

Article

Influence of Laser Treatment on Ginseng Seed Vigor and Non-Destructive Prediction of Emergence Rate

Mingxuan Xue^{1,2}, Jing Guo¹, Yonghua Xu³, Zhaocong Lyu¹, Yansong Li¹, Leijinyu Zhou¹, Qingyang Han² and Helong Yu^{1,2,*} 

¹ Smart Agriculture Research Institute, Jilin Agricultural University, Changchun 130118, China; xuemx@jlau.edu.cn (M.X.); guojing0862@163.com (J.G.); 14743883337@163.com (Z.L.); 20241612@mails.jlau.edu.cn (Y.L.); 20242136@mails.jlau.edu.cn (L.Z.)

² College of Information Technology, Jilin Agricultural University, Changchun 130118, China; 15143204513@163.com

³ College of Chinese Medicinal Materials, Jilin Agricultural University, Changchun 130118, China; xuyonghua@jlau.edu.cn

* Correspondence: yuhelong@jlau.edu.cn

Abstract

Seed vigor directly affects emergence rate, early emergence capacity, and seedling uniformity in ginseng. This study integrated pre-sowing laser treatment with hyperspectral prediction for rapid evaluation of ginseng seed vigor. Cracked-stage Mountain Cultivated Ginseng Seeds and American Ginseng Seeds were treated with four semiconductor laser conditions, including red light, blue light, red–blue 3:1, and red–blue 6:1, for 5, 15, and 30 min. Emergence rate, emergence energy, and emergence index were evaluated, and 350–2500 nm hyperspectral reflectance data were collected. Different preprocessing, feature selection, and machine learning methods were compared to construct an optimized support vector regression model for emergence rate prediction. The results showed that laser effects were seed-type-specific and parameter-dependent. Mountain Cultivated Ginseng Seeds responded strongly, with 5 min blue light treatment producing the best performance, achieving a 100.0% emergence rate, 40.4 percentage points higher than the control, with significantly higher emergence energy and emergence index. In American Ginseng Seeds, no treatment differed significantly from the control, although the 30 min red light treatment showed the highest emergence rate trend. Savitzky–Golay smoothing combined with the second derivative was the optimal preprocessing method. The Caterpillar Fungus Optimizer-based support vector regression (CFO-SVR) joint optimization model achieved the best prediction performance, with R_p^2 of 0.9517, $RMSEP$ of 2.4766, and $MAEP$ of 1.0730. These results provide a feasible technical reference for pre-sowing treatment of ginseng seeds and rapid, non-destructive evaluation of emergence vigor.

Keywords: ginseng seeds; laser pre-sowing treatment; seed vigor; hyperspectral spectroscopy; CFO-SVR; non-destructive prediction



Academic Editor: Nikolaos Katsenios

Received: 31 May 2026

Revised: 12 July 2026

Accepted: 14 July 2026

Published: 16 July 2026

Copyright: © 2026 by the authors.

Licensee MDPI, Basel, Switzerland.

This article is an open access article distributed under the terms and conditions of the [Creative Commons Attribution \(CC BY\)](https://creativecommons.org/licenses/by/4.0/) license.

1. Introduction

Ginseng (*Panax ginseng* C.A. Meyer) is a traditional and valuable medicinal plant in China, with a long history of medicinal use and high economic value. Its roots are rich in ginsenosides, polysaccharides, amino acids, volatile oils, and various bioactive compounds, and are widely used in traditional Chinese medicine preparations, functional foods, and health products [1]. In artificial ginseng cultivation, seed quality is an important factor

affecting emergence stability, seedling uniformity, and subsequent yield formation, among which seed vigor is directly related to germination potential and field stand establishment. However, ginseng seeds exhibit morphological after-ripening and physiological dormancy. Their emergence process is affected by multiple factors, including seed maturity, storage conditions, pregermination treatment, and post-sowing environment, often resulting in a long germination period, uneven emergence, and unstable emergence rate [2]. Therefore, exploring appropriate pre-sowing vigor treatment methods for ginseng seeds and establishing rapid, non-destructive, and accurate vigor evaluation techniques are important for improving ginseng seedling quality and promoting precision cultivation.

In recent years, laser treatment has received increasing attention in crop breeding and seed vigor enhancement as a green, non-contact, and parameter-controllable physical regulation technology. Lasers are characterized by strong directionality, good monochromaticity, and high energy density. At an appropriate dose, laser treatment can promote seed germination and seedling growth by affecting seed energy status, enzyme activity, membrane system stability, and reserve mobilization [3]. Previous studies have shown that He–Ne laser treatment can enhance germination-related enzyme activities and improve germination characteristics in safflower seeds [4]; He–Ne red light irradiation at an appropriate time can promote seed germination and seedling physiological performance in triticale [5], and low-dose Nd:YAG laser treatment can improve seed germination, photosynthetic pigments, and morphological traits in maize [6]. In seeds of *Arabidopsis thaliana* [7], *Erigeron breviscapus* [8], *Elaeagnus mollis* Diels [9], eggplant [10], and soybean [11], appropriate laser treatment has also been shown to increase germination rate, germination energy, or seedling growth quality. However, the biological effects of laser irradiation generally show clear dose dependence and species specificity. An insufficient treatment time may fail to induce effective stimulation, whereas excessive intensity or prolonged treatment may inhibit germination or even damage seeds. At present, systematic studies on the response patterns of ginseng seeds to semiconductor laser treatments with different light-quality combinations and irradiation times remain limited, and suitable laser regulation parameters for ginseng seeds have not yet been clarified.

Seed vigor evaluation is an important component of seed quality control and pre-sowing grading. Although conventional germination tests can directly reflect seed germination capacity, they require a long testing period, are strongly affected by environmental conditions, and are difficult to meet the practical requirements of rapid grading and precision seedling production. Hyperspectral technology can simultaneously acquire spectral response information from samples across multiple continuous wavelengths and characterize changes in internal seed components, such as water, proteins, lipids, sugars, and other storage reserves, without destroying the samples. Therefore, it has high application potential for the non-destructive detection of seed vigor.

Previous studies have combined hyperspectral technology with machine learning or deep learning methods to identify the vigor or viability of seeds such as maize [12], common bean [13], *Sophora japonica* [14], cucumber [15], strawberry [16], wheat [17], peanut [18], and rice [19], achieving favorable classification or prediction performance. These studies indicate that spectral preprocessing, feature wavelength selection, and modeling algorithm optimization are key steps for improving the prediction accuracy of seed vigor. However, existing studies have mainly focused on vigor grading of artificially aged seeds or conventional crop seeds, whereas studies on the spectral response and quantitative prediction of medicinal plant seeds, especially the emergence vigor of ginseng seeds after laser treatment, remain limited. Laser treatment may induce internal physiological changes, such as water migration, membrane system recovery, and reserve transformation, which are expected to be reflected by spectral features in the near-infrared and shortwave-infrared regions.

Therefore, combining laser-based pre-sowing treatment with hyperspectral non-destructive detection may help address two key issues: seed vigor improvement and rapid vigor evaluation in ginseng seeds.

Based on this research background, semiconductor laser irradiation, as a low-energy physical stimulus, has potential for regulating seed vigor, while hyperspectral reflectance technology provides a new technical approach for the rapid and non-destructive evaluation of seed vigor status. However, studies on the effects of semiconductor laser treatment on the emergence vigor of ginseng seeds remain limited. The response differences in different ginseng seed types to light-quality combinations and irradiation times are still unclear, and methods for rapidly predicting the emergence rate (ER) of laser-treated ginseng seeds based on spectral information require further improvement.

Therefore, this study used cracked-stage Mountain Cultivated Ginseng Seeds (MCGS) and American Ginseng Seeds (AGS) as materials to systematically analyze the effects of semiconductor laser treatments with different light-quality combinations and irradiation times on ER, emergence energy (EE), and emergence index (EI). Hyperspectral reflectance data of the treated seeds were collected, and spectral preprocessing, feature wavelength selection, and machine learning modeling methods were combined to construct a non-destructive prediction model for ginseng seed ER. The innovations of this study are mainly reflected in three aspects: first, semiconductor laser pre-sowing treatment was applied to regulate ginseng seed vigor, and the emergence response differences between MCGS and AGS under different light-quality combinations and irradiation times were compared; second, hyperspectral reflectance data were used to construct a rapid and non-destructive prediction model for the ER of laser-treated ginseng seeds, providing a new technical approach for ginseng seed vigor evaluation; third, the Caterpillar Fungus Optimizer (CFO) was introduced for feature wavelength selection and coordinated optimization of SVR parameters, and a CFO-SVR joint optimization framework was constructed to improve model prediction accuracy and feature interpretability.

2. Materials and Methods

2.1. Materials and Experimental Equipment

The ginseng seeds used in this study were purchased from a ginseng seed supplier in Jilin Province, China. To ensure the reliability and reproducibility of the experimental results, only healthy cracked-stage MCGS and AGS with uniform appearance, no disease spots, and no mechanical damage were selected as experimental materials for laser irradiation treatment.

The light source used in the experiment was a laser plant growth regulation lamp (model: CXL1-X0130-0001; Changxin Optoelectronic Technology Co., Ltd., Hangzhou, China). Different light-quality combinations were obtained by adjusting the light-emitting modules.

The hyperspectral data acquisition system consisted of a portable field spectroradiometer (FieldSpec4-Hi-Res NG; ASD Inc., Boulder, CO, USA), a fixed reflectance probe, a matching halogen light source, optical fibers, and a computer. The field spectroradiometer had a wavelength range of 350–2500 nm and could acquire 2150 spectral data points. The RS3 software (version 6.4.3) supplied with the spectroradiometer was used to collect hyperspectral data from ginseng seeds. Data organization and processing in this study were performed using ViewSpecPro 6.2, Microsoft Excel, and PyCharm 2024.3.1. Statistical analysis was conducted using IBM SPSS Statistics 27.0.1, and figure plotting and visualization were performed using Origin and Microsoft PowerPoint.

2.2. Laser Treatment Design

Different light-quality combinations were obtained by adjusting the light-emitting modules. Four light-quality conditions were set in the experiment, namely monochromatic red light, monochromatic blue light, red–blue 3:1, and red–blue 6:1, to compare the effects of different light-quality ratios on seed emergence rate.

As shown in Figure 1, during irradiation, all seeds were evenly spread in Petri dishes to ensure uniform light exposure and avoid shading among seeds. Three irradiation times, namely 5, 15, and 30 min, were set for each light-quality treatment to investigate the effects of different irradiation times on the emergence rate of ginseng seeds. The laser irradiation intensity was uniformly controlled at $50 \mu\text{mol}\cdot\text{m}^{-2}\cdot\text{s}^{-1}$ (PPFD), ensuring that all experimental groups were conducted under the same light-intensity conditions.



Figure 1. Laser irradiation setup: (a) blue light irradiation; (b) red light irradiation.

The control group consisted of seeds without any laser treatment, which were directly subjected to subsequent hyperspectral data acquisition and planting experiments. All treatments were strictly performed according to the experimental design to ensure the scientific validity and comparability of the results.

2.3. Sowing Experiment and Emergence Measurement

This experiment was conducted in the Science and Technology Courtyard Laboratory of the Institute of Smart Agriculture, Jilin Agricultural University. Sowing was performed at 10:00 a.m. on 21 April 2025. For each ginseng species, 12 experimental treatments and 1 blank control (CK) were established, resulting in a total of 26 experimental groups for the two species.

Each experimental group included three biological replicates, with 33 seeds per replicate. A total of 2574 seeds were sown across all 26 groups. The experimental substrate was prepared under the guidance of agronomic experts by mixing peat, vermiculite, and perlite at a ratio of 5:3:1, followed by the addition of 20 g of carbendazim solution for thorough mixing and disinfection. During sowing, three seeds were placed in each seedling cup and evenly distributed in a triangular arrangement. Specifically, each seedling cup was first filled with substrate to two-thirds of its volume. The seeds were then placed and covered with a thin layer of substrate, followed by watering until the substrate was fully moistened. During the cultivation period, regular watering was applied to maintain the substrate's relative water content at approximately 50%.

Seed emergence was monitored and recorded daily until no new seedlings appeared. MCGS began to emerge on the 9th day after sowing, and emergence ended on the 25th day; AGS began to emerge on the 10th day after sowing, and emergence ended on the 36th day.

2.4. Hyperspectral Data Acquisition and Preprocessing

2.4.1. Hyperspectral Data Acquisition

Hyperspectral data acquisition of the laser-irradiated ginseng seeds, including MCGS and AGS, was conducted in a completely dark indoor environment to prevent interference from external light sources. The hyperspectral data acquisition system is shown in Figure 2. The hyperspectral data in this study were not collected immediately after laser irradiation. All laser treatments were completed one day before hyperspectral data acquisition, and hyperspectral reflectance data were collected on the following day. Before spectral acquisition, seeds from all laser treatment groups and the control group were kept overnight under the same indoor environmental conditions to ensure that seeds from different treatment groups were in a relatively consistent environmental state before acquisition, and to minimize the effects of instantaneous surface temperature differences and the photothermal cooling process after laser irradiation on the spectral acquisition results.

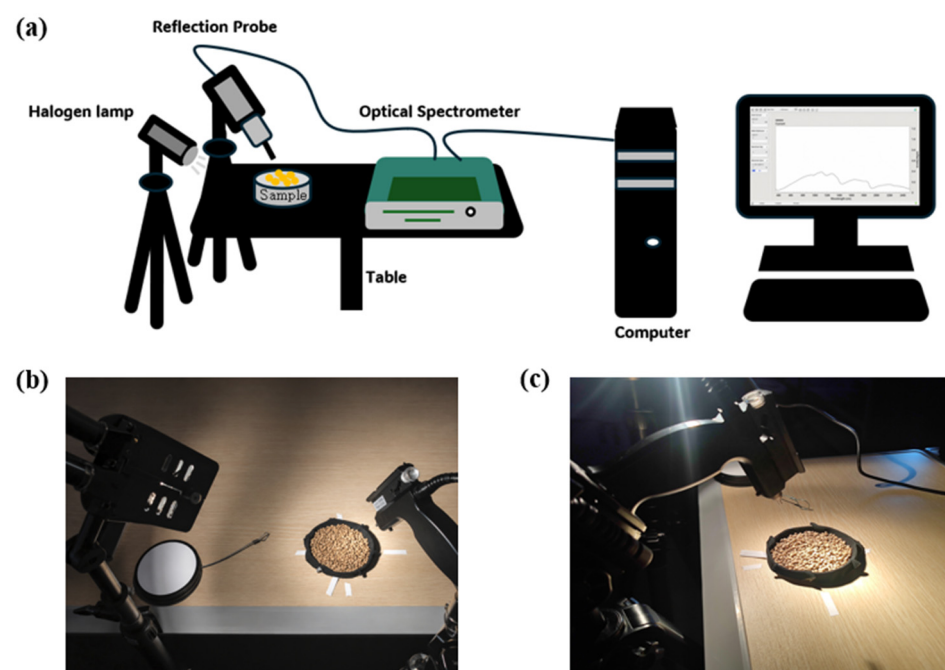


Figure 2. Hyperspectral data acquisition system. (a) Schematic diagram of the hyperspectral data acquisition system; (b) top view of the actual measurement setup; (c) side view of the actual measurement setup.

To ensure instrument stability and reduce the influence of external environmental changes on data acquisition, the spectroradiometer was warmed up for 10 min before the experiment to reach a stable operating state, followed by white reference calibration. During the experiment, white reference calibration was repeated whenever a sample was replaced and before data acquisition began. After calibration, the instrument was considered successfully calibrated when the spectral reflectance curve appeared as a smooth straight line.

During hyperspectral data acquisition, each sample was evenly spread in a tray with a black background to reduce the influence of background reflection on the spectral acquisition results. During measurement, the probe was positioned approximately 10 cm above the seeds and fixed at a 45° angle relative to the tray surface. A 20 W halogen light source was fixed on one side of the sample and above the seed-spreading area, with the lamp body positioned approximately 22 cm above the tabletop where the sample was placed, providing stable illumination for the seed-spreading area. The illuminated area

covered the seed samples to be measured. Throughout the acquisition process, the probe height, probe angle, light source position, sample background, and sample placement method were kept consistent to ensure the consistency and reproducibility of spectral acquisition conditions among different treatment groups. For each treatment group, data were collected at 13 different points, and 10 spectral curves were acquired at each point. The 10 collected spectral curves were then averaged into one spectrum, resulting in a total of 338 spectral samples, thereby ensuring the representativeness and diversity of the data. It should be noted that the ER in this study was a population-level indicator at the treatment-group level, rather than an independent indicator at the single-seed or single-spectrum level. Therefore, the 338 spectral samples represented 338 spectral observations, but not 338 mutually independent ER observations. During model development, spectra obtained from different sampling points within the same treatment group were assigned the corresponding ER label of that treatment group to establish the predictive relationship between hyperspectral features and treatment-level ER. The dataset was randomly divided at the spectral-sample level at a ratio of 7:3, rather than at the treatment-group level. Therefore, different spectra from the same treatment group may have entered the training and test sets separately, but the same spectrum did not appear in both datasets.

2.4.2. Hyperspectral Data Preprocessing

During hyperspectral data acquisition, spectral curves are easily affected by light source fluctuations, slight deviations in probe angle, and instrumental noise, resulting in noise interference, scattering effects, and baseline drift. These non-systematic errors may obscure feature wavelengths related to seed emergence rate, thereby affecting modeling accuracy. Therefore, before modeling analysis, the raw spectral data must be preprocessed to reduce irrelevant interference and improve the correlation and representational ability between the spectra and emergence rate. In this study, several classical preprocessing methods were adopted, including Savitzky–Golay (SG) smoothing [20], which was used to reduce high-frequency random noise while preserving spectral profile characteristics; standard normal variate (SNV) transformation [21], which eliminates amplitude variations caused by particle size or scattering differences through mean centering and variance standardization; multiplicative scatter correction (MSC) [22], which was used to correct spectral baseline drift and scattering effects, thereby making the spectra of different samples more comparable; first derivative (1st Der) [23], which was used to eliminate partial baseline shifts and enhance spectral variation trends; second derivative (2nd Der) [24], which was used to further highlight weak absorption peaks and subtle spectral features; and detrending/baseline drift correction (DT) [25], which was used to remove slowly varying trends in spectral curves and reduce baseline shifts caused by instrument drift, illumination changes, or differences in sample surface conditions. Building upon the single preprocessing method, this study further explored two combined preprocessing strategies: SG smoothing combined with the first derivative (S-G 1st) [26] and SG smoothing combined with the second derivative (S-G 2nd) [27], aiming to enhance spectral detail information and provide a more favorable data basis for subsequent feature selection and model construction.

2.5. Feature Selection and Modeling Strategy

2.5.1. Feature Selection Methods

After spectral data preprocessing was completed and the optimal preprocessing method was determined, further selection of spectral feature variables was required. The reason is straightforward: on the one hand, this process can reduce interference caused by redundant information during modeling; on the other hand, when the number of high-

dimensional spectral variables is excessive, the model is more prone to overfitting. Based on this consideration, two typical feature selection algorithms, Random Frog and competitive adaptive reweighted sampling, together with the Caterpillar Fungus Optimization algorithm, were further employed in this study to select and optimize feature wavelengths, thereby reducing model complexity and improving computational efficiency.

Random Frog (RF) is a stochastic feature selection method suitable for high-dimensional spectral data [28]. This method continuously generates and updates candidate variable subsets randomly in the variable space and evaluates each subset in combination with partial least squares regression (PLSR). During multiple iterations, the selection frequency of each variable is counted to characterize its importance. The variables are then ranked according to their selection probabilities, and different variable subsets are compared based on the root mean square error of k-fold cross-validation ($RMSE_{CV}$) on the full training set. When $RMSE_{CV}$ reaches its minimum value, the corresponding set of variables with high selection probabilities is regarded as the optimal feature set. This method can effectively extract key wavelengths that make substantial contributions to modeling and reduce the interference of redundant variables and noise information, thereby decreasing model complexity and improving model prediction performance.

Competitive adaptive reweighted sampling (CARS) [29] is a feature selection method that combines Monte Carlo sampling with partial least squares regression. This method establishes submodels through repeated random sampling, uses an exponential decay function to control the number of retained variables, and gradually selects candidate variables according to the magnitude of regression coefficients. Meanwhile, the root mean square error of k-fold cross-validation ($RMSE_{CV}$) for each variable subset on the full training set is recorded throughout the process. When $RMSE_{CV}$ reaches its minimum value, the corresponding variable set is regarded as the optimal feature set, thereby enabling effective extraction of key wavelengths.

The Caterpillar Fungus Optimizer (CFO) is a swarm intelligence optimization algorithm inspired by the growth, search, and parasitic behaviors of caterpillar fungus [30]. This algorithm simulates the processes by which caterpillar fungus searches for hosts, parasitizes hosts, and continuously adjusts its search strategy in complex environments, and abstracts these processes into an optimization mechanism that combines global exploration and local exploitation. The core logic of the algorithm is constructed around two search modes and two parasitic behaviors. The two search modes, namely wave propagation and spiral ascent, can effectively enhance the global exploration ability of the population, whereas the two parasitic behaviors, namely repeated parasitism and optimal parasitism, can improve the approximation accuracy of individuals toward high-quality solution regions, thereby enabling efficient search for the global optimal solution. Compared with traditional swarm intelligence algorithms, this method has strong advantages in escaping local optima and maintaining convergence stability.

Let the dimension of the optimization problem be D , the population size be N , the position vector of the i -th individual at the t -th iteration be denoted as X_i^t , and the current global best position be denoted as X_{best}^t . Each iteration of CFO mainly consists of two stages: the exploration stage and the parasitic stage. The exploration stage is used to expand the search range, whereas the parasitic stage is used to strengthen local optimization ability.

(1) In the exploration stage, the algorithm first ranks the population according to fitness. Then, each individual selects either the “wave propagation operator” or the “spiral ascent operator” with equal probability to update its position.

The wave propagation operator simulates the random undulating behavior of caterpillar fungus when searching for hosts in soil, and its position update formula is shown in Equation (1):

$$X_i^{so} = \begin{cases} X_i^t - \text{rand}(1, D) \odot (X_{best}^t - X_i^t) + \alpha \odot (X_{best}^t - X_i^t), & i = 1 \\ X_i^t - \text{rand}(1, D) \odot (X_{i-1}^t - X_i^t) + \alpha \odot (X_{best}^t - X_i^t), & i = 2, 3, \dots, N \end{cases} \quad (1)$$

where $\odot$ denotes the Hadamard element-wise product, $\text{rand}(1, D)$ denotes a D -dimensional uniform random vector within $[0, 1]$, and X_{i-1}^t represents the position of the adjacent individual ranked better than the i -th individual after sorting. The parameter α is the random search step size in the wave propagation stage, and its expression is shown in Equation (2):

$$\alpha = 2.5 \text{rand}(1, D) \odot |\cos(\pi \text{rand}(1, D))| \quad (2)$$

This mechanism guides individuals toward the current optimal solution while exploring new candidate regions globally through random step sizes, thereby enhancing the global search capability of the algorithm.

The spiral ascent operator simulates the behavior of caterpillar fungus gradually narrowing its search range and spirally approaching dominant regions when it approaches a suitable environment. Its update method is shown in Equation (3):

$$X_i^{so} = \begin{cases} X_{best}^t - \text{rand}(1, D) \odot (X_{best}^t - X_i^t) + \beta (X_{best}^t - X_i^t), & i = 1 \\ X_{best}^t - \text{rand}(1, D) \odot (X_{i-1}^t - X_i^t) + \beta (X_{best}^t - X_i^t), & i = 2, 3, \dots, N \end{cases} \quad (3)$$

where β is the search step size in the spiral ascent stage, and its mathematical expression is shown in Equation (4):

$$\beta = 2 \cos(\pi r_1) \left(\frac{t}{T_{\max}} \right)^{r_1 r_2} \quad (4)$$

where $r_1 \in [0, 1]$ is a uniformly distributed random number, r_2 is a random integer taking the value of 1 or 2, and $T_{\max}$ is the maximum number of iterations. This parameter dynamically adjusts the search radius, allowing the algorithm to maintain strong search flexibility in the early iterations and gradually narrow the search range in the later iterations. This enables a natural transition from coarse-grained exploration to fine-grained exploitation, thereby enhancing the convergence ability and solution precision of the algorithm in the later stage.

(2) After the exploration update is completed, the algorithm enters the parasitic stage. This stage imitates the parasitic behavior of caterpillar fungus on host larvae and constructs two position update strategies: “repeated parasitism” and “optimal parasitism”. The former emphasizes jump search under random perturbation, whereas the latter focuses on directional contraction around the global optimal region. The two strategies cooperate with each other, helping to improve local exploitation efficiency while maintaining global search capability.

Repeated parasitism is used to describe the process in which an individual continues to correct its position based on an existing parasitic relationship. Its update formula is shown in Equation (5):

$$X_i^{pl} = X_i^{so} + 3r_3 (\text{rand} \cdot X_{best}^t - \text{rand} \cdot X_i^{so}), \quad i = 1, 2, \dots, N \quad (5)$$

where X_i^{pl} denotes the new position of the individual in the parasitic stage, and r_3 is a random scalar following a standard normal distribution. Since r_3 is close to zero in most cases, this strategy usually produces only small adjustments near the current dominant

region. In a few cases, however, a larger r_3 may cause a substantial displacement, enabling the individual to escape the current search region and move toward a new potential solution space. Therefore, repeated parasitism has the dual functions of local intensification and global jumping.

Optimal parasitism focuses on further guiding individuals toward the vicinity of the current optimal solution, and its update method is shown in Equation (6):

$$X_i^{pl} = X_{best}^t + \lambda(\text{rand} \cdot X_{best}^t - \text{rand} \cdot X_i^{so}), \quad i = 1, 2, \dots, N \quad (6)$$

where λ is the adaptive contraction factor, and its mathematical expression is shown in Equation (7):

$$\lambda = r_4 \left[\left(\frac{t}{T_{\max}} \right)^2 - 2 \left(\frac{t}{T_{\max}} \right) + 1 \right] \quad (7)$$

where $r_4 \in [0, 1]$ is a uniformly distributed random number. It can be seen that λ is relatively large in the early iterations and gradually approaches 0 in the later iterations. This indicates that the algorithm has a larger search step size in the early stage, which is beneficial for maintaining population diversity; in the later stage, the search range is gradually narrowed to improve the exploitation accuracy of the optimal region, ultimately promoting stable convergence of the algorithm toward the global optimum.

2.5.2. Support Vector Regression (SVR)

SVR [31] is a supervised regression method based on statistical learning theory. It maps input variables into a high-dimensional feature space through a kernel function and seeks the optimal regression hyperplane under the constraint of the ε -insensitive loss function, thereby achieving high prediction accuracy while controlling model complexity. SVR is particularly suitable for small-sample, high-dimensional, and nonlinear problems, and therefore has strong applicability in hyperspectral regression studies. Kernel function and parameter selection have a substantial influence on prediction accuracy. The kernel function defines a way to map input data into a high-dimensional space. In this study, the radial basis function (RBF) kernel was used to construct the SVR model, and the penalty parameter C , kernel parameter γ , and insensitive loss parameter ε were optimized to further improve the ability of the model to characterize complex nonlinear relationships in hyperspectral data and enhance regression prediction accuracy.

2.5.3. CFO-SVR-Based Joint Optimization Modeling Framework

To improve the accuracy and stability of the ginseng seed emergence rate prediction, this study constructed a joint optimization modeling framework based on the Caterpillar Fungus Optimizer (CFO) and SVR. This method takes spectral wavelength feature selection and SVR parameter optimization as its core components and incorporates both into the same optimization process, thereby achieving coordinated optimization of the feature subset and model parameters.

As shown in Figure 3, data reading and feature extraction were first performed for the training set and test set to obtain the feature matrix X and target variable y . Correlation analysis was then used to preselect the original spectral variables to reduce the interference of redundant information in modeling. Subsequently, CFO was used to jointly encode and iteratively optimize the selection status of candidate features and the key SVR parameters C , γ , and ε . During the optimization process, the root mean square error of five-fold cross-validation ($RMSE_{CV}$) was used as the main evaluation metric, and a penalty term for the feature selection ratio was introduced to construct the fitness function, which is expressed as follows:

$$Fitness = RMSE_{CV} + \alpha \times \frac{N_s}{q} \quad (8)$$

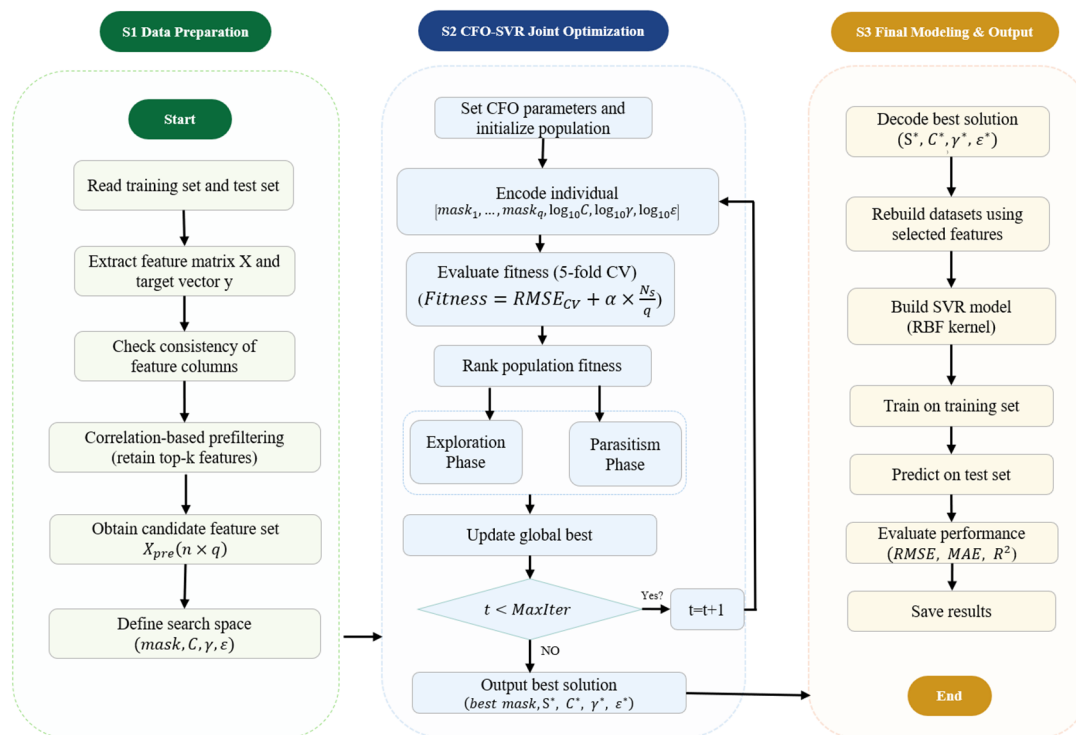


Figure 3. Framework of the CFO-SVR joint optimization modeling approach. The superscript “*” denotes the optimal solution obtained after CFO iteration; S^* represents the optimal feature subset, and C^* , γ^* , and ε^* represent the optimized SVR hyperparameters.

In Equation (8), $RMSE_{CV}$ represents the average root mean square error obtained from five-fold cross-validation, N_s represents the number of features selected by the current individual, q represents the total number of candidate features, and α is the penalty coefficient. This fitness function minimizes prediction error while constraining the number of selected features, thereby preventing excessive redundant wavelengths from participating in modeling and enabling the model to achieve a balance between prediction accuracy and structural simplicity.

After the maximum number of iterations was reached, the global optimal feature subset and the corresponding optimal SVR parameters were output. The input matrices of the training set and test set were then reconstructed based on the optimal feature subset, and the final prediction model was established. This joint optimization framework can effectively reduce redundant wavelengths and improve parameter optimization efficiency, thereby enhancing the accuracy and generalization ability of the model for predicting ginseng seed emergence rate.

2.5.4. SHAP Interpretability Analysis

SHAP (Shapley Additive exPlanations) is a commonly used model interpretation method [32]. Its basic principle is to decompose the model prediction into the sum of the contributions of each input feature based on the Shapley value from cooperative game theory, thereby quantifying the influence of different features on the prediction results. Compared with traditional analysis methods that focus only on model prediction accuracy, SHAP can provide interpretability information at both global and local levels. It can not only evaluate the importance ranking of features but also reveal the positive or negative effects of features on model outputs. Owing to its good interpretability and strong general applicability, SHAP has been widely used for the interpretability analysis of complex machine learning models.

2.6. Seed Vigor Evaluation Indicators

After the planting experiment was completed, seed emergence of each ginseng treatment group was recorded under uniform cultivation conditions. ER, EE, and EI were used as evaluation indicators [33–35] of seed vigor level for subsequent construction and performance evaluation of the hyperspectral-based seed vigor detection model. ER was calculated according to Equation (9).

$$G = \frac{M_1}{M} \times 100\% \quad (9)$$

In Equation (9), G represents the ER, M_1 represents the total number of emerged seedlings in a single sample, and M represents the total number of experimental seeds in a single sample.

$$F = \frac{H_1}{M} \times 100\% \quad (10)$$

In Equation (10), F represents EE, H_1 represents the cumulative number of emerged seeds at the peak emergence period (May 3 for MCGS and May 5 for AGS), and M represents the total number of experimental seeds in a single sample.

$$GI = \sum \left(\frac{E_t}{D_t} \right) \quad (11)$$

In Equation (11), (GI) represents the EI, (E_t) represents the number of newly and normally emerged seeds on day (t), and (D_t) represents the number of days after sowing on day (t).

2.7. Model Performance Evaluation Metrics

In model evaluation, the coefficient of determination (R^2), root mean square error (RMSE), and mean absolute error (MAE) were used to evaluate the model fit and prediction performance [36]. Among these metrics, R^2 was used to characterize the degree of fit between the predicted and actual values, whereas RMSE and MAE were used to measure model prediction errors. An R^2 value closer to 1 indicates better model fitting performance, while smaller RMSE and MAE values indicate lower prediction errors. The calculation formulas for RMSE and MAE are shown in Equations (12) and (13), respectively:

$$RMSE = \sqrt{\frac{1}{N} \sum_{i=1}^N (y_i - \hat{y}_i)^2} \quad (12)$$

$$MAE = \frac{1}{N} \sum_{i=1}^N |y_i - \hat{y}_i| \quad (13)$$

where y_i is the actual value of the i -th sample, $\hat{y}_i$ is the predicted value of the i -th sample, and N is the number of samples.

3. Results

3.1. Effects of Laser Treatment on Seed Vigor Indicators of Ginseng

As shown in Tables 1 and 2, different laser treatments markedly affected the emergence performance of MCGS. Compared with CK, several laser treatments significantly increased the ER, EE, and EI of MCGS. Among them, blue light treatment for 5 min showed the best overall performance, with ER reaching $100.0 \pm 0.0\%$, representing an increase of 40.4 percentage points compared with CK ($59.6 \pm 3.5\%$). The EE and EI reached $90.9 \pm 3.0\%$ and 3.057 ± 0.044 , respectively, both of which were significantly higher than those of CK ($44.4 \pm 4.6\%$ and 1.714 ± 0.129). In addition to blue light treatment for 5 min, red–blue

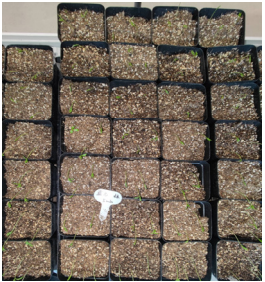
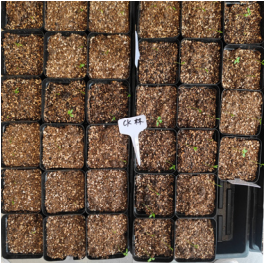
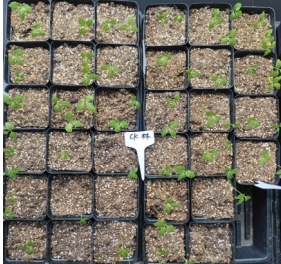
3:1 for 5 min, blue light for 15 min, red–blue 6:1 for 15 min, red light for 30 min, blue light for 30 min, and red–blue 3:1 for 30 min also showed significantly higher ER and EI than CK, indicating that these treatments had a certain promoting effect on the emergence vigor of MCGS. In terms of EE, blue light for 5 min, red–blue 3:1 for 5 min, blue light for 15 min, red–blue 6:1 for 15 min, and blue light for 30 min were all significantly higher than CK, suggesting that these treatments not only improved the final emergence level but also enhanced early emergence capacity. Overall, MCGS responded sensitively to laser treatment, with blue light treatments generally showing better performance, especially blue light for 5 min, which performed prominently across all three emergence vigor indicators. However, some red–blue composite light treatments did not show significant promoting effects, indicating that the effect of laser treatment was closely related to the light-quality combination and irradiation time.

Table 1. Emergence vigor indicators of MCGS after different laser treatments.

Light Quality	Time (min)	ER (%) (<i>n</i> = 3)	EE (%) (<i>n</i> = 3)	EI (<i>n</i> = 3)
Red	5	69.7 ± 3.0	53.5 ± 7.6	1.979 ± 0.151
Blue	5	100.0 ± 0.0 *	90.9 ± 3.0 *	3.057 ± 0.044 *
Red–blue 3:1	5	94.9 ± 3.5 *	72.7 ± 6.1 *	2.750 ± 0.127 *
Red–blue 6:1	5	69.7 ± 3.0	42.4 ± 10.9	1.877 ± 0.077
Red	15	70.7 ± 7.6	56.6 ± 4.6	2.154 ± 0.162
Blue	15	99.0 ± 1.7 *	84.8 ± 6.1 *	3.046 ± 0.074 *
Red–blue 3:1	15	65.7 ± 6.3	43.4 ± 7.6	1.843 ± 0.195
Red–blue 6:1	15	98.0 ± 1.7 *	85.9 ± 1.7 *	2.965 ± 0.037 *
Red	30	92.9 ± 7.0 *	85.9 ± 11.5	2.927 ± 0.277 *
Blue	30	98.0 ± 1.7 *	83.8 ± 4.6 *	2.969 ± 0.039 *
Red–blue 3:1	30	89.9 ± 3.5 *	72.7 ± 8.0	2.725 ± 0.171 *
Red–blue 6:1	30	63.6 ± 5.2	54.5 ± 9.1	1.950 ± 0.191
CK		59.6 ± 3.5	44.4 ± 4.6	1.714 ± 0.129

Note: Values are presented as the mean ± standard deviation of three biological replicates. * indicates a significant difference compared with CK after Welch's *t*-test followed by Holm correction at ($p < 0.05$); values without an asterisk indicate no significant difference.

Table 2. Comparison of cumulative emergence dynamics between the blue light treatment for 5 min and CK in MCGS.

Time (9–25 d)				
	9 d	15 d	20 d	25 d
Blue light 5 min				
CK				

As shown in Figure 4a–c, the cumulative emergence process of MCGS generally followed an approximately S-shaped growth pattern, with differences among treatment groups in emergence initiation, the rapid growth stage, and the final stable level. Compared with CK, some laser treatments shifted the cumulative emergence curves forward and showed higher increasing rates during the rapid growth stage, indicating that appropriate laser treatment may help improve the early emergence capacity and final emergence level of MCGS. Among them, the cumulative emergence curve of blue light treatment for 5 min started earlier, increased more rapidly, and ultimately reached a higher emergence level, which was consistent with the higher ER, EE, and EI of this treatment shown in Table 1. Overall, the cumulative emergence dynamics further indicated that MCGS showed a clear response to laser treatment, and short-duration blue light treatment exhibited a relatively prominent overall effect.

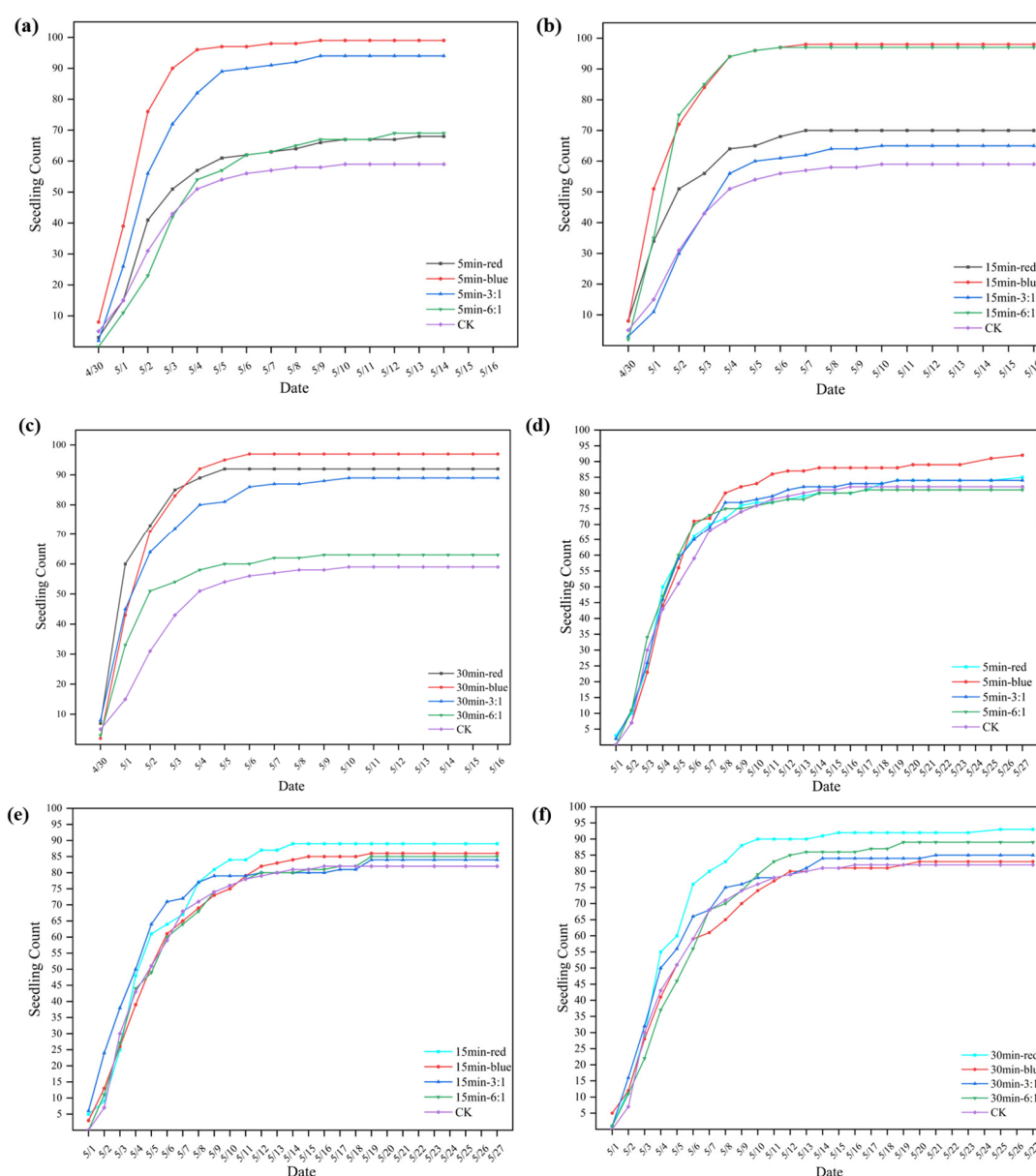


Figure 4. Dynamic changes in the number of emerged seeds of MCGS and AGS: (a) MCGS, 5 min; (b) MCGS, 15 min; (c) MCGS, 30 min; (d) AGS, 5 min; (e) AGS, 15 min; (f) AGS, 30 min.

As shown in Tables 3 and 4, the ER, EE, and EI of AGS showed certain numerical changes under different laser treatments, but none of the laser treatments differed sig-

nificantly from CK. Compared with MCGS, the differences among AGS treatments were relatively small, indicating that AGS responded more moderately to the laser treatment parameters used in this study. Numerically, red light treatment for 30 min showed the highest ER, reaching $93.9 \pm 3.0\%$, which was higher than that of CK ($82.8 \pm 4.6\%$). The ER values of blue light for 5 min, red light for 15 min, and red–blue 6:1 for 30 min were also higher than those of CK, but these differences did not reach a significant level after Welch's *t*-test followed by Holm multiple correction. Therefore, treatments such as red light for 30 min only indicated an increasing trend in the ER of AGS, rather than a significant promoting effect. In terms of early emergence capacity, red–blue 3:1 for 15 min showed the highest EE value, reaching $64.6 \pm 1.7\%$, while red light for 30 min showed the highest EI value, reaching 2.288 ± 0.108 . Overall, AGS showed certain numerical increasing trends under some laser treatments, but no statistically significant improvement was observed, suggesting that its response to the laser parameters used in this study was weaker than that of MCGS.

Table 3. Emergence vigor indicators of AGS after different laser treatments.

Light Quality	Time (min)	ER (%) (<i>n</i> = 3)	EE (%) (<i>n</i> = 3)	EI (<i>n</i> = 3)
Red	5	85.9 ± 9.3	59.6 ± 9.3	2.058 ± 0.206
Blue	5	92.9 ± 9.7	56.6 ± 18.5	2.167 ± 0.339
Red–blue 3:1	5	84.8 ± 12.1	59.6 ± 12.2	2.057 ± 0.298
Red–blue 6:1	5	81.8 ± 8.0	60.6 ± 13.2	2.031 ± 0.216
Red	15	89.9 ± 7.6	61.6 ± 20.2	2.164 ± 0.296
Blue	15	86.9 ± 6.3	51.5 ± 8.0	2.046 ± 0.198
Red–blue 3:1	15	84.8 ± 6.1	64.6 ± 1.7	2.151 ± 0.141
Red–blue 6:1	15	85.9 ± 11.5	49.5 ± 11.5	2.011 ± 0.296
Red	30	93.9 ± 3.0	60.6 ± 9.1	2.288 ± 0.108
Blue	30	83.8 ± 3.5	51.5 ± 5.2	1.991 ± 0.063
Red–blue 3:1	30	85.9 ± 1.7	56.6 ± 8.7	2.092 ± 0.028
Red–blue 6:1	30	89.9 ± 4.6	46.5 ± 6.3	2.064 ± 0.089
CK		82.8 ± 4.6	51.5 ± 8.0	1.982 ± 0.136

Note: Values are presented as the mean $\pm$ standard deviation of three biological replicates.

As shown in Figure 4d–f, the cumulative emergence curves of AGS also followed an approximately S-shaped growth pattern, but the degree of curve separation among different treatment groups was relatively small, and the overall differences were less pronounced than those observed in MCGS. This indicates that AGS itself had a relatively high basal emergence capacity, and laser treatment mainly resulted in slight numerical fluctuations during the emergence process rather than significant promotion. Some treatments, such as red light for 30 min, showed a relatively high final emergence level during the later stage of cumulative emergence, while red–blue 3:1 for 15 min showed certain advantages during the early emergence stage. However, these changes mainly reflected trend-level differences and were not sufficient to demonstrate a significant improvement in the emergence vigor of AGS. Further studies are needed to increase the number of replicates, include more seed batches, and further optimize laser parameters to verify the stability of these trends.

A three-way ANOVA was further conducted for the laser-treated groups to clarify the effects of seed type, light quality, irradiation time, and their interactions on the emergence vigor of ginseng seeds. As shown in Table 5, light quality had significant effects on ER, EE, and EI, indicating that light quality was an important factor affecting the emergence vigor of ginseng seeds. Seed type had significant effects on EE and EI, whereas its main effect on ER did not reach a significant level. The main effect of irradiation time was not significant for any of the three emergence vigor indicators, suggesting that its influence on emergence vigor was relatively limited when irradiation time was considered alone. In

terms of interactions, seed type $\times$ light quality, light quality $\times$ irradiation time, and seed type $\times$ light quality $\times$ irradiation time had significant effects on ER, EE, and EI. Among them, seed type $\times$ irradiation time had a significant effect only on EE, whereas its effects on ER and EI did not reach a significant level. Notably, the three-way interaction of seed type $\times$ light quality $\times$ irradiation time was significant for all three indicators, indicating that the effects of light quality and irradiation time on the emergence vigor of ginseng seeds depended on seed type. Based on the above three-way interaction and the changes in treatment means shown in Tables 1 and 3, the response patterns of different seed types to the combinations of light quality and irradiation time were inconsistent. These results indicate that the effects of laser treatment on the emergence vigor of ginseng seeds cannot be explained by a single light quality or a single irradiation time alone, but should be interpreted by comprehensively considering the interactions among seed type, light-quality combination, and irradiation time. In other words, the same laser treatment parameter may produce different effects in different seed types, which is also an important reason for the differences in emergence responses between MCGS and AGS.

Table 4. Comparison of cumulative emergence dynamics between the red light treatment for 30 min and CK in AGS.

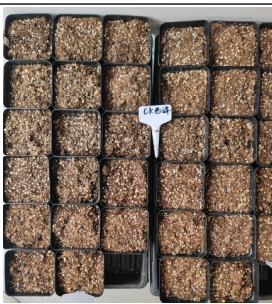
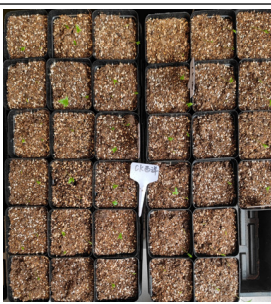
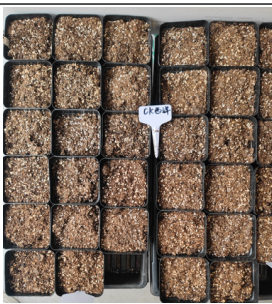
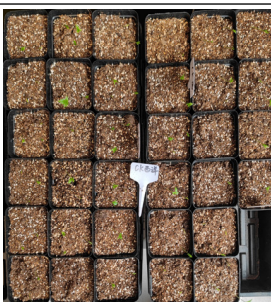
		Time (10–36 d)			
		10 d	18 d	28 d	36 d
Red light 30 min					
					
CK					

Table 5. Three-way ANOVA results for seed emergence vigor indicators.

Source	ER	EE	EI
Seed type	3.750	29.695 ***	96.719 ***
Light quality	12.822 ***	5.835 **	15.097 ***
Irradiation time	1.007	0.312	2.977
Seed type $\times$ Light quality	12.259 ***	9.795 ***	17.092 ***
Seed type $\times$ Irradiation time	0.063	3.590 *	1.750
Light quality $\times$ Irradiation time	11.187 ***	4.823 ***	13.242 ***
Seed type $\times$ Light quality $\times$ Irradiation time	9.319 ***	6.859 ***	12.246 ***

Note: Values are F values obtained from three-way ANOVA. *, **, and *** indicate significant effects of the factor or interaction at $p < 0.05$, $p < 0.01$, and $p < 0.001$, respectively. “ $\times$ ” indicates interaction between factors. CK was not included in the three-way ANOVA because it had no light-quality or irradiation-time levels.

Overall, under the treatment conditions used in this study, the effects of laser treatment on the emergence vigor of the two types of ginseng seeds showed clear seed-type differences and parameter dependence. Significant promoting effects were mainly observed in MCGS, especially under blue light treatment for 5 min, which showed prominent performance in terms of ER, EE, and EI. In contrast, AGS did not show statistically significant improvement in vigor under the laser treatment conditions used in this study, although some treatments showed numerical increasing trends. These results indicate that laser treatment does not exert a universal promoting effect on all ginseng seeds, and its effects are jointly influenced by seed type, light-quality combination, and irradiation time. Therefore, laser treatment parameters should be further optimized for different types of ginseng seeds to achieve more stable and targeted emergence-promoting effects.

3.2. SVR-Based Hyperspectral Prediction of Ginseng Seed Emergence Rate

3.2.1. Hyperspectral Data Preprocessing Results

As shown in Figure 5, relatively obvious noise and baseline drift were observed in the short-wavelength region (500–800 nm) of the unprocessed raw spectral data (RAW). These problems may have resulted from random noise of the instrument itself, scattering effects of the light source, and slight deviations in probe angle during data acquisition. After Savitzky–Golay (SG) smoothing, the high-frequency random noise in the curves was effectively reduced, and the overall spectral profiles became smoother, while the characteristic spectral differences among ginseng seeds from different laser treatment groups were well preserved. Multiplicative scatter correction (MSC) clearly reduced the scattering effects caused by differences in sample surface roughness and measurement angle, resulting in a more consistent spectral baseline. First derivative (1st Der) processing enhanced subtle spectral features and enabled some overlapping absorption peaks to be distinguished, whereas second derivative (2nd Der) processing further improved spectral peak resolution while reducing the influence of baseline drift. Standard normal variate (SNV) transformation could also correct scattering effects and effectively reduce overall reflectance differences among different samples. Wavelet denoising (DT) showed a relatively strong ability to suppress random noise and improved curve stability while retaining the main spectral features. On this basis, combined preprocessing methods further improved data quality. The combination of SG and derivatives, namely S-G 1st and S-G 2nd, highlighted detailed spectral features while reducing noise, and the enhancement effect of S-G 2nd on weak spectral absorption peaks was particularly evident.

Before modeling, the optimal preprocessing method was selected from multiple spectral preprocessing approaches to improve the accuracy and robustness of subsequent model prediction. For this purpose, the raw spectra were compared with spectra processed using several commonly used preprocessing methods, and SVR was used for model-based evaluation. As shown in Table 6, compared with the raw spectra, the prediction performance of SVR models constructed using preprocessed spectra improved to varying degrees. This indicates that appropriate spectral preprocessing can effectively reduce spectral noise, extract useful information, and improve model fitting and generalization performance. Among all preprocessing methods, SG combined with the second derivative (S-G 2nd) showed the best performance, with an average R_p^2 , $RMSEP$, and $MAEP$ values of 0.8881, 3.6622, and 1.6839, respectively. Therefore, S-G 2nd was selected as the optimal preprocessing method for subsequent spectral data analysis and modeling.

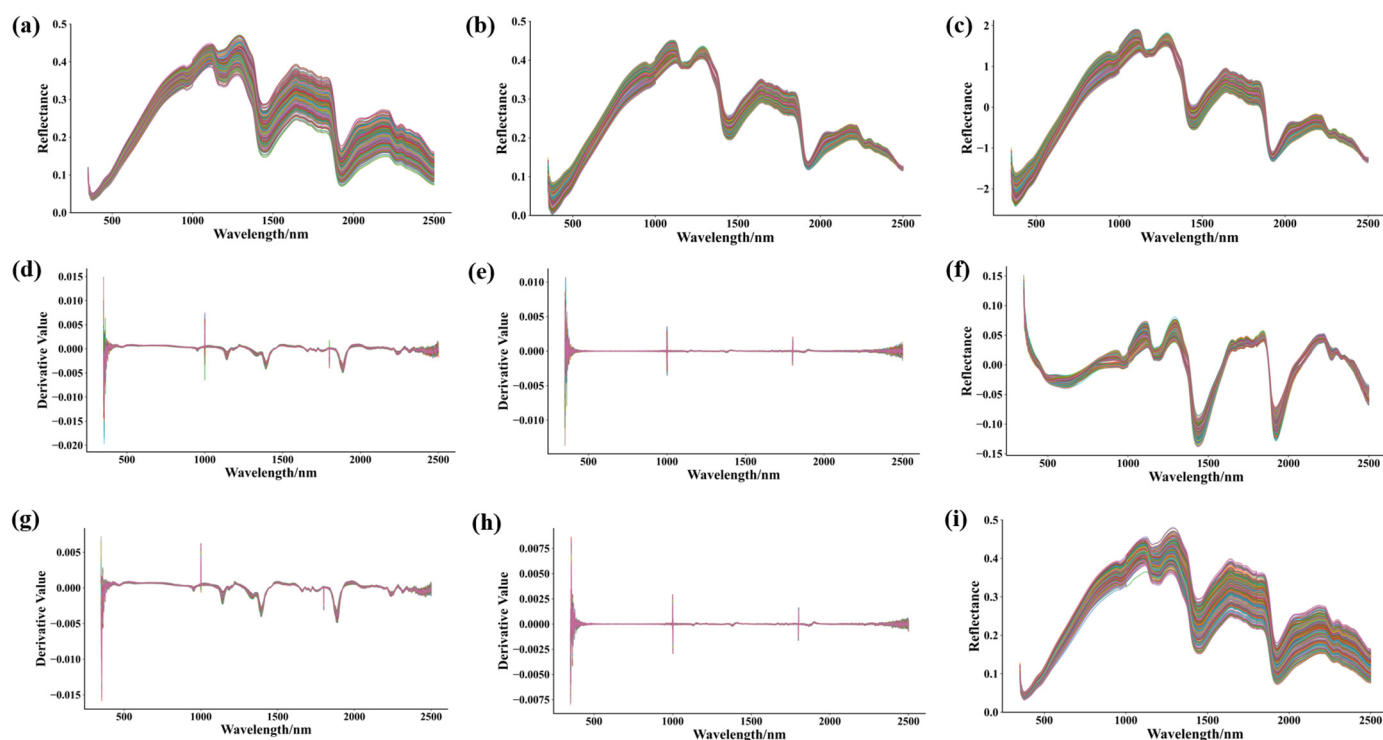


Figure 5. Visualization of all preprocessed spectra. (a) SG; (b) MSC; (c) SNV; (d) 1st deriv; (e) 2nd deriv; (f) DT; (g) SG + 1st; (h) SG + 2nd; (i) RAW.

Table 6. Comparison of SVR model performance for predicting ginseng seed emergence rate under different spectral preprocessing methods.

Preprocessing Method	Testing Set		
	R_p^2	RMSEP	MAEP
S-G 2nd	0.8881	3.6622	1.6839
S-G 1st	0.7985	4.8787	2.5342
1st Der	0.5974	7.0288	4.3276
2nd Der	0.4748	8.0828	5.3684
DT	0.2916	9.3151	5.7735
SG	0.2747	9.4479	6.3116
Raw	0.2261	9.7813	6.5863
MSC	0.1412	10.3016	6.8823
SNV	0.1193	10.4351	6.9831

3.2.2. Feature Selection Results

Because hyperspectral data are inherently characterized by high dimensionality, high redundancy, and severe multicollinearity, directly using all wavelengths for modeling can easily lead to excessive model complexity, low computational efficiency, and overfitting. In this study, two typical spectral feature selection methods, Random Frog (RF) and competitive adaptive reweighted sampling (CARS), together with the Caterpillar Fungus Optimizer (CFO), were used to select the most representative key wavelengths for emergence rate prediction from different perspectives, thereby reducing redundancy and improving model accuracy and robustness. Feature extraction was performed based on the spectral data preprocessed by SG combined with the second derivative (S-G 2nd).

The Random Frog (RF) algorithm was used for feature selection of spectral variables. In this study, the number of RF iterations was set to 2000, and the initial variable subset size was set to 10% of the total number of spectral variables. Variable selection was

performed based on the training set to avoid the involvement of test set information in the feature selection process. According to the variable selection probabilities obtained by the RF algorithm, the wavelength variables were ranked in descending order of selection probability. PLS models were then established by sequentially increasing the number of feature variables, and the root mean square error of five-fold cross-validation ($RMSE_{CV}$) on the training set was used to compare subsets with different numbers of variables and determine the optimal number of feature variables. The results showed that when the number of feature variables was 60, the model obtained the minimum $RMSE_{CV}$, with a value of 2.5514. Therefore, this variable subset was determined as the optimal feature combination selected by RF. The RF-based feature wavelength selection results are shown in Figure 6, where Figure 6a shows the distribution of variable selection probabilities, and Figure 6b shows the distribution of the selected feature wavelengths on the mean spectral curve.

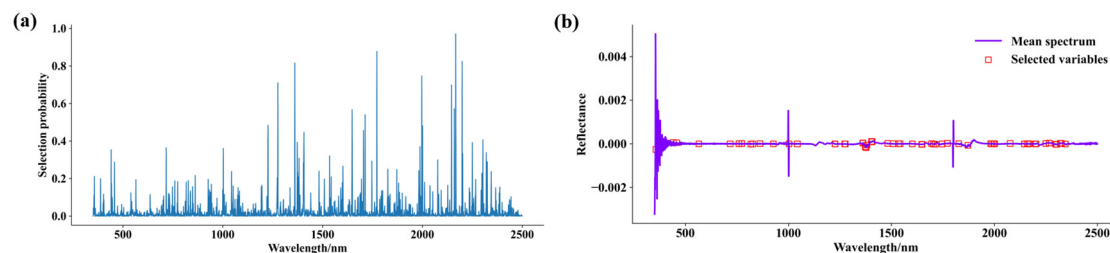


Figure 6. Feature wavelength selection results based on the RF algorithm: (a) distribution of variable selection probabilities; (b) distribution of selected feature wavelengths.

Competitive adaptive reweighted sampling (CARS) was used for feature selection of spectral bands, and the results are shown in Figure 7. CARS establishes PLS submodels based on Monte Carlo sampling and gradually reduces the number of retained variables using an exponentially decreasing function (EDF). As shown in Figure 7a, with the increase in sampling iterations, the number of retained variables decreased rapidly and then gradually leveled off, indicating that a large number of redundant variables were progressively eliminated during the iterative process. After each sampling iteration, five-fold cross-validation on the training set was used to calculate $RMSE_{CV}$ to evaluate the modeling performance of the corresponding variable subset. Figure 7b shows the variation curve of $RMSE_{CV}$ with sampling iterations. The $RMSE_{CV}$ first decreased and then increased with the sampling iterations, reaching the minimum value of 2.6444 at the 26th iteration, indicating that the variable subset corresponding to this iteration had the best modeling performance. Figure 7c shows the regression coefficient path, where the dashed line corresponds to the 26th iteration, consistent with the position of the minimum $RMSE_{CV}$. Finally, CARS selected 61 feature variables, which were used as the feature band combination for subsequent model construction. Figure 7d shows the distribution of the final selected feature bands on the mean spectral curve.

The CFO algorithm was used to select spectral feature wavelengths. First, preliminary screening was performed using the absolute correlation coefficients between the spectral variables and the target values in the training set, and the top 120 variables with relatively high correlations were retained as the candidate feature set. Subsequently, with the feature selection mask as the optimization object, the CFO algorithm was employed to search for the optimal subset of candidate variables. During feature subset evaluation, a fitness function was constructed by comprehensively considering the prediction error of the candidate feature subset and a penalty term for the number of variables, so as to control the number of selected variables while ensuring feature effectiveness. In this study, the population size

of the CFO algorithm was set to 20, the maximum number of iterations was set to 60, and the penalty coefficient for feature number was set to 0.06.

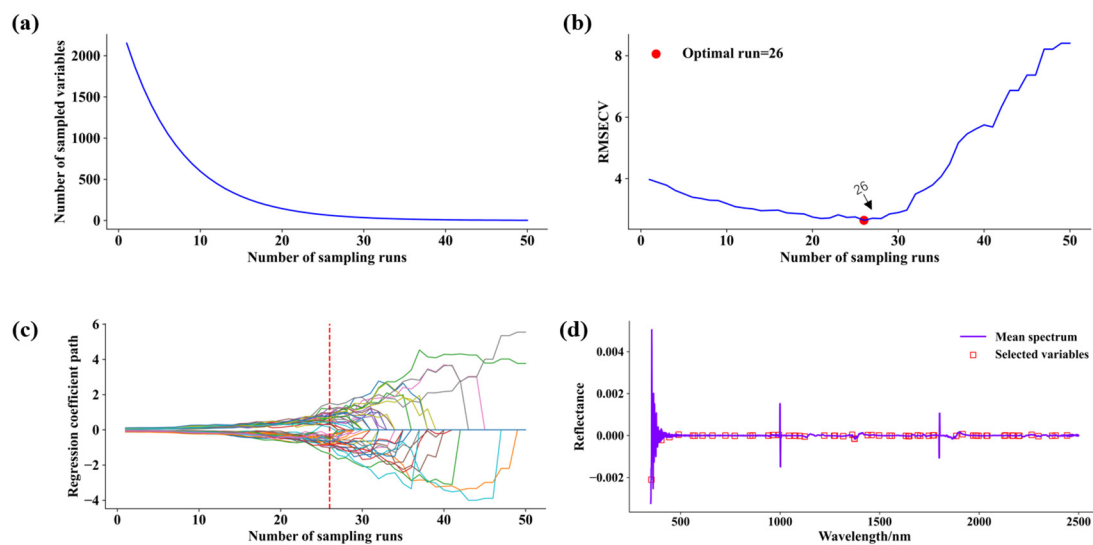


Figure 7. Feature wavelength selection results based on the CARS algorithm: (a) relationship between sampling iterations and the number of sampled variables; (b) variation in $RMSE_{CV}$ with sampling iterations; (c) regression coefficient path; (d) distribution of selected feature wavelengths.

According to the CFO iteration results, the algorithm showed a clear downward trend in fitness within the first 20 iterations, with the optimal fitness value decreasing from 2.7090 to 2.5552, indicating that the CFO algorithm could rapidly search for a relatively optimal feature combination region within a small number of iterations. After 20 iterations, the optimal fitness value remained basically stable, indicating that the algorithm had gradually converged and that further iterations produced no obvious improvement in fitness. Ultimately, the CFO algorithm selected 57 feature wavelengths from the 120 candidate variables, corresponding to a feature retention rate of 47.5%. This indicates that the method can effectively retain the major spectral information related to the target variable while reducing the number of input variables. The results of CFO feature wavelength selection are shown in Figure 8, where Figure 8a presents the CFO iterative convergence curve, and Figure 8b shows the distribution of the selected feature wavelengths on the mean spectral curve.

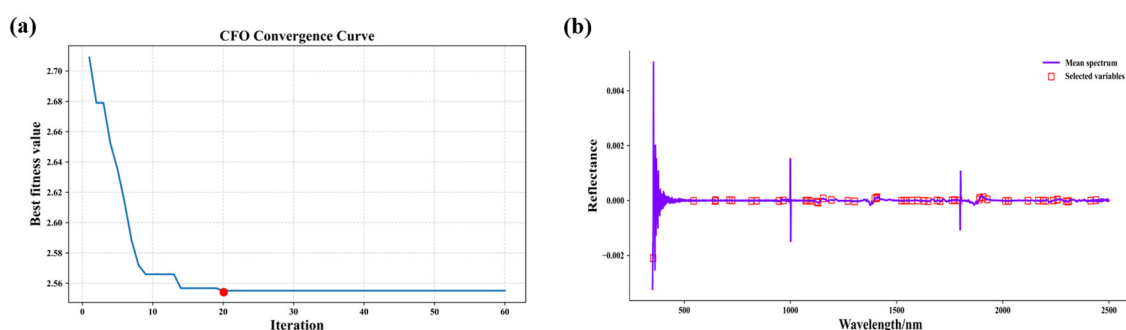


Figure 8. Feature wavelength selection results based on the CFO algorithm: (a) CFO iterative convergence curve; (b) selected feature wavelengths after processing by the CFO algorithm.

As shown in Table 7, under the same SVR modeling conditions, different feature selection methods had certain effects on model prediction performance. Compared with RF and CARS, the SVR model established using the feature bands selected by CFO achieved the best prediction performance on the test set, with R_p^2 of 0.9383, and both $RMSE_p$ and

MAE_p were lower than those of the other two methods. This result indicates that the feature subset selected by the CFO could more effectively retain spectral information related to the prediction of ginseng seed ER, thereby improving the test-set prediction performance of the model within the experimental data range.

Table 7. Modeling results for predicting ginseng seed emergence rate.

Model	Feature Selection	Training Set			Testing Set		
		R ²	RMSE	MAE	R _p ²	RMSE _p	MAE _p
SVR	RF	0.9738	1.8088	0.7938	0.9219	3.1495	1.5256
	CARS	0.9739	1.8052	0.7528	0.9203	3.1804	1.5044
	CFO	0.9652	2.0833	0.8123	0.9383	2.7993	1.4264

From the perspective of algorithmic mechanisms, RF and CARS are essentially conventional variable selection methods. Their advantage lies in their ability to rapidly identify locally significant wavelengths, but they focus more on the contribution of individual variables or local competitive relationships, with relatively limited consideration of the strong correlations, information redundancy, and synergistic effects among variables that are commonly present in hyperspectral data. In contrast, the CFO adopts a swarm intelligence optimization strategy and performs overall optimization of feature subsets in a larger combinatorial search space. It can not only avoid premature convergence to local optima through global exploration, but also dynamically balance the relationship between feature compression and prediction performance during iteration. This also means that the core value of CFO is not limited to selecting individual important wavelengths; rather, it is more capable of mining complementary relationships among wavelengths and ultimately selecting a wavelength combination with stronger adaptability and the best overall performance. Therefore, although all three methods achieved feature dimensionality reduction, CFO obtained a higher coefficient of determination and lower prediction errors on the test set, indicating that it has stronger robustness and better generalization potential in hyperspectral feature selection.

3.3. Prediction Results and Analysis of Ginseng Seed Emergence Rate Based on the CFO-SVR Joint Optimization Model

The prediction performance of the SVR model is greatly affected by hyperparameters such as the penalty factor (C), kernel parameter (γ), and insensitive loss parameter (ϵ). Traditional manual parameter tuning or grid search methods have limitations such as low search efficiency, dependence on empirical parameter combinations, and difficulty in simultaneously considering global search and local optimization. To further improve the prediction accuracy of the ginseng seed emergence rate, this study jointly optimized the hyperparameters of the SVR model after completing hyperspectral feature selection based on the CFO algorithm and constructed a CFO-SVR joint optimization modeling framework. During modeling, the sample data were divided into a training set and a test set at a ratio of 7:3. The training set was used for model training, feature selection, and parameter optimization, whereas the test set was used only for the final evaluation of model prediction performance.

3.3.1. Analysis of Emergence Rate Prediction Results

To verify the effectiveness and superiority of the CFO-SVR joint optimization model in predicting ginseng seed emergence rate, GA-SVR and PSO-SVR were selected as comparison models in this study. A comprehensive analysis was conducted in terms of fitness iteration changes, model fitting ability, prediction accuracy, and generalization performance.

To ensure the fairness of the comparison among different swarm intelligence optimization algorithms, CFO-SVR, GA-SVR, and PSO-SVR all adopted the same joint optimization framework. Specifically, on the same candidate wavelength set, the wavelength selection mask and the three key hyperparameters of the SVR model, namely C , γ , and ε , were optimized simultaneously. The three models used the same parameter search ranges, five-fold cross-validation-based fitness function, population size, maximum number of iterations, and feature number penalty coefficient.

The variation pattern of fitness during iteration showed that all three optimization algorithms could effectively reduce the fitness value of the joint optimization model and achieve simultaneous search in the parameter space and feature subspace. Among them, CFO-SVR showed a clear convergence advantage. The fitness value decreased rapidly in the early stage of iteration, and convergence was completed within far fewer iterations than those required by the other two algorithms, indicating significantly higher optimization efficiency. In contrast, PSO-SVR tended to fall into local optima during the middle and late stages of iteration, with the decrease in fitness gradually stagnating. Although GA-SVR had continuous search ability, its convergence process was slow, resulting in a higher overall computational cost. These results verify that the CFO algorithm is more suitable for the complex search scenario involving simultaneous optimization of features and parameters in this study and can efficiently and stably construct an optimal modeling framework.

As shown in Table 8, among the different joint optimization models, CFO-SVR achieved favorable prediction performance. Compared with GA-SVR and PSO-SVR, CFO-SVR obtained the highest R_p^2 (0.9517) and the lowest $RMSE_p$ and MAE_p on the test set, indicating better consistency between the predicted and measured values. This result suggests that the CFO algorithm had good optimization capability during the simultaneous optimization of feature selection and SVR parameters, thereby improving the prediction performance of the model for ginseng seed ER within the experimental data range.

Table 8. Comparison of prediction performance of different joint optimization models for ginseng seed emergence rate.

Joint Optimization	Training Set			Testing Set		
	R^2	RMSE	MAE	R_p^2	$RMSE_p$	MAE_p
GA-SVR	0.9778	1.6631	0.5195	0.9399	2.7611	1.2324
PSO-SVR	0.9758	1.7374	0.5571	0.9319	2.9410	1.2797
CFO-SVR	0.9820	1.4978	0.4221	0.9517	2.4766	1.0730

To further exclude the potential influence of inherent emergence differences between MCGS and AGS on model performance in mixed-data modeling, the CFO-SVR model was independently trained and tested using the MCGS and AGS datasets. As shown in Table 9, the R_p^2 values of MCGS-only CFO-SVR and AGS-only CFO-SVR on the test set were 0.9474 and 0.9397, respectively, both close to that of the global CFO-SVR model (0.9517), indicating that CFO-SVR still maintained good prediction ability within a single seed type. This result suggests that the high prediction performance of the global CFO-SVR model was not solely caused by the inherent emergence differences between the two seed types, but that the model could capture the association between hyperspectral features and ER within the same seed type. From the scatter plot distribution in Figure 9, further comparison showed that the scatter points of both the training and test sets in the global CFO-SVR model shown in Figure 9a were more concentrated, with smaller deviations between the predicted points and the ideal fitting line, especially in the medium-to-high ER range, where good fitting consistency was observed. In contrast, some prediction points of GA-SVR and PSO-SVR in Figure 9b,c deviated more obviously from the ideal fitting line,

indicating relatively larger prediction errors. Meanwhile, Figure 9d,e further showed that MCGS-only CFO-SVR and AGS-only CFO-SVR also achieved good prediction consistency within the single-seed-type datasets, further verifying the stability of the CFO-SVR model.

Table 9. Validation of ER prediction performance of the CFO-SVR model based on independent seed-type datasets.

Validation Strategy	Training Set			Testing Set		
	R ²	RMSE	MAE	R _p ²	RMSE _p	MAE _p
Global CFO-SVR	0.9820	1.4978	0.4221	0.9517	2.4766	1.0730
MCGS-only CFO-SVR	0.9858	1.8020	0.5189	0.9474	3.4890	1.7434
AGS-only CFO-SVR	0.9658	0.6628	0.3708	0.9397	0.8806	0.5915

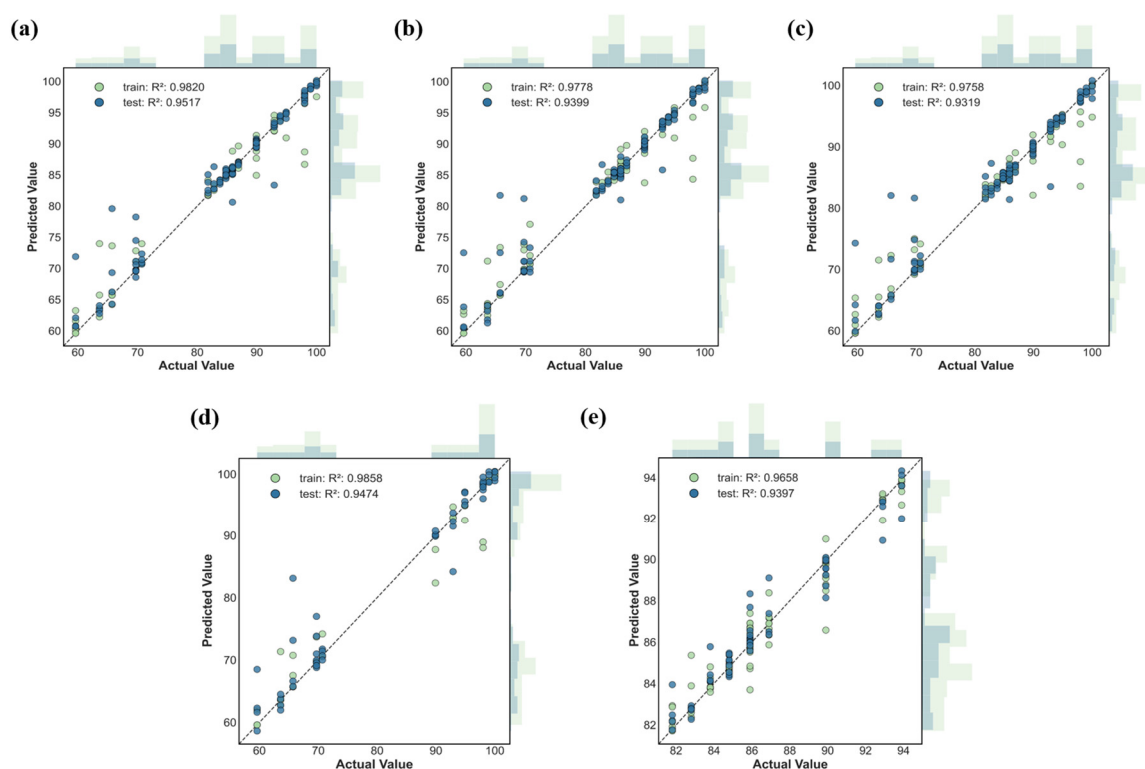


Figure 9. Scatter plots of different optimization frameworks: (a) CFO-SVR; (b) GA-SVR; (c) PSO-SVR; (d) MCGS-only CFO-SVR; (e) AGS-only CFO-SVR.

The results showed that the CFO-SVR joint optimization modeling framework proposed in this study could simultaneously achieve feature wavelength selection and SVR parameter optimization. By reducing high-dimensional spectral redundancy and noise interference, the framework improved the model fitting ability and test-set prediction performance for ginseng seed ER. Compared with GA-SVR and PSO-SVR, CFO-SVR showed faster convergence and lower prediction errors and maintained good prediction performance in the MCGS-only and AGS-only datasets, indicating that this model could effectively extract hyperspectral feature information related to ER. Overall, CFO-SVR exhibited good ER prediction ability within the experimental data range and can provide a methodological reference for rapid and non-destructive evaluation of ginseng seed vigor.

3.3.2. Analysis of Spectral Response Characteristics of Ginseng Seeds

In this study, the feature wavelengths selected by CFO-SVR were mapped onto the mean spectral curves of ginseng seeds, as shown in Figure 10a,b. The results showed that the

selected feature wavelengths were not randomly distributed across the full spectral range, but were mainly concentrated in spectrally sensitive regions of the shortwave infrared range, particularly within 1080–1250 nm, 1370–1470 nm, 1640–1770 nm, 1880–1960 nm, and 2200–2350 nm. To further analyze the physicochemical significance of these feature wavelengths, their main functional group assignments and possible corresponding physicochemical characteristics were summarized. As shown in Table 10, the feature wavelengths selected by the CFO-SVR model were mainly associated with the absorption of functional groups such as O–H, C–H, and N–H and may reflect changes in internal seed components and status, including water status, lipids, proteins, starch, and sugars. Among them, 1370–1470 nm and 1880–1960 nm may be mainly related to water and bound water status, 1080–1250 nm and 1640–1770 nm may be mainly associated with lipids and membrane lipid status, and 2200–2350 nm may reflect information on complex components such as proteins, starch, and sugars. These results indicate that these feature wavelengths contributed to model prediction and may also provide a reference for analyzing differences in the internal physicochemical status of ginseng seeds after laser treatment.

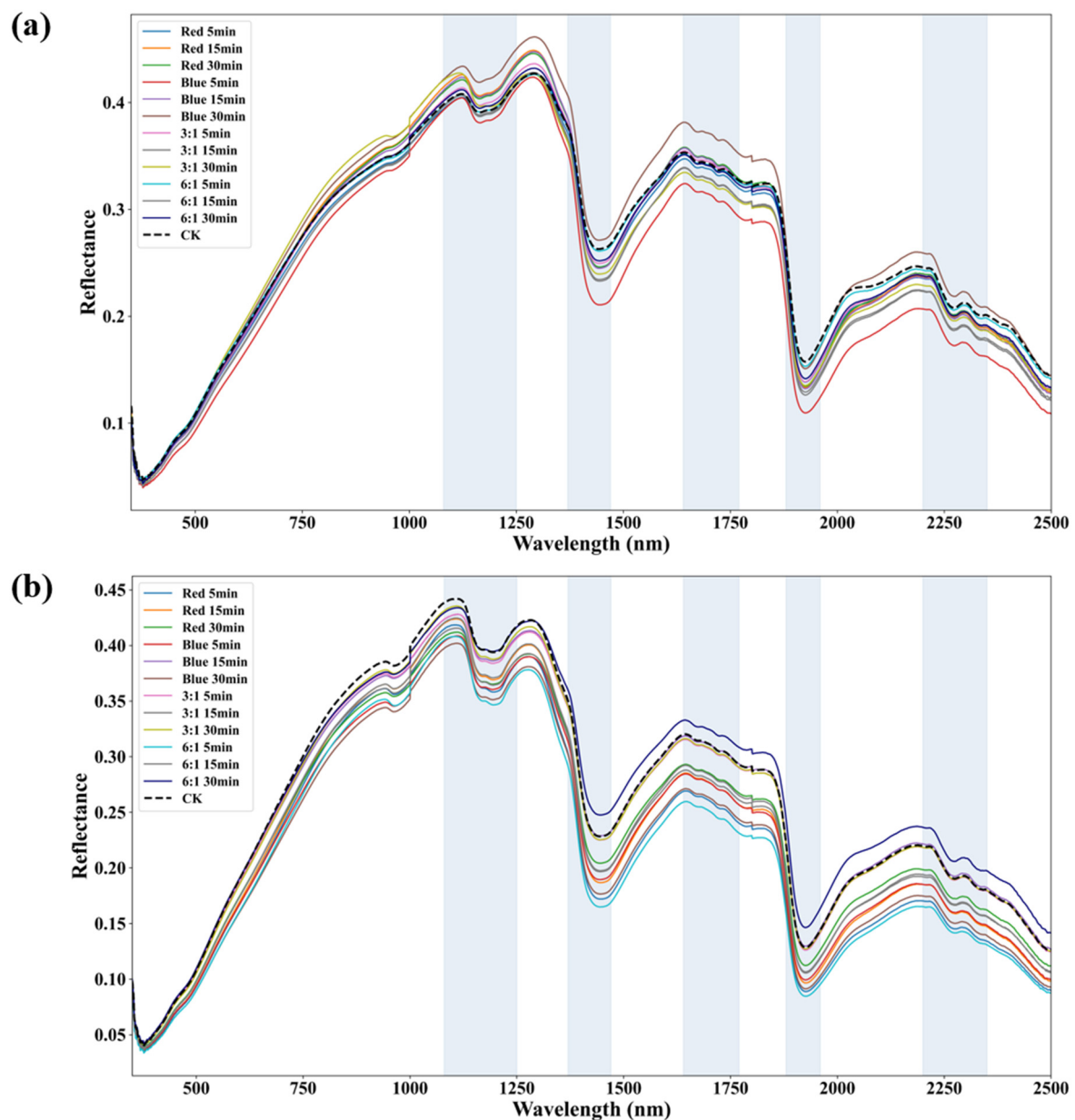


Figure 10. Mean spectral curves of ginseng seeds: (a) MCGS; (b) AGS.

Table 10. Main functional group assignments of feature wavelengths and their corresponding physicochemical characteristics.

Wavelength Region	Main Functional Group Assignment	Possible Corresponding Physicochemical Characteristics
1080–1250 nm	Mainly C–H second overtone, primarily involving –CH, –CH ₂ , and –CH ₃ vibrations [37,38]	Lipids, crude fat, and some reserve carbohydrates
1370–1470 nm	Mainly O–H absorption, with contributions from C–H and N–H [39,40]	Water content, bound water, and hydrogen-bonding environment may also partially reflect changes in proteins and carbohydrates
1640–1770 nm	Mainly C–H first overtone [37,41]	Lipids and membrane lipid-related information
1880–1960 nm	O–H combination band/strong water band [37,42]	Water status and bound water changes
2200–2350 nm	Overlapping region of N–H/C–H/O–H combination bands [41]	Complex components such as proteins, starch, cellulose, sugars, and some lipids

The emergence results in Table 1, the mean spectral curve in Figure 10a, and the feature band assignments in Table 10 showed that MCGS had relatively high ER, EE, and EI under blue light irradiation for 5 min, indicating that short-duration blue light treatment markedly improved the emergence vigor of MCGS. From the perspective of spectral response, the blue light treatment for 5 min showed clear reflectance differences from the other treatments in O–H-related sensitive regions, such as 1370–1470 nm and 1880–1960 nm. These regions are mainly associated with seed water status, bound water, and the hydrogen-bonding environment, suggesting that the improvement in emergence vigor of MCGS under short-duration blue light treatment may be related to changes in water-related spectral information.

In addition, certain spectral differences among treatments were also observed in the wavelength regions of 1080–1250 nm, 1640–1770 nm, and 2200–2350 nm, which are related to C–H, N–H, and O–H combination absorptions. These regions may involve information on components such as lipids, membrane lipids, proteins, starch, and sugars. The ER of the MCGS control group was lower than that of AGS, indicating a relatively low basal emergence level. After appropriate laser treatment, the emergence performance of MCGS improved markedly, and these changes may be associated with spectral information related to water status, membrane lipid structure, and storage reserve status. However, the specific physiological processes require further verification using physiological and biochemical indicators.

Combined with the emergence results in Table 3, the mean spectral curves in Figure 10b, and the feature band assignments in Table 10, the AGS control group showed a relatively high ER, indicating a favorable basal emergence level of its seeds. Compared with MCGS, AGS showed more distinct stage-dependent responses to different light qualities and irradiation times. Under blue light irradiation for 5 min, AGS achieved a relatively high final ER, suggesting that short-duration blue light treatment may have certain regulatory potential for its emergence process. This effect may be related to water status and partial storage reserve information reflected by O–H- and C–H-related wavelengths. The red–blue 3:1 treatment for 15 min showed relatively high EE, indicating that an appropriate composite light ratio may be more conducive to early emergence initiation in AGS, which may also be associated with spectral information related to water status and storage reserves. Under red light irradiation for 30 min, AGS showed better ER and EI, suggesting that longer-

duration red light may be more beneficial for later emergence accumulation. Its spectral response may be more closely reflected in wavelength regions such as 1640–1770 nm and 2200–2350 nm, which are associated with complex components including lipids, proteins, sugars, and starch.

Therefore, the spectral information captured by the CFO-SVR model can be used not only for ER prediction, but also as a reference for understanding changes in ginseng seed vigor after laser treatment. The selected key wavelengths were mainly associated with physicochemical components such as water, lipids, proteins, and carbohydrates, indicating that the spectral information contained in these wavelengths may reflect differences in the internal component status of seeds after laser treatment. MCGS and AGS showed different response trends within similar feature wavelength ranges, further suggesting that different ginseng seed types differ in their adaptability to light quality and irradiation time. These findings also provide a reference for establishing differentiated laser regulation parameters for different ginseng seed types in future studies.

3.3.3. SHAP Analysis

To further analyze the prediction basis and feature contributions of the optimized CFO-SVR model, SHAP was used to interpret the final model. The contributions of each input feature were visualized using SHAP bar and beeswarm plots. In the bar plot, the bar length represents the mean absolute SHAP value of each feature across all explained samples, reflecting its importance in the prediction dataset. The beeswarm plot further shows the distribution range of SHAP values for each feature and the direction of its influence on model output. Features are ranked from top to bottom according to importance. A wider SHAP value distribution indicates a stronger association with the output and a greater potential influence on the prediction results.

As shown in Figure 11a, among the top 20 important feature wavelengths, 17 were located in the shortwave infrared region, including 1221, 2238, 1123, 2027, 1549, 2257, 1227, 2167, 1193, 1082, 1648, 2202, 1226, 1408, 1373, 1768, and 1705 nm. Among these wavelengths, 1221 nm had the greatest influence on the prediction of ginseng seed ER.

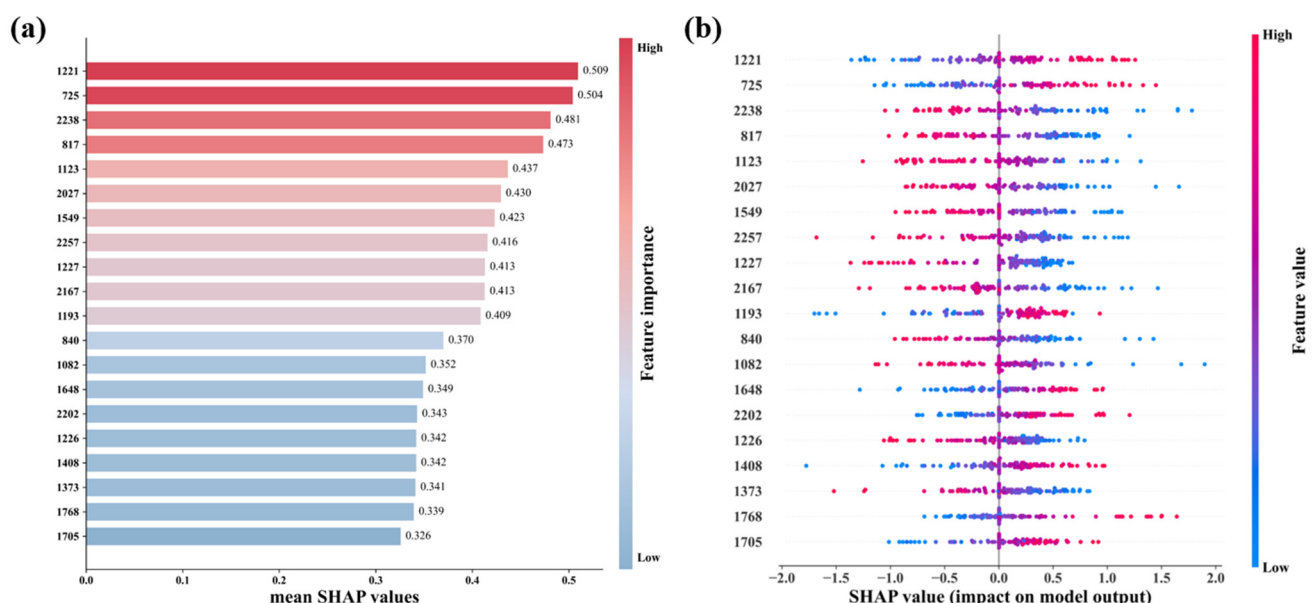


Figure 11. Feature importance plots for interpreting the CFO-SVR model: (a) SHAP bar plot; (b) SHAP beeswarm plot.

As shown in Figure 11b, different feature wavelengths showed different directions of influence on the model output. High feature values at 1221, 725, 1193, 1648, 2202, 1408,

1768, and 1705 nm were mainly distributed in the positive SHAP value region, indicating that higher values at these wavelengths generally increased the model-predicted ER and made positive contributions to ER prediction. In contrast, high feature values at 2238, 817, 1123, 2027, 1549, 2257, 1227, 2167, 840, 1082, 1226, and 1373 nm were mainly distributed in the negative SHAP value region, indicating that higher values at these wavelengths may reduce the model-predicted ER and thus contribute negatively to ER prediction.

According to the near-infrared band assignments listed in Table 10, bands such as 1221, 1226, 1227, and 1193 nm were mainly distributed near 1200 nm. This region is generally associated with C–H second-overtone absorption and partial high-order overtone absorption of O–H and may reflect spectral information related to lipids, carbohydrates, and bound water status in seeds. The bands at 1408 and 1373 nm are located near O–H-related absorption regions and may be associated with seed water status and the hydrogen-bonding environment. Bands such as 1549, 1648, 1705, and 1768 nm mainly involve N–H- and C–H-related overtone absorptions and may be related to information on storage substances such as proteins, lipids, and some carbohydrates. Long-wave near-infrared bands such as 2027, 2167, 2202, 2238, and 2257 nm are mostly located in the combination absorption regions of O–H, N–H, and C–H, and may reflect information on complex components such as water, proteins, starch, sugars, and lipids.

These results indicate that the key bands identified by SHAP were not randomly distributed, but were mainly concentrated in typical near-infrared regions associated with seed water status, storage substances, and organic functional group absorption, providing a possible spectral explanation for the prediction of ginseng seed ER by the CFO-SVR model. Because the information reflected by these wavelengths overlaps, a single wavelength cannot be directly assigned to a specific physiological process. Therefore, this study provides only a preliminary interpretation from the perspective of spectral assignment.

Overall, the SHAP analysis showed that the predictions of the CFO-SVR model were jointly influenced by multiple feature wavelengths, mainly located in the shortwave infrared region. This indicates that the model did not rely on random wavelengths for prediction, but primarily used spectral information related to seed components such as water, lipids, proteins, and carbohydrates. SHAP analysis provided supplementary evidence for the importance ranking and influence direction of key wavelengths, thereby improving the interpretability of the CFO-SVR prediction results.

4. Discussion

The results of this study indicate that semiconductor laser treatment affected the emergence vigor of ginseng seeds. However, this effect was not a uniform promotion response, but was jointly influenced by seed type, light-quality combination, and irradiation time. MCGS showed greater vigor enhancement under blue light treatment, whereas AGS performed better under longer-duration red light irradiation. These results suggest that the effects of laser treatment on ginseng seed vigor are not single or fixed promoting effects, but depend on seed type, light-quality characteristics, and irradiation dose.

This difference may be related to variations between MCGS and AGS in dormancy depth, seed coat structure, embryo developmental status, and endogenous hormone balance. Although cracked-stage seeds have acquired a certain basis for germination, they remain at a critical transitional stage from dormancy release to germination initiation, during which different ginseng seed types may differ in their perception of and response to external light signals. In general, blue light is more commonly associated with the activation of energy metabolism, antioxidant regulation, and enhanced cellular activity, whereas red light is more involved in phytochrome-mediated dormancy release and radicle elongation regulation [43–45]. Therefore, the higher sensitivity of MCGS to blue light may indicate

that its vigor enhancement depends more on metabolic activation and enhanced enzymatic reactions. In contrast, the better performance of AGS under long-duration red light suggests that its germination process may be more strongly affected by phytochrome signaling and early growth regulation. Meanwhile, the biological effects of laser irradiation also showed clear dose dependence. An appropriate irradiation time may act as a low-energy physical stimulus to promote reserve transformation and embryo growth, whereas insufficient or excessive stimulation may weaken the promoting effect. Therefore, the different treatment combinations reflected the differential physiological responses of different ginseng seed types within specific laser dose windows. The hyperspectral modeling results further demonstrated that changes in ginseng seed vigor after laser treatment could be effectively characterized using spectral information. Preprocessing with SG combined with the second derivative (S-G·2nd) significantly improved model prediction performance, indicating that this method reduced the effects of noise and baseline drift while enhancing weak absorption features. Compared with conventional feature selection methods such as RF and CARS, the feature subset selected by CFO achieved higher prediction accuracy when combined with SVR, suggesting that swarm intelligence optimization has strong global search capability for high-dimensional and highly collinear hyperspectral data. Furthermore, the CFO-SVR joint optimization model achieved the highest coefficient of determination and the lowest prediction errors on the test set, indicating that simultaneous optimization of feature wavelengths and model parameters improved model generalization performance.

It should be noted that the time interval between laser irradiation and hyperspectral acquisition may affect the interpretation of the spectral results. In this study, hyperspectral reflectance data were not collected immediately after laser irradiation, but were collected on the second day after all laser treatments were completed. Seeds from all laser treatment groups and the control group were kept overnight under the same indoor environmental conditions before spectral acquisition. Therefore, the instantaneous surface temperature differences and immediate photothermal cooling process that may have been caused by laser irradiation were largely eliminated, and the hyperspectral sensor was unlikely to mainly measure differences in photothermal cooling curves among different treatment groups. Meanwhile, this study collected visible–near-infrared–shortwave infrared reflectance information within the range of 350–2500 nm, rather than thermal infrared radiation signals. Therefore, this system was not a device for directly monitoring seed surface temperature or the thermal cooling process. Considering the overnight interval between laser treatment and spectral acquisition in the experimental procedure of this study, the influence of instantaneous photothermal effects on the hyperspectral acquisition results can be considered relatively small. However, this study did not directly record the dynamic changes in seed surface temperature after laser irradiation. Future studies may further verify the photothermal effects after laser treatment by combining infrared thermal imaging or surface temperature monitoring.

The key bands selected by the CFO were mainly concentrated in the regions of 1080–1250 nm, 1370–1470 nm, 1640–1770 nm, 1880–1960 nm, and 2200–2350 nm. In terms of near-infrared spectral assignment, these bands are mainly associated with overtone and combination absorptions of functional groups such as O–H, C–H, and N–H, and may reflect spectral information related to seed water status, surface scattering characteristics, lipids, proteins, starch, sugars, and other storage substances. It should be noted that near-infrared absorption peaks are characterized by overlap and complex assignments, and a single feature wavelength cannot be directly interpreted as a change in a specific biochemical component. Therefore, the spectral interpretation in this study mainly indicates that these bands are associated with information related to seed emergence vigor, but cannot directly prove that stable biochemical reconstruction of internal seed macromolecular structures

occurred after laser treatment. SHAP analysis further showed that bands such as 1221, 2238, 1123, and 2027 nm made important contributions to model prediction, providing a basis for feature interpretation of the hyperspectral model.

Although this study established a relatively high-accuracy non-destructive prediction method for ginseng seed vigor, several limitations remain. First, only two types of seeds, MCGS and AGS, were used in this study. The samples were obtained from the same production area, the experimental data were collected in the same experimental season, and hyperspectral data acquisition was completed using the same measurement system. Therefore, the sample sources, year-to-year variation, and instrumental conditions were still limited. Second, this study used ER and related emergence indicators as the basis for vigor evaluation, and the prediction model was mainly established based on the statistical association between hyperspectral features and ER. Continuous spectral monitoring at different time points after laser treatment, seed surface temperature recording, and physiological and biochemical indicator measurements were not conducted simultaneously. Therefore, the biological interpretation of the key bands in this study remains mainly a speculative analysis based on near-infrared spectral assignments, and the present spectral evidence alone is insufficient to directly demonstrate stable biochemical reconstruction of internal macromolecular structures, such as proteins, lipids, starch, or sugars, after laser treatment. Future studies may further combine indicators such as enzyme activity, respiration intensity, soluble sugars, membrane lipid peroxidation, infrared thermal imaging, and dynamic monitoring of the imbibition process to verify the relationship between spectral features and internal physiological changes in seeds. In addition, model validation in this study was mainly used to evaluate the ER prediction ability within the established laser treatment system. Because the training and test sets were divided at the spectral-sample level, different spectra from the same treatment group may have entered the training and test sets separately. Therefore, the R^2 , RMSE, and MAE reported in this study should mainly be interpreted as internal prediction performance within the treatment range of this experiment, and should not be regarded as a completely independent validation for new seed batches or new treatment groups. Further independent external validation should be conducted under different years, production areas, seed batches, larger sample sizes, and different measurement systems to improve the stability and applicability of the model in practical seed quality detection and production applications. Overall, this study provides a feasible technical pathway for laser-based pre-sowing treatment of ginseng seeds and hyperspectral nondestructive detection and also offers a reference for the rapid quality evaluation of medicinal plant seeds.

5. Conclusions

In this study, cracked-stage MCGS and AGS were used to investigate the effects of semiconductor laser treatments with different light-quality combinations and irradiation times on the emergence vigor of ginseng seeds, and a non-destructive prediction model for ER was constructed based on hyperspectral technology and machine learning. The results showed that the effects of laser treatment were characterized by clear seed-type differences and parameter dependence. MCGS responded most markedly to short-duration blue light treatment, and blue light treatment for 5 min significantly increased ER, EE, and EI. In contrast, no significant differences were observed between any laser treatment group and CK in AGS, with only certain numerical trends. Hyperspectral results showed that changes in seed emergence vigor could be characterized by spectral information, and the CFO-SVR model showed good ER prediction ability within the treatment range of this experiment. However, because model evaluation was based on spectral-sample-level division within the same experimental system, the obtained prediction results should be regarded as internal

validation results within the treatment range of this experiment, rather than as completely independent validation for new seed batches or unknown treatment groups. Future studies should further introduce independent samples from different years, production areas, and seed batches, and adopt independent validation at the treatment-group or batch level to further evaluate the robustness and practical applicability of the model.

Author Contributions: Conceptualization, J.G., M.X. and H.Y.; methodology, J.G. and M.X.; validation, J.G., M.X. and H.Y.; formal analysis, J.G.; investigation, J.G.; resources, J.G., M.X., Y.X., Z.L., Y.L., L.Z. and Q.H.; data curation, J.G.; writing—original draft preparation, J.G.; writing—review and editing, J.G., M.X., H.Y., Y.X., Z.L., Y.L., L.Z. and Q.H.; visualization, J.G.; supervision, H.Y.; project administration, Y.L. and L.Z.; funding acquisition, H.Y. All authors have read and agreed to the published version of the manuscript.

Funding: This study was financially supported by the “Changchun Talent Support Program” Science and Technology Innovation Project of Changchun Municipal Science and Technology Bureau (Grant No. 2025ccyc15) and the Jilin Province Major Issues Identification Project for Seed Industry Revitalization and Security, funded by the Jilin Provincial Department of Science and Technology (Grant No. JL2023-03).

Data Availability Statement: The data presented in this study are available on request from the corresponding author.

Acknowledgments: The authors express their sincere gratitude to all co-authors for their continuous support and contributions throughout this research.

Conflicts of Interest: The authors affirm unequivocally that no discernible competing financial interests or personal relationships exist that could conceivably have exerted any influence on the research findings presented in this paper.

Abbreviations

The following abbreviations are used in this manuscript:

MCGS	Mountain Cultivated Ginseng Seeds
AGS	American Ginseng Seeds
ER	Emergence rate
EE	Emergence energy
EI	Emergence index
SG	Savitzky–Golay smoothing
SNV	Standard normal variate
MSC	Multiplicative scatter correction
RF	Random Forest
CARS	Competitive adaptive reweighted sampling
CFO	Caterpillar Fungus Optimizer
SVR	Support vector regression
CFO-SVR	Caterpillar Fungus Optimizer-based support vector regression
SHAP	Shapley additive explanations
RBF	Radial basis function
R ²	Coefficient of determination
RMSE	Root mean square error
MAE	Mean absolute error

References

1. Hao, L.; Hou, X.; Li, C. Study on medicinal value of ginseng. *Asia-Pac. Tradit. Med.* **2012**, *8*, 213–214.
2. Zhang, S.; Cheng, W.; Guo, J.; Jiang, X.; Li, J. Technique of ginseng seed germination in autumn harvest and spring sowing and causes of non-emergence after sowing. *Ginseng Res.* **2022**, *6*, 34–35. [[CrossRef](#)]

3. Zhou, L.; Yang, M.; Bi, C.; Yu, H. Principles and applications of crop laser breeding and intensification technology. *J. Jilin Agric. Univ.* **2025**, *47*, 581–593. [[CrossRef](#)]
4. Perveen, R.; Wang, X.; Jamil, Y.; Ali, Q.; Ali, S.; Zakaria, M.Q.; Afzaal, M.; Kasana, R.A.; Saleem, M.H.; Fiaz, S. Quantitative determination of the effects of He–Ne laser irradiation on seed thermodynamics, germination attributes and metabolites of safflower (*Carthamus tinctorius* L.) in relation with the activities of germination enzymes. *Agronomy* **2021**, *11*, 1411. [[CrossRef](#)]
5. Możdżeń, K.; Barabasz-Krasny, B.; Zandi, P. Effect of long-term of He-Ne laser light irradiation on selected physiological processes of triticale. *Plants* **2020**, *9*, 1703. [[CrossRef](#)] [[PubMed](#)]
6. Perveen, R.; Jamil, Y.; Al-Huqail, A.A.; Alsudays, I.M.; Alghanem, S.M.S.; Ali, Q.; Saeed, F.; Azeem, M.; Rizwan, M.; Al-Robai, S.A. Effects of pulsed Nd:YAG laser kernel irradiation on maize (*Zea mays* L.): Insights into germination, gas exchange, photosynthetic pigments, and morphological modifications. *J. Photochem. Photobiol. B Biol.* **2024**, *253*, 112876. [[CrossRef](#)] [[PubMed](#)]
7. Li, X.; Chen, H.; Han, R. Effects of He-Ne laser irradiation on the seeds germination and seedling growth of *Arabidopsis thaliana*. *Acta Bot. Boreal.-Occident. Sin.* **2012**, *32*, 131–135.
8. Li, Y.; Wang, W.; Yang, L. Effect of CO₂ laser treatment on seed germination of *Erigeron breviscapus*. *Seed* **2019**, *38*, 114–116. [[CrossRef](#)]
9. Yang, L.; Gao, J.; Shen, S.; Chen, Y.; Tian, J. The physiological effects of He-Ne laser irradiation on seed germination of *Elaeagnus mollis* Diels. *Appl. Laser* **2018**, *38*, 812–816. [[CrossRef](#)]
10. Swathy, P.S.; Kiran, K.R.; Joshi, M.B.; Mahato, K.K.; Muthusamy, A. He-Ne laser accelerates seed germination by modulating growth hormones and reprogramming metabolism in brinjal. *Sci. Rep.* **2021**, *11*, 7948. [[CrossRef](#)] [[PubMed](#)]
11. Sarreta, Y.; de Castro Neto, J.C. Effects of 660 nm laser irradiation of soybean seeds on germination, emergence and seedling growth. *Acta Agrophys.* **2021**, *28*, 5–18. [[CrossRef](#)]
12. Xu, P.; Fu, L.; Pan, Y.; Chen, D.; Yang, S.; Yang, R. Identification of maize seed vigor based on hyperspectral imaging and deep learning. *Bull. Natl. Res. Cent.* **2024**, *48*, 84. [[CrossRef](#)]
13. Li, S.; Sun, L.; Jin, X.; Feng, G.; Zhang, L.; Bai, H. Research on identification of common bean seed vigor based on hyperspectral and deep learning. *Microchem. J.* **2025**, *211*, 113133. [[CrossRef](#)]
14. Pang, L.; Wang, L.; Yuan, P.; Yan, L.; Xiao, J. Rapid seed viability prediction of *Sophora japonica* by improved successive projection algorithm and hyperspectral imaging. *Infrared Phys. Technol.* **2022**, *123*, 104143. [[CrossRef](#)]
15. Zhang, M.; Song, J.; Jia, H.; Zhang, X.; Yang, W.; Wang, Y.; Wang, H. Prediction of vigor of naturally aged seeds from Xishuangbanna cucumber (*Cucumis sativus* L. var. *xishuangbannanensis*) using hyperspectral imaging. *Agriculture* **2025**, *15*, 1043. [[CrossRef](#)]
16. Sun, M.; Jiang, F.; Wang, C.; Zhang, A.; Tian, Y. Development of a comprehensive quality index for strawberry and its non-destructive prediction using hyperspectral imaging. *Food Control* **2026**, *181*, 111754. [[CrossRef](#)]
17. Zhang, T.; Li, J.; Tong, J.; Song, Y.; Wang, L.; Wu, R.; Wei, X.; Song, Y.; Zeng, R. End-to-end deep fusion of hyperspectral imaging and computer vision techniques for rapid detection of wheat seed quality. *Artif. Intell. Agric.* **2025**, *15*, 537–549. [[CrossRef](#)]
18. Zou, Z.; Chen, J.; Wu, W.; Luo, J.; Long, T.; Wu, Q.; Wang, Q.; Zhen, J.; Zhao, Y.; Wang, Y.; et al. Detection of peanut seed vigor based on hyperspectral imaging and chemometrics. *Front. Plant Sci.* **2023**, *14*, 1127108. [[CrossRef](#)] [[PubMed](#)]
19. Qi, H.; Huang, Z.; Sun, Z.; Tang, Q.; Zhao, G.; Zhu, X.; Zhang, C. Rice seed vigor detection based on near-infrared hyperspectral imaging and deep transfer learning. *Front. Plant Sci.* **2023**, *14*, 1283921. [[CrossRef](#)] [[PubMed](#)]
20. Hu, Q.; Lu, W.; Guo, Y.; He, W.; Luo, H.; Deng, Y. Vigor detection for naturally aged soybean seeds based on polarized hyperspectral imaging combined with ensemble learning algorithm. *Agriculture* **2023**, *13*, 1499. [[CrossRef](#)]
21. Wonggasem, K.; Wongchaisuwat, P.; Chakranon, P.; Onwimol, D. Utilization of machine learning and hyperspectral imaging technologies for classifying coated maize seed vigor: A case study on the assessment of seed DNA repair capability. *Agronomy* **2024**, *14*, 1991. [[CrossRef](#)]
22. Zaid, A.; Abu-Khalaf, N.; Mudalal, S.; Petracci, M. Differentiation between normal and white striped turkey breasts by visible/near infrared spectroscopy and multivariate data analysis. *Food Sci. Anim. Resour.* **2020**, *40*, 96–105. [[CrossRef](#)] [[PubMed](#)]
23. Zhang, Y.; Huan, K.; Liu, X.; Fan, Y.; Cao, X.; Han, X. Multi-scale feature attention network for rapid and non-destructive quantification of aflatoxin B1 in maize using hyperspectral imaging. *Foods* **2025**, *14*, 3769. [[CrossRef](#)] [[PubMed](#)]
24. Yang, X.; Wei, Z.; Li, L.; Yang, X.; Han, J.; Ming, M.; Wang, G.; Cao, F.; Zhou, K.; Fu, F. Large-scale non-destructive crown-level assessment of Ginkgo pigments via hyperspectral and machine learning techniques. *Plant Methods* **2025**, *21*, 130. [[CrossRef](#)] [[PubMed](#)]
25. Wang, X.; Han, J.; Liu, C.; Feng, T. Non-Destructive Assessment of Apple Internal Quality Using Rotational Hyperspectral Imaging. *Front. Plant Sci.* **2024**, *15*, 1432120. [[CrossRef](#)] [[PubMed](#)]
26. Choi, J.-Y.; Kim, M.; Park, S.; Cho, J.-S.; Lim, J.H.; Moon, K.-D. Rapid discrimination of *Panax ginseng* powder adulterated with various root plants by FT-IR spectroscopy coupled with multivariate analysis. *Food Sci. Biotechnol.* **2024**, *33*, 805–815. [[CrossRef](#)] [[PubMed](#)]
27. Al Siam, A.; Salehin, M.M.; Alam, M.S.; Ahamed, S.; Islam, M.H.; Rahman, A. Paddy Seed Viability Prediction Based on Feature Fusion of Color and Hyperspectral Image with Multivariate Analysis. *Heliyon* **2024**, *10*, e36999. [[CrossRef](#)] [[PubMed](#)]

28. Chen, Y.; Sun, W.; Jiu, S.; Wang, L.; Deng, B.; Chen, Z.; Jiang, F.; Hu, M.; Zhang, C. Soluble Solids Content Binary Classification of Miyagawa Satsuma in Chongming Island Based on Near Infrared Spectroscopy. *Front. Plant Sci.* **2022**, *13*, 841452. [CrossRef] [PubMed]
29. Zhang, S.; Lv, C.; Cui, C.; Wang, J.; Wu, J.; Mao, W. Near-Infrared Spectral Analysis for Assessing Germination Rate of Rapeseed Seeds: An Applied Sciences Approach. *Appl. Sci.* **2023**, *13*, 11001. [CrossRef]
30. Yang, B.; Liang, B.; Zhou, S.; Qian, Y.; Zheng, R.; Shu, H.; He, P.; Wang, J.; Jiang, L.; Sang, Y.; et al. A Novel Bio-Inspired Caterpillar Fungus (*Ophiocordyceps sinensis*) Optimizer for SOFC Parameter Identification via GRNN. *Renew. Energy* **2026**, *256*, 123995. [CrossRef]
31. Cheng, T.; Chen, G.; Wang, Z.; Hu, R.; She, B.; Pan, Z.; Zhou, X.-G.; Zhang, G.; Zhang, D. Hyperspectral and Imagery Integrated Analysis for Vegetable Seed Vigor Detection. *Infrared Phys. Technol.* **2023**, *131*, 104605. [CrossRef]
32. Ahmed, M.T.; Villordon, A.; Kamruzzaman, M. Hyperspectral Imaging and Explainable Deep-Learning for Non-Destructive Quality Prediction of Sweetpotato. *Postharvest Biol. Technol.* **2025**, *222*, 113379. [CrossRef]
33. Lan, Z.; Hong, Z.; Pan, Y.; Liu, Y. Preliminary Report on the Experiment of 6-BA Enhanced Germination of Multi-Flowered Solomon's Seal Primary Rhizomes. *J. Zhejiang For. Sci. Technol.* **2024**, *44*, 56–60. [CrossRef]
34. Zhao, J.-Z.; Kong, D.-S.; Wang, L.; Guo, Y.-Y. Effects of Low Temperature Stratification on Emergence of *Lycium ruthenicum* under Drought at Different Sowing Depths. *Acta Pratacult. Sin.* **2017**, *26*, 56–66. Available online: <http://cyxb.magtech.com.cn/CN/10.11686/cyxb2017088> (accessed on 1 July 2026).
35. Wu, L.; Zhu, M.; Yan, J.; Zhu, W. Effects of Solid Matrix Priming on Spinach Seed Emergence and Physiological and Biochemical Indices of Seedlings. *Acta Agric. Shanghai* **2025**, *41*, 40–44. [CrossRef]
36. Nagy, A.; Szabó, A.; Elbeltagi, A.; Nxumalo, G.S.; Budayné Bódi, E.; Tamás, J. Hyperspectral Indices Data Fusion-Based Machine Learning Enhanced by MRMR Algorithm for Estimating Maize Chlorophyll Content. *Front. Plant Sci.* **2024**, *15*, 1419316. [CrossRef] [PubMed]
37. Kong, H.; Wang, J.; Lin, G.; Chen, J.; Xie, Z. Analysis of Nutritional Content in Rice Seeds Based on Near-Infrared Spectroscopy. *Photonics* **2025**, *12*, 481. [CrossRef]
38. Ma, L.; Peng, Y.; Pei, Y.; Zeng, J.; Shen, H.; Cao, J.; Qiao, Y.; Wu, Z. Systematic Discovery about NIR Spectral Assignment from Chemical Structural Property to Natural Chemical Compounds. *Sci. Rep.* **2019**, *9*, 9503. [CrossRef] [PubMed]
39. Yasmin, J.; Ahmed, M.R.; Wakholi, C.; Lohumi, S.; Mukasa, P.; Kim, G.; Kim, J.; Lee, H.; Cho, B.-K. Near-Infrared Hyperspectral Imaging for Online Measurement of the Viability Detection of Naturally Aged Watermelon Seeds. *Front. Plant Sci.* **2022**, *13*, 986754. [CrossRef] [PubMed]
40. Tang, Y.; Nie, W.; Zhang, Y.; Li, C.; Wei, Z.; Chen, H.; Hu, Y.; Li, W. Rapid Identification of *Fritillaria* spp. Using Multi-Wavebands Spectroscopy and Multi-Source Data Fusion Strategies. *J. Appl. Res. Med. Aromat. Plants* **2025**, *46*, 100636. [CrossRef]
41. Shruti; Shukla, A.; Rahman, S.S.; Suneja, P.; Yadav, R.; Hussain, Z.; Singh, R.; Yadav, S.K.; Rana, J.C.; Yadav, S.; et al. Developing an NIRS Prediction Model for Oil, Protein, Amino Acids and Fatty Acids in Amaranth and Buckwheat. *Agriculture* **2023**, *13*, 469. [CrossRef]
42. Xue, H.; Xu, X.; Yang, Y.; Hu, D.; Niu, G. Rapid and Non-Destructive Prediction of Moisture Content in Maize Seeds Using Hyperspectral Imaging. *Sensors* **2024**, *24*, 1855. [CrossRef] [PubMed]
43. Ahsan, S.M.; Injamum-Ul-Hoque, M.; Shaffique, S.; Ayoobi, A.; Rahman, M.A.; Rahman, M.M.; Choi, H.W. Illuminating *Cannabis sativa* L.: The Power of Light in Enhancing *C. sativa* Growth and Secondary Metabolite Production. *Plants* **2024**, *13*, 2774. [CrossRef] [PubMed]
44. Trivellini, A.; Toscano, S.; Romano, D.; Ferrante, A. The Role of Blue and Red Light in the Orchestration of Secondary Metabolites, Nutrient Transport and Plant Quality. *Plants* **2023**, *12*, 2026. [CrossRef] [PubMed]
45. Farrokh Tehrani, P.; Majd, A.; Mahmoodzadeh, H.; Nejad Satari, T. Effect of Red and Blue Light-Emitting Diodes on Germination, Morphological and Anatomical Features of *Brassica napus*. *Adv. Stud. Biol.* **2016**, *8*, 173–180. [CrossRef]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.