photonics

MDPI

Article

Non-Destructive Assessment of Rice Seed Vigor and Extraction of Characteristic Spectra Based on Near-Infrared Spectroscopy

Qing Huang [1,2], Jinxing Wei [3], Jiale Cheng [2,4], Mingdong Zhu [3], Wei Nie [2], Xingping Wang [2], Mai Hu [2], Zhenyu Xu [2], Ruifeng Kan [2,*] and Wenqing Liu [2]

1 School of Environmental Science and Optoelectronic Technology, University of Science and Technology of China, Hefei 230026, China
2 Anhui Institute of Optics and Fine Mechanics, Hefei Institute of Physical Science, Chinese Academy of Sciences, Hefei 230031, China
3 State Key Laboratory of Hybrid Rice, Hunan Rice Research Institute, Changsha 410125, China
4 College of Information Science and Engineering, Northeastern University, Shenyang 110819, China
* Correspondence: kanruifeng@aiofm.ac.cn

Abstract

Rice seed vigor is one of the critical factors determining rice yield and quality. Identifying substances related to seed vigor and rapidly assessing seed vigor by non-destructive methods are of great significance for increasing rice production. This study employed near-infrared diffuse reflectance spectroscopy (NIR-DRS) and transmission spectroscopy (NIR-TS) to evaluate the vigor of naturally aged rice seeds. The NIR-DRS failed to establish a reliable relationship between spectral data and seed vigor, proving ineffective in distinguishing seed vigor. After enhancing the spectral differences between viable and non-viable seeds, the NIR-TS successfully identified high-vigor and non-viable seeds, with a partial least squares discriminant analysis (PLS-DA) model achieving accuracy and germination rates of 84.52% and 88.57% on the test set, respectively. Furthermore, three algorithms, including interval partial least squares (iPLS), genetic algorithm (GA), and competitive adaptive reweighted sampling (CARS), were applied to extract characteristic spectral wavelengths associated with seed vigor. Among these, the CARS algorithm performed the best, identifying 38 characteristic wavelengths. Wavelength analysis indicated that rice seed vigor is primarily influenced by molecules such as starch, protein, moisture, and lipids. Using the characteristic wavelengths selected by the CARS algorithm, a PLS-DA prediction model for rice seed vigor was constructed, achieving high accuracy and germination rates of 90.47% and 95.38% on the test set, respectively. This study demonstrates that NIR-TS outperforms NIR-DRS in assessing rice seed vigor. Moreover, wavelength selection techniques can effectively identify characteristic spectral features related to seed vigor and significantly enhance the prediction accuracy of the model.

Keywords: rice seed; vigor prediction; near-infrared diffuse reflectance spectroscopy; near-infrared transmission spectroscopy; extraction of characteristic spectra

Check for updates

Received: 14 January 2026
Revised: 3 February 2026
Accepted: 25 February 2026
Published: 26 February 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) license.

1. Introduction

Rice is considered to be one of the most significant food crops in the world. According to statistics, the global cultivated area of rice spans approximately 160 to 165 million hectares, accounting for about 28% of the total cereal cultivation area worldwide. Asia is the primary region for rice cultivation, accounting for 90% of the global total area [1]. Seed vigor is a pivotal indicator of seed quality, exerting a direct influence on the performance

of rice fields and the ultimate yield [2]. High-vigor seeds characteristically demonstrate robust germination potential, uniform seedling emergence, robust seedlings, high stress tolerance, shorter growth periods and significant yield potential [3]. However, during storage or transportation, improper conditions can easily lead to a decline in seed vigor. Conversely, low-vigor seeds demonstrate a diminished capacity for stress tolerance, exhibit stunted growth, and are susceptible to seedling mortality [4]. Consequently, the rapid and accurate assessment of rice seed vigor is of great practical significance for ensuring agricultural production efficiency [5].

Currently, the seed vigor testing methods fall into two categories: destructive and non-destructive. Destructive testing methods mainly include standard germination tests [6], enzyme activity assays [7], tetrazolium (TTC) staining [8], and molecular marker analysis [9]. However, these methods typically cause irreversible damage to seeds, permitting only sampling-based inspection rather than comprehensive seed-by-seed screening of large quantities. Moreover, their lengthy testing cycles make them impractical for direct selection of high-vigor seeds for field sowing [5]. Non-destructive testing methods include image analysis [10], oxygen sensor technology [11], impedance testing [12], and spectroscopic techniques [13]. In recent years, with the rapid development of computer science and spectroscopy technology, spectral non-destructive testing has demonstrated significant advantages and application potential in the assessment of seed vigor [5]. A range of spectroscopic techniques, including Fourier-transform near-infrared spectroscopy (FT-NIR) [14], Raman spectroscopy (RS) [14], tunable diode laser absorption spectroscopy (TDLAS) [15] photoacoustic spectroscopy (PAS) [16], and hyperspectral imaging (HSI) [17], have been employed in the domain of rice seed vigor detection, showing considerable development potential. Fan et al. employed a supercontinuum laser source coupled with a near-infrared (NIR) spectrometer to acquire transmission spectra from individual rice seeds. Following preprocessing by normalization (Norm), second derivative (SD), and orthogonal signal correction (OSC), a vigor prediction model was constructed using PLS-DA. The model demonstrated an accuracy rate of 91.67% on the test set, thereby validating the effectiveness and feasibility of this method for non-destructive, rapid prediction of rice seed vigor [18]. Guo et al. utilized PAS technology to assess vigor by measuring carbon dioxide evolution rates in rice seeds. The results demonstrated a strong positive correlation ($r = 0.979$) between seed respiration rate and vigor. The system has been demonstrated to complete vigor testing within 8 h, which is approximately one-ninth of the time required for standard germination tests [16]. Fan et al. implemented near-infrared spectroscopy (NIRS) to measure key parameters determining rice quality—the content of amylose and fat. Comparative analysis of NIR-DRS versus NIR-TS modes revealed: for amylose content prediction in individual rice seeds, models based on NIR-TS achieved superior predictive accuracy; conversely, NIR-DRS demonstrated greater efficacy for fat content prediction [19]. Abdullah et al. applied hyperspectral imaging (HSI) technology to predict rice seed vigor. The spectral data were preprocessed using the Savitzky–Golay (SG) second derivative method and integrated with both spectral features and key color image features to construct a PLS-DA model. The model achieved high accuracy rates of 93.3% on the calibration set and 90.9% on the test set, respectively [20].

NIRS and HSI represent core non-destructive technologies for the assessment of rice seed vigor [18,21–25]. NIRS typically employs reflectance and transmission measurement modes, while HSI universally utilizes diffuse reflectance to acquire whole-seed spectral data by extracting regions of interest [17]. Current research exhibits three critical limitations: systematic evaluation of comparative performance between NIR-DRS and NIR-TS in vigor detection remains unavailable; studies predominantly rely on artificially aged seeds with insufficient investigation under natural aging conditions; modeling generally depends

on full-spectrum data without adequate exploration of vigor-specific spectral features. In response to the aforementioned limitations, this study: (1) The experimental samples were selected to be naturally aged rice seeds, with the aim of enhancing the practical application value of the conclusions. (2) The possibility of collecting diffuse reflectance and transmission spectra of rice seeds simultaneously to predict seed vigor was explored. (3) The characterizations of physiological differences between viable and non-viable seeds were strengthened by combining phenotypic data after seed germination. Spectral bands significantly related to vigor were analyzed in depth using algorithms such as iPLS [26] and CARS [27].

2. Materials and Methods

2.1. Selection of Rice Seed Sample

The experimental materials were provided by the Hunan Rice Research Institute. Seeds of variety 9311 from three production years (2005, 2018, and 2021) were selected as the research objects. The selected samples were stored in a controlled-environment repository (4 °C, 44% relative humidity) prior to analysis. After rigorous screening to remove individuals exhibiting visible lesions, glume-gapping, pre-harvest sprouting, mechanical damage to the seed coat, or abnormal coloration, the final usable sample sizes were determined to be 288 seeds for the 2005 production year, 288 seeds for 2018, and 192 seeds for 2021. The entire data was divided into calibration and test sets at a ratio of 7:3 by using the Kennard-Stone (KS) method.

2.2. Spectral Data Acquisition of Rice Seed Sample

2.2.1. Data Acquisition of Diffuse Reflectance Spectroscopy

Diffuse reflectance spectra of rice seeds were acquired using an MPA Fourier transform near-infrared (FT-NIR) spectrometer (Bruker, Ettlingen, Germany). The wavelength detection range was 834.169–2502.508 nm with an optical resolution of 1.074 nm, yielding 1037 discrete wavelengths. Before data acquisition, the spectrometer was warmed up for 30 min to stabilize. A single rice seed was placed at the center of the circular detection window, and spectra were collected in diffuse reflectance mode. A single rice seed was placed at the center of the circular detection window. Each seed was manually replaced for every measurement. Spectra were collected in diffuse reflectance mode [28]. Each seed underwent 32 repeated scans, with data acquired from both sides (two opposite sides). The average spectrum was calculated as the final spectral data for that seed.

2.2.2. Data Acquisition of Transmission Reflectance Spectroscopy

This study established a transmission spectroscopy detection system for rice seeds, and its working principle is shown in Figure 1. The system consists of a supercontinuum laser source (SuperK Compact, NKT Photonics, Tokyo, Japan), a grating spectrometer (NIRQuest512-2.2, Ocean Optics, Orlando, FL, USA), focusing lenses, a circular aperture, and other components. The spectral acquisition range spans 899.149–2132.355 nm with an optical resolution of 2.408 nm, yielding 512 discrete wavelengths. With the light source covering a wavelength range of 450–2400 nm and the spectrometer integration time set to 50 ms, this configuration yielded rice seed spectra with a relatively high signal-to-noise ratio. Before data acquisition, both the light source and spectrometer were preheated for 30 min. Each seed was manually replaced for measurement, with five repeated acquisitions performed for each side. Subsequently, the average transmission spectrum was calculated and used as the final spectral data for the seed.

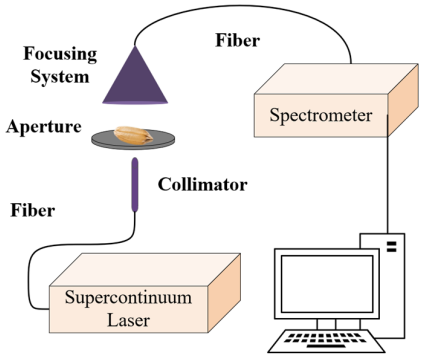


Figure 1. Transmission Reflectance Spectroscopy Acquisition System.

2.3. Standard Germination Test

To accurately measure rice seed vigor, all seeds underwent initial dry weight measurement after spectral data collection. Subsequently, germination tests were conducted. Seeds were cultivated in an environmental chamber with controlled temperature (30 ± 0.5 °C), humidity (80%), and a 16 h light/8 h dark photoperiod for 7 days. All procedures were strictly conducted in accordance with the Chinese National Standard GB/T 3543.4-1995) [6]. Four phenotypic parameters per seed were systematically recorded: sprout length (coleoptile protrusion from seed coat), root length (longest primary root), root count (number of roots), and fresh weight (whole seedling weight). Based on germination criteria, seeds with radicle or coleoptile emergence of 2 mm or more from the seed coat were classified as viable (labeled 1), (https://www.seedtest.org/en/publications/international-rules-seed-testing.html (accessed on 28 August 2025)) while ungerminated or malformed germinated seeds [28] were classified as non-viable (labeled 0). This established a binary vigor classification dataset, and the germination rate of this seed batch was calculated. This dataset was then linked with the corresponding spectral data to enable non-destructive prediction of seed viability.

2.4. Data Preprocessing

During spectral acquisition, electronic noise from the spectrometer and environmental interference cause signal fluctuations, necessitating preprocessing of the raw spectral data. The raw spectral data was systematically processed through four dimensions: scale scaling, scatter correction, smoothing, and baseline correction [29]. The optimal preprocessing strategy was explored by comparing the suitability of different methods for rice seed spectra. The objective of scale scaling is to eliminate dimensional differences and suppress strong signal interference. The primary methods include Norm, standardization (Std), and vector normalization (VN) [30]. Multiple scattering effects induced by surface textures of rice seeds can introduce noise that is unrelated to the target components. The multiplicative scatter correction (MSC) and standard normal variate (SNV) have been demonstrated to be effective in the elimination of interference caused by particle size differences and light scattering [31,32]. Smoothing processing removes high-frequency noise and enhances the signal-to-noise ratio through methods including moving average (MA) smoothing, Gaussian filtering, SG filtering, and median filtering [33]. Baseline correction is employed to suppress background drift and enhance spectral quality, with conventional methods encompassing second derivative (FD), SD, and fractional-order differentiation [34]. The specific combination of preprocessing methods adopted in this study is outlined in Table 1.

Table 1. Methods of Spectral Preprocessing.

Category	Method
scale scaling	Norm, VN
Scatter correction	MSC, SNV
smoothing	MA, SG
baseline correction	FD, SD

2.5. Enhancement of Spectral Differences Between Viable and Non-Viable Rice Seeds

Currently, no unified grading standard exists for assessing rice seed vigor, making accurate quantitative evaluation challenging. The rice seeds were classified into viable and non-viable groups based on the results of the germination test. However, while some seeds meet germination criteria, their significantly low biomass accumulation or retarded growth indicates substantially reduced actual vigor. Conversely, seeds with germination potential may fail to sprout due to biotic stresses such as bacterial/fungal infestation or embryonic micro-damage, and thus be misclassified as non-viable. When such anomalous samples are included in modeling datasets, the noisy features may be introduced and compromise the classifier's generalization capability and prediction accuracy.

This study implemented a dual-screening strategy to amplify inter-group differences in seed vigor, indirectly enhancing spectral feature distinctions between viable and non-viable seeds. First, high-vigor seeds were selected from germinated populations while simultaneously eliminating false-negative samples from non-germinated groups. The rationale for this strategy lies in the significant positive correlation between rice seed vigor and early phenotypic parameters (sprout length, root length, root count, and fresh-to-dry weight ratio), which serve as effective variables for vigor assessment. The quartile method was employed for screening [35], with the specific strategy outlined as follows: for germinated seeds, individuals exhibiting sprout length, root length, root count, and fresh-to-dry weight ratio all above median values were classified as high-vigor seeds. For non-germinated seeds, since no roots or sprouts developed, individuals with fresh-to-dry weight ratios in the third quartile were considered as false-negative samples and were removed because of having potential germination capability.

2.6. Search for Characteristic Spectra of Rice Seed Vigor

While full-wavelength spectral modeling preserves comprehensive spectral information, the high-dimensional data contains substantial unrelated noise, which will significantly increase computational complexity, and may induce overfitting risks due to the curse of dimensionality, consequently reducing model generalizability. Therefore, spectral dimensionality reduction through wavelength selection is essential to eliminate redundant information while precisely capturing feature bands associated with biochemical processes of seed vigor. Compared to feature extraction methods like principal component analysis (PCA) [36], wavelength selection algorithms directly retain the physical significance of raw spectra, enabling traceability of selected wavelengths to specific molecular bond vibrations [37]. This study employed three wavelength selection methods (iPLS, GA, and CARS) to systematically mine key characteristic spectra correlated with rice seed vigor. The fundamental principles of the three algorithms are as follows:

2.6.1. iPLS

The full spectrum is partitioned into K equal-width sub-intervals. A local partial least squares (PLS) model is built within each sub-interval, with the accuracy of cross-validation (ACC_{CV}) serving as the fitness value to evaluate sub-interval quality. The sub-interval with the highest fitness value is designated as the primary interval. Remaining sub-intervals

are sequentially incorporated into this primary interval in descending order of fitness value. Sub-intervals that improve the fitness value are retained, while those degrading performance are eliminated [26].

2.6.2. GA

As a swarm intelligence optimization algorithm simulating natural evolutionary mechanisms, GA exhibits strong adaptability and global search capabilities, granting it significant advantages in spectral feature wavelength selection. The algorithm comprises four core steps: population initialization, selection operation, crossover operation, and mutation operation. Through iterative evolution, the wavelength combination with the highest fitness value is ultimately identified [38].

2.6.3. CARS

The CARS algorithm is inspired by the principle of survival of the fittest from Darwin's theory of evolution and combines it with the regression coefficients of the partial least squares (PLS) model for variable selection. At each iteration, samples are first randomly selected in proportion using Monte Carlo sampling. Then, variables with smaller absolute values of regression coefficients are removed using an exponential decay function. Subsequently, the algorithm further selects the remaining variables using an adaptive weighted sampling strategy, retains the set with larger regression coefficient weights, constructs the PLS model and calculates the RMSECV value under this feature wavelength combination. After multiple iterations, the feature wavelength combination with the smallest RMSECV value is selected as the final optimal subset [27]. In this study, partial least squares-discriminant analysis (PLS-DA) was used as the calibration model, with ACC_{CV} selected as the fitness value.

2.7. Prediction Model of Rice Seed Vigor: PLS-DA

PLS-DA is a supervised learning method based on partial least squares regression, and is particularly well-suited for analyzing high-dimensional small-sample data. The PLS-DA algorithm comprises two components: partial least squares regression and linear discriminant analysis applied to the values obtained from the PLS fitting [39]. The primary computational steps are as follows:

The first step is to decompose the seed spectral matrix X and the vigor marker matrix Y. The mathematical model is:

$$X = TP^T + E \tag{1}$$

$$Y = UQ^T + F \tag{2}$$

where T and U are the score matrix of X and Y respectively, P and Q are the load matrix of X and Y respectively, and E and F are the PLS fitting residual matrix of X and Y respectively.

The second step is to perform linear regression on T and U:

$$U = TB \tag{3}$$

$$B = \left(T^T T\right)^{-1} T^T Y \tag{4}$$

The third step is to reconstruct the data and obtain the relationship between matrix Y and spectral data matrix:

$$Y = TBQ \tag{5}$$

The fourth step is classified prediction. For the new prediction value $\check{Y}$, the category where the maximum value is located is taken as the classification result.

2.8. Evaluation Metrics and Software

2.8.1. Evaluation Indices

This study employed Accuracy (ACC) and Germination Rate (GR) as evaluation metrics for the rice seed vigor prediction model [40]. A 10-fold cross-validation was applied to the validation set to ensure model robustness and generalizability across subsets. Four cumulative evaluation metrics were recorded: cross-validated accuracy in the validation set, cross-validated germination rate in the validation set (GR_{CV}), accuracy in the test set, and germination rate in the test set. The calculation formulas for *ACC* and *GR* are as follows:

$$ACC = \frac{TP + TN}{TP + TN + FP + FN} \tag{6}$$

$$GR = \frac{TP}{TP + FP} \tag{7}$$

where *TP* (true positive) refers to the number of seeds that actually germinated and were correctly predicted by the model as having germinated; *TN* (true negative) refers to the number of seeds that did not germinate and were correctly predicted by the model as not germinating; *FP* (false positive) refers to the number of seeds that did not actually germinate but were incorrectly predicted by the model as germinating; *FN* (false negative) refers to the number of seeds that actually germinated but were incorrectly predicted by the model as not germinating.

2.8.2. Software

All code and calculations are programmed and implemented using Python 3.8, with some functions imported from the Sklearn library. The computer CPU is an Intel(R) Core (TM) i5-14600K, at 3.50 GHz with 32 GB of RAM.

3. Results and Discussion

3.1. Phenotypic Data Analysis of Rice Seeds

Following a 7-day standard germination test, fresh weight, sprout length, root length, and root count were measured for germinated seeds. From a biomass perspective, the fresh-to-dry weight ratio comprehensively reflects macro-manifestations of seed vigor. Statistical analysis was performed on the four phenotypic indicators after seed germination to reveal the influence of storage time on seed vigor.

Statistical distributions of phenotypic data for rice seeds stored in 2005, 2018, and 2021 are presented in Figure 2. According to the distribution of sprout length measured 7 days after seed germination, the longest sprout was observed in 2018, followed by 2021, while the shortest was recorded in 2005. The proportions of seeds with sprout length exceeding 60 mm for these three years were 68%, 50.3%, and 46.7%, respectively. Furthermore, the proportion of seeds that sprouted lengths between 70–80 mm after seven days of germination was as high as 34% in 2018, approximately twice that of 2005 and 2021. The root length distribution exhibited that the proportion of seeds with root lengths greater than 60 mm seven days after germination was 81.8%, 84.3%, and 88% in the three years, respectively, with relatively minor variations. The root number distribution indicates that the proportions of seeds with more than five roots seven days after germination in 2005, 2018, and 2021 were 36.4%, 53.8%, and 50.5%, respectively, with the lowest proportion in 2005. The fresh-to-dry weight ratio distribution indicates that in 2021, the proportion of seeds exhibiting a weight gain ratio greater than 2.5 seven days following germination reached 88.3%, which is significantly higher than the figures observed in 2005 and 2018.

Collectively, the four phenotypic parameters demonstrate significantly reduced vigor in 2005-stored seeds compared to 2018 and 2021. However, while 2018-stored seeds

achieved optimal sprout length and 2021-stored seeds showed the highest fresh-to-dry weight ratio, the relative vigor between these two years cannot be conclusively ranked based on any single phenotypic metric.

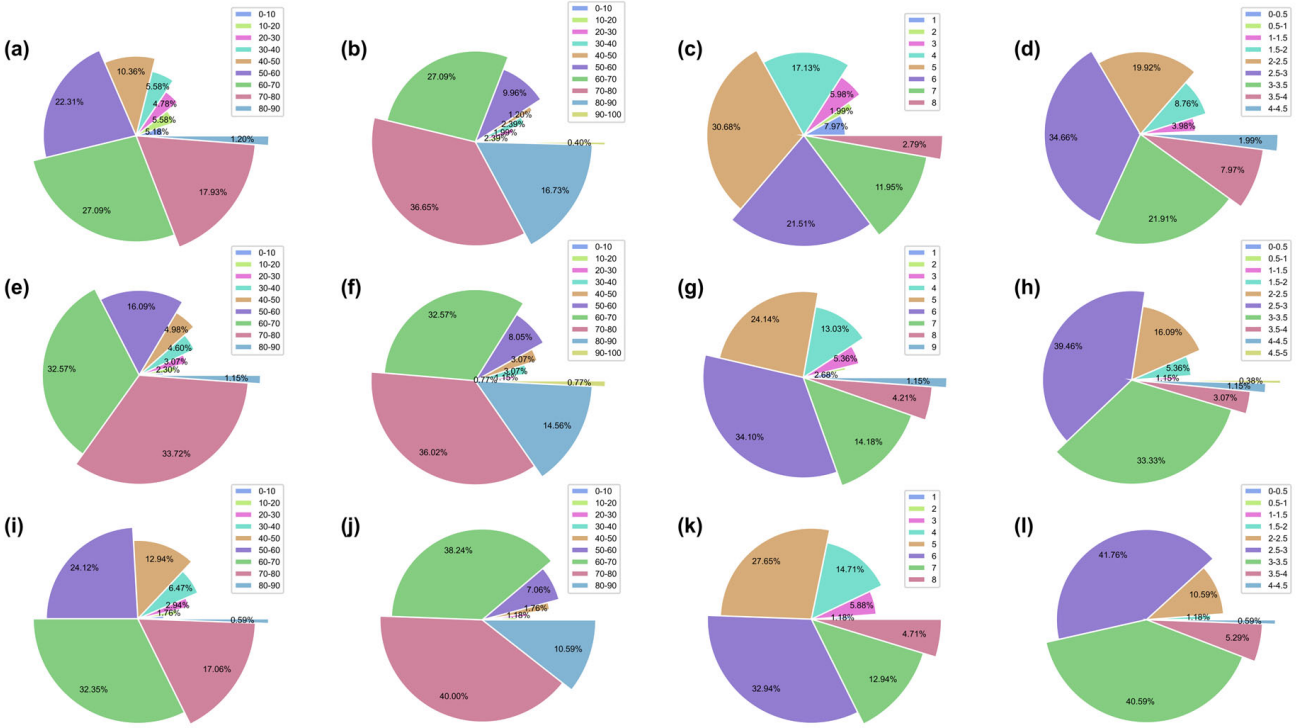


Figure 2. Phenotypic data of rice seeds across three years. (**a**) Proportion of different sprout length ranges of rice seeds from 2005. (**b**) Proportion of different root length ranges of rice seeds from 2005. (**c**) Proportion of different root count ranges of rice seeds from 2005. (**d**) Proportion of different fresh-to-dry weight ratio ranges of rice seeds from 2005. (**e**) Proportion of different sprout length ranges of rice seeds from 2018. (**f**) Proportion of different root length ranges of rice seeds from 2018. (**g**) Proportion of different root count ranges of rice seeds from 2018. (**h**) Proportion of different fresh-to-dry weight ratio ranges of rice seeds from 2018. (**i**) Proportion of different sprout length ranges of rice seeds from 2021. (**j**) Proportion of different root length ranges of rice seeds from 2021. (**k**) Proportion of different root count ranges of rice seeds from 2021. (**l**) Proportion of different fresh-to-dry weight ratio ranges of rice seeds from 2021.

3.2. Spectral Analysis

3.2.1. Diffuse Reflectance Near-Infrared Spectra

The normalized NIR-DRS of rice seeds from the 2005, 2018, and 2021 storage years are shown in Figure 3. Significant noise was observed in the 834–1132 nm spectral range, and the data in this range has been excluded during modelling. The spectral reflectance trends of viable and non-viable seeds are closely similar, making it difficult to distinguish them directly. The average spectra for each year show that the average spectral differences between viable and non-viable seeds are minimal, with their curves highly overlapping. As shown in Figure 3m, the average diffuse reflectance spectra of seeds in 2005 are significantly lower than those in 2018 and 2021. The average diffuse reflectance spectra of seeds in 2021 are slightly lower than those in 2018, but since the storage periods of the two are close, the differences between them are small.

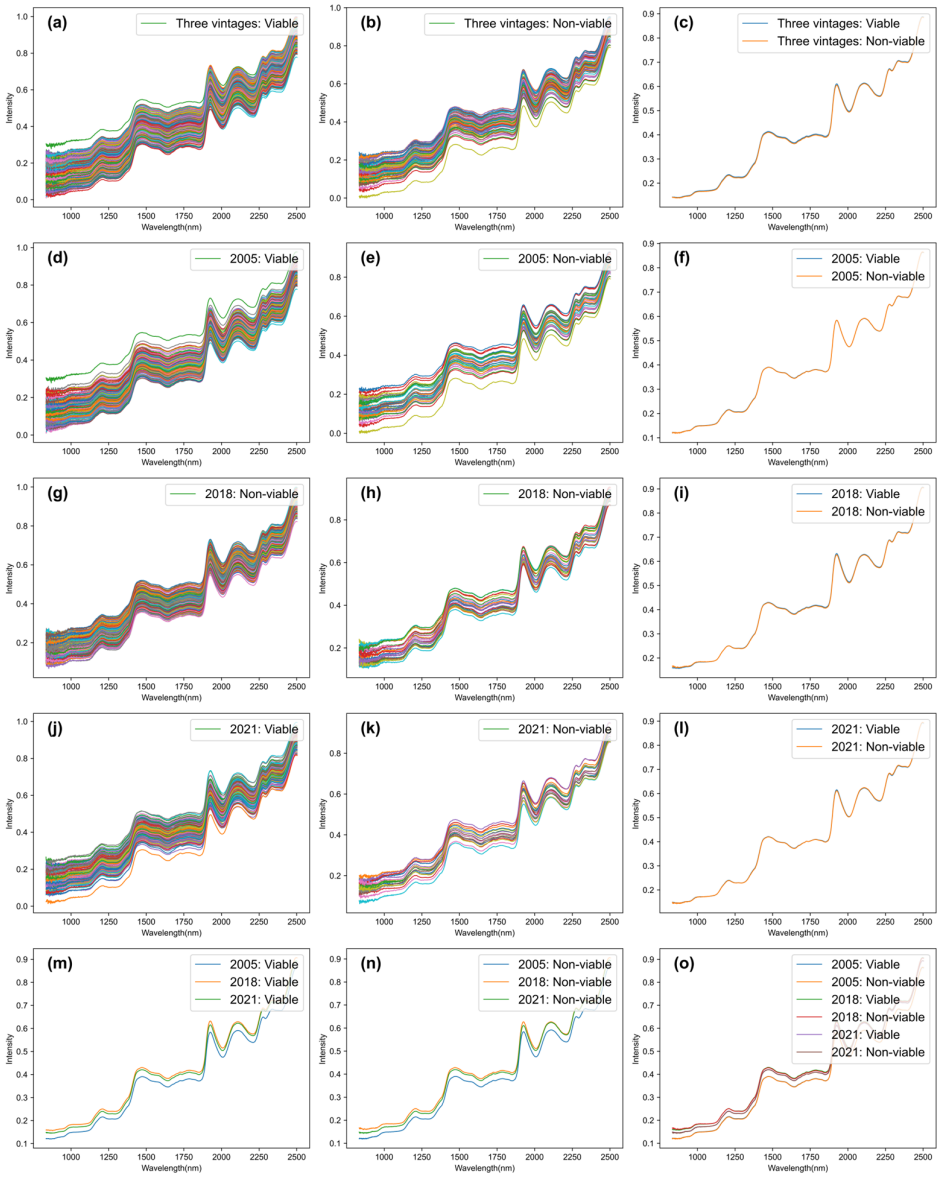


Figure 3. NIR-DRS curves of rice seeds from three years. (**a**) spectral curves of viable seeds from three years. (**b**) spectral curves of non-viable seeds from three years. (**c**) average spectral curves of viable and non-viable seeds from three years. (**d**) Spectral curves of viable seeds from 2005. (**e**) Spectral curves of non-viable seeds from 2005. (**f**) Average spectral curves of viable and non-viable seeds from 2005. (**g**) Spectral curves of viable seeds from 2018. (**h**) Spectral curves of non-viable seeds from 2018. (**i**) Average spectral curves of viable and non-viable seeds in 2018. (**j**) Spectral curves of viable seeds in 2021. (**k**) Spectral curves of non-viable seeds in 2021. (**l**) Average spectral curves of viable and non-viable seeds in 2021. (**m**) Average spectral curves of viable seeds across three years. (**n**) Average spectral curves of non-viable seeds across three years. (**o**) Average spectral curves of seeds across three years.

3.2.2. Transmission Near-Infrared Spectra

Figure 4 displays the normalized NIR-TS of rice seeds from the 2005, 2018, and 2021 storage years. Consistent with the diffuse reflectance spectral analysis, the spectral waveforms between viable and non-viable seeds are strikingly similar, making visual differentiation impractical. Observation of the average spectra across years reveals that viable seeds consistently exhibit higher transmission spectral intensity than non-viable seeds in all three storage years (2005, 2018, and 2021). Furthermore, as the duration of storage increased, the spectral differences between the two types of seeds gradually

decreased. As observed in Figure 4m, comparison of the average transmittance spectra of viable seeds across the three years reveals that the average spectral intensity of rice seeds in 2005 is significantly lower than that in 2018 and 2021. Within the wavelength ranges of 800–1300 nm and 1900–2200 nm, the average transmittance spectral intensity in 2021 is slightly lower than that in 2018, whereas the inverse pattern is observed in the 1300–1900 nm range. Furthermore, the transmittance spectra of non-viable seeds exhibit analogous trends across different years. This phenomenon aligns with the seed vigor characteristics reflected in the aforementioned phenotypic data.

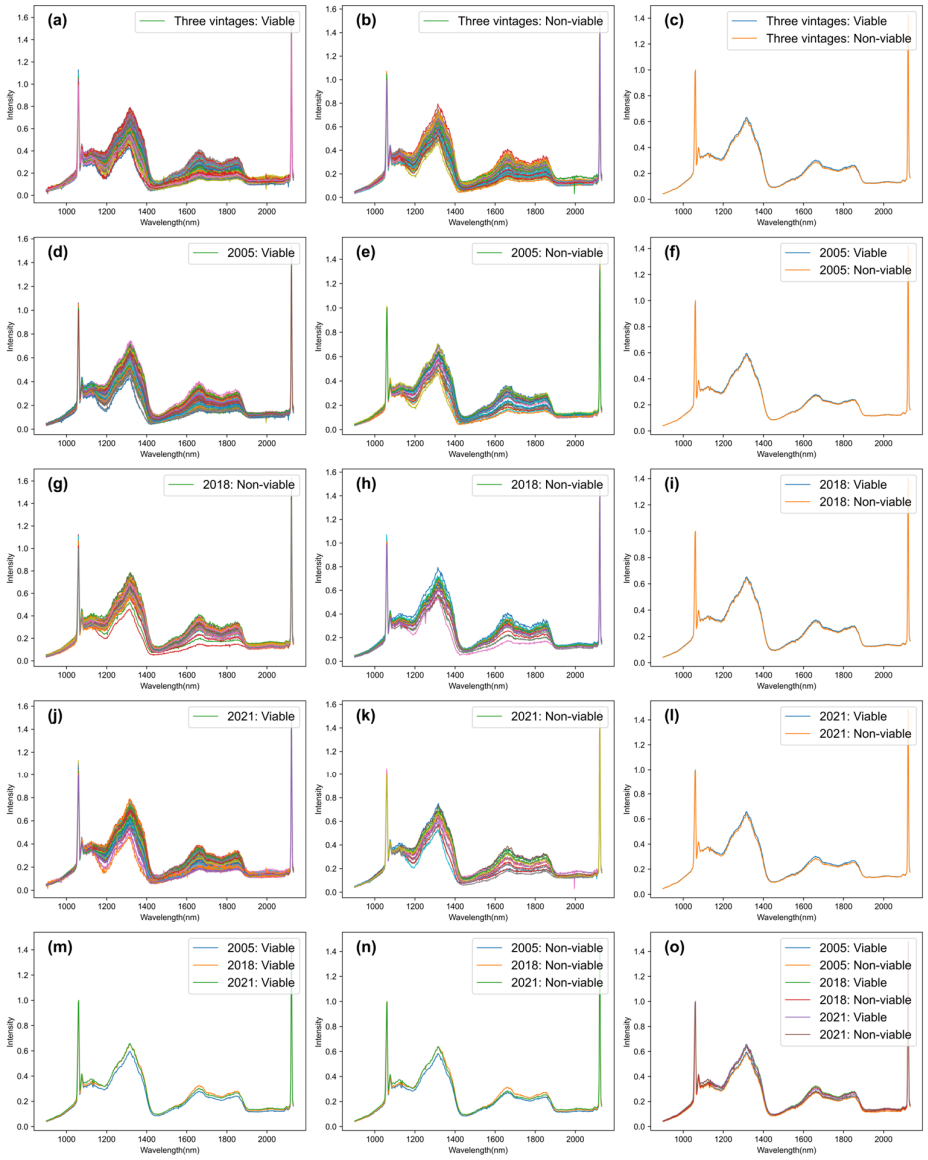


Figure 4. NIR-TS curves of rice seeds from three years. (**a**) spectral curves of viable seeds from three years. (**b**) spectral curves of non-viable seeds from three years. (**c**) average spectral curves of viable and non-viable seeds from three years. (**d**) Spectral curves of viable seeds from 2005. (**e**) Spectral curves of non-viable seeds from 2005. (**f**) Average spectral curves of viable and non-viable seeds from 2005. (**g**) Spectral curves of viable seeds from 2018. (**h**) Spectral curves of non-viable seeds from 2018. (**i**) Average spectral curves of viable and non-viable seeds in 2018. (**j**) Spectral curves of viable seeds in 2021. (**k**) Spectral curves of non-viable seeds in 2021. (**l**) Average spectral curves of viable and non-viable seeds in 2021. (**m**) Average spectral curves of viable seeds across three years. (**n**) Average spectral curves of non-viable seeds across three years. (**o**) Average spectral curves of seeds across three years.

NIR-DRS measures the intensity of light reflected and scattered from the sample surface, primarily capturing spectral features of the rice seed husk. Existing studies confirm that the husk significantly influences spectral measurements. In contrast, NIR-TS measures the intensity of light passing through the entire seed, providing insights into internal structural information. Comparison of average diffuse reflectance and transmission spectra across the three storage years reveals that the differences between viable and non-viable seeds are more pronounced in transmission spectra, indicating that rice seed vigor primarily depends on changes in their internal composition. It should be noted that the differences in seed vigor across different years arise not only from storage duration but also from combined effects of environmental factors during seed development (low/high temperature, drought, rice disease and insect pest), seed maturity, and storage conditions (temperature, humidity, etc. These factors may induce complex physiological changes within the seeds, ultimately manifesting as differences in vigor.

3.3. Data Preprocessing

The Kennard-Stone (KS) algorithm [41] was employed to divide all seeds into validation and test sets at a ratio of 3:1. The germination rate of the validation set was 88.08%, and that of the test set was 90.47%. The original spectral data underwent a series of preprocessing steps involving Norm, VN, MSC, SNV, FD, SD, MA, and SG, culminating in the establishment of PLS-DA models. As shown in Table 2, all preprocessing methods, whether applied to diffuse reflectance or transmission spectra, exhibited model collapse during cross-validation. Despite achieving an overall accuracy of 88.05%, the model predicted all seeds as belonging to the viable category. This issue stemmed from two factors. Firstly, the initial germination rate of the seeds was found to be excessively high, resulting in an alarmingly low proportion of non-viable samples and causing severe class imbalance. Secondly, the spectral differences between viable and non-viable seeds are minimal, thereby complicating direct differentiation through spectroscopy. Consequently, enhancing the spectral differences between viable and non-viable seeds is imperative.

Table 2. PLS-DA Prediction Results After Preprocessing of NIR-DRS and NIR-TS Data.

Method		LV	Validation Sets		Test Sets	
			ACC_{CV} (%)	GR_{CV} (%)	ACC (%)	GR (%)
NIR-DRS	Raw	6	88.05	88.05 *	89.61	90.39
	Norm	6	88.05	88.05 *	89.61	90.39
	VN	5	87.86	88.03 *	90.04	90.43
	MSC	6	87.86	88.17 *	90.47	90.47 *
	SNV	6	87.86	88.17 *	90.47	90.47 *
	FD	6	82.64	88.48 *	82.25	90
	SD	5	79.67	88.87	80.95	90.64
	MA	5	88.05	88.05 *	90.47	90.47
	SG	5	88.05	88.05 *	89.61	90.39
NIR-TS	Raw	7	87.48	87.93 *	90.47	90.47 *
	Norm	7	87.48	87.93 *	90.47	90.47 *
	VN	9	87.30	88.40 *	87.87	90.95
	MSC	8	87.12	88.21 *	90.04	90.43
	SNV	5	86.92	87.98 *	89.61	90.39
	FD	5	85.45	88.53 *	87.87	90.58
	SD	5	84.90	88.09 *	88.74	90.66
	MA	6	87.48	87.93 *	90.47	90.47 *
	SG	5	87.48	87.93 *	90.47	90.47 *

* All rice seeds were predicted as viable. ACC denotes accuracy, ACC_{CV} represents cross-validated accuracy, GR stands for germination rate, and GR_{CV} indicates cross-validated germination rate.

3.4. Enhancement of Vigor-Related Spectral Differences and Data Preprocessing

3.4.1. Amplification of Within-Group Differences in Seed Vigor Using the Interquartile Range Method

In order to enhance the spectral differences between viable and non-viable seeds, a quartile method was employed for seed screening. By increasing the inter-group differences in seed vigor, the spectral differences in the two seed types were indirectly enhanced. The average spectra of viable and non-viable seeds following screening are shown in Figure 5. Comparison of Figures 3c and 5a reveals enhanced differences in the average diffuse reflectance spectra between viable and non-viable seeds within the 1172–1381 nm and 1424–1861 nm ranges, as well as near 2006 nm, 2209 nm, and 2389 nm. Similarly, comparison of Figures 4c and 5a demonstrates that the average transmission spectra of the two seed types exhibit a more pronounced separation trend in the 1139–1393 nm and 1492–1878 nm wavelength bands. By amplifying the significant differences between the seed vigor groups, the spectral differences between the two seed types are further magnified, with the enhancement effect being particularly pronounced in the transmission spectra.

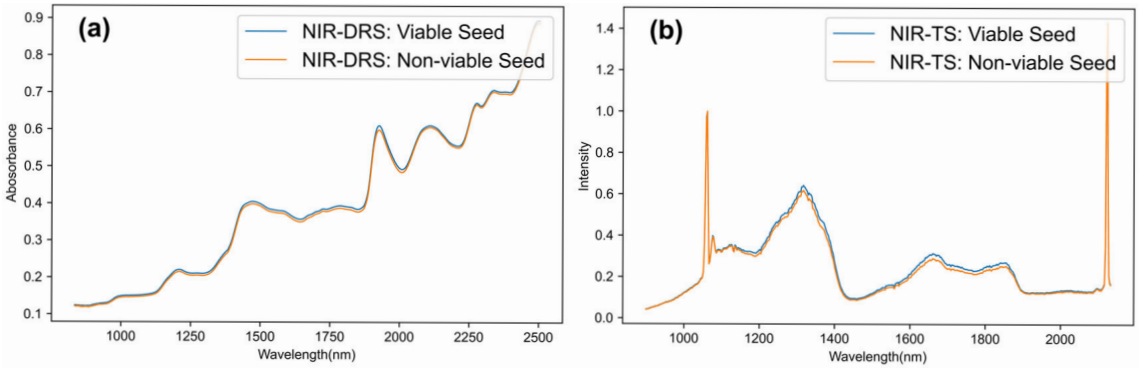


Figure 5. Average spectra after inter-group seed vigor differences enhancement. (**a**) NIR-DRS. (**b**) NIR-TS.

3.4.2. Data Preprocessing

The spectral data of seeds screened using the quartile method were preprocessed using Norm, VN, MSC, SNV, FD, SD, MA, and SG. The data were divided into validation and testing sets in a 3:1 ratio, with a germination rate of 73.58% for the validation set and 79.76% for the test set. The results of establishing the PLS-DA models based on the preprocessed data are shown in Table 3. The results indicate that even after spectral differences enhancement processing, the NIR-DRS of rice seeds cannot effectively predict vigor levels. On one hand, the problem of class imbalance persists. On the other hand, the differences in NIR-DRS responses between viable and non-viable seeds did not reach the model's identifiable threshold, preventing the establishment of a statistical association model between spectral features and vigor status.

The transmission spectra after spectral-differences enhancement could effectively distinguish between high-vigor and non-viable seeds. However, all preprocessing methods resulted in a decrease in model accuracy. After normalization, the germination rate in the test set reached 95.59%, but the accuracy on the test set was only 75%. This indicates that the model suffers from overfitting and fails to meet the requirements for precise detection. Although data preprocessing can suppress effects such as baseline drift and noise interference, the minimal spectral differences between viable and non-viable seeds causes the preprocessing to paradoxically discard critical discriminatory information, ultimately degrading model prediction accuracy.

Table 3. PLS-DA Prediction Results Using Different Preprocessing Methods After Spectral Differences Enhancement.

Method		LV	Validation Sets		Test Sets	
			ACC_{CV} (%)	GR_{CV} (%)	ACC (%)	GR (%)
NIR-DRS	Raw	5	73.07	75.95 *	80.95	84.00
	Norm	5	73.07	75.95 *	84.52	84.61
	VN	5	71.02	75.95 *	79.76	93.78
	MSC	6	72.55	78.80	79.76	79.76 *
	SNV	6	72.07	78.32	79.76	86.76
	FD	5	67.52	75.74	77.38	84.28
	SD	8	62.71	73.16	75.00	81.94
	MA	5	75.63	77.99 *	78.57	78.04
	SG	5	76.65	79.05 *	77.38	77.77
NIR-TS	Raw	8	80.39	84.85	84.52	88.57
	Norm	8	80.39	84.85	75.00	95.59
	VN	6	75.26	80.51	79.76	84.72
	MSC	7	77.31	84.29	79.76	79.76 *
	SNV	5	77.78	82.22	78.57	82.66
	FD	7	74.13	81.43	78.57	88.88
	SD	8	70.97	80.45	72.61	87.93
	MA	9	77.76	89.11	82.14	83.11
	SG	9	79.34	84.78	84.52	88.57

* All rice seeds were predicted as viable. ACC denotes accuracy, ACC_{CV} represents cross-validated accuracy, GR stands for germination rate, and GR_{CV} indicates cross-validated germination rate.

3.5. Extraction of Characteristic Spectra of Rice Seed Vigor

Feature wavelength selection was performed on the spectral data amplified for vigor differences to identify characteristic spectra related to seed vigor. Three algorithms (iPLS, GA, and CARS) were employed for wavelength selection on the full- spectrum data, and PLS-DA model was ultimately established, with the modeling results shown in Table 4. iPLS effectively reduced the number of wavelengths required for modeling and significantly improved the predictive performance of the model. Although GA exhibited some effectiveness on the test set, its accuracy reached only 77.38%, indicating overfitting. Among all the methods, CARS demonstrated superior performance, achieving impressive accuracy rates of over 90% on both the training and test sets, representing improvements of 11.32% and 5.95%, respectively, compared to the full-wavelength model. The germination rate saw even more notable enhancement, reaching 93.54% for the training set and 95.38% for the test set. These figures reflect increases of 9.09% and 6.81% over the full-wavelength model, and 19.96% and 15.71% over the true germination rates, respectively. Furthermore, CARS required the fewest wavelengths, at just 7.42% of the full-wavelength.

Table 4. PLS-DA Prediction Results Using Different Wavelength Selection Methods.

Method	nVar	Validation Sets		Test Sets	
		ACC_{CV} (%)	GR_{CV} (%)	ACC (%)	GR (%)
/	512	80.39	84.85	84.52	88.57
iPLS	247	80.89	85.24	85.71	91.04
GA	242	91.68	93.00	80.95	93.22
CARS	38	91.71	93.54	90.47	95.38

ACC denotes accuracy, ACC_{CV} represents cross-validated accuracy, GR stands for germination rate, and GR_{CV} indicates cross-validated germination rate.

The wavelength points selected for iPLS, GA, and CARS are shown in Figure 6. The iPLS primarily selected wavelengths within 1125–1235 nm, 1463–1684 nm, and 1906–2118 nm. Specifically, the 1125–1235 nm range corresponds to the second overtone stretching vibration absorption of C-H, O-H, and N-H bonds, while the 1463–1684 nm range is primarily linked to the first overtone stretching vibration absorption of C-H, O-H, and N-H bonds, as well as the combination bands absorption of C-H, O-H, and C-H. Absorption in the 1906–2118 nm range originates predominantly from combination and overtone absorptions of functional groups including O-H, N-H, C-H, and C=O in molecular structures. Based on the functional group information corresponding to wavelengths selected by iPLS, these spectral absorptions primarily reflect major organic components in rice, such as lipid, starch, protein, saccharide, moisture, and cellulose. Starch, proteins, and lipids constitute the principal storage substances in rice seeds. Starch serves as the main component of rice endosperm, accounting for 70–80% of dry seed weight and providing the primary energy source during early germination [42,43]. Proteins represent the second largest storage substance at approximately 15% [44], participating in nitrogen metabolism during early germination. Lipids comprise about 2% of seed content and also contribute to energy supply [45]. The wavelength points selected by the GA algorithm were uniformly distributed across the entire wavelength range, while effectively eliminating points with high collinearity.

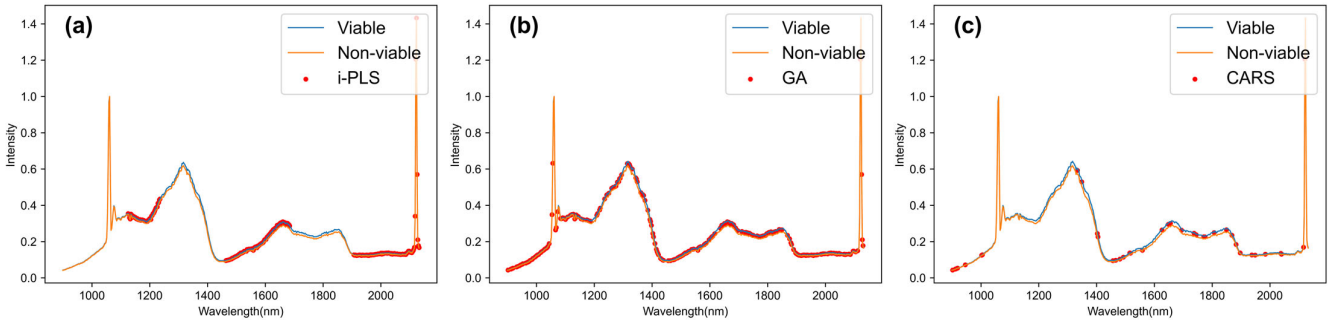


Figure 6. Characteristic wavelength points selected by different algorithms. (**a**) iPLS. (**b**) GA. (**c**) CARS.

The wavelength points selected by the CARS algorithm were relatively discrete. Compared to iPLS, CARS selected additional points in the short-wave near-infrared region. These wavelength points primarily correspond to the third overtone stretching vibration absorptions of C-H bonds and the second overtone stretching vibration absorptions of O-H and N-H bonds. Although this region demonstrates relatively weak absorption intensity, it shows heightened sensitivity to variations in starch and protein content, retaining significant value for constructing high-sensitivity spectral models. The CARS algorithm eliminated all wavelengths in the 1005–1332 nm range, suggesting that these wavelengths either lack a significant contribution to seed vigor prediction or may introduce spectral noise. The majority of CARS-selected wavelengths were concentrated in the medium-wave near-infrared region. Specifically, 1333.9 nm and 1348.6 nm likely correspond to the second overtone stretching vibration absorptions of C-H bonds and their combination band absorptions. 1402.5 nm and 1404.9 nm, may be associated with combination band absorptions of C-H bonds and the first overtone stretching vibration absorptions of O-H bonds. 1456.2 nm, 1466 nm, 1480.6 nm, 1492.8 nm, 1514.8 nm, 1517.2 nm, and 1558.6 nm, primarily correspond to the first overtone stretching vibration absorptions of O-H and N-H bonds. This region is critical for moisture detection and also contains rich information related to starch and protein. 1626.5 nm, 1648.2 nm, 1655.5 nm, 1694 nm, 1737.3 nm, 1744.5 nm, 1770.9 nm, and 1775.7 nm, mainly attributed to the first overtone stretching vibration absorptions of C-H bonds. This region is highly sensitive to structural and concentration changes in

starch, serving as a core spectral zone for starch analysis. 1849.6 nm, 1868.5 nm, 1882.8 nm, 1896.9 nm, 1937 nm, and 1948 nm: Among these, 1937 nm and 1948 nm align with strong water absorption peaks, while the remaining wavelengths may be associated with weak absorptions of starch or lipids. Other wavelengths cannot be unambiguously assigned to specific chemical bonds, yet they hold significant value for seed vigor prediction.

Spectral analysis of characteristic wavelengths of seed vigor reveals that vigor is influenced by the synergistic effects of multiple substances, with starch, protein, lipids, and moisture being the primary contributors. Although certain substances directly related to vigor have not yet been clearly identified through current analysis of spectral wavelengths, the spectral absorption characteristics formed by these substances collectively play a decisive role in constructing high-precision vigor prediction models. Appendix A provides a detailed list of all characteristic wavelengths selected by the three algorithms (iPLS, GA, and CARS), which serve as key spectral reference data for breeding experts to further explore the material basis of rice seed vigor.

A minor limitation of this study is that only one rice variety was tested. In subsequent research, we will gradually expand the scope to multiple rice varieties to further verify the stability and applicability of the methods employed in this experiment.

4. Conclusions

This study explored the potential of NIR-DRS and NIR-TS for assessing seed vigor. Wavelength selection algorithms were combined to extract characteristic spectra that can be associated with substances related to seed vigor. Spectral analysis revealed there were minimal spectral differences between viable and non-viable seeds. As storage time increases, the spectral absorption of seeds gradually intensifies, while the spectral differences between viable and non-viable seeds gradually diminish. All preprocessing methods exhibit model collapse during cross-validation, making it challenging to directly determine seed vigor using the two spectroscopic techniques. After amplifying spectral differences between viable and non-viable seeds, near-infrared transmission spectroscopy can effectively distinguish between high-viable and non-viable seeds, achieving test set accuracy and germination rates of 84.52% and 88.57%, respectively. Selecting wavelengths from the full-wavelength spectrum using three methods (iPLS, GA, and CARS) significantly improved model prediction accuracy. The CARS algorithm performed best among these, as shown by the results. The CARS algorithm selected 38 feature wavelengths related to seed vigor, and the PLS-DA model constructed with these wavelengths achieved high test set accuracy and germination rates of 90.47% and 95.38%, respectively. Spectral analysis of characteristic wavelengths revealed that seed vigor is influenced by the synergistic effects of multiple substances, with starch, proteins, lipids, and moisture serving as primary contributing factors. Although some wavelengths have not yet been associated with specific substances, they are significant for predicting seed vigor. Breeding experts can identify additional substances related to seed vigor by analyzing characteristic wavelengths, ultimately advancing production standards for high-vigor seeds.

Author Contributions: Conceptualization, Q.H.; Methodology, Q.H., J.W., M.Z. and W.N.; Software, Q.H.; Validation, Q.H. and J.W.; Formal analysis, Q.H.; Investigation, Q.H. and M.Z.; Resources, W.N. and Z.X.; Data curation, Q.H. and J.W.; Writing—original draft, Q.H. and J.W.; Writing—review & editing, J.C., M.Z., W.N., M.H., Z.X., R.K. and W.L.; Supervision, W.N.; Funding acquisition, X.W., M.H., Z.X. and R.K. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the HFIPS Director's Fund (No.YZJJ202302-CX), the National Key R&D Program of China (No. 2023YFF0614002), the Jianghuai Frontier Technology Synergetic Innovation Center Autonomous Research Project (No. 00QK0054), and the Youth Innovation Promo-

tion Association of Chinese Academy of Sciences (No. 2022451). The APC was funded by the HFIPS Director's Fund (YZJJ202302-CX).

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

Conflicts of Interest: The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Appendix A. Characteristic Wavelengths Selected by Three Algorithms

Table A1. Wavelength selection results of iPLS, GA and CARS algorithms.

Methods	Wavelength (nm)
iPLS	1125–1235, 1463–1684, 1906–2132
GA	901, 911, 916, 918, 931, 933, 940, 945, 950, 953, 962, 967, 970, 972, 975, 977, 985, 987, 994, 997, 1002, 1004, 1007, 1014, 1016, 1024, 1029, 1034, 1036, 1039, 1043, 1053, 1056, 1066, 1068, 1073, 1083, 1088, 1090, 1095, 1100, 1107, 1117, 1122, 1127, 1132, 1137, 1142, 1144, 1154, 1161, 1174, 1176, 1181, 1186, 1208, 1211, 1218, 1220, 1235, 1243, 1250, 1260, 1265, 1272, 1275, 1279, 1284, 1287, 1289, 1294, 1314, 1319, 1321, 1324, 1331, 1343, 1346, 1348, 1360, 1368, 1370, 1375, 1385, 1392, 1395, 1397, 1402, 1404, 1409, 1414, 1417, 1419, 1422, 1429, 1431, 1436, 1439, 1456, 1463, 1473, 1478, 1488, 1492, 1495, 1497, 1500, 1505, 1507, 1509, 1512, 1514, 1517, 1519, 1522, 1526, 1529, 1536, 1541, 1544, 1556, 1558, 1561, 1563, 1568, 1580, 1582, 1587, 1597, 1604, 1607, 1609, 1616, 1636, 1638, 1641, 1643, 1648, 1650, 1655, 1660, 1665, 1667, 1669, 1674, 1682, 1689, 1696, 1710, 1720, 1722, 1725, 1730, 1734, 1737, 1739, 1744, 1749, 1763, 1766, 1768, 1770, 1773, 1775, 1778, 1780, 1782, 1785, 1792, 1797, 1804, 1809, 1811, 1816, 1823, 1828, 1832, 1840, 1842, 1847, 1849, 1868, 1875, 1880, 1882, 1889, 1896, 1899, 1904, 1906, 1908, 1915, 1918, 1920, 1922, 1925, 1930, 1932, 1937, 1939, 1941, 1951, 1955, 1958, 1960, 1967, 1972, 1984, 1986, 1988, 1991, 1995, 1998, 2007, 2012, 2014, 2016, 2019, 2021, 2023, 2028, 2033, 2037, 2040, 2046, 2049, 2051, 2063, 2065, 2067, 2070, 2072, 2081, 2086, 2095, 2097, 2104, 2111, 2120, 2125, 2127, 2130
CARS	901, 911, 916, 918, 945, 1002, 1004, 1333, 1348, 1402, 1404, 1456, 1466, 1480, 1492, 1514, 1517, 1558, 1626, 1648, 1655, 1694, 1737, 1744, 1770, 1775, 1806, 1849, 1868, 1882, 1896, 1937, 1948, 1984, 2037, 2040, 2116, 2120

References

1. FAO. Crops and Livestock Products. 2023. Available online: https://www.fao.org/statistics/en (accessed on 28 August 2025).
2. Yamane, K.; Garcia, R.; Imayoshi, K.; Mabesa-Telosa, R.C.; Banayo, N.P.M.C.; Vergara, G.; Yamauchi, A.; Sta Cruz, P.; Kato, Y. Seed vigour contributes to yield improvement in dry direct-seeded rainfed lowland rice. *Ann. Appl. Biol.* **2018**, *172*, 100–110. [CrossRef]
3. Xing, M.; Long, Y.; Wang, Q.; Tian, X.; Fan, S.; Zhang, C.; Huang, W. Physiological Alterations and Nondestructive Test Methods of Crop Seed Vigor: A Comprehensive Review. *Agriculture* **2023**, *13*, 527. [CrossRef]
4. Zhao, J.; He, Y.; Huang, S.; Wang, Z. Advances in the Identification of Quantitative Trait Loci and Genes Involved in Seed Vigor in Rice. *Front. Plant Sci.* **2021**, *12*, 659307. [CrossRef] [PubMed]
5. Zhang, J.; Fang, W.; Xu, C.; Xiong, A.; Zhang, M.; Goebel, R.; Bo, G. Current Optical Sensing Applications in Seeds Vigor Determination. *Agronomy* **2023**, *13*, 1167. [CrossRef]
6. *GB/T 3543.4-1995*; Rules for Agricultural Seed Testing—Germination Test. The State Bureau of Quality and Technical Supervision: Beijing, China, 1995; p. 24.
7. An, T.; Fan, Y.; Tian, X.; Wang, Q.; Wang, Z.; Fan, S.; Huang, W. Green analytical assay for the viability assessment of single maize seeds using double-threshold strategy for catalase activity and malondialdehyde content. *Food Chem.* **2024**, *455*, 139889. [CrossRef]
8. Wang, S.; Wu, M.; Zhong, S.; Sun, J.; Mao, X.; Qiu, N.; Zhou, F. A Rapid and Quantitative Method for Determining Seed Viability Using 2,3,5-Triphenyl Tetrazolium Chloride (TTC): With the Example of Wheat Seed. *Molecules* **2023**, *28*, 6828. [CrossRef]
9. Sahoo, S.; Sanghamitra, P.; Nanda, N.; Pawar, S.; Pandit, E.; Bastia, R.; Muduli, K.C.; Pradhan, S.K. Association of molecular markers with physio-biochemical traits related to seed vigour in rice. *Physiol. Mol. Biol. Plants* **2020**, *26*, 1989–2003. [CrossRef]
10. Qiao, J.; Liao, Y.; Yin, C.; Yang, X.; Tu, H.M.; Wang, W.; Liu, Y. Vigour testing for the rice seed with computer vision-based techniques. *Front. Plant Sci.* **2023**, *14*, 1194701. [CrossRef]
11. Zhao, G.W.; Zhong, T.L. Improving the assessment method of seed vigor in *Cunninghamia lanceolata* and *Pinus massoniana* based on oxygen sensing technology. *J. For. Res.* **2012**, *23*, 95–101. [CrossRef]
12. Feng, L.; Hou, T.; Wang, B.; Zhang, B. Assessment of rice seed vigour using selected frequencies of electrical impedance spectroscopy. *Biosyst. Eng.* **2021**, *209*, 53–63. [CrossRef]

13. Zhang, H.; Kang, K.; Wang, C.; Sun, Q.; Luo, B. Cross-variety seed vigor detection using new spectral analysis techniques and ensemble learning methods. *J. Food Compos. Anal.* **2024**, *136*, 106845. [CrossRef]
14. Ambrose, A.; Lohumi, S.; Lee, W.-H.; Cho, B.K. Comparative nondestructive measurement of corn seed viability using Fourier transform near-infrared (FT-NIR) and Raman spectroscopy. *Sens. Actuators B Chem.* **2016**, *224*, 500–506. [CrossRef]
15. Jia, L.; Qi, H.; Hu, W.; Zhao, G.; Kan, R.; Gao, L.; Zheng, W.; Xu, Q. Rapid Nondestructive Grading Detection of Maize Seed Vigor Using TDLAS Technique. *Chin. J. Lasers* **2019**, *46*, 911002. [CrossRef]
16. Guo, Z.; Fan, Y.; Zhai, B.; Xie, R.; Shang, Z.; Tian, Y.; Qiu, X.; Li, C. Detection of seed viability by photoacoustic carbon dioxide sensing. *Opt. Precis. Eng.* **2025**, *33*, 367–376. [CrossRef]
17. Wu, N.; Weng, S.; Chen, J.; Xiao, Q.; Zhang, C.; He, Y. Deep convolution neural network with weighted loss to detect rice seeds vigor based on hyperspectral imaging under the sample-imbalanced condition. *Comput. Electron. Agric.* **2022**, *196*, 106850. [CrossRef]
18. Fan, X.; Zhu, M.; Yang, C.; Xie, H.; Tang, G.; Deng, H.; Zeng, X.; Kan, R.; He, Y.; Yu, Y. Assessment of Rice Seed Vigor Using Near Infrared Spectroscopy. *Hybrid Rice* **2019**, *34*, 62–67.
19. Fan, S.; Xu, Z.; Cheng, W.; Wang, Q.; Yang, Y.; Guo, J.; Zhang, P.; Wu, Y. Establishment of Non-Destructive Methods for the Detection of Amylose and Fat Content in Single Rice Kernels Using Near-Infrared Spectroscopy. *Agriculture* **2022**, *12*, 1258. [CrossRef]
20. 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]
21. Qi, H.; Huang, Z.; Jin, B.; Tang, Q.; Jia, L.; Zhao, G.; Cao, D.; Sun, Z.; Zhang, C. SAM-GAN: An improved DCGAN for rice seed viability determination using near-infrared hyperspectral imaging. *Comput. Electron. Agric.* **2024**, *216*, 108473. [CrossRef]
22. 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]
23. Yang, Y.; Chen, J.; He, Y.; Liu, F.; Feng, X.; Zhang, J. Assessment of the vigor of rice seeds by near-infrared hyperspectral imaging combined with transfer learning. *RSC Adv.* **2020**, *10*, 44149–44158. [CrossRef] [PubMed]
24. Jin, W.-l.; Cao, N.-l.; Zhu, M.-d.; Chen, W.; Zhang, P.-g.; Zhao, Q.-l.; Liang, J.-q.; Yu, Y.-h.; Lv, J.-g.; Kan, R.-f. Nondestructive grading test of rice seed activity using near infrared super-continuum laser spectrum. *Chin. Opt.* **2020**, *13*, 1032–1043. [CrossRef]
25. Chen, J.; Li, M.; Pan, T.; Pang, L.; Yao, L.; Zhang, J. Rapid and non-destructive analysis for the identification of multi-grain rice seeds with near-infrared spectroscopy. *Spectrochim. Acta Part A Mol. Biomol. Spectrosc.* **2019**, *219*, 179–185. [CrossRef] [PubMed]
26. Zou, X.; Zhao, H.; Li, Y. Selection of the efficient wavelength regions in FT-NIR spectroscopy for determination of SSC of 'Fuji' apple based on BiPLS and FiPLS models. *Vib. Spectrosc.* **2007**, *44*, 220–227. [CrossRef]
27. Li, H.; Liang, Y.; Xu, Q.; Cao, D. Key wavelengths screening using competitive adaptive reweighted sampling method for multivariate calibration. *Anal. Chim. Acta* **2009**, *648*, 77–84. [CrossRef]
28. Xu, Z.; Fan, S.; Cheng, W.; Liu, J.; Zhang, P.; Yang, Y.; Xu, C.; Liu, B.; Liu, J.; Wang, Q.; et al. A correlation-analysis-based wavelength selection method for calibration transfer. *Spectrochim. Acta Part A Mol. Biomol. Spectrosc.* **2020**, *230*, 118053. [CrossRef]
29. Kan, X.; Li, Y.; Wang, L.; Xie, G.; Meng, Y.; Li, C.; Xie, J.; Li, Y. Moisture content detection of Fraxinus mandshurica logs at low temperatures based on different spectrum pretreatments. *J. Cent. South Univ. For. Technol.* **2022**, *42*, 154–163.
30. Sun, J.; Xia, Y. Pretreating and normalizing metabolomics data for statistical analysis. *Genes Dis.* **2024**, *11*, 100979. [CrossRef]
31. Maleki, M.R.; Mouazen, A.M.; Ramon, H.; De Baerdemaeker, J. Multiplicative scatter correction during on-line measurement with near infrared spectroscopy. *Biosyst. Eng.* **2007**, *96*, 427–433. [CrossRef]
32. Fearn, T.; Riccioli, C.; Garrido-Varo, A.; Guerrero-Ginel, J.E. On the geometry of SNV and MSC. *Chemom. Intell. Lab. Syst.* **2009**, *96*, 22–26. [CrossRef]
33. Wang, J.; Lin, T.; Ma, S.; Ju, J.; Wang, R.; Chen, G.; Jiang, R.; Wang, Z. The qualitative and quantitative analysis of industrial paraffin contamination levels in rice using spectral pretreatment combined with machine learning models. *J. Food Compos. Anal.* **2023**, *121*, 105430. [CrossRef]
34. Bhadra, S.; Sagan, V.; Maimaitijiang, M.; Maimaitiyiming, M.; Newcomb, M.; Shakoor, N.; Mockler, T.C. Quantifying Leaf Chlorophyll Concentration of Sorghum from Hyperspectral Data Using Derivative Calculus and Machine Learning. *Remote Sens.* **2020**, *12*, 2082. [CrossRef]
35. Žerovnik, J.; Rupnik Poklukar, D. Elementary methods for computation of quartiles. *Teach. Stat.* **2017**, *39*, 88–91. [CrossRef]
36. Greenacre, M.; Groenen, P.J.F.; Hastie, T.; D'Enza, A.L.; Markos, A.; Tuzhilina, E. Principal component analysis. *Nat. Rev. Methods Primers* **2022**, *2*, 100. [CrossRef]
37. Yun, Y.-H.; Li, H.-D.; Deng, B.-C.; Cao, D.-S. An overview of variable selection methods in multivariate analysis of near-infrared spectra. *TrAC Trends Anal. Chem.* **2019**, *113*, 102–115. [CrossRef]
38. Lucasius, C.B.; Beckers, M.L.M.; Kateman, G. Genetic algorithms in wavelength selection—A comparative-study. *Anal. Chim. Acta* **1994**, *286*, 135–153. [CrossRef]
39. Ballabio, D.; Consonni, V. Classification tools in chemistry. Part 1: Linear models. *PLS-DA. Anal. Methods* **2013**, *5*, 3790–3798. [CrossRef]

40. Zhang, T.-t.; Xiang, Y.-y.; Yang, L.-m.; Wang, J.-h.; Sun, Q. Wavelength Variable Selection Methods for Non-Destructive Detection of the Viability of Single Wheat Kernel Based on Hyperspectral Imaging. *Spectrosc. Spectr. Anal.* **2019**, *39*, 1556–1562.
41. Solout, M.V.; Zade, S.V.; Abdollahi, H.; Ghasemi, J.B. Enhanced data point importance for subset selection in partial least squares regression: A comparative study with Kennard-Stone method. *Chemom. Intell. Lab. Syst.* **2025**, *263*, 105416. [CrossRef]
42. He, Z. Study on the Starch Properties of Rice Kernels. *Seed* **1982**, *3*, 41–43. [CrossRef]
43. Nie, L.; Song, S.; Yin, Q.; Zhao, T.; Liu, H.; He, A.; Wang, W. Enhancement in Seed Priming-Induced Starch Degradation of Rice Seed Under Chilling Stress via GA-Mediated α-Amylase Expression. *Rice* **2022**, *15*, 19. [CrossRef]
44. Guo, J. Investigating the Molecular Mechanism and Genetic Application of OsMBF1a on Regulating Seed Germination and Stress Response in Rice. Master's Thesis, Hubei University, Wuhan, China, 2024.
45. Khin, O.M.; Sato, M.; Li-Tao, T.; Matsue, Y.; Yoshimura, A.; Mochizuki, T. Close Association between Aleurone Traits and Lipid Contents of Rice Grains Observed in Widely Different Genetic Resources of Oryza sativa. *Plant Prod. Sci.* **2013**, *16*, 41–49. [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.