Nondestructive discrimination of moldy pear core based on the recurrence plots of vibration acoustic signals and deep convolutional neural networks

Yue Yang[1], Kang Zhao[1,2,3*], Jin Zhao[1], Ye Song[2,3], Tao Shen[1]

(1. *School of Electrical Engineering, University of Jinan, Jinan 250022, China*;
2. *Jinan Fruit Research Institute, All-China Federation of Supply and Marketing Cooperatives, Jinan 250220, China*;
3. *China Technology and Research Center for Storage and Processing of Fruit and Vegetables, Jinan 250220, China*)

Abstract: Moldy core is a serious internal defect in pears. Since there is no significant difference in appearance between the healthy pears and those with mild moldy core, it is still a great challenge for the early detection of moldy pear core. This study transformed the vibration acoustic signals (VA signal) of pears into recurrence plots and Markov transition field to enable image-based classification of moldy cores. In addition to traditional machine-learning baselines (Random Forest and k-Nearest Neighbors) trained on LBP-extracted texture features from RP/MTF, the deep models were constructed and compared, which include ResNet-101, DenseNet-121, SqueezeNet, Vision Transformer (ViT), and an improved SqueezeNet (ISqueezeNet). Hyperparameters were tuned via Bayesian optimization over optimizer type, learning rate, batch size, and L2 weight decay, yielding model-specific optimal settings. Under these configurations, the ISqueezeNet achieved the highest test accuracy of 93.05%, with class-wise accuracies of 89.28% (healthy), 96.15% (slight), and 94.44% (moderate and severe). Comparisons with lightweight networks (MobileNetV1 and ShuffleNetV2) further showed that ISqueezeNet attains superior accuracy with favorable parameter efficiency and inference speed. Grad-CAM visualizations confirmed that the model focuses on lesion-relevant regions, supporting interpretability and practical reliability. These results indicate that the proposed approach is promising for early, nondestructive detection of moldy pear cores.

Keywords: moldy pear core, deep convolutional neural network, vibration acoustic signals, recurrence plots, Markov transition field, nondestructive detection

DOI: 10.25165/j.ijabe.20251806.9916

Citation: Yang Y, Zhao K, Zhao J, Song Y, Shen T. Nondestructive discrimination of moldy pear core based on the recurrence plots of vibration acoustic signals and deep convolutional neural networks. Int J Agric & Biol Eng, 2025; 18(6): 230–240.

1 Introduction

Korla pear (*Pyrus bretschneideri Rehd*) is popular with worldwide consumers due to its rich nutritional components and juicy texture[1]. However, fresh pears are susceptible to postharvest diseases caused by pathogenic infections during growth and storage, which increases the risk of quality loss and export rate drop. Moldy core is one of the most serious pathological diseases found in pear fruit[2]. As a typical internal disease, the development of a moldy core begins with the pear kernel followed by the pulp eroding outward. Early development of the pears infected with moldy core has no visible external symptoms and they are difficult to distinguish from healthy pears in appearance[3]. Due to the severe infectivity of moldy core, the diseased pears will further decay and face the risk of infection for the whole consignment of healthy pears, leading to serious economic losses[4]. Noteworthy, when the moldy pear core is mixed into the healthy fruit at the deep processing stage, the moldy tissue with many toxins could exceed the prescribed limit of mycotoxins in derived products such as pear juice, pear sauce, pear cans, etc., posing potential food safety hazards[5]. Traditionally, the pears with moldy core are detected by destructive sampling combined with visual inspection by trained personnel, followed by overall assessment[6]. However, these methods are time-consuming and laborious and suffer from drawbacks of inefficiency and subjectivity[1,7]. Thus, it would be of practical significance to develop a rapid and nondestructive detection technique for moldy pear core, so as to remove the diseased pears from the postharvest storage and marketing chain.

With the advent of emerging technologies, much literature has been published on non-destructive detection of internal diseases in fruit using X-ray technology[8,9], near-infrared spectroscopy[10-12], hyperspectral imaging[13,14], electronic nose[15,16]. Compared with other nondestructive detection technologies, the vibro-acoustic method is increasingly adopted to evaluate fruit quality attributes due to its simplicity, speed, and low-cost[7]. However, there are few studies on VA methods for discriminating moldy pear core. Notably, the feasibility of using the acoustic vibration technique for identifying internal disease in fruit has been demonstrated by some researchers. Zhang et al.[17] applied the acoustic vibration method to realize an overall recognition accuracy of 86.4% for discriminating mild-browning pears. Han et al.[18] used this method to detect moldy kernels of in-shell hickory nuts with 91.67% discrimination accuracy. For these reported studies, most of the researchers mainly focused on using traditional machine learning algorithms. The effectiveness of these algorithms relied largely on the extraction and

Received date: 2025-05-23 **Accepted date:** 2025-09-03

Biographies: Yue Yang, MS candidate, research interest: intelligent detection and intelligent information processing, Email: yangyue9395@163.com; **Jin Zhao**, MS candidate, research interest: intelligent detection and intelligent information processing, Email: ZJzhaojin1561@163.com; **Ye Song**, Researcher, research interest: rapid identification and traceability of fruit and vegetable quality, Email: songye_sy@163.com; **Tao Shen**, Professor, research interest: intelligent perception and signal processing, Email: cse_st@ujn.edu.cn.

***Corresponding author: Kang Zhao**, Postdoc Researcher, research interest: intelligent perception and signal processing. School of Automation and Electrical Engineering, University of Jinan, Jinan 250022, China, Tel: +86-13589000213, Email: cse_zhaok@ujn.edu.cn.

selection of features that can reflect disease information. While these handcrafted features required the knowledge and experience of domain experts, which has a negative effect on the improvement of discrimination performance and generalization ability of the classification model[19].

To address these challenges, deep learning has been proposed and successfully employed in various fields. A strong advantage of deep learning is feature learning. It can adaptively learn and extract deeper and more discriminative feature information from raw data. At present, deep learning has gradually become a powerful tool for feature mining in existing studies[20-22]. Despite this, there are still few reports on the multi-classification discrimination of deep learning algorithms represented by deep convolutional neural networks in recognition of internal infections of fruits and vegetables. Cai et al.[23] established a general CNN model for the classification of citrus species with a detection accuracy of 95%. Ünal et al.[24] demonstrated that the deep learning network had high accuracy for fruit and vegetable classification applications. Shin et al.[25] optimized and evaluated several well-established learners, which proved that deep convolutional neural networks had great potential in image processing.

In recent years, one of the development trends in the field of signal processing is the combination of signal visualization and deep learning. Seo and Nam[26] converted the wireless signals into recurrence plots (RP) and fed them into a CNN network for discrimination with 99% classification accuracy. It has been observed that the recurrence plots have the ability to fully describe the energy distribution of VA signal. This method has been successfully implemented in various studies related to machine fault diagnosis[27-28]. Compared with traditional handcrafted features, RP-based representations provide a more comprehensive description of nonlinear and dynamic characteristics in signals, reducing dependence on expert knowledge and improving generalization. Currently, there are relatively few studies on the utilization of deep learning combined with the recurrence plots of acoustic vibration signals for discriminating internal diseases in fruits.

The goal of this research was to identify moldy pear core using VA method and deep learning. Specific objectives were to: (1) Develop a moldy pear core detection device based on piezoelectric transducers that can rapidly acquire VA signals; (2) Convert VA signals into recurrence plots and establish classification models using deep learning algorithms for identifying the moldy pear core. (3) Evaluate the classification performance of the established models and determine the best discrimination algorithm.

2 Materials and methods

2.1 Sample preparation

The Korla pears were selected as experimental subjects and all samples were purchased from fruit wholesale market on Dikou Road, Jinan, China. The samples without physical defects were stored in a refrigerator compartment (0°C and RH: 85%-90%). Since there was no visible difference between healthy and slight moldy-core pears, it was difficult to procure the diseased samples that met the requirements in supermarket. Thus, the pears with different extents of moldy core were artificially prepared by referring to the method mentioned by Zhang et al.[2]. The basic steps of diseased sample preparation were as follows: (1) According to Koch's postulates, five lesion tissues were cut from the core of naturally diseased pears and cultured on the PDA medium. (2) Pour an appropriate amount of sterile water into the *Alternaria* culture dish. After stirring, filtering, and counting, the spore concentration was adjusted to 2.0×10^8/mL of bacterial suspension for later use. (3) Using the wound inoculation method, 0.05 mL of composite bacterial suspension was injected from the calyx into the core of the pears. (4) The injected samples were stored in the conditions of an intelligent artificial climate chamber (25°C and 95% RH). After 3 d, the moldy tissue of the treated samples was generated in the inoculated core area. The prepared samples were subjected to VA test to collect the VA signals of pears with varying extents of moldy core.

2.2 VA signals collection and preprocessing of pears

The VA signal testing system was used to collect the VA response signals from pears with different extents of moldy core. As shown in Figure 1, the system consisted of four components: a vibration control and dynamic signal analyzer, a voltage amplifier, a test bench with three identical piezoelectric sensors, a computer with the SA 1804B2 software. Referring to Li et al.[29], a half-sine pulse signal V_E with a peak voltage of 2.5 V and a time history of

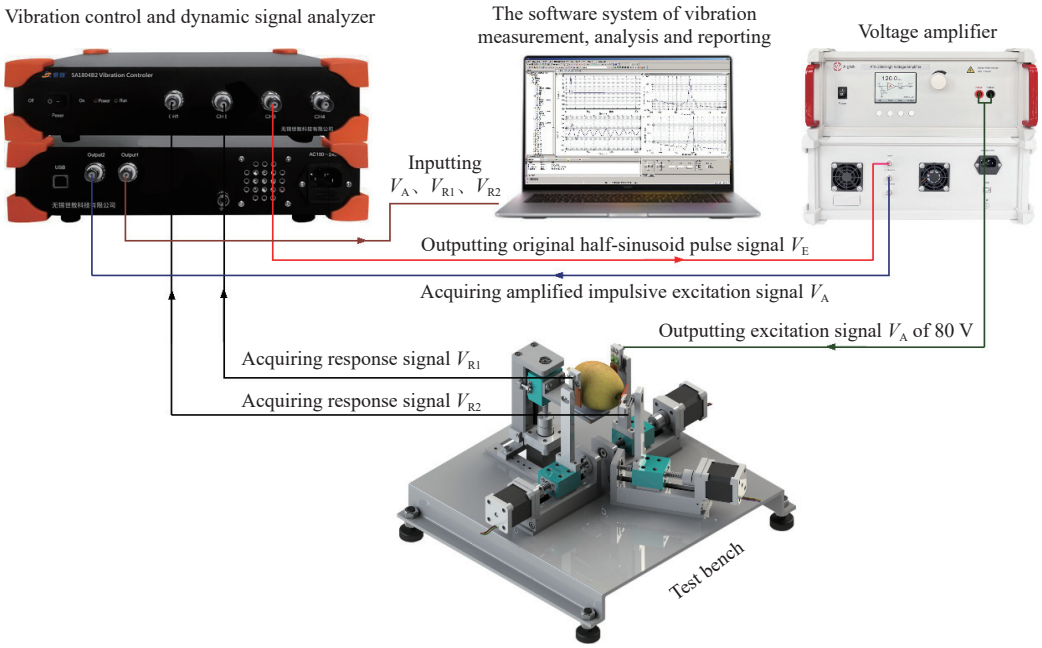


Figure 1 Nondestructive detection system based on the VA method for moldy pear core

0.16 s was recorded as the excitation signal. To enhance the signal excitation and improve anti-interference ability, the ATA-2000 voltage amplifier (Aigtek Electronic Technology Co., Ltd., Xi'an, China) was used to linearly amplify the original 2.5 *V* signal V_E into an excitation signal V_A with a peak voltage of 80 *V*. The amplified excitation signal V_A was output to the excitation sensor in contact with the equator of the pear samples. Subsequently, the response signals V_{R1} and V_{R2} were sensed and collected by the piezoelectric sensors in contact with the equator and calyx on the other side of the pear. Ultimately, the V_A, V_{R1}, and V_{R2} signals were collected by the vibration control and dynamic signal analyzer.

2.3 Pear moldy-core degree measurement

To measure the extent of moldy core, the pear samples were cut horizontally along the equator after the acoustic vibration testing. The cross-section of pears was photographed with a camera. The cross-section images were processed by the specific procedure as shown in Figure 2. The RGB three-channel components of the pear color photos were extracted and then the B channel images were segmented using the edge detection and segmentation method. The moldy-core region in the B channel images was extracted using the 8-connected region labeling algorithm. Finally, the ratio value of the moldy-core area S_1 to the pear cross-sectional area S_2 was defined as the disease extent *S* of moldy-core in pears.

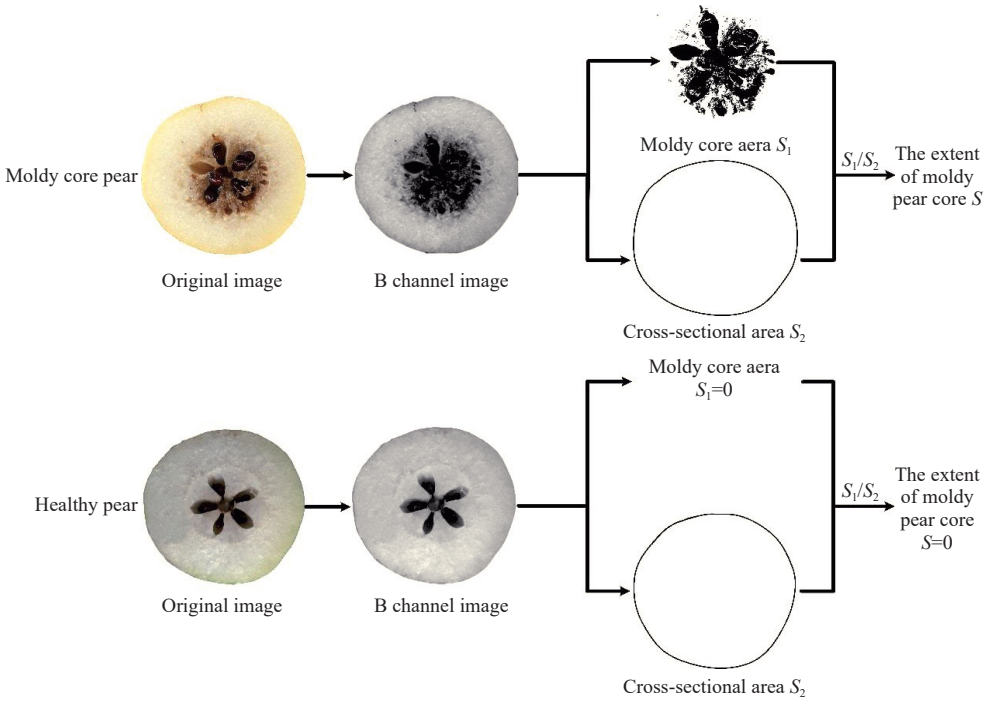


Figure 2 Diagrammatic sketch of the calculation procedure for the extent of moldy pear core.

2.4 Conversion of VA signal

2.4.1 Recurrence plot

Recurrence plot (RP) was an important method for feature enhancement that can transform the one-dimensional signals into two-dimensional images. Thus, the VA signals of both healthy pears and pears with different extents of moldy-core were subjected to the recurrence plot transformation. The Taken embedding theorem was used to reconstruct the original VA signals in phase space. The principle is as shown in Equation (1):

$$X_i = \{x_i, x_{i+\tau}, \ldots, x_{i+(m-1)\tau}\}(i = 1, 2, \ldots, N) \quad (1)$$

where τ and m present the delay coefficient and embedded dimension, $N = (m-1)\tau$, X_i is the vector of x_i after mapping in phase space. The recursive matrix R_{ij} is calculated as follows:

$$R_{ij} = \theta(\varepsilon - \|X_i - X_j\|) \quad (2)$$

where, ε is the distance threshold between two points in phase space, $\|\cdot\|$ is the Euclidean distance, $\theta(x)$ is the Heaviside function. The calculation equation of $\theta(x)$ is as follows:

$$\theta(x) = \begin{cases} 1, & x > 0 \\ 0, & x \le 0 \end{cases} \quad (3)$$

In recurrence plot, the embedding dimension m, the delay coefficient τ, and the distance threshold ϵ are three key parameters. When m is small, the reconstructed phase space may overlap or even self-intersect, leading to significant discrepancies between the reconstructed phase space and the original signal. A larger m increases computational complexity and amplifies the noise in the signal. The delay coefficient τ determines the amount of information retained in the reconstructed phase space. Smaller values of τ result in excessive correlation between the delayed coordinates, causing the phase space trajectory to concentrate around the main diagonal. Conversely, a larger τ can lead to the mutual correlation between the delayed coordinates becoming independent, making the phase space trajectory more dispersed. Currently, there is no adaptive algorithm to select an optimal threshold for ϵ. Therefore, in this study, ϵ was empirically set to 5% of the maximum phase space diameter. The study thus explored the effects of m and τ on the discrimination ability of different signals. Based on the preliminary experiment, the optimal spectral parameters were determined as m=3 and τ=2. The original VA signal was then transformed into a recurrence plot of size 224×224 or 227×227, which was used as input to different CNN models for constructing a classification model for moldy pear cores.

2.4.2 MTF

Markov Transition Field (MTF) was an image coding method that converts a one-dimensional signal into a two-dimensional image by embedding the temporal dynamics of state transitions. In

this approach, the original VA sequence is first discretized into Q quantile-based bins, and each point is assigned to one of the discrete states. A first-order Markov transition matrix W is then constructed, where each entry w_{ij} indicates the probability of transitioning from state i to state j:

$$w_{ij} = \frac{\text{Number of transitions from state } i \text{ to state } j}{\text{Total number of transitions from state } i}, \quad \sum_{j=1}^{Q} w_{ij} = 1 \quad (4)$$

Unlike recurrence plots, which capture spatial proximity in phase space, the MTF preserves sequential dependencies by mapping the transition probabilities back to the temporal domain. The resulting field is defined as:

$$\mathrm{MTF}_{x,y} = w_{q_x q_y}, \quad x, y \in [1, T] \quad (5)$$

where, T is the signal length; and q_x, q_y are the quantized states at times x and y. The resolution of the field is mainly determined by quantization number Q. A very small Q may lose fine-grained information, while an excessively large Q can introduce redundancy and noise. In this study, Q is set to 8.

2.5 Establishment of machine learning models

2.5.1 LBP feature extraction

Local Binary Pattern (LBP) is a widely used texture descriptor that transforms local structural information of an image into binary codes, thereby providing a compact and discriminative representation. The basic idea of LBP is to compare each pixel with its neighboring pixels within a defined window. If the neighbor intensity is greater than or equal to the center pixel, the corresponding position is assigned a value of 1; otherwise, it is assigned 0. The resulting binary sequence is then converted into a decimal number to represent the local texture pattern. Mathematically, the LBP operator for a pixel at position (x, y) can be expressed as:

$$\mathrm{LBP}_{P,R}(x, y) = \sum_{P=0}^{P-1} s(g_p - g_c) \cdot 2^p \quad (6)$$

where, g_c denotes the gray value of the center pixel, g_p is the gray value of the p-th neighbor, P is the number of sampling points on a circle of radius R, and the function $s(x)$ is defined as:

$$s(x) = \begin{cases} 1, & x \geq 0 \\ 0, & x < 0 \end{cases} \quad (7)$$

The parameters P and R jointly determine the resolution and scale of the LBP operator. A larger P captures more detailed neighborhood information, while a larger R extends the receptive field to encode broader spatial structures. In this study, LBP is used to extract texture features of RP and MTF for subsequent machine learning model training and testing.

2.5.2 Classifier construction and training

In this study, two widely used machine learning classifiers, Random Forest (RF) and k-Nearest Neighbors (KNN), were employed in this study to evaluate the classification performance of the extracted features. RF is an ensemble learning method that integrates multiple decision trees based on bootstrap aggregation and random feature selection. By constructing a large number of decision trees and combining their outputs through majority voting, RF effectively reduces variance and improves classification stability. The corresponding basic principle can be described as follows:

$$\hat{y} = \arg\max_{c} \frac{1}{T} \sum_{t=1}^{T} \mathbb{I}(h_t(\boldsymbol{x}) = c) \quad (8)$$

where, $\hat{y}$ is the predicted class label, T is the number of decision trees, $h_t(\cdot)$ denotes the prediction of the t-th tree, and $\mathbb{I}(\cdot)$ is the indicator function.

KNN is a non-parametric classification algorithm that predicts the category of a test sample according to the majority class among its k nearest neighbors in the feature space. The principle can be expressed as:

$$\hat{y} = \arg\max_{c} \sum_{i=1}^{k} \mathbb{I}(y_i = c) \quad (9)$$

where, $\hat{y}$ is the predicted class label, k is the number of nearest neighbors, y_i is the label of the i-th nearest neighbor, and $\mathbb{I}(\cdot)$ is the indicator function.

2.6 Establishment of deep classification models

In this study, the deep convolutional neural networks (including ResNet-101, DenseNet-121, SqueezeNet, Vision Transformer and ISqueezeNet) were used to construct the classification models of moldy pear core. The basic properties and configurations of these networks, such as image input size, depth, size, and parameter count, are listed in Table 1.

Table 1 Basic properties and configurations of the deep convolutional neural networks

Networks	Image dimension	Depth	Size/MB	Parameters (×10⁶)
Squeeze Net	227×227	18	4.6	1.2
ResNet-101	224×224	101	45.5	44.6
DenseNet-121	224×224	121	32	7.98
Vision Transformer	224×224	12	88	5.7
ISqueezeNet	227×227	18	0.18	0.05

2.6.1 Vision transformer

Vision Transformer (ViT) is a transformer-based neural network architecture that applies the self-attention mechanism to image recognition tasks. The network structure of ViT is shown in Figure 3a. The patch embedding module was the basic unit of ViT, in which the input image was divided into a sequence of fixed-size patches and linearly projected into embeddings with position information. These patch embeddings, together with a learnable classification token, were then fed into multiple stacked Transformer encoder layers, each composed of multi-head self-attention and feed-forward networks. Finally, the representation was passed through a fully connected layer to generate the class prediction.

2.6.2 ResNet-101

ResNet-101 with residual block can effectively address problems related to gradient disappearance and explosion[30]. The specific structure of ResNet-101 network is shown in Figure 3b. The ResNet-101 comprised a 7×7 convolutional layer, an average pooling layer, and 4 stages. Each stage contains the residual blocks consisting of two 3×3 convolutional layers, the batch normalization, and the ReLU activation function.

2.6.3 DenseNet-121

DenseNet-121 is a deep neural network architecture based on dense connections. The densely connected structure was realized by connecting the outputs of all previous layers between each layer. DenseNet-121 has been shown to be suitable for mining deep features related to classification[31]. As shown in Figure 3c, the DenseNet-121 can be described as consisting of multiple densely connected dense blocks. The feature maps between dense blocks were connected through transition layers to control the size of the feature maps.

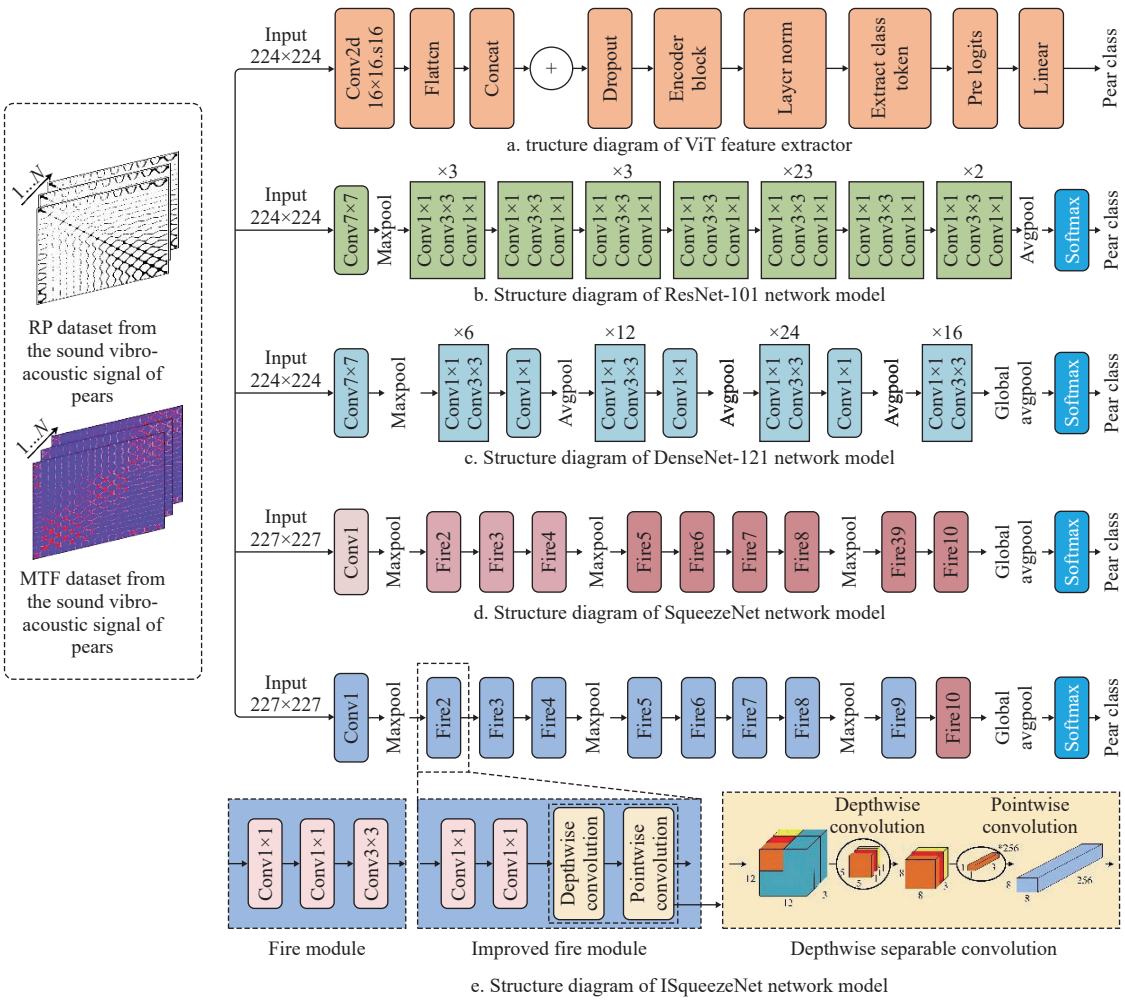


Figure 3 Flow chart of the construction of classification model for identifying the pear with moldy core

2.6.4 SqueezeNet

SqueezeNet is a lightweight neural network architecture that can not only reduce computational load but also maintain good computational performance. The network structure of SqueezeNet is shown in Figure 3d. The fire module was the basic unit of SqueezeNet, which was composed with the squeeze layer using 1x1 convolution kernels and the expand layer with a mixture of 1x1 and 3x3 convolution kernels.

2.6.5 ISqueezeNet

Depthwise separable convolution, as a key technology in convolutional neural networks, has achieved remarkable results in the field of computer vision in recent years. Inspired by this, the traditional convolution of the Fire module in SqueezeNet was replaced by a depth-wise separable convolution in this study. As shown in Figure 3e, the deep separable convolution divided the traditional convolution operation of the convolution kernel into two parts: depthwise convolution and pointwise convolution. For the depthwise convolution, M convolution kernels of size $k \times k$ were used to convolve the feature maps input from each channel and obtain M feature maps. Subsequently, the point-wise convolution was performed on the feature maps obtained by the depth convolution. A 1×1 convolution kernel was used to linearly combine the feature maps on each channel to generate the final feature maps. Through the above two steps, the depth-wise separable convolution can learn spatial features and channel features, and reduce the computational complexity.

2.7 Hyperparameter optimization methods

The model training environment used in the experiment was a Windows 10 64-bit operating system. The graphics card was an Nvidia 4060Ti GPU with 16 G video memory. The hardware configuration CPU was an Intel (R) Core (TM) i9-13900HX with a 2.20 GHz main frequency and 32 GB memory. The Python programming language was implemented in PyCharm software. During training, Bayesian optimization was used to automatically adjust the hyperparameters of the model. The search parameters included the initial learning rate (ranging from 1e-5 to 1e-2), batch size (32, 64, and 128), L2 regularization coefficient (1e-6 to 1e-3), and optimizer (Adam, SGD).

2.8 Model performance evaluation

To evaluate the classification performance of four models, the accuracy, Precision, Recall, and F_1 were utilized as evaluation metrics and calculated by the confusion matrix analysis. The specific calculation formulas are as follows:

$$\text{Accuracy} = \frac{\text{TP}+\text{TN}}{\text{TP}+\text{TN}+\text{FP}+\text{FN}} \tag{10}$$

$$\text{Recall} = \frac{\text{TP}}{\text{TP}+\text{FN}} \tag{11}$$

$$\text{Recall} = \frac{\text{TP}}{\text{TP}+\text{FN}} \tag{12}$$

$$F_1 = 2 \times \frac{\text{precision} \times \text{recall}}{\text{precision} + \text{recall}} \tag{13}$$

3 Results and discussion

3.1 Analysis of RP images of VA response signals

Figure 4 shows the typical VA responses signal curves of healthy pears and pears with varying extents of moldy core. It can be found that there was no significant differentiation in the VA signal curves of the three categories of pears at the first peak of approximately 0.018 s. In the following peaks, the difference in signal peak values between healthy pears and slight moldy-core pears was very small. With the severity of lesion tissue intensified, the energy of the signal curves of moderate and severe moldy-core pears gradually decreased. Nevertheless, the corresponding peak value was significantly higher than that of the healthy and slight moldy-core pears. Therefore, it was still difficult to extract the effective characteristics from VA curves to establish the classification model of moldy pear core. However, the RP and MTF converted from VA signals of the three-class pears were notably different in visualized texture distribution. To demonstrate the effectiveness of the RP and MTF for discriminating the moldy pear core, the recurrence plots of three-class pears were sent into the deep learning algorithm for model training and classification.

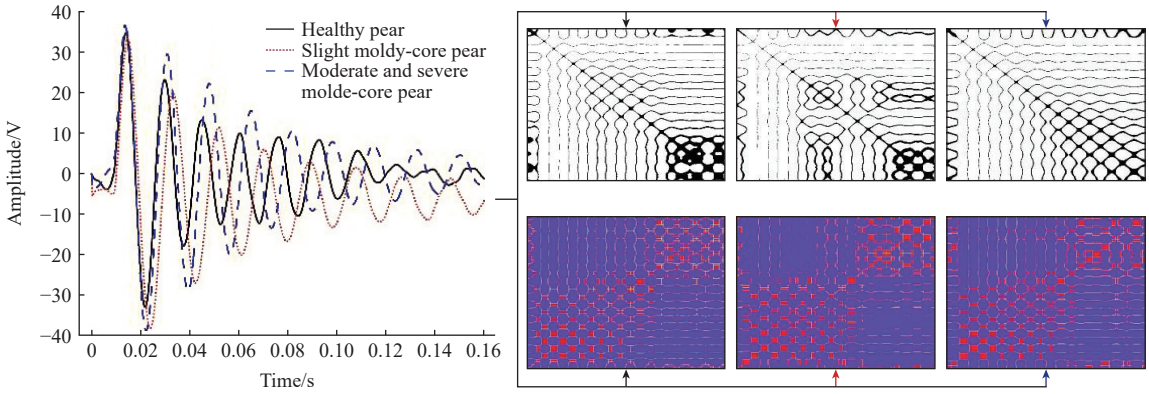


Figure 4 The RP and MTF converted from the typical VA signals of the three types of pears

3.2 Training and parameter optimization results of different classification models

3.2.1 Training results of machine learning models

LBP-RP and LBP-MTF texture features were employed as inputs to construct RF and KNN classification models, respectively. The corresponding classification results are summarized in Table 2. A comparative analysis indicated that models trained with LBP-RP texture features achieved superior overall classification accuracy compared with those trained on LBP-MTF features, suggesting that RP provides a more discriminative representation of temporal dynamic information. Specifically, the RF and KNN models based on LBP-RP features attained overall accuracies of 95.65% and 98.33%, respectively, whereas the RF and KNN models based on LBP-MTF features achieved 92.67% and 96.52%, respectively. These findings confirm that LBP-RP features can more effectively capture the texture distribution patterns of signals after phase-space mapping. Consequently, this highlights the importance of selecting appropriate feature representations to enhance classification performance in the nondestructive detection of moldy pears. Based on these results, the subsequent analysis focused on the two classifiers trained with LBP-RP texture features.

Table 2 Different model classification results of LBP-RP and LBP-MTF features on training set

Networks	Input features	Overall accuracy/%
RF	LBP-RP	95.65
	LBP-MTF	92.67
KNN	LBP-RP	98.33
	LBP-MTF	96.52

3.2.2 Training results of deep learning models

The transformed RP and MTF were selected as input variables for the classification model. In this study, five classification models, namely ResNet-101, DenseNet-121, SqueezeNet, ISqueezeNet and ViT were constructed to validate the effectiveness of RP and MTF in the discrimination of moldy pear core. During the training process, the uniform sampling algorithm was utilized to randomly divide sample set into the training set and test set. Table 3 illustrates the sample set division results of different categories of pears. The pears with varying moldy-core extents were classified into three types: healthy (S = 0%), slight moldy-core (0 < S < 10%), and moderate and severe moldy core ($S \geq 10\%$). The train set consisted of 288 samples including 112 healthy pears, 104 slight moldy-core pears and 72 moderate and severe moldy-core pears. The test set consisted of 72 samples including 28 healthy pears, 26 slight moldy-core pears and 18 moderate and severe moldy-core pears.

Table 3 Sample set division results of different categories of pears

Classes	Moldy-core extents	Total	Training set	Test set
Healthy	0	140	112	28
Slight	< 10%	130	104	26
Moderate and severe	≥10%	90	72	18

Hyperparameter optimization was a key step for the ViT, ResNet-101, DenseNet-121, SqueezeNet, and ISqueezeNet classification models. To accomplish this task, Bayesian optimization was employed to search the space of optimizer type, learning rate, batch size, and L2 regularization coefficient. The optimization results indicated that ISqueezeNet and DenseNet-121 achieved their best performance with the Adam optimizer (learning rate of 1×10^{-4}, batch size of 64, and L2 coefficients of 5×10^{-4} and 1×10^{-5} respectively), while SqueezeNet converged best with Adam using a smaller learning rate of 5×10^{-5}, a batch size of 32, and an L2 coefficient of 1×10^{-5}. In contrast, ResNet-101 achieved the highest accuracy with the SGD optimizer (learning rate of 1×10^{-3}, batch size of 128, and L2 coefficient of 1×10^{-4}), whereas ViT obtained optimal performance with the Adam optimizer (learning rate of 3×10^{-5}, batch size of 64, and L2 coefficient of 1×10^{-4}). Under these model-specific optimal configurations, the final training accuracies of the

five models with RP input reached 100%, 97.22%, 99.65%, 94.44%, and 99.65%, respectively. When MTF features were used as input, the corresponding training accuracies were 97.22%, 94.44%, 100%, 94.44%, and 99.65%. The training curves of the five models with both RP and MTF inputs are presented in Figure 5. Although both inputs yielded high accuracies, models trained with RP features achieved overall more stable and slightly superior performance. Therefore, RP features were selected as the input for subsequent analysis.

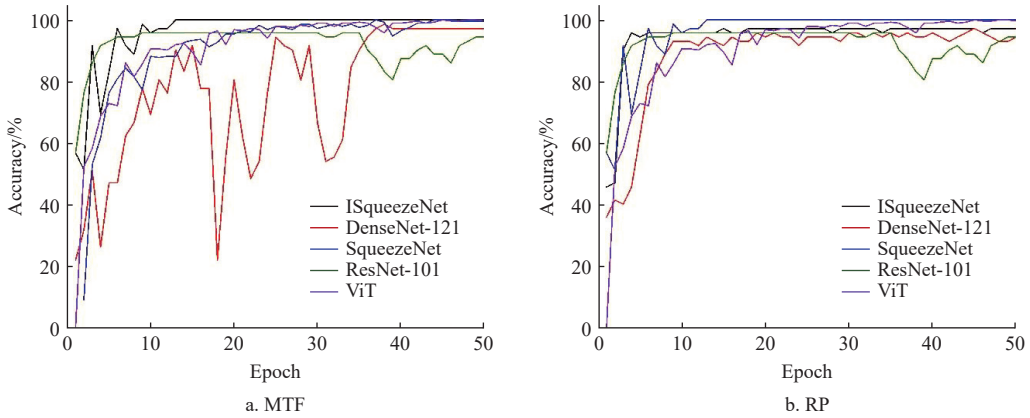


Figure 5 Training performance of five models under RP and MTF inputs

3.3 Visualization analysis of deep features

To verify the effects of deep feature extraction of the convolutional layer, the t-SNE algorithm was utilized to visualize the deep features extracted from the output of the second-last layer of the five networks. Figure 6 shows the cluster visualization results using the t-SNE. It can be found that there are few overlapping regions among the three clusters and the sample points of three-class pears were closely clustered in each model. Obviously, the low-level original RP images in the input layer were converted into high-level features layer by layer, which can improve the robustness of discrimination model. The feature visualization results indicated that the trained models have excellent feature learning capacity and can capture effective moldy-core features from the original RP images adaptively.

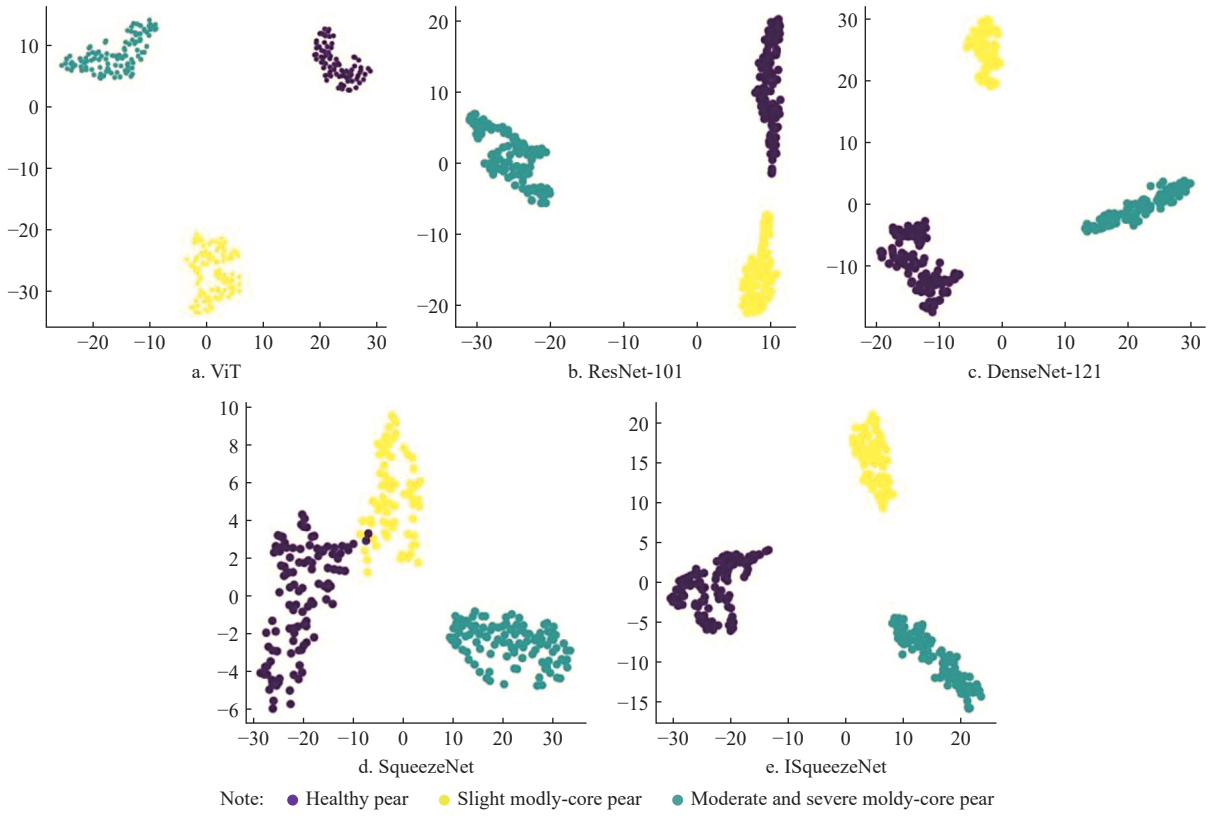


Figure 6 Feature visualization through t-SNE for the convolution layer in four models

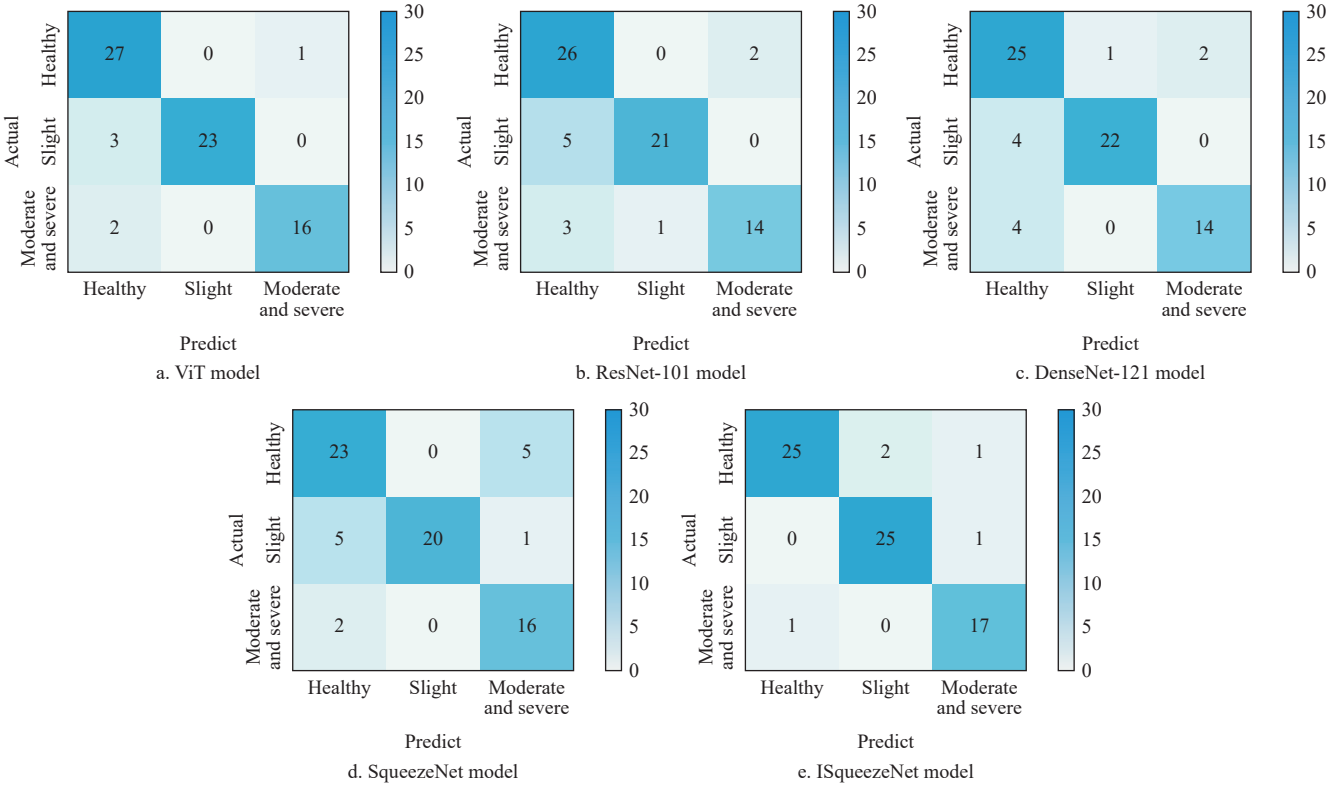
3.4 Verification and evaluation of classification performance of models

To further verify the classification performance of the constructed model, a test data set consisting of 72 samples was used as input, including 28 healthy pears, 26 slight moldy-core pears, and 18 moderate and severe moldy-core pears. Figure 7 illustrates the confusion matrices of the RF and KNN models. For the RF model, 60 out of 72 samples were correctly classified, corresponding to an overall discrimination accuracy of 83.33%. Specifically, the recognition accuracies were 85.71% for healthy

pears, 76.92% for slight moldy-core pears, and 88.89% for moderate and severe moldy-core pears.For the KNN model, 62 out of 72 samples were correctly identified, yielding an overall accuracy of 86.11%. The class-wise accuracies were 82.14% for healthy pears, 92.31% for slight moldy-core pears, and 83.33% for moderate and severe moldy-core pears.These results suggest that both RF and KNN can effectively discriminate moldy pear cores to a certain extent, with KNN showing better performance in identifying slight moldy-core pears. Figure 8 depicts the confusion matrix of four classification models for three-class pear samples on the test set. Both ResNet-101 model and DenseNet-121 model can correctly identify 61 out of 72 pears, with an overall discrimination accuracy of 84.72%. The above results were even worse when the SqueezeNet model was applied, with the overall discrimination accuracy being as low as 80%, meaning that < 60 out of 72 fruits were correctly classified. High misjudgment rate could be attributed to the insignificant difference of vibro-acoustic signal curves between the healthy and slight moldy-core fruit with similar. Compared with the above-mentioned models, the ISqueezeNet model exhibited a significant improvement in classification results. As seen in Figure 9, the overall discrimination accuracy for ISqueezeNet model reached 93.05%. Specifically, the discrimination accuracy was 89.28%, 96.15%, and 94.44% for healthy pears, slight moldy-core pears, and moderate and severe moldy-core pears, respectively. The comparison of the discrimination effects of above models indicated that the ISqueezeNet model exhibited the superior performance for discriminating moldy pear cores.

a. RF model b. KNN model

Figure 7 Classification results of three types of pears in test set



a. ViT model b. ResNet-101 model c. DenseNet-121 model

d. SqueezeNet model e. ISqueezeNet model

Figure 8 Classification results of three types of pears in test set

Table 4 presented the performance evaluation results of the ISqueezeNet model. The value of Precision, Recall, and F_1 exceeded 92.7%, indicating that the ISqueezeNet model had higher generalization ability for identifying moldy pear core. F_1 score is the weighted average of precision and recall. A higher F_1 score indicated that the model performs well in the classification task and

has a higher discrimination ability for the minority class of early moldy-core fruits. The F_1 score of the ISqueezeNet model for early moldy pears core reached 92.59% for healthy pears, 94.33% for slight moldy-core pears, and 91.89% for moderate and severe pears. The results show that the ISqueezeNet model is quite robust for identifying moldy pear core using vibro-acoustic testing method. It can be concluded that the ISqueezeNet model developed in this study can be employed to identify moldy pear cores with high classification accuracy and generalization performance through a homemade VA detection system.

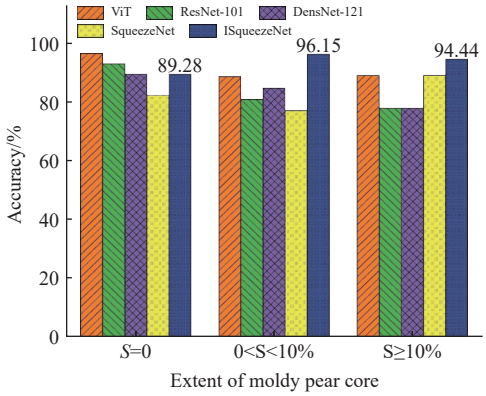


Figure 9 The model discrimination results based on confusion matrix analysis

At present, some scholars have studied the identification of internal diseases of pears, and the detection accuracy can reach 95% or above[32,33]. However, these studies only identify the presence or absence of internal diseases of pears, and do not consider the early diseases of fruits. Liu et al.[6] converted one-dimensional acoustic vibration multi-domain images (AVMDI), and used visual transformer (ViT) for feature extraction to construct

Table 4 Performance evaluation of the ISqueezeNet model

Moldy core degree	Performance evaluation		
	Precision/%	Recall/%	F_1/%
Healthy	96.15	89.28	92.59
Slight	92.59	96.15	94.33
Moderate and severe	89.47	94.44	91.89
Total	92.74	93.29	92.94

classification model of pear moldy-core. The recognition accuracy of AVMDI-ViT model was 94.74% for slight moldy-core apples. In comparison, the recognition accuracy of early moldy pear core can reach 96.15% by the ISqueezeNet model in this work, which was an improvement of 1.41%. Moreover, compared with other models, ISqueezeNet effectively reduces the number of model parameters and computational complexity through depth-wise separable convolution, significantly improving the inference speed and efficiency of model deployment while ensuring high accuracy.

3.5 Misclassification analysis based on Grad-CAM

To investigate the causes of misclassification, we employed Grad-CAM to visualize the regions attended by the model during decision-making (Figure 10). The heatmaps reveal that healthy and slight samples exhibit highly similar high-response regions in both spatial location and intensity, which makes them prone to confusion. In addition, slight samples often show an expansion of attention to larger areas and oversensitivity to local peaks, leading to their misclassification as moderate and severe. Nevertheless, such cases remain limited, and the overall attention patterns align well with the true categories, underscoring the robustness and reliability of the proposed method.

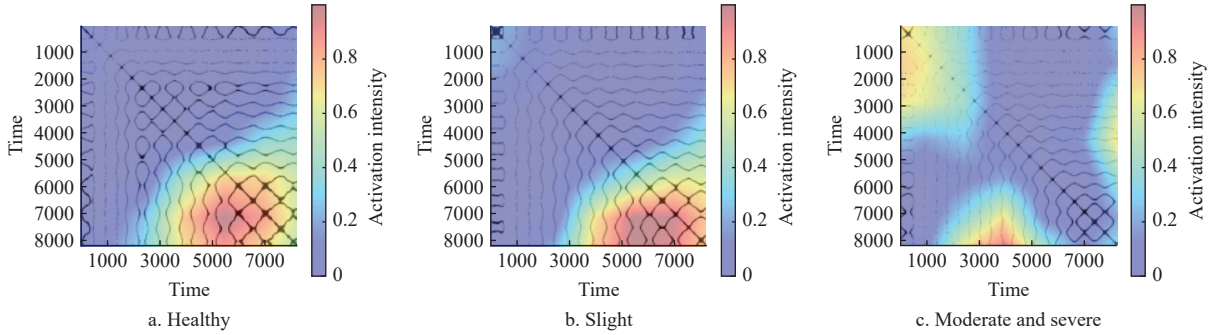


Figure 10 Grad-CAM visualization of the model's attention area

3.6 Comparative analysis of the ISqueezeNet network model and other lightweight network models

In this study, we conducted a comparative analysis between the ISqueezeNet model and other popular lightweight network models, including SqueezeNet, MobileNetV1, and ShuffleNetV2. This analysis focuses on several key performance metrics: classification accuracy, model size, FLOPs, and inference speed. The goal is to highlight the advantages of ISqueezeNet, especially in terms of maintaining high performance while reducing computational complexity.

As shown in Table 5, ISqueezeNet shows the strongest classification performance, with an overall classification accuracy of 93.05%. In addition to the classification accuracy, as a lightweight network, the ISqueezeNet model also demonstrates good real-time deployment capabilities. The model achieves an inference time of 0.009 s per sample, with a model size of 0.18 MB and 0.06 GFLOPs FLOPs, which outperforms other lightweight models. These indicators confirm that the ISqueezeNet model can be deployed on embedded devices such as sorting lines, providing efficient, high-accuracy detection without compromising processing speed. The ability to balance high classification accuracy with low resource consumption makes the ISqueezeNet model a viable solution for real-time moldy pear core detection in production settings.

Table 5 Classification results comparison of the network models of the ISqueezeNet, SqueezeNet, MobileNetV1 and ShuffleNetV2

Networks	DSC	OA/%	Size/MB	FLOP/GFLOPs	Speed/s
SqueezeNet	-	81.94	4.6	0.39	0.040
MobileNetV1	√	83.58	0.34	0.07	0.089
ShuffleNetV2	-	90.62	7.79	0.29	0.050
ISqueezeNet	√	93.05	0.18	0.06	0.009

4 Conclusions

This work develops a nondestructive classification framework for moldy pear cores by converting VA signals into recurrence-plot andmarkov transition field and training both machine-learning and deep-learning models. Machine-learning baselines (RF and KNN) delivered reasonable performance, but deep architectures exhibited stronger generalization. Among them, the proposed ISqueezeNet achieved the best overall accuracy (93.05%) with balanced class-wise performance. A systematic Bayesian optimization of optimizer, learning rate, batch size, and L2 regularization contributed to stable convergence and superior accuracy across models. In head-to-head comparisons with ResNet-101, DenseNet-121, SqueezeNet, ViT, and lightweight CNNs (MobileNetV1, ShuffleNetV2), ISqueezeNet provided a favorable trade-off between accuracy, parameter count, and inference speed. Grad-CAM analyses showed consistent attention over disease-related regions, enhancing interpretability and user trust. Taken together, these findings substantiate the suitability of the proposed method for accurate and efficient online detection of moldy cores in practical sorting and quality-inspection scenarios. In future work, we will incorporate naturally infected samples collected from different regions and under varied storage conditions, enabling more realistic evaluation of the proposed method and contributing to the development of reliable non-destructive detection techniques. Ultimately, this will support postharvest disease management and promote food safety and quality assurance in the fruit industry.

Acknowledgements

The authors are grateful for the support of Key R&D Projects in Shandong Province (Grant No. 2022TZXD007) and PhD Start-up Fund of University of Jinan (Grant No. XBS2494).

[References]

[1] Pan T, Chyngyz E, Sun D W, Paliwal J, Pu H. Pathogenetic process monitoring and early detection of pear black spot disease caused by *Alternaria alternata* using hyperspectral imaging. Postharvest Biology and Technology, 2019; 154: 96–104.

[2] Zhang Q, Huang W, Wang Q, Wu J, Li J. Detection of pears with moldy core using online full-transmittance spectroscopy combined with supervised classifier comparison and variable optimization. Computers and Electronics in Agriculture, 2022; 200: 107231.

[3] Zhang Z, Liu H, Chen D, Zhang J, Li H, Shen M. SMOTE-based method for balanced spectral nondestructive detection of moldy apple core. Food Control, 2022; 141: 109100.

[4] Liu H L, Wei Z Y, Lu M, Gao P, Li J K, Zhao J, Hu J. A Vis/NIR device for detecting moldy apple cores using spectral shape features. Computers and Electronics in Agriculture, 2024; 220: 108898.

[5] Li Y P, Zhang X N, Nie J Y, Bacha S A S, Yan Z, Gao G W. Occurrence and co-occurrence of mycotoxins in apple and apple products from China. Food Control, 2020; 118: 107354.

[6] Liu Z, Chen N, Le D X Lai Q R, Li B, Wu J, et al. Acoustic vibration multi-domain images vision transformer (AVMDI-ViT) to the detection of moldy apple core: Using a novel device based on micro-LDV and resonance speaker. Postharvest Biology and Technology, 2024; 211: 112838.

[7] Tian S, Wang J, Xu H. Firmness measurement of kiwifruit using a self-designed device based on acoustic vibration technology. Postharvest Biology and Technology, 2022; 187: 111851.

[8] Tempelaere A, Phan H M, Van De, Looverbosch T, Verboven P, Nicolai B. Non-destructive internal disorder segmentation in pear fruit by X-ray radiography and AI. Computers and Electronics in Agriculture, 2023; 212: 108142. doi: 10.1016/j.compag. 2023.108142

[9] Van D, Looverbosch T, Raeymaekers E, Verboven P, Sijbers J, Nicolai B. Non-destructive internal disorder detection of Conference pears by semantic segmentation of X-ray CT scans using deep learning. Expert Systems with Applications, 2021; 176: 114925.

[10] Velásquez C, Prieto F, Palou L, Cubero S, Blasco J, Aleixos N. New model for the automatic detection of anthracnose in mango fruits based on Vis/NIR hyperspectral imaging and discriminant analysis. Journal of Food Measurement and Characterization, 2024; 18(1): 560–570.

[11] Mogollón M, R, Contreras C, De Freitas, S T, Zoffoli J P. NIR spectral models for early detection of bitter pit in asymptomatic 'Fuji' apples. Scientia Horticulturae, 2021; 280: 109945.

[12] Ghooshkhaneh N G, Golzarian M R, Mollazade K. VIS-NIR spectroscopy for detection of citrus core rot caused by Alternaria alternata. Food Control, 2023; 144: 109320.

[13] Tang Y, Yang J P, Zhuang J J, Hou C J, Miao A M, Ren J C. Early detection of citrus anthracnose caused by *Colletotrichum gloeosporioides* using hyperspectral imaging. Computers and Electronics in Agriculture, 2023; 214: 108348.

[14] Chun S W, Song D J, Lee K H, Kim M J, Kim M S, Kim K S. Deep learning algorithm development for early detection of Botrytis cinerea infected strawberry fruit using hyperspectral fluorescence imaging. Postharvest Biology and Technology, 2024; 214: 112918.

[15] Qin Y, Jia W, Sun X, Lv H. Development of electronic nose for detection of micro-mechanical damages in strawberries. Frontiers in Nutrition, 2023; 10: 1222988.

[16] Zhang J C, Zhang P, Xue Y L, Jia X Y, Li J K. Characterization of characteristic odor of rotten core apples based on electronic nose and establishment of non-destructive detection model. Food and Fermentation Industries, 2022; 48(2): 267–273.

[17] Zhang H, Zha Z H, Kulasiri D, Wu J. Detection of early core browning in pears based on statistical features in vibro-acoustic signals. Food and Bioprocess Technology, 2021; 14: 887–897.

[18] Han Q L, Long B X, Yan X J, Wang W, Liu F, Chen X L. Exploration of using acoustic vibration technology to non-destructively detect moldy kernels of in-shell hickory nuts (*Carya cathayensis Sarg.*) Computers and Electronics in Agriculture, 2023; 212: 108137.

[19] Mohammed A, Kora R. A comprehensive review on ensemble deep learning: Opportunities and challenges. Journal of King Saud University-Computer and Information Sciences, 2023; 35(2): 757–774.

[20] Al-Fraihat D, Sharrab Y, Alzyoud F, Qahmash A, Tarawneh M, Maaita A. Speech recognition utilizing deep learning: A systematic review of the latest developments. Human-Centric Computing and Information Sciences, 2024; 14. doi: 10.22967/HCIS.2024.14.015

[21] Jiang H, Diao Z, Shi T, Zhou Y, Wang F, Hu W. A review of deep learning-based multiple-lesion recognition from medical images: classification, detection and segmentation. Computers in Biology and Medicine, 2023; 157: 106726.

[22] Zhu Z, Lei Y, Qi G, Chai Y, Mazur N, An Y, Huang X. A review of the application of deep learning in intelligent fault diagnosis of rotating machinery. Measurement, 2023; 206: 112346.

[23] Cai Z, Sun C, Zhang H, Zhang Y, Li J. Developing universal classification models for the detection of early decayed citrus by structured-illumination reflectance imaging coupling with deep learning methods. Postharvest Biology and Technology, 2024; 210: 112788. doi: 10.1016/j.postharvbio.2024.112788

[24] Ünal Z, Kızıldeniz T, Özden M, Aktaş H, Karagöz Ö. Detection of bruises on red apples using deep learning models. Scientia Horticulturae, 2024; 329: 113021.

[25] Shin J, Chang Y K, Heung B, Nguyen-Quang T, Price G W, Al-Mallahi A. A deep learning approach for RGB image-based powdery mildew disease detection on strawberry leaves. Computers and Electronics in Agriculture, 2021; 183: 106042.

[26] Seo D, Nam H. Deep rp-cnn for burst signal detection in cognitive radios. IEEE Access, 2020; 8: 167164–167171.

[27] Zhou Y, Wang Z, Zuo X, Zhao H. Identification of wear mechanisms of main bearings of marine diesel engine using recurrence plot based on CNN model. Wear, 2023; 520: 204656.

[28] Petrauskiene V, Pal M, Cao M, Wang J, Ragulskis M. Color recurrence plots for bearing fault diagnosis. Sensors, 2022; 22(22): 8870.

[29] Li H, Zhao K, Zha Z H, Zhang H, Wu J. Early nondestructive detection of blackheart disease in pear based on empirical mode decomposition of acoustic vibration signals. Food Science, 2023; 44(20): 357–371.

[30] He K M, Zhang X Y, Ren S Q, Sun J. Deep residual learning for image recognition. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2016; pp.770–778. doi: 10.48550/arXiv.1512.03385

[31] Huang G, Liu Z, Van Der, Maaten L, Weinberger K Q. Densely connected convolutional networks. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2017; pp.4700–4708. doi: 10.

1109/CVPR.2017.243

[32] Wang N, Yu S K, Qi Z P, Ding X Y, Wu X, Hu N. Pears classification by identifying internal defects based on X-ray images and neural networks. Advances in Manufacturing, 2025; 13: 552–561.

[33] Seo H-J, Song J-H. Detection of internal browning disorder in 'Greensis' pears using a portable non-destructive instrument. Horticulturae, 2023; 9(8): 944.