

Multispectral imaging-based detection of *Acidovorax citrulli*: from colony identification to infested seed discrimination

Yinuo Xin¹

Vegetable Research Center, Beijing Academy of Agriculture and Forestry Sciences

Yanhong Qiu

Vegetable Research Center, Beijing Academy of Agriculture and Forestry Sciences

Yang Xia

Vegetable Research Center, Beijing Academy of Agriculture and Forestry Sciences

Haijun Zhang

Vegetable Research Center, Beijing Academy of Agriculture and Forestry Sciences

Hongyang Wang

Vegetable Research Center, Beijing Academy of Agriculture and Forestry Sciences

Yongtao Yu

Vegetable Research Center, Beijing Academy of Agriculture and Forestry Sciences

Laixin Luo

China Agricultural University

Xiulan Xu

xuxiulan@nercv.org

Vegetable Research Center, Beijing Academy of Agriculture and Forestry Sciences

Research Article

Keywords: multispectral imaging, *Acidovorax citrulli*, colony identification, seed infestation detection, artificially inoculated seeds, naturally infested seeds

Posted Date: August 6th, 2026

DOI: <https://doi.org/10.21203/rs.3.rs-10488281/v1>

License: © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License. [Read Full License](#)

Additional Declarations: No competing interests reported.

Abstract

Bacterial fruit blotch (BFB) caused by *Acidovorax citrulli*, is a destructive seed-transmitted disease that seriously threatens global cucurbit production. To address the need for detecting *A. citrulli*-infested seeds, this study developed a colony identification model and a seed infestation detection model based on multispectral imaging. The combined nMahalanobis and nCDA colony identification models achieved a high recall of 0.999 and a low false-positive rate of 0.149 when tested on samples. For infested melon seed detection, we evaluated and compared the classification performance of seven machine learning models. The results showed that LDA, logistic regression, and MLP exhibited stable performance on artificially infested seed samples. Furthermore, multi-cultivar modeling improved model generalizability and demonstrated the feasibility of using multispectral imaging to identify naturally infested seeds. When a qPCR Ct threshold of 37 was used to define seed infestation status, the logistic regression model achieved a validation accuracy of 0.82. Overall, these findings demonstrate the potential of multispectral imaging for colony identification and seed infestation detection, providing a new technical approach and a scientific basis for seed health testing of bacterial fruit blotch in cucurbit crops.

Introduction

Cucumis melo L. ranks among the most economically important horticultural crops worldwide. China is the world's leading melon producer and accounts for more than half of global production^[1]. In recent years, advances in cultivar breeding, field cultivation, and postharvest management have substantially improved melon yield and market value. However, the expansion of large-scale cultivation and the rapid growth of international seed trade have promoted the development of the melon industry while increasing the risk of long-distance transmission of seed-borne pathogens, posing a serious threat to melon production and seed quality.

Bacterial fruit blotch (BFB), caused by *Acidovorax citrulli*, is a destructive bacterial disease of cucurbit crops, mainly infecting watermelon and melon^[2]. This disease is characterized by rapid epidemic spread and severe damage in field production^[3]. Seed infestation serves as the primary pathway for the long-distance spread of this pathogen^[4]. *Acidovorax citrulli* can survive for long periods in seeds and infect seedlings during germination, leading to disease outbreaks in the field^[4, 5]. Therefore, detecting infested seed is critical for preventing pathogen transmission and enabling effective source control.

Current detection of *Acidovorax citrulli* mainly relies on symptom observation, pathogen isolation and culture, serological assays (e.g., ELISA), and molecular methods (e.g., PCR and qPCR)^[6, 7]. Although these methods yield high sensitivity and specificity, they generally require destructive sample processing, specialized equipment, and skilled operators, limiting their applicability for rapid and large-scale seed screening. Therefore, the development of rapid, non-destructive, and high-throughput seed health detection techniques has become a critical research priority.

In recent years, optical sensing techniques such as Raman spectroscopy, near-infrared spectroscopy, and hyperspectral imaging have been extensively explored for agricultural product quality assessment and pathogen detection. Raman spectroscopy delivers detailed molecular structural information but is often affected by fluorescence background interference^[8]. Near-infrared spectroscopy allows rapid detection with minimal sample preparation; however, it suffers from limited spatial resolution and strong dependence on chemometric

models^[9]. Hyperspectral imaging simultaneously captures spectral and spatial information but is limited by large data volumes, high computational burden, and high hardware cost^[10].

In contrast, compared with these techniques, multispectral imaging captures images at a limited number of selected wavelengths, retaining key spectral information while substantially reducing data dimensionality and computational burden. In addition to spectral information, this technique enables the extraction of complementary phenotypic features, including color, texture, and morphology, thereby improving sample discrimination and classification accuracy. Furthermore, multispectral imaging offers several practical advantages, including rapid detection, efficient data processing, and relatively low operational cost^[11]. To date, multispectral imaging has been extensively explored for seed quality assessment, including cultivar discrimination^[12], damage detection^[13], seed vigor evaluation^[14], and fungal infection detection^[15].

However, spectroscopic techniques remain underutilized for the detection of seed-borne bacterial pathogens. Representative studies include hyperspectral imaging for identifying rice seeds infected with bacterial panicle blight^[14] and Raman–hyperspectral imaging for detecting *Acidovorax citrulli*-infected watermelon seeds^[16]. These studies demonstrate that spectroscopic techniques can effectively capture subtle changes in seed physicochemical properties induced by pathogen infection. Nevertheless, most existing studies have been conducted using artificially inoculated seeds. Systematic evaluations of model generalizability across different cultivars, detection performance on naturally infested seeds, and practical field applicability remain limited^[17]. In addition, pathogen isolation and culture remain an indispensable procedure in international seed testing and disease diagnosis. During pathogen isolation, plating, and incubation from seed samples, target colonies are typically small and often accompanied by abundant background microbes and morphologically similar colonies. These issues severely reduce the efficiency of traditional manual identification, making it highly dependent on operator expertise.

Although substantial studies have focused on microbial colony classification and identification using spectral imaging techniques^[18, 19], limited research has addressed the precise identification and direct localization of target colonies on culture plates.

Therefore, to address the aforementioned limitations, this study adopted multispectral imaging to develop an integrated detection strategy for rapid *Acidovorax citrulli* colony identification and non-destructive detection of infested melon seeds. Specifically, discriminative models based on normalized Mahalanobis distance and normalized canonical discriminant analysis (nCDA) were developed to enable rapid identification and precise localization of *A. citrulli* colonies on culture plates. Machine learning-based classification models were subsequently constructed using artificially inoculated melon seeds, and their cross-cultivar transferability was systematically evaluated. Finally, naturally infested seed samples were incorporated to validate the practical applicability of multispectral imaging under realistic infection conditions. Collectively, this study aims to provide an efficient strategy for rapid seed health screening, offering technical support for the early detection and management of bacterial fruit blotch in cucurbit crops, while laying a foundation for the application of multispectral imaging in seed-borne bacterial disease detection.

Materials and methods

Preparation of *A. citrulli* colony plates

The *Acidovorax citrulli* strains used in this study were sourced from the Vegetable Seed Quality Inspection and Testing Center of Beijing Academy of Agriculture and Forestry Sciences and the Seed Pathology Laboratory, College of Plant Protection, China Agricultural University. To evaluate the specificity of the colony identification models, *Acidovorax avenae*, *Pseudomonas syringae* pv. *lachrymans*, and several morphologically similar bacterial isolates derived from cucurbit seed testing samples were included for specificity assessment as listed in Table 1.

Table 1
Bacterial strains used for multispectral model development

Model construction	Bacteria	Strain
Target in Model 1, 2, 3, and 4	<i>Acidovorax citrulli</i>	Group Ⅰ: Xu201802, Xu201803, Xu201804, Xu201805, Xu201806, Xu201807
		Group Ⅱ: Xu8-14, Xu3-15, Xu9-15, Xu21-15, Xu25-15, Xu202202
Non-target in Model 1	<i>Agrobacterium</i> sp.	Xu2025-02, Xu2025-25
	<i>Acidovorax avenae</i>	Xu2025-05
	<i>Staphylococcus</i> sp.	Xu2025-33
Non-target in Model 2	<i>Pseudomonas syringae</i> pv. <i>lachrymans</i>	Xu2025-04
	<i>Acidovorax avenae</i>	Xu2025-05
Non-target in Model 3	<i>Agrobacterium</i> sp.	Xu2025-02, Xu2025-25
	<i>Delftia</i> sp.	Xu2025-23
	<i>Stenotrophomonas</i> sp.	Xu2025-24
	<i>Enterobacter</i> sp.	Xu2025-26
	<i>Achromobacter</i> sp.	Xu2025-28
	<i>Brevundimonas</i> sp.	Xu2025-29
	<i>Sphingomonas</i> sp.	Xu2025-31
	<i>Staphylococcus</i> sp.	Xu2025-33
Non-target in Model 4	<i>Agrobacterium</i> sp.	Xu2025-02, Xu2025-25, Xu2025-27, Xu2025-32

All *A. citrulli* strains were preserved at – 80°C. Before experimental use, the strains were streaked on Luria–Bertani (LB) medium to obtain single colonies. Strain identity and genetic subgroups were confirmed by PCR analysis^[20] and BOX-PCR fingerprinting^[21]. The validated strains were subsequently used for multispectral image acquisition and colony identification model development.

To evaluate the practical applicability of the established models in real samples, pathogen isolation by plating on LB medium was conducted on 11 naturally infested melon and watermelon seed and seedling samples. This process generated three types of culture plates: pure *A. citrulli* colonies (23 plates), mixed plates containing both

A. citrulli and other bacteria (32 plates), and plates containing only other bacteria (50 plates). All plate datasets were subsequently used for model evaluation.

Preparation of *A. citrulli*-infested melon seeds

Artificially infested seeds were prepared using a fermentation-based inoculation method to simulate natural *A. citrulli* seed infestation during seed harvest. Three melon cultivars, including 'Xizhoumi', 'Yunaixiang', and 'Hami melon', were used in this study. Mature, healthy melon fruits were surface-sterilized with 75% ethanol and aseptically dissected to collect intact seeds. For every 25 g of seeds, 50 mL of fruit juice was added. Activated *A. citrulli* cultures were adjusted to an OD₆₀₀ of approximately 0.5 (10⁸ CFU mL⁻¹) and inoculated into the fruit juice system, followed by fermentation at 30°C with shaking at 60 rpm min⁻¹. Seed samples with graded infestation levels were generated by varying two experimental variables: inoculation volume (2 or 4 mL) and fermentation duration (24 or 48 h). After fermentation, all seeds were thoroughly rinsed with sterile water and air-dried under ambient conditions^[22]. Non-inoculated seeds were used as healthy controls.

Naturally infested melon seeds were collected from commercial seed batches provided by Jinyan Yinong (Beijing) Seed Sci-Tech. Co., Ltd. Two pairs of species-specific primers for real-time quantitative PCR (qPCR) were used to determine seed infestation status^[23, 24]. The cycle threshold (Ct) value was used as the diagnostic criterion for seed infestation, and presence of *A. citrulli* were confirmed by grow-out assays. Specifically, artificially infested seed samples were used for model development and cross-cultivar transferability evaluation, whereas naturally infested seeds were used to validate the practical applicability of the developed multispectral detection models.

Multispectral imaging system

All multispectral image acquisition was performed using a VideometerLab multispectral imaging system (Videometer A/S, Hørsholm, Denmark). The system comprises an integrating sphere illumination unit, a CCD camera, and computer-controlled software. It is equipped with 19 LED light sources that are sequentially activated for multispectral image acquisition, producing a three-dimensional multispectral data cube (x, y, λ) spanning a spectral range of 365–960 nm. The corresponding central wavelengths are 365, 405, 430, 450, 470, 490, 515, 550, 570, 590, 630, 645, 660, 690, 780, 850, 880, 940, and 960 nm. The system supports both reflectance and fluorescence imaging modes. In fluorescence mode, samples are excited at specific wavelengths, and emitted signals are collected through a long-pass filter.

Multispectral images were acquired under controlled and uniform illumination conditions. During image acquisition, environmental conditions were kept constant to minimize external interference. After acquisition, foreground–background segmentation was performed to extract regions of interest (ROIs), which were subsequently used for spectral feature extraction and model development (Fig. 1).

Colony discriminant model

Colony discriminant models were constructed using the built-in discriminant analysis modules of VideometerLab software, including the normalized Mahalanobis distance (nMahalanobis) model and normalized canonical discriminant analysis (nCDA) model. Both models were constructed based on manually annotated

training datasets, while differing in their underlying algorithmic principles. The nMahalanobis model characterizes each class by modeling the statistical distribution of target colony regions using the mean vector and covariance matrix, and performs classification by comparing the Mahalanobis distance between test samples and the target class distribution against a predefined threshold. In contrast, the nCDA model constructs a discriminant space using both target and non-target colony regions and performs classification by simultaneously maximizing inter-class variance and minimizing intra-class variance.

Feature extraction from melon seeds

Following background removal, individual seed was segmented and treated as independent sample for feature extraction. In total, 41 quantitative variables encompassing spectral, color, and texture features were extracted for model development.

The spectral feature set comprised 24 variables, including 19 diffuse reflectance bands and five fluorescence bands. Color features were derived from the CIELab and CIELCh color spaces, including L^* , a^* , b^* , hue, and chroma. Texture features were used to characterize the spatial heterogeneity of seed surface morphology. The extracted texture descriptors consisted of directional texture parameters, local reflectance standard deviation, autocorrelation features, and histogram-based directional statistics, resulting in 12 texture variables in total. All features were automatically extracted using VideometerLab software and converted into numerical feature vectors for machine learning-based modeling and classification.

Infested melon seed classification model

A classification model for identifying *A. citrulli*-infested melon seeds was developed using Python(3.10.10). Each seed sample was assigned to two classes, healthy or infested, based on infestation status verified by qPCR. To minimize the effects of illumination variation and sample-to-sample variability, spectral features were preprocessed using standard normal variate (SNV) transformation to correct for scattering effects and baseline drift, whereas color and texture features were used without further preprocessing. Subsequently, all extracted features were standardized using StandardScaler to achieve zero mean and unit variance. The dataset was randomly split into training and validation sets at an 8:2 ratio using stratified sampling to preserve the original class distribution.

Seven machine learning classification models were developed, including support vector machine (SVM)^[25], multilayer perceptron (MLP)^[26], linear discriminant analysis (LDA)^[27], logistic regression^[28], random forest (RF)^[29], XGBoost^[30] and LightGBM^[31]. Feature importance was evaluated using the random forest (RF) model, and the 30 highest-ranked features were selected to train the SVM and MLP models. In contrast, all extracted features were used to develop the LDA, logistic regression, RF, XGBoost, and LightGBM models.

The hyperparameters of each model were determined based on published literature and empirical experience. The SVM model employed a radial basis function (RBF) kernel, with probability estimation enabled to generate class probabilities. The MLP model employed a two-hidden-layer architecture with 100 and 50 neurons in the first and second hidden layers, respectively, and a maximum of 500 training iterations. The logistic regression model was trained for a maximum of 500 iterations. The RF model consisted of 300 decision trees. The XGBoost

model was configured with 200 decision trees and used log loss as the evaluation metric during training. The LDA and LightGBM models were implemented using default parameter settings. All models were trained on the training set and subsequently evaluated on the validation and independent test sets.

Model evaluation and comparison

To comprehensively evaluate model performance, different evaluation metrics were selected according to the characteristics of each classification task. For *A. citrulli* colony recognition, classification performance was assessed using precision, recall, F1-score, specificity, and Cohen's kappa coefficient. For seed-borne *A. citrulli* detection, accuracy, precision, recall, F1-score, and the area under the receiver operating characteristic curve (AUC) were calculated to evaluate classification performance and generalization capability. Model performance was evaluated using the training, validation, and independent test sets. Comparative analyses were performed among different models according to previously described evaluation criteria^[32, 33, 34].

Results

Feasibility assessment and method evaluation for *A. citrulli* colony identification

To evaluate the feasibility of multispectral imaging for identifying *A. citrulli* colonies, an nMahalanobis discriminant model was constructed. *A. citrulli* was designated as the target class, whereas *Pseudomonas syringae* pv. *lachrymans* and five additional non-target bacterial isolates from watermelon samples representing different genera were used as the non-target class. As shown in Fig. 2A&B, the established model accurately identified *A. citrulli* colonies on mixed-culture plates. The identification results were fully consistent with PCR detection using *A. citrulli*-specific primers, achieving 100% accuracy with no false positives or false negatives. These results indicate that multispectral features can effectively discriminate bacterial colonies at the genus level and enable accurate identification of *A. citrulli* under complex microbial backgrounds.

Furthermore, the discriminative capability of multispectral imaging for closely related bacterial species was evaluated using *A. citrulli* and *A. avenae*, two species within the same genus. An nCDA discriminant model was established using pure-cultured colonies and applied to mixed-colony images for classification. As shown in Fig. 2C&D, the model achieved an average classification accuracy of 98.75% for both species on LB medium. These results demonstrate distinct spectral signatures between closely related bacterial species, enabling their reliable discrimination via multispectral imaging.

In addition, two genetic subgroups of *A. citrulli* (Group I and Group II) were analyzed to evaluate the influence of genetic background on model performance. The nCDA-based discriminant model failed to effectively distinguish the two subgroups (Fig. 2E), with evident cross-classification observed between training and validation isolates. The two subgroups exhibited highly similar spectral response profiles across all tested wavelengths, with an average Pearson correlation coefficient of 0.99, indicating no significant spectral differences between the two genetic lineages. These findings suggest that *A. citrulli* strains from different genetic backgrounds exhibit spectral consistency and can be incorporated into a unified classification model, thereby ensuring model robustness and generalizability.

Development and performance of colony identification models

An nMahalanobis discriminant model was established using *A. citrulli* colony images acquired from two independent plating experiments. The training dataset consisted of six Group I and six Group II strains, yielding a total of 24 multispectral images. As shown in the identification maps, the model achieved stable recognition of the training colonies, with no false negative detections (Fig. 3A). The model was further validated using colony plate images obtained from practical diagnostic samples (Fig. 3B). A total of 105 colony plates from 11 batches of test samples were included in the statistical analysis (Table 2). The nMahalanobis model achieved recall values of 1.000 and 0.999 for pure *A. citrulli* plates and mixed plates containing accompanying microorganisms, respectively, demonstrating robust target colony detection capability. However, for mixed plates containing non-target bacteria, the model achieved a precision of 0.787 and a Specificity of 0.692, indicating that false-positive detections still occurred under complex microbial backgrounds.

Table 2
Classification performance of the nMahalanobis model and four nCDA models on three types of culture plates.

Plate type	Plates with <i>A. citrulli</i> and other bacteria			Plates with other bacteria	Plates with pure <i>A. citrulli</i>
	Recall	Precision	Specificity	Specificity	Recall
nMahalanobis	0.999	0.787	0.692	0.728	1.000
Model 1-nCDA	0.659	0.845	0.867	0.846	0.933
Model 2-nCDA	0.437	0.824	0.895	0.798	0.438
Model 3-nCDA	0.542	0.885	0.922	0.860	0.704
Model 4-nCDA	0.637	0.904	0.924	0.753	0.929

Meanwhile, four nCDA-based discriminant models were established using different non-target bacterial species and subsequently validated using practical diagnostic samples. As shown in Fig. 3C, all four models successfully identified *A. citrulli* colonies, although their classification performance differed. On mixed plates containing *A. citrulli* and accompanying microorganisms, Model 1 and Model 4 achieved relatively higher recall values of 0.659 and 0.637, respectively. In contrast, Model 3 and Model 4 exhibited lower false positive rates (0.078 and 0.076, respectively) together with higher precision values (0.885 and 0.904, respectively) (Table 2). Furthermore, validation using plates containing only non-target bacteria and pure *A. citrulli* cultures further highlighted differences among the four models in balancing target detection capability and false-positive control.

Identification results for all test samples were comprehensively evaluated using four quantitative metrics, including recall, precision, F1-score, and Specificity (Fig. 4). The nMahalanobis model achieved the highest recall (0.999) but also exhibited the lowest precision (0.620) and the highest Specificity (0.719). In contrast, the four nCDA-based discriminant models generally exhibited lower false positive rates. Among them, Model 1 achieved the best overall performance, with recall, precision, and F1-score values of 0.761, 0.700, and 0.728, respectively. Taken together, these results suggest that the nMahalanobis model is more suitable for high-sensitivity preliminary screening, whereas the nCDA-based models are more effective at reducing false-positive detections.

Based on the comprehensive evaluation of the performance metrics, a hybrid discriminant strategy integrating the nMahalanobis model and Model 1 of the nCDA-based models was established. In this strategy, the two models based on distinct discriminant mechanisms were applied in parallel to the same multispectral images. The nMahalanobis model was used to maximize target colony detection while minimizing the risk of false negatives, whereas Model 1 of the nCDA-based models was used to improve classification precision by reducing false-positive detections. Owing to their complementary discriminative capabilities, the integration of these two models balanced high sensitivity with effective false-positive control, thereby providing more reliable candidate colonies for subsequent PCR validation and improving the overall efficiency of *A. citrulli* colony screening.

Development and validation of detection models for artificially infested seeds

To evaluate the performance of different machine learning algorithms for detecting fermentation-inoculated melon seeds, seven classification models were developed using either the first batch of fermentation-inoculated ‘Xizhoumi’ seeds or the combined dataset from two batches, and their performance was systematically compared.

When trained using the first batch of fermentation-inoculated ‘Xizhoumi’ melon seeds, the seven models exhibited clear differences in classification performance. Specifically, 360 infested and 300 healthy seeds were used for training, whereas an independent test set consisting of 360 infested and 240 healthy seeds was used for evaluation. Among the seven models, LDA, logistic regression, and MLP exhibited superior performance. Notably, both LDA and logistic regression achieved perfect results on the training and validation sets, with accuracy, precision, recall, and F1-score all reaching 1.000. On the independent test set, LDA and logistic regression achieved accuracies of 0.90 and 0.89, respectively, with recall values of 1.00 and 0.99, demonstrating strong generalization ability. MLP also performed well, achieving an accuracy of 0.89 and a recall of 0.99. In contrast, SVM, RF, XGBoost, and LightGBM showed relatively lower performance on the independent test set.

To further evaluate cross-batch generalizability, the two fermentation-inoculated seed datasets were combined to retrain the models (Table 3). The merged dataset included 720 infested and 600 healthy seeds for training, while 660 infested and 300 healthy seeds were used for testing. With increased sample size and greater variability in infestation levels, overall model performance remained consistent with that obtained from the single-batch dataset. LDA, logistic regression, and MLP continued to show superior performance. On the independent test set, LDA achieved an accuracy of 0.93 and a recall of 0.97, logistic regression achieved an accuracy of 0.93 and a recall of 0.99, whereas MLP achieved an accuracy of 0.91 and a recall of 0.96. These results indicate that LDA, logistic regression, and MLP can accommodate variations in data distribution caused by different fermentation durations, inoculum concentrations, and experimental batches, demonstrating strong robustness and generalization capability.

Table 3
Performance comparison of seven machine learning models using the combined fermentation-inoculated seed dataset.

Model	Train Accuracy	Validation Accuracy	Precision	Recall	F1	AUC	Test Accuracy	Precision	Recall
SVM	0.93	0.89	0.89	0.91	0.90	0.96	0.82	0.89	0.84
MLP	1.00	0.96	0.97	0.96	0.97	1.00	0.91	0.91	0.96
LDA	0.99	0.98	0.99	0.98	0.99	1.00	0.93	0.93	0.97
LR	0.97	0.96	0.97	0.96	0.97	0.99	0.93	0.91	0.99
RF	1.00	0.83	0.84	0.85	0.85	0.90	0.74	0.85	0.77
XGBoost	1.00	0.90	0.91	0.90	0.91	0.96	0.86	0.90	0.90
LGBM	1.00	0.92	0.93	0.93	0.93	0.97	0.86	0.90	0.90

To evaluate cross-variety transferability, the seven models trained on ‘Xizhoumi’ seeds were directly applied to fermentation-inoculated seeds of ‘Yunaixiang’ (480 seeds) and ‘Hami’ (360 seeds). As shown in Fig. 5A&B, all models exhibited a substantial decline in performance when applied to cross-variety datasets. Classification accuracies ranged from 0.56 to 0.69 for ‘Yunaixiang’ seeds and from 0.46 to 0.69 for ‘Hami’ seeds, indicating limited cross-variety generalization ability of single-variety models.

To improve model generalization, seed datasets from three melon varieties (‘Xizhoumi’, ‘Yunaixiang’, and ‘Hami’) were integrated to construct multi-variety models. For each variety, 240 seeds (120 infested and 120 healthy) were used for training, while an independent test set of 120 seeds (60 infested and 60 healthy) was used for evaluation. Compared with single-variety models Fig. 5C&D, multi-variety models showed improved overall performance. Among all models, LDA, logistic regression, and MLP achieved the best performance, with test accuracies of 0.85, 0.83, and 0.80, respectively, while maintaining a good balance between precision and recall.

Overall, multi-variety modeling expanded the spectral variability in the training data, thereby improving cross-variety classification performance. Compared with single-variety models, multi-variety models showed superior generalization ability.

Model evaluation using naturally infested seeds

To validate the ability of multispectral imaging to detect naturally infested melon seeds, the classification performance of seven machine learning models was evaluated under two quantitative real-time PCR (qPCR)-based infestation criteria, with $Ct < 40$ and $Ct < 37$ used as thresholds to define seed infestation status (Table 4).

Table 4
Performance comparison of seven machine learning models for identifying naturally infested melon seeds under two Ct thresholds.

Ct threshold	Model	Train Accuracy	Val Accuracy	Precision	Recall	F1	AUC
Ct < 40	SVM	0.72	0.55	0.73	0.55	0.63	0.59
	MLP	1.00	0.56	0.59	0.58	0.58	0.57
	LDA	0.73	0.65	0.63	0.68	0.65	0.68
	LR	0.67	0.61	0.64	0.62	0.63	0.67
	RF	1.00	0.53	0.64	0.54	0.59	0.58
	XGBoost	1.00	0.60	0.69	0.60	0.64	0.61
	LightGBM	1.00	0.61	0.69	0.61	0.65	0.66
Ct < 37	SVM	0.82	0.69	0.58	0.73	0.64	0.72
	MLP	1.00	0.69	0.55	0.75	0.63	0.71
	LDA	0.85	0.79	0.73	0.83	0.77	0.85
	LR	0.80	0.82	0.79	0.84	0.81	0.83
	RF	1.00	0.69	0.55	0.75	0.63	0.71
	XGBoost	1.00	0.67	0.64	0.68	0.66	0.70
	LightGBM	1.00	0.70	0.64	0.72	0.68	0.72

When Ct < 40 was used as the infestation criterion, all models exhibited relatively weak performance, with validation accuracies ranging from 0.53 to 0.65. In contrast, when the stricter criterion of Ct < 37 was applied, all models showed markedly improved performance and achieved substantially higher validation accuracies. Among the seven models, logistic regression achieved the best performance, with a validation accuracy of 0.82, followed by LDA at 0.79. These results indicate that lower Ct values correspond to higher pathogen loads, resulting in more pronounced spectral responses and improved classification performance for infested seed detection.

Key wavelength identification and feature importance analysis

To quantify the contributions of different features to seed classification, feature importance analysis was conducted for the three best-performing models: LDA, logistic regression, and MLP. Due to differences in their inherent algorithms, the feature importance rankings and selected features varied across models. Specifically, LDA assessed feature contributions based on discriminant coefficients, logistic regression quantified feature importance using regression coefficients, while MLP adopted SHAP values to interpret feature contributions to model predictions.

Although feature importance rankings differed across the three models, several spectral features consistently contributed substantially to model performance(Fig. 6). Notably, the 490 nm wavelength was consistently

ranked among the top 10 most important features across all models. In addition, wavelengths at 430, 450, 470, 570, 690, and 940 nm, along with fluorescence features and color descriptors such as Chroma and CIE b*, were repeatedly identified by multiple models. Overall, the dominant features were primarily distributed across the blue, red, and near-infrared spectral regions. These findings demonstrate that spectral and color information can reliably distinguish healthy seeds from *A. citrulli*-infested seeds, providing critical discriminative features for intelligent detection of infested melon seeds.

Discussion

Applications and limitations of multispectral imaging for *A. citrulli* colony identification

During the experiments, culture medium composition and colony growth conditions were identified as key factors influencing model performance. The LB medium, due to its higher transparency and more homogeneous background, facilitated colony segmentation and spectral feature extraction. In contrast, models were also developed using PF medium; however, its darker background reduced the reflectance contrast between colonies and the surrounding medium, thereby decreasing spectral separability and weakening model discrimination performance. In addition, high colony density, colony coalescence, and metabolite accumulation from accompanying microorganisms can alter local spectral characteristics, thereby increasing the risk of false-positive classification. Therefore, culture medium preparation, incubation conditions, colony density, and image acquisition procedures should be strictly standardized in practical applications to ensure model robustness and reproducibility.

Although the proposed multispectral imaging-based colony identification models enable rapid screening of *A. citrulli* colonies, several limitations remain. The imaging system used in this study included only 19 spectral bands and did not fully exploit complementary information such as colony morphology, texture, and fluorescence characteristics. Moreover, both the nMahalanobis and nCDA models are conventional statistical discriminant methods with limited capacity to capture complex nonlinear relationships in the data. Future studies should integrate multimodal features and incorporate nonlinear modeling approaches, such as deep learning methods, while expanding the diversity of *A. citrulli* isolates and accompanying microorganisms to further improve model performance and generalization under more complex sample conditions.

Challenges and prospects of multispectral detection of infested melon seeds

The transferability analysis demonstrated that models developed using a single melon variety exhibited substantially reduced performance when applied to other varieties. This decline was primarily attributed to varietal differences in seed coat color, microstructure, and chemical composition, which altered the intrinsic spectral characteristics of seeds and, in some cases, exceeded the spectral variations induced by *A. citrulli* infestation. Similar cultivar-dependent spectral variability has been reported in previous multispectral imaging studies of crop seed quality assessment and disease detection, where inherent differences among cultivars often compromise model transferability and generalization performance^[14]. In contrast, multi-variety modeling significantly improved model generalization, indicating that expanding the spectral variability in the training dataset enables models to learn more robust and representative pathogen-specific spectral signatures. Therefore, multi-variety modeling represents an effective strategy for improving cross-variety applicability. Future

studies should further incorporate naturally infested seed samples collected across different production years, geographic regions, and storage conditions to enhance model robustness under practical agricultural environment.

The classification performance for fermentation-inoculated seeds was markedly superior to that for naturally infested seeds, which can be explained by differences in infection patterns and pathogen load. In artificially inoculated seeds, pathogens were more uniformly distributed on the seed surface, resulting in stable and consistent spectral differences. In contrast, naturally infested seeds typically exhibited lower pathogen loads and heterogeneous distribution, leading to weaker spectral responses. In some low-infection samples, spectral features overlapped considerably with those of healthy seeds, increasing classification difficulty. The use of a more stringent infestation threshold ($Ct < 37$) significantly enhanced model performance, further indicating that multispectral imaging is currently more suitable for detecting seeds with relatively high pathogen loads, whereas its ability to identify low-level or latent infections remains limited and requires further optimization.

Interestingly, simple linear classifiers (LDA and logistic regression) consistently outperformed more complex nonlinear algorithms, including Random Forest, XGBoost, and LightGBM. This may be attributed to the high separability of the selected multispectral features after preprocessing, allowing linear decision boundaries to effectively distinguish infested and healthy seeds. Similar findings have been reported in multispectral seed analysis, where linear discriminant models achieved high classification performance when spectral features showed strong class separability^[14].

The key discriminative wavelengths were mainly distributed in the blue, red, and near-infrared regions, reflecting multi-scale physiological and microstructural alterations in seeds following pathogen infestation. The blue region (450–500 nm) is highly sensitive to phenolic compounds and stress-related metabolites^[34], and its variation may be associated with defense responses induced by *A. citrulli*. The band at 570 nm may be related to changes in carotenoid absorption and seed coat pigment composition^[35], reflecting stress-induced pigment regulation. The red region (690 nm), located near the chlorophyll a absorption feature^[36], may reflect structural changes in the seed coat and residual pigment effects. The near-infrared region (940 nm) is a typical water-sensitive band^[37], associated with changes in seed moisture content and tissue scattering properties. In addition, wavelengths in the 430–470 nm range are closely related to carotenoid and other pigment absorption features^[38, 39], further supporting the contribution of seed coat chemical alterations to spectral responses. These spectral variations are consistent with previous pathological studies reporting that *A. citrulli* can invade seeds through the seed coat or stigmatic pathway, colonize internal tissues, and accumulate in cuticle folds and microstructural regions, thereby altering local moisture and chemical composition^[40, 41]. Collectively, these changes result in stable spectral differences in the blue, red, and near-infrared regions, providing a solid theoretical basis for the identification of infested seeds.

Future studies should further expand the scale of naturally infested seed samples and establish comprehensive datasets covering multiple varieties and infection levels. In addition, integrating optimal wavelength selection, feature fusion strategies, and deep learning approaches may further improve model robustness and generalization ability. Based on the identified discriminative wavelengths, the development of compact multispectral sensing devices could also be explored, facilitating the practical application of this technology in seed health inspection, quarantine, and rapid pathogen screening.

Conclusion

This study explored the feasibility of multispectral imaging for rapid identification of *Acidovorax citrulli* colonies and non-destructive detection of infested melon seeds. For colony recognition, discriminative models based on normalized Mahalanobis distance and normalized CDA were established, and a combined decision strategy was further proposed. This integrated approach effectively reduced false-positive rates while maintaining high detection sensitivity, enhanced the overall reliability of colony identification, and demonstrated strong potential for rapid auxiliary detection of *A. citrulli*.

For seed infestation detection, seven machine learning models were systematically compared for their classification performance. Among these models, LDA, logistic regression, and MLP yielded the best overall performance on artificially inoculated seed samples. Compared with single-variety modeling, multi-variety modeling significantly improved cross-variety generalization capability. Furthermore, the classification results obtained from naturally infested seeds validated the practical applicability of multispectral imaging for seed health evaluation, and a Ct threshold of < 37 was identified as a more reliable criterion for distinguishing infested seeds.

Collectively, this study developed an integrated multispectral detection strategy spanning rapid *A. citrulli* colony identification and non-destructive detection of infested melon seeds. The proposed methodology provides an efficient technical approach for bacterial fruit blotch screening and routine melon seed health evaluation. Moreover, this work lays a theoretical and technical foundation for the application of multispectral imaging in seed-transmitted bacterial disease detection and supports the future development of targeted intelligent sensing systems.

Declarations

Author contributions

XX conceived and designed the experiment. Y. Xin and YQ performed the experiments.

Y. Xia and HZ analyzed the data. HW and YY collected seed material and contribute to the experiment. LL provided writing suggestions. Y. Xin and YQ wrote the manuscript, and XX revised the paper. and All authors contributed to the article and approved the submitted version.

Funding

This research was supported by Innovation and Development Program of Vegetable Research Center, Beijing Academy of Agriculture and Forestry Sciences (KJCX20261412 and KYCX202605).

Data availability

No datasets were generated or analyzed during the current study.

Code availability

Not applicable.

Ethical approval

Not applicable.

Consent to participate

Not applicable.

Consent for publication

Not applicable.

Competing interests

The authors declare no competing interests.

References

1. Food and Agriculture Organization of the United Nations. FAOSTAT: Crops and livestock products database. <https://www.fao.org/faostat/>. Accessed 18 Jul 2026.
2. Schaad NW, Postnikova E, Sechler A, Claflin LE, Vidaver AK, Jones JB, Agarkova I, Ignatov A, Dickstein E, Ramundo BA. Reclassification of subspecies of *Acidovorax avenae* as *A. avenae*, *A. cattleyae*, *A. citrulli*, and proposal of *A. oryzae* sp. nov. *Syst Appl Microbiol*. 2008;31:434–43. <https://doi.org/10.1016/j.syapm.2008.09.003>.
3. Zhao TC, Sun FZ, Wang BW. Research progress on bacterial fruit blotch of watermelon. *Plant Prot Technol*. 2001.
4. Block CC, Shepherd LM. Long-term survival and seed transmission of *Acidovorax avenae* subsp. *citrulli* in melon and watermelon seed. *Plant Health Prog*. 2008;9(1):1219–01. <https://doi.org/10.1094/PHP-2008-1219-01-BR>.
5. Dutta B, Schneider RW, Robertson CL, Walcott RR. Embryo localization enhances the survival of *Acidovorax citrulli* in watermelon seeds. *Phytopathology*. 2016;106:330–8. <https://doi.org/10.1094/PHYTO-09-15-0232-R>.
6. Yu HB, Bi XY, Xia B, An MN, Xia ZH, Wu YH. Identification of the pathogen causing bacterial fruit blotch of watermelon and melon in Liaoning Province. *J Shenyang Agric Univ*. 2019;50(3):345–50. <https://doi.org/10.13926/j.cnki.apps.000130>.
7. Hong XX, Wu XQ, Luo JY, Wang YL, Sun GC, An QL, Qiu W, Li B. First report of bacterial fruit blotch of cucurbits caused by *Acidovorax citrulli* on sweet potato in Zhejiang Province. *Zhejiang Agric J*. 2019;31(7):1112–8. <https://doi.org/10.3969/j.issn.1004-1524.2019.07.11>.
8. Qin J, Kim MS, Chao K, Dhakal S, Cho BK, Lohumi S, Mo C, Peng Y, Huang M. Advances in Raman spectroscopy and imaging techniques for quality and safety inspection of horticultural products. *Postharvest Biol Technol*. 2019;148:1–14. <https://doi.org/10.1016/j.postharvbio.2018.11.004>.
9. Cozzolino D. Advantages and limitations of using near infrared spectroscopy in plant phenomics applications. *Comput Electron Agric*. 2023;212:108078. <https://doi.org/10.1016/j.compag.2023.108078>.

10. Bhargava A, Sachdeva A, Sharma K, Alsharif MH, Uthansakul P, Uthansakul M. Hyperspectral imaging and its applications: A review. *Heliyon*. 2024;10(12):e33208. <https://doi.org/10.1016/j.heliyon.2024.e33208>.
11. Boelt B, Shrestha S, Salimi Z, Jørgensen JR, Nicolaisen M, Carstensen JM. Multispectral imaging—a new tool in seed quality assessment? *Seed Sci Res*. 2018;28(Suppl 3):222–8. <https://doi.org/10.1017/S0960258518000235>.
12. Shrestha S, Deleuran LC, Olesen MH, Gislum R. Use of multispectral imaging in varietal identification of tomato. *Sens (Basel)*. 2015;15(2):4496–512. <https://doi.org/10.3390/s150204496>.
13. Salimi Z, Boelt B. Classification of processing damage in sugar beet (*Beta vulgaris*) seeds by multispectral image analysis. *Sens (Basel)*. 2019;19(10):2360. <https://doi.org/10.3390/s19102360>.
14. Baek I, Kusumaningrum DM, Kandpal LM, Lohumi S, Mo C, Kim MS, et al. Rapid measurement of soybean seed viability using kernel-based multispectral image analysis. *Sens (Basel)*. 2019;19:271. <https://doi.org/10.3390/s19020271>.
15. Weng H, Tian Y, Wu N, Li X, Yang B, Huang Y, Ye D, Wu R. Development of a low-cost narrow band multispectral imaging system coupled with chemometric analysis for rapid detection of rice false smut in rice seed. *Sens (Basel)*. 2020;20(4):1209. <https://doi.org/10.3390/s20041209>.
16. Lee H, Kim MS, Qin J, Park E, Song YR, Oh CS, Cho BK. Raman hyperspectral imaging for detection of watermelon seeds infected with *Acidovorax citrulli*. *Sens (Basel)*. 2017;17(10):2188. <https://doi.org/10.3390/s17102188>.
17. Xin YN, Qiu YH, Luo LX, et al. Research progress on the application of spectral technology in seed health detection. *Seed*. 2025;44(10):105–12. <https://doi.org/10.16590/j.cnki.1001-4705.2025.10.105>.
18. Li Y, Hu X, Zhang F, Shi J, Qiu B. Colony image segmentation and counting based on hyperspectral technology. *Trans Chin Soc Agric Eng*. 2020;36(20):326–32. <https://doi.org/10.11975/j.issn.1002-6819.2020.20.038>.
19. Sekhon AS, Unger P, Sharma S, Singh B, Chen X, Ganjyal GM, Michael M. Hyperspectral imaging of foodborne pathogens at colony and cellular levels for rapid identification in dairy products. *Food Sci Nutr*. 2023;12(1):239–54. <https://doi.org/10.1002/fsn3.3766>.
20. Kan YM, Yun XM, Li YW, Li JQ, Luo LX. Screening of specific primers for Bio-PCR detection of *Acidovorax citrulli* in cucurbit seeds. *Acta Phytopathol Sin*. 2018;48(2):263–70. <https://doi.org/10.13926/j.cnki.apps.000130>.
21. Walcott RR, Fessehaie A, Castro AC. Differences in pathogenicity between two genetically distinct groups of *Acidovorax avenae* subsp. *citrulli* on cucurbit hosts. *J Phytopathol*. 2004;152(5):277–85. <https://doi.org/10.1111/j.1439-0434.2004.00841.x>.
22. Xu XL, Lu Y, Zhao ZJ, et al. Effects of different fermentation conditions on infection of watermelon seeds by *Acidovorax citrulli*. *Plant Prot*. 2018;44(4):60–64. <https://doi.org/10.16688/j.zwbh.2017424>.
23. Tian Q, Feng J, Hu J, et al. Selective detection of viable seed-borne *Acidovorax citrulli* by real-time PCR with propidium monoazide. *Sci Rep*. 2016;6:35457. <https://doi.org/10.1038/srep35457>.
24. Kleinhesselink K. *Validation report: Seed Extract and qPCR Method for the Detection of Acidovorax citrulli on Cucurbit Seeds*. Nyon (Switzerland): International Seed Federation; 2017.
25. Cortes C, Vapnik V. Support-vector networks. *Mach Learn*. 1995;20:273–97. <https://doi.org/10.1007/BF00994018>.

26. Rumelhart DE, Hinton GE, Williams RJ. Learning representations by back-propagating errors. *Nature*. 1986;323:533–6. <https://doi.org/10.1038/323533a0>.
27. Fisher RA. The use of multiple measurements in taxonomic problems. *Ann Eugen*. 1936;7(2):179–88. <https://doi.org/10.1111/j.1469-1809.1936.tb02137.x>.
28. Good IJ. *The Estimation of Probabilities: An Essay on Modern Bayesian Methods*. Cambridge (MA): MIT Press; 1965.
29. Breiman L. Random forests. *Mach Learn*. 2001;45:5–32. <https://doi.org/10.1023/A:1010933404324>.
30. Chen T, Guestrin C. XGBoost. A scalable tree boosting system. *Proc ACM SIGKDD Int Conf Knowl Discov Data Min*. 2016;785–94. <https://doi.org/10.1145/2939672.2939785>.
31. Ke G, Meng Q, Finley T, Wang T, Chen W, Ma W, Ye Q, Liu TY. LightGBM: a highly efficient gradient boosting decision tree. *Adv Neural Inf Process Syst*. 2017;30:3146–54.
32. Sokolova M, Lapalme G. A systematic analysis of performance measures for classification tasks. *Inf Process Manag*. 2009;45(4):427–37. <https://doi.org/10.1016/j.ipm.2009.03.002>.
33. Fawcett T. An introduction to ROC analysis. *Pattern Recognit Lett*. 2006;27(8):861–74. <https://doi.org/10.1016/j.patrec.2005.10.010>.
34. Cohen J. A coefficient of agreement for nominal scales. *Educ Psychol Meas*. 1960;20(1):37–46. <https://doi.org/10.1177/001316446002000104>.
35. Carter GA. Responses of leaf spectral reflectance to plant stress. *Am J Bot*. 1993;80:239–43. <https://doi.org/10.1002/j.1537-2197.1993.tb13796.x>.
36. Gamon JA, Peñuelas J, Field CB. A narrow-waveband spectral index that tracks diurnal changes in photosynthetic efficiency. *Remote Sens Environ*. 1992;41:35–44. [https://doi.org/10.1016/0034-4257\(92\)90059-S](https://doi.org/10.1016/0034-4257(92)90059-S).
37. Gitelson AA, Merzlyak MN. Remote sensing of chlorophyll concentration in higher plant leaves. *Adv Space Res*. 1998;22(5):689–92. [https://doi.org/10.1016/S0273-1177\(97\)01133-2](https://doi.org/10.1016/S0273-1177(97)01133-2).
38. Peñuelas J, Baret F, Filella I. Semiempirical indexes to assess carotenoids/chlorophyll-a ratio from leaf spectral reflectance. *Photosynthetica*. 1995;32(3):221–5. <https://doi.org/10.1007/BF00029464>.
39. Lichtenthaler HK. Chlorophylls and carotenoids: pigments of photosynthetic biomembranes. *Methods Enzymol*. 1987;148:350–82. [https://doi.org/10.1016/0076-6879\(87\)48036-1](https://doi.org/10.1016/0076-6879(87)48036-1).
40. Merzlyak MN, Gitelson AA, Chivkunova OB, Rakitin VY. Non-destructive optical detection of pigment changes during leaf senescence and fruit ripening. *Physiol Plant*. 1999;106(1):135–41. <https://doi.org/10.1034/j.1399-3054.1999.106119.x>.
41. Silva Neto EB, Silveira EB, Mariano RLR, Nogueira NL, Rossi ML, Santos LA. Penetração e colonização de *Acidovorax avenae* subsp. *citrulli* em folhas, frutos e sementes de melão amarelo. *Fitopatol Bras*. 2006;31(1):84–8. <https://doi.org/10.1590/S0100-41582006000100015>.
42. Dutta B, Avci U, Hahn MG, Walcott RR. Location of *Acidovorax citrulli* in infested watermelon seeds is influenced by the pathway of bacterial invasion. *Phytopathology*. 2012;102:461–8. <https://doi.org/10.1094/PHYTO-10-11-0286-R>.

Figures

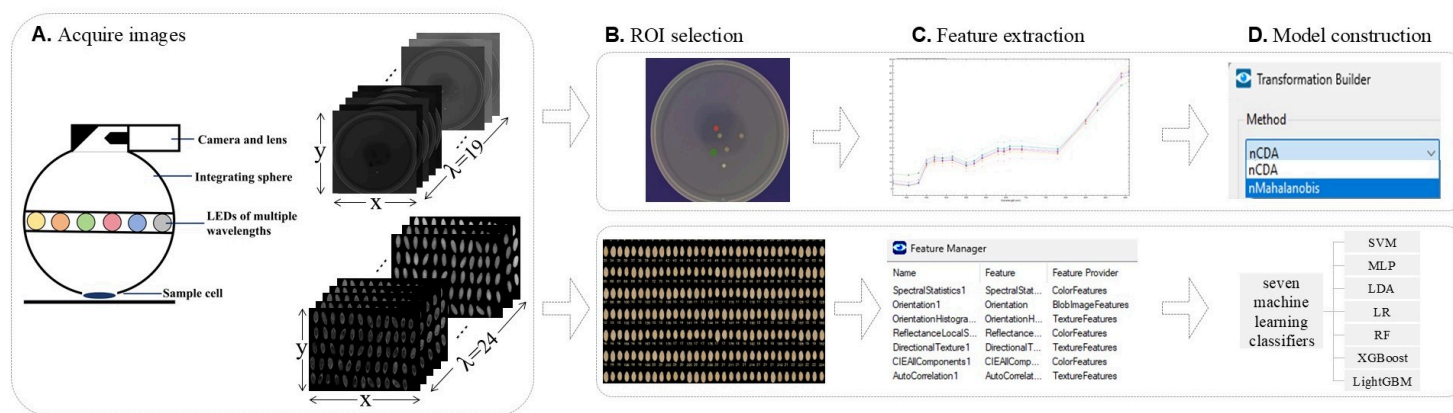


Figure 1

Figure 1

Flowchart of the multispectral imaging-based detection procedure.

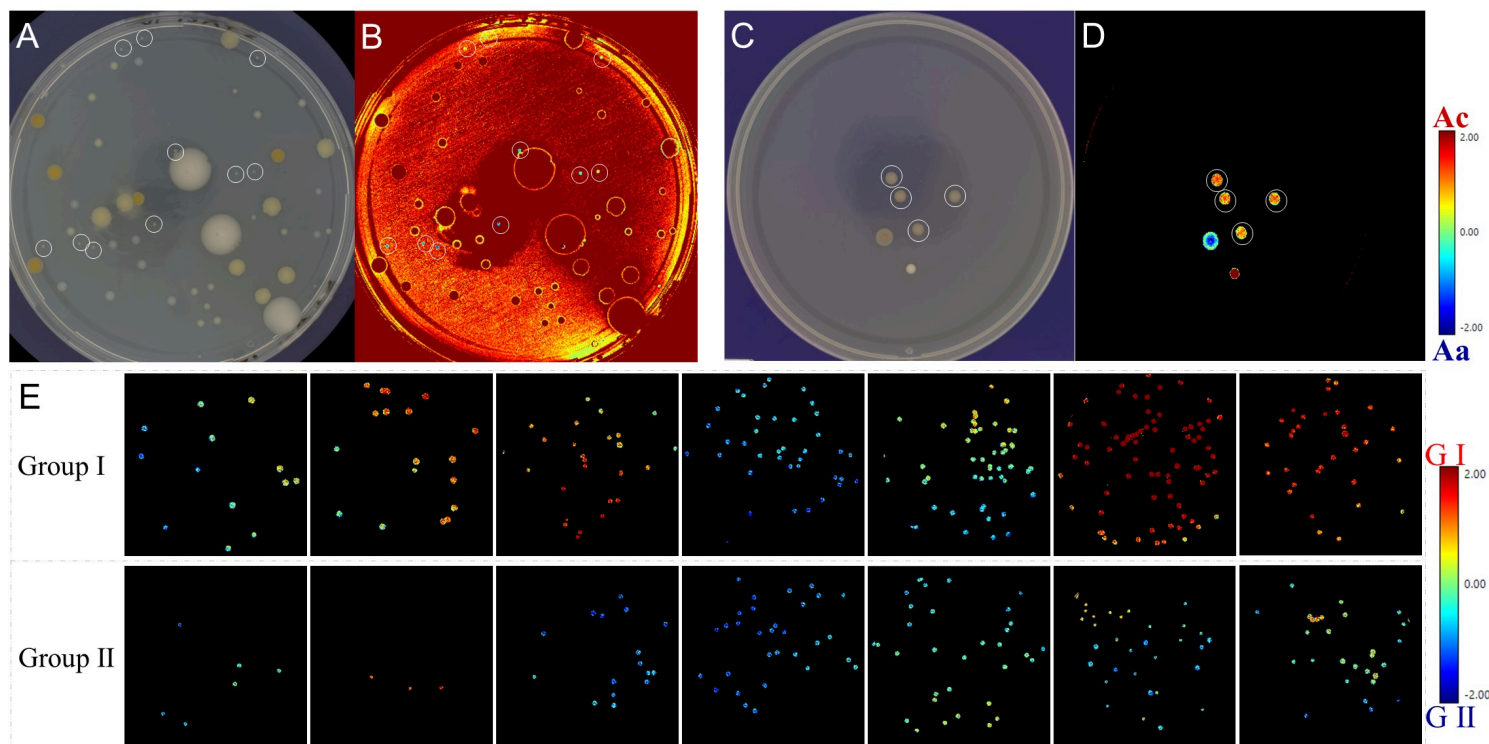


Figure 2

Figure 2

Feasibility evaluation of *A. citrullicolony* identification using the nMahalanobis and nCDA methods. (A) Mixed bacterial culture plate containing *A. citrulli*, *Pseudomonas syringae* pv. *lachrymans*, and five other bacterial species isolated from watermelon seed. White circles indicate true *A. citrulli* colonies. (B) Identification results of

the mixed bacterial culture plate using the nMahalanobis method. Target *A. citrulli* colonies are highlighted in blue and indicated by white circles. **(C)** Co-culture plate containing *A. citrulli* and *A. avenae*. White circles indicate *A. citrulli* colonies. **(D)** Identification results of *A. citrulli* and *A. avenae* obtained using the nCDA method. Target *A. citrulli* colonies are highlighted in red and indicated by white circles. **(E)** Discrimination of the two genetic groups of *A. citrulli* using the nCDA method. When the two groups are correctly distinguished, Group I colonies are shown in red and Group II colonies are shown in blue.

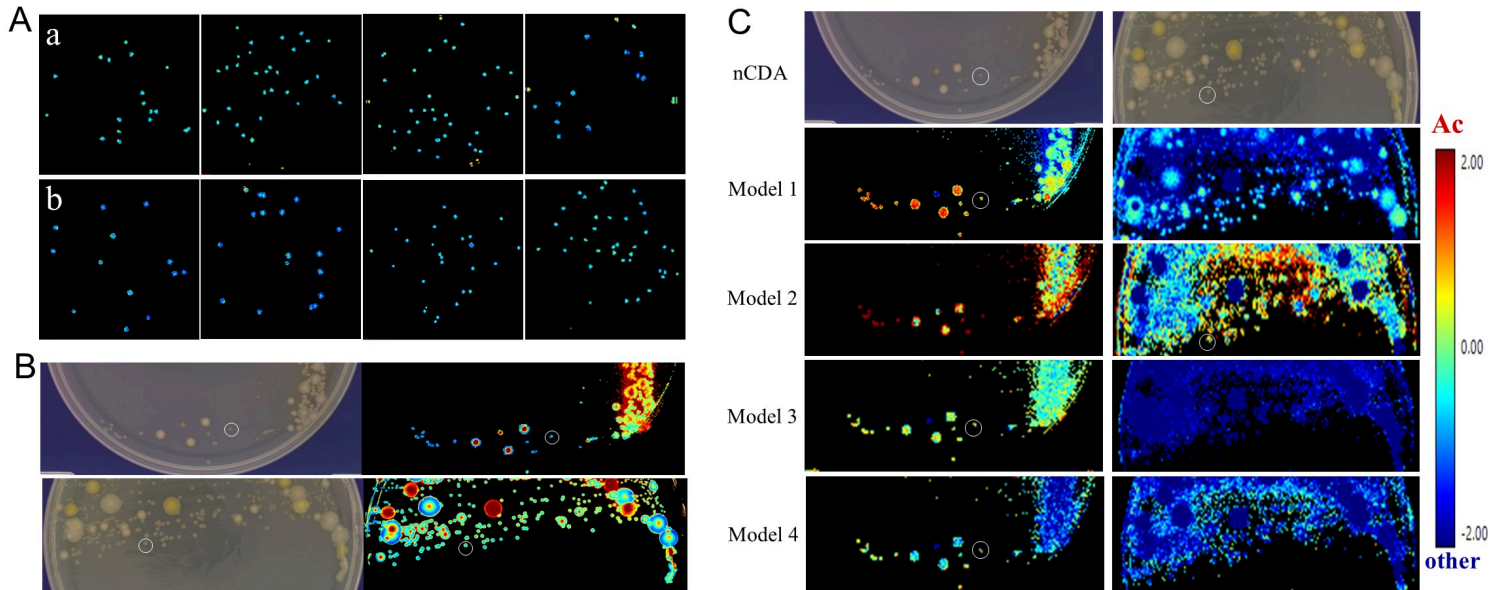


Figure 3

Figure 3

Colony identification of *A. citrulli* using the nMahalanobis and nCDA methods. **(A)** Identification results of the nMahalanobis discriminant model established using 12 representative *A. citrulli* strains from Group I and Group II, validating its discrimination capability between the two groups. (a) Identification results for Group I strains. (b) Identification results for Group II strains. **(B)** Colony identification results for bacterial culture plates from two batches of test samples using the nMahalanobis discriminant model established in VideometerLab software. The left panels show sRGB images of the culture plates, with white circles indicating PCR-confirmed *A. citrulli* colonies. The right panels show the corresponding identification maps generated by VideometerLab software analysis, where blue-highlighted regions represent the target colonies predicted by the model. **(C)** Colony identification results for bacterial culture plates from two batches of test samples using the nCDA discriminant model established in VideometerLab software. The left panels show sRGB images of the culture plates, with white circles indicating PCR-confirmed *A. citrulli* colonies. The right panels show the corresponding identification

maps generated by VideometerLab software analysis, where blue-highlighted regions represent the target colonies predicted by the model.

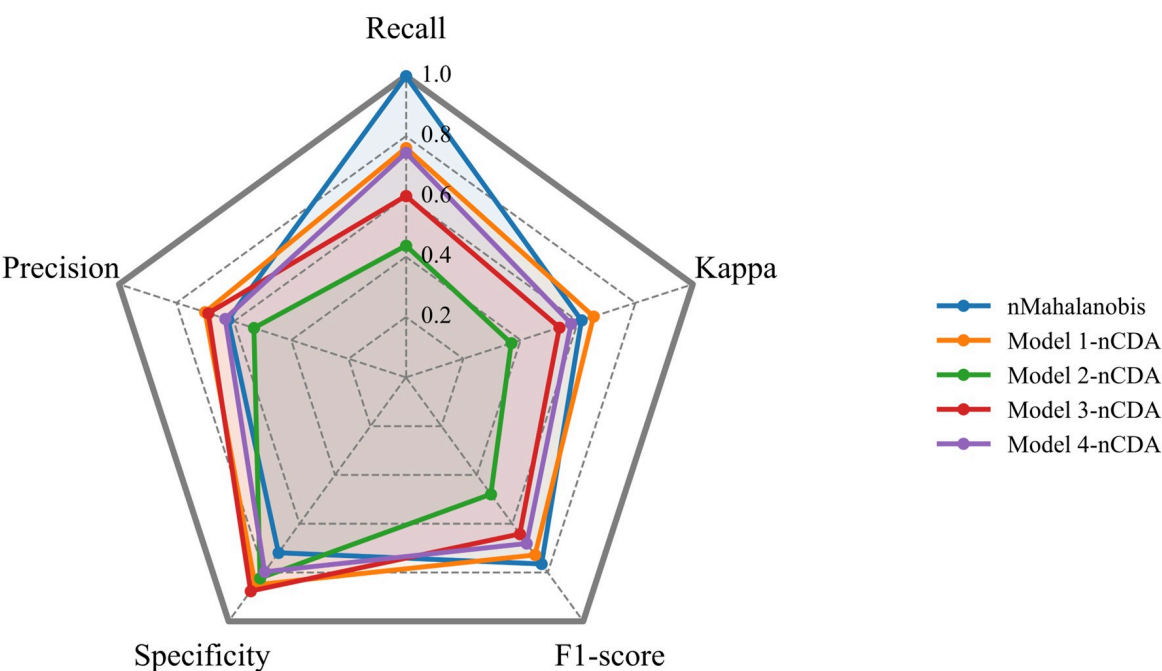


Figure 4

Figure 4
Radar chart comparing the performance of five models for identifying *Acidovorax citrulli* colonies on LB medium.

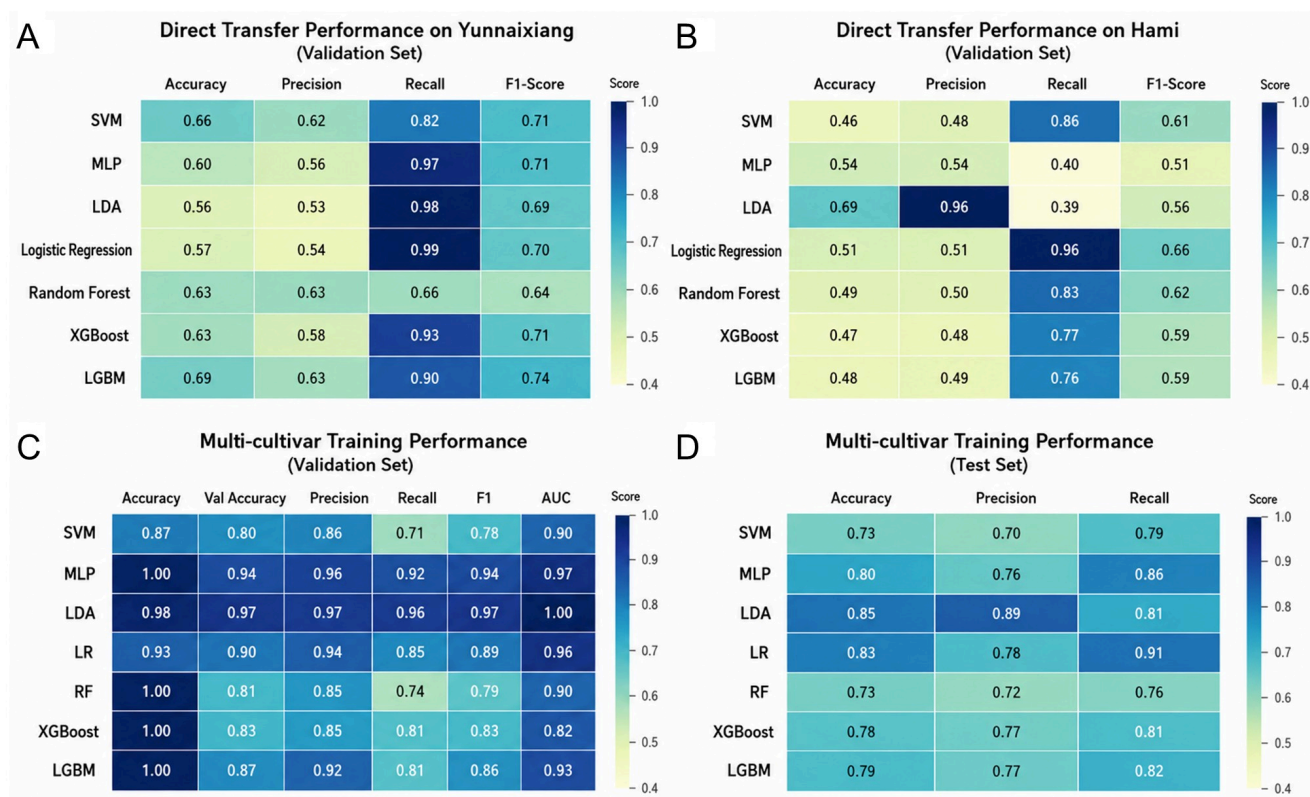


Figure 5

Figure 5

Comparison of direct transfer and multi-cultivar training strategies for the detection of *A. citrulli*-infested melon seeds. (A) Direct transfer performance of different classification models on Yunnaxiang seeds. (B) Direct transfer performance of different classification models on Hami seeds. (C) Validation performance of models developed using multi-cultivar training strategy. (D) Independent test performance of models developed using multi-cultivar training strategy.

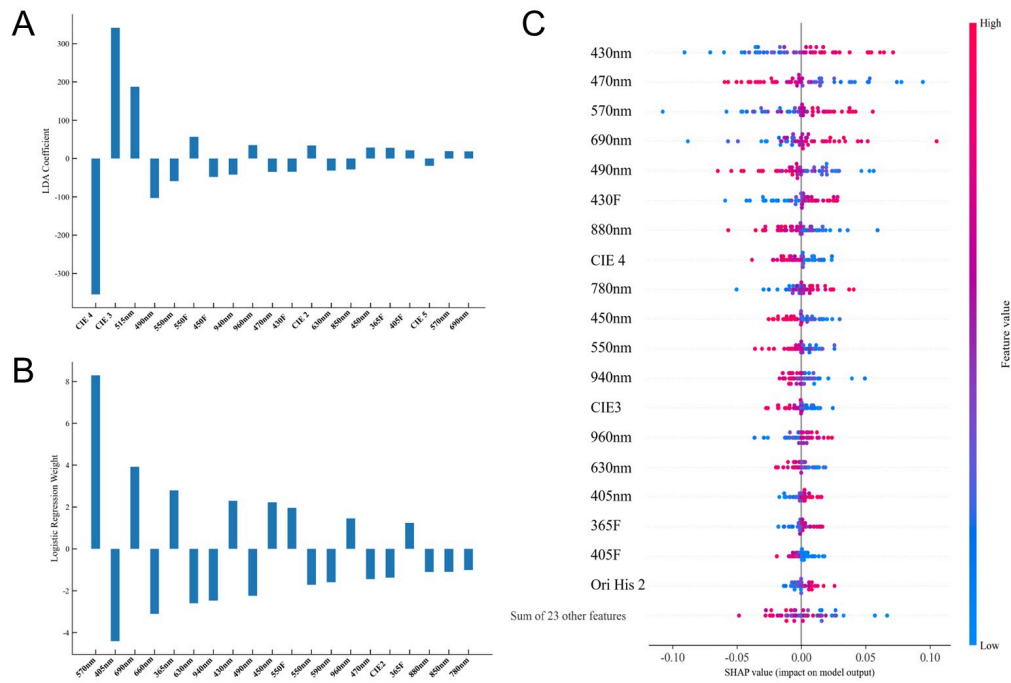


Figure 6

Figure 6

Feature importance analysis of three machine learning models for detection of *A. citrulli*-infested melon seeds. (A) Top 20 important features derived from the LDA model. (B) Top 20 important features derived from the logistic regression model. (C) SHAP summary plot illustrating the feature contributions of the MLP model.