

RESEARCH

Open Access



# Soybean seed pest damage detection method based on spatial frequency domain imaging combined with RL-SVM

Xuanyu Chen<sup>1</sup>, Wei He<sup>2</sup>, Zhihao Ye<sup>3</sup>, Junyi Gai<sup>3</sup>, Wei Lu<sup>1\*</sup> and Guangnan Xing<sup>3\*</sup>

## Abstract

Soybean seeds are susceptible to damage from the *Riptortus pedestris*, which is a significant factor affecting the quality of soybean seeds. Currently, manual screening methods for soybean seeds are limited to visual inspection, making it difficult to identify seeds that are phenotypically defect-free but have been punctured by stink bugs on the sub-surface. To facilitate the convenient and efficient identification of healthy soybean seeds, this paper proposes a soybean seed pest detection method based on spatial frequency domain imaging combined with RL-SVM. Firstly, soybean optical data is obtained using single integration sphere technique, and the vigor index of soybean seeds is obtained through germination experiments. Then, based on the above two data items using feature extraction algorithms (the successive projections algorithm and the competitive adaptive reweighted sampling algorithm), the characteristic wavelengths of soybeans are identified. Subsequently, the spatial frequency domain imaging technique is used to obtain the sub-surface images of soybean seeds in a forward manner, and the optical coefficients such as the reduced scattering coefficient  $\mu'_s$  and absorption coefficient  $\mu_a$  of soybean seeds are inverted. Finally, RL-MLR, RL-GRNN, and RL-SVM prediction models are established based on the ratio of the area of insect-damaged sub-surface to the entire seed, soybean varieties, and  $\mu_a$  at three wavelengths (502 nm, 813 nm, and 712 nm) for predicting and identifying soybean the stinging and sucking pest damage levels of soybean seeds. The experimental results show that the spatial frequency domain imaging technique yields small errors in the optical coefficients of soybean seeds, with errors of less than 15% for  $\mu_a$  and less than 10% for  $\mu'_s$ . After parameter adjustment through reinforcement learning, the Macro-Recall metrics of each model have improved by 10%-15%, and the RL-SVM model achieves a high Macro-Recall value of 0.9635 for classifying the pest damage levels of soybean seeds.

**Keywords** Spatial frequency domain imaging, Soybean seed testing, Reinforcement learning, Neural network

\*Correspondence:

Wei Lu  
njaurobot@njau.edu.cn  
Guangnan Xing  
xinggn@njau.edu.cn

<sup>1</sup> College of Artificial Intelligence, Nanjing Agricultural University, Nanjing 210031, China

<sup>2</sup> College of Engineering, Nanjing Agricultural University, Nanjing 210031, China

<sup>3</sup> Soybean Research Institute, MARA National Center for Soybean Improvement, MARA Key Laboratory of Biology and Genetic

Improvement of Soybean, National Key Laboratory for Crop Genetics

& Germplasm Enhancement and Utilization, Jiangsu Collaborative Innovation Center for Modern Crop Production, College of Agriculture, Nanjing Agricultural University, Nanjing 210095, China



© The Author(s) 2024, corrected publication (2024). **Open Access** This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

## Introduction

Soybean has high nutritional value, with high content of protein, unsaturated fatty acids, isoflavones, and other nutrients. It is also an important oilseed crop [1]. However, soybean seeds are susceptible to damage by *Riptortus pedestris* (Order Hemiptera, family Pentatomidae) while growing. These stink bugs produce saliva containing digestive enzymes while feeding, extracting nutrients and moisture from soybeans [2], leading to a decrease in soybean seed quality [3]. Healthy soybean seeds have higher germination rates and can grow into more resistant seedlings, ultimately improving crop yield [4]. Therefore, seed quality detection is crucial [5]. The timing of pest detection directly affects the final crop yield. This is because the phenotypes of crops at the early stage of infestation are not significantly different from healthy soybean seeds. Manual selection is time-consuming, labor-intensive, and less stable, relying on the experience of workers. Machine recognition methods are limited by the small size and hidden nature of soybean pest characteristics [6], resulting in many infected seeds going undetected until they are planted or used. Moreover, *R. pedestris*, as a vector for yeast spot disease [7], often leads to complete sterility of seeds in entire field [8]. Therefore, it is crucial to screen for soybean seeds minor pest damage to improve seed quality and ensure emergence rate.

At present, large seed companies both domestically and internationally have established seed processing procedures. These procedures primarily rely on the physical differences in seed size, density, and other characteristics. Impurities are removed using equipment such as air screen cleaners and liquid separators. Seeds that are not plump, have abnormal colors, severe deterioration, obvious signs of pest damage, or surface contamination by pests and diseases are eliminated [9]. This process mainly targets physically damaged or visually abnormal seeds. It cannot selectively identify deteriorated seeds that appear normal but have substandard germination rates and vigor. It also fails to detect internal defects, damage, contaminants, and pathogen invasion within the endosperm and embryo of seeds. Some researchers have explored the use of scattering spectra obtained through near-infrared [10], hyperspectral [11], fluorescence [12], as well as X-ray [13] and terahertz [14] techniques to establish seed quality models based on optical and density information. These methods aim to achieve non-destructive detection of seed damage. However, due to factors such as long detection cycles and high costs, these methods are still limited to laboratory-scale applications. Moreover, they cannot accurately determine pest damage hidden beneath the soybean phenotype. In this regard, spatial frequency domain imaging technology has emerged as a promising approach in optical

detection due to its capability for depth discrimination in imaging.

In recent years, structured light reflection imaging technology has experienced rapid development. Based on the frequency-dependent attenuation of light in biological tissue [15], this technology projects light with different frequencies onto the surface of an object. By analyzing the modulated direct and alternating current images containing depth information, and further processing the images, the contrast of the information is enhanced. Structured light reflection imaging technology was initially applied in the field of biomedical sciences. In 2005, Cuccia proposed this technique [16], and it has since been widely used for measuring optical properties of biological tissues, such as burn severity assessment [17], prediction of foot ulcers in diabetic patients [18], and oxygenation measurement [19, 20]. In 2007, Anderson et al. [21] first discovered the potential of spatial frequency domain imaging technology in the agricultural field. They found that the reduced scattering coefficient of damaged apples was higher than that of normal apples. Subsequently, Lu et al. [15] proposed structured light imaging technology based on spatial frequency domain imaging. They applied this technique to detect bruises on the subsurface of fresh apples and demodulated the transmitted images of the apples. Sun et al. [6] used spatial frequency domain imaging to obtain alternating current images and ratio images of peaches. They combined the watershed algorithm with partial least squares discriminant analysis and convolutional neural network (CNN) methods to classify infected peaches. The detection rates ranged from 65 to 87% with the former method and reached 98.6% with the latter. Luo et al. [22] studied the calibration, correction, and application of spatial frequency domain imaging in surface damage detection of pears. They measured the reduced scattering coefficient and absorption coefficient of pears and used a linear discriminant analysis model to distinguish different surface-damaged pears, achieving results superior to traditional methods' results.

Above studies have demonstrated the tremendous potential of spatial frequency domain imaging technology in the agricultural field. However, current agricultural experiments are still limited to the measurement of fruits and vegetables, which have similar physicochemical properties and provide certain reference points for each other. There is currently a lack of research on spatial frequency domain imaging for seed crops. Due to long-term moisture loss and wrinkling of the soybean seed coat during storage and friction between seed coats during transportation, soybean seeds often exhibit dark abrasions on the surface [23]. These abrasions are very similar to non-apparent pest damage and are difficult

to distinguish with the naked eye [24], causing manual selection to heavily rely on experience and not guarantee accuracy. However, surface abrasions do not affect the subsurface, and healthy seeds have noticeable differences from pest-damaged seeds in the subsurface. By directly observing and comparing the subsurface, they can be easily and accurately differentiated.

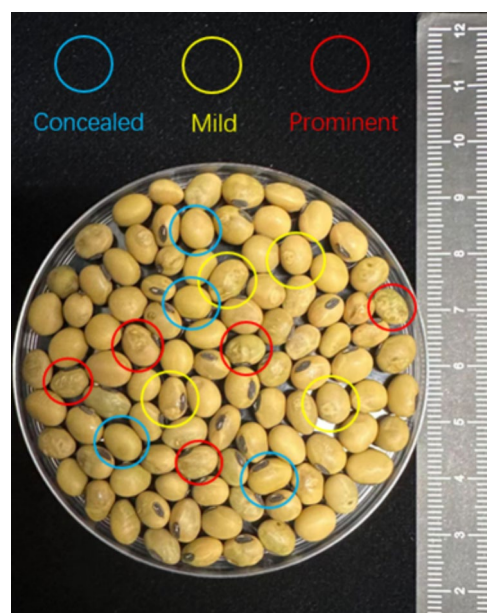
Currently, shallow neural networks have become a popular direction in deep learning due to their small data set [25]. Compared to traditional deep learning models that require multiple hidden layers and a large amount of training data, shallow neural networks have a simpler model structure and perform better on small datasets [26]. Given the significant differences in information and the relative difficulty in acquiring information in agricultural product detection, shallow neural network models are more suitable for agricultural applications [27]. During the training process of neural networks, there are many invariant parameters, such as the sigma of Generalized Regression Neural Network (GRNN), the kernel and C of Support Vector Machine (SVM), and the alpha of Mixed Logistic Regression (MLR). The selection of these parameters directly affects the performance of the final model [28]. These parameters are determined by both empirical knowledge, selecting appropriate ranges for setting, and reliable methods to accurately determine their final values [29]. The fundamental principle of reinforcement learning is that an intelligent agent learns autonomously through continuous interaction with the environment, guided by a reward mechanism, to achieve optimal behavioral choices [30]. To optimize the model, applying reinforcement learning techniques and treating the parameter selection of the neural network model as a behavior choice in reinforcement learning is an emerging approach to improving neural network performance [31].

Therefore, this study focuses on soybean seeds and aims to validate the feasibility of the algorithm by comparing spatial frequency domain imaging data with single integrating sphere measurement data. Subsequently, structured light imaging methods are employed for pest damage detection imaging of soybeans. Finally, in combination with reinforcement learning, SVM, MLR, and GRNN neural network models are optimized. Based on the size of the subsurface erosion area of soybean seeds, a predictive classification model for soybean seeds pest damage levels is constructed, exploring the feasibility of pest damage detection in soybean seeds.

## Materials and methods

### Experimental materials and data acquisition

In this experiment, seeds of three soybean varieties, namely SuXian26, NanNong1138-2, and QY1808 (referred to as SX26, NN1138-2, and QY1808

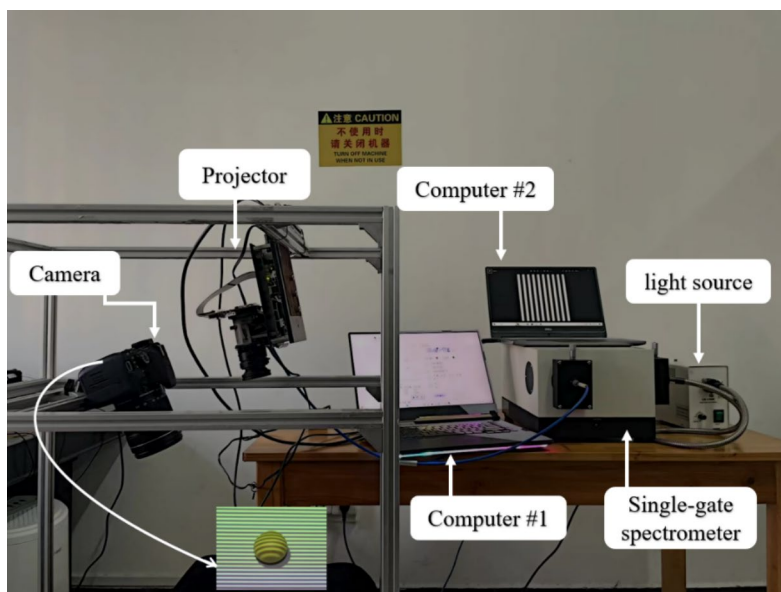


**Fig. 1** Soybean Seeds Classification. (The blue, yellow, and red circles in the diagram correspond to the concealed type, mild type, and severe type of seeds, respectively)

respectively), were selected after being infested with *R. pedestris* at the R3 stage at the Baima base of Nanjing Agricultural University (longitude: 119.180691, latitude: 31.613777). Some soybean seeds may not show obvious signs of damage on the surface after being bitten by *R. pedestris* (Fig. 1). They are classified into three categories based on surface identification: concealed type, mild type, and prominent type, corresponding to the blue, yellow, and red circles in the image, respectively. For each variety, 100 healthy seeds were selected, which were similar in size, color, and had no obvious surface defects. In total, there were 300 samples, with 210 seeds used as the training set and 90 seeds as the testing set. Firstly, the soybean seeds were washed with a 1000 mg/L concentration of sodium hypochlorite solution to ensure that the surface was free from bacteria. Then, the soybean seeds were rinsed with distilled water to remove the sodium hypochlorite solution, followed by air drying in a drying room.

Finally, soybean seeds of the concealed type were selected for subsequent experiments, as this type of seeds are difficult to screen manually by naked eyes and the screening accuracy is relatively low. Seeds that did not meet the requirements were discarded, and new soybean seeds were selected to repeat the above steps until the required number of seeds was obtained.

The SFDI (Spatial Frequency Domain Imaging) system used in this experiment is based on a halogen light source [32]. The physical configuration of the system is shown in

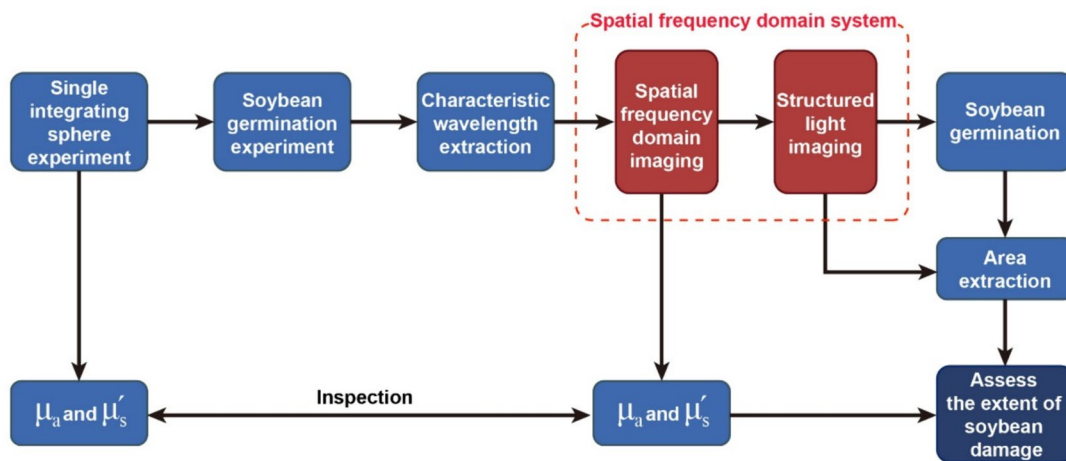


**Fig. 2** Physical image of spatial frequency domain imaging system

Fig. 2. The system consists of two main components: the projection unit and the imaging unit. The projection unit includes a dark box, a GX-100 halogen light source (Beijing Padiwei Instrument Co., Ltd., China) with adjustable light intensity, a WDG single-grating spectrometer (Beijing Beiguang Century Instrument Co., Ltd., China), a lifting platform, and a DLI6500 projector (Texas Instruments, USA). The halogen light source has good color rendering properties, and its spectrum is close to the continuous spectrum of sunlight. The maximum power of the instrument is 200W, but for the sake of experimental stability, a power of 150W was used. The dimensions of the lifting platform are 300 × 300 mm, and its height

range is 100–460 mm. Spectrometer is used to screen the wavelength of light sources and is directly connected and controlled by the computer#1. The images required by the projector are provided by a connected computer#2, and the desired sine images are generated using MATLAB R2020a software. To reduce the influence of surface reflection from soybean seeds on the captured images, a polarizing film is installed in front of the projector.

The imaging unit primarily consists of a lifting platform, computers, and a Canon EOS 700D camera (Canon, Japan). The lifting platform is used to control the distance between the soybean seed and the projector and camera, ensuring uniform projection onto the



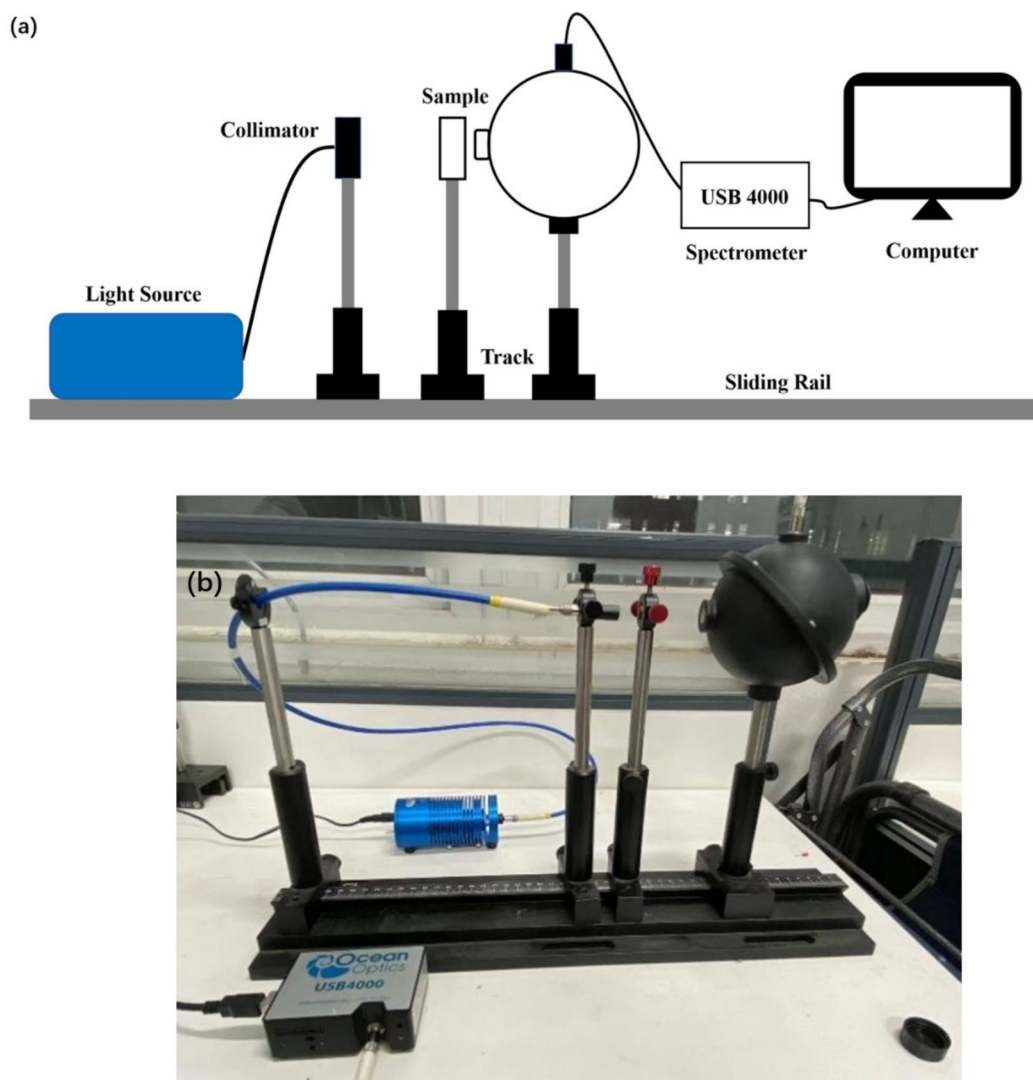
**Fig. 3** Process of soybean early detection method

soybean seed and clear image capture. The Canon camera is controlled by a smartphone via software to allow for focusing and capturing photos. The captured images are then transferred to the computer for demodulation processing.

**Process of soybean seeds pest detection**

This study utilizes the SFDI (Spatial Frequency Domain Imaging) method to generate depth images of soybean seeds based on three-phase demodulation. The detection method workflow is illustrated in Fig. 3 and consists of five steps: (1) using a single integrating sphere to measure the transmittance and reflectance of different bands of soybean and calculate the absorption

coefficient  $\mu_a$  and the reduced scattering coefficient  $\mu'_s$ ; (2) conducting soybean germination experiment; (3) extracting feature wavelengths of soybean based on successive projection algorithm and competitive adaptive reweighting algorithm; (4) obtaining the structured light reflection image of soybean in the forward direction by spatial frequency domain imaging system; (5) performing inverse reconstruction of  $\mu_a$  and  $\mu'_s$  by spatial frequency imaging system, and the feasibility of the experimental algorithm is verified; (6) classifying soybean of damage based on gnawing area; and (7) comparing and analyzing the models and evaluating the extent of soybean damage.



**Fig. 4** single integrating sphere system. **a** Schematic diagram. **b** Physical diagram

### Measurement of soybean optical properties by single integrating sphere

This study employed an iterative indirect measurement method to calibrate soybean seeds of three different varieties. The PTFE integrating sphere system (Jinan Chuangpu Instrument Co., Ltd., China) was used to measure the total reflectance and total transmittance of each soybean using SpectraSuite software. As shown in Fig. 4.

When measuring the transmission spectrum, the experimental setup is first placed in a dark environment without opening any ports, and the dark field spectrum is measured. Then, the halogen lamp light source is turned on and preheated for 10 min. The input port is opened to allow light to enter the interior of the single integrating sphere and obtain the transmission reference spectrum. The test sample is then placed tightly against the input port of the integrating sphere, and the distance between the light source and the sample is adjusted to minimize and maximize the brightness of the light spot on the surface of the test sample. Some of the light passes through the sample and enters the detection end of the integrating sphere, where the detector measures the transmitted light intensity.

When measuring the reflection spectrum, both the input port and the output port of the instrument are opened simultaneously, allowing the light to pass through and measure the dark field spectrum. Then, a standard reflection board (HSIA-CT-SRT-99-050, Jiangsu Dualix Spectral Imaging Technology Co., Ltd., China) is placed at the output port to obtain the reflection reference spectrum. The reflection board is then replaced with the test sample, and the distance between the light source and the sample is adjusted to minimize and maximize the brightness of the light spot on the sample surface. The sample reflects some of the light, which enters the detection end of the integrating sphere again, and the detector measures the reflected light intensity.

During the experimental process, it is important to ensure that the distance between the light source and the sample remains constant, and the incident light spot on the sample surface maintains a diameter of 1.9 mm [33]. Using the scale at the right end of the optical rail as a reference, for measuring the transmission spectrum, the light source is positioned at 10 cm, the integrating sphere is positioned at 20 cm, and the test sample is positioned at 14 cm. When measuring the reflection spectrum, only the position of the test sample is changed to 25 cm. The soybean spectral data are preprocessed using SpectraSuite software, including noise removal and data smoothing, and ultimately calculated to obtain the total reflectance and total transmittance. The integration time

is set to 1 s, and the smoothing and averaging are set to 3 times.

To ensure the reliability and consistency of the experimental data, the experiments are repeated multiple times and the results are compared. If the difference exceeds the threshold set, the experiments are continued until the difference in data is within the threshold for at least three or more repetitions [34].

For most food and agricultural products, the anisotropy factor has a small impact on the calculation of the  $\mu_a$  and  $\mu'_s$  values. The range of  $g$  values is typically [0.7, 0.9] [35]. In the subsequent IAD iterative computations in this study, an anisotropy factor of 0.7 is used. The reflectance and transmittance data were then used in the iterative process based on the Inverse Adding Doubling (IAD) algorithm [36] to calculate the values of  $\mu_a$  and  $\mu'_s$  [37]. The accuracy of the measurements obtained using the single integrating sphere method has been previously demonstrated [38], and thus, the results obtained from the single integrating sphere measurements were used as the reference for comparison with the subsequent spatial frequency domain imaging experiments. This was done to validate the feasibility of the spatial frequency domain algorithm.

### Germination experiment of soybean seed samples

After the completion of single integrating sphere measurements, the soybean seeds were immediately subjected to germination experiments. A total of 216 seeds (24 seeds for each variety) from three different varieties, namely SX26, NN1138-2, and QY1808, were taken from the past year (2022, since the soybean seeds were harvested in the second half of 2023, and the experiments were conducted in the first half of the year). The soybean seeds were first soaked in room temperature water (23 °C) for 4–5 h. Then, they were classified and placed in different culture dishes. The culture dishes had an upper cover with a diameter of 81 mm, a lower cover with a diameter of 75 mm, a height of 17.6 mm, a wall thickness of 2.3 mm, and weighed 59 g. Labels were attached to the culture dishes for identification. The germination process was conducted under a light–dark cycle of 16 h of light and 8 h of darkness. The number of germinated soybeans was recorded daily, and timely watering was done to maintain moisture levels. Moldy seeds were removed, and if more than half of the seeds in a culture dish were moldy, the germination filter paper was replaced. Finally, based on the data collected for five consecutive days, the seed germination rate, germination index (GI), and vigor index (VI) were calculated.

The germination rate refers to the percentage of germinated soybean seeds out of all the experimental seeds under specific environmental conditions and within a certain period. A higher germination rate indicates higher seed vigor, while a lower germination rate indicates lower seed vigor. It can be obtained using the following formula:

$$G = \frac{Num_{spr}}{Num_{all}} \times 100\% \quad (1)$$

where  $G$  represents the germination rate,  $Num_{spr}$  represents the number of germinated seeds, and  $Num_{all}$  represents the total number of seeds.

The GI is a more detailed indicator than the germination rate, as it represents the sum of germinated seeds on a daily basis. It provides a finer analysis of the germination process by considering each day individually, thus better reflecting the seed vigor. The calculation formula is as follows:

$$GI = \sum_{i=1}^n \frac{Num_{spr}}{Day_i} \quad (2)$$

where  $Num_{spr}$  represents the number of germinated seeds and  $Day_i$  represents the  $i$ -th day.

VI is a comprehensive index of germination rate and growth rate of seeds. More importantly, this index can reflect each individual condition of seeds. Seed vigor is influenced by three main factors: genetic factors, environmental conditions during seed development, and storage conditions. Genetic factors determine the strength of seed vigor, while environmental conditions (including temperature, light, mineral nutrition, and pests) affect the extent of seed vigor expression [39]. Therefore, the vigor index is associated with the degree of damage caused by pests to soybean seeds, and the measurement of vigor index will provide data for the subsequent extraction of characteristic wavelengths. Its calculation formula is as follows:

$$VI = S \times GI \quad (3)$$

where  $S$  stands for No. of seedlings length (the length excluding the cotyledon but including the root system, and it is measured directly with a standard ruler).

**The structured light reflection image of soybean is obtained in the forward direction.**

#### *Image acquisition conditions*

Before the formal experiment, it is important to note that there is a negative correlation between the depth

of penetration and the resolution of the final image based on light transmission. Choosing the appropriate frequency directly affects the accuracy of the results. Therefore, we conducted pre-experiments to explore the imaging effects of soybean seeds at different frequencies. We found that clear imaging and moderate penetration depth were achieved at frequencies of 65, 75, and 85 (unit:  $m^{-1}$ ). Therefore, we selected these three frequencies for the current experiment.

During the capturing process, the soybean samples were placed on a manually adjustable lifting platform. The height of the experimental platform was set to 146 mm, and the distance from the projector was 500 mm. The camera and projector were tilted at a 10-degree angle in the vertical direction. This setup was intended to minimize the specular reflection of light on the surface of the soybean.

Under the sinusoidal illumination mode, each sample was photographed under three different phases for each of the three wavelength bands: 502 nm, 813 nm, and 712 nm. The specific phases were set as 0,  $\pi/3$ , and  $2\pi/3$ .

#### *Image preprocessing*

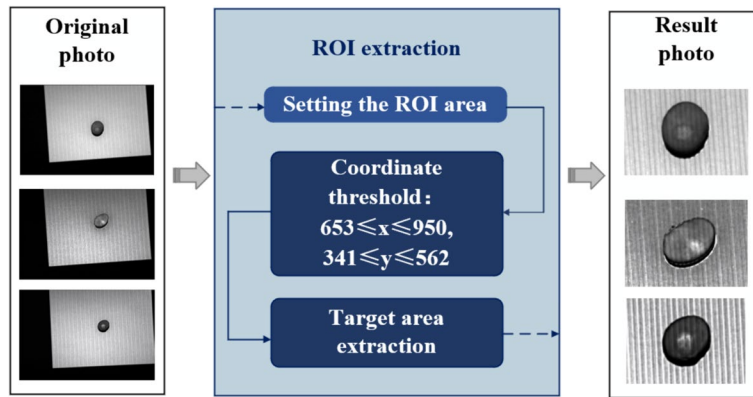
Based on various reasons, the preliminary obtained images suffer from issues such as high noise levels, large data volume, and inconsistent data scales. To eliminate differences in light intensity caused by variations in experimental time and stabilize the reflectance variations in the images, an initial black-white calibration is performed. The calculation formula for the calibration is as follows:

$$I_{relatively} = \frac{I_{source} - I_{dark}}{I_{bright} - I_{dark}} \quad (4)$$

where  $I_{relatively}$  is the result image after black and white correction,  $I_{source}$  is the original image,  $I_{bright}$  is the white-board image obtained with the light source turned on,  $I_{dark}$  is the dark light field image with the light source turned off.

To improve the accuracy of data processing while reducing the amount of data processing, we extracted the Region of Interest (ROI) of the image after black and white correction (see Fig. 5), which not only reduced the resolution of the image, but also increased the proportion of soybean seeds in the whole image [40].

Finally, the images are further processed with a smoothing technique using a Gaussian low-pass filter to preserve the low-frequency information and eliminate high-frequency segmental noise. The kernel size for the filter is selected as  $3 \times 3$ .



**Fig. 5** Soybean seeds ROI extraction process

### Image demodulation

The diffuse reflection of the image captured by the camera consists of two components, AC component and DC component, which are calculated as follows:

$$I(x, f_x) = I_{DC}(x, f_x) + I_{AC}(x, f_x) \quad (5)$$

where  $I_{AC}(x, f_x)$  and  $I_{DC}(x, f_x)$  represent the reflection intensities of AC and DC images, respectively. The DC component is only related to the intensity of the light source, while the AC component is not only related to the intensity of the light source, but also to the frequency of the projection [41]. Based on obtaining the three phase images, we can obtain the AC and DC images by the following equation:

$$M_{AC}(x, f_x) = \frac{\sqrt{2}}{3} [(I_1 - I_2)^2 + (I_1 - I_3)^2 + (I_2 - I_3)^2]^{\frac{1}{2}} \quad (6)$$

$$I_{AC}(x, f_x) = M_{AC}(x, f_x) \cdot \cos(2\pi f_x + \varphi) \quad (7)$$

$$I_{DC}(x, f_x) = \frac{1}{3}(I_1 + I_2 + I_3) \quad (8)$$

where  $I_1, I_2$  and  $I_3$  are the images with offsets of  $0, \pi/3$  and  $2\pi/3$  respectively.  $M_{AC}(x, f_x)$  is referred to as the amplitude envelope of the diffuse intensity. Due to the significantly higher attenuation rate in tissues compared to A, we introduce a ratio image [15] to enhance the contrast of the image. The calculation formula for the ratio image is as follows:

$$\gamma = \frac{I_{AC}}{I_{DC}} = \frac{\sqrt{2}}{I_1 + I_2 + I_3} [(I_1 - I_2)^2 + (I_1 - I_3)^2 + (I_2 - I_3)^2]^{\frac{1}{2}} \quad (9)$$

where  $\gamma$  is the ratio image, where the value of each pixel ranges from 0 to 1, resulting in improved imaging performance.

### The feasibility of the algorithm is tested by reverse inversion.

Before performing inverse inversion, to mitigate the influence of systematic errors, we can employ a single integrating sphere to directly measure the diffuse reflection from the surface of a specific sample. This measured sample can serve as a reference sample. Then, by utilizing the following formula (10), we can determine the corrected diffuse reflection of the test sample.

$$R(x, f_x) = \frac{M_{AC}(x, f_x)}{M_{AC,ref}(x, f_x)} \cdot R_{ref}(f_x) \quad (10)$$

where  $R_{ref}(f_x)$  is the diffuse reflection of the reference sample and  $M_{AC,ref}(x, f_x)$  is the amplitude envelope of the diffuse reflection intensity of the reference sample. After correcting the diffuse reflection of the sample to be measured,  $\mu_a$  and  $\mu'_s$  can be obtained by nonlinear fitting Eq. (11) using the following Eq. (11–14):

$$R(f_x) = \frac{3A\mu'_s/\mu_{tr}}{(\mu'_{eff}/\mu_{tr} + 1)(\mu'_{eff}/\mu_{tr} + 3A)} \quad (11)$$

$$A = \frac{1 - R_{eff}}{2(1 + R_{eff})} \quad (12)$$

$$R_{eff} \approx 0.0636n + 0.668 + 0.71/n - 1.44/n^2 \quad (13)$$

$$\mu'_{eff} = \sqrt{3\mu_a\mu_{tr} + 4\pi^2 f_x^2} \quad (14)$$

$$\mu_{tr} = \mu_a + \mu'_s \quad (15)$$

$$e_r(x) = \frac{|x - x^*|}{x^*} \times 100\% \quad (16)$$



where  $A$  is the scaling constant,  $R_{eff}$  is the effective reflection coefficient,  $n$  is the sample refractive index,  $\mu'_{eff}$  is the reduced attenuation coefficient,  $\mu_{tr}$  is the full attenuation coefficient,  $e_r(x)$  is the relative error,  $x$  is the measured value and  $x^*$  is the reference value.

### Modeling insect pest levels in soybean seeds

Due to the challenging nature of acquiring spatial frequency domain images of soybean seeds and the limited availability of experimental datasets, this experiment employed three shallow neural networks, namely Support Vector Machine (SVM), Generalized Regression Neural Network (GRNN), and Mixed Logistic Regression (MLR), to establish models for three different soybean varieties (NN1138-2, SX26, QY1808). To obtain more reliable evaluation results and maintain class balance, we employed a stratified sampling method to partition the dataset, ensuring a similar proportion of samples from each class in the training and testing sets. The division ratio was set at 7:3.

SVM [42] is a binary classification model (Dichotomous models, DM). In comparison to traditional neural networks, SVM offers advantages such as handling small sample sizes, good generalization ability, handling interactions between non-linear features, and solving high-dimensional problems.

GRNN [43] belongs to the Radial Basis Function (RBF) neural network family and represents an improved version of RBF. It possesses a strong ability to learn nonlinear mappings. In comparison to RBF, GRNN offers faster learning speed and exhibits notable advantages in prediction performance when dealing with small datasets and high levels of data noise [44].

MLR [45] is an optimization of the linear regression model (Logistic regression, LR), and its fundamental optimization concept is the divide-and-conquer approach [46]. By partitioning high-dimensional data into regions, each region corresponds to a linear LR model. The MLR model utilizes these linear models fitted to different partitions to achieve a nonlinear pattern for high-dimensional data.

During the training of the classification model, the input data consisted of a (210×5) array representing the five detection features of the soybean seeds. These features include the ratio of the insect-damaged sub-surface area to the entire seed, soybean variety, and the measured values of  $\mu_a$  at 502 nm, 813 nm, and 712 nm wavelengths. The output array was a (210×1) array corresponding to the pest damage level based on the ratio of damaged area after final germination to the entire seed.

To achieve optimal results for the predictive model, parameter tuning is crucial. In this regard, we utilize a reinforcement learning-based multi-armed bandit

approach to autonomously learn and adjust the parameters of the model to obtain the best possible results.

A multi-armed bandit [47] is an algorithm in reinforcement learning that is based on the probability theory. It can be described as a slot machine with multiple lever arms. Each arm corresponds to a different probability of winning, and our goal is to maximize the winnings within a limited number of trials [48]. The fundamental principle of reinforcement learning is to continuously explore and learn from experiences, where different actions yield varying rewards. Higher rewards are given when the results are closer to the expected values, while lower rewards are assigned otherwise [49]. In contrast to exhaustive trial and error, the agent in reinforcement learning does not equally explore all parameters. Instead, it evaluates different parameters based on past experiences and selects the ones that offers the highest rewards. By trying out new parameters and updating its experiences, the agent, much like a human, learns continuously and eventually becomes an "expert" in those parameters, capable of selecting better ones.

In this paper, the intelligent agent (Agent) refers to the algorithm itself, the environment (Environment) refers to the neural network model, the action (Action) refers to trying different model parameters, and the reward (Reward) is based on the results of cross-validation. Our ultimate objective is the value function, specifically the state-action value function  $Q(s, a)$  ( $s$  stands for state and  $a$  stands for action). Since the state in a multi-armed bandit is fixed and unchanging [50], we can simplify  $Q(s, a)$  to the action-value function  $Q(a)$ , which is solely dependent on the action. The calculation formula for  $Q(a)$  is as follows:

$$Q_t(a) = \frac{\sum_{i=1}^{t-1} \text{Reward}}{\sum_{i=1}^{t-1} 1} \quad (17)$$

where  $Q_t(a)$  refers to the value estimate of selecting action  $a$  at time  $t$ , and its value depends on the synthesis of previous value estimates. The selection of actions in the multi-armed bandit employs the Upper Confidence Bound Algorithm (UCB) [51]. The specific formula for UCB is as follows:

**Table 1** Seeds germination of different soybean varieties

Germination indicators	NN1138-2	QY1808	SX26
Germination rate	0.916	0.875	0.830
Germination index	41.2	30.25	24.33
Vigor index	5001.0	3384.2	3218.4



**Fig. 6** Grading of seed germination and cotyledon damage symptoms. **a** original image. **b** damage level classification image

$$A_t = \operatorname{argmax}(Q(a) + \sqrt{\frac{2 \ln t}{N_t(a)}}) \quad (18)$$

where  $A_t$  represents the action selected at time  $t$ , and  $N_t(a)$  represents the number of times the action has been selected at time  $t$ .

## Results

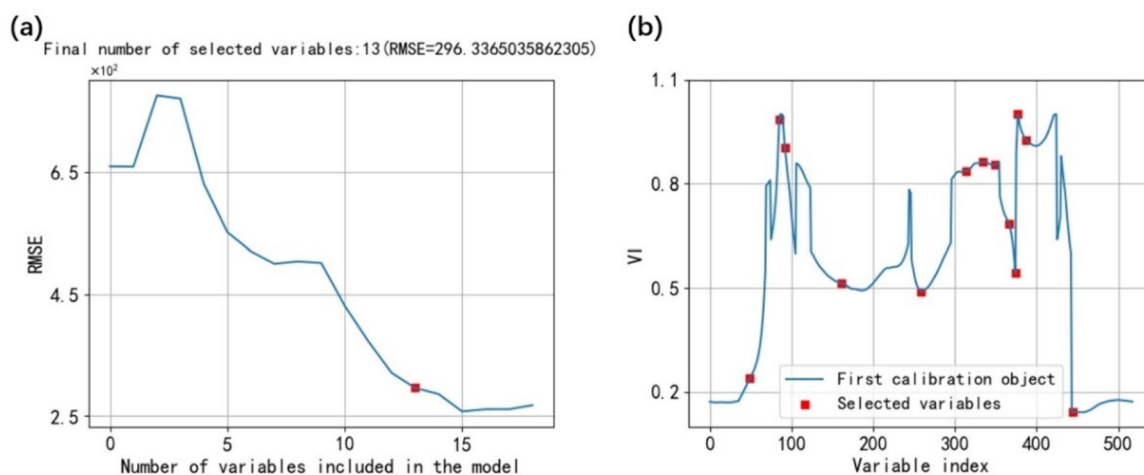
### Germination data of soybean seeds.

The germination experimental data of soybean seeds is shown in Table 1. This data will serve as the foundational data for extracting soybean characteristic wavelengths.

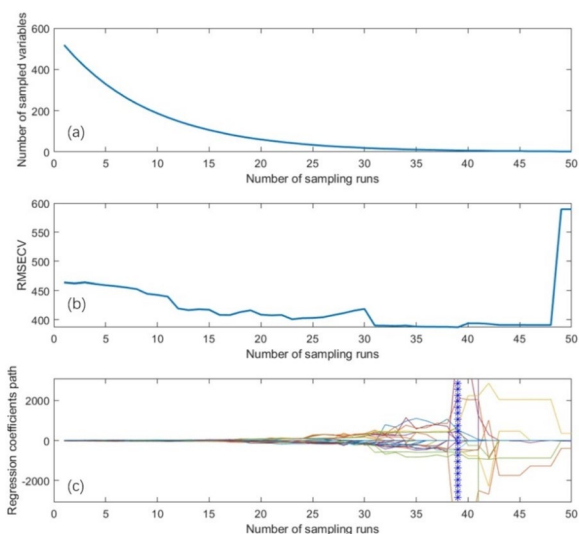
Meanwhile, we classified the extent of damage to soybean buds. The classification criterion was based on the proportion of the black damaged area (including detached parts) to the entire soybean after germination. Based on the proportion, it was classified into three categories: mild damage, moderate damage, and severe damage (see Fig. 6).

### Extraction of characteristic wavelength of soybean seed

The feature wavelength extraction was performed using two algorithms: the successive projections algorithm (SPA) and the competitive adaptive reweighted sampling (CARS) algorithm. In these algorithms, the operation matrix represents the samples, while the columns represent the spectral bands. The wavelength range spans from 450 to 950 nm and is divided into 518 intervals. The initial iteration vector was set as the soybean seed vigor index, and the remaining vectors in the matrix corresponded to the reflectance or transmittance values for different bands. The step size was selected as 1. Root mean square error (RMSE) is a common index to measure the prediction accuracy of regression models. It quantifies the difference between the predicted value and the actual observed value. The root-mean-square error of cross-validation (RMSECV), a subclass of RMSE, is obtained by repeating the cross-validation process several times and calculating the average of the RMSE



**Fig. 7** **a** RMSE changes with the number of set variables in SPA wavelength selection and **b** Characteristic wavelength selection



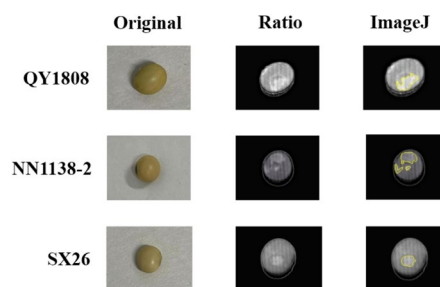
**Fig. 8** CARS characteristic wavelength selection. **a** Variation trend of the number of variables with the number of samples; **b** RMSECV; **c** The change process of regression coefficient of each variable with sampling times

values obtained in each iteration [52]. In SPA and CARS, RMSE and RESECV are selected as the criteria to evaluate the relationship between features and target variables according to the algorithm properties [53]. After each iteration, the change in RMSE(RMSECV) is monitored to evaluate the impact of adding or removing variables. A smaller RMSE(RMSECV) value indicates smaller prediction errors, indicating better model fit and prediction accuracy.

Based on the SPA [54], the dimensionality reduction results for the soybean reflectance dataset are shown in Fig. 7. The RMSE reaches its minimum value when the number of feature wavelengths is 13, resulting in a compression rate of 97.49%. The selected 13 feature wavelengths are as follows: 501.55 nm, 798.08 nm, 712.34 nm, 813.91 nm, 615.42 nm, 545.12 nm, 884.94 nm, 784.02 nm, 765.12 nm, 833.29 nm, 821.31 nm, 824.08 nm, and 537.9 nm.

Based on the CARS algorithm [55], the dimensionality reduction results for the soybean reflectance dataset are shown in Fig. 8. The RMSECV value is achieved in the 39th run. Therefore, the wavelength variables retained in the 39th run are considered as the selected feature wavelengths, with a count of 7. The specific values are 520.29 nm, 813.02 nm, 700.66 nm, 701.64 nm, 774.59 nm, 784.02 nm, and 874.18 nm.

Referring to the five characteristic wavelengths with the highest weight coefficients of the two methods, three wavelengths of 502 nm, 813 nm and 712 nm were finally selected as the following experimental wavelengths. The



**Fig. 9** The original images, ratio images and images processed using ImageJ of soybean seeds for varieties QY1808, NN1138-2 and SX26

400-700 nm region is affected by the absorption light of carotenoids and other pigments [56], while 710 nm is close to the vibration of H element [57], and soybeans exhibit an absorption peak near 820 nm, which represents carotenoid content [58]. Therefore, the three wavelengths selected in this paper have a theoretical basis for soybeans and are used as experimental wavelengths for subsequent experiments.

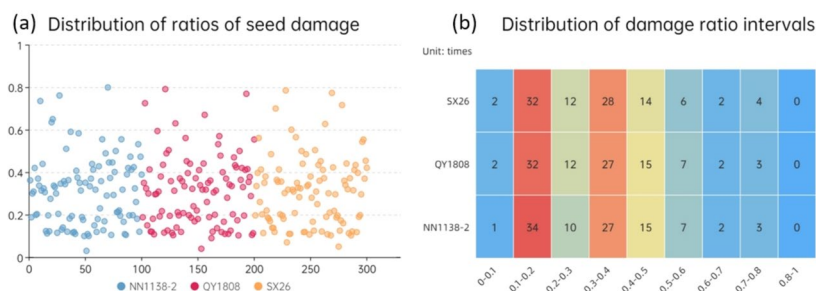
#### Measurement of the subsurface area of soybean seeds

After obtaining Ratio images of QY1808, NN1138-2 and SX26, the soybean seeds were subjected to damage area extraction using professional software ImageJ. As shown in Fig. 9. The ratio of the damaged area to the entire seed was calculated, and these ratios will be used as inputs for the neural network model.

In this process, we measured 100 seeds for each variety, amounting to a total of 300 soybean seeds. The overall distribution of the measurement data is shown in Fig. 10. The measurement results of the three soybean varieties are relatively evenly distributed, and the distribution patterns among different varieties are also similar. From the distribution heatmap in the right figure (b), it can be observed that the distribution of damage ratio intervals is primarily concentrated between 0.1 and 0.5.

#### Measurement results and analysis of spatial frequency domain method

In this measurement experiment, the samples are divided into healthy soybean seeds and soybean seeds that have been subjected to piercing and sucking by *R. pedestris*, but with a concealed appearance of damage on the surface. Sample values are obtained through spatial frequency domain inversion, while reference values are measured using the integrating sphere method. Overall, the measurement results obtained in the experiment have errors within an acceptable range (15%). This confirms that the experimental algorithm is feasible (see Table 2).



**Fig. 10** The distribution of ratios of NN1138-2, QY1808 and SX26 damaged area to the entire seed. **a** Scatter plot of ratios of seed damage; **b** Heatmap of distribution of damage ratio intervals, warm colors indicate a larger quantity, while cool colors indicate a smaller quantity

**Table 2** Optical properties of normal and damaged soybean seeds

Variety	Measurement target	$\mu_a(mm^{-1})$			$\mu'_s(mm^{-1})$		
		502	813	712	502	813	712
NN 1138-2	Normal sample	0.07802	0.03961	0.03004	0.82389	0.82731	0.80683
	Reference value	0.09107	0.04571	0.03478	0.91493	0.90724	0.88469
	Relative error (%)	14.33	13.35	13.62	9.95	8.81	8.80
	Damaged sample	0.11498	0.06772	0.04944	0.87472	0.75930	0.89752
	Reference value	0.13190	0.07571	0.05781	0.93217	0.82343	0.98469
	Relative error (%)	12.83	10.55	14.48	6.17	7.79	8.86
QY 1808	Normal sample	0.05531	0.01769	0.00952	0.76929	0.74397	0.81923
	Reference value	0.06086	0.02033	0.01090	0.83873	0.82367	0.88179
	Relative error (%)	9.12	12.99	12.63	8.28	9.71	7.10
	Damaged sample	0.16160	0.05213	0.03284	0.81472	0.70931	0.78346
	Reference value	0.17976	0.05792	0.03799	0.88294	0.75724	0.83469
	Relative error (%)	10.11	10.00	14.56	7.73	6.33	7.14
SX 26	Normal sample	0.06189	0.01687	0.01462	0.72189	0.72385	0.71936
	Reference value	0.07043	0.01935	0.01647	0.78294	0.77539	0.77926
	Relative error (%)	12.13	12.84	11.26	7.80	6.65	7.69
	Damaged sample	0.16166	0.05320	0.04267	0.73899	0.73196	0.75159
	Reference value	0.19002	0.06155	0.04998	0.80471	0.80001	0.81938
	Relative error (%)	14.93	13.57	14.62	8.17	8.51	8.28

**Table 3** Comparison of modeling performance for pest damage levels of soybean seeds

Model	Test set		
	Macro-recall	f1-score	Accuracy
MLR	0.8603	0.8931	0.9213
GRNN	0.8269	0.8678	0.9444
SVM	0.8828	0.9079	0.9325
RL-MLR	0.9556	0.9509	0.9437
RL-GRNN	0.9583	0.9737	0.9889
RL-SVM	0.9635	0.9754	0.9883

**Prediction results of soybean pest damaged degree**

In this study, three methods, namely SVM, GRNN, and MLR, were used to build a prediction model for soybean pest levels. The training set and test set were set at a ratio of 7:3. The parameters of the three models, including (kernel, C) for SVM, sigma for GRNN, and alpha for MLR, were optimized and adjusted using a multi-armed bandit approach. The final modeling results are compared and shown in Table 3.

Macro-Recall is the average Recall, which measures the ability of a model to correctly predict positive samples for each class [59]. Accuracy measures the ratio of

correctly predicted samples to the total number of samples. F1-score is the harmonic mean of Accuracy and Recall, providing a balanced evaluation of the model's performance.

Due to having three categories in this classification task with imbalanced class distribution, Accuracy may be dominated by the predictions of the majority class, thereby being less sensitive to the classification performance of minority classes [60]. In contrast, Macro-Recall considers the recall of each individual class. Therefore, the Macro-Recall metric is of greater importance. Prior to optimizing the parameters of the neural network model, the Macro-Recall values were relatively low. However, after optimizing the parameters using reinforcement learning, the RL-SVM classification model exhibited the best performance, with a Macro-Recall of 0.9635, F1-Score of 0.9754, and Accuracy of 0.9883.

## Discussion

The three soybean seed varieties exhibit varying degrees of increase in  $\mu_a$  after being attacked by the *R. pedestris*. The presence of the *R. pedestris* begins during the R2 to R4 stages of soybean growth [61]. During the R3 stage, the uppermost four nodes on the main stem of the soybean plant, each with sufficient leaf growth, possess pods that are at least 5 mm in length [62]. At this stage, *R. pedestris* pierces and sucks the soybean pods using its mouthparts, and its saliva contains a significant number of digestive enzymes. This saliva causes the soybean tissue to become uneven and discontinuous, with numerous air-filled cavities. Multiple scattering during the transmission process enhances the tissue's light absorption, resulting in an increase in  $\mu_a$  [63]. Among the three varieties, the increment in  $\mu_a$  is smaller in NN1138-2 pest-resistant seeds compared to SX26 and QY1808. This is due to NN1138-2 exhibiting a lower level of pest infestation. In our field and greenhouse insect resistance identification experiments, NN1128-2 was identified as a resistant variety, while QY1808 and SX26 were found to be susceptible varieties. Additionally, we discovered a significant negative correlation between soybean's insect resistance and hundred-grain weight. In our experiments, the pod damage index and kernel damage index were utilized to quantify the extent of *R. pedestris* damage on soybeans. The severity of damage was directly proportional to higher values of both indices. The correlation coefficient between the pod damage index and hundred-grain weight was found to be 0.47, indicating a significant relationship. Similarly, the correlation coefficient between the kernel damage index and hundred-grain weight was determined to be 0.74, also signifying a strong statistical significance. Research indicates that hundred-grain weight is also a mechanism of insect resistance, defined

as damage dilution, which helps reduce losses caused by insect feeding [64]. Generally, lower hundred-grain weight corresponds to a higher number of grains per plant and a smaller proportion of damaged grains. The hundred-grain weights of NN1138-2, QY1808, and SX26 are 18.9, 20.0, and 33.3 g, respectively.

Using the spatial frequency domain measurement method, the optical parameter analysis of soybeans was conducted. The experiment measured the  $\mu_a$  and  $\mu'_s$  of three soybean varieties: SX26, NN1138-2, and QY1808. It was found that there was a significant difference between the measured  $\mu_a$  and the results obtained using a single integrating sphere method. This discrepancy may be due to the small numerical value of  $\mu_a$ , which poses statistical challenges [32]. The highest value of  $\mu_a$  for soybeans was observed at 502 nm, followed by 813 nm, while the lowest value was at 712 nm. This variation is attributed to the differential light absorption capacity of chemical substances at different wavelengths. Both 502 nm and 813 nm are influenced by the absorption of carotenoids [44], with greater absorption at 502 nm compared to 813 nm. Near 712 nm, there is no strong light-absorbing substance, resulting in a lower  $\mu_a$  value. In soybeans damaged by insects, both  $\mu_a$  and  $\mu'_s$  showed varying degrees of increase. However, the change in  $\mu_a$  was much larger than the change in  $\mu'_s$ , indicating that  $\mu_a$  is an important parameter for studying soybean seeds pest damage.

Currently, researchers have been using infrared spectroscopy and hyperspectral imaging techniques to detect seeds. Different components of a substance have distinct spectral characteristics, which means they absorb, reflect, or scatter light differently at different wavelengths. By analyzing the spectral information in the spectral image, we can infer the composition, concentration, texture, and other internal characteristics of the substance [65]. However, these methods require high experimental requirements, as variations and non-uniformity in lighting can lead to inconsistency or noise in the spectral data, affecting the accuracy and reliability of the data. Moreover, hyperspectral imaging generates a large amount of data, with each pixel containing a significant amount of spectral information [66]. Therefore, data processing and analysis require complex algorithms and computational methods, demanding significant computational resources and expertise.

Some scholars have also used X-rays and terahertz for seed detection. However, X-rays have high energy and ionizing radiation characteristics, which may pose certain radiation hazards to biological tissues [67]. Terahertz detectors, on the other hand, are expensive, and their sensitivity and dynamic range need further improvement to meet the requirements of high-quality imaging [68]. Considering factors such as detection cycle length and cost, these methods remain at the laboratory stage.

In comparison, spatial frequency imaging methods project light of different frequencies onto the surface of an object, capturing direct current and alternating current images that contain depth information. These images are further processed to enhance the contrast of the information. Not only is the equipment cost low, but the detection cycle is also short. Additionally, it allows for non-destructive detection of subsurface information in crops. It has excellent prospects for application.

Due to objective factors such as a series of strict reviews and controls required for inoculating pests on soybean seeds, all soybeans used in this experiment were yellow soybeans. The current model in this paper is still limited to the detection of yellow soybeans. We will further increase the detection of soybeans of different colors in the future.

## Conclusion

The methods for manually detecting soybean seeds pest damage are primarily limited to visual identification and measuring overall germination rates. These methods can only assess soybean seeds with visible surface defects or evaluate the overall performance of a particular variety. They also suffer from subjectivity. In this study, a spatial frequency domain imaging method was employed to acquire sub-surface images of soybeans and calculate their optical characteristic parameters. Based on this, a method for early detection of soybean pest damage was investigated. The following conclusions were drawn:

1. At the three wavelengths of 502 nm, 813 nm, and 712 nm, the  $\mu_a$  of soybeans is highest at 502 nm, followed by 813 nm, and lowest at 712 nm, while  $\mu'_s$  shows no significant difference. Feeding by *R. pedestris* causes a significant increase in  $\mu_a$  of soybeans without affecting  $\mu'_s$ . Among the three varieties, NN1138-2 exhibits a noticeably lower increase in  $\mu_a$  compared to the other two, indicating it is less affected by pest damage.
2. Autonomous adjustment of neural network parameters based on reinforcement learning methods can effectively improve the performance of neural networks and has broad adaptability across different models. The Macro-Recall metric for different models is improved by 10–15%. Among the optimized models, RL-SVM performs well in predicting soybean seeds pest damage levels, achieving a Macro-Recall of 0.9635.

Overall, the spatial frequency domain imaging technique can accurately measure the optical characteristic parameters of soybeans and acquire sub-surface images that carry depth information. The construction of an RL-SVM

classification model for soybean pest detection and identification is feasible, and it has the potential for further expansion to quality assessment of seeds from other crops.

## Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s13007-024-01257-5>.

Additional file 1

### Author contribution

XC was responsible for coding the spatial frequency domain imaging system and neural network model, as well as drafting the manuscript. WH participated in the design and construction of the spatial frequency domain imaging system. ZY was responsible for soybean inoculation with *Riptortus pedestris* and collected germination experiment data. WL proposed the idea of exploring soybean pest detection methods based on the spatial frequency domain system and provided guidance for the experiments. JG and GX were responsible for the biochemical analysis of soybean seeds and guided the research on the relationship between soybean seed properties and pest damage. All authors read and approved the final manuscript.

### Funding

This work was supported by the National Natural Science Foundation of China (32071896).

### Availability of data and materials

No datasets were generated or analysed during the current study.

## Declarations

### Ethics approval and consent to participate

Not applicable.

### Competing interests

The authors declare no competing interests.

Received: 24 November 2023 Accepted: 2 August 2024

Published: 20 August 2024

## References

1. Zhang T, Wu T, Wang L, et al. A combined linkage and GWAS analysis identifies QTLs linked to soybean seed protein and oil content. *Int J Mol Sci.* 2019;20(23):5915.
2. Li W, Gao Y, Hu Y, et al. Field cage assessment of feeding damage by *Riptortus pedestris* on soybeans in China. *Insects.* 2021;12(3):255.
3. Wei Z, Guo W, Jiang S, et al. Transcriptional profiling reveals a critical role of GmFT2a in soybean staygreen syndrome caused by the pest *Riptortus pedestris*. *New Phytol.* 2023;237(5):1876–90.
4. Matsumoto H, Fan X, Wang Y, et al. Bacterial seed endophyte shapes disease resistance in rice. *Nat Plants.* 2021;7(1):60–72.
5. Rezvyakova S, Eremin L, Matveychuk P, et al. The influence of biofungicide and chemical fungicides on the manifestation of diseases and the yield of soybeans[C]//E3S Web of Conferences. EDP Sciences. 2021;247:01046.
6. Sun Y, Lu R, Lu Y, et al. Detection of early decay in peaches by structured-illumination reflectance imaging. *Postharvest Biol Technol.* 2019;151:68–78.
7. Kimura S, Tokumaru S, Kikuchi A. Carrying and transmission of *Eremothecium coryli* (Peglion) Kurtzman as a causal pathogen of yeast-spot disease in soybeans by *Riptortus clavatus* (Thunberg), *Nezara antennata* Scott, *Piezodorus hybneri* (Gmelin) and *Dolycoris baccarum* (Linnaeus). *Jpn J Appl Entomol Zool.* 2008;52(1):13–8.
8. Li K, Zhang X, Guo J, et al. Feeding of *Riptortus pedestris* on soybean plants, the primary cause of soybean staygreen syndrome in the Huang-Huai-Hai river basin. *Crop J.* 2019;7(3):360–7.

9. Gao L, Sun S, Li K, et al. Spatio-temporal characterisation of changes in the resistance of widely grown soybean cultivars to Soybean mosaic virus across a century of breeding in China. *Crop Pasture Sci.* 2018;69(4):395–405.
10. Huang Y, Lu R, Chen K. Detection of internal defect of apples by a multi-channel Vis/NIR spectroscopic system. *Postharvest Biol Technol.* 2020;161:111065.
11. Sun Y, Li Y, Pan L, et al. Authentication of the geographic origin of Yangshan region peaches based on hyperspectral imaging. *Postharvest Biol Technol.* 2021;171:111320.
12. Barboza da Silva C, Oliveira NM, de Carvalho MEA, et al. Autofluorescence-spectral imaging as an innovative method for rapid, non-destructive and reliable assessing of soybean seed quality. *Sci Rep.* 2021;11(1):17834.
13. de Medeiros AD, Bernardes RC, da Silva LJ, et al. Deep learning-based approach using X-ray images for classifying *Crambe abyssinica* seed quality. *Ind Crops Prod.* 2021;164:113378.
14. Hu J, Lv H, Qiao P, et al. Research on rice seed fullness detection method based on terahertz imaging technology and feature extraction method. *J Infrared, Millim Terahertz Waves.* 2023;44:407–29.
15. Lu Y, Li R, Lu R. Structured-illumination reflectance imaging (SIRI) for enhanced detection of fresh bruises in apples. *Postharvest Biol Technol.* 2016;117:89–93.
16. Cuccia DJ, Bevilacqua F, Durkin AJ, et al. Modulated imaging: quantitative analysis and tomography of turbid media in the spatial-frequency domain. *Opt Lett.* 2023;30(11):1354–6.
17. Ponticorvo A, Rowland R, Baldado M, et al. Spatial frequency domain imaging (SFDI) of clinical burns: a case report. *Burns Open.* 2020;4(2):67–71.
18. Lee S, Mey L, Szymanska AF, et al. SFDI biomarkers provide a quantitative ulcer risk metric and can be used to predict diabetic foot ulcer onset. *J Diabetes Complicat.* 2020;34(9):107624.
19. Schmidt M, Aguénounon E, Nahas A, et al. Real-time, wide-field, and quantitative oxygenation imaging using spatiotemporal modulation of light. *J Biomed Opt.* 2019;24(7):071610–071610.
20. Chen MT, Durr NJ. Rapid tissue oxygenation mapping from snapshot structured-light images with adversarial deep learning. *J Biomed Opt.* 2020;25(11):112907–112907.
21. Anderson ER, Cuccia DJ, Durkin AJ. Detection of bruises on golden delicious apples using spatial-frequency-domain imaging. *Adv Biomed Clin Diagn Syst.* 2007;6430:308–18.
22. Luo Y, Jiang X, Fu X. Spatial frequency domain imaging system calibration, correction and application for pear surface damage detection. *Foods.* 2021;10(9):2151.
23. Zhao G, Quan L, Li H, et al. Real-time recognition system of soybean seed full-surface defects based on deep learning. *Comput Electron Agric.* 2021;187:106230.
24. Yan S, Xu J, Zhang S, et al. Effects of flexibility and surface hydrophobicity on emulsifying properties: ultrasound-treated soybean protein isolate. *LWT.* 2021;142:110881.
25. Mao T, Zhou DX. Rates of approximation by ReLU shallow neural networks. *J Complex.* 2023;79:101784.
26. Agliari E, Alemanno F, Barra A, et al. The emergence of a concept in shallow neural networks. *Neural Netw.* 2022;148:232–53.
27. Kamilaris A, Prenafeta-Boldú FX. Deep learning in agriculture: a survey. *Comput Electron Agric.* 2018;147:70–90.
28. Bejani MM, Ghatte M. A systematic review on overfitting control in shallow and deep neural networks. *Artif Intell Rev.* 2021. <https://doi.org/10.1007/s10462-021-09975-1>.
29. Xiuxia C, Pin Z, Shuaibin D. Imitation camouflage synthesis based on shallow neural network. *Multimed Syst.* 2023. <https://doi.org/10.1007/s00530-023-01149-z>.
30. Zhang K, Yang Z, Başar T. Multi-agent reinforcement learning: A selective overview of theories and algorithms. In: Vamvoudakis KG, Wan Y, Lewis FL, Cansever D, editors. *Handbook of reinforcement learning and control*. Cham: Springer International Publishing; 2021. p. 321–84.
31. Perrusquía A, Yu W. Identification and optimal control of nonlinear systems using recurrent neural networks and reinforcement learning: an overview. *Neurocomputing.* 2021;438:145–54.
32. Hu D, Fu X, He X, et al. Noncontact and wide-field characterization of the absorption and scattering properties of apple fruit using spatial-frequency domain imaging. *Sci Rep.* 2016;6(1):37920.
33. Lu R, Van Beers R, Saeys W, et al. Measurement of optical properties of fruits and vegetables: a review. *Postharvest Biol Technol.* 2020;159:111003.
34. Debije MG, Evans RC, Griffini G. Laboratory protocols for measuring and reporting the performance of luminescent solar concentrators. *Energy Environ Sci.* 2021;14(1):293–301.
35. He X, Fu X, Rao X, et al. Assessing firmness and SSC of pears based on absorption and scattering properties using an automatic integrating sphere system from 400 to 1150 nm. *Postharvest Biol Technol.* 2016;121:62–70.
36. Pahl SA, van Gemert MJC, Welch AJ. Determining the optical properties of turbid media by using the adding–doubling method. *Appl Opt.* 1993;32(4):559–68.
37. Cen H, Lu R. Optimization of the hyperspectral imaging-based spatially-resolved system for measuring the optical properties of biological materials. *Opt Express.* 2010;18(16):17412–32.
38. Hu D, Lu R, Ying Y. A two-step parameter optimization algorithm for improving estimation of optical properties using spatial frequency domain imaging. *J Quant Spectrosc Radiat Transfer.* 2018;207:32–40.
39. Cheng H, Ye M, Wu T, Ma H. Evaluation and heritability analysis of the seed vigor of soybean strains tested in the Huanghuaihai regional test of China. *Plants.* 2023;12(6):1347.
40. He W, Ye Z, Li M, et al. Extraction of soybean plant trait parameters based on SFM-MVS algorithm combined with GRNN. *Front Plant Sci.* 2023;14:1181322.
41. Cuccia DJ, Bevilacqua F, Durkin AJ, et al. Quantitation and mapping of tissue optical properties using modulated imaging. *J Biomed Optics.* 2009;14(2):024012.
42. Vukovic DB, Romanyuk K, Ivashchenko S, et al. Are CDS spreads predictable during the Covid-19 pandemic? Forecasting based on SVM, GMDH, LSTM and Markov switching autoregression. *Expert Syst Appl.* 2022;194:116553.
43. Lu W, Du R, Niu P, et al. Soybean yield preharvest prediction based on bean pods and leaves image recognition using deep learning neural network combined with GRNN. *Front Plant Sci.* 2022;12:791256.
44. Wang J, Zhu L, Dai H. An efficient state-of-health estimation method for lithium-ion batteries based on feature-importance ranking strategy and PSO-GRNN algorithm. *J Energy Storage.* 2023;72:108638.
45. Chen L, Fakharian P, Eidgahee DR, et al. Axial compressive strength predictive models for recycled aggregate concrete filled circular steel tube columns using ANN, GEP, and MLR. *J Build Eng.* 2023;77:107439.
46. Shams SR, Jahani A, Kalantary S, et al. The evaluation on artificial neural networks (ANN) and multiple linear regressions (MLR) models for predicting SO<sub>2</sub> concentration. *Urban Clim.* 2021;37:100837.
47. Xu X, Tao M. Decentralized multi-agent multi-armed bandit learning with calibration for multi-cell caching. *IEEE Trans Commun.* 2020;69(4):2457–72.
48. Bhuyan AK, Dutta H, Biswas S. Federated multi-armed bandit learning for caching in UAV-aided content dissemination. *Ad Hoc Netw.* 2023;151:103306.
49. Matsuo Y, LeCun Y, Sahani M, et al. Deep learning, reinforcement learning, and world models. *Neural Netw.* 2022;152:267–75.
50. Gupta S, Chaudhari S, Joshi G, et al. Multi-armed bandits with correlated arms. *IEEE Trans Inf Theory.* 2021;67(10):6711–32.
51. Shen B, Gnanasambandam R, Wang R, et al. Multi-task Gaussian process upper confidence bound for hyperparameter tuning and its application for simulation studies of additive manufacturing. *IJSE Trans.* 2023;55(5):496–508.
52. Chicco D, Warrens MJ, Jurman G. The coefficient of determination R-squared is more informative than SMAPE, MAE, MAPE, MSE and RMSE in regression analysis evaluation. *Peer J Comput Sci.* 2021;7: e623.
53. Li Y, Yang X. Quantitative analysis of near infrared spectroscopic data based on dual-band transformation and competitive adaptive reweighted sampling. *Spectrochim Acta Part A Mol Biomol Spectrosc.* 2023;285:121924.
54. Qin Y, Song K, Zhang N, et al. Robust NIR quantitative model using MIC-SPA variable selection and GA-ELM. *Infrared Phys Technol.* 2023;128:104534.
55. Hu Q, Lu W, Guo Y, et al. Vigor detection for naturally aged soybean seeds based on polarized hyperspectral imaging combined with ensemble learning algorithm. *Agriculture.* 2023;13(8):1499.

56. Pu H, Liu D, Wang L, et al. Soluble solids content and pH prediction and maturity discrimination of lychee fruits using visible and near infrared hyperspectral imaging. *Food Anal Methods*. 2016;9:235–44.
57. Sun Y, Wang Y, Xiao H, et al. Hyperspectral imaging detection of decayed honey peaches based on their chlorophyll content. *Food Chem*. 2017;235:194–202.
58. Salas EAL, Henebry GM. Separability of maize and soybean in the spectral regions of chlorophyll and carotenoids using the Moment Distance Index. *Israel J Plant Sci*. 2012;60(1–2):65–76.
59. Fränti P, Mariescu-Istodor R. Soft precision and recall. *Pattern Recogn Lett*. 2023;167:115–21.
60. Bagui S, Li K. Resampling imbalanced data for network intrusion detection datasets. *J Big Data*. 2021;8(1):6.
61. Kim S, Lim UT. Seasonal occurrence pattern and within-plant egg distribution of bean bug, *Riptortus pedestris* (Fabricius)(Hemiptera: Alydidae), and its egg parasitoids in soybean fields. *Appl Entomol Zool*. 2010;45(3):457–64.
62. Li J, Li Q, Yu C, et al. A model for identifying soybean growth periods based on multi-source sensors and improved convolutional neural network. *Agronomy*. 2022;12(12):2991.
63. Vanoli M, Rizzolo A, Grassi M, et al. Studies on classification models to discriminate 'Braeburn' apples affected by internal browning using the optical properties measured by time-resolved reflectance spectroscopy. *Postharvest Biol Technol*. 2014;91:112–21.
64. da Rocha F, Vieira CC, Ferreira MC, et al. Selection of soybean lines exhibiting resistance to stink bug complex in distinct environments. *Food Energy Secur*. 2015;4(2):133–43.
65. Mortensen AK, Gislum R, Jørgensen JR, et al. The use of multispectral imaging and single seed and bulk near-infrared spectroscopy to characterize seed covering structures: methods and applications in seed testing and research. *Agriculture*. 2021;11(4):301.
66. Wang Y, Song S. Variety identification of sweet maize seeds based on hyperspectral imaging combined with deep learning. *Infrared Phys Technol*. 2023;130: 104611.
67. França-Silva F, Gomes-Junior FG, Rego CHQ, et al. Advances in imaging technologies for soybean seed analysis. *J Seed Sci*. 2023;45: e202345022.
68. Ge H, Lv M, Lu X, et al. Applications of THz spectral imaging in the detection of agricultural products photonics. *MDPI*. 2021;8(11):518.

## Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.