Check for updates

Detection of peanut seed vigor based on hyperspectral imaging and chemometrics

Zhiyong Zou[1†], Jie Chen[1†], Weijia Wu[1], Jinghao Luo[1], Tao Long[1], Qingsong Wu[1], Qianlong Wang[1], Jiangbo Zhen[1], Yongpeng Zhao[1], Yuchao Wang[1], Yongming Chen[2], Man Zhou[3*] and Lijia Xu[1*]

[1]College of Mechanical and Electrical Engineering, Sichuan Agricultural University, Yaan, China,
[2]School of Electrical Engineering and Automation, Hubei Normal University, Huangshi, Hubei, China,
[3]Food Academy, Sichuan Agricultural University, Yaan, China

Rapid nondestructive testing of peanut seed vigor is of great significance in current research. Before seeds are sown, effective screening of high-quality seeds for planting is crucial to improve the quality of crop yield, and seed vitality is one of the important indicators to evaluate seed quality, which can represent the potential ability of seeds to germinate quickly and whole and grow into normal seedlings or plants. Meanwhile, the advantage of nondestructive testing technology is that the seeds themselves will not be damaged. In this study, hyperspectral technology and superoxide dismutase activity were used to detect peanut seed vigor. To investigate peanut seed vigor and predict superoxide dismutase activity, spectral characteristics of peanut seeds in the wavelength range of 400-1000 nm were analyzed. The spectral data are processed by a variety of hot spot algorithms. Spectral data were preprocessed with Savitzky-Golay (SG), multivariate scatter correction (MSC), and median filtering (MF), which can effectively to reduce the effects of baseline drift and tilt. CatBoost and Gradient Boosted Decision Tree were used for feature band extraction, the top five weights of the characteristic bands of peanut seed vigor classification are 425.48nm, 930.8nm, 965.32nm, 984.0nm, and 994.7nm. XGBoost, LightGBM, Support Vector Machine and Random Forest were used for modeling of seed vitality classification. XGBoost and partial least squares regression were used to establish superoxide dismutase activity value regression model. The results indicated that MF-CatBoost-LightGBM was the best model for peanut seed vigor classification, and the accuracy result was 90.83%. MSC-CatBoost-PLSR was the optimal regression model of superoxide dismutase activity value. The results show that the R^2 was 0.9787 and the RMSE value was 0.0566. The results suggested that hyperspectral technology could correlate the external manifestation of effective peanut seed vigor.

1 Introduction

Peanut is an important oil crop mainly used for oil production, and by-products of peanut also contain many other functional compounds such as protein, fiber, polyphe-nols, antioxidants, vitamins and minerals, which can be added as functional ingredients in many processed foods (Arya et al., 2016). Factors affecting the yield and quality of peanuts include climate temperature and humidity, fertilization ratio, sowing density, degree of pests and diseases, and seed vigor of peanut seeds(Gomes et al., 2009). Among many factors, seed vigor plays a crucial role (Tu et al., 2022). If the problem of aging or vigor decline of peanut seeds during storage is not solved, it will cause delayed seed germination (Wang et al., 2021), poor growth potential, weak stress resistance, and biological Yields and economic yields are reduced, thereby affecting agricultural production. Therefore, in the production and application of peanut crops, judging the vigor of peanut seeds has a guiding role in improving the yield and quality of peanuts.

Generally speaking, the seed vigor of peanut is difficult to judge by manual observation. The appearance of aged peanut seeds is almost the same as that of fresh peanut seeds except that the gloss of the seed surface is slightly rough. Therefore, methods such as standard seed germination detection method (Zhang et al., 2020), field emergence test method (Kaya et al., 2019), conductivity measurement method (Xiang et al., 2020), TTC detection method (Liu et al., 2009), and red ink measurement method (Mattioni et al., 2015) are usually used to identify whether peanut seeds are fresh or aged. However, these methods require experienced operators, and farmers cannot master this skill well. At the same time, these methods are highly destructive, inefficient, time-consuming, and complicated in operation procedures, and are not suitable for rapid non-destructive testing of peanut seed viability. Compared with traditional destructive inspection methods, machine vision technology (Tan et al., 2022) and near-infrared spectroscopy technology (Jasinski et al., 2016) have been widely used in seed quality inspection, such as variety (Kotwaliwale et al., 2019), mildew (Fu et al., 2017), vigor (Liu et al., 2021), and defect (Wang et al., 2022). However, machine vision only uses phenotypic features, such as color, size, shape and surface texture, and is not suitable for predicting the chemical composition of a sample, because the internal chemical content of seeds changes after aging (Huang & Chien, 2017); Near-infrared spectroscopy can be used to assess the chemical composition of a sample, but it is only used to obtain spectral information from a single point of light, and is always affected by the uniformity of the sample distribution, and is not the best choice (ElMasry et al., 2019).

Composed of spatial imaging, spectroscopy (Kucha et al., 2021), and chemical measurement tools, hyperspectral imaging techniques (Xu et al., 2021) can provide information on seed quality characteristics and characterization parameters(Jun Yang et al., 2021), overcoming the limitations of machine vision and near-infrared spectroscopy techniques(Laborde et al., 2021). In recent years, some studies have used hyperspectral imaging technology as a powerful tool for seed vigor monitoring. Compared with the traditional seed vigor detection technology based on biological properties, the non-destructive seed vigor detection technology based on hyperspectral imaging technology is gradually attracting the attention of the seed industry. Some scholars only use hyperspectral imaging technology to discriminate seed vigor. Such as He Xiantao et al. used hyperspectral imaging technology to predict rice seed viability within three years, combined with various preprocessing and feature band extraction algorithms, the results showed that the classification accuracy reached 94.38% (He et al., 2019). Some studies combine hyperspectral imaging techniques (Zou et al., 2022) and texture feature methods. Such as Wang Zheli et al. used hyperspectral imaging technology and chemometrics to quickly and nondestructively classify new and old corn seeds, and established a model of spectral and image texture features, with a classification accuracy rate of 95% (Wang et al., 2022). There are also studies that integrate hyperspectral imaging technology and mapping technology. For example, Yan Lei et al. used hyperspectral imaging technology combined with machine learning and convolutional neural network to identify seed germination with an accuracy of 99.96% (Pang et al., 2020).

Previous studies have demonstrated the potential of hyperspectral imaging techniques and provided a good reference for the field of seed quality inspection. However, most of the current research stays at the macroscopic stage, and only uses hyperspectral technology to identify whether the seeds are aging or not, and has not analyzed the factors affecting the seed vigor. In fact, factors that affect seed vigor include seed moisture content, protein content, superoxide dismutase content (SOD) (Matłok et al., 2022), catalase (CAT), ascorbate peroxidase (As A-POD) and guaiacol peroxidase (G-POD) (Zhao et al., 2021). Among them, SOD is involved in catalyzing the disproportionation reaction with reactive oxygen species and free radicals as substrates, and its activity level directly affects the seed vigor (Bandeira et al., 2014). Therefore, the SOD of peanut seeds was measured and modeled in this paper.

The overall goal of this study was to examine the potential of hyperspectral imaging in the detection of aged peanut seeds using samples of varying degrees of aging, while establishing a microscopic content analysis of superoxide dismutase in peanut seeds. The specific goals are (1) To establish a hyperspectral-based classification model for peanut seed viability detection through standard roll paper germination tests; (2) To identify and evaluate the best characteristic wavelengths for peanut seed detection; (3) Determination of peanut seed SOD value by tetrazolium blue (NBT) method, and establishment of a peanut seed SOD regression model based on hyperspectral spectrum. From the perspective of spectroscopy, it is revealed that the characteristic band, seed vigor index and SOD value have strong correlation, which provides a new method for seed quality evaluation.

2 Materials and methods

The process of raw material processing is shown in Figure 1. Figure 1A shows that all peanut seeds are divided into 5 types of peanut original seed treatments after removing the factors such as damage and mildew, namely A0~A4, A0 peanut seeds represent

FIGURE 1
Flow chart of raw material processing steps: **(A)** Peanut aging treatment; **(B)** Two types of germination and SOD determination; **(C)** Hyperspectral determination; **(D)** Spectral data analysis; **(E)** Roll paper germination test; **(F)** Peanut superoxide dismutase enzyme activity assay.
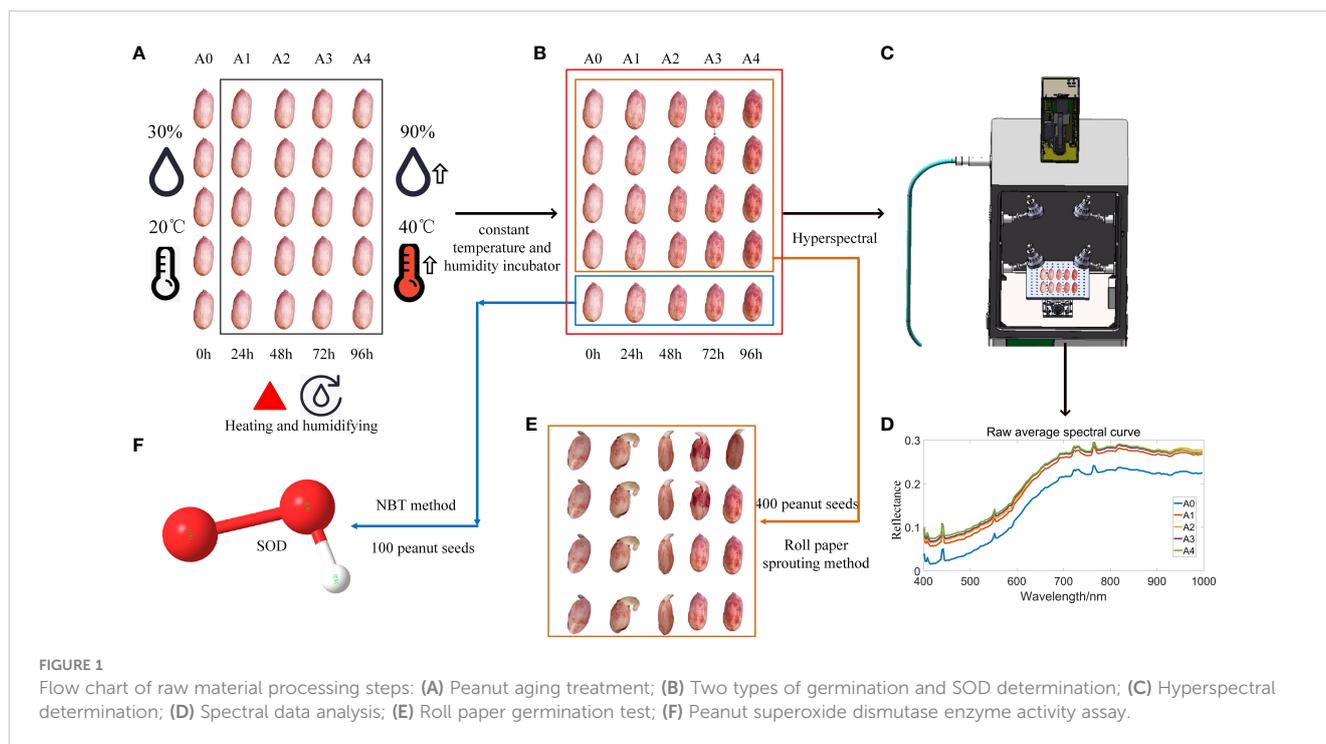
fresh and unaged peanut seeds, which are cultivated in an environment with a temperature of 30°C and a humidity of 30%. At the same time, the peanut seeds of groups A1~A4 were artificially aged and placed in a constant temperature and humidity incubator with a temperature of 40°C and a humidity of 90% to simulate natural aging. Among them, the aging time of peanut seeds in groups A1~A4 ranged from 24h to 96h. As shown in Figure 1B, the peanut seeds obtained after treatment were divided into two groups, and 400 peanut seeds in one group were used for the roll paper germination test, which was used for the establishment of the subsequent peanut seed vigor classification model. Another group of 100 peanut seeds was used for the determination of superoxide dismutase value and used for the establishment of regression model of peanut seed superoxide dismutase value. Collect hyperspectral images of all peanut seeds to obtain their spectral data information, as shown in Figure 1C, and obtain the spectral curve of peanut seeds, as shown in Figure 1D. The peanut seeds tested in the roll paper germination test are shown in Figure 1E, and the relevant indicators of seed vigor were recorded, including germination rate, germination potential, germination index, average germination time, vitality index, and simple vitality index. As shown in Figure 1F, the SOD value was measured and its data information was recorded. Finally, the correlation analysis of the average spectral reflectance value, germination rate, germination potential, germination index, average germination time, vigor index, simple vigor index, and the average value of superoxide dismutase in each group of A0~A4 groups was established.

2.1 Sample preparation

In order to expand peanut seed samples with different vigor, it is necessary to perform aging of peanut seeds. Studies have shown that the overall metabolic pathways of natural aging and artificial aging were similar (Xu et al., 2013), and there was no significant difference in the aging mechanism. Therefore, artificial aging was used to treat peanut seeds. 2.5 kg of Sichuan Tianfu peanuts were purchased in the year of 2021. The peanut seeds with no mildew and damage were selected, and were divided into 5 equal parts. Each part of 0.5 kg was evenly spread in the storage tray. One group of peanuts was used as the reference group and placed at room temperature (temperature of 20°C, relative humidity of 30%) as unaged peanut seeds (A0), while the other 4 groups of peanuts were placed in an intelligent constant temperature and humidity incubator with a constant temperature. The humidity of the constant temperature and humidity incubator was set to 90%, and the temperature was set to 40°C. On the first day, the first group of peanuts was placed in the incubator and labeled as aging treatment 4d (A4). On the second day, the second group of peanuts was placed in the incubator and labeled as aging treatment 3d (A3). On the third day, the third group of peanuts was placed in the incubator and labeled as aging treatment 2d (A2). On the fourth day, the fourth group of peanuts was placed in the incubator and labeled as aging treatment 1d (A1). When the aging treatment reached the fifth day, all peanuts were uniformly taken out and placed at room temperature for 2 days to make their moisture content reach a similar level, and then placed in a low temperature box of use.

2.2 Hyperspectral imaging system and analysis and processing

The hyperspectral system tested is the GaiaSorter Hyperspectral Sorter from Zolix Co., ltd. (Beijing, China). And integrated Imspector series of Imaging spectrometers from Spectral Imaging Ltd. (Oulu, Finnish). The hyperspectral camera model is Gaiafluo-VN-HR(Zolix Co., ltd., Beijing, China), the spectrometer adopts a transmission grating (PGP) structure, the spectral sampling rate is 0.6 nm, and the scanning mode is built-in push-broom. The spectral range was 384-1034nm, the spectral resolution was 2.8nm, the spectral collection point was 0.65nm, the effective slit length was 8.7mm, the relative aperture was F/2.4, the slit width was 30um, A/D The output is 12, the exposure time range is 0.01-10000ms, the camera pixel is 1344×1024, the pixel size is 6.45×6.45um, the imager power consumption is 8V·A, the rated power of halogen light source is 200W, and the scanning stroke of the sample station is 400mm. The control software name is SpecView (SpecView ltd., Uckfield, UK). Hyperspectral equipment core components included a uniform light source, a spectral camera, an electronically controlled mobile platform (or conveyor belt), a computer and control software. Its working principle was to illuminate the object to be measured (sample) placed on the electronically controlled moving platform (or conveyor belt) through the light source, and the emitted light of the sample was captured by the spectral camera through the lens, and an one-dimensional image and spectral information were obtained. The moving platform (or conveyor belt) drove the sample to run continuously, so that continuous one-dimensional images and real-time spectral information can be obtained. All data were recorded by computer software to obtain a three-dimensional data cube containing image information and spectral information. The hyperspectral instrument is located in an open room of 15 square meters, Before the experiment starts, the curtains in the room are drawn to block all external light, the halogen light source inside the hyperspectral instrument is turned on, and the brightness of the indoor halogen light source is adjusted, and after the hyperspectral image is not saturated and distorted, the official collection of hyperspectral data collection information begins.

Before the image acquisition experiment, the instrument was warmed up for 30 mins, and then the instrument was adjusted and the exposure time of the camera was determined to be 10 ms, and the moving speed of the displacement platform was 1.5 mm/s. The peanut seed samples were placed on the black cardboard of the sample tray in 4 rows and 5 columns. After movement towards the electronically controlled platform, the hyperspectral imaging instrument captured the hyperspectral image of the sample and transmitted it to the computer for storage, and scanned each aging process. Gradient peanut seeds were 100, and a total of 500 seed samples were scanned. After the acquisition was completed, the original hyperspectral image was subjected to black and white correction processing to remove the dark current noise introduced in camera. The formula for black and white correction is as follows:

$$R_{ci} = \frac{Sample_{ci} - dark_{ci}}{White_{ci} - dark_{ci}} \qquad (1)$$

In the formula, $Sample_{ci}$ is the original sample data, $dark_{ci}$ is the dark background data, $White_{ci}$ is the whiteboard data, and R_{ci} is the corrected sample data.

2.3 Standard germination test

Two layers of wet filter paper were spread in the germination box, 80 peanut seeds were selected from each group of the collected seeds, and placed them in the germination box in the order of spectrum collection according to the arrangement of 10×8 in each group. After being sprayed with a small amount of water, the box was closed and placed in an intelligent constant temperature and humidity box, under light (28℃, 8 h) and dark (20℃, 16 h) conditions, and then taken out and germinated every day. The seeds were sprayed and watered, and the relevant vigor indexes were recorded.

The remaining 100 peanut seeds were subjected to SOD superoxide dismutase physical and chemical index test. The relevant calculation formula of the vitality index is as follows:

$$\text{Germination potential}(GP) = \frac{m_1}{M} \times 100\,\% \qquad (2)$$

$$\text{Germination rate}(GE) = \frac{m_2}{M} \times 100\,\% \qquad (3)$$

$$\text{Germination index}(GI) = \sum \frac{Gt}{Dt} \qquad (4)$$

$$\text{Mean germination time}(MGT) = \frac{\sum Gt \times Dt}{\sum Gt} \qquad (5)$$

$$\text{Vitality index}(VI) = GI \times S \qquad (6)$$

$$\text{Simple vitality index}(SVI) = GP \times S \qquad (7)$$

In the formula, m_1 is the number of normally germinated seeds within 3d; m_2 is the number of normally germinated seeds within 8d; M is the total number of test seeds; Dt is the number of days of germination; Gt is the number of germinated seeds per day corresponding to the end of germination; average seedling length (cm); S is the root weight after 8 days of germinated peanuts.

2.4 Superoxide dismutase(SOD) activity assay

SOD activity was closely related to peanut seed vigor and played a key role in its subsequent germination. 20 peanut seeds of 5 aging groups were randomly sampled for SOD activity determination. The method of determination was nitroblue tetrazolium (NBT) method (Kono, 2022). 1 g of peanut seeds was taken, water was

absorbed on filter paper, and then put it in a mortar, 5 ml of pre-cooled phosphate buffer was added, and fully ground into a homogenate on an ice bath, then the homogenate was poured into a 10 mL centrifuge tube, which was shaken for homogenization, and frozen and centrifuged at 4°C and 12000 rpm for 15min with a high-speed centrifuge. The supernatant was the crude SOD extract. The supernatant was poured into a test tube and stored at 0°C-4°C for later use. After mixing the solutions, one control group was placed under dark conditions, while the other groups were placed under 4000 Lux light, and reacted at 20°C for 10-20 min. After the reaction, a small amount of samples were taken in a 96-well microplate, and the absorbance at 560 nm was measured with a multi-function microplate reader Spark10M (A) (Hou et al., 2019).

The formula for calculating SOD (superoxide dismutase) activity (Zhao, 2017) is as follows:

$$\text{SOD activity value } \frac{\text{U}}{\text{gFW}} = \frac{[(A_0 - A_b) - [(A_s - A_b)] \times VT}{\frac{1}{2} \times (A_0 - A_b) \times W \times V_s} \quad (8)$$

In the formula: A_b is the absorbance value of the dark control; A_0 is the absorbance value under light; A_s is the absorbance value of rice seeds; VT is the total volume of enzyme solution (mL); V_s is the volume of enzyme solution used for measurement (ml); W is the fresh weight of the sample.

2.5 Modeling method

Hyperspectral seed vigor classification algorithm mainly uses Extreme gradient boosting(XGBoost), Light Gradient Boosting Machine(LightGBM), Support vector machine(SVM), Random forest(RF); SOD prediction algorithm adopts Partial least squares regression(PLSR) and XGBoost (Chen et al., 2016). XGBoost is an iterative tree-like algorithm that combines multiple weak classifiers together into a strong classifier, which is an implementation of Gradient Boosting Decision Number (Zhang et al., 2021). LightGBM (Ke et al., 2017) is an iterative boosting system, which is an improved variant of gradient decision tree (GBDT). The gradient boosting decision tree in the LightGBM algorithm is obtained by multiple iterations of the given training data set, and in each iteration, a new tree is refitted with the gradient information to join the previous iteration tree, and this in the function space can be regarded as an iterative linear combination process (Wang & Wang, 2020). SVM (Cortes & Vapnik, 1995) constructs the hyperplane as the decision surface and maximizes the isolation edge between the two classes in the classification process (Yang & Gao, 2020). The kernel function used in this study is the radial basis function. RF (Chen & Ishwaran, 2012) classifier is an ensemble classifier model composed of many decision tree classification models. RF classifier is insensitive to parameters, not easy to over fit, and has a fast training speed, which is more suitable for multi-classification problems (Wang & Chen, 2020). PLSR (Deal, 2005) is a new multivariate statistical data analysis method. It mainly studies the regression modeling of multiple dependent variables to multiple independent variables (Burnett et al., 2021).The model structure is shown in Figure 2.

2.6 Spectral preprocessing methods

In the acquisition of raw spectral data, it not only contains useful information, but also is interfered by stray light, instrument
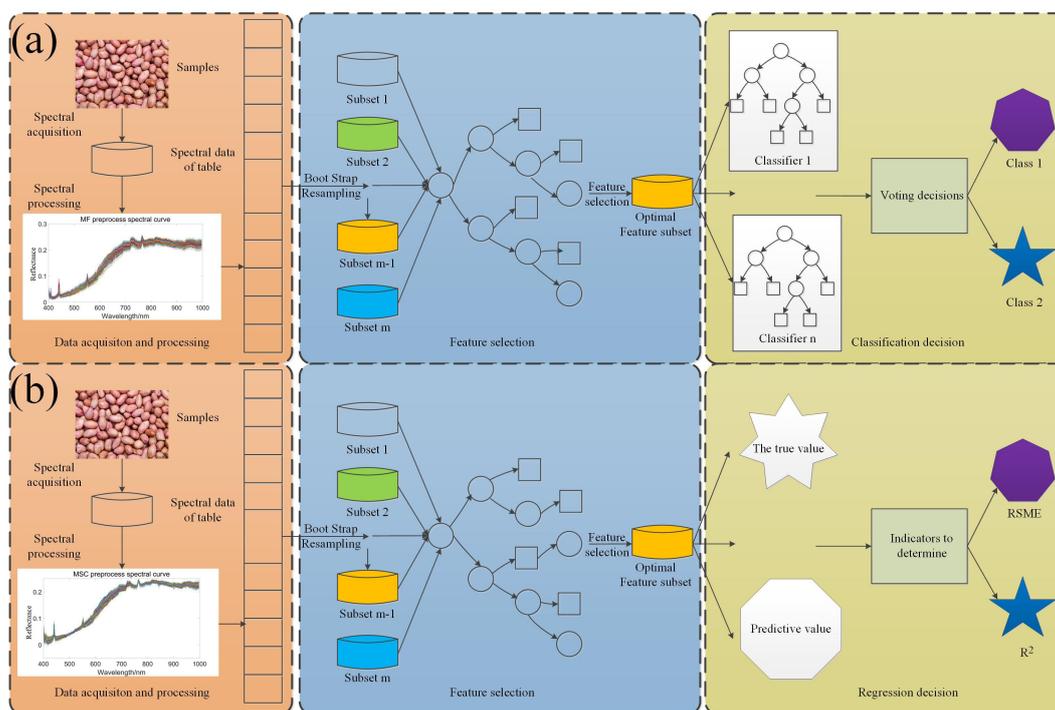
FIGURE 2
Algorithm flow chart: **(A)** Peanut seed viability classification model; **(B)** SOD regression model.

noise, sample background, baseline drift and other factors, all of which affect the selection of characteristic wavelengths, thereby affecting the quantitative and qualitative analysis results of the spectrum. Therefore, in order to make the extracted spectral information more accurately reflect the change of the sample curve, it is necessary to preprocess the original spectrum to eliminate or reduce the influence of light intensity, environmental factors and noise interference on the spectral information as much as possible. In this experiment, the Savitzky-Golay(SG) (Wang et al., 2021), multivariate scatter correction(MSC) (Liang et al., 2018) and median filter(MF) (Adams, 2021) methods were used to preprocess the original data.

2.7 Feature band extraction methods

The spectral data collected by hyperspectral had a large number of bands, high in-formation redundancy, large space required for data storage, long processing time, and was prone to the phenomenon of dimensional disaster, that was, the classification accuracy was reduced. Therefore, it is necessary to extract spectral data below full wavelength. The characteristic wavelengths with strong correlation with peanut seed vigor index were obtained. In this paper, CatBoost (Bentejac et al., 2021) and GBDT (Zhou et al., 2020) were used to select characteristic wavelengths, and the characteristic wavelengths with the top 15 weights were extracted to simplify the establishment of subsequent models and reduce the amount of calculation.

3 Results and discussion

3.1 Standard germination results

After 8 days of germination and culture, the test results of the vitality index were as shown in Table 1.

The results in Table 1 show that, with increase of aging time, the germination potential, germination rate, vigor index and simple vigor index of peanut seeds de-creased significantly, and the average germination time increased, indicating that the aging treatment changed the relevant physical and chemical indexes of seed vigor, thus affecting the vigor of peanut seeds.

3.2 Superoxide dismutase(SOD) activity results

Figure 3 is Boxplot of SOD values of peanuts with different aging degrees. The overall trend that the SOD activity value of aged peanut seeds was significantly lower than that of unaged peanut seeds. The average SOD activity value of unaged peanut seeds is above 3, and the average SOD activity value of aged peanut seeds is below 3, and with the increase of aging time, the SOD activity value is lower.

3.3 Spectral description

In the full band range of 384-1034nm, a large amount of noise in the band range of 384-400nm and 1000nm-1034nm will directly affect the subsequent model establishment results. After removing the head-to-tail noise bands, spectral data of 400-1000nm will be retained for curve drawing. Figure 4A indicate that the raw average spectral curve of peanut seeds show an upward trend, which may be due to the fact that aging accelerated the degradation of the internal storage substances of peanut seeds and reduced the content of organic matter, thereby increasing the spectral reflectance. In the spectral band of 400 nm-700 nm, the spectral reflectance of peanut seeds have a certain separability, which is due to the difference in the content of pigments and organic matter inside the peanut seeds; in the spectral band of 700 nm-1000 nm, the spectral reflectance have obvious differences mainly due to cause by the organic chemical bonds of peanut seeds. Figure 4B is a partial enlargement of Figure 4A. Figure 4B shows there is an obvious difference in spectral reflectance between aged and unaged peanuts, but there is little difference in the spectrum of aged peanut seeds. The peak at 450 nm is due to the strong light reflectivity of starch at 455 nm, 465 nm, and 495 nm (Wang & Wang, 2022). There is a peak change at 550 nm, which is due to the distribution of the characteristic peaks of soluble sugars (Sen, 2016) here. The spectral change around 920 nm is related to the absorption of protein. With the deepening of aging, the protein content in peanut seeds decreased, resulting in the increase of spectral reflectance. The spectral changes around 970 nm corresponded to the secondary ubiquitination stretching of the O-H bond (Raj et al., 2021) and the tertiary stretching of the C-H bond (Choi et al., 2021), causing the lipid peroxide reaction to occur

TABLE 1   Seed vigor index test result table.

aging group	total number of samples	seed vigor	seedless vitality	GP	GE	GI	MGT	VI	SVI
A0	80	77	3	96.25%	96.25%	25.66	2.78	43.37	1.63
A1	80	68	12	76.25%	85%	21.20	3.24	37.52	1.50
A2	80	52	28	53.75%	65%	15.62	4.30	24.68	1.03
A3	80	31	49	28.75%	38.75%	8.95	5.74	14.05	0.61
A4	80	18	62	11.25%	22.5%	3.17	7.21	4.82	0.32

A0, Unaged peanuts; A1, Peanuts aged 24 hours; A2, Peanuts aged 48 hours; A3, Peanuts aged 72 hours; A4, Peanuts aged 96 hours; GP, germination potential; GE, germination rate; GI, germination index; MGT, average germination time; VI, vitality index; SVI, simple vitality index.

FIGURE 3
Boxplot of SOD values of peanuts with different aging degrees.
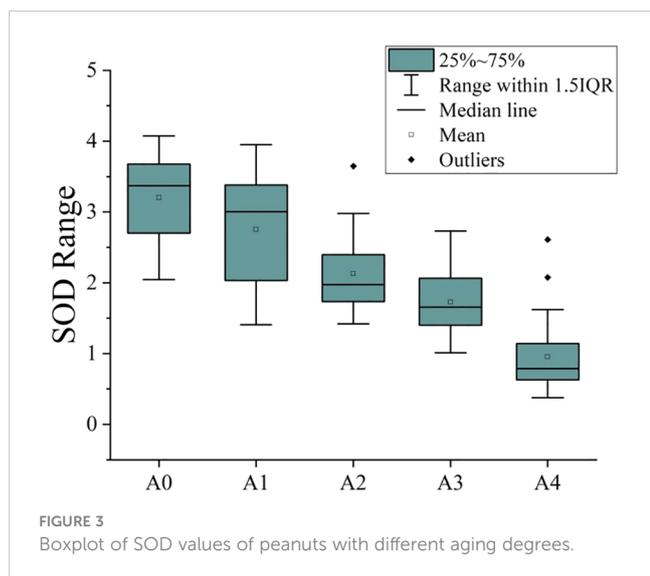
inside the peanut seeds, resulting in the decomposition of organic matter into CO_2 and H_2O (Kou et al., 2022). Figure 4C is the original absorption spectrum extracted from the hyperspectral images of all peanut seeds with different aging gradients, all spectra have the same trend, indicating that peanut seeds have

the same absorption characteristics at full wavelengths, five main absorption peaks are caused by the function of O—H (Liu et al., 2014), C—H (Pjw & Sk, 2016) and other functional groups. However, a large amount of useful information is masked by the irregular appearance of peanuts, and it is necessary to use spectral preprocessing methods to improve the signal-to-noise ratio of the model and display more information. Figure 4D shows the spectrum curve of peanuts after Savitzky-Golay preprocessing. It can be seen that the preprocessed spectrum becomes smoother than the original spectrum curve, which eliminates part of the noise, which is conducive to subsequent modeling and analysis. Figure 4E is the peanut spectrum curve after multivariate scatter correction preprocessing. It can be seen that the preprocessed spectrum becomes more shrunk. This is because the multivariate scatter correction can effectively eliminate the spectral differences caused by different scattering levels, and correct the baseline shift and offset phenomenon of spectral data, thereby enhancing the correlation between spectra and data. Figure 4F is the peanut spectrum curve after preprocessing by median filter. The peaks in the spectrum data tend to be flat, and the spectrum curve is optimized, so that the position where the spectrum curve transitions from the peak to the smooth band can better perform baseline simulation. The fitted baseline changes more gently at this position, which can effectively reduce the occurrence of under-fitting.

FIGURE 4
Spectral reflectance curve of peanuts. (A) Raw average spectral curves of peanut seeds with different aging gradients; (B) Partial enlarged view of spectral reflectance curve; (C) Raw spectral curve of peanut; (D) SG preprocess spectral curve of peanut; (E) MSC preprocess spectral curve of peanut; (F) MF preprocess spectral curve of peanut.
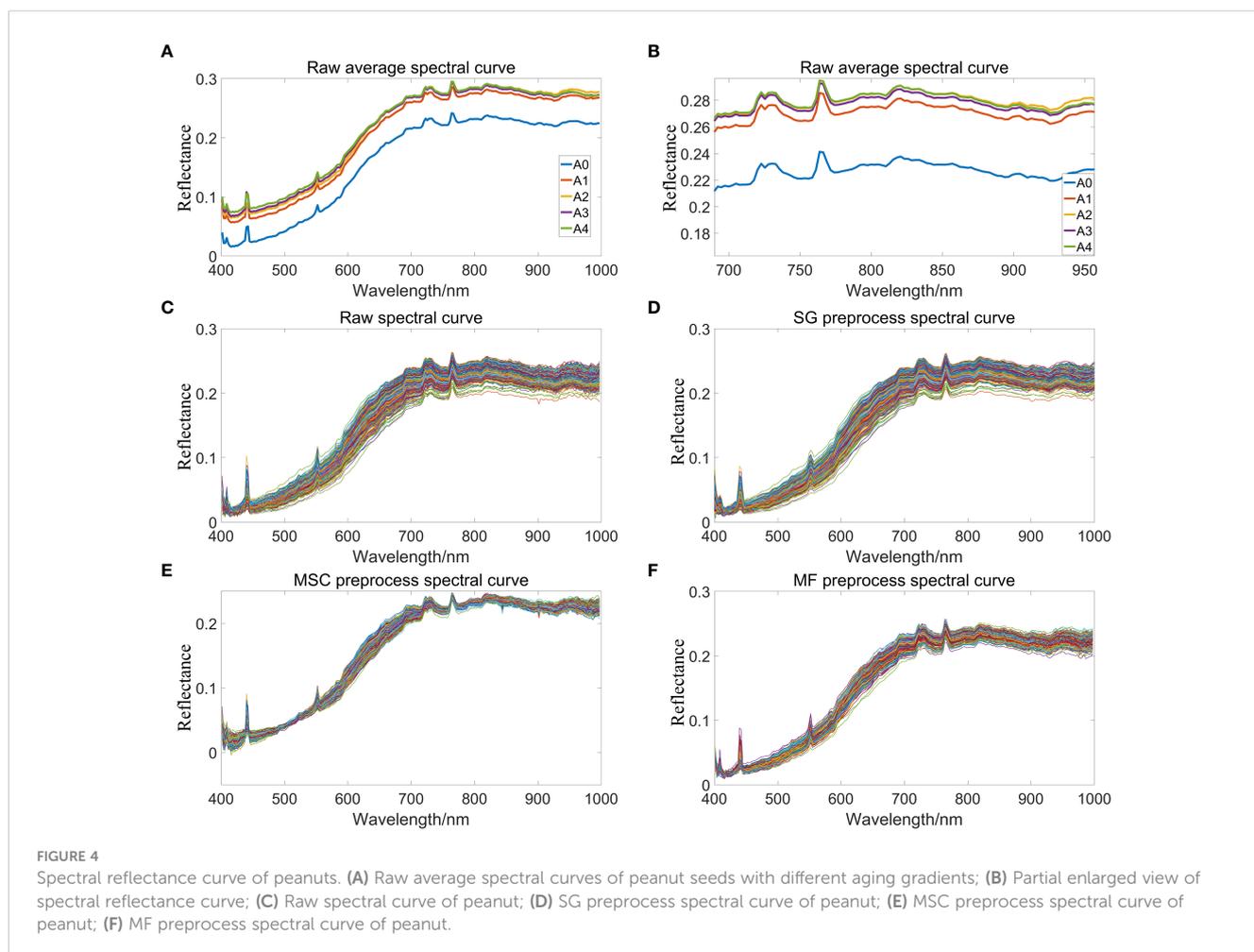
3.4 Seed vigor classification model results

If the length of the germ exceeds 1.5 cm within 8 days of the specified cultivation time, it is judged that the seed is germinating, that is, it has vigor. Otherwise, it is judged that the seed has lost its vigor, and 400 peanut seeds germinated and cultivated with rolled paper were randomly selected according to the ratio of 7:3. They were divided into the training set and the prediction set to train the model and test the performance of the model. There were 280 peanut seeds in the training set, including 173 peanut seeds with seed vigor, 107 peanut seeds without seed vigor, and 120 peanut seeds in the prediction set, including 73 seed vigor peanut seeds and 47 peanut seeds without seed vigor. The model classification results are as shown in Table 2.

After the feature band extraction, the classification accuracy of the modeling results is generally improved. MF is the abbreviation of median filter preprocessing method. In the optimal model MF-CatBoost-LightGBM for CatBoost feature band extraction, the feature wavelengths extracted by CatBoost included 802.51 nm, 878.08 nm, 907.02 nm, 930.8 nm, 938.75 nm, 946.71 nm, 965.32 nm, 967.99 nm, 970.65 nm, 973.32 nm, 981.33 nm, 984.0 nm, 989.35 nm, 992.02 nm, and 994.7 nm. The characteristic wavelength distribution mainly exist in the near-infrared region of 900 nm-1000 nm, because there is a certain correlation between the spectral reflectance and pigment, water and protein in this band region (Liu et al., 2014) as shown in Figure 5D. This result showed that this band region contained more feature information, and the correlation with peanut seed vigor is strong. Table 2 shows that after the feature band extraction of CatBoost method, the accuracy of modeling and classification is generally improved within 5%. The weight map of CatBoost feature band is as shown in Figure 5A. The characteristic wavelengths extracted by GBDT in the optimal model MF-GBDT-XGBoost for GBDT feature band extraction included 403.88 nm, 411.07 nm, 425.48 nm, 549.73 nm, 799.92 nm, 807.69 nm, 833.66 nm, 849.3 nm, 930.8 nm, 944.05 nm, 965.32 nm, 978.66 nm, 984.0 nm, 981.33 nm and, 992.02 nm. The characteristic wavelength distribution is mainly around 430 nm and 900-1000 nm. This is because starch have a strong reflectivity near the 450nm

wavelength (G.-l. Wang et al., 2021), and amino acids, lactose, biological enzymes and other substances have a strong reflectivity in the 900-1030 nm band (Pan et al., 2019), which shows that these substances and bands are strongly related to the vigor of peanut seeds as shown in Figure 5B. After the feature bands are extracted by GBDT method, the accuracy of modeling and classification is generally improved within the range of 1%-10%. The weights of GBDT feature bands are as shown in Figure 5C. However, Table 2 shows that there are still some models due to too few modeling bands after the feature band extraction, and the classification accuracy is reduced. The least modeling effect is with the SVM model. The classification accuracy is between 50% and 70%. The classification effects of XGBoost, LightGBM and Random Forest are all suitable, and the classification accuracy is about 80%. Among them, MF-CatBoost-LightGBM built the model performs the best, with a classification result of 91.07% for the prediction set and 90.83% for the prediction set. MF-CatBoost-LightGBM training set confusion matrix and prediction set confusion matrix are as shown in Figure 6. Figure 6 shows the classification effect of the MF-CatBoost-LightGBM model on peanut seed viability. In the training set, 280 peanut seeds, of which 255 peanut seeds were correctly classified, and in the prediction set, 120 peanut seeds, of which 110 peanut seeds were classified correctly. The results show that hyperspectral technology combined with MF-CatBoost-LightGBM model has high classification performance for peanut seed vigor.

3.5 Prediction result of superoxide dismutase activity value

After the SOD activity of 100 peanut seeds was measured, SG (Savitzky-Golay), MSC(multivariate scatter correction) and MF (median filter) were used to preprocess the raw data, CatBoost and GBDT were used to extract characteristic bands from the original bands, and PLSR and XGBoost regression algorithms were used to predict SOD activity. The SOD activity prediction results are as shown in Table 3. The results in Table 3 show that the data modeling effect after MSC preprocessing is significantly

TABLE 2 Seed vitality classification results table.

		XGBoost			LightGBM			SVM			Random Forest		
		Full	Sel1	Sel2	Full	Sel1	Sel2	Full	Sel1	Sel2	Full	Sel1	Sel2
Raw	Cal	77.50	78.33	85.83	80.83	82.50	81.66	61.79	62.14	65.83	76.07	78.21	77.50
	Pre	76.78	75.71	80.00	78.21	79.64	79.64	58.33	60.83	61.42	73.57	73.93	75.00
SG	Cal	77.5	75.83	85.83	80.83	82.50	81.66	61.79	62.14	65.83	80.00	82.5	80.83
	Pre	76.78	75.71	80.00	78.21	79.64	79.64	58.33	60.83	58.33	76.07	78.21	79.64
MSC	Cal	85.00	88.33	86.66	83.33	85.83	86.66	61.79	62.14	62.50	79.29	83.33	80.83
	Pre	80.36	84.29	84.64	76.43	82.85	75.71	58.33	60.83	59.16	76.67	81.79	79.16
MF	Cal	84.64	90.00	87.14	88.57	91.07	89.16	61.78	62.14	65.83	84.64	85.83	86.66
	Pre	82.50	87.14	88.92	87.14	90.83	87.14	58.33	60.83	62.14	80.71	82.85	86.07

Cal, calibration; Pre, prediction; Raw , raw data; Full , full wavelengths; Sel1 , selected wavelengths by CatBoost; Sel2 , selected wavelengths by GBDT; SG , Savitzky-Golay; MSC , multivariate scatter correction; MF , median filter.

FIGURE 5

Classification model feature band extraction map: **(A)** CatBoost feature band extraction weight map; **(B)** CatBoost feature band extraction in all bands; **(C)** GBDT feature band extraction weight map; **(D)** GBDT feature band extraction in all bands.
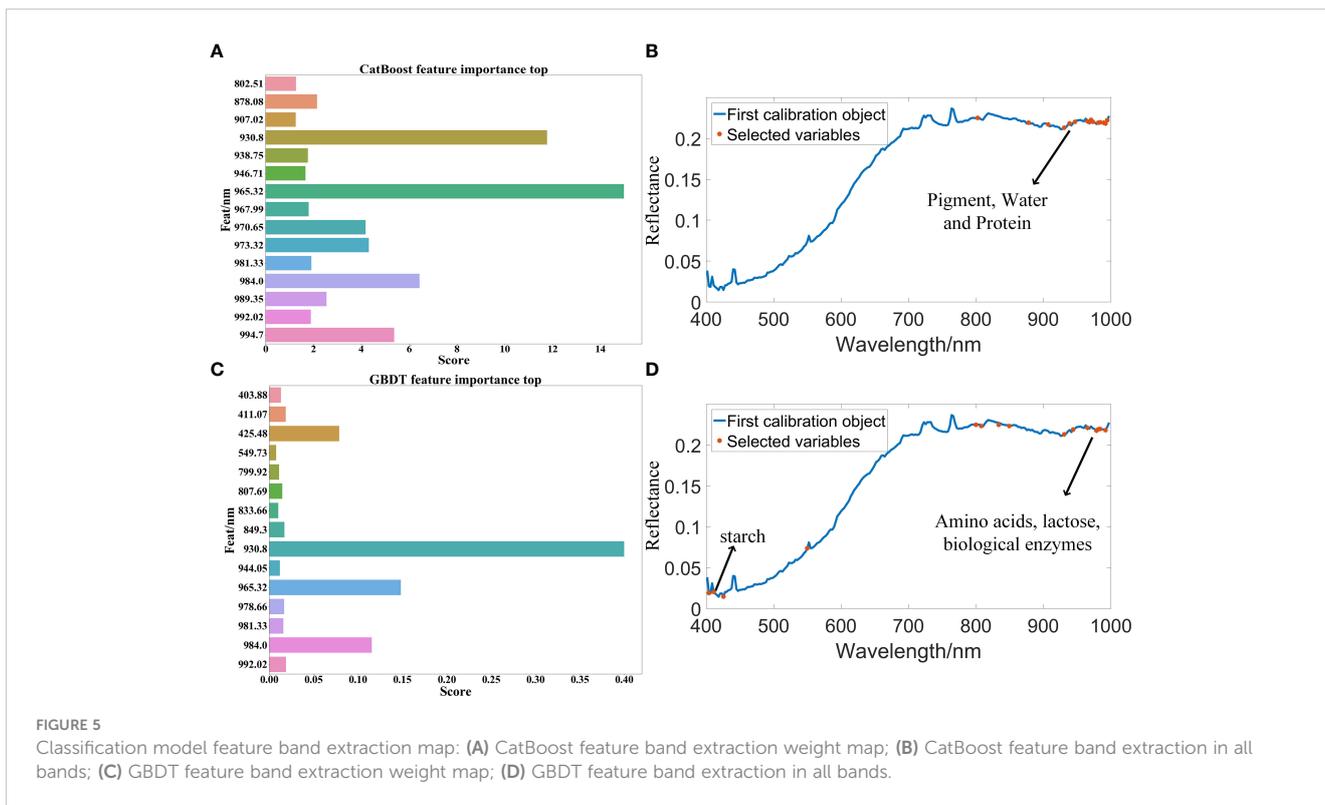
improved. After preprocessing, the R^2 is about 90%, and the RMSE value is also significantly reduced. The optimal model is extracted from the CatBoost feature band. In MSC-CatBoost-PLSR, the extracted 15 characteristic bands are 406.28 nm, 408.67 nm, 442.35 nm, 466.55 nm, 468.97 nm, 505.52 nm, 601.81 nm, 611.79 nm, 684.72 nm, 733.06 nm, 758.68 nm, 823.26 nm, 907.02 nm, 941.4 nm, and 962.66 nm, the weight map of CatBoost feature band extraction is as shown in Figure 7A. Figure 7B shows the full-band distribution of CatBoost feature band extraction, and the bands near 530 nm are related to chlorophyll content. In the optimal model MSC-GBDT-PLSR for GBDT feature band extraction, the

extracted 15 feature bands are 478.69 nm, 483.56 nm, 493.3 nm, 654.42 nm, 661.98 nm, 684.72 nm, 781.84 nm, 794.75 nm, 805.1 nm, 825.85 nm, 862.36 nm, 880.71 nm, 917.58 nm, 922.86 nm, and 975.99 nm, the GBDT feature band extraction weight map is as shown in Figure 7C, and the characteristic areas are mainly concentrated in the vicinity of 400 nm and 900 nm, indicating that the SOD activity of peanut seeds is related to related substances in this band, such as starch, protein and other substances. Figure 7D is the characteristic band distribution map of GBDT, the vicinity of 760 nm is related to the oxygen content (Yuan et al., 2022), and 470 nm-490 nm is related to the nitrogen content (Li et al., 2022).

FIGURE 6

MF-CatBoost-LightGBM confusion matrix: **(A)** Training set confusion matrix; **(B)** prediction set confusion matrix.
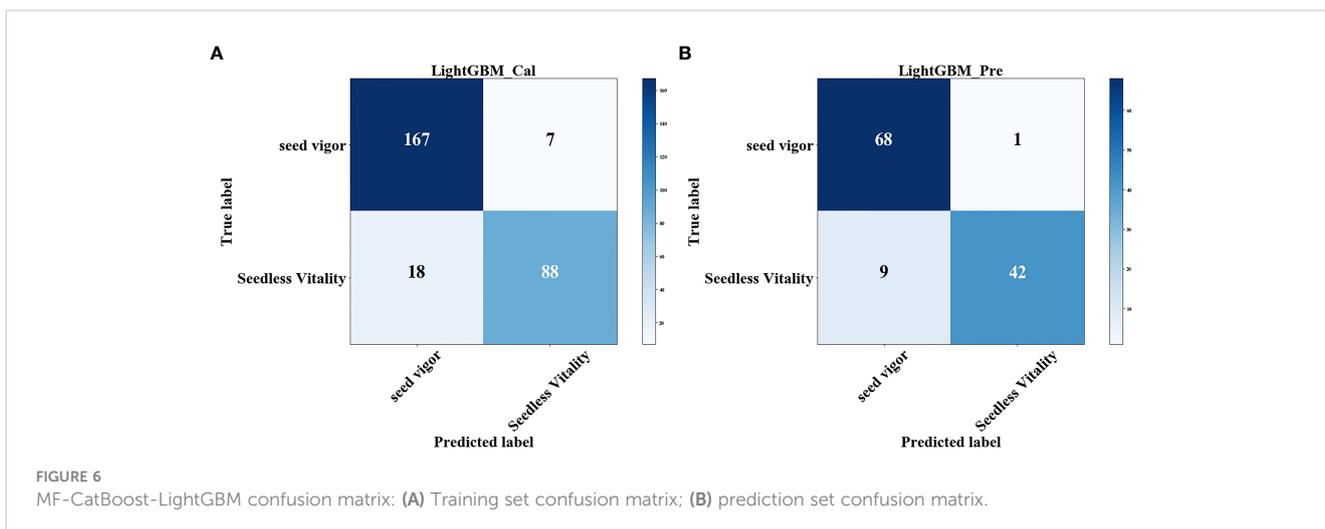
TABLE 3 SOD activity prediction result table.

|		PLSR						XGBoost					
		Full.		Sel1.		Sel2.		Full.		Sel1.		Sel2.	
		R^2	RMSE	R^2	RMSE	R^2	RMSE	R^2	RMSE	R^2	RMSE	R^2	RMSE
Raw	Cal	0.5942	0.6622	0.6433	0.5817	0.6358	0.5645	0.5927	0.5355	0.7107	0.5278	0.6491	0.4630
	Pre	0.5366	0.6397	0.6062	0.6260	0.6073	0.5862	0.5322	0.6189	0.5935	0.6147	0.5549	0.5899
SG	Cal	0.4875	0.6820	0.5991	0.5842	0.5839	0.6873	0.4716	0.6376	0.5270	0.6726	0.6343	0.5738
	Pre	0.4510	0.7112	0.5313	0.5849	0.5580	0.5679	0.4278	0.7138	0.4278	0.7138	0.4278	0.7138
MSC	Cal	0.9040	0.1309	0.9834	0.0541	0.9769	0.0676	0.9658	0.0736	0.9724	0.0666	0.9592	0.0804
	Pre	0.8701	0.1483	0.9787	0.0566	0.9659	0.0755	0.9542	0.0852	0.9563	0.0825	0.9247	0.1094
MF	Cal	0.6474	0.5272	0.6428	0.5528	0.6890	0.5701	0.7352	0.4853	0.7368	0.4838	0.7809	0.4414
	Pre	0.5302	0.6447	0.6358	0.5645	0.6881	0.5643	0.6959	0.5655	0.7009	0.5158	0.7363	0.4842

Cal, calibration; Pre, prediction; Raw, raw data; Full, full wavelengths; Sel1, selected wavelengths by CatBoost; Sel2, selected wavelengths by GBDT; R2, Correlation coefficient; RMSE, root mean square error; SG, Savitzky-Golay; MSC, multivariate scatter correction; MF, median filter.

Among them, the optimal model MSC-CatBoost-PLSR training set R^2 is 98.34%, RMSE value is 5.41%. The training set SOD activity prediction result is as shown in Figure 8A, the prediction set R^2 is 97.87%, RMSE value is 5.66%. The SOD activity prediction result of the prediction set is as shown in Figure 8B, which shows that the hyperspectral imaging system could accurately predict the SOD activity of peanut seeds. Figure 7 shows that the maximum error of the training set is lower than 0.3, and the maximum error of the prediction set is lower than 0.2. The results show that the hyperspectral technology combined with the MSC-CatBoost-PLSR model has a high prediction performance for the SOD activity value of peanut seeds.

3.6 Correlation analysis

The top five weights of the characteristic bands of peanut seed vigor classification are 425.48nm, 930.8nm, 965.32nm, 984.0nm, and

FIGURE 7
Regression model feature band extraction map: **(A)** CatBoost feature band extraction weight map; **(B)** CatBoost feature band extraction in all bands; **(C)** GBDT feature band extraction weight map; **(D)** GBDT feature band extraction in all bands.
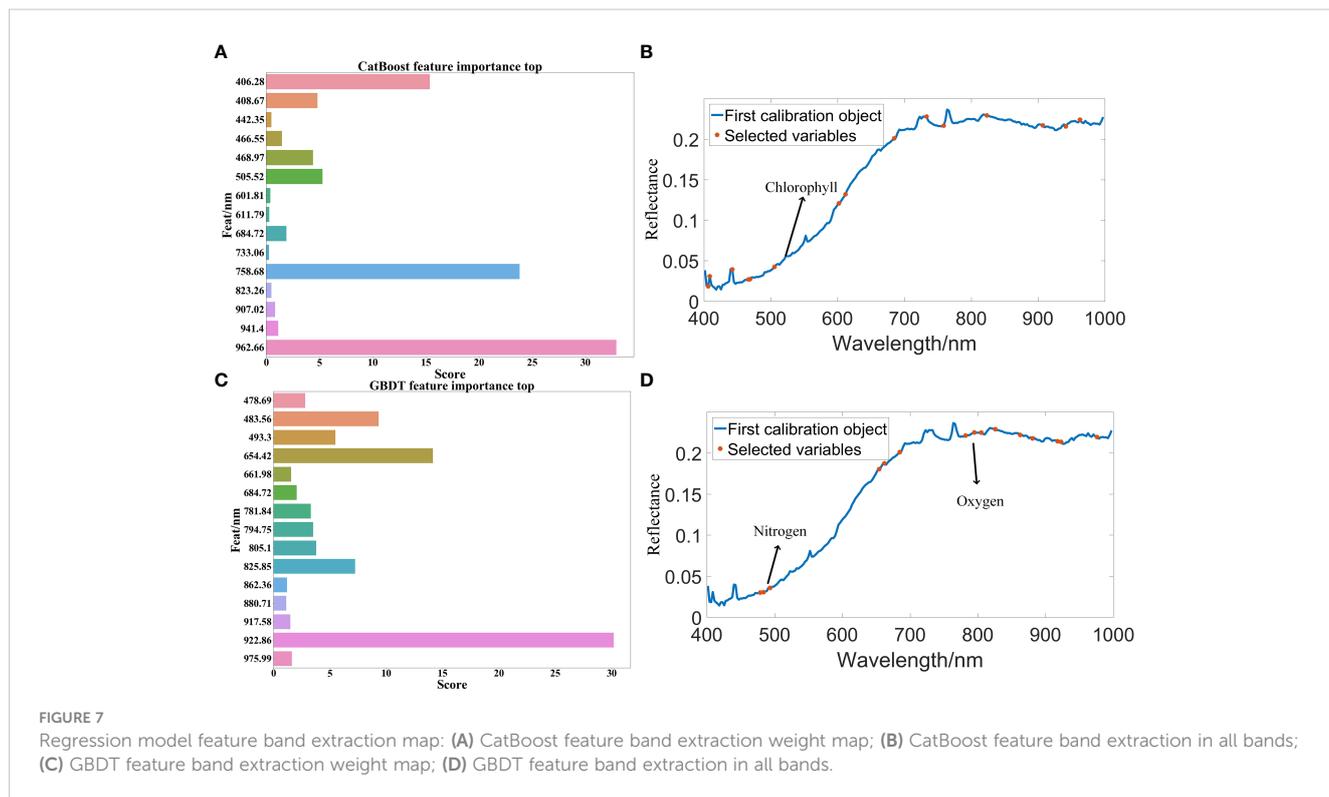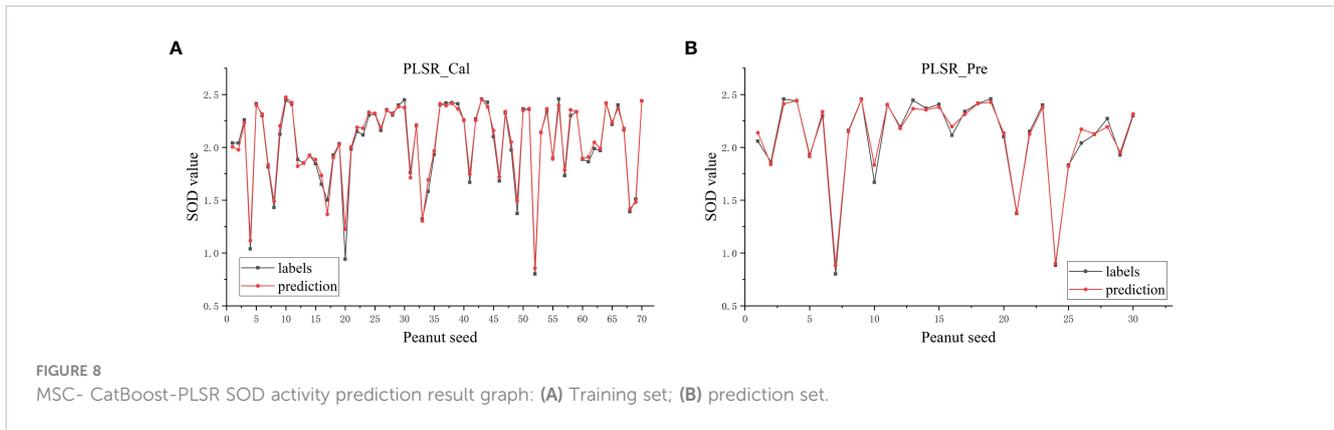
FIGURE 8
MSC- CatBoost-PLSR SOD activity prediction result graph: **(A)** Training set; **(B)** prediction set.

994.7nm, respectively. The top five predicted characteristic band weights of SOD value of peanut seeds are 406.28nm, 654.42nm, 758.68nm, 922.86nm and 962.66nm respectively. The average value of spectral reflectance and the average value of SOD value of peanut seeds in each group of A0~A4 groups were calculated. The correlation analysis of peanut seed vigor index(germination rate, germination potential, germination index, average germination time, vitality index, and simple vitality index), average spectral reflectance, average SOD, top five seed vigor classification characteristic band weights, and SOD value prediction characteristic band weight top five in groups A0~A4 were established. Figure 9 shows the heat map of the correlation analysis. It can be seen from the figure that the self-growth index of seed vigor has a high correlation, almost all reaching 1, and GE, GP, GI, VI, and SVI are positively correlated, while MGT is negatively correlated with other seed vigor indexes. This is due to the calculation formula of the seed vigor index, and the average germination time is a negative index of the seed vigor index. Generally, the higher the vigor of peanut seeds, the shorter the average germination time. The bands at 922.86nm, 930.8nm,

962.66nm, 965.32nm, 984nm and 994.7nm had strong correlation with peanut seed vigor index, and the correlation with GE was 0.35~0.46, and GP and GI was 0.37~0.48. The correlations with MGT were -0.41~-0.52, and with VI and SVI were 0.36~0.47. It is not difficult to see that most of these bands come from the characteristic bands of peanut seed vigor classification, which are consistent with the previous description of peanut seed vigor classification. All characteristic bands are also highly correlated with the SOD value of peanut seeds, up to 0.52, which is consistent with the regression analysis of peanut seeds SOD value described above. The most important thing is that the SOD value of peanut seeds has a strong correlation with the peanut vigor index, reaching 0.8, this also verifies the fact that SOD participates in the catalytic disproportionation reaction with reactive oxygen species and free radicals as substrates, and its activity level directly affects the seed vigor.

4 Conclusion

Seed vigor plays a crucial role in crop growth stage, and SOD enzyme activity is highly correlated with seed vigor indicators. In this paper, hyperspectral imaging technology was used to identify the vigor of peanut seeds and predict the activity of SOD inside the seeds. After the peanut seeds were germinated and cultivated by the method of rolling paper germination, the correlation analysis of characteristic bands, SOD activity and seed vigor indexes were made. After comparing and analyzing the results of various classification models, it was discovered that the MF-CatBoost-LightGBM model has the best effect and the best discrimination effect on peanut seed vigor. The accuracy rate of the training set reach 91.07% and the accuracy rate of the prediction set reach 90.83%. The MSC-CatBoost-PLSR model has the best effect when the SOD activity inside the seeds is analyzed by regression, in which the R^2 of the training set is 0.9834, the RMSE value is 0.0541, the R^2 of the prediction set is 0.9787, and the RMSE value is 0.0566. The correlation analysis of characteristic bands, vigor index and SOD activity shows that SOD activity of peanut seeds has strong correlation with seed vigor index, the correlation coefficient between the vigor index and SOD value reaches 0.8. In summary,
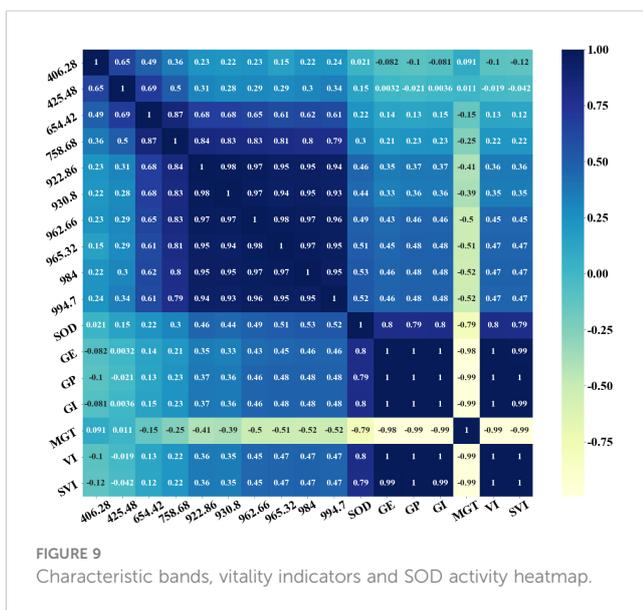
FIGURE 9
Characteristic bands, vitality indicators and SOD activity heatmap.

the use of hyperspectral imaging technology can accurately discriminate peanut seed vigor indexes, and can predict the SOD activity inside seeds. At the same time, the strong correlation between SOD activity and seed vigor indexes can provide new opportunities for future seed quality testing machinery ideas.

Data availability statement

The original contributions presented in the study are publicly available. This data can be found here: https://github.com/cjkka/cjkka/tree/main.

Author contributions

Conceptualization, ZZ and QiaW. Data curation, JZ and MZ. Formal analysis, YZ and QinW. Funding acquisition, YW and YC. Methodology, TL and JC. Project administration, MZ. Writing—original draft, JC and JL. Writing—review and editing, ZZ and WW. All authors contributed to the article and approved the submitted version.

Conflict of interest

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

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

References

Adams, A. (2021). Fast median filters using separable sorting networks. *ACM Trans. Graphics* 40 (4), 1-11. doi: 10.1145/3450626.3459773

Arya, S. S., Salve, A. R., and Chauhan, S. (2016). Peanuts as functional food: A review. *J. Food Sci. Technology-Mysore* 53 (1), 31–41. doi: 10.1007/s13197-015-2007-9

Bandeira, J. d. M., Marini, P., Gouvea de Borba, I. C., Noguez Martins, A. B., do Amarante, L., et al. (2014). Antioxidative metabolism: A tool to detect small differences in the vigor of soybean seeds. *Iheringia Serie Botanica* 69 (2), 285–292.

Bentejac, C., Csorgo, A., and Martinez-Munoz, G. (2021). A comparative analysis of gradient boosting algorithms. *Artif. Intell. Rev.* 54 (3), 1937–1967. doi: 10.1007/s10462-020-09896-5

Burnett, A. C., Anderson, J., Davidson, K. J., Ely, K. S., Lamour, J., Li, Q., et al. (2021). A best-practice guide to predicting plant traits from leaf-level hyperspectral data using partial least squares regression. *J. Exp. Bot.* 72 (18), 6175–6189. doi: 10.1093/jxb/erab295

Chen, T., Guestrin, C.Assoc Comp M (2016). "XGBoost: A scalable tree boosting system," in *Paper presented at the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD)*, San Francisco, CA (New York, NY, United States: Association for Computing Machinery).

Chen, X., and Ishwaran, H. (2012). Random forests for genomic data analysis. *Genomics* 99 (6), 323–329. doi: 10.1016/j.ygeno.2012.04.003

Choi, J.-Y., Kim, H.-C., and Moon, K.-D. (2021). Geographical origin discriminant analysis of chia seeds (Salvia hispanica l.) using hyperspectral imaging. *J. Food Composition Anal.* 101, 1-6. doi: 10.1016/j.jfca.2021.103916

Cortes, C., and Vapnik, V. (1995). Support-vector networks. *Mach. Learn.* 20 (3), 273–297. doi: 10.1023/a:1022627411411

Deal, K. (2005). Soft modeling with partial least squares - use partial least squares regression when hard assumptions get in the way of analysis. *Marketing Res.* 17 (3), 40–41.

ElMasry, G., Mandour, N., Al-Rejaie, S., Belin, E., and Rousseau, D. (2019). Recent applications of multispectral imaging in seed phenotyping and quality MonitoringAn overview. *Sensors* 19 (5), 1090. doi: 10.3390/s19051090

Fu, H., Jiang, D., Zhou, R., Yang, T., Chen, F., Li, H., et al. (2017). Predicting mildew contamination and shelf-life of sunflower seeds and soybeans by Fourier transform near-infrared spectroscopy and chemometric data analysis. *Food Analytical Methods* 10 (5), 1597–1608. doi: 10.1007/s12161-016-0726-9

Gomes, D. P., Barrozo, L. M., Souza, A. L., Sader, R., and Silva, G. C. (2009). Effect of the vigour and of the fungicide treatment in the germination and sanity tests of soybean seeds. *Bioscience J.* 25 (6), 59–65.

He, X. T., Feng, X. P., Sun, D. W., Liu, F., Bao, Y. D., and He, Y. (2019). Rapid and nondestructive measurement of rice seed vitality of different years using near-infrared hyperspectral imaging. *Molecules* 24 (12), 2227. doi: 10.3390/molecules24122227

Hou, Z., Zhao, L., Wang, Y., and Liao, X. (2019). Effects of high pressure on activities and properties of superoxide dismutase from chestnut rose. *Food Chem.* 294, 557–564. doi: 10.1016/j.foodchem.2019.05.080

Huang, K.-Y., and Chien, M.-C. (2017). A novel method of identifying paddy seed varieties. *Sensors* 17 (4), 809. doi: 10.3390/s17040809

Jasinski, S., Lecureuil, A., Durandet, M., Bernard-Moulin, P., and Guerche, P. (2016). Arabidopsis seed content QTL mapping using high-throughput phenotyping: The assets of near infrared spectroscopy. *Front. Plant Sci.* 7. doi: 10.3389/fpls.2016.01682

Kaya, M. D., Kulan, E. G., Ileri, O., and Avci, S. (2019). Prediction of viability and emergence capacity of safflower seed lots. *J. Anim. Plant Sciences-Japs* 29 (3), 714–720.

Ke, G., Meng, Q., Finley, T., Wang, T., Chen, W., Ma, W., et al. (2017). "LightGBM: A highly efficient gradient boosting decision tree," in *Paper presented at the 31st Annual Conference on Neural Information Processing Systems (NIPS)*, Long Beach, CA (Montreal, Canada: Neural Information Processing Systems).

Kono, Y. (2022). Reprint of: Generation of superoxide radical during autoxidation of hydroxylamine and an assay for superoxide dismutase. *Arch. Biochem. biophysics* 726, 109247–109247. doi: 10.1016/j.abb.2022.109247

Kotwaliwale, N., Singh, K., Chakrabarty, S. K., Joshi, M. A., Kalne, A., Tiwari, M., et al. (2019). Machine vision for characterisation of some phenomic features of plant parts in distinguishing varieties-a review. *Int. J. Bio-Inspired Comput.* 14 (4), 201–212. doi: 10.1504/ijbic.2019.103960

Kou, X., Zhao, Y., Xu, L., Kang, Z., Wang, Y., Zou, Z., et al. (2022). Controlled fabrication of core-shell gamma-Fe2O3@C-Reduced graphene oxide composites with tunable interfacial structure for highly efficient microwave absorption. *J. Colloid Interface Sci.* 615, 685–696. doi: 10.1016/j.jcis.2022.02.023

Kucha, C. T., Liu, L., Ngadi, M., and Claude, G. (2021). Hyperspectral imaging and chemometrics as a non-invasive tool to discriminate and analyze iodine value of pork fat. *Food Control* 127, 127. doi: 10.1016/j.foodcont.2021.108145

Laborde, A., Puig-Castellvi, F., Bouveresse, D. J.-R., Eveleigh, L., Cordella, C., and Jaillais, B. (2021). Detection of chocolate powder adulteration with peanut using near-infrared hyperspectral imaging and multivariate curve resolution. *Food Control* 119, 107454. doi: 10.1016/j.foodcont.2020.107454

Li, M., Zhu, X., Li, W., Tang, X., Yu, X., and Jiang, Y. (2022). Retrieval of nitrogen content in apple canopy based on unmanned aerial vehicle hyperspectral images using a modified correlation coefficient method. *Sustainability* 14 (4), 1039110. doi: 10.3390/su14041992

Liang, K., Liu, Q. X., Xu, J. H., Wang, Y. Q., Okinda, C. S., and Shen, M. X. (2018). Determination and visualization of different levels of deoxynivalenol in bulk wheat kernels by hyperspectral imaging. *J. Appl. Spectrosc.* 85 (5), 953–961. doi: 10.1007/s10812-018-0745-y

Liu, X., Gao, C., and Xing, D. (2009). A non-invasive and rapid seed vigor biosensor based on quantitative measurement of superoxide generated by aleurone cell in intact seeds. *Biosensors Bioelectronics* 24 (6), 1537–1542. doi: 10.1016/j.bios.2008.06.040

Liu, M.-B., Li, X.-L., Liu, Y., Huang, J.-F., and Tang, Y.-L. (2014). Detection of crude protein, crude starch, and amylose for rice by hyperspectral reflectance. *Spectrosc. Lett.* 47 (2), 101–106. doi: 10.1080/00387010.2013.779283

Liu, W., Liu, J., Jiang, J., and Li, Y. (2021). Comparison of partial least squares-discriminant analysis, support vector machines and deep neural networks for spectrometric classification of seed vigour in a broad range of tree species. *J. near Infrared Spectrosc.* 29 (1), 33–41. doi: 10.1177/0967033520963759

Liu, D., Sun, D.-W., and Zeng, X.-A. (2014). Recent advances in wavelength selection techniques for hyperspectral image processing in the food industry. *Food Bioprocess Technol.* 7 (2), 307–323. doi: 10.1007/s11947-013-1193-6

Matłok, N., Piechowiak, T., Królikowski, K., and Balawejder, M. (2022). Mechanism of reduction of drought-induced oxidative stress in maize plants by fertilizer seed coating. *Agriculture* 12 (5), 662. doi: 10.3390/agriculture12050662

Mattioni, N. M., Mertz, L. M., Piccinin Barbieri, A. P., Haesbaert, F. M., Giordani, W., and Lopes, S. J. (2015). Individual electrical conductivity test for the assessment of soybean seed germination. *Semina-Ciencias Agrar.* 36 (1), 31–38. doi: 10.5433/1679-0359.2015v36n1p31

Pan, Q.-m., Zhang, J.-s., Zhang, J.-p., Meng, P., Wang, G.-b., Yang, H.-g., et al. (2019). Analysis of correlation and differences between leaf moisture and hyperspectral reflectance among different walnut varieties. *For. Res.* 32 (6), 1–6. doi: 10.13275/j.cnki.lykxyj.2019.06.001

Pang, L., Men, S., Yan, L., and Xiao, J. (2020). Rapid vitality estimation and prediction of corn seeds based on spectra and images using deep learning and hyperspectral imaging techniques. *IEEE Access* 8, 123026–123036. doi: 10.1109/access.2020.3006495

Pjw, A., and Sk, B. (2016). Classification of maize kernels using NIR hyperspectral imaging. *Food Chem.* 209, 131–138. doi: 10.1016/j.foodchem.2016.04.044

Raj, R., Walker, J. P., Vinod, V., Pingale, R., Naik, B., and Jagarlapudi, A. (2021). Leaf water content estimation using top-of-canopy airborne hyperspectral data. *Int. J. Appl. Earth Observation Geoinformation* 102, 102393. doi: 10.1016/j.jag.2021.102393

Sen, F. (2016). Combination of visible and mid-infrared spectra for the prediction of chemical parameters of wines. *Talanta: Int. J. Pure Appl. Analytical Chem.* 161, 130-137. doi: 10.1016/j.talanta.2016.08.057

Tan, S., Liu, J., Lu, H., Lan, M., Yu, J., Liao, G., et al. (2022). Machine learning approaches for rice seedling growth stages detection. *Front. Plant Sci.* 13. doi: 10.3389/fpls.2022.914771

Tu, K., Cheng, Y., Pan, T., Wang, J., and Sun, Q. (2022). Effects of seed priming on vitality and preservation of pepper seeds. *Agriculture* 12 (5), 603. doi: 10.3390/agriculture12050603

Wang, Q., and Chen, H. (2020). Optimization of parallel random forest algorithm based on distance weight. *J. Intelligent Fuzzy Syst.* 39 (2), 1951–1963. doi: 10.3233/jifs-179965

Wang, Z., Huang, W., Tian, X., Long, Y., Li, L., and Fan, S. (2022). Rapid and non-destructive classification of new and aged maize seeds using hyperspectral image and chemometric methods. *Front. Plant Sci.* 13. doi: 10.3389/fpls.2022.849495

Wang, Z. J., Liu, S. J., Yang, Q. Q., and Peng, H. L. (2021). Classification of typical tree species in a karst area of guizhou province based on principal component analysis and support vector machine. *Spectrosc. Lett.* 54 (4), 305–315. doi: 10.1080/00387010.2021.1931790

Wang, L., Liu, J., Zhang, J., Wang, J., and Fan, X. (2022). Corn seed defect detection based on watershed algorithm and two-pathway convolutional neural networks. *Front. Plant Sci.* 13. doi: 10.3389/fpls.2022.730190

Wang, Y., and Wang, T. (2020). Application of improved LightGBM model in blood glucose prediction. *Appl. Sciences-Basel* 10 (9), 3227. doi: 10.3390/app10093227

Wang, F., and Wang, C. (2022). Improved model for starch prediction in potato by the fusion of near-infrared spectral and textural data. *Foods* 11 (19). doi: 10.3390/foods11193133

Wang, H., Wang, K., Wu, J.-z., and Han, P. (2021). Progress in research on rapid and non-destructive detection of seed quality based on spectroscopy and imaging technology. *Spectrosc. Spectral Anal.* 41 (1), 52–59. doi: 10.3964/j.issn.1000-0593(2021)01-0052-08

Wang, G.-l., Yu, K.-q., Cheng, K., Liu, X., Wang, W.-j., Li, H., et al. (2021). Hyperspectral technique coupled with chemometrics methods for predicting alkali spreading value of millet flour. *Spectrosc. Spectral Anal.* 41 (10), 3189–3193. doi: 10.3964/j.issn.1000-0593(2021)10-3189-05

Xiang, Y., Li, H., Zhang, T., Wang, J., and Sun, Q. (2020). Study on early detection of seed vigor of maize and wheat seeds by conductivity method. *J. China Agric. Univ.* 25 (6), 12–19. [电导率法早期检测玉米和小麦种子>活力].

Xu, M., Sun, J., Zhou, X., Tang, N., Shen, J., and Wu, X. (2021). Research on nondestructive identification of grape varieties based on EEMD-DWT and hyperspectral image. *J. Food Sci.* 86 (5), 2011–2023. doi: 10.1111/1750-3841.15715

Xu, H., Wei, Y., Lian, L., Zhu, Y., Xie, H., Wang, Z., et al. (2013). Comparative analysis of artificial aging and natural aging with rice seeds. *Mol. Plant Breed.* 11 (5), 552–556. [水稻种子人工老化与自然老化的分析比较].

Yang, J., and Gao, H. (2020). Cultural emperor penguin optimizer and its application for face recognition. *Math. Problems Eng.* 2020, 1-6. doi: 10.1155/2020/9579538

Yang, J., Sun, L., Xing, W., Feng, G., Bai, H., and Wang, J. (2021). Hyperspectral prediction of sugarbeet seed germination based on gauss kernel SVM. *Spectrochimica Acta Part a-Molecular Biomolecular Spectrosc.* 253, 119585. doi: 10.1016/j.saa.2021.119585

Yuan, Z., Ye, Y., Wei, L., Yang, X., and Huang, C. (2022). Study on the optimization of hyperspectral characteristic bands combined with monitoring and visualization of pepper leaf SPAD value. *Sensors* 22 (1), 183. doi: 10.3390/s22010183

Zhang, T., Fan, S., Xiang, Y., Zhang, S., Wang, J., and Sun, Q. (2020). Non-destructive analysis of germination percentage, germination energy and simple vigour index on wheat seeds during storage by Vis/NIR and SWIR hyperspectral imaging. *Spectrochimica Acta Part a-Molecular Biomolecular Spectrosc.* 239, 118488. doi: 10.1016/j.saa.2020.118488

Zhang, S., Zhu, X., Anduv, B., Jin, X., and Du, Z. (2021). Fault detection and diagnosis for the screw chillers using multi-region XGBoost model. *Sci. Technol. Built Environ.* 27 (5), 608–623. doi: 10.1080/23744731.2021.1877966

Zhao, L. (2017). "Measurement of superoxide dismutase-like activity in peel and pulp of apple from anshan acres," in *Paper presented at the 6th International Conference on Biotechnology and Bioengineering (ICBB)*, Offenburg, GERMANY (Barcelona, Spain: AIP publishing).

Zhao, J., He, Y., Huang, S., and Wang, Z. (2021). Advances in the identification of quantitative trait loci and genes involved in seed vigor in rice. *Front. Plant Sci.* 12. doi: 10.3389/fpls.2021.659307

Zhou, S., Wang, S., Wu, Q., Azim, R., and Li, W. (2020). Predicting potential miRNA-disease associations by combining gradient boosting decision tree with logistic regression. *Comput. Biol. Chem.* 85, 107200. doi: 10.1016/j.compbiolchem.2020.107200

Zou, Z., Wu, Q., Wang, J., Xu, L., Zhou, M., Lu, Z., et al. (2022). Research on non-destructive testing of hotpot oil quality by fluorescence hyperspectral technology combined with machine learning. *Spectrochimica Acta Part A Mol. biomolecular Spectrosc.* 284, 121785–121785. doi: 10.1016/j.saa.2022.121785