



OPEN ACCESS

EDITED BY

Jiban Shrestha,
Nepal Agricultural Research Council, Nepal

REVIEWED BY

Yongpeng Zhao,
Sichuan Agricultural University, China
Renato Herrig Furlanetto,
University of Florida, United States

*CORRESPONDENCE

Jun Zhang
✉ 15931896691@163.com
Bo Peng
✉ bopeng@hebau.edu.cn

RECEIVED 27 February 2025

ACCEPTED 15 May 2025

PUBLISHED 03 June 2025

CITATION

Yang B, Liu X, Zhang D, Fan X, Peng B and Zhang J (2025) Optimized wavelength selection for eggplant seed vitality classification using information acquisition techniques.
Front. Plant Sci. 16:1584269.
doi: 10.3389/fpls.2025.1584269

COPYRIGHT

© 2025 Yang, Liu, Zhang, Fan, Peng and Zhang. This is an open-access article distributed under the terms of the [Creative Commons Attribution License \(CC BY\)](#). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

Optimized wavelength selection for eggplant seed vitality classification using information acquisition techniques

Bing Yang¹, Xuyang Liu², Dongfang Zhang^{3,4}, Xiaofei Fan^{1,3}, Bo Peng^{1*} and Jun Zhang^{1,3*}

¹College of Mechanical and Electrical Engineering, Hebei Agricultural University, Baoding, China,

²College of Economics and Management, Hebei Agricultural University, Baoding, China, ³State Key Laboratory of North China Crop Improvement and Regulation, Hebei Agricultural University, Baoding, China, ⁴College of Horticulture, Hebei Agricultural University, Baoding, China

Eggplant seed vigor is a crucial indicator of its germination rate and seedling growth quality. In response to the need for efficient and nondestructive assessment methods, this study explores the use of hyperspectral imaging combined with advanced feature selection and classification algorithms to evaluate eggplant seed viability. Hyperspectral imaging was employed to collect spectral data from eggplant seeds, covering 360 bands within a wavelength range of 395.24–1008.20 nm. The seeds underwent microwave heating and constant-temperature water bath aging treatments. Data preprocessing involved three techniques: Multiplicative Scatter Correction (MSC), Savitzky–Golay (SG) smoothing, and Standard Normal Variate (SNV) transformation. An Enhanced Information Acquisition Optimization (EIAO) algorithm was proposed for feature selection, which successfully identified a minimal set of 23 key wavelengths. Seed vigor classification models were developed using Extreme Learning Machine (ELM), Random Forest (RF), and Support Vector Machine (SVM). The optimal classification accuracies achieved were 90.0% for ELM, 91.45% for RF, and 90.5% for SVM. The MSC-EIAO-RF model demonstrated the best performance, achieving an accuracy of 91.45%, which is 9.04% higher than the MSC-IAO model (82.41%). Validation on four UCI datasets further confirmed the EIAO algorithm's superiority over conventional feature selection methods. These results verify the robustness and generalizability of hyperspectral imaging combined with EIAO for nondestructive seed viability detection, offering an intelligent and efficient solution for seed quality assessment.

KEYWORDS

information acquisition techniques, wavelength selection, hyperspectral, eggplant seed, vitality classification

1 Introduction

Eggplant is a perennial herb widely cultivated in warm tropical and subtropical regions worldwide. Due to its high mineral content and low caloric value, it is considered one of the healthiest fruits and vegetables (Abubakar et al., 2023; Asafew and Chandravanshi, 2021; Yamaguchi et al., 2019). Eggplant seeds are flat, shiny, and either reddish-black or yellow, with fine lines on the seed coat but no hairs, exhibiting strong vitality. Seed quality is commonly assessed based on germination potential and vigor (Xing et al., 2023). Among these, seed vigor is the most crucial indicator of seed quality, as it directly influences seedling germination rates and the overall health of the plant (Ventura et al., 2012). For seeds stored over extended periods, especially those used for breeding or conserving genetic diversity, seed vigor remains a prominent research focus (De Vitis et al., 2020). Even under optimal storage conditions, seed viability inevitably declines over time, a phenomenon referred to as seed aging (Ebene et al., 2019). Traditional methods for assessing seed vigor, such as conductivity and red ink tests, are often complex, inefficient, and somewhat destructive. These methods play a critical role in determining the quality grade and shelf life of seeds, making them essential for evaluating seed vitality and longevity (Wang et al., 2021a). Therefore, developing more advanced and innovative testing methods to enhance the efficiency of existing seed testing technologies has become a critical area of research.

In recent years, Hyperspectral imaging (HSI) technology has garnered widespread attention in seed quality assessment due to its non-destructive and rapid characteristics. By providing spatial and spectral information related to plant and biochemistry, the technology has excellent capabilities in seed variety classification and grading, seed viability and damage detection, and seed composition determination (Feng et al., 2019). In seed classification, researchers have combined hyperspectral imaging technology in combination with the firefly optimization algorithm to optimize deep learning parameters, successfully achieving vitality detection of sweet corn seeds (Wang and Song, 2024). Huang et al. employed HSI to classify corn seeds from four different years and developed a classification model based on the average spectral features of the seeds, utilizing the least squares support vector machine (LSSVM) (Huang et al., 2016). Wang et al. employed near-infrared hyperspectral imaging (NIR-HSI) technology to study the maturity classification of corn seeds. They extracted the average spectra from the embryo side (T1) and the endosperm side (T2) of the seeds, and calculated the average spectrum of both sides (T3). Principal component analysis (PCA) was applied to select the characteristic wavelengths. When T1 and T2 were used as inputs for the optimal model, the classification accuracy reached 98.7% and 100%, respectively (Wang et al., 2021b). In classifying camellia oil grades, researchers collected hyperspectral images of camellia oil samples from three different grades. The successive projections algorithm (SPA) and competitive adaptive reweighted sampling (CARS) were employed to extract spectral and texture features. Subsequently, the genetic algorithm (GA) was used to optimize the kernel function of the support vector machine (SVM), along with its corresponding kernel function parameters and penalty factors. The results demonstrated that the model classification performance was best when GA was used to

optimize the SVM (Gu et al., 2024). These studies highlight the significance and potential of hyperspectral imaging (HSI) in seed vigor detection. By providing spatial and spectral information related to plants and their biochemical characteristics, HSI holds great promise for the non-destructive assessment of seed vigor and other key quality parameters.

However, wavelength selection remains a critical challenge in applying hyperspectral imaging technology to seed identification (Huang et al., 2024; Huang and Xia, 2023). Selecting characteristic wavelengths can enhance data processing efficiency, highlight important features, and improve model robustness (Huang et al., 2022; Ong et al., 2023). In the classification of rapeseed maturity, the continuous projection algorithm (SPA), competitive adaptive reweighted sampling (CARS), and interval variable iterative spatial shrinkage (IVISSA) are combined to select spectral wavelengths. The results indicate that the algorithm combining SPA and IVISSA achieves an accuracy of 97.86% (Feng et al., 2024). In addition to conventional characteristic wavelength selection methods, biologically inspired meta-heuristic algorithms have emerged as effective tools for hyperspectral band selection. A representative application can be found in the research on chlorophyll content prediction in Chinese cabbage using hyperspectral technology. In this study, the reflectance data underwent comprehensive preprocessing through standard normal variate (SNV) transformation, Savitzky-Golay (SG) smoothing, and second derivative (2D) analysis. Subsequently, a genetic algorithm (GA) was employed to identify optimal spectral characteristic bands, which were then utilized to construct a sophisticated one-dimensional convolutional neural network (1D-CNN) prediction model (Zhang et al., 2023). For the identification of wheat grain varieties, the interval random frog (iRF) algorithm, an advanced wavelength selection method, was implemented to optimize the spectral wavelength intervals through an iterative jumping mechanism (Que et al., 2023). Although existing studies have proposed several effective wavelength selection methods, such as the integration of convolutional neural network architectures with traditional feature selection algorithms (e.g., SPA, CARS, and IVISSA), these methods have demonstrated significant improvements in classification accuracy. However, they still exhibit certain limitations.

Traditional methods may fail to fully capture key features in spectral data and tend to be slower when handling large datasets with many redundant features. To address these challenges, this study presents a novel wavelength selection algorithm (EIAO) based on information acquisition optimization, designed to enhance the efficiency of spectral data processing and improve classification accuracy. By processing and optimizing the hyperspectral data of eggplant seeds, the EIAO algorithm effectively selects the most informative spectral bands, offering an innovative and efficient solution for the non-destructive detection of seed vigor.

2 Materials and methods

2.1 Data collection

In the experiment, four varieties of eggplant with complete shape and uniform size were selected for artificial aging treatment, with

three experimental groups set up (Figure 1B), namely the microwave group, the water bath group, and the control group with no treatment. Each aging treatment was repeated three times to eliminate the influence of seed variability and treatment-induced effects, ensuring that no visible changes appeared on the seed surface. After treatment, the seeds were air-dried at room temperature to restore their original weight. The experiment utilized an HG101 portable hyperspectral imaging system (Zhongchuan Optoelectronics Precision Machinery Co., Ltd., Beijing), as shown in Figure 1A. Hyperspectral data were collected from these eggplant seed samples using a hyperspectral camera, which covers a spectral range of 395–1008 nm with 360 bands, and includes components such as a halogen lamp, a sample board, and a light shield. To ensure clear and accurate imaging, the camera lens-to-sample distance was fixed at 30 cm, and the exposure time was set to 15 milliseconds. Prior to capturing hyperspectral images, a whiteboard calibration was performed to correct for any potential lighting inconsistencies. After imaging, each seed, categorized by its aging level, was sealed in a bag and numbered for subsequent identification and analysis of seed vigor. Following these procedures, the spectral images of the samples were successfully captured and stored.

2.2 Software and performance evaluation

This study used ENVI (version 5.6, Harris Corporation, USA) to extract the reflectance of eggplant seeds and MATLAB (version 2023a) for characteristic wavelength selection and classification modeling tasks. The dataset was divided into an 8:2 ratio, resulting in 800 training samples and 200 test samples. When analyzing the results of data dimensionality reduction, the fitness score defined in Eq. (5) was used to evaluate the quality of potential optimal feature subsets. Accuracy represents the classification performance of the feature subset on the classifier, while the

number of dimensions of the feature subset indicates how many important features contribute to the classification.

In the EIAO framework, the iteration count K was set to 50. The parameter settings for other comparison algorithms are provided in Table 1. These parameters are essential for guiding the algorithm's search process, balancing exploration and exploitation, and ensuring efficient convergence toward the optimal solution.

2.3 Data analysis methods

2.3.1 Data preprocessing

Spectral data often contains noise from the environment and instruments. Preprocessing the spectral data can enhance the accuracy of the model (Chen et al., 2024). In this study, we applied three preprocessing methods: MSC, SG, and SNV.

2.3.2 Feature extraction

Hyperspectral data often contains redundant information due to the large number of spectral bands. Data dimensionality reduction can help select effective characteristic wavelengths, thereby reducing model computation and improving operational efficiency. In this study, the Enhanced Information Acquisition Optimization (EIAO) algorithm is used to extract characteristic wavelengths, with improvements made to the original Information Acquisition Optimization algorithm. A detailed explanation of the algorithm is provided below.

The original Information Acquisition Optimization Algorithm (IAO) algorithm was recently proposed (Wu et al., 2024). The basic idea is to simulate the ability of humans to process massive amounts of information. It includes three steps: collecting information, filtering and evaluating information, and analyzing and organizing information. When EIAO is applied to the feature selection problem,

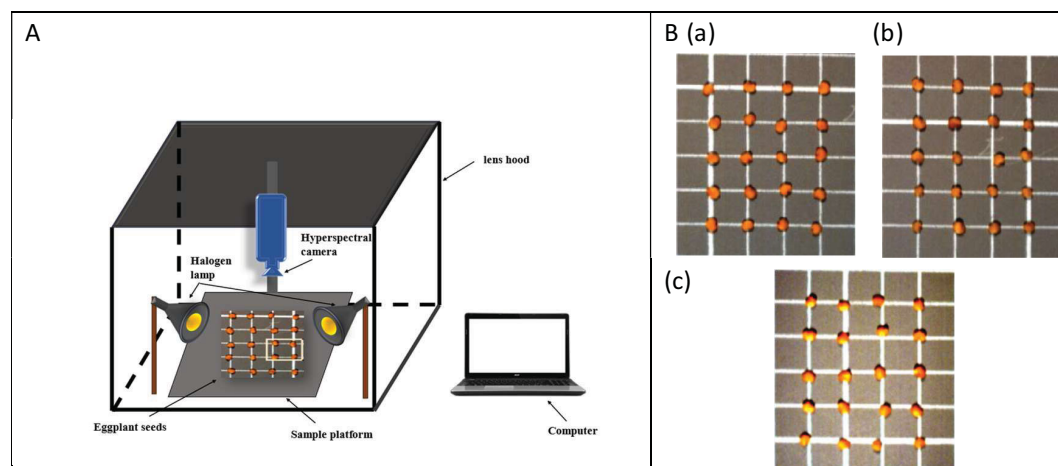


FIGURE 1

(A) Hyperspectral data acquisition system. (B) Artificially aged eggplant seed samples (a) control group without any treatment (b) water bath group (c) microwave group.

TABLE 1 Parameter setting.

Algorithms	Parameter values
EIAO	v, β, γ , and δ random numbers over $[0,1]$ are set as the literature [15]
IAO	v, β, γ , and δ random numbers over $[0,1]$ are set as the literature [15]
PSO	$c1=c2=2, \omega=1$
CSA	$\alpha=\delta=0.1, r1=10, U=0.00565, \omega=0.005$
SO	$Q=0.25, T=0.6, c1=0.5, c2=0.05, c3=2$
SCA	$a=2$

We define a population of N human individuals as a two-dimensional matrix $X = \{x_1, \dots, x_i, \dots, x_N\} = \{x_j^i | 1 \leq i \leq N, 1 \leq j \leq D\}$. x_j^1 represents the individual at the first position in the j th dimension. x_j^i represents the i -th individual in the j dimension. When IAO is used for feature selection problems, all solutions are restricted to binary values. That is, $x_j^i \in [0,1]$. At this time, x_j^i represents the j -th feature in the i -th seed sample. If $x_j^i=1$, it means that the feature is selected. The feature selection framework based on the EIAO algorithm is shown in Figure 2.

As shown in Figure 3, the first step of EIAO is population initialization. Building on the initialization method of the original

algorithm, this study selects 10% of the population size as the number of elite individuals and applies chaotic reverse learning to the selected elites. Elite chaotic reverse learning is an initialization technique that combines elite strategies, chaotic disturbances, and reverse learning. It is primarily used to enhance the diversity and exploration ability of the population.

The expression for the initial state is shown in Equation 1.

$$x_j^i(k) = (ub_j(k) - lb_j(k)) * \theta^k + lb_j(k) \quad (1)$$

The expression of logistic mapping is

$$\theta^{k+1} = \mu \theta^k (1 - \theta^k) \quad (2)$$

According to Equation 2, chaos mapping is a nonlinear dynamic system that generates random numbers with unique dynamic characteristics. These random numbers are distinguished by their non-repeatability, accessibility, regularity, and unpredictability (Zhang and Feng, 2018). The reverse learning strategy is an intelligent calculation method introduced (Tizhoosh, 2005) in 2005. It has been widely applied in other group-based intelligent optimization algorithms to enhance their search performance. Building on the effectiveness of reverse learning strategies, Wang et al. introduced general reverse factors and proposed the concept of general reverse learning strategies (Wang et al., 2011). Li et al. introduced the concept of elite learning and proposed an elite reverse learning strategy based on general reverse learning strategies. Experimental results demonstrated that the elite reverse learning strategy outperforms the general reverse learning strategies (Li, 2012).

Since the band selection of the spectrum belongs to the feature selection of binary problems, it is necessary to convert the EIAO continuous form into the EIAO-FS binary discrete form. Thus converted from Equation 1 to Equation 3. The expression of T in Equation 3 is as shown in Equation 4.

$$x_j^i(k) = \begin{cases} 1, & T(x_j^i(k)) \geq C \\ 0, & T(x_j^i(k)) < C \end{cases} \quad (3)$$

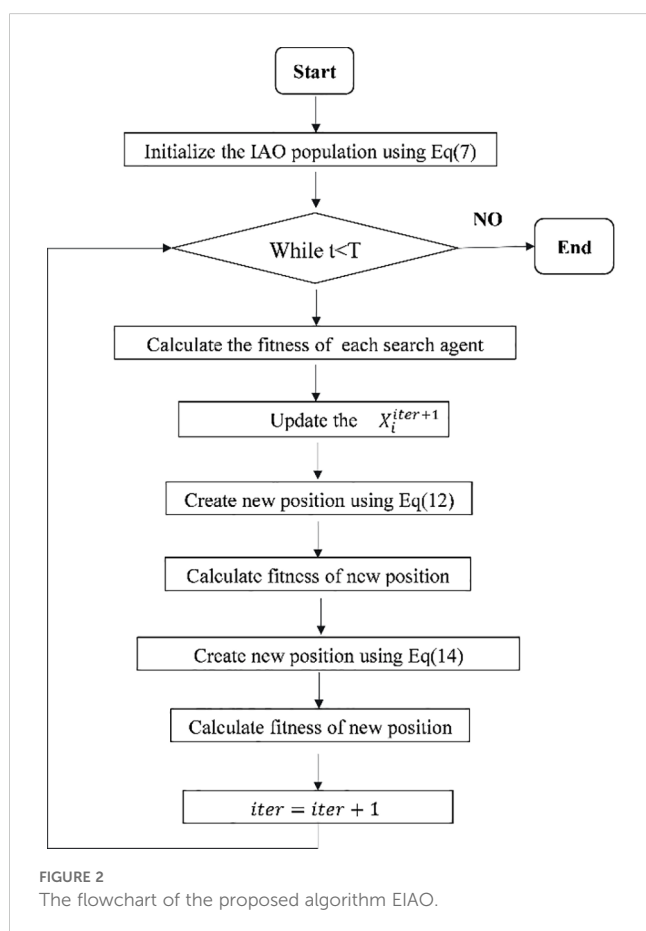
$$T(x_j^i(k)) = \frac{1}{1 + \exp^{-x_j^i(k)}} \quad (4)$$

Fitness calculation

$$fitness = \alpha \rho_R(Y) + \beta \frac{|K|}{|N|} \quad (5)$$

As shown in Equation 5, among them, $\rho_R(Y)$ represents the error rate of the KNN classifier, and α and β meet the $\alpha \in [0, 1]$ and $\alpha + \beta = 1$, respectively. $|K|$ is the number of selected features, $|N|$ is the number of data concentration original features. Under normal circumstances, the smaller the adaptation value, the better.

K-NN is a non-parametric supervised classification learning algorithm. The category of a new sample is determined by the K nearest training samples to the new sample (Pernkopf, 2005). Here, the K-NN model is used as a classification method to evaluate the features generated by EIAO, where $K = 3$.



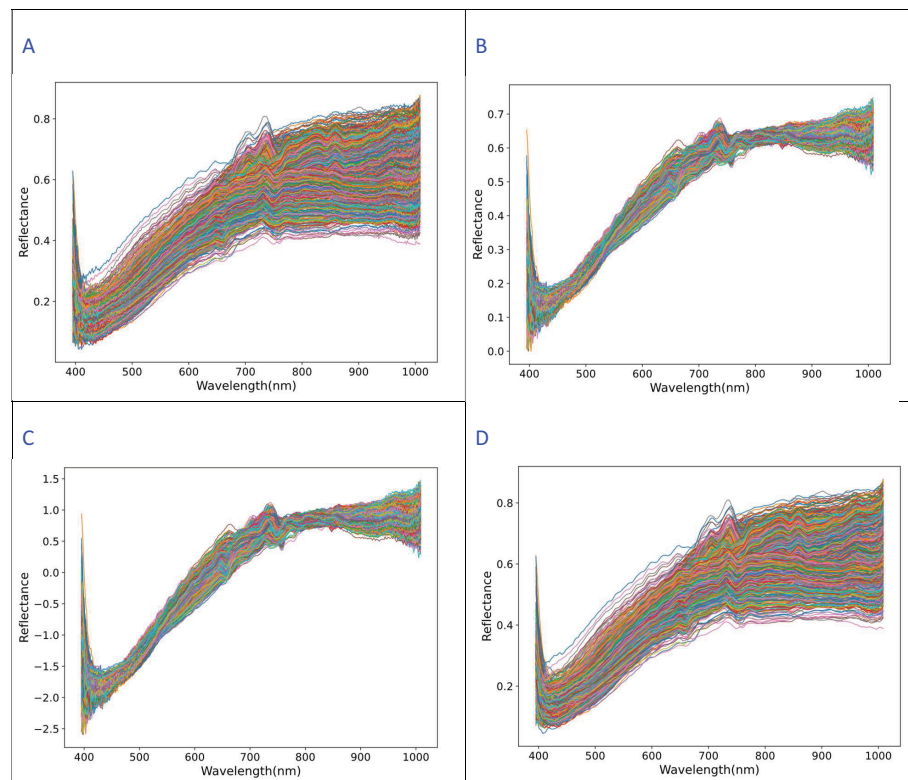


FIGURE 3
Hyperspectral reflectance and preprocessing. (A) The raw average reflectance (B) MSC preprocessing (C) SNV preprocessing (D) SG smoothing Preprocessing image.

The first stage involves collecting information and establishing an initial information system. The mathematical model is expressed as Equation 6.

$$x_i^{iter+1} = x_i^{iter} + \vartheta \times (x_i^{r1} - x_i^{r2}) \quad (6)$$

Where $iter$ represents the current iteration number, ϑ is a random number between $[0,1]$, and x_i^{r1} and x_i^{r2} are two information bodies randomly generated during iteration. Consider $(x_i^{r1} - x_i^{r2})$ as a differential variable, and refer to DE/rand/2/bin to add another differential variable $(x_i^{r3} - x_i^{r4})$. At this time, the information collection phase is updated through Eq (7). By adding more information bodies, more randomness and diversity are introduced during the update, thereby improving the breadth of the search. It helps to avoid premature convergence to the local optimal solution. Multiple information bodies can make the exploration of the search space more comprehensive, thereby increasing the diversity of the population. The updated first-stage mathematical model is as shown in Equation 7.

$$x_i^{iter+1} = x_i^{iter} + \vartheta \times (x_i^{r1} - x_i^{r2}) + \vartheta \times (x_i^{r3} - x_i^{r4}) \quad (7)$$

The second stage is information filtering evaluation, the expression is as in Equation 8:

$$\begin{cases} x_i^{iter+1} = x_i^{iter} - \Delta \times rand \times (x_i^{rand} - x_i^{iter}), & \text{if } rand < 0.5 \\ x_i^{iter+1} = x_i^{iter} + \Delta \times rand \times (x_i^{rand} - x_i^{iter}), & \text{otherwise} \end{cases} \quad (8)$$

Among them, $\Delta = \frac{\cos(\frac{\pi}{2} \times \sqrt{|r|})}{\Xi}$ represents the error caused by subjective factors in filtering and evaluating information, and Ξ is the subjective influencing factor, which reflects the individual's less correct judgment on information processing due to preferences, experience, emotions, etc. Its value is calculated by the following Equation 9:

$$\Xi = 2 \times \text{mod}(3.648 \times v \times (1 - \beta) \times (a \cos(\gamma \times 10^4))), 1) \quad (9)$$

v , β , and γ are random numbers generated between $[0, 1]$. Γ is defined as a reliability factor, which characterizes the ability of the algorithm to optimize its behavior by self-adjusting the quality of information at different stages. This design enhances the adaptability and flexibility of the algorithm. Γ is calculated by the following Equation 10.

$$\Gamma = \sin\left(\left(\frac{\pi}{4}\right)^{\frac{iter}{Max_{iter}}}\right) + \varnothing + \frac{\log_{10} \frac{iter}{Max_{iter}}}{8} \quad (10)$$

where Φ is the information quality factor, the expression is as in Equation 11.

$$\varnothing = \cos(2 \times \delta + 1) \times \left(1 - \frac{\text{iter}}{\text{Max}_{\text{iter}}}\right) \quad (11)$$

From Eq (8)-Eq (11), it can be seen that the iterative update is mainly determined by the difference between the current individual state x_i^{iter} and the randomly selected reference point x_i^{rand} . Although this update method introduces randomness, its search range and direction are limited and lack other dynamic adjustment mechanisms. Therefore, in complex multi-peak functions, it may not be able to effectively jump out of the local optimum. To avoid the above problems, we introduce the sine-cosine optimization algorithm (SCA) and update by the following formula (Equation 12). $r5$ is a random number between $[0, 2\pi]$.

$$\begin{cases} x_i^{\text{iter}+1} = x_i^{\text{iter}} - \Delta \times \sin(r5) \times \text{rand} \times (x_i^{\text{rand}} - x_i^{\text{iter}}), & \text{if } \text{rand} < 0.5 \\ x_i^{\text{iter}+1} = x_i^{\text{iter}} + \Delta \times \cos(r5) \times \text{rand} \times (x_i^{\text{rand}} - x_i^{\text{iter}}), & \text{otherwise} \end{cases} \quad (12)$$

The final stage is to identify valuable information from the filtered data and increase the likelihood of obtaining the optimal information set. The mathematical model is expressed as Equation 13.

$$\begin{cases} x_i^{\text{iter}+1} = x_i^{\text{best}} \times \cos\left(\frac{\pi}{2} \times \sqrt{\Lambda^{\frac{1}{3}}}\right) - \varepsilon \times \left(\frac{1}{D} \sum_{i=1}^D x_i^{\text{best}} - x_i^{\text{best}}\right) & \text{if } \varnothing \geq 0.5 \\ x_i^{\text{iter}+1} = x_i^{\text{best}} \times \cos\left(\frac{\pi}{2} \times \sqrt{\Lambda^{\frac{1}{3}}}\right) - 0.8 \times \left(\frac{\zeta \times \kappa \times \frac{1}{D} \sum_{i=1}^D x_i^{\text{best}} -}{(2 \times \omega - 1) \times x_i^{\text{best}}}\right), & \text{otherwise} \end{cases} \quad (13)$$

Where $\Lambda = 2(\sqrt{|f|}^{-2})$, represents the control factor for analyzing and organizing information. The update formula of Eq (13) mainly relies on the local neighborhood search of the current x_i^{best} optimal solution, combined with some dynamic adjustment factors. Although it can speed up the convergence, it may easily fall into the local optimum due to the lack of long-distance jumps. The addition of Levy flight allows individuals to make long-distance jumps with a certain probability. This jump helps to discover new potential optimal solutions, thereby balancing the ability of global exploration and local development. Updated by Eq (14).

$$\begin{cases} x_i^{\text{iter}+1} = x_i^{\text{best}} \times \cos\left(\frac{\pi}{2} \times \sqrt{\Lambda^{\frac{1}{3}}}\right) - \varepsilon \times \left(\frac{1}{D} \sum_{i=1}^D x_i^{\text{best}} - x_i^{\text{best}}\right) + \text{Levy}, & \text{if } \varnothing \geq 0.5 \\ x_i^{\text{iter}+1} = x_i^{\text{best}} \times \cos\left(\frac{\pi}{2} \times \sqrt{\Lambda^{\frac{1}{3}}}\right) - 0.8 \times \left(\frac{\zeta \times \kappa \times \frac{1}{D} \sum_{i=1}^D x_i^{\text{best}} -}{(2 \times \omega - 1) \times x_i^{\text{best}}}\right) + \text{Levy}, & \text{otherwise} \end{cases} \quad (14)$$

Levy flight is a random walk mechanism (Li et al., 2022) that enables large jumps from a local position with a high probability. The probability density distribution of Levy flight is characterized by sharp peaks, asymmetry, and heavy tails. Its movement pattern alternates between frequent short-distance jumps and occasional long-distance leaps, allowing it to escape local optima and expand the search area. Metaheuristic algorithms, inspired by natural processes, are used to solve NP-hard problems. Levy flight can serve as an operator in these metaheuristic search algorithms.

At the end of the iteration, we try to add the Laplace operator. Deep and Thakur proposed the original Laplace operator idea in 2007 (Deep and Thakur, 2007), the idea is to generate children $y_1 = (y_1^1, y_1^2, \dots, y_1^m)$ and $y_2 = (y_2^1, y_2^2, \dots, y_2^m)$ from the parent $x_1 = (x_1^1, x_1^2,$

$\dots, x_1^m)$, and the two pairs of children are generated in a symmetric way around the position of their parents.

The random distribution of Laplace $l_i X$ is as Equation 15:

$$l_i = \begin{cases} p - q \log_e(u_i), & u_i \leq \frac{1}{2} \\ p + q \log_e(u_i), & u_i > \frac{1}{2} \end{cases} \quad (15)$$

Among them, u_i is randomly and uniformly distributed on $[0, 1]$; $p, q > 0$ represent the position and measurement parameters respectively. The resulting descendant relationship expression is as shown in Equation 16:

$$y_1^i = x_1^i + l_i |x_1^i - x_2^i|, y_2^i = x_2^i + l_i |x_1^i - x_2^i| \quad (16)$$

As an effective local search technique, some researchers have combined Levy flight (LX) with the Harris Hawk Optimization (HHO) algorithm, demonstrating its unique and superior performance on certain complex optimization problems (Nasab and Abualigah, 2024). This shows us the potential of this local search.

2.3.3 Performance evaluation

After obtaining the optimal feature subset using the EIAO algorithm, we constructed three classification models to evaluate their classification performance.

The extreme learning machine (ELM) is an advanced one-way feedback neural network algorithm based on a feedforward neural network (Huang et al., 2006). When using the ELM to build the seed activity discrimination model, the activation function employed is the S-shaped function. The number of neurons in the extreme learning machine is set within the range of 30 to 100, with a step size of 10. Simultaneously, the number of hidden layer neurons is adjusted to determine the optimal configuration for different spectral data. ELM exhibits strong generalization ability when handling high-dimensional data and is particularly well-suited for large-scale classification tasks. Therefore, in this study, ELM was chosen to develop the seed vigor classification model, aiming to enhance both classification accuracy and processing efficiency.

RF is a decision tree ensemble model based on the bagging strategy (Breiman, 2001). RF integrates multiple decision trees, each of which is trained on the original training set. The final classification result is obtained by voting among all decision trees (Si et al., 2023). In this study, the parameter “N” of the bagging framework is set to 300, the maximum depth “M” of the decision tree ranges from 1 to 20, and the step size of the grid search is 1. RF is effective in handling datasets with high noise and complex patterns, and it can automatically assess feature importance, offering enhanced interpretability. In this study, RF was employed to evaluate the stability of seed vigor classification, particularly in addressing noise and variations in spectral data.

SVM is a classic supervised machine learning model capable of classifying both linear and nonlinear data. It is a well-established binary classification model in supervised learning (Picciulli et al., 2024). It is widely used across various fields. In this study, the SVM algorithm utilizes the radial basis function (RBF) as the kernel function, with the penalty factor ‘C’ and kernel parameter ‘G’ optimized through 5-fold cross-validation.

2.3.4 Verification of the generalization capability

This section aims to evaluate the generalization capability of the EIAO algorithm. To this end, EIAO, along with five other optimization algorithms, is tested on four datasets from the UCI Machine Learning Repository. The UCI Repository, created by the University of California, Irvine, is a widely recognized collection of machine learning datasets. It includes a diverse range of datasets designed for various machine learning tasks such as classification, regression, and clustering. Each dataset typically provides details such as the dataset name, features, labels, descriptive documentation, and data formats. The Ionosphere dataset was collected using a ground-based radar system; The Arrhythmia dataset was obtained using electrocardiogram sensors that recorded cardiac signals; The Vehicle dataset contains features extracted from vehicle images using image processing techniques; The Vote dataset is based on survey records of U.S. Congressional voting behavior. These datasets are all commonly used for classification tasks. For this study, four classification datasets are selected, as summarized in Table 2.

3 Results

3.1 Hyperspectral data preprocessing analysis

The resulting processed spectral curve is shown in Figure 3. The hyperspectral reflectance curve of eggplant seeds shows a clear upward trend in the 500–700 nm range, while the reflectance in the near-infrared band remains relatively flat, with high reflectance values. This is attributed to the presence of water and other tissue components in the seeds, which strongly reflect near-infrared light. The reflectance characteristics of eggplant seeds vary in this range depending on the treatment. Around 750 nm, the curve fluctuates, initially rising and then falling, due to differences in seed water content (Figure 3A). The changes in light source from the two halogen lamps in the hyperspectral equipment and variations in instrument response cause surface scattering of the eggplant seed samples. To correct for these effects, the multivariate scatter correction (MSC) method is applied (Figure 3B). Since the hyperspectral data spans a large range, the standard normal variate (SNV) method is used to eliminate large gaps in the data (Figure 3C). The influence of the external environment introduces errors in the original hyperspectral reflectance, which are corrected by the Savitzky-Golay (SG) smoothing filter to remove irrelevant signals and smooth the data (Figure 3D).

TABLE 2 UCI datasets used in this study.

Dataset	Number of Samples	Number of Features	Class
Ionosphere	351	34	2
Arrhythmia	452	279	16
vehicle	846	18	4
vote	435	16	2

3.2 Feature extraction result analysis

3.2.1 Comparative performance of EIAO and baseline algorithms

In this study, the Enhanced Information Acquisition Optimization (EIAO) algorithm was used to extract feature bands from the collected seed hyperspectral data. PSO (Kocak and Orkcu, 2024), Chameleon Optimization Algorithm (CSA) (Braik et al., 2023a), Snake Optimization Algorithm (SO) (Braik et al., 2023b), Sin-Cosine Optimization Algorithm (SCA) (Sun et al., 2022) and the standard IAO, as outlined in Table 1, are employed as baseline algorithms for comparative analysis to effectively demonstrate the overall advantages of the proposed method. Based on the fitness function in Eq. (5) from section 2.3.2, we applied the KNN model as a classification method to evaluate the quality of the feature subsets generated by each algorithm. The evaluation was carried out using the classification accuracy of the KNN model and the number of selected features.

To validate the feature selection capability of the EIAO algorithm, it was applied to four different eggplant varieties (DS_1 to DS_4) for comparison. As shown in Table 3, EIAO achieves the highest classification accuracy on the DS_4 dataset and selects the fewest features on DS_3. For example, in DS_4, EIAO only needs 70 features to achieve an accuracy of 97.56%, while IAO requires 90 features. In addition, the fitness function value of EIAO is significantly lower than that of other algorithms, indicating its excellent ability in balancing classification accuracy and feature redundancy. This result confirms that EIAO is highly efficient in feature selection, effectively eliminating redundant bands while maintaining excellent classification performance.

At the same time, from Table 4, we can analyze which spectral regions of the full wavelength the selected features are mainly concentrated in. Among the selected wavelength features, the 905.70–1008.20 nm range accounted for the highest proportion (23.17%), followed by 694.92–798.40 nm (16.52%) and 490.95–589.60 nm (15.94%). These frequently selected regions were mainly located in the near-infrared range, particularly in the short-wave near-infrared region above 900 nm. This spectral region is typically associated with absorption peaks of functional groups such as O–H, N–H, and C–H, which correspond to key seed components like moisture, proteins, and starch. These components are known to change during seed aging treatments, providing a clear biochemical basis for the frequent selection of this range as effective feature wavelengths. Overall, the selected wavelengths were primarily concentrated in the information-rich near-infrared region, highlighting both the effectiveness of the feature extraction algorithm and the physical relevance of the selected features.

3.2.2 Robustness analysis of EIAO algorithm

To assess the robustness of the proposed EIAO method, we further evaluated it from two aspects. On one hand, we examined whether the convergence curve of EIAO remains stable across multiple repetitions. From Figure 4, we observe that the shaded region for DS_1 and DS_3 shows minimal fluctuations. The shaded area represents the standard deviation, which indicates the range of

TABLE 3 Performance comparisons between EIAO and 5 baseline algorithms on Seed Spectrum datasets.

Metric	Algorithm	DS_1	DS_2	DS_3	DS_4
Accuracy (%)	IAO	90.50	85.71	72.22	95.12
	EIAO	92.00	85.71	73.33	97.56
	CSA	90.50	85.00	71.11	92.68
	SCA	90.50	85.71	71.11	95.12
	PSO	92.00	85.00	70.00	92.68
	SO	91.50	85.00	71.11	95.12
Selected features	IAO	85	50	71	90
	EIAO	33	40	23	70
	CSA	186	40	162	230
	SCA	51	60	95	150
	PSO	137	100	107	220
	SO	98	70	95	260
Best fitness function	IAO	0.0950	0.1441	0.2776	0.0498
	EIAO	0.0816	0.1435	0.2674	0.0266
	CSA	0.0992	0.1506	0.2918	0.0763
	SCA	0.0955	0.1446	0.2868	0.0495
	PSO	0.0830	0.1538	0.3008	0.0761
	SO	0.0869	0.1522	0.2894	0.0526

Bold numbers indicate the best performance among all compared methods.

variation in the fitness values at different iterations. Standard deviation is a measure of the dispersion of data, with a larger value indicating greater differences between data points, and a smaller value indicating more concentrated data with less variation. On the other hand, Figure 5 examines whether the EIAO algorithm quickly converges to the optimal solution during iterations. By comparing with other algorithms, we can see that EIAO not only provides better solutions than the other five algorithms, but also converges significantly faster.

3.2.3 Modeling and analysis based on feature extraction

Although the analysis of four independent eggplant varieties confirmed the feature selection capability of EIAO, in practical applications, seeds are often stored in a mixed form consisting of multiple varieties. Therefore, this section combines the data from the four varieties to assess the practicality of EIAO in more complex scenarios. To evaluate the effect of feature selection on the classification model, the wavelength subsets selected by EIAO were input into the ELM, RF, and SVM models, and the results were compared with those obtained using the full wavelength data. Accuracy was used as the evaluation metric for the models. In this section, we compare the classification accuracy of three classifiers (ELM, RF, and SVM) using three sets of spectral data: raw data, preprocessed data, and data obtained after wavelength selection. This comparison aims to evaluate the effects of data preprocessing

TABLE 4 The proportion of the feature wavelength to the full-band.

Band range (nm)	Proportion (%)
395.25–489.30	13.83%
490.95–589.60	15.94%
591.34–693.20	15.58%
694.92–798.40	16.52%
800.20–903.90	14.96%
905.70–1008.20	23.17%

and feature selection on classification performance, providing insights into how these factors influence the overall model efficacy.

Based on the accuracy evaluation coefficient presented in Table 5, the following analysis can be made: First, when comparing different preprocessing methods, SG-EIAO and SNV-EIAO demonstrate the most significant optimization effects. These methods have achieved consistently favorable results across all classifiers, particularly on the test set, where the accuracy has reached 90% and 90.5%, respectively. This suggests that combining EIAO with spectral preprocessing methods can further enhance classifier performance and improve model classification accuracy across different datasets. Second, the improvement effects of EIAO optimization vary across different classifiers. In the ELM and RF models, accuracy significantly improved after optimization, especially on the test set. In the SVM model, although the improvement in test accuracy is relatively stable, its overall performance is not as high as that of the RF and ELM models. Furthermore, it was observed that, in the SVM model, the accuracy after MSC preprocessing with IAO and EIAO wavelength selection decreased notably, particularly for MSC-EIAO, where the test accuracy dropped to only 73%, significantly lower than other models. This may be attributed to the possibility that certain features, valuable to the SVM model, were discarded during feature extraction, leading to the loss of key information and a subsequent reduction in test accuracy.

To further evaluate the performance of the classification models, additional metrics including precision, recall, and F1-score were introduced alongside overall accuracy. As a key indicator of a model, the F1-score effectively reflects the balance between precision and recall across different classes. As shown in Table 6, among the three classifiers, the Random Forest (RF) model achieved the best overall performance, with the highest F1-score of 91.45% under the MSC-EIAO method. The Support Vector Machine (SVM) model also performed well under the SNV-EIAO method, reaching a test accuracy of 90.50% and an F1-score of 90.49%. However, SVM was found to be more sensitive to feature selection strategies, with its performance significantly declining under methods such as MSC-EIAO. Overall, the MSC-EIAO method demonstrated the greatest potential among all preprocessing and feature selection strategies. It significantly improved the performance of the RF model, showed stable effectiveness in the ELM model, but may suppress performance when applied to SVM.

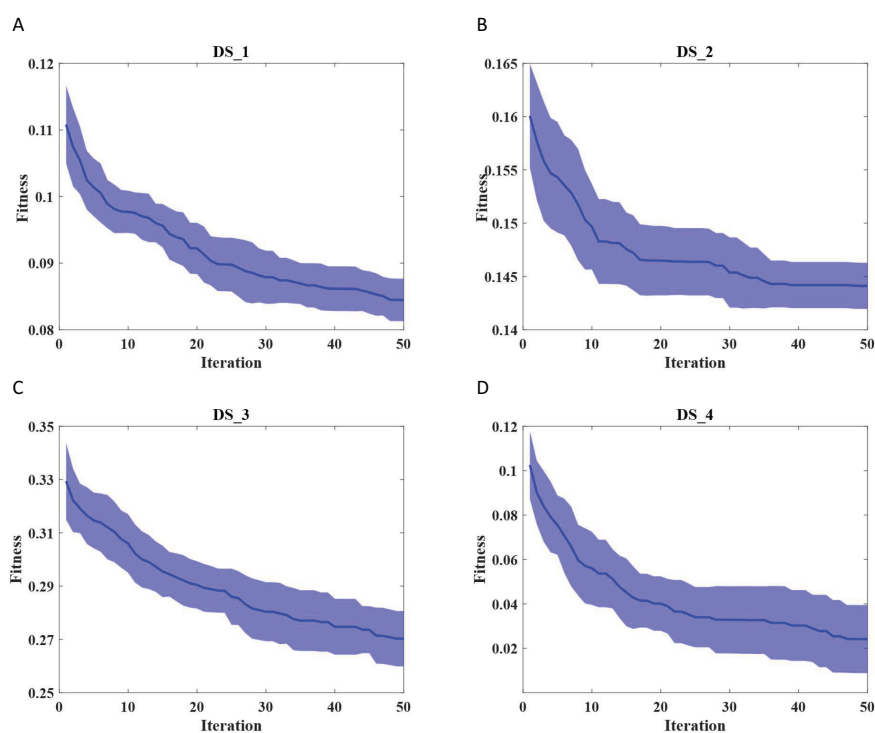


FIGURE 4

Comparison of the convergence performance of EIAO on four datasets. (A) DS_1, (B) DS_2, (C) DS_3, (D) DS_4.

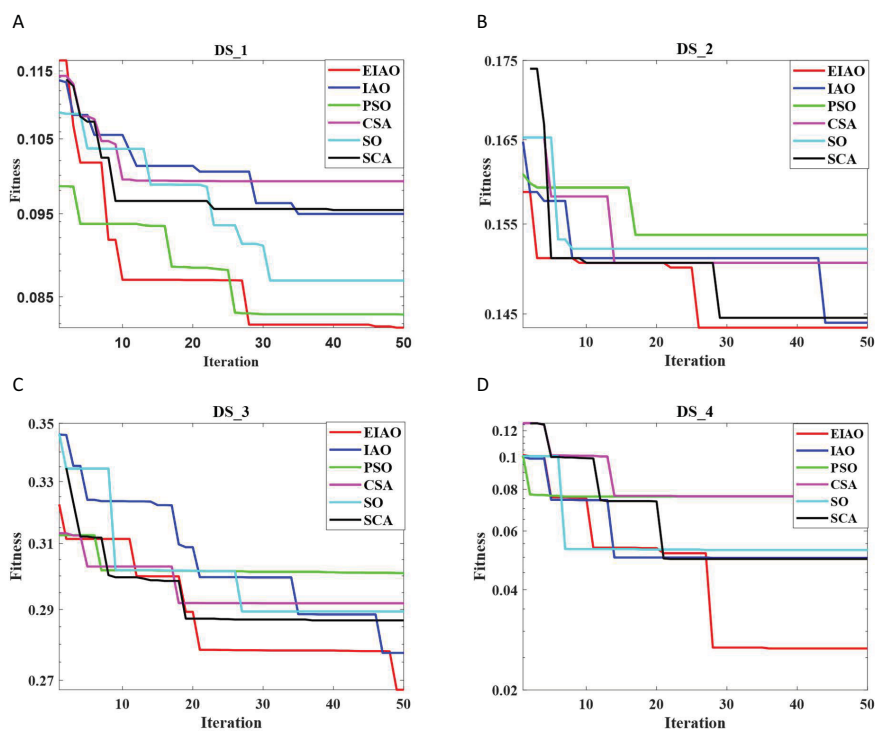


FIGURE 5

Stability analysis of EIAO and five baseline methods on four datasets. (A) DS_1, (B) DS_2, (C) DS_3, (D) DS_4.

TABLE 5 Accuracy of full and feature wavelength classification model.

Methods	Training accuracy (%)			Test accuracy (%)		
	ELM	RF	SVM	ELM	RF	SVM
RAW	91.87	94.87	87.22	89	83.50	87.50
MSC	92.87	96.12	86.50	85.50	87.50	85.50
SG	92	94.25	87.37	87.50	84	85
SNV	91	96.12	86.62	89.50	87	85.50
MSC-IAO	91.36	96.25	76.09	87	90.45	73.50
SG-IAO	92.49	93.87	86.50	88	82.41	84
SNV-IAO	92.61	95.87	90.86	87.50	86.93	88
MSC-EIAO	92.74	95.87	74.50	89	91.45	73
SG-EIAO	92.99	93.37	83.37	90	87.43	80.50
SNV- EIAO	92.24	95.37	90.25	90	89.44	90.50

Bold numbers indicate the best performance among all compared methods.

3.3 EIAO performs well on the UCI dataset

To illustrate the generality of the algorithm, we tried to test EIAO and other five optimization algorithms on four UCI datasets. Table 7 shows the average performance of EIAO and other five optimization algorithms on four UCI datasets. Due to the low feature dimension of the Vote dataset, the classification accuracy and the number of selected relevant features of the six algorithms are very close. On the Arrhythmia and Vehicle datasets, the classification performance of the algorithm is slightly inferior to that of the other two datasets. The reason may be that there are too many classification levels. For the Ionosphere dataset, the lowest optimal fitness value can be found, that is, the classification accuracy in Eq (5) is the highest and the least features are selected. At the same time, in order to analyze the robustness of

EIAO, it is checked whether the convergence curve of the EIAO algorithm is stable in multiple repetitions and whether it converges quickly to the optimal solution in iterations. In Figure 6, it is observed that EIAO achieves good performance compared with the other five swarm intelligence algorithms.

4 Discussion

Seed vigor is a crucial factor in crop cultivation and growth. Traditional methods for measuring vigor are often highly destructive and inefficient. Numerous studies have demonstrated that spectroscopy can effectively detect seed vigor parameters, while machine learning can be employed for modeling these parameters (Yang et al., 2017). For example, in chilling injury detection of eggplant, researchers have collected hyperspectral data that reflect internal physiological indicators and integrated them with machine learning to achieve consistent detection of eggplants stored under harmful cold conditions (Tsouvaltzis et al., 2020). This demonstrates that hyperspectral data can be used not only for quality evaluation but also hold great potential for monitoring changes in seed physiological status, thereby playing a significant role in improving the accuracy of seed vigor identification. While several studies have explored the integration of hyperspectral data with machine learning algorithms for the non-destructive analysis of seed quality parameters, research focused on vigor detection in this context remains limited. For instance, some studies have utilized hyperspectral data in conjunction with machine learning to assess the vigor of naturally aged rice seeds (Jin et al., 2022) (Xu et al., 2022). Another study combined visible and near-infrared hyperspectral data with machine learning techniques to non-destructively detect the vigor of watermelon seeds (Sun et al., 2024). These studies have provided preliminary evidence supporting the feasibility of using the full wavelength range for classifying seed vigor.

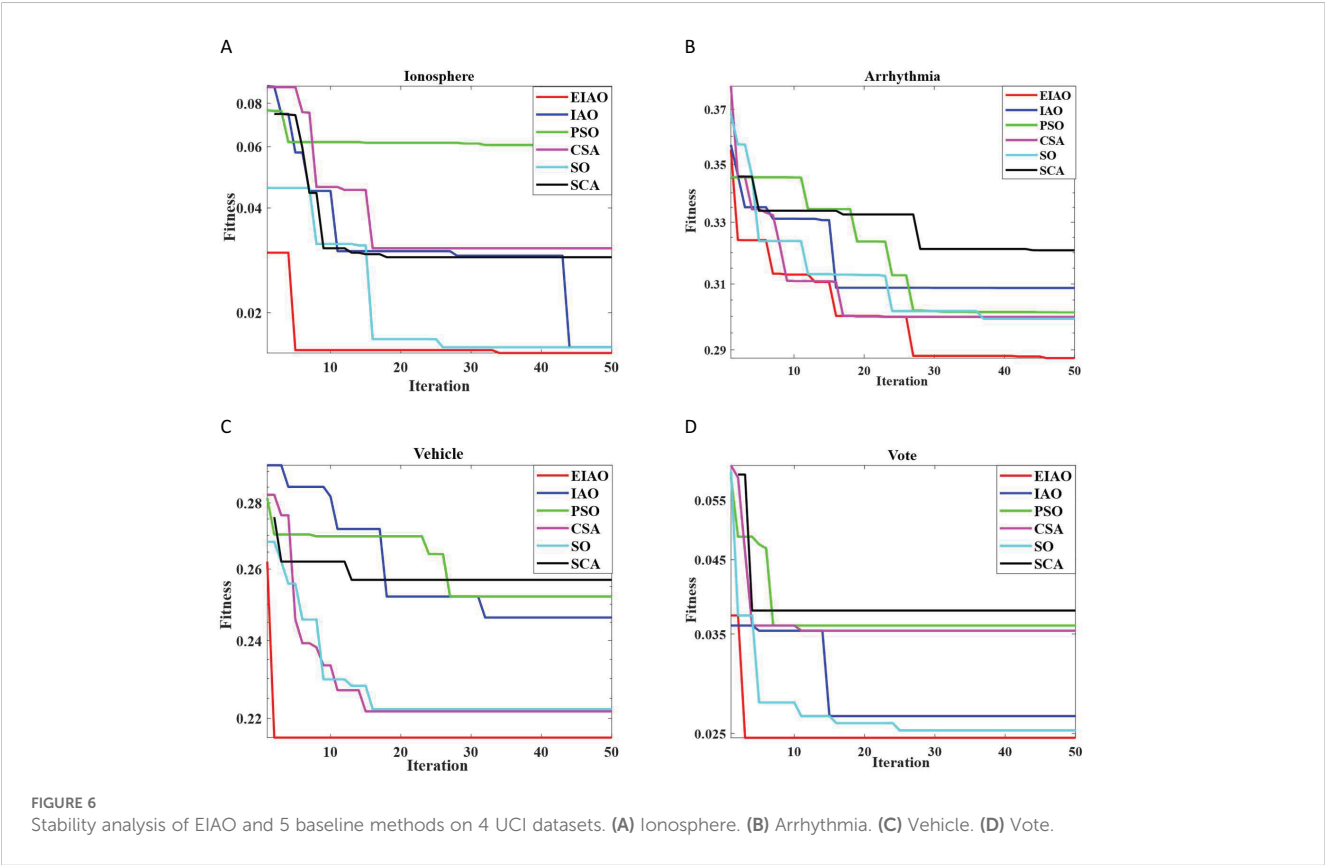
A series of feature selection algorithms were subsequently applied to extract characteristic wavelengths from the full spectral range,

TABLE 6 Comparison of average precision, recall, and F1-score under different methods.

Methods	ELM			RF			SVM		
	Average Precision	Average Recall	Average F1-score	Average Precision	Average Recall	Average F1-score	Average Precision	Average Recall	Average F1-score
MSC	85.86%	85.00%	85.43%	87.53%	87.50%	87.39%	85.51%	85.50%	85.50%
SG	87.88%	87.00%	87.44%	83.88%	84.05%	83.93%	85.01%	85.00%	84.96%
SNV	89.90%	89.00%	89.45%	87.04%	87.05%	87.03%	85.59%	85.50%	85.49%
MSC-IAO	85.58%	89.00%	87.25%	90.47%	90.45%	90.46%	73.53%	73.50%	73.50%
SG-IAO	86.54%	90.00%	88.24%	82.32%	82.35%	82.33%	84.06%	84.00%	84.00%
SNV-IAO	87.13%	88.00%	87.56%	86.95%	87.00%	86.90%	88.06%	88.00%	88.00%
MSC-EIAO	89.00%	89.00%	89.00%	91.46%	91.45%	91.45%	73.01%	73.00%	72.99%
SG-EIAO	90.82%	89.00%	89.00%	87.45%	87.45%	87.45%	80.53%	80.50%	80.50%
SNV- EIAO	90.00%	90.00%	90.00%	89.42%	89.45%	89.43%	90.54%	90.50%	90.49%

TABLE 7 Performance comparisons between EIAO and 5 baseline algorithms on UCI datasets.

Metric	Algorithm	Ionosphere	Arrhythmia	Vehicle	Vote
Accuracy (%)	IAO	95.71	70.00	75.74	97.70
	EIAO	97.14	72.22	78.70	97.70
	CSA	94.29	67.78	78.11	96.55
	SCA	97.14	70.00	74.56	96.55
	PSO	95.71	65.56	75.15	96.55
	SO	95.71	70.00	78.11	97.70
Selected features	IAO	6	32	11	6
	EIAO	4	30	8	2
	CSA	13	122	9	3
	SCA	5	36	9	6
	PSO	10	108	11	3
	SO	3	74	10	4
Best fitness function	IAO	0.0159	0.3088	0.2463	0.0265
	EIAO	0.0153	0.2876	0.2153	0.0241
	CSA	0.0306	0.2998	0.2217	0.0354
	SCA	0.0289	0.3208	0.2569	0.0379
	PSO	0.0607	0.3012	0.2521	0.036
	SO	0.0159	0.2993	0.2223	0.0253



aiming to reduce data redundancy while preserving classification accuracy. In seed vitality studies using spectral detection, traditional wavelength screening methods such as SPA and CARS are commonly employed. However, these methods are prone to high sensitivity to noise and unstable selected bands. To address these issues, this study proposes a novel wavelength selection algorithm, EIAO, and applies it to the band selection problem. Compared to other wavelength screening methods, the EIAO algorithm not only reduced the number of selected features but also demonstrated superior performance in terms of classification accuracy. Specifically, EIAO achieved an accuracy of 92.0%, outperforming benchmark algorithms such as IAO, PSO, and SCA.

Furthermore, this study utilized three classifiers, ELM, RF, and SVM, combined with various wavelength selection methods for performance evaluation. As shown in the results presented in Table 5, wavelength screening and feature subset selection significantly enhanced classifier performance, particularly in the RF and SVM models. After wavelength screening optimized by EIAO, the accuracy of the RF model on the test set reached 91.45%, and the accuracy of the SVM model was 90.50%, both of which were substantially higher than those of the unscreened original data model (83.50% and 87.50%, respectively). These findings indicate that wavelength screening not only improves classification accuracy but also reduces unnecessary features, thereby enhancing model performance. This series of results underscore the importance of wavelength screening in hyperspectral data analysis, particularly in seed vigor classification. The improved accuracy following wavelength screening aligns with the findings of Bi (Bi et al., 2024).

The generalization capability of the EIAO algorithm is further evaluated in the final part of this study. Compared to other optimization algorithms, EIAO demonstrates strong classification performance across multiple datasets. Notably, on the Ionosphere dataset, the classification accuracy achieved by EIAO is 97.14%, which is significantly higher than that of IAO (95.71%) and other optimization algorithms. This result aligns with previous research (Desuky et al., 2021), where the performance of the newly proposed EAOA algorithm was validated on a dataset with 16 feature selections. In that study, the new algorithm achieved the highest classification accuracy in selecting the optimal feature subset from the training data.

5 Conclusions

This study investigates the potential of hyperspectral technology for the non-destructive detection of eggplant seed vitality. Hyperspectral data from the region of interest of eggplant seeds were extracted, and preprocessing algorithms, including MSC, SNV, and SG, were applied to reduce noise. The EIAO algorithm proposed in this study was then employed to select feature wavelengths, and classification models using ELM, RF, and SVM were constructed. The results demonstrated that data preprocessing and feature selection substantially enhanced classification performance compared to the original data, with the EIAO-based feature selection method yielding particularly favorable outcomes. The RF and SVM models exhibited strong generalization ability under the optimized features, especially

on the test set. While the model based on the selected wavelengths resulted in the loss of some spectral information, it effectively reduced data redundancy and significantly improved classification accuracy. This study highlights the importance of selecting characteristic wavelengths for accurate modeling. Future research could further explore the integration of additional spectral information, such as texture features, to further enhance model performance.

Data availability statement

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

Author contributions

BY: Data curation, Formal Analysis, Writing – original draft, Writing – review & editing. XL: Supervision, Writing – review & editing. DZ: Supervision, Writing – review & editing. JZ: Supervision, Writing – review & editing. XF: Validation, Resources, Writing – review & editing. BP: Funding acquisition, Supervision, Writing – review & editing.

Funding

The author(s) declare that financial support was received for the research and/or publication of this article. This research was supported by the earmarked fund for CARS (CARS-23), Hebei Youth Talents Support Project, China (BJK2024038), the Talent Recruiting Program of Hebei Agricultural University (YJ2021056) and Hebei Province, Innovative Research Group Project (C2024204246).

Acknowledgments

Thanks to the teachers in the project room for their guidance and the help of the students. And we would like to thank Editage (www.editage.cn) for English language editing.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Generative AI statement

The author(s) declare that no Generative AI was used in the creation of this manuscript.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated

organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

References

- Abubakar, M., Yadav, D., Koul, B., and Song, M. (2023). Efficacy of Eco-Friendly Bio-Pesticides against the Whitefly *Bemisia tabaci* (Gennadius) for Sustainable Eggplant Cultivation in Kebbi State, Nigeria. *Agronomy-Basel* 13, 2083. doi: 10.3390/agronomy13123083
- Asafew, K., and Chandravanshi, B. (2021). Levels of major and trace metals in eggplant and soil. *Bull. Chem. Soc. Ethiopia* 35, 243–255. doi: 10.4314/bcse.v35i2.3
- Bi, C. G., Zhang, S., Chen, H., Bi, X. H., Liu, J. J., Xie, H., et al. (2024). Non-destructive classification of maize seeds based on RGB and hyperspectral data with improved grey wolf optimization algorithms. *Agronomy-Basel* 14, 645. doi: 10.3390/agronomy14040645
- Braik, M. S., Hammouri, A. I., Awadallah, M. A., Al-Betar, M. A., and Alzubi, O. A. (2023b). Improved versions of snake optimizer for feature selection in medical diagnosis: a real case COVID-19. *Soft Computing* 27, 17833–17865. doi: 10.1007/s00500-023-09062-3
- Braik, M. S., Hammouri, A. I., Awadallah, M. A., Al-Betar, M. A., and Khtatneh, K. (2023a). An improved hybrid chameleon swarm algorithm for feature selection in medical diagnosis. *Biomed. Signal Process. Control* 85, 105073. doi: 10.1016/j.bspc.2023.105073
- Breiman, L. (2001). Random forests. *Mach. Learn.* 45, 5–32. doi: 10.1023/A:1010933404324
- Chen, Y., Li, S., Zhang, X., Gao, X., Jiang, Y., Wang, J., et al. (2024). Prediction of apple moisture content based on hyperspectral imaging combined with neural network modeling. *Scientia Hort.* 338, 113739. doi: 10.1016/j.scienta.2024.113739
- Deep, K., and Thakur, M. (2007). A new crossover operator for real coded genetic algorithms. *Appl. Mathematics Comput.* 188, 895–911. doi: 10.1016/j.amc.2006.10.047
- Desuky, A. S., Hussain, S., Kausar, S., Islam, M. A., and Bakrawy, L. M. E. (2021). EAOA: an enhanced archimedes optimization algorithm for feature selection in classification. *IEEE Access* 9, 120795–120814. doi: 10.1109/ACCESS.2021.3108533
- De Vitis, M., Hay, F., Dickie, J., Trivedi, C., Choi, J., and Fiegner, R. (2020). Seed storage: maintaining seed viability and vigor for restoration use. *Restor. Ecol.* 28, S249–S255. doi: 10.1111/rec.13174
- Ebone, L. A., Caverzan, A., and Chavarria, G. (2019). Physiologic alterations in orthodox seeds due to deterioration processes. *Plant Physiol. Biochem.* 145, 34–42. doi: 10.1016/j.plaphy.2019.10.028
- Feng, H., Chen, Y., Song, J., Lu, B., Shu, C., Qiao, J., et al. (2024). Maturity classification of rapeseed using hyperspectral image combined with machine learning. *Plant Phenomics* 6, 139. doi: 10.34133/plantphenomics.0139
- Feng, L., Zhu, S. S., Liu, F., He, Y., Bao, Y. D., and Zhang, C. (2019). Hyperspectral imaging for seed quality and safety inspection: a review. *Plant Methods* 15, 91. doi: 10.1186/s13007-019-0476-y
- Gu, Y., Wu, J., Guo, Y., Hu, S., Li, K., Shang, Y., et al. (2024). Grade classification of camellia seed oil based on hyperspectral imaging technology. *Foods (Basel Switzerland)* 13, 3331. doi: 10.3390/foods13203331
- Huang, M., Tang, J. Y., Yang, B., and Zhu, Q. B. (2016). Classification of maize seeds of different years based on hyperspectral imaging and model updating. *Comput. Electron. Agric.* 122, 139–145. doi: 10.1016/j.compag.2016.01.029
- Huang, X., and Xia, L. (2023). Application of iterative distance correlation and PLS for wavelength interval selection in near infrared spectroscopy. *Chemometrics Intelligent Lab. Syst.* 234, 104756. doi: 10.1016/j.chemolab.2023.104756
- Huang, Q., Xue, H. R., Liu, J. P., and Jiang, X. H. (2022). A new method for spectral wavelength selection based on multiple linear regression combined with ant colony optimization and genetic algorithm. *J. Spectrosc.* 2022, 2440518. doi: 10.1155/2022/2440518
- Huang, G.-B., Zhu, Q.-Y., and Siew, C.-K. (2006). Extreme learning machine: Theory and applications. *Neurocomputing* 70, 489–501. doi: 10.1016/j.neucom.2005.12.126
- Huang, Q., Zhu, M. D., Xu, Z. Y., and Kan, R. F. (2024). A new framework for interval wavelength selection based on wavelength importance clustering. *Analytica Chimica Acta* 1326, 343153. doi: 10.1016/j.aca.2024.343153
- Jin, B., Qi, H., Jia, L., Tang, Q., Gao, L., Li, Z., et al. (2022). Determination of viability and vigor of naturally-aged rice seeds using hyperspectral imaging with machine learning. *Infrared Phys. Technol.* 122, 104097. doi: 10.1016/j.infrared.2022.104097
- Kocak, E., and Orkcü, H. H. (2024). Particle swarm optimization based feature selection using factorial design. *Haceteppe J. Mathematics Stat* 53, 879–896. doi: 10.15672/hujms.1346686
- Li, J. (2012). “A hybrid differential evolution algorithm with opposition-based learning,” in *2012 4th International Conference on Intelligent Human-Machine Systems and Cybernetics*. Nanchang, China: IEEE, 85–89.
- Li, J., An, Q., Lei, H., Deng, Q., and Wang, G.-G. (2022). Survey of levy flight-based metaheuristics for optimization. *Mathematics* 10, 2785. doi: 10.3390/math10152785
- Nasab, S. T. M., and Abualigah, L. (2024). Improve Harris Hawkes optimizer algorithm via Laplace crossover. *J. Ambient Intell. Humanized Computing* 15, 2057–2072. doi: 10.1007/s12652-023-04734-2
- Ong, P., Jian, J. B., Yin, J. H., and Ma, G. D. (2023). Characteristic wavelength optimization for partial least squares regression using improved flower pollination algorithm. *Spectrochimica Acta Part a-Molecular Biomolecular Spectrosc.* 302, 123095. doi: 10.1016/j.saa.2023.123095
- Pernkopf, F. (2005). Bayesian network classifiers versus selective k-NN classifier. *Pattern Recognition* 38, 1–10. doi: 10.1016/j.patcog.2004.05.012
- Piccialli, V., Schwidessen, J., and Sudoso, A. M. (2024). Optimization meets machine learning: an exact algorithm for semi-supervised support vector machines. *Math. PROGRAMMING.* 1–43. doi: 10.1007/s10107-024-02175-z
- Que, H. T., Zhao, X., Sun, X. L., Zhu, Q. B., and Huang, M. (2023). Identification of wheat kernel varieties based on hyperspectral imaging technology and grouped convolutional neural network with feature intervals. *Infrared Phys. Technol.* 131, 104653. doi: 10.1016/j.infrared.2023.104653
- Si, Y., Brumerick, F., Yang, C., Glowacz, A., Ma, Z., Siarry, P., et al. (2023). Prediction and evaluation of energy and exergy efficiencies of a nanofluid-based photovoltaic-thermal system with a needle finned serpentine channel using random forest machine learning approach. *Eng. Anal. Boundary Elements* 151, 328–343. doi: 10.1016/jenganabound.2023.03.009
- Sun, J., Nirere, A., Dusable, K. D., Zhong, Y., and Adrien, G. (2024). Rapid and nondestructive watermelon (*Citrullus lanatus*) seed viability detection based on visible near-infrared hyperspectral imaging technology and machine learning algorithms. *J. Food Sci.* 89, 4403–4418. doi: 10.1111/1750-3841.17151
- Sun, L., Qin, H., Przystupa, K., Cui, Y., Kochan, O., Skowron, M., et al. (2022). A hybrid feature selection framework using improved sine cosine algorithm with metaheuristic techniques. *Energies* 15, 3485. doi: 10.3390/en15103485
- Tizhoosh, H. R. (2005). “Opposition-based learning: A new scheme for machine intelligence,” in *International Conference on Computational Intelligence for Modelling, Control and Automation and International Conference on Intelligent Agents, Web Technologies and Internet Commerce (CIMCA-IAWTIC'06)*. Vienna, Austria: IEEE, 695–701.
- Tsouvaltzis, P., Babellahi, F., Amodio, M. L., and Colelli, G. (2020). Early detection of eggplant fruit stored at chilling temperature using different non-destructive optical techniques and supervised classification algorithms. *Postharvest Biol. Technol.* 159, 111001. doi: 10.1016/j.postharvbio.2019.111001
- Ventura, L., Donà, M., Macovei, A., Carbonera, D., Buttafava, A., Mondoni, A., et al. (2012). Understanding the molecular pathways associated with seed vigor. *Plant Physiol. Biochem.* 60, 196–206. doi: 10.1016/j.plaphy.2012.07.031
- Wang, Y. L., Peng, Y. K., Qiao, X., and Zhuang, Q. B. (2021a). Discriminant analysis and comparison of corn seed vigor based on multiband spectrum. *Comput. Electron. Agric.* 190, 106444. doi: 10.1016/j.compag.2021.106444
- Wang, Y., and Song, S. (2024). Detection of sweet corn seed viability based on hyperspectral imaging combined with firefly algorithm optimized deep learning. *Front. Plant Sci.* 15. doi: 10.3389/fpls.2024.1361309
- Wang, Z., Tian, X., Fan, S., Zhang, C., and Li, J. (2021b). Maturity determination of single maize seed by using near-infrared hyperspectral imaging coupled with comparative analysis of multiple classification models. *Infrared Phys. Technol.* 112, 103596. doi: 10.1016/j.infrared.2020.103596
- Wang, H., Wu, Z., Rahnamayan, S., Liu, Y., and Ventresca, M. (2011). Enhancing particle swarm optimization using generalized opposition-based learning. *Inf. Sci.* 181, 4699–4714. doi: 10.1016/j.ins.2011.03.016
- Wu, X., Li, S., Jiang, X., and Zhou, Y. (2024). Information acquisition optimizer: a new efficient algorithm for solving numerical and constrained engineering optimization problems. *J. Supercomputing* 80, 25736–25791. doi: 10.1007/s11227-024-06384-3

- Xing, M., Long, Y., Wang, Q., Tian, X., Fan, S., Zhang, C., et al. (2023). Physiological alterations and nondestructive test methods of crop seed vigor: A comprehensive review. *Agriculture* 13, 527. doi: 10.3390/agriculture13030527
- Xu, L., Xie, Y., Chen, X., Chen, Y., Kang, Z., Huang, P., et al. (2022). Design of an efficient combined multipoint picking scheme for tea buds. *Front. Plant Sci. Volume* 13. doi: 10.3389/fpls.2022.1042035
- Yamaguchi, S., Matsumoto, K., Koyama, M., Tian, S., Watanabe, M., Takahashi, A., et al. (2019). Antihypertensive effects of orally administered eggplant (*Solanum melongena*) rich in acetylcholine on spontaneously hypertensive rats. *Food Chem.* 276, 376–382. doi: 10.1016/j.foodchem.2018.10.017
- Yang, S., Zhu, Q.-b., and Huang, M. (2017). Application of joint skewness algorithm to select optimal wavelengths of hyperspectral image for maize seed classification. *Spectrosc. Spectral Anal.* 37, 990–996. doi: 10.3964/j.issn.1000-0593(2017)03-0990-07
- Zhang, X., and Feng, T. (2018). Chaotic bean optimization algorithm. *Soft Computing* 22, 67–77. doi: 10.1007/s00500-016-2322-8
- Zhang, D., Zhang, J., Peng, B., Wu, T., Jiao, Z., Lu, Y., et al. (2023). Hyperspectral model based on genetic algorithm and SA-1DCNN for predicting Chinese cabbage chlorophyll content. *Scientia Hort.* 321, 112334. doi: 10.1016/j.scienta.2023.112334