Two-stage machine vision and near-infrared spectroscopy for grading leafhopper damage in fresh tea leaves

Linsen Li
Jiangnan University
Feihu Song
214191024@qq.com
Jiangnan University
Nailiang Zhang
Jiangnan University
Chunfang Song
Jiangnan University
Zhenfeng Li
Jiangnan University
Guangyuan Jin
Jiangnan University
Jing Li
Jiangnan University

Research Article

Keywords:

Posted Date: September 14th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-10691752/v1

Additional Declarations: No competing interests reported.

Two-stage machine vision and near-infrared spectroscopy for grading leafhopper damage in fresh tea leaves

Linsen Li[a], Feihu Song[a*], Nailiang Zhang[a], Chunfang Song[a], Zhenfeng Li[a], Guangyuan Jin[a], Jing Li[a]

[a] *Jiangsu Key Laboratory of Advanced Food Manufacturing Equipment & Technology, School of Intelligent Manufacturing, Jiangnan University, Wuxi, 214122, Jiangsu, China*

*Corresponding author.
E-mail address: 214191024@qq.com
Phone: +86-18362360100

Abstract

Leafhopper feeding-damage severity is an important factor in the raw-material grading and final quality of Zijin Chan tea, but conventional assessment relies heavily on subjective visual judgment. This study developed a two-stage workflow combining machine-vision screening with offline near-infrared (NIR) spectroscopic reassessment. First, an improved small You Only Look Once version 8 (YOLOv8s) model localized tea-leaf targets and classified slight, moderate, and severe feeding damage. Physical samples corresponding to machine-vision no-result targets were then reassessed using an NIR model integrating standard normal variate preprocessing, Pearson correlation-based spectral-region selection, the successive projections algorithm, and a support vector machine. On 200 test images containing 4,000 annotated targets, the improved YOLOv8s model achieved precision of 88.6%, recall of 87.7%, and mean average precision at an intersection-over-union threshold of 0.5 of 89.3%, representing improvements of 12.2, 9.3, and 8.7 percentage points over the baseline, respectively. Developed from 240 independent physical samples, the NIR model achieved macro-averaged recall of 94.03% on the cross-validated training set and 95.12% on the spectral model-selection set. When the fixed model was applied to physical samples corresponding to 492 machine-vision no-result targets, 473 targets were classified correctly, yielding a reassessment accuracy of 96.14%. Across all targets, standalone machine vision achieved an end-to-end accuracy of 77.7%, whereas the two-stage workflow achieved 89.5%. These results support the feasibility of combining visual and spectral information for laboratory-based grading. Validation using in situ spectra and samples from multiple batches and seasons remains necessary before online deployment.

Keywords: Zijin Chan tea; leafhopper feeding-damage severity; machine vision; YOLOv8s; near-infrared spectroscopy; quality grading

1 Introduction

Zijin Chan tea is a distinctive tea product manufactured from fresh tea leaves subjected to feeding by the tea green leafhopper (Liao et al. 2019; Scott et al. 2020). Its characteristic honey-like aroma and flavor profile are closely associated with the severity of leafhopper feeding damage in the fresh leaves (Scott et al. 2020). The degree of feeding damage not only affects the transformation of quality-related constituents during subsequent processing but also, to some extent, determines the quality and market value of the finished tea (Liao et al. 2019; Shao et al. 2021). Therefore, rapid and objective assessment of feeding-damage severity at the initial stage of Zijin Chan tea production is essential for raw-material grading, consistent product quality, and standardized production (Wang et al. 2024).

At present, the severity of leafhopper feeding damage in fresh Zijin Chan tea leaves is assessed primarily on the basis of human experience. Tea growers typically rely on visible characteristics such as leaf color, feeding lesions, marginal scorching, and the degree of leaf curling (Gill et al. 2013; Wang et al. 2022). However, such assessments are readily influenced by operator experience, illumination conditions, and fatigue, resulting in substantial subjectivity, limited consistency, and low classification efficiency (Wang et al. 2023). Therefore, developing an objective and rapid method capable of evaluating feeding-damage severity under complex fresh-leaf conditions is essential for improving the consistency of raw-material grading and the stability of processing quality.

Machine vision offers several advantages, including rapid, nondestructive analysis and a high degree of automation, making it well suited for capturing external leaf characteristics such as color, feeding lesions, marginal scorching, and curling (Ren et al. 2021; Song et al. 2024; Guo et al. 2024). Owing to their high detection speed and ability to perform object localization and classification simultaneously, models in the You

Only Look Once (YOLO) family have been widely applied to tasks such as tea-bud detection, disease and pest identification, and tea-quality grading (Lee et al. 2020; Xue et al. 2023; Lin et al. 2023). However, fresh Zijin Chan tea leaves often exhibit irregular morphologies, mutual occlusion, and substantial visual similarity between adjacent feeding-damage classes, which may lead to missed detections or class confusion when a standalone machine-vision model is used (Lee et al. 2020; Lin et al. 2023). Near-infrared spectroscopy provides information on the internal chemical composition of samples and the associated absorption responses, thereby complementing the external morphological features captured by machine vision (Huang et al. 2020; Luo et al. 2022; Tian et al. 2023). Spectral variations arising from changes in quality-related constituents, including tea polyphenols, catechins, caffeine, free amino acids, and soluble sugars, can provide supplementary evidence for classifying samples with insufficient visual information (Huang et al. 2020; Luo et al. 2022). However, applying spectroscopic analysis to every fresh-leaf sample may be impractical because of constraints associated with sample preparation, analytical throughput, and implementation cost (Gharibzahedi et al. 2022). Therefore, although near-infrared spectroscopy can provide rich information on internal quality attributes, machine vision was prioritized in this study for the rapid initial screening of large sample batches. Spectroscopic reassessment was performed only for the physical samples corresponding to machine-vision no-result targets, thereby balancing classification capability, processing efficiency, and analytical cost (Jin et al. 2020).

Prior studies on tea-quality assessment have used machine vision, spectroscopy, or multisensor fusion, but have generally focused on direct quality grading or process monitoring rather than staged reassessment of machine-vision no-result targets (Jin et al. 2020). Research specifically addressing staged classification of leafhopper feeding-damage severity in fresh Zijin Chan tea leaves therefore remains limited (Gharibzahedi et al. 2022; Wang et al. 2023). Accordingly, this study does not seek to replace machine-vision detection with near-infrared spectroscopy. Instead, spectroscopy is used as a complementary reassessment tool for the physical samples corresponding to machine-vision no-result targets, thereby combining efficient initial screening with further discrimination of challenging samples.

Building on this framework, the present study developed a two-stage workflow for fresh Zijin Chan tea leaves, comprising initial screening by machine vision followed by reassessment using near-infrared spectroscopy. First, an improved YOLOv8s model was employed to localize and initially classify tea-leaf targets with slight, moderate, and severe leafhopper feeding damage. The physical samples corresponding to machine-vision no-result targets were subsequently reassessed spectroscopically through correlation analysis of quality-related constituents, characteristic-wavelength selection, and classification modeling. Finally, the overall performance of the proposed workflow was evaluated based on the target-to-sample flow through the two stages and the number of correct classifications obtained at each stage. This study aimed to evaluate the feasibility of integrating complementary visual appearance and internal spectral information for grading raw materials used in Zijin Chan tea production, and to provide an experimental basis for the subsequent development of in situ fresh-leaf detection and online inspection systems.

2 Materials and Methods

2.1 Experimental Samples and Overall Experimental Design

The experimental samples consisted of fresh Zijin Chan tea shoots with one bud and two leaves, collected from a tea plantation in Zijin County, Guangdong Province, China. Based on the visible changes induced by tea green leafhopper feeding, the samples were classified into three feeding-damage categories: slight, moderate, and severe, as shown in Figure 1. The grading criteria primarily included leaf color, feeding lesions, the extent of marginal scorching, leaf-curling characteristics, and tea-bud color. Reference grades were assigned by local tea growers experienced in grading Zijin Chan tea based on these visible characteristics and were used as the expert-assigned reference labels for evaluating both the machine-vision and near-infrared spectroscopy models. Samples with slight feeding damage were predominantly green, with relatively firm leaves and bright-green tea buds. Moderately damaged samples retained a generally green leaf color but exhibited some reddish-black lesions and marginal scorching, while the tea buds appeared yellow. Severely damaged samples showed larger areas of desiccation, withered yellow tea buds, and more pronounced leaf curling and marginal scorching.

This study developed a two-stage classification workflow comprising initial screening by machine vision and subsequent reassessment using near-infrared spectroscopy. The machine-vision dataset comprised 2,000 images of fresh Zijin Chan tea leaves and was divided into training, validation, and test sets at a ratio of 8:1:1, containing 1,600, 200, and 200 images, respectively. The training set was used for model training, the validation set for hyperparameter tuning, and the test set for evaluating model performance and the overall two-stage workflow. The 200 test images contained a total of 4,000 annotated tea-leaf targets; therefore, each individual annotated target was treated as the basic statistical unit for workflow evaluation.

In the first stage, the improved YOLOv8s model was used to localize tea-leaf targets and classify their feeding-damage severity. In this study, a machine-vision no-result target was defined as an annotated target for which the model produced no valid detection-and-classification result under the inference settings used in this study. In addition to precision, recall, and mean average precision at an intersection-over-union threshold of 0.5 (mAP@0.5) for evaluating object-detection performance, the machine-vision initial-screening coverage and the classification accuracy for successfully detected targets were calculated based on the actual flow of test targets through the workflow. Machine-vision initial-screening coverage was defined as the proportion of all test targets for which the machine-vision stage produced valid detection-and-classification results. The classification accuracy for successfully detected targets was defined as the proportion of successfully detected targets whose predicted feeding-damage classes agreed with the expert-assigned reference labels.

The near-infrared spectroscopy model for three-class classification was developed using a separate set of 240 independent physical samples comprising only the slight, moderate, and severe feeding-damage classes. Unfed control samples were used solely to analyze variations in quality-related constituents and spectral responses and were excluded from classification modeling. The 240 samples were partitioned using the Kennard–Stone (K–S) algorithm at a ratio of 3:1, yielding a training set of 180 samples and a spectral model-selection set of 60 samples. Model fitting, variable selection, and hyperparameter optimization were performed exclusively on the training set using ten-fold cross-validation. The spectral model-selection set was excluded from model fitting and hyperparameter optimization, but its results were used to compare the performance of the candidate models. The final model configuration was selected by jointly considering cross-validation performance on the training set, performance on the spectral model-selection set, and the degree of variable reduction. The finalized model was then fixed and applied to the physical samples corresponding to machine-vision no-result targets.

2.2 Image Data Acquisition and Development of the Improved YOLOv8s Model

After harvesting, the fresh Zijin Chan tea-leaf samples were spread evenly within a designated area, and an industrial camera was mounted vertically above the samples for image acquisition, thereby simulating an industrial flat-bed sorting scenario for fresh Zijin Chan tea leaves. Images were acquired in mid-October 2025, and the samples were collected from a tea plantation in Zijin County, Guangdong Province, China. A Daheng Imaging MERCURY2 MER2-160-227-U3C industrial camera was used, with an image resolution of 3,264 × 2,448 pixels. The vertical distance between the camera and the leaf-spreading area was 35 cm. A total of 2,000 original images of fresh Zijin Chan tea leaves were acquired, comprising samples with slight, moderate, and severe leafhopper feeding damage.

Tea-leaf targets in the images were annotated with rectangular bounding boxes using LabelImg (Das et al. 2026) and assigned to the Slight, Moderate, or Severe class. The annotations were saved in TXT format, with each bounding-box record containing the class index and the normalized center coordinates, width, and height of the target box. Following annotation, the 2,000 images were divided into training, validation, and test sets at a ratio of 8:1:1, comprising 1,600, 200, and 200 images, respectively. The dataset contained a total of 40,960 bounding-box annotations. To improve the model's robustness to variations in sample orientation, degree of overlap, and illumination conditions, online data augmentation was applied during training, including random affine transformations, horizontal flipping, scaling, and MixUp augmentation (Zhang et al. 2018).

Given the requirements of fresh Zijin Chan tea-leaf sorting for detection efficiency, compact model size, and ease of deployment, YOLOv8s was selected as the baseline object-detection model (Sapkota et al. 2025). To address the irregular morphology of Zijin Chan tea leaves, the localized nature of leafhopper feeding-damage features, the subtle visual differences among feeding-damage classes, and overlap and occlusion among tea leaves, YOLOv8s was improved at three architectural levels: the backbone, neck, and detection head. The squeeze-and-excitation (SE) attention mechanism (Hu et al. 2018) was embedded into the C2f modules of both the backbone and neck to construct an SE_C2f module, thereby enhancing the model's sensitivity to key visual features such as leafhopper feeding lesions, marginal scorching, and leaf browning. A linear deformable convolution (LDConv) module (Zhang et al. 2024) was introduced into the neck to improve feature extraction from irregularly shaped, curled, overlapping, and occluded leaves. In addition, an additional spatial pyramid pooling–fast (SPPF) module (He et al. 2015; Sapkota et al. 2025) was inserted before the Detect head to strengthen multiscale contextual feature fusion and improve the detection of tea-leaf targets with complex spatial distributions.

The object-detection models were evaluated using precision, recall, and mAP@0.5 to compare the performance of the baseline model, ablation variants, and alternative detection models. The two-stage workflow was evaluated separately according to the actual flow of test targets. Machine-vision initial-screening coverage was defined as the proportion of all annotated targets for which the machine-vision stage produced valid detection-and-classification results. Classification accuracy for successfully detected targets was defined as the proportion of successfully detected targets whose predicted feeding-damage classes agreed with the reference labels assigned by experienced tea growers. The overall accuracy of the two-stage workflow was calculated as the sum of the targets correctly classified at the machine-vision stage and those correctly reassessed at the near-infrared spectroscopy stage, divided by the total number of test targets. The image acquisition settings, dataset partitioning, and model training parameters are presented in Table 1.

2.3 Near-Infrared Spectral Acquisition and Determination of Quality-Related Constituents

To compensate for the limited information available to machine vision under conditions of occlusion, missing localized features, and similar visual appearances, near-infrared spectroscopy was incorporated into the workflow. This stage was applied to the physical samples corresponding to machine-vision no-result targets. Spectral responses associated with quality-related constituents, including tea polyphenols, catechins, caffeine, free amino acids, and soluble sugars, were used to provide discriminatory information complementary to visual appearance features.

Near-infrared spectra were acquired after the samples had undergone rapid freezing, cold-chain transportation, freeze-drying, grinding, and sieving. The spectral dataset comprised two subsets: samples used for classification modeling and unfed control samples. The classification-modeling dataset comprised 240 independent physical samples representing only the slight, moderate, and severe feeding-damage classes. These samples were used for preprocessing-method comparison, selection of quality-related spectral regions, feature-variable optimization, and development of the three-class classification model. Unfed control samples were used solely to interpret changes in quality-related constituents and spectral responses and were excluded from model training and performance evaluation. The modeling samples were collected separately and were independent of the physical samples corresponding to machine-vision no-result targets. Because spectra were acquired from processed tea samples, the second stage constituted offline reassessment under laboratory conditions. Its results were used to verify the complementary discriminatory value of internal quality information and should not be interpreted as direct evidence of online fresh-leaf detection performance.

Sample spectra were acquired using a portable near-infrared spectroscopy system. The original wavelength range was 858–2,590 nm; after excluding the noise-prone regions at both ends, the 1,000–2,400 nm range was retained for modeling. Each spectrum was obtained by averaging 32 consecutive scans. Three spectra were collected for each sample and subsequently averaged, yielding one representative spectrum per physical sample. The 240 samples were partitioned using the Kennard–Stone (K–S) algorithm at a ratio of 3:1, resulting in a training set of 180 samples and a spectral model-selection set of 60 samples (Galvao et al. 2005). Model fitting, variable selection, and hyperparameter optimization were performed exclusively on the training set using ten-fold cross-validation. The spectral model-selection set was excluded from model fitting and hyperparameter optimization, but its results were used to compare the performance of the candidate models.

Model fitting, variable selection, and hyperparameter optimization for the different spectral preprocessing methods, selection of quality-related spectral regions based on the Pearson product–moment correlation coefficient, feature-variable optimization based on the successive projections algorithm, and classifier parameter combinations were performed exclusively on the training set. The 60 spectral model-selection samples were excluded from all of these procedures, and their results were used to compare the performance of the candidate models. Macro-averaged recall was reported for both the training set and the spectral model-selection set.

The optimal model combination was selected by jointly considering the ten-fold cross-validation results on the training set, performance on the spectral model-selection set, and the degree of variable reduction. Once the model combination had been determined, the spectral preprocessing pipeline, selected feature variables, and classifier parameters were fixed and subsequently applied to the practical reassessment of the physical samples corresponding to machine-vision no-result targets. These external reassessment samples were excluded from spectral-model fitting, variable selection, hyperparameter optimization, and candidate-model comparison.

To further characterize these differences and provide a chemical basis for selecting quality-related spectral regions, the contents of tea polyphenols, catechins, caffeine, free amino acids, and soluble sugars were determined using the methods described in (Yemm and Willis 1954; Musci and Yao 2017; Jabeen et al. 2019; Kalidass et al. 2021; Vuletic et al. 2021). These constituents reflect changes in the internal chemical composition of Zijin Chan tea following leafhopper feeding (Liao et al. 2019; Scott et al. 2020) and were subsequently correlated with the near-infrared spectral data to identify quality-related spectral regions closely associated with the classification of feeding-damage severity (Wang et al. 2018).

2.4 Spectral Data Processing, Feature Selection, and Evaluation of Classification Models

Based on the spectral data and quality-related constituent measurements obtained in Section 2.3, spectral preprocessing, selection of quality-related spectral regions, optimization of characteristic wavelengths, and three-class classification modeling were conducted sequentially. Modeling was performed over the wavelength range of 1,000–2,400 nm, with slight, moderate, and severe feeding damage used as the class labels. The spectral model for second-stage reassessment was determined by comparing different combinations of preprocessing methods, variable-selection strategies, and classifiers. Unfed control samples were used solely for result interpretation and were not included as a classification class.

Spectral preprocessing was performed to minimize the effects of sample-particle scattering, baseline drift, and random noise on model performance. The evaluated methods included Savitzky–Golay (S–G) smoothing (Sheng et al. 2019; Tian et al. 2023), standard normal variate (SNV) transformation (Barnes et al. 1989), multiplicative scatter correction (MSC) (Geladi et al. 1985), and first- and second-derivative transformations. The most appropriate preprocessing method was selected according to the classification performance of the subsequent models. For feature-variable selection, the Pearson product–moment correlation coefficient (PPMCC; r) was first used to assess the relationships between the near-infrared spectral variables and quality-related constituents, including tea polyphenols, catechins, caffeine, free amino acids, and soluble sugars. An r value closer to 1 or −1 indicates a stronger positive or negative correlation, respectively. Wavelength variables with an absolute correlation coefficient greater than 0.6 were considered strongly correlated with the corresponding quality-related constituents, and the associated wavelength regions were retained as candidate quality-related spectral regions. Within these candidate regions, principal component analysis (PCA) (Bro and Smilde 2014), the successive projections algorithm (SPA) (Araújo et al. 2001), and Random Frog (Li et al. 2012) were subsequently applied for further feature-variable selection to reduce spectral dimensionality, variable redundancy, and multicollinearity.

For the near-infrared three-class classification models, the recall values for the Slight, Moderate, and Severe classes were calculated from the normalized confusion matrix. The arithmetic mean of the recall values across the three classes was used to evaluate the model's overall ability to identify samples with different levels of leafhopper feeding damage. This metric was defined as macro-averaged recall and calculated as follows:

$$R_{macro} = \frac{R_{Slight} + R_{Moderate} + R_{Severe}}{3}$$

where R_{Slight}, $R_{Moderate}$, and R_{Severe} denote the recall values for the Slight, Moderate, and Severe classes, respectively. The values of 94.03% and 95.12% reported below both represent macro-averaged recall calculated according to this evaluation criterion.

For each candidate model, spectral preprocessing, PPMCC-based spectral-region selection, feature-variable optimization, model fitting, and classifier hyperparameter optimization were performed exclusively using the 180-sample training set. Training-stage performance was evaluated by ten-fold cross-validation. The 60-sample spectral model-selection set was excluded from model fitting and hyperparameter optimization, and its macro-averaged recall was used to compare the performance of the candidate models. The use of linear discriminant analysis (LDA) and random forest (RF) was grounded in standard descriptions of these classifiers, while support vector machines (SVMs) have also been applied to tea-leaf recognition tasks (Breiman 2001; Tharwat et al. 2017; Hossain et al. 2018). The optimal model was selected by jointly considering ten-fold cross-validation performance on the training set, performance on the spectral model-selection set, and the degree of variable reduction. Once the model combination had been determined, the preprocessing pipeline, selected feature variables, and classifier parameters were fixed. The 492 machine-vision no-result targets were used solely for final reassessment evaluation. The methods used for spectral data processing, feature selection, and model evaluation are summarized in Table 3.

3 Results

3.1 Initial Screening Performance of the Improved YOLOv8s Model

The test results showed that the improved model achieved precision, recall, and mAP@0.5 values of 88.6%, 87.7%, and 89.3%, respectively, representing increases of 12.2, 9.3, and 8.7 percentage points over the baseline YOLOv8s model. These improvements indicate that the introduced modules enhanced target localization and class recognition across different levels of leafhopper feeding damage. The object-detection metrics were used for model performance comparison, whereas machine-vision initial-screening coverage and the overall accuracy of the two-stage workflow were calculated separately in Section 3.4 based on the actual flow of annotated targets through the workflow.

To further assess the contribution of each introduced module to the improvement in model performance, ablation experiments were conducted on the SE_C2f, LDConv, and SPPF modules. The results are presented in Table 4.

As shown in Table 4, the introduction of the SE_C2f, LDConv, and SPPF modules improved, to varying degrees, the model's ability to recognize tea-leaf targets exhibiting feeding lesions, marginal scorching, browning, curling, and occlusion. The model incorporating all three modules achieved the best performance, indicating that these modules exerted complementary effects in feature enhancement, spatial adaptation, and multiscale information representation.

To evaluate the overall performance of the improved YOLOv8s model relative to other object-detection models, it was further compared with SSD (Liu et al. 2016), Faster R-CNN (Ren et al. 2017), YOLOv5, YOLOv11, and YOLOv12 (Sapkota et al. 2025, 2026). The results are presented in Table 5.

To further investigate the basis for the improved discriminative performance, gradient-weighted class activation mapping (Grad-CAM) heatmap visualization (Selvaraju et al. 2017) was performed to compare the feature-attention regions of the baseline YOLOv8s model and the improved YOLOv8s model. The results are shown in Figure 2. As shown in Figure 2, the feature-attention regions of the baseline YOLOv8s model were relatively dispersed across the tea-leaf targets. By contrast, the improved YOLOv8s model focused more strongly on visual regions closely associated with feeding-damage severity, including tea buds, feeding lesions, marginal scorching, browning, and curled leaf edges. This finding indicates that the features learned by the improved model were highly consistent with the key appearance characteristics used by human evaluators to assess leafhopper feeding-damage severity in Zijin Chan tea, thereby enhancing the model's ability to discriminate among tea-leaf targets with different levels of feeding damage.

Figure 3 presents representative missed-detection and misclassification cases from the machine-vision stage. Tea-leaf targets with moderate feeding damage were prone to confusion with targets showing slight feeding damage when relatively large green regions remained exposed or when feeding lesions, marginal scorching, and browning were occluded. Mutual overlap or curling of the leaves could also result in missed detections or localization errors by obscuring target boundaries and key appearance features. These cases indicate that the principal challenge faced by the model was insufficient effective appearance information, rather than merely limited network capacity.

3.2 Near-Infrared Spectral Feature Analysis and Selection of Quality-Related Feature Wavelengths

To investigate the chemical basis of the near-infrared reassessment stage, the contents of tea polyphenols, catechins, caffeine, free amino acids, and soluble sugars were compared among unfed control samples and samples exhibiting slight, moderate, and severe leafhopper feeding damage. The results are shown in Figure 4. The unfed samples were used solely to facilitate interpretation of compositional trends before and after leafhopper feeding and were not treated as a class in the three-class classification model.

As shown in Figure 4, the major quality-related constituents exhibited distinct trends across the different feeding-damage classes. Samples with moderate feeding damage had relatively high tea polyphenol and catechin contents, samples with slight feeding damage had relatively high soluble sugar content, and samples with severe feeding damage had relatively high free amino acid content.

To further characterize the spectral responses of samples with different levels of leafhopper feeding damage, near-infrared spectra were acquired from the Zijin Chan tea samples. Because the two ends of the original spectral range contained considerable noise, only the spectral data within 1,000–2,400 nm were retained for subsequent analysis. To reduce the effects of sample-surface heterogeneity, particle scattering, baseline drift, and other interfering factors on model performance, multiple spectral preprocessing methods were compared, and standard normal variate (SNV) transformation was ultimately selected for subsequent feature selection and modeling. The original spectra and the spectra after SNV preprocessing are shown in Figure 5.

As shown in Figure 5, the Zijin Chan tea samples exhibited pronounced near-infrared absorption responses over the range of 1,000–2,400 nm.

The full-spectrum data contained a substantial number of redundant variables. PPMCC was first used to identify candidate wavelength regions strongly correlated with the contents of tea polyphenols, catechins, caffeine, free amino acids, and soluble sugars ($|r| > 0.6$). PCA, SPA, and Random Frog were then applied within these candidate regions to compare the effects of different dimensionality-reduction strategies on classification performance.

SPA primarily reduces redundancy and collinearity among variables (Araújo et al. 2001), whereas Random Frog evaluates wavelength importance based on variable-selection probabilities obtained through stochastic iterations (Li et al. 2012). Figure 6 shows the PPMCC-derived candidate wavelength regions and the distributions of feature wavelengths selected by SPA and Random Frog.

This two-step procedure—prescreening quality-related spectral regions followed by selection of representative variables—integrated chemical relevance with variable reduction. Its effectiveness was ultimately evaluated by jointly considering ten-fold cross-validation performance on the training set, classification performance on the spectral model-selection set, and the degree of variable reduction, rather than relying solely on a single correlation coefficient.

3.3 Development of the Near-Infrared Reassessment Model and Evaluation on the Spectral Model-Selection Set

Based on the feature-processing results described above, a near-infrared three-class classification model was developed using 240 samples representing slight, moderate, and severe feeding damage. The samples were partitioned by the Kennard–Stone (K–S) algorithm at a ratio of 3:1 into a training set of 180 samples and a spectral model-selection set of 60 samples. Model fitting and hyperparameter optimization for LDA, SVM, and RF, together with their respective variable-processing combinations, were performed exclusively on the training set, using ten-fold cross-validation for training-stage evaluation. The spectral model-selection set was excluded from model fitting and hyperparameter optimization, and its results were used to compare the performance of the candidate models. The optimal model was ultimately selected by jointly considering cross-validation performance on the training set, performance on the spectral model-selection set, and the degree of variable reduction. The 492 machine-vision no-result targets were excluded from model development and candidate-model comparison. The parameter settings of each model, together with the macro-averaged recall values obtained on the training set and the spectral model-selection set, are presented in Table 6.

As shown in Table 6, the different spectral variable-processing methods and classification models exhibited marked differences in performance on both the training set and the spectral model-selection set. Compared with the use of PPMCC, PCA, SPA, or Random Frog alone, first applying PPMCC to identify candidate wavelength regions associated with quality-related constituents and then using SPA to reduce variable redundancy and collinearity yielded more representative model inputs.

Among the candidate models evaluated, the combination of SNV preprocessing, PPMCC-based selection of quality-related spectral regions, SPA-based feature-variable optimization, and an SVM classifier achieved the best performance, with macro-averaged recall values of 94.03% on the training set and 95.12% on the spectral model-selection set.

To further examine the classification behavior of this model across the different feeding-damage classes, normalized confusion matrices were generated for the training and spectral model-selection sets, as shown in Figure 7.

As shown in Figure 7, the recall values for the Slight, Moderate, and Severe classes in the training set were 88.49%, 93.60%, and 100.00%, respectively, yielding a macro-averaged recall of 94.03%. For the spectral model-selection set, the corresponding recall values were 90.34%, 95.02%, and 100.00%, respectively, with a macro-averaged recall of 95.12%. The model achieved the best recognition performance for Severe samples, with recall reaching 100.00% on both the training set and the spectral model-selection set. Although a small degree of confusion remained between the Slight and Moderate classes, the overall classification performance was stable. Based on a comprehensive assessment of ten-fold cross-validation performance on the training set, macro-averaged recall on the spectral model-selection set, and the degree of variable reduction, the SNV–PPMCC–SPA–SVM model was selected as the near-infrared reassessment model for the second stage.

Once the optimal model combination had been determined, the spectral processing pipeline, selected feature variables, and classifier parameters were fixed and subsequently applied to the reassessment evaluation of the physical samples corresponding to machine-vision no-result targets.

3.4 Two-Stage Classification Performance of Machine-Vision Initial Screening and Near-Infrared Spectroscopic Reassessment

The 200 test images contained a total of 4,000 annotated targets. The machine-vision stage produced valid detection and feeding-damage classification results for 3,508 targets, corresponding to a machine-vision initial-screening coverage of 87.7%. Using the feeding-damage classes assigned by experienced tea growers as the reference labels, 3,108 of these targets were classified correctly, yielding a classification accuracy of 88.6% among successfully detected targets; the remaining 400 targets with valid detection-and-classification results were misclassified. When all 4,000 test targets were used as the common denominator, the end-to-end accuracy of the standalone machine-vision stage was 77.7%.

The remaining 492 tea-leaf targets were categorized as machine-vision no-result targets, accounting for 12.3% of all test targets, and were therefore routed to the near-infrared spectroscopic reassessment stage. The SNV–PPMCC–SPA–SVM model was developed using a separate set of 240 spectral samples, which were partitioned by the Kennard–Stone (K–S) algorithm at a ratio of 3:1 into a training set of 180 samples and a spectral model-selection set of 60 samples. Model fitting, feature-variable selection, and hyperparameter optimization were performed exclusively on the training set, using ten-fold cross-validation for training-stage evaluation. The spectral model-selection set was excluded from model fitting and hyperparameter optimization, and its results were used to compare the performance of the candidate models. The finalized model achieved macro-averaged recall values of 94.03% on the training set and 95.12% on the spectral model-selection set. Once the model combination, selected feature variables, and classifier parameters had been determined, they were fixed for subsequent application.

After the fixed model reassessed the physical samples corresponding to the 492 targets, the resulting target-level classifications were correct for 473 targets and incorrect for 19, yielding an actual reassessment accuracy of 96.14%. Across the two stages, a total of 3,581 targets were classified correctly, corresponding to an overall two-stage workflow accuracy of 89.53%, reported as 89.5% to one decimal place. Compared with the standalone machine-vision end-to-end accuracy of 77.7%, the two-stage workflow improved accuracy by 11.8 percentage points. This improvement was primarily attributable to the recovery and classification of machine-vision no-result targets.

4 Discussion

Taken together, these results demonstrate the feasibility of the proposed two-stage workflow, while also revealing the respective contributions, operational boundaries, and practical limitations of the machine-vision and spectroscopic stages.

The improvements obtained by the three modules may be attributed to their complementary functions. The SE_C2f module recalibrates channel-wise feature responses, thereby strengthening damage-related features while suppressing irrelevant background information. LDConv adjusts its sampling positions according to local geometric variations and may therefore improve feature extraction from irregular, curled, overlapping, and partially occluded tea leaves. The additional SPPF module placed before the detection head enlarges the effective receptive field and aggregates multiscale contextual information. These complementary functions provide a possible explanation for the performance gains observed in the ablation experiments.

As shown in Table 5, the improved YOLOv8s model exhibited superior precision, recall, and mAP@0.5. Although Faster R-CNN demonstrated strong object-detection capability, its substantially larger parameter count and computational cost would hinder practical deployment in subsequent online sorting applications. By contrast, the improved YOLOv8s model achieved better discriminative performance while maintaining relatively low parameter and computational requirements. These results indicate that the improved model provides a favorable balance between detection accuracy and computational efficiency, making it more suitable for the initial screening of leafhopper feeding-damage severity in Zijin Chan tea.

Direct numerical comparison with previous tea-related detection studies should be interpreted cautiously because of differences in tea materials, class definitions, imaging environments, and evaluation metrics. Previous studies have nevertheless demonstrated the effectiveness of attention mechanisms, multiscale feature fusion, and adaptive feature extraction for identifying visually similar tea diseases and quality classes. The present results extend these approaches to the classification of leafhopper feeding-damage severity under conditions involving irregular morphology, dense distribution, overlap, and partial occlusion.

The missed-detection and misclassification cases shown in Fig. 3 explain why near-infrared spectroscopy was incorporated as a complementary reassessment method only for the physical samples corresponding to machine-vision no-result targets. This design enabled the recovery of some machine-vision no-result targets but could not correct targets that had already been assigned an incorrect class by the machine-vision stage. Moreover, targets with low-confidence predictions, similar class probabilities, or severe occlusion were not yet automatically routed to spectroscopic reassessment. This operational boundary means that the subsequent two-stage results should be interpreted primarily as demonstrating complementary discrimination for machine-vision no-result targets, rather than as evidence that the workflow corrects all machine-vision misclassifications.

The compositional trends shown in Figure 4 are consistent with prior reports that leafhopper attack and other stressors can alter tea-leaf metabolite profiles (Liao et al. 2019; Scott et al. 2020; Shao et al. 2021). These findings suggest that different levels of leafhopper feeding damage may be accompanied by differences in internal chemical composition, thereby providing an interpretive basis for spectroscopic reassessment.

The spectral variations shown in Figure 5 may be associated with overtone and combination-band absorptions of hydrogen-containing functional groups, such as O–H, C–H, and N–H bonds, in constituents including moisture, soluble sugars, tea polyphenols, and amino acids (Huang et al. 2020; Tian et al. 2023). Because the spectral samples were freeze-dried, ground, and sieved before measurement, differences in particle characteristics, sample-packing uniformity, and surface scattering may have affected the original spectra, resulting in baseline drift and fluctuations in spectral intensity. SNV preprocessing reduced, to some extent, the effects of intersample scattering differences and baseline shifts, thereby improving the stability of spectral variations associated with chemical composition (Sheng et al. 2019; Tian et al. 2023). Therefore, the use of SNV as the preprocessing method for subsequent feature selection and classification modeling was considered appropriate.

More specifically, noticeable spectral variations were observed at approximately 1,140–1,160, 1,480–1,550, and 1,850–2,050 nm. The variations in the latter two regions may be associated with overtone and combination-band absorptions of O–H-containing constituents, including residual moisture, soluble sugars, and polyphenolic compounds. Because the samples had been freeze-dried before spectral acquisition, these assignments should be interpreted as possible contributions from residual moisture and other O–H-containing constituents rather than as direct measurements of water content.

The superior performance of the PPMCC–SPA–SVM combination may be explained by the complementary functions of the successive processing steps. PPMCC first restricted the analysis to wavelength regions associated with the measured quality-related constituents, thereby reducing the influence of chemically irrelevant variables. SPA then minimized redundancy and collinearity among the retained wavelength variables. In contrast, considerable overlap among the three feeding-damage classes remained in the PCA score space, indicating that the directions explaining the greatest overall spectral variance were not necessarily the most discriminative. The generally better performance of SVM and RF relative to LDA further suggests that the relationship between spectral features and feeding-damage severity may involve nonlinear class boundaries.

These results demonstrate that the near-infrared spectroscopic model developed using SNV preprocessing, PPMCC-based selection of correlated spectral regions, SPA-based feature-wavelength optimization, and an SVM classifier exhibited good discriminatory performance for leafhopper feeding-damage severity in the independently collected spectral modeling samples. The findings indicate that Zijin Chan tea samples with different levels of feeding damage are distinguishable based on spectral features associated with quality-related constituents, and that near-infrared spectral information can provide an effective basis for the subsequent reassessment of physical samples corresponding to machine-vision no-result targets. Related near-infrared spectroscopy studies have likewise demonstrated the discrimination or quantification of tea constituents and quality classes (Huang et al. 2020; Yang et al. 2021; Luo et al. 2022).

The value of 95.12% represents the macro-averaged recall obtained by the model on the 60 spectral model-selection samples, whereas 96.14% represents the accuracy achieved when the fixed model was applied to the practical reassessment of the 492 machine-vision no-result targets. These two values therefore correspond to different sample sources and evaluation stages. The overall two-stage workflow accuracy of 89.5% reflects the aggregate classification performance of the predefined procedure comprising machine-vision initial screening followed by reassessment of physical samples corresponding to machine-vision no-result targets. Because the current near-infrared stage did not reassess targets that had already been assigned a machine-vision class but were incorrectly classified, this result applies specifically to the reassessment trigger scope defined in this study. Future inclusion of successfully detected targets with low-confidence predictions or similar class probabilities in the reassessment stage may provide further opportunities to improve overall system performance.

Overall, this study verified the complementarity of visual appearance information and near-infrared spectral information at the level of target-to-sample routing. However, the second stage currently relies on processed samples that have been freeze-dried, ground, and sieved, and the triggering of spectroscopic reassessment has not yet been automated online. The present findings should therefore be regarded as evidence of the feasibility of a laboratory-based offline framework. Future work should focus on in situ spectral acquisition from fresh tea leaves, independent validation across batches and seasons, and the development of automated reassessment rules incorporating prediction confidence, differences in class probabilities, and the degree of occlusion, thereby enabling evaluation of the workflow's stability and efficiency under practical sorting conditions.

5 Conclusions

This study established a two-stage workflow for classifying leafhopper feeding-damage severity in fresh Zijin Chan tea leaves. First, the improved YOLOv8s model was used for machine-vision initial screening through target localization and feeding-damage classification. The physical samples corresponding to machine-vision no-result targets were then subjected to near-infrared spectroscopic reassessment. This design was intended to exploit the ability of machine vision to rapidly capture appearance information while using internal spectral information to provide complementary discriminatory evidence for physical samples corresponding to machine-vision no-result targets, rather than replacing machine-vision detection with spectroscopy.

The improved YOLOv8s model achieved precision, recall, and mAP@0.5 values of 88.6%, 87.7%, and 89.3%, respectively, representing increases of 12.2, 9.3, and 8.7 percentage points over the baseline model. The 240 spectral samples were partitioned using the Kennard–Stone (K–S) algorithm at a ratio of 3:1 into a training set of 180 samples and a spectral model-selection set of 60 samples. Model development was conducted on the training set using ten-fold cross-validation, whereas the spectral model-selection set was used to compare the performance of the candidate models. The SNV–PPMCC–SPA–SVM model achieved macro-averaged recall values of 94.03% on the training set and 95.12% on the spectral model-selection set. When the fixed model was applied to reassess the physical samples corresponding to 492 machine-vision no-result targets, 473 were classified correctly, corresponding to a reassessment accuracy of 96.14%.

Across all 4,000 test targets, the standalone machine-vision stage correctly classified 3,108 targets, yielding an end-to-end accuracy of 77.7%. After near-infrared spectroscopic reassessment was introduced, the number of correctly classified targets increased from 3,108 to 3,581, and the overall accuracy of the two-stage workflow reached 89.5%, representing an improvement of 11.8 percentage points over the standalone machine-vision end-to-end accuracy. These results demonstrate that near-infrared spectroscopic reassessment can effectively recover and classify a proportion of machine-vision no-result targets.

The near-infrared data used in this study were acquired from processed samples that had been freeze-dried, ground, and sieved. At the present stage, the results primarily demonstrate the feasibility of a laboratory-based offline two-stage framework and should not be interpreted as direct evidence of online fresh-leaf sorting performance. Future studies should conduct in situ spectral acquisition from fresh tea leaves, perform independent validation across batches and seasons, and establish automated reassessment triggers for targets with low-confidence predictions, similar class probabilities, or severe occlusion. Following completion of these validations, the proposed framework may provide a useful reference for raw-material grading and the development of intelligent detection systems for Zijin Chan tea.

Statements and Declarations

Funding:

The authors did not receive support from any organization for the submitted work.

Competing Interests:

The authors have no relevant financial or non-financial interests to disclose.

Author Contributions

Linsen Li: Conceptualization, Methodology, Software, Formal analysis, Investigation, Data curation, Visualization, and Writing—original draft preparation.

Feihu Song: Conceptualization, Methodology, Supervision, Project administration, and Writing—review and editing.

Nailiang Zhang: Investigation, Validation, and Data curation.

Chunfang Song: Investigation and Validation.

Zhenfeng Li: Resources and Investigation.

Guangyuan Jin: Resources and Data curation.

Jing Li: Validation, Visualization, and Writing—review and editing.

All authors read and approved the final manuscript.

Data Availability:

The datasets generated and/or analyzed during the current study are available from the corresponding author on reasonable request.

Code Availability:

The custom code used in the current study is available from the corresponding author on reasonable request.

Ethics Approval:

Not applicable. This study involved plant materials only and did not involve human participants or animals.

Consent to Participate

Not applicable.

Consent for Publication

Not applicable.

References

Araújo MCU, Saldanha TCB, Galvão RKH, et al (2001) The successive projections algorithm for variable selection in spectroscopic multicomponent analysis. Chemom Intell Lab Syst 57:65–73. https://doi.org/10.1016/S0169-7439(01)00119-8

Barnes RJ, Dhanoa MS, Lister SJ (1989) Standard Normal Variate Transformation and De-Trending of Near-Infrared Diffuse Reflectance Spectra. Appl Spectrosc 43:772–777. https://doi.org/10.1366/0003702894202201

Breiman L (2001) Random Forests. Mach Learn 45:5–32. https://doi.org/10.1023/A:1010933404324

Bro R, Smilde AK (2014) Principal component analysis. Anal Methods 6:2812–2831. https://doi.org/10.1039/C3AY41907J

Das A, Subburaj VH, Yang Y, Bednarz CW (2026) A UAV image dataset for object detection with annotations generated using LabelImg and Roboflow. Data Brief 65:112483. https://doi.org/10.1016/j.dib.2026.112483

Galvao R, Araujo M, Jose G, et al (2005) A method for calibration and validation subset partitioning. Talanta 67:736–740. https://doi.org/10.1016/j.talanta.2005.03.025

Geladi P, MacDougall D, Martens H (1985) Linearization and Scatter-Correction for Near-Infrared Reflectance Spectra of Meat. Appl Spectrosc 39:491–500. https://doi.org/10.1366/0003702854248656

Gharibzahedi SMT, Barba FJ, Zhou J, et al (2022) Electronic Sensor Technologies in Monitoring Quality of Tea: A Review. Biosensors 12:356. https://doi.org/10.3390/bios12050356

Gill GS, Kumar A, Agarwal R (2013) Nondestructive grading of black tea based on physical parameters by texture analysis. Biosyst Eng 116:198–204. https://doi.org/10.1016/j.biosystemseng.2013.08.002

Guo J, Liang J, Xia H, et al (2024) An Improved Inception Network to classify black tea appearance quality. J Food Eng 369:111931. https://doi.org/10.1016/j.jfoodeng.2023.111931

He K, Zhang X, Ren S, Sun J (2015) Spatial Pyramid Pooling in Deep Convolutional Networks for Visual

Recognition. IEEE Trans Pattern Anal Mach Intell 37:1904–1916. https://doi.org/10.1109/TPAMI.2015.2389824

Hossain S, Mou RM, Hasan MM, et al (2018) Recognition and detection of tea leaf's diseases using support vector machine. In: 2018 IEEE 14th International Colloquium on Signal Processing & Its Applications (CSPA). IEEE, Batu Feringghi, pp 150–154

Hu J, Shen L, Sun G (2018) Squeeze-and-Excitation Networks. In: 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. IEEE, Salt Lake City, UT, pp 7132–7141

Huang Y, Dong W, Sanaeifar A, et al (2020) Development of simple identification models for four main catechins and caffeine in fresh green tea leaf based on visible and near-infrared spectroscopy. Comput Electron Agric 173:105388. https://doi.org/10.1016/j.compag.2020.105388

Jabeen S, Alam S, Saleem M, et al (2019) Withering timings affect the total free amino acids and mineral contents of tea leaves during black tea manufacturing. Arab J Chem 12:2411–2417. https://doi.org/10.1016/j.arabjc.2015.03.011

Jin G, Wang Y, Li L, et al (2020) Intelligent evaluation of black tea fermentation degree by FT-NIR and computer vision based on data fusion strategy. LWT 125:109216. https://doi.org/10.1016/j.lwt.2020.109216

Kalidass S, Daiyarvijaya K, Kumar RR (2021) Comparative Study on Quantification of Total Catechins Using UV-Vis Spectrophotometric Method and High Performance Liquid Chromatography Techniques. Orient J Chem 37:136–142. https://doi.org/10.13005/ojc/370118

Lee S, Lin S, Chen S (2020) Identification of tea foliar diseases and pest damage under practical field conditions using a convolutional neural network. Plant Pathol 69:1731–1739. https://doi.org/10.1111/ppa.13251

Li H-D, Xu Q-S, Liang Y-Z (2012) Random frog: An efficient reversible jump Markov Chain Monte Carlo-like approach for variable selection with applications to gene selection and disease classification. Anal Chim Acta 740:20–26. https://doi.org/10.1016/j.aca.2012.06.031

Liao Y, Yu Z, Liu X, et al (2019) Effect of Major Tea Insect Attack on Formation of Quality-Related Nonvolatile

Specialized Metabolites in Tea (*Camellia sinensis*) Leaves. J Agric Food Chem 67:6716–6724. https://doi.org/10.1021/acs.jafc.9b01854

Lin J, Bai D, Xu R, Lin H (2023) TSBA-YOLO: An Improved Tea Diseases Detection Model Based on Attention Mechanisms and Feature Fusion. Forests 14:619. https://doi.org/10.3390/f14030619

Liu W, Anguelov D, Erhan D, et al (2016) SSD: Single Shot MultiBox Detector. In: Leibe B, Matas J, Sebe N, Welling M (eds) Computer Vision – ECCV 2016. Springer International Publishing, Cham, pp 21–37

Luo W, Tian P, Fan G, et al (2022) Non-destructive determination of four tea polyphenols in fresh tea using visible and near-infrared spectroscopy. Infrared Phys Technol 123:104037. https://doi.org/10.1016/j.infrared.2022.104037

Musci M, Yao S (2017) Optimization and validation of Folin–Ciocalteu method for the determination of total polyphenol content of Pu-erh tea. Int J Food Sci Nutr 68:913–918. https://doi.org/10.1080/09637486.2017.1311844

Ren G, Gan N, Song Y, et al (2021) Evaluating Congou black tea quality using a lab-made computer vision system coupled with morphological features and chemometrics. Microchem J 160:105600. https://doi.org/10.1016/j.microc.2020.105600

Ren S, He K, Girshick R, Sun J (2017) Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks. IEEE Trans Pattern Anal Mach Intell 39:1137–1149. https://doi.org/10.1109/TPAMI.2016.2577031

Sapkota R, Flores-Calero M, Qureshi R, et al (2025) YOLO advances to its genesis: a decadal and comprehensive review of the You Only Look Once (YOLO) series. Artif Intell Rev 58:274. https://doi.org/10.1007/s10462-025-11253-3

Sapkota R, Meng Z, Churuvija M, et al (2026) Comprehensive performance evaluation of YOLOv12, YOLO11, YOLOv10, YOLOv9 and YOLOv8 on detecting and counting fruitlet in complex orchard environments. Agric Commun 4:100125. https://doi.org/10.1016/j.agrcom.2026.100125

Scott ER, Li X, Wei J-P, et al (2020) Changes in Tea Plant Secondary Metabolite Profiles as a Function of

Leafhopper Density and Damage. Front Plant Sci 11:636. https://doi.org/10.3389/fpls.2020.00636

Selvaraju RR, Cogswell M, Das A, et al (2017) Grad-CAM: Visual Explanations from Deep Networks via Gradient-Based Localization. In: 2017 IEEE International Conference on Computer Vision (ICCV). IEEE, Venice, pp 618–626

Shao C, Zhang C, Lv Z, Shen C (2021) Pre- and post-harvest exposure to stress influence quality-related metabolites in fresh tea leaves (Camellia sinensis). Sci Hortic 281:109984. https://doi.org/10.1016/j.scienta.2021.109984

Sheng R, Cheng W, Li H, et al (2019) Model development for soluble solids and lycopene contents of cherry tomato at different temperatures using near-infrared spectroscopy. Postharvest Biol Technol 156:110952. https://doi.org/10.1016/j.postharvbio.2019.110952

Song F, Lu X, Lin Y, et al (2024) Evaluation of black tea appearance quality using a segmentation-based feature extraction method. Food Biosci 58:103644. https://doi.org/10.1016/j.fbio.2024.103644

Tharwat A, Gaber T, Ibrahim A, Hassanien AE (2017) Linear discriminant analysis: A detailed tutorial. AI Commun 30:169–190. https://doi.org/10.3233/AIC-170729

Tian W, Li Y, Guzman C, et al (2023) Quantification of food bioactives by NIR spectroscopy: Current insights, long-lasting challenges, and future trends. J Food Compos Anal 124:105708. https://doi.org/10.1016/j.jfca.2023.105708

Vuletic N, Bardic L, Odzak R (2021) Spectrophotometric determining of caffeine content in the selection of teas, soft and energy drinks available on the Croatian market. Food Res 5:325–330. https://doi.org/10.26656/fr.2017.5(2).482

Wang H, Gu J, Wang M (2023) A review on the application of computer vision and machine learning in the tea industry. Front Sustain Food Syst 7:1172543. https://doi.org/10.3389/fsufs.2023.1172543

Wang J, Wang Y, Cheng J, et al (2018) Enhanced cross-category models for predicting the total polyphenols, caffeine and free amino acids contents in Chinese tea using NIR spectroscopy. LWT 96:90–97. https://doi.org/10.1016/j.lwt.2018.05.012

Wang Y, Cui Q, Jin S, et al (2022) Tea Analyzer: A low-cost and portable tool for quality quantification of postharvest fresh tea leaves. LWT 159:113248. https://doi.org/10.1016/j.lwt.2022.113248

Wang Y, Ren Y, Kang S, et al (2024) Identification of tea quality at different picking periods: A hyperspectral system coupled with a multibranch kernel attention network. Food Chem 433:137307. https://doi.org/10.1016/j.foodchem.2023.137307

Xue Z, Xu R, Bai D, Lin H (2023) YOLO-Tea: A Tea Disease Detection Model Improved by YOLOv5. Forests 14:415. https://doi.org/10.3390/f14020415

Yang J, Wang J, Lu G, et al (2021) TeaNet: Deep learning on Near-Infrared Spectroscopy (NIR) data for the assurance of tea quality. Comput Electron Agric 190:106431. https://doi.org/10.1016/j.compag.2021.106431

Yemm EW, Willis AJ (1954) The estimation of carbohydrates in plant extracts by anthrone. Biochem J 57:508–514. https://doi.org/10.1042/bj0570508

Zhang H, Cisse M, Dauphin YN, Lopez-Paz D (2018) mixup: Beyond Empirical Risk Minimization. In: International Conference on Learning Representations

Zhang X, Song Y, Song T, et al (2024) LDConv: Linear deformable convolution for improving convolutional neural networks. Image Vis Comput 149:105190. https://doi.org/10.1016/j.imavis.2024.105190

Tables

Table 1. Image Acquisition, Dataset Partitioning, and Model Training Parameters

Item	Parameter or Description
Image Acquisition Equipment	Daheng Imaging MERCURY2 MER2-160-227-U3C industrial camera
Image Resolution	3,264×2,448
Camera-to-Sample Distance	35 cm
Number of Original Images	2,000 images
Annotation Tool	LabelImg
Annotated Classes	Slight, Moderate, Severe
Dataset Division	1,600 images in the training set, 200 images in the validation set, and 200 images in the test set
Number of Annotated Targets in the Test Set	4,000 annotated targets
Number of Bounding-Box Annotations	40,960 bounding boxes
Data Augmentation Methods	Online random affine transformation, horizontal flipping, scaling, and MixUp
Baseline Model	YOLOv8s
Improvement Modules	SE_C2f, LDConv, additional SPPF
Operating System	Windows 10
GPU	NVIDIA GeForce RTX 3080

Item	Parameter or Description
CPU	12th Gen Intel Core i5-12400F
Programming Environment	Python 3.10.15, CUDA 12.4, PyTorch 2.5.1
Training Epochs	200
Batch size	32
Initial Learning Rate	0.016
Object-Detection Evaluation Metrics	Precision, Recall, mAP@0.5
Machine-Vision Initial-Screening Metrics	Machine-vision initial-screening coverage Classification accuracy for successfully detected targets
Overall Evaluation of the Two-Stage Workflow	The overall workflow accuracy was calculated from the numbers of targets correctly classified by machine vision and correctly reassessed by spectroscopy

Table 2. Near-Infrared Spectral Acquisition, Sample Preparation, and Data Partitioning

Item	Content
Classes of Samples Used for Classification Modeling	Zijin Chan tea samples with slight, moderate, and severe leafhopper feeding damage
Purpose of Unfed Control Samples	Used exclusively to analyze changes in quality-related constituents and spectral responses; not included in classification modeling
Classification Task	Three-class classification of slight, moderate, and severe leafhopper feeding damage
Modeling Objective	To establish in advance a near-infrared spectroscopic model for classifying the severity of leafhopper feeding damage in Zijin Chan tea and apply it to the physical samples corresponding to machine-vision no-result targets
Sample Processing Method	Rapidly frozen and stored at −80°C, transported under cold-chain conditions, freeze-dried at −55°C for 24 h, ground, and passed through a 40-mesh sieve
Spectral Acquisition System	NIR2500 portable near-infrared spectrometer, HL2000 halogen light source, and FIB-Y-200-NIR seven-channel Y-shaped optical fiber
Reference and Spectral Acquisition Software	STD-WS standard white reference panel and Morpho spectral processing software
Acquisition Mode	Absorbance mode
Ambient Temperature	24 °C
Integration Time	100 ms

Item	Content
Smoothing Window	0
Original Spectral Range	858–2,590 nm
Modeling Spectral Range	1,000–2,400 nm
Spectral Modeling Samples	A total of 240 independent physical samples were partitioned at a ratio of 3:1 into a training set of 180 samples and a spectral model-selection set of 60 samples
Spectra Acquired Per Sample	Three consecutive spectra acquired and averaged
Training-Set Evaluation	Ten-fold cross-validation was performed on the 180-sample training set
Training-Set Data Processing	Spectral preprocessing, PPMCC-based wavelength selection, SPA-based feature-variable selection, and hyperparameter optimization were all performed using the training set
Spectral Model-Selection Set Evaluation	The 60-sample spectral model-selection set was excluded from model fitting and hyperparameter optimization, and its results were used to compare the performance of the candidate models
Purpose of the External Reassessment Set	Physical samples corresponding to 492 machine-vision no-result targets; used solely to evaluate the fixed model and excluded from model selection
Quality-Related Constituent Measurement Indices	Tea polyphenols, catechins, caffeine, free amino acids, and soluble sugars

Table 3. Spectral Data Processing, Feature Selection, and Model Evaluation Methods

Processing Step	Methods or Settings	Purpose or Evaluation Basis
Spectral Range Selection	1,000–2,400 nm	Exclusion of noise-prone regions at both ends of the raw spectra to improve modeling data quality
Spectral Preprocessing	S–G smoothing, SNV, MSC, first derivative, and second derivative	Compare the correction effects of different methods on noise, scattering effects, and baseline drift
Screening of Quality-Related Spectral Bands	PPMCC	Screen spectral bands that are strongly correlated with tea polyphenols, catechins, caffeine, free amino acids, and soluble sugars
Correlation Criterion	|r| > 0.6	Use spectral bands that meet the correlation threshold as candidate quality-related bands
Feature Variable Selection	PCA, SPA, Random Frog	Compare the effects of different feature selection methods on variable dimensionality reduction, redundancy elimination, and model discrimination performance

Processing Step	Methods or Settings	Purpose or Evaluation Basis
Construction of Discriminant Models	linear discriminant analysis (LDA), support vector machine (SVM), and random forest (RF)	Compare the applicability of different classification models for reassessing leafhopper feeding-damage severity
Model Evaluation	Training-Set Ten-Fold Cross-Validation and Spectral Model-Selection Set	Candidate-model comparison based on training-set cross-validation and spectral model-selection set macro-averaged recall
Training/Model-Selection-Set Partitioning Principle	The 240 samples were partitioned at a ratio of 3:1 into a training set of 180 samples and a spectral model-selection set of 60 samples	The spectral model-selection set was excluded from model fitting and hyperparameter optimization, and its results were used to compare the performance of the candidate models
Optimal Model Selection	Training-set ten-fold cross-validation, spectral model-selection set macro-averaged recall, and number of variables	Selection of the Final Spectral Model Combination

Table 4. Ablation Results for the Improved YOLOv8s Model

Model	P (%)	R (%)	mAP@0.5 (%)	Parameters (M)	GFLOPs
YOLOv8s	76.4	78.4	80.6	11.13	28.4
YOLOv8s+SE_C2f	78.6	81.8	81.7	11.15	28.5
YOLOv8s+LDConv	77.5	80.5	81.3	11.07	28.3
YOLOv8s+SPPF	85.2	81.6	82.1	11.78	29.0
YOLOv8s+SE_C2f+LDConv	82.4	84.1	86.3	11.02	28.3
YOLOv8s+SE_C2f+SPPF	87.4	86.3	88.1	11.81	29.9
YOLOv8s+LDConv+SPPF	86.1	85.6	87.6	11.09	28.9
Improved YOLOv8s	88.6	87.7	89.3	11.04	29.1

Note: P, R, and mAP@0.5 are evaluation metrics for the object-detection model and were used to compare the performance of the ablation variants. Machine-vision initial-screening coverage and classification accuracy for successfully detected targets, which were used to calculate the overall accuracy of the two-stage workflow, were determined according to the actual flow of annotated targets through the workflow, as described in Section 3.4

Table 5. Performance Comparison of Different Object-Detection Models

Model	P (%)	R (%)	mAP@0.5 (%)	Parameters (M)	GFLOPs
SSD	72.2	74.1	71.6	11.02	28.9
Faster R-CNN	83.5	85.1	83.9	178.35	452.6
YOLOv5	74.1	78.1	74.2	19.87	29.3
YOLOv11	80.7	81.4	81.7	11.09	33.7
YOLOv12	79.8	80.9	81.5	11.06	34.9
Improved YOLOv8s	88.6	87.7	89.3	11.04	29.1

Note: P, R, and mAP@0.5 are evaluation metrics for the object-detection models and were used solely to compare the performance of different detection models. They were not used directly to calculate the overall accuracy of the two-stage workflow

Table 6. Macro-Averaged Recall on the Training Set and Spectral Model-Selection Set for Different Spectral Variable-Processing Methods and Three-Class Classification Models

Classification model	Spectral preprocessing	Variable processing methods	Parameters	Training-Set Cross-Validated Macro-Averaged Recall (%)	Spectral Model-Selection Set Macro-Averaged Recall (%)
LDA	SNV	PPMCC	—	67.29	66.06
LDA	SNV	PCA	—	74.24	75.65
LDA	SNV	PPMCC+PCA	—	86.03	87.23
LDA	SNV	SPA	—	73.03	72.13
LDA	SNV	PPMCC+SPA	—	83.25	79.89
LDA	SNV	Random Frog	—	72.29	73.24
LDA	SNV	PPMCC+Random Frog	—	81.89	82.97
SVM	SNV	PPMCC	c=0.7695 g=0.02	67.26	68.33
SVM	SNV	PCA	c=0.5895 g=0.03	70.24	71.65
SVM	SNV	PPMCC+PCA	c=0.5899 g=0.01	80.29	79.06

Classification model	Spectral preprocessing	Variable processing methods	Parameters	Training-Set Cross-Validated Macro-Averaged Recall (%)	Spectral Model-Selection Set Macro-Averaged Recall (%)
SVM	SNV	SPA	c=0.6105 g=0.02	78.03	80.17
SVM	SNV	PPMCC+SPA	c=0.7013 g=0.03	94.03	95.12
SVM	SNV	Random Frog	c=0.5905 g=0.05	76.29	77.24
SVM	SNV	PPMCC+Random Frog	c=0.4721 g=0.02	87.62	89.56
RF	SNV	PPMCC	N_{tree}=140	66.47	67.23
RF	SNV	PCA	N_{tree}=130	71.24	70.65
RF	SNV	PPMCC+PCA	N_{tree}=110	83.03	83.18
RF	SNV	SPA	N_{tree}=145	76.03	77.18
RF	SNV	PPMCC+SPA	N_{tree}=130	89.24	89.65
RF	SNV	Random Frog	N_{tree}=160	79.29	80.03
RF	SNV	PPMCC+Random Frog	N_{tree}=138	92.29	93.24

Note: The 240 spectral samples representing the three feeding-damage classes were partitioned using the Kennard–Stone (K–S) algorithm at a ratio of 3:1 into a training set of 180 samples and a spectral model-selection set of 60 samples. Model fitting, variable selection, and hyperparameter optimization were performed exclusively on the training set, using ten-fold cross-validation for training-stage evaluation. The spectral model-selection set was excluded from model fitting and hyperparameter optimization, and its results were used to compare the performance of the candidate models. The values reported in the table are the macro-averaged recall values across the Slight, Moderate, and Severe classes. Unfed control samples were excluded from classification modeling

Figure Captions

Fig. 1 Representative fresh Zijin Chan tea shoots with different levels of leafhopper feeding damage

Fig. 2 Heatmap Visualization Results of the Improved YOLOv8s Model

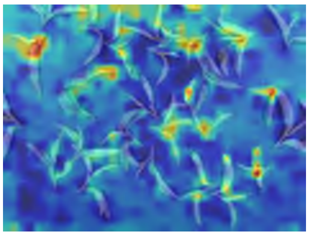
Fig. 3 Missed Detection and Misclassification Cases of the Improved YOLOv8s Model

Fig. 4 Changes in Quality-Related Constituents of Unfed Control and Zijin Chan Tea Samples with Different Levels of Leafhopper Feeding Damage

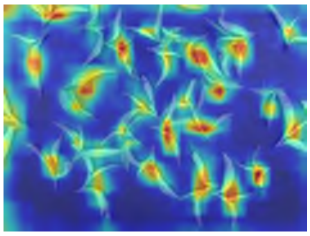
Fig. 5 Raw Spectra and SNV-Preprocessed Spectra

Fig. 6 Selection Results for Quality-Related Spectral Regions and Feature Wavelengths

Fig. 7 Normalized Confusion Matrices of the SNV + PPMCC + SPA + SVM Model for the Training Set and Spectral Model-Selection Set



(a) Baseline YOLOv8s



(b) Improved YOLOv8s