important to understand the distribution of the samples. Figure 6 illustrates the distribution of hollow-sounding and solid-sounding pineapples, as determined by farmers' expert judgments, which serve as reference data

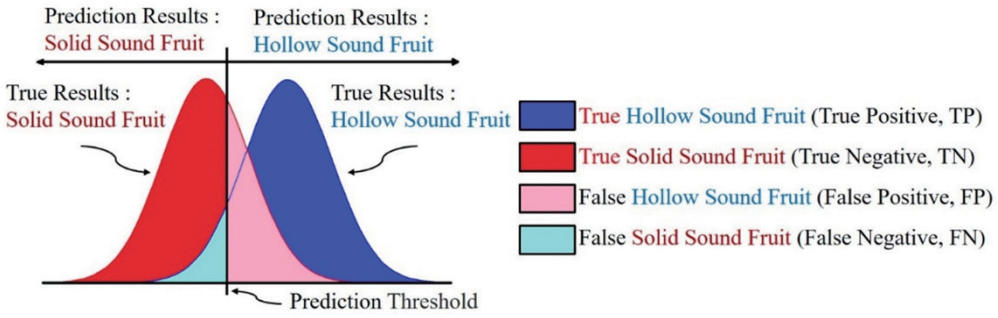


Fig. 6. (Color online) Distribution plot of hollow-sounding and solid-sounding fruits.

for the control group. A prediction threshold is defined to separate hollow-sounding and solid-sounding fruits, and predictions are generated on the basis of this threshold. Given the ground truth and predicted results, four possible outcomes are defined: true positive (*TP*), true negative (*TN*), false positive (*FP*), and false negative (*FN*). In this study, these outcomes are interpreted in the context of pineapple classification as follows: true hollow-sounding fruit, true solid-sounding fruit, false hollow-sounding fruit, and false solid-sounding fruit, respectively. To clearly present the confusion matrix and the corresponding numerical results, a confusion matrix is shown in Table 1, where a true hollow-sounding fruit refers to the case where both the actual condition and the predicted result indicate a hollow-sounding fruit, a true solid-sounding fruit refers to the case where both the actual condition and the predicted result indicate a solid-sounding fruit, a false hollow-sounding fruit refers to the case where the actual condition is a solid-sounding fruit, but the model predicts a hollow-sounding fruit, and a false solid-sounding fruit refers to the case in which the actual condition is a hollow-sounding fruit, but the model predicts a solid-sounding fruit. These four classification outcomes can be organized into a 2 × 2 confusion matrix.

To evaluate the feasibility of the classification model, the confusion matrix data—true hollow-sounding fruit, true solid-sounding fruit, false hollow-sounding fruit, and false solid-sounding fruit—were used to calculate the classification accuracy of pineapple texture. The true positive rate (*TPR*) for hollow-sounding fruits and the true negative rate (*TNR*) for solid-sounding fruits are defined as

$$TPR = \frac{TP}{TP + FN}, \tag{13}$$

$$TNR = \frac{TN}{FP + TN}, \tag{14}$$

where *TPR* represents the proportion of correctly predicted hollow-sounding fruits among all actual hollow-sounding fruit samples. The closer the number of true hollow-sounding samples is to the total number of actual hollow-sounding fruits, the closer the *TPR* value approaches 1. Similarly, *TNR* represents the proportion of correctly predicted solid-sounding fruits among all actual solid-sounding fruit samples. A higher number of true solid-sounding samples lead to a *TNR* value closer to 1. To determine the optimal classification threshold for distinguishing pineapple texture and to effectively assess the reliability of the confusion matrix data, we employed additional evaluation metrics, including accuracy, F_1 score, *MCC*, and ROC curve.

Table 1
Confusion matrix.

Predicted results	Ground truth	
	Hollow-sounding fruit	Solid-sounding fruit
Hollow-sounding fruit	True hollow-sounding fruit (TP)	False hollow-sounding fruit (FP)
Solid-sounding fruit	False solid-sounding fruit (FN)	True solid-sounding fruit (TN)

Accuracy (*acc*) is one of the most commonly used metrics for evaluating classification performance. It is calculated as the proportion of samples where both the actual condition and the predicted result are identified as either hollow-sounding or solid-sounding fruit, relative to the total number of samples. The formula for accuracy is

$$acc = \frac{TP + TN}{TP + FP + FN + TN}, \tag{15}$$

where *acc* ranges from 0 to 1. A value closer to 1 indicates higher classification performance, meaning that more samples are correctly classified as true hollow-sounding and true solid-sounding fruits. Conversely, a value closer to 0 indicates lower classification performance, meaning that more samples are misclassified as false hollow-sounding and false solid-sounding fruits.

We used *acc* to verify the number of pineapple texture samples, but it does not account for the classification performance specific to hollow-sounding fruit. Therefore, the F_1 score was adopted for evaluation. The F_1 score or *F*-measure is defined in this study as the harmonic mean of precision and recall for hollow-sounding fruit. The F_1 score is defined as

$$F_1 = \frac{2TP}{2TP + FP + FN}. \tag{16}$$

The F_1 score ranges from 0 to 1. A value closer to 1 indicates higher classification performance in terms of both precision and recall, meaning that more samples are correctly classified as true hollow-sounding fruit. Conversely, a value closer to 0 indicates lower classification performance, meaning that more samples are misclassified as false hollow-sounding and false solid-sounding fruits.

Acc and the F_1 score, which are commonly used evaluation metrics derived from the confusion matrix, do not fully take into account all elements of the matrix, especially in cases of imbalanced data. As a result, both metrics may lead to misleading conclusions. Therefore, *MCC* is adopted as a basis for distinguishing between different pineapple textures, as well as an indicator for determining the optimal decision threshold for the classification model. *MCC* is defined as

$$MCC = \frac{TP \times TN - FP \times FN}{\sqrt{(TP + FP)(TP + FN)(TN + FP)(TN + FN)}}. \tag{17}$$

Its value ranges between −1 and 1. A coefficient of 1 indicates perfect prediction, while −1 indicates complete disagreement between actual and predicted results. The closer the value is to 1, the greater the number of samples correctly classified as true hollow-sounding and true solid-sounding fruits. Conversely, the closer the value is to −1, the greater the number of samples misclassified as false hollow-sounding and false solid-sounding fruits, with these misclassified

samples exceeding the number of correctly classified ones. When the numbers of *TP*s for hollow-sounding fruit and *TN*s for solid-sounding fruit increase, and the numbers of *FP*s and *FN*s decrease, *MCC* becomes larger. Therefore, *MCC* can effectively be used to distinguish pineapple textures and serve as a criterion for determining the optimal decision threshold.

To understand the discriminative ability of the classification model, the ROC curve is used for evaluation. By varying the threshold, the numbers of true hollow-sounding fruit, true solid-sounding fruit, false hollow-sounding fruit, and false solid-sounding fruit samples can be obtained, from which the hollow-sounding and solid-sounding fruit accuracies are calculated. Figure 7 illustrates the threshold-based distinction of pineapple textures. The figure shows the ROC curve relationship at three threshold points. By adjusting the threshold from point A through point B to point C, each threshold corresponds to a specific pineapple texture classification result. For each threshold, the hollow-sounding and solid-sounding fruit accuracies are estimated. Each hollow-sounding fruit accuracy corresponds to one solid-sounding fruit accuracy; connecting all pairs of these accuracies forms the ROC curve, as shown in Fig. 8.

In Fig. 7, the blue distribution corresponds to samples with the actual label of hollow-sounding fruit, whereas the red distribution corresponds to samples with the actual label of solid-sounding fruit. In this schematic, samples with values above the threshold are classified as hollow-sounding fruit, while those below the threshold are classified as solid-sounding fruit. At threshold position A, the ratio of true to false hollow-sounding fruit samples is 10:0, yielding a *TPR* of 1.0. On the other hand, the ratio of true to false solid-sounding fruit samples is 2:8, resulting in a *TNR* of 0.2, and consequently, a false positive rate (*FPR*) of $1 - TNR = 0.8$. At threshold position B, the true to false hollow-sounding fruit ratio is 9:1, corresponding to a *TPR* of 0.9. The true to false solid-sounding fruit ratio is 5:5, yielding a *TNR* of 0.5 and an FPR of 0.5. At threshold position C, the true to false hollow-sounding fruit ratio is 7:3, resulting in a *TPR* of 0.7. The true to false solid-sounding fruit ratio is 8:2, resulting in a *TNR* of 0.8 and an FPR of 0.2. By varying the threshold and classifying pineapple textures accordingly, the resulting points are connected to form the ROC curve, as depicted in Fig. 8, where the *x*-axis represents *FPR* ($1 - TNR$) and the *y*-axis represents *TPR*. To assess the discriminative capability of the ROC

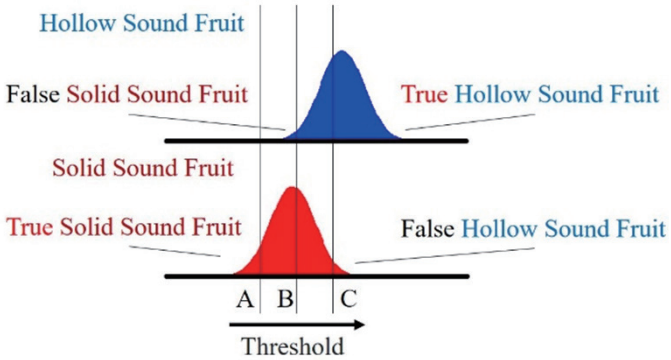


Fig. 7. (Color online) Illustration of threshold classification.

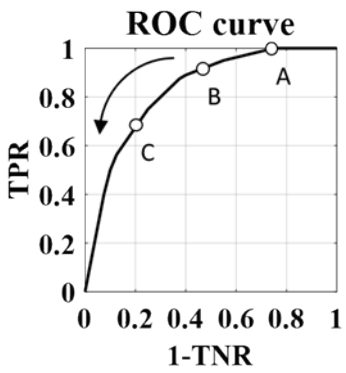


Fig. 8. Schematic of ROC curve.

curve, AUC is computed using the trapezoidal rule. Three ROC curve quantifications are illustrated in Fig. 9.

In Fig. 9, the blue curve represents hollow-sounding fruit, while the red curve represents solid-sounding fruit. AUC ranges from 0 to 1, where an AUC value closer to 1 indicates higher discriminative ability, and a value closer to 0.5 indicates no discriminative power. When AUC equals 1, the pineapple texture distributions are perfectly distinguishable. An AUC of 0.8 indicates effective discrimination, although some data overlap makes classification challenging in certain cases. An AUC of 0.5 suggests that the pineapple textures cannot be effectively distinguished. Therefore, the ROC curve shown in Fig. 8 has an AUC value of 0.8, indicating that most pineapple textures can be effectively classified, while a small portion remains difficult to identify.

3. Results and Discussion

In this section, we primarily describe the classification methods for pineapple texture and explain how machine learning is applied to discriminate pineapple textures. The classification results are evaluated using a confusion matrix to determine the optimal decision threshold using evaluation metrics. The discriminative ability of the classification model is further assessed through the ROC curve and AUC. Additionally, PCA and LDA are employed for the dimensionality reduction and classification of the dataset used in this study. A total of 240 pineapple samples were analyzed, with 182 samples classified as hollow-sounding fruit and 58 samples classified as solid-sounding fruit.

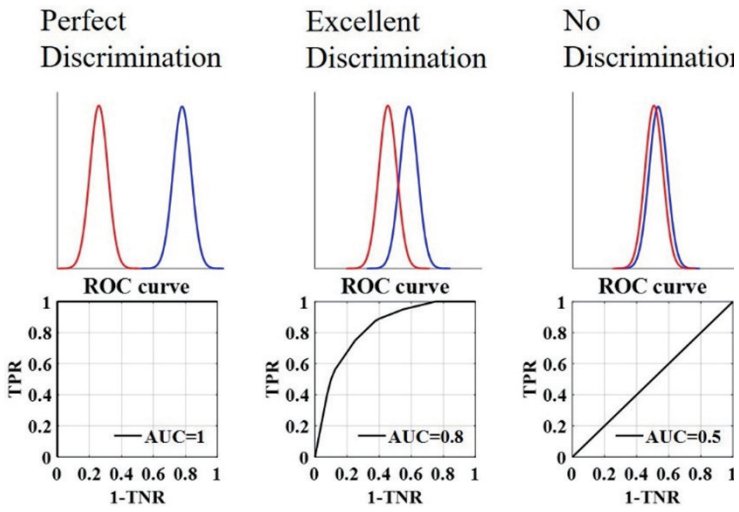


Fig. 9. (Color online) Three ROC curves and their corresponding AUCs.

3.1 Identification by PCA

To reduce the dimensionality of the dataset, PCA was employed for dimensionality reduction. Table 2 presents the eigenvalues of PCA. According to the explained variance, the first two principal components account for nearly 70% of the total variance (42.70% + 28.45%), indicating effective dimensionality reduction. The classification of hollow-sounding and solid-sounding fruits was performed on the basis of the 1st and 2nd PCs.

We utilized the 1st and 2nd PCs to compute the confusion matrix. Figure 10 and Table 3 present the classification results of the dataset after PCA. The classification of pineapple texture was determined using *MCC*, with threshold values set at −0.49 for the 1st PC and −0.80 for the 2nd PC. Samples with PC values above the threshold were classified as solid-sounding fruit, while those below were classified as hollow-sounding fruit. The overall accuracy was 0.61, indicating that 147 samples were correctly classified as either hollow-sounding or solid-sounding

Table 2
PCA eigenvalues.

Decision parameter	Eigenvalue	Explained variance
1st PC	2.172	42.70
2nd PC	1.416	28.45
3rd PC	0.880	17.67
4th PC	0.523	10.99
5th PC	0.010	0.19

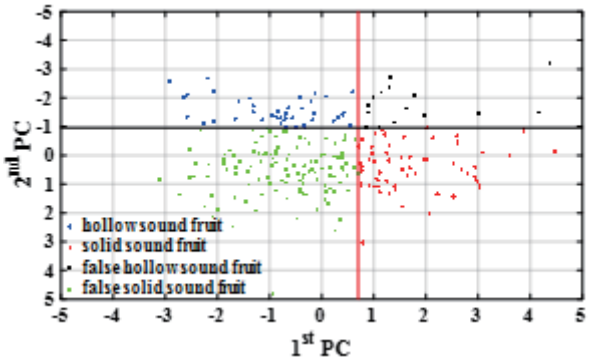


Fig. 10. (Color online) Classification results of 1st and 2nd PCs.

Table 3
Confusion matrix and evaluation metrics of 1st and 2nd PCs.

		1st PC	
		Hollow-sounding fruit	Solid-sounding fruit
2nd PC	Hollow-sounding fruit	33	33
	Solid-sounding fruit	60	114

Accuracy	0.61
*F*1 score	0.42
MCC	0.14

fruit. However, the PCA classification tended to be biased towards solid-sounding fruit. The F_1 score was 0.42, reflecting lower classification performance for hollow-sounding fruit. Figure 11 shows the ROC curves for PCA, where the AUC values for the 1st and 2nd PCs were 0.56 and 0.55, respectively, indicating that neither principal component exhibited significant discriminative power.

On the basis of the results of PCA, the optimal decision threshold was determined and further validated using the criteria employed by farmers in distinguishing pineapple texture. Figure 12 illustrates the classification results using the identified threshold. According to the threshold-based classification, 30 samples were identified as hollow-sounding fruit and 48 samples as solid-sounding fruit. The accuracy of the farmer-based classification was 0.33, indicating low identification performance when evaluated under the PCA-derived threshold criteria.

To fairly evaluate whether the first two principal components enable the classification of pineapples without label leakage, we also trained our model on 80% of the data and applied the learned thresholds to the 20% hold-out set. Using fivefold cross-validation on the training set to maximize *MCC*, the 1st PC achieved a test accuracy of 0.81, *MCC* 0.44, F_1 (hollow) 0.88, and F_1

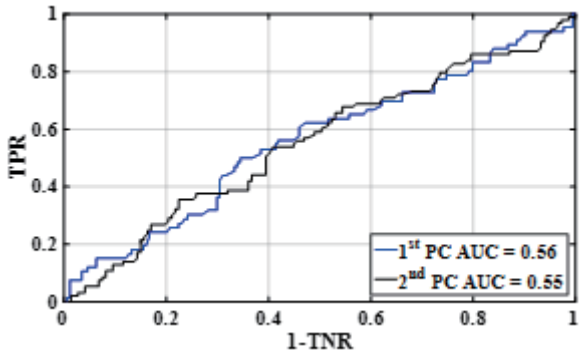


Fig. 11. (Color online) ROC curves for 1st and 2nd PCs.

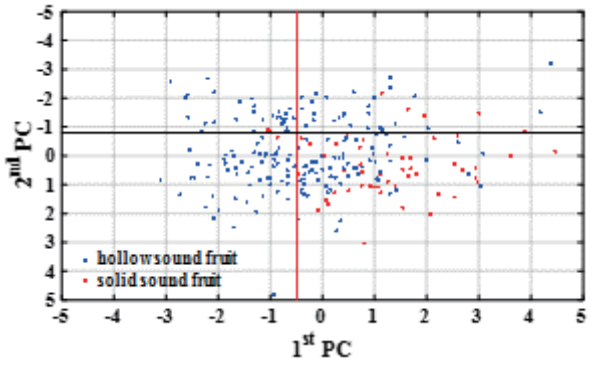


Fig. 12. (Color online) Classification results of pineapple texture using PCA-derived decision threshold.

(solid) 0.53, whereas the 2nd PC achieved 0.58 accuracy with *MCC* 0.10, confirming that individual PCs provide limited discriminative power. These test-set metrics are lower than the in-sample values and emphasize that PCA serves mainly for visualizing the data structure rather than for classification.

3.2 Identification by LDA

The results of PCA revealed that the classification tended to favor solid-sounding fruit, which does not align with the actual distribution. However, the model still maintained a classification accuracy of approximately 60%. Therefore, LDA was conducted for further analysis of pineapple texture. Table 4 presents the eigenvalues obtained from LDA. As shown, the 1st linear discriminant (LD) reaches a value of 1, while the eigenvalues for the 2nd LD and beyond approach zero. This indicates that the 1st LD captures the majority of the variance in the original data, making it sufficient to represent the dataset and to replace the original five variables for subsequent classification analysis.

The 1st and 2nd LDs were used to compute the confusion matrix. Figure 13 and Table 5 present the classification results of the dataset after applying LDA. *MCC* was used to determine

Table 4
LDA eigenvalues.

Decision parameter	Eigenvalue	Explained variance
1st LD	3654.681	1.00
2nd LD	8.19E−13	2.24E−16
3rd LD	5.71E−14	1.56E−17
4th LD	1.20E−18	3.29E−22
5th LD	2.72E−20	7.45E−24

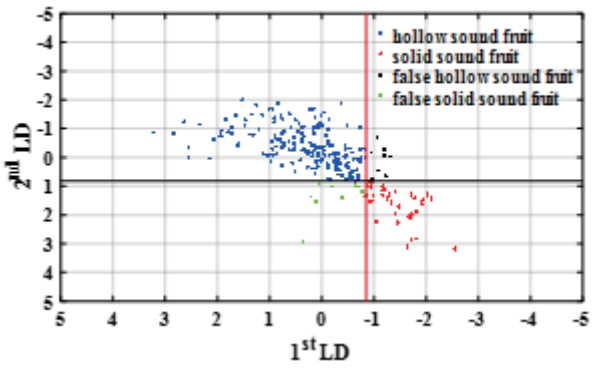


Fig. 13. (Color online) Classification results of 1st and 2nd LDs.

Table 5
Confusion matrices and evaluation metrics for 1st and 2nd LDs.

		1st LD	
		Hollow-sounding fruit	Solid-sounding fruit
2nd LD	Hollow-sounding fruit	186	9
	Solid-sounding fruit	10	35

Accuracy	0.92
*F*1 score	0.95
MCC	0.74

the decision thresholds for distinguishing pineapple texture. On the basis of the *MCC* values, threshold values were set at −0.85 for the 1st LD and 0.83 for the 2nd LD. For the 1st LD, samples with values above the threshold were classified as hollow-sounding fruit, while those below were classified as solid-sounding fruit. In contrast, for the 2nd LD, samples with values above the threshold were classified as solid-sounding fruit, while those below as hollow-sounding fruit. The overall classification accuracy was 0.92, indicating that 221 samples were correctly classified. The F_1 score reached 0.95, reflecting excellent classification performance for hollow-sounding fruit. Figure 14 shows the ROC curves for LDA, where both the 1st and 2nd LDs achieved an AUC value of 0.92, indicating that both discriminants have outstanding discriminative ability.

From the results of LDA, the optimal classification threshold was determined and further verified using the farmers' criteria for distinguishing pineapples. Figure 15 shows the classification results based on the threshold. Using this threshold, 160 hollow-sounding pineapples and 25 solid-sounding pineapples were identified. The farmers' classification accuracy was 0.77, indicating that LDA yields higher identification accuracy in line with the farmers' judgment.

To prevent data leakage, we then performed a group-stratified 80/20 split (ensuring all repeated measurements of a fruit stay in one set), trained LDA on the training set, selected the decision threshold via fivefold cross-validation maximizing *MCC*, and evaluated the classifier on the hold-out set. Under this rigorous protocol, the test accuracy dropped to 0.81, *MCC* 0.44, with F_1 scores of 0.88 (hollow) and 0.53 (solid), and a ROC-AUC of 0.91 with a PR-AUC of 0.97. Although lower than the original 0.92 accuracy, these results demonstrate good generalization without label leakage. To ensure that the discriminative performance observed in the full-data evaluation does not arise from label leakage, an additional independent test set experiment is conducted and presented in Sect. 3.3. The corresponding confusion matrix and ROC curve of this independent evaluation are summarized in Table 6 and Fig. 16.

3.3. Independent test set evaluation

To ensure fair evaluation and eliminate label leakage, we further assessed the proposed method on an independent 20% hold-out test set. All three repeated measurements from each

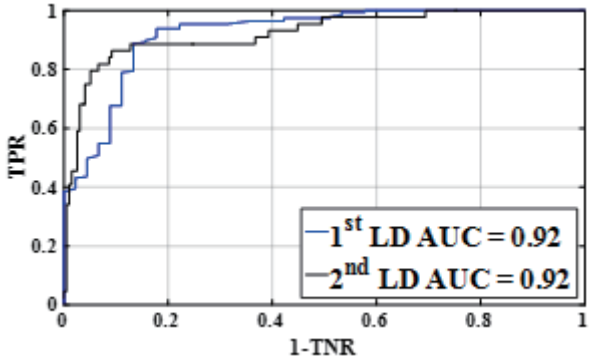


Fig. 14. (Color online) ROC curves for 1st and 2nd LDs.

Table 6
Confusion matrices and evaluation metrics for 1st and 2nd LDs.

		1st LD	
		Hollow-sounding fruit	Solid-sounding fruit
2nd LD	Hollow-sounding fruit	34	2
	Solid-sounding fruit	7	5

Accuracy	0.81
*F*1 score	0.71
MCC	0.44

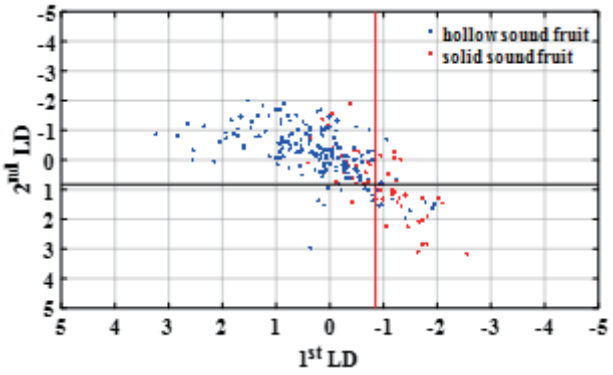


Fig. 15. (Color online) Classification result diagram of pineapple texture based on the discriminant threshold (LDA).

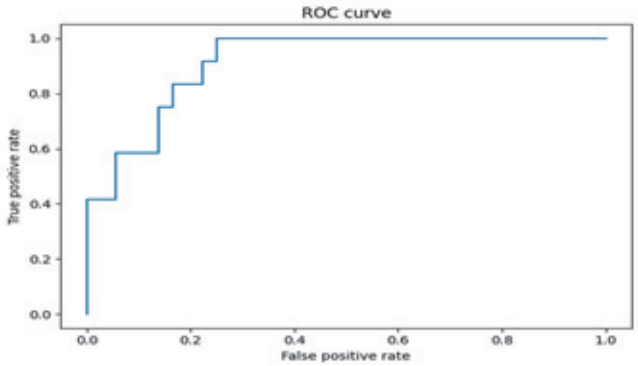


Fig. 16. (Color online) Precision–recall curve of the LDA classifier on the hold-out test set, with PR-AUC of 0.97.

fruit were grouped together and randomly assigned so that each fruit appeared entirely in either the training (80%) or test (20%) set. Within the training set, a fivefold cross-validation was performed to select the decision threshold that maximized *MCC* on the posterior probabilities. This threshold was then fixed and applied once to the hold-out test set.

On the test set, the first principal component (PC1) achieved an accuracy of 0.81, with an *MCC* of 0.44, F_1 scores of 0.88 for hollow-sounding fruits and 0.53 for solid-sounding fruits (average F_1 score of 0.71), a ROC-AUC of 0.91, and a PR-AUC of 0.97. PC2 achieved an accuracy of 0.58 and an *MCC* of 0.10 (results not tabulated), confirming that PCA serves primarily for

feature visualization rather than classification. The LDA classifier, trained and validated on the 80% training set and evaluated on the test set using the *MCC*-optimized threshold, reached a test accuracy of 0.81, an *MCC* 0.44, and F_1 score of 0.88 (hollow), and 0.53 (solid), consistent with PC1 performance. These figures are lower than the 92% accuracy reported in the full-data evaluation, but demonstrate the model's generalization ability without label leakage. The confusion matrix and key metrics for the independent test set are summarized in Table 6, and the corresponding ROC curve is shown in Fig. 16.

4. Conclusions

In this study, a machine learning approach was employed to classify pineapples on the basis of their measured physical and dynamic characteristics. A confusion matrix was used for discriminant analysis, and *MCC* was applied to determine the optimal classification threshold for distinguishing between hollow-sounding and solid-sounding fruits. The discriminative power of the classification model was further validated using the ROC curve and AUC. Finally, by introducing the concepts of PCA and LDA, the classification model was implemented to generate new projected data for the improved differentiation of pineapple texture types.

From the findings of this study, the following conclusions can be drawn:

1. Pineapple texture is commonly evaluated by tapping the fruit's outer surface. However, owing to the rough and complex structure of the skin, we selected the crown area of the pineapple for vibration excitation and signal measurement. The measured primary resonance frequency and corresponding amplitude were shown to enable the effective identification of the pineapple texture, indicating that the selected region can serve as a reliable reference for texture classification.
2. The confusion-matrix-based method used in this study successfully distinguished between hollow-sounding and solid-sounding pineapples. This approach can replace manual grading based on farmers' experience, significantly reducing labor requirements and classification errors, thereby improving the accuracy of texture discrimination.
3. PCA and LDA were introduced to process the measured pineapple features, allowing for the dimensionality reduction and extraction of relevant classification information. The results of PCA indicated that, although the analysis slightly favored solid-sounding fruits, the method effectively performed automatic feature selection while maintaining reasonable performance. When labels for hollow- and solid-sounding fruits were included, LDA yielded a clearly separable classification model. Verification using farmer-judged texture labels demonstrated good identification performance.

Acknowledgments

This research was supported by funding from the Industrial Technology Research Institute (ITRI) of Taiwan under Project No. DBET10902, for which the authors express their sincere gratitude. The authors also thank ITRI for providing comprehensive experimental facilities and a suitable research environment. The use of the internal texture detection module facilitated the

successful execution of the experiments and related analyses in this study. The authors also wish to thank the anonymous reviewers for their valuable comments and suggestions in revising the paper.

Author Contributions

The author contributions declaration of this paper is provided as follows. Chang-Sheng Lin: Conceptualization, methodology, supervision, and manuscript drafting. Chih-Chung Hsu: Data curation, software development, and formal analysis. Chin-Hui Yeh: Experimental setup, data acquisition, and validation. Chi-Han Tsai: Data processing and visualization. Hung-Tse Chiang: Resources, project administration, and technical support.

Competing Interest Declaration

The authors of this paper declare no competing interests as defined by Springer, or other interests that might be perceived to influence the results and/or discussion reported in this paper.

Data availability

This manuscript does not report data generation or analysis.

References

1 S. Pathaveerat, A. Terdwongworakul, and A. Phaungsombut: J. Food Eng. **89** (2008) 112. https://doi.org/10.1016/j.jfoodeng.2008.04.012
2 C. Ay: Proc. 2000 ASAE Annual International Meeting (ASAE, Milwaukee, WI, USA, 2000).
3 C. Ay: Proc. 2004 CIGR International Conf. **1** (2004).
4 R. Abbaszadeh and A. Moosavian: J. Food Sci. Technol. **52** (2015) 1075. https://doi.org/10.1007/s13197-013-1086-5
5 M. Taniwaki and N. Sakurai: J. Jpn. Soc. Hortic. Sci. **79** (2010) 113. https://doi.org/10.2503/jjshs1.79.113
6 T. Tiplica, P. Vandewalle, S. Verron, and C. Grémy-Gros: Proc. Int. Metrology Conf. (2010) 1–9.
7 H. Zhang, J. Wu, and H. Ma: Postharvest Biol. Technol. **148** (2019) 151. https://doi.org/10.1016/j.postharvbio.2018.10.011
8 D. Ballabio, F. Grisoni, and R. Todeschini: Chemom. Intell. Lab. Syst. **174** (2018) 33. https://doi.org/10.1016/j.chemolab.2017.12.018
9 M. Bekkar, H. K. Djemaa, and T. A. Alitouche: J. Inf. Eng. Appl. **3** (2013) 27
10 A. L. B. Brito, L. R. Brito, F. A. Honorato, M. J. C. Pontes, and L. F. B. L. Pontes: Food Res. Int. **51** (2013) 924. https://doi.org/10.1016/j.foodres.2013.02.025
11 D. Chicco and G. Jurman: BMC Genomics **21** (2020) 6. https://doi.org/10.1186/s12864-019-6413-7
12 T. Fawcett: Pattern Recognit. Lett. **27** (2006) 861. https://doi.org/10.1016/j.patrec.2005.10.010
13 A. Tharwat: Appl. Comput. Inform. **15** (2019) 1. https://doi.org/10.1016/j.aci.2018.08.003
14 T.-W. Huang, S. A. Bhat, N.-F. Huang, C.-Y. Chang, P.-C. Chan, and A. R. Elepano: Agriculture **12** (2022) 129. https://doi.org/10.3390/agriculture12020129
15 Y. J. Chen, Y. C. Liou, W. H. Ho, J. T. Tsai, C. C. Liu, and K. S. Hwang: Sci. Prog. **104** (2022) 1. https://doi.org/10.1177/00368504221110856
16 S. Arwatchananukul, S. Chaiwong, N. Aunsri, S. Kittiwachana, K. Luengwilai, T. Trongsatitkul, and R. Saengrayap: Smart Agric. Technol. **9** (2024) 100601. https://doi.org/10.1016/j.atech.2024.100601

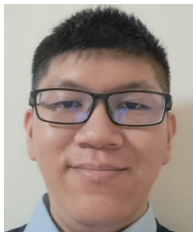
About the Authors



Chang-Sheng Lin received his Ph.D. degree from the Department of Aeronautics and Astronautics of National Cheng Kung University, Taiwan, in 2011. He joined the Faculty of National Kaohsiung University of Science and Technology, Taiwan, in 2024 where he is currently an associate professor at the Department of Mechanical Engineering. His research interests are random vibration and modal analysis. (changsheng@nkust.edu.tw)



Chih-Chung Hsu received his Ph.D. degree in Electrical Engineering from National Tsing Hua University, Taiwan, in 2014. He is currently an associate professor at the Institute of Smart Industry and Green Energy, National Yang Ming Chiao Tung University, Taiwan, and serves as the NYCU Young Chair Professor. His research interests include computer vision, trustworthy artificial intelligence, efficient deep learning, autonomous visual perception, and hyperspectral image analysis.



Chin-Hui Yeh received his B.S. and M.S. degrees from the Department of Vehicle of National Pingtung University of Science and Technology, Taiwan in 2019 and 2021, respectively. He is currently an engineer at LIMO System Technology Co., Ltd., Kaohsiung, Taiwan.



Chi-Han Tsai received his M.S. degree in Physics from National Chiao Tung University, Taiwan. He is currently pursuing his Ph.D. degree at the College of Artificial Intelligence, National Yang Ming Chiao Tung University, Taiwan, and is a member of the Advanced Computer Vision Lab (ACVLab). His research interests include computer vision, artificial intelligence, statistical mechanics, and population dynamics.



Hung-Tse Chiang received his B.S. and M.S. degrees from the Department of Vehicle of National Pingtung University of Science and Technology, Taiwan in 2019 and 2021, respectively. He is currently an engineer at Chroma ATE Inc., Taoyuan, Taiwan.