Hyperspectral imaging with machine learning for non-destructive classification of *Astragalus membranaceus* var. *mongholicus*, *Astragalus membranaceus*, and similar seeds

Yanan Xu[1], Weifeng Wu[1], Yi Chen[1], Tingting Zhang[2], Keling Tu[1], Yun Hao[1], Hailu Cao[3], Xuehui Dong[1]* and Qun Sun[1]*

[1]College of Agronomy and Biotechnology, Department of Plant Genetics & Breeding and Seed Science/Chinese Medicinal Herbs Research Center, China Agricultural University/The Innovation Center (Beijing) of Crop Seeds whole-process Technology Research, Ministry of Agriculture and Rural Affairs/Beijing Key Laboratory of Crop Genetic Improvement, Beijing, China, [2]Key Laboratory of Ministry of Education for Genetics, Breeding and Multiple Utilization of Crops, College of Agriculture, Fujian Agriculture and Forestry University, Fuzhou, China, [3]Hengde Materia Medica (Beijing) Agricultural Technology Co., Ltd., Beijing, China

The roots of *Astragalus membranaceus* var. *mongholicus* (AMM) and A. membranaceus (AM) are widely used in traditional Chinese medicine. Although AMM has higher yields and accounts for a larger market share, its cultivation is fraught with challenges, including mixed germplasm resources and widespread adulteration of commercial seeds. Current methods for distinguishing Astragalus seeds from similar (SM) seeds are time-consuming, laborious, and destructive. To establish a non-destructive method, AMM, AM, and SM seeds were collected from various production areas. Machine vision and hyperspectral imaging (HSI) were used to collect morphological data and spectral data of each seed batch, which was used to establish discriminant models through various algorithms. Several preprocessing methods based on hyperspectral data were compared, including multiplicative scatter correction (MSC), standard normal variable (SNV), and first derivative (FD). Then selection methods for identifying informative features in the above data were compared, including successive projections algorithm (SPA), uninformative variable elimination (UVE), and competitive adaptive reweighted sampling (CARS). The results showed that support vector machine (SVM) modeling of machine vision data could distinguish Astragalus seeds from SM with >99% accuracy, but could not satisfactorily distinguish AMM seeds from AM. The FD-UVE-SVM model based on hyperspectral data reached 100.0% accuracy in the validation set. Another 90 seeds were tested, and the recognition accuracy was 100.0%, supporting the stability of the model. In summary, HSI data can be applied to

discriminate among the seeds of AMM, AM, and SM non-destructively and with high accuracy, which can drive standardization in the Astragalus production industry.

1 Introduction

Astragalus commonly refers to *Astragalus membranaceus* var. *mongholicus* (AMM) or *A. membranaceus* (AM), a legume which is harvested for the medicinal properties of its roots, which are purported to confer anti-inflammatory, anti-oxidative, and anti-cancer effects (Bai et al., 2018; Chen et al., 2020; Zhang et al., 2021). Notably, Astragalus is among the 40 most commonly used traditional Chinese medicines, and its demand is continuously increasing.

However, the seed quality of traditional Chinese medicines is often markedly lower than that of other crops, with problems including seeds of mixed and unknown origins, contamination with debris and foreign matter, inconsistent maturity, low germination rates, and slow or irregular emergence. The adulteration of traditional Chinese medicines is common due to the high value and high demand for these products, which has aroused widespread concerns for public safety (Reid et al., 2006; Yang et al., 2021b). Because it is primarily propagated by seed, Astragalus production faces particular challenges, such as mixed germplasm resources, uneven seed quality, and substantial adulteration of commercial seeds. These cumulative factors adversely affect Astragalus cultivation, and present the challenge of distinguishing between Astragalus and similar (SM) seeds.

AMM and AM are both included in the 2020 edition of the Chinese Pharmacopoeia, and although *A. complanatus*, *Melilotus officinalis*, *A. sinicus*, and *Hedysarum polybotrys* are also considered authentic medicines, they are common SM seeds found as contaminants in AM or AMM seed lots. In addition to contamination with SM seeds, AM and AMM seeds are often mixed, which is problematic due to differences in growth habit,

Abbreviations: AMM, *Astragalus membranaceus* var. *mongholicus*; AM, *A. membranaceus*; SM, Similar; SVM, Support vector machine; HSI, Hyperspectral imaging; MSC, Multiplicative scatter correction; SNV, Standard normal variable; FD, First derivative; SPA, Successive projections algorithm; UVE, Uninformative variable elimination; CARS, Competitive adaptive reweighted sampling; PLS-DA, Partial least squares discriminant analysis; MLP, Multilayer perceptron; VIS/NIR, Visible/near-infrared; ROI, region of interest; EWs, Effective Wavelengths; RMSE, The root means square error; LVs, latent variables.

planting adaptability, cultivation and management techniques, chemical composition, and commercial value. AMM plants are morphologically shorter than AM, but the roots are typical taproot, with characteristically few root branches and high yield (since the root is the harvested portion). By contrast, AM plants are taller, with shorter roots that exhibit higher branch number, and lower yield (Wang and Liu, 1996; Zhang et al., 2009; Yang et al., 2020). Therefore, AMM is the most widely cultivated variety, and mixing AMM and AM seeds will lead to different heights during planting, which limits the effectiveness of AMM management practices and decreases yield. This ongoing problem in AMM cultivations points to a need for a system that can accurately sort AMM and AM seeds.

Conventional methods of distinguishing between AMM, AM, and SM seeds include observations with an electron microscope, physical and chemical methods, ultraviolet spectroscopy, and molecular labeling (Yan et al., 2001; Yan et al., 2005; Wang et al., 2005; Duan et al., 2012; Zheng et al., 2019). However, these methods generally produce qualitative results that depend heavily on experience, especially visual evaluation, and a non-destructive, accurate, and simple method for discriminating among AMM, AM and SM seeds is urgently needed by Astragalus producers and market regulators alike.

Machine vision technology combines computational analysis with image recognition and processing technologies (Savakar and Anami, 2009; Patel et al., 2012). Machine vision with image processing (typically RGB images) is currently widely used in agriculture, while research on seed classification is also developing (Cheng et al., 2010; Huang and Cheng, 2017; Tu et al., 2021). Although machine vision can identify morphological and textural variation well, it does not provide any non-visual trait data, such as internal composition, which has limited the application of this approach for distinguishing seeds from genetically and phenotypically similar varieties. This problem has been overcome through hyperspectral imaging (HSI), which simultaneously integrates spatial data with spectral information to highlight differences in chemical composition that affect light reflection or transmittance through the sample, providing additional analytical layers for each sample. In addition, HSI provides high spatial resolution, generating continuous and narrow-band spectral information

for a given object (Guo et al., 2017; Paoletti et al., 2019). HSI techniques are increasingly tested for application in the identification of crops such as staple grains, fruits, and vegetables (Choudhary et al., 2009; Nansen et al., 2015; Sun et al., 2016; Wang et al., 2016; Wang et al., 2018; Xiao et al., 2020; Liu et al., 2022), supporting its feasibility for distinguishing different seed types.

It should be noted that analyzing the large datasets obtained by machine vision or HSI have presented a non-trivial challenge for data scientists and researchers. Machine learning methods have been developed that are currently the most efficient approaches for image processing and analysis. Common machine learning algorithms, including support vector machine (SVM), partial least squares discriminant analysis (PLS-DA), and multilayer perceptron (MLP), have been successfully applied to a range of classification tasks (Yang et al., 2015a; Sun et al., 2016; Sun et al., 2021; Nazari et al., 2021).

Machine vision and HSI approaches have been combined with machine learning algorithms to classify different crop seeds (Hong et al., 2016; Sun et al., 2017a; Zhao et al., 2018; Nie et al., 2019; Nazari et al., 2021; Ruslan et al., 2022; Tu et al., 2022). Table 1 shows details of these classification tasks. For Astragalus, Xiao et al. (2020) used visible/short-wave near-infrared and near infrared hyperspectral imaging with a convolutional neural network to identify Radix Astragali from five geographical origins, showing an accuracy of >98%. Despite the success of machine vision and HSI technologies in different crop seeds, to our knowledge, no studies have yet reported the application of machine vision and HSI technologies in AMM, AM and SM seed identification.

In this study, machine vision and HSI techniques were applied to the identification of AMM, AM and SM seeds of different origins. The specific objectives of this work include: (1) to establish detection models for AMM, AM and SM seeds using machine vision or HSI image data combined with a machine learning algorithm; (2) to determine the optimal classification model based on the predictive accuracy of different models with the validation dataset; (3) to compare the detection accuracy of the strongest machine vision- or HSI-based models for sample seeds not included in the training data; (4) to compare machine vision and HSI methods to determine which imaging method is better suited for AMM, AM and SM seed identification.

2 Materials and methods

2.1 Materials

The seed samples used in this study were divided into "Astragalus seeds" and "SM seeds". Astragalus seeds include AMM and AM seeds, while SM includes *Astragalus complanatus*, *Melilotus officinalis*, *A. sinicus*, and *Hedysarum polybotrys* seeds. After the seeds were collected, they were sealed

and fumigated with aluminum phosphide for 4-5 days, placed in a ventilated place for 7-10 days, bagged, and stored in a ventilated place at room temperature. All samples were identified by field planting and their known origin. The source and quantity of seed samples collected in this study are shown in Table 2.

2.2 RGB image acquisition and feature extraction

Several AMM, AM, and SM seeds were randomly selected and scanned with a ScanMaker i360/i460 scanner (Shanghai, China) at a resolution of 600dpi. The images were saved in TIFF lossless format. Hundreds of seeds were scanned each time, though there was no contact between seeds.

The Phenoseed automatic extraction system, which was jointly developed by the Seed Science and Technology Research Center of China Agricultural University (Beijing, China) and Nanjing Zhinong Yunxin Big Data Technology Co., Ltd. (Nanjing, China), was used to extract phenotypic features of the seeds. Shape features included length (mm), width (mm), L/W ratio, area (mm2), perimeter (mm), and roundness (mm). Color features included R (red in the primary color light spectrum), G (green in the primary color light spectrum), B (blue in the primary color light spectrum), L (luminosity), a (range from red to green), b (range from blue to yellow), hue, saturation, value, gray, and standard deviation. Texture features included the average value and standard deviation of contrast, dissimilarity, homogeneity, energy, correlation, ASM, and entropy under gray, R, G, and B. There were a total of 54 phenotypic features.

2.3 Hyperspectral reflectance data extraction

2.3.1 HSI system and analysis software

Hyperspectral reflectance images of seeds were collected using a prototype visible/near-infrared (VIS/NIR) HSI system installed at the Beijing Key Laboratory of Crop Genetic Improvement, China Agricultural University. The spectral range of the system was 311- 1090 nm, the bandwidth was 0.78 nm, and the image resolution was 1004 × 1002 pixels. The software for collecting spectral information was Spectral Image-VNIR (Isuzu Optics Corp., Taiwan, China). Other detailed descriptions of the VIS/NIR HSI system are listed in a paper published by Zhang et al. (2020).

2.3.2 Hyperspectral image acquisition and data extraction

Before collecting hyperspectral images, we calibrated the black-and-white board to produce the corrected image and then

TABLE 1 Application of machine vision and HSI for different crop seed classification tasks.

Method	Seed	Varieties	Classifier(s)	Result	Reference
Machine vision	Sorghum	10 sorghum cultivars	MLP, MLT	99%	Nazari et al., 2021
	Weedy rice	5 cultivated rice varieties and a weedy rice seed	LR	92.40%	Ruslan et al., 2022
	Rice	6 common cultivated rice seed varieties	KNN, SVM, RF	90.54%	Hong et al., 2016
HSI	Grape	3 grape seed varieties	SVM	94.30%	Zhao et al., 2018
	Maize	Jingke 968 and Non-Jingke 968	RF, SVM, MLP	~99%	Tu et al., 2022
	Hybrid okra seeds and hybrid loofah seeds	6 hybrid okra seed varieties and 6 hybrid loofah seed varieties	PLS-DA, SVM, DCNN	~95%	Nie et al., 2019
	Rice	Rice seeds from 4 different regions	SVM	91.67%	Sun et al., 2017a

set the moving speed of the electric control platform to 1.7 mm/s and the exposure time of the camera to 6 ms. When collecting spectral information from the sample, we placed each seed on the electric control displacement table. The camera scanned the whole platform as it moved. We then collected 50 seeds from each batch, for a total of 1150 seeds. The hyperspectral image acquisition process was completed in the dark box.

Before extracting spectral information, an HSI analyzer (Isuzu Optics Corp., Taiwan, China) was used to correct the spectral image, as shown in equation (1), where I is the corrected hyperspectral image. I_0 is the original hyperspectral image; B is the blackboard file image (reflectivity close to 0%), and W is the whiteboard file image (reflectivity close to 100%).

$$I = \frac{I_0 - B}{W - B} \tag{1}$$

After the black-and-white plate was corrected, the single seed was separated from the background of the hyperspectral image. The background and seeds were separated by setting a threshold and obtaining the region of interest (ROI) by

TABLE 2 Table of materials.

Category	Species	Identifier	Source	Genotypes	Number of seeds	
					Machine vision	Hyperspectral
Astragalus seeds	*A. membranaceus* var. *mongholicus* (AMM)	AMM-1	Tangshan, Hebei	Cultivated	234	50
		AMM-2	Datong, Shanxi	Cultivated	249	50
		AMM-3	Chifeng, Inner Mongolia	Cultivated	199	50
		AMM-4	Yulin, Shaanxi	Cultivated	215	50
		AMM-5	Guyuan, Ningxia	Cultivated	184	50
		AMM-6	Shangzhuang, Beijing	Cultivated	361	50
		AMM-7	Anguo, Hebei	Cultivated	85	50
		AMM-8	Market purchase	Cultivated	186	50
		AMM-9	Longnan, Gansu	Wild	234	50
	A. membranaceus (AM)	AM-1	Tangshan, Hebei	Cultivated	290	50
		AM-2	Zhangjiakou, Hebei	Cultivated	256	50
		AM-3	Hezheng, Gansu	Cultivated	237	50
		AM-4	Guyuan, Ningxia	Cultivated	197	50
		AM-5	Shangzhuang, Beijing	Cultivated	239	50
		AM-6	Longnan, Gansu	Wild	202	50
		AM-7	Zhangjiakou, Hebei	Wild	240	50
Similar (SM) seeds	*A. complanatus*	AC-1	Anguo, Hebei	Cultivated	427	50
		AC-2	Bozhou, Anhui	Cultivated	325	50
		AC-3	Suqian, Jiangsu	Cultivated	407	50
		AC-4	Laboratory samples	/	352	50
	M. officinalis	MO-1	Laboratory samples	/	132	50
	A. sinicus	AS-1	Laboratory samples	/	214	50
	H. polybotrys	HP-1	Laboratory samples	/	145	50

morphological filtering and mask processing. Due to the large noise interference in the head and tail bands, only the reflection spectra of 765 bands within 400-1000 nm of each seed were extracted for subsequent modeling and analysis. Among them, the VIS included 490 reflectance data points in the range of 400-780 nm, and the NIR included 275 reflectance data points in the range of 780-1000 nm.

2.4 Spectra preprocessing

Spectral information is inevitably affected by various factors, and spectral preprocessing is used to improve the usefulness of spectral data (Esquerre et al., 2012). This study compared three spectral preprocessing methods, including multiplicative scatter correction (MSC), standard normal variable (SNV), and first derivative (FD). From the latter, MSC and SNV were used to consider the addition/multiplication effect and scattering effect in spectral data (Silalahi et al., 2018; Wu et al., 2019). FD helps delete baseline offset (Qu and Liu, 2017). These preprocessing methods eliminate the external interference generated during the acquisition of hyperspectral images. In this study, a suitable preprocessing method was selected based on the preprocessing effect of various preprocessing methods on the raw spectral data.

2.5 Selection of effective wavelengths

The presence of high dimensional data and a large amount of redundant information in hyperspectral images can affect the modelling speed. Therefore, it is important to use the variable selection method during the analysis and processing of hyperspectral data. In this study, successive projection algorithms (SPA), uninformative variable elimination (UVE), and competitive adaptive reweighted sampling (CARS) were used to select EWs.

SPA can extract low collinearity and low redundancy variables to avoid the influence of information overlap and collinearity. When the SPA method is used to optimize the wave band, multiple linear regression models can be established one by one for different wave segment subsets, and the root means square error (RMSE) value can be calculated, in which the number of variables corresponding to the lowest RMSE is the optimal EWs (Galvão et al., 2008; Zhang et al., 2018b). UVE can remove the wavelength variables that contribute less to the modeling and select the characteristic wavelength variables. The removed wavelength variables are called non-information variables. UVE and CARS establishment are based on the PLS algorithm. To select non-informative variables, the UVE algorithm adds a group of white noise variables with the same number of original variables to the PLS model and obtains the regression coefficient corresponding to each variable based on the cross-leave method of the PLS model. The stable value of

each variable coefficient is divided by the standard deviation, their quotient is compared with the stable value obtained from the random variable matrix, and those wavelength variables that are invalid for modeling are deleted as random variables. (Put et al., 2006; Wang et al., 2020) In CARS algorithm, adaptive weighted sampling (ARS) is used to retain the points with large absolute values of regression coefficients in the PLS model as a new subset each time, and remove the points with small weights. Then, the PLS model is established based on the new subset. After multiple calculations, the subset with the smallest cross-validation root mean square error (RMSECV) of the PLS model is selected as the characteristic wavelength. (Zhang et al., 2018a; Zhou et al., 2020a)

2.6 Data–driven modeling

Support vector machine (SVM), partial least squares discriminant analysis (PLS-DA), and multilayer perceptron (MLP) are widely used classification methods that have been well validated for seed detection (Kujawa et al., 2014; Nazari et al., 2021; Yang et al., 2021a). In this study, three classification models of AMM, AM, and SM seeds were established using SVM, PLS-DA, and MLP, respectively. SVM is widely used to solve linearly differentiable and linearly indistinguishable classification problems, and the radial basis kernel function (RBF) kernel is the most common and effective method for classification problems. Optimization of the hyperparameters is necessary before the real model can be trained. In current practical applications, hyperparameters are usually determined empirically or by grid search. (Liu et al., 2022) PLS-DA is a typical classification method, which is considered as a supervision method to distinguish samples to the maximum extent. (Nie et al., 2019; Zhang et al., 2022b) MLP is a feedforward neural network. It maps a set of input vectors to a set of output vectors. The inputs and outputs can be connected by multilayer weighting, with strong self-learning, adaptive, associative memory and parallel processing of things and environments. (Xu et al., 2021)

To avoid the effect of default parameters on the prediction accuracy of the classification model, the internal parameters of the classification model must be separately adjusted. In the SVM algorithm, the RBF kernel was selected, and it carried out the 5-fold internal cross-validation and grid search method to calculate optimal penalty coefficient c and the kernel parameter g. The searching range was both set to -10 to 10 with the step of 0.2 (a total of 101*101 combinations were used to search the best parameters). In the PLS-DA model, the number of latent variables (LVs) changes, and the model correctly identifies the highest percentage of seeds. The MLP network with two hidden layers was selected, and the hidden layer adopted the hyperbolic tangent activation function of SPSS. The output layer adopted the Softmax activation function.

2.7 Analyzing

MSC, SNV, FD spectral preprocessing, SPA, UVE, CARS feature band extraction, and SVM and PLS-DA model driving processes were implemented in Matlab R2020b. The MLP modeling process was using IBM SPSS statistics 26. For each model, the ratio of three model training sets and validation sets was 7:3. The specific quantity of seeds in each batch is shown in Table 2. Origin 2022 and R 4.1.2 were used to visualize data. The experimental flow of this study is shown in Figure 1.

3 Results

3.1 Identification of AMM, AM, and SM seeds based on machine vision

In order to develop a machine learning algorithm capable of distinguishing AMM, AM, and other SM seeds, we first tested untrained machine vision using a mixed set of 5610 seeds. The machine vision technology based on Phenoseed automatic extraction system extracted a set of 54 potentially informative morphological features for discriminating among seed types, including shape, color, and texture. A probability density map was then generated to examine the distributions of these phenotypic features across the 5610 combined AMM, AM and SM seeds (Figure 2), which showed high overlap in their features related to color and texture, especially between AMM and AM. These results indicated that these features might not be sufficiently different between seed types to form a basis for distinguishing between them. It is worth noting that the size of SM was generally smaller than that of AMM or AM seeds, although the size distributions showed substantial overlap (Figure 2). Thus, SM resembled small AMM and AM seeds, which is germane to distinguishing SM seeds mixed into Astragalus seed lots in the market.
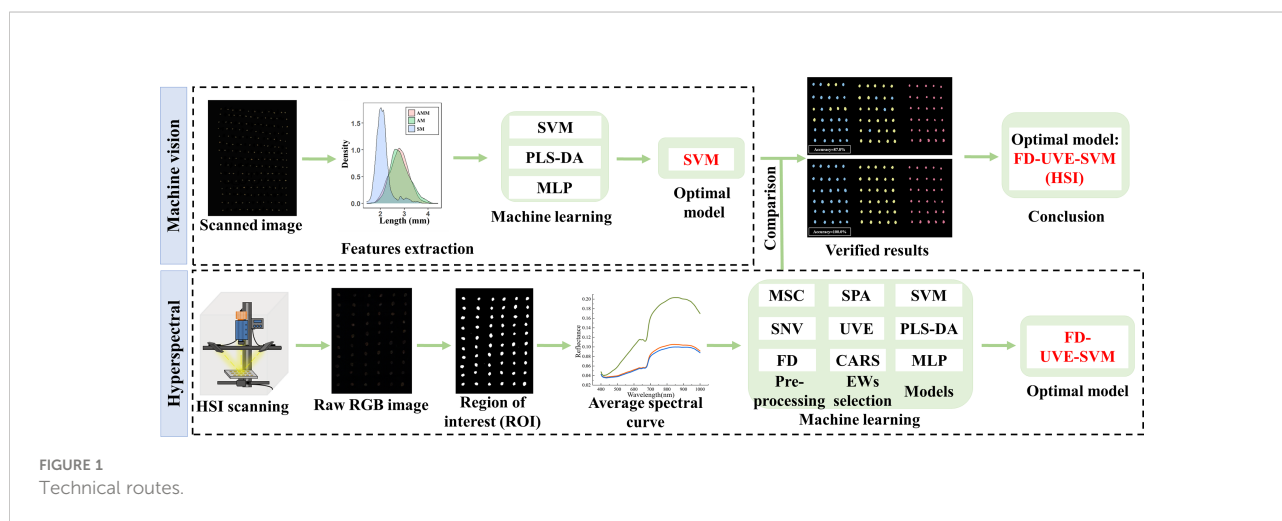
In order to establish a model for classifying AMM, AM, and SM seeds, the 54 morphological features detected by machine vision were used as inputs for SVM, PLS-DA, and MLP networks, with 3927 seeds in the training set and 1683 seeds in the validation set. The accuracy in distinguishing between seeds was then calculated for each model (Figure 3), with 48 latent variables (LVs) selected for the PLS-DA model. Classification accuracy of SM seeds using these features reached >98.2%, but the classification effect of AMM and AM seeds were not ideal. In particular, SVM showed 83.3% accuracy for identifying AMM seeds, 89.2% accuracy for detecting AM seeds, and 99.5% accuracy for SM seeds, resulting in 91.1% average accuracy for this model.
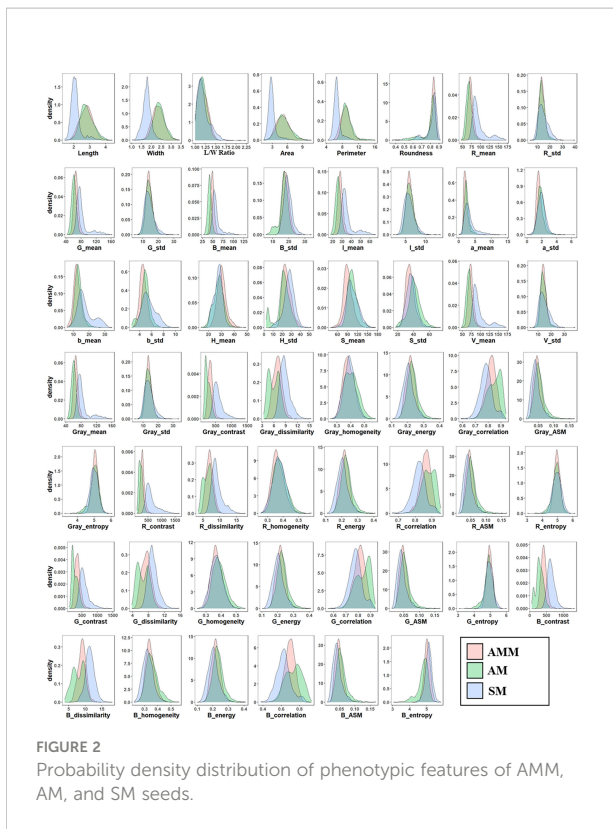
3.2 Identification of AMM, AM, and SM seeds using HSI

Since machine vision with different machine learning algorithms could not effectively distinguish AMM seeds from AM seeds, we next explored means of improving accuracy by VIS/NIR hyperspectral imaging of reflectance spectra between 400 nm and 1000 nm bands for each seed type.

3.2.1 Spectral characteristics

To identify wavelengths that were distinct among seed types, the reflectance spectra were obtained for 1150 seeds (Figure 4A) and the average spectrum of AMM, AM, and SM seeds were calculated (Figure 4B), which revealed that the average spectra of AMM and AM seeds was significantly lower than that of SM seeds. By contrast, AMM and AM seeds displayed similar spectra, which agreed well with previous studies examining the variety and viability of seeds from other crops (Yang et al., 2017; Wakholi et al., 2018; Yang et al., 2021a). However, the reflectance of AMM was generally higher than that of AM seeds, with greater differences in the NIR region (780-1000
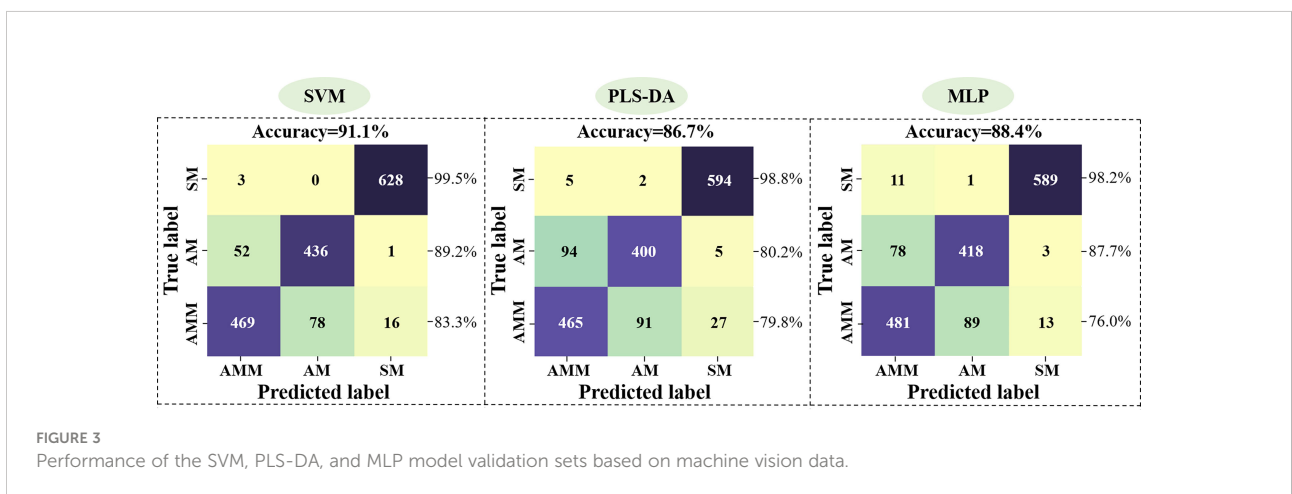

FIGURE 1
Technical routes.

FIGURE 2
Probability density distribution of phenotypic features of AMM, AM, and SM seeds.

nm) than in the VIS region (400-780 nm). The VIS region may be related to β-carotene and anthocyanin in the seeds (Sun et al., 2021), and the difference in the NIR region may be related to protein, starch and other organic matter in the seeds (Cen and He, 2007; Awanthi et al., 2019). These results indicated that HSI could capture differences in texture, pigment and other physical and chemical properties between AMM and AM seeds. And the differences between AMM and AM seeds in terms of organic matter such as protein and starch are greater than the differences in pigment content, which explains the inability of models based on VIS data (machine vision) to correctly distinguish between AMM and AM. Further analysis showed that in the range of 660-750 nm, the spectral curve showed an obvious upward trend, and the average spectral curve of the two seeds gradually separated. The reason is that this wavelength corresponds to the vibration of N-H chemical bond of amino acid in seeds, which can be used to verify the difference of amino acid content in AMM and AM seeds. (Yang et al., 2021a; Wang et al., 2022) In addition, there are four absorption peaks in the average reflectance spectra of AMM and AM seeds (valleys at 415 nm, 640, 680 and 885 nm). The carotenoid (Yang et al., 2021a) and proanthocyanidin content of the seed coat (Wang et al., 2022) can be determined at about 415 nm; The bands at about 640 nm and 680 nm may be associated with the absorption of chlorophyll b and chlorophyll a (Zhang et al., 2016). The band at about 885 nm may be associated with C-O, N-H, C-H and O-H bonding vibrations in proteins, carbohydrates and fats (Caporaso et al., 2018). However, the spectral curves were not unique enough to distinguish AMM, AM, and SM seeds. More specifically, while a large proportion of SM overlapped with Astragalus seeds in the original spectral curve, the reflectance spectra of AMM and AM showed extremely high overlap, making them indistinguishable by curve shape. To identify specific spectral features or effective wavelengths that could be used for classification of AMM, AM, or SM seeds, it was first necessary to generate discriminant analytical models to test different methods of preprocessing to maximize the accuracy of discriminating among AMM, AM and SM seeds.

3.2.2 Spectral preprocessing and effect analysis

In order to establish the most effective discriminant model, we first tested three methods for preprocessing the seed reflectance spectra, including MSC, SNV and FD, as well as raw data. The SVM, PLS-DA, and MLP methods were each used to generate models with the processed and raw data, using a training set of 805 seeds and a validation set of 345 seeds (i.e., a 7:3 ratio) (see
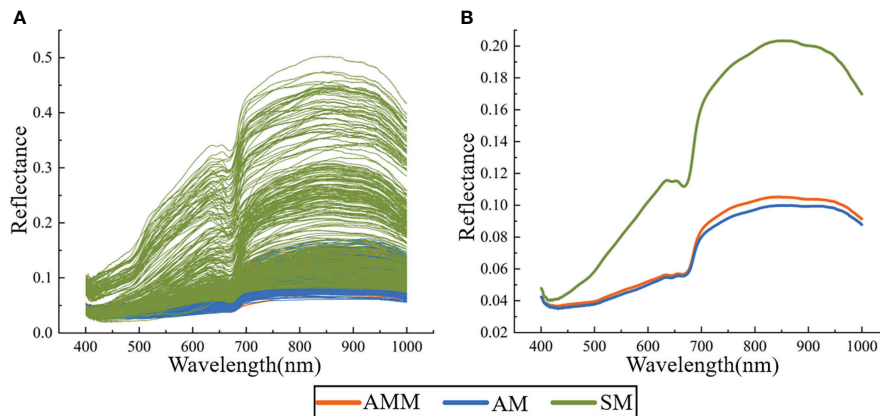

FIGURE 3
Performance of the SVM, PLS-DA, and MLP model validation sets based on machine vision data.

FIGURE 4
Reflection spectrum **(A)** and average spectrum **(B)** of AMM, AM, and SM seeds.

section 2.6 for parameter details). The PLS-DA model selected 10, 10, 10, and 9 LVs from raw spectra or spectra preprocessed by MSC, SNV, FD, respectively. Validation of each classifier model with each respective preprocessing method showed that accuracy ranged from 81.7%, in the PLS-DA model with raw data, to 100.0% in the SVM model with the FD-processed spectra (Figure 5). Among the three pre-processing methods, the model built after FD pre-processing was the optimal, while MSC and SNV showed suboptimal accuracy compared to FD, but still better than RAW. Among the three classifiers, the performance of the model showed SVM > MLP > PLS-DA, regardless of the pre-processing based. While accuracy differed to a limited extent

between each combination of classifier and data processing method, it warrants mention that all preprocessing methods increased the accuracy over that of raw data input for all classifiers by denoising the original reflectance spectra. Since the FD-SVM denoising/classifier combination provided the highest accuracy in discriminating among AMM, AM, and SM seeds, this method was used in subsequent analyses of effective wavelengths HSI reflectance data.

3.2.3 Selection of effective wavelengths

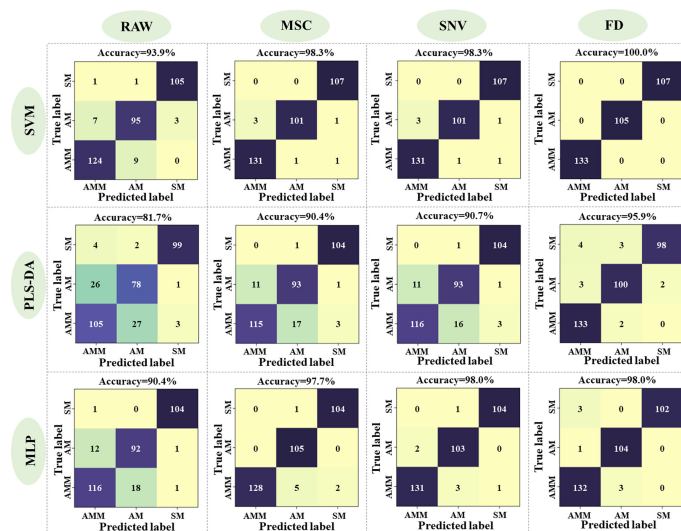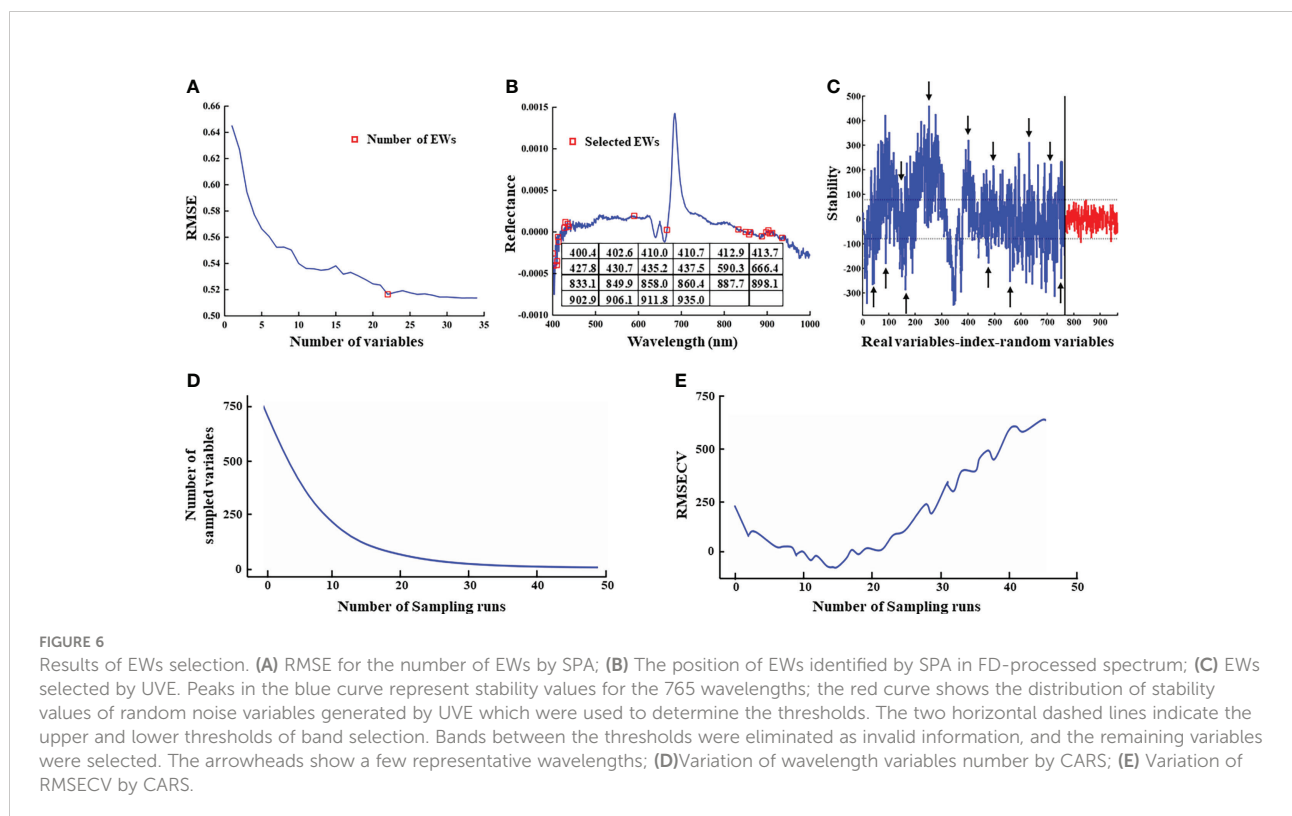Based on the above results showing highly accurate classification of AMM, AM, and SM seeds obtained by full

FIGURE 5
Accuracy of SVM, PLS-DA, and MLP models with spectral data that was preprocessed or not.

spectra models, we next sought to increase computational efficiency and reduce processing time by identifying effective wavelengths in FD-processed data that were informative for classifying seed types.

Since the algorithmic principles underlying wavelength selection can impact modeling results (Zhang et al., 2020), we tested three approaches to EW extraction, SPA, UVE, and CARS. In SPA extraction, the root means square error (RMSE) values decreased with increasing number of variables (i.e., EWs). The lowest RMSE value coincided with 22 variables, above which RMSE remained stable (Figure 6A). Therefore, an RMSE threshold of 0.5164 with 22 effective wavelengths in the denoised reflectance spectrum was selected for subsequent tests (Figure 6B). By contrast, UVE analysis identified 391 potentially informative EW features (Figure 6C) for classifier analysis. In CARS extraction, the number of EWs decreased rapidly with the exponential decay function, but decreased at a slower rate as the number of samples increases (Figure 6D). The cross-validation root mean square error (RMSECV) tended to decrease and then increase as the number of samples increased, with the smallest RMSECV value when the number of samples reached 14 (Figure 6E) (i.e., the subset of EW selected for this sampling was the key variable for predicting AMM, AM and SM seeds). Therefore, 140 potentially informative EW features (Figure 6D) were identified for classifier analysis after CARS.

The EWs selected by SPA were further examined to better understand their relevance as a theoretical basis for distinguishing AMM, AM and SM seeds. Several EWs were located around specific regions, which suggested that seed reflectance was related to the presence of specific chemicals. Some specific wavelengths in the visible light spectrum are reportedly related to plant pigments, such as absorption peaks for carotenoids at 412.9 and 413.7 nm, chlorophyll a at 427.8, 430.7, and 666.4nm, and anthocyanin at around 435.2 and 437.5 nm (Li et al., 2013; Nansen et al., 2015; Yang et al., 2015b; Zhang et al., 2016). Among the EWs screened by SPA, the selection of bands near pigment absorption peaks indicated that accurate classification of AMM, AM, and SM seeds may rely on seed color. By contrast, several EWs in the NIR were attributable to various chemical bonds. For example, EWs located near 833.1 nm were related to the vibrations of C-O, N-H, C-H, and O-H bonds in proteins, carbohydrates, and fats, respectively (Cen and He, 2007; Sun et al., 2017b; Caporaso et al., 2018). Tannin absorption peaks were also detected near 887.7 and 898.1 nm, and cellulose absorption peaks were present near 935.0 nm (Wang et al., 2022). In addition, peaks between 860-970 nm reflected the vibration of N-H bonds in proteins and amino acids, and suggested that these seeds differed in protein or amino acid contents (Zhou et al., 2020b). Since tannin, cellulose, and protein contents in the seed coat affect the thickness and hardness of the seed coat, these results indicated that AM, AMM, and SM seeds differed in their seedcoat hardness.

FIGURE 6
Results of EWs selection. **(A)** RMSE for the number of EWs by SPA; **(B)** The position of EWs identified by SPA in FD-processed spectrum; **(C)** EWs selected by UVE. Peaks in the blue curve represent stability values for the 765 wavelengths; the red curve shows the distribution of stability values of random noise variables generated by UVE which were used to determine the thresholds. The two horizontal dashed lines indicate the upper and lower thresholds of band selection. Bands between the thresholds were eliminated as invalid information, and the remaining variables were selected. The arrowheads show a few representative wavelengths; **(D)**Variation of wavelength variables number by CARS; **(E)** Variation of RMSECV by CARS.

3.2.4 SVM modeling based on different EW selection methods

In light of our above findings, we next tested the accuracy of the SVM model in discriminating between AM, AMM, and SM seeds using FD-denoised EWs selected by SPA or UVE or CARS as inputs. Validation of model accuracy is shown in Figure 7. The changes in model performance indicate that UVE screening of EWs results in higher accuracy, with CARS showing the sub-optimal accuracy and SPA the worst performance. It potentially due to an insufficiently large sample set of EWs selected by SPA, leading to the exclusion of informative wavelengths related to seed classification. Our results thus showed that UVE method is more suitable than the SPA and CARS algorithm for building SVM models to classify AMM, AM, and SM seeds and that the FD-UVE-SVM combination provides an optimal model for this task, providing 100.0% accuracy in the validation set.

3.3 Visual comparison between machine vision and HSI prediction results
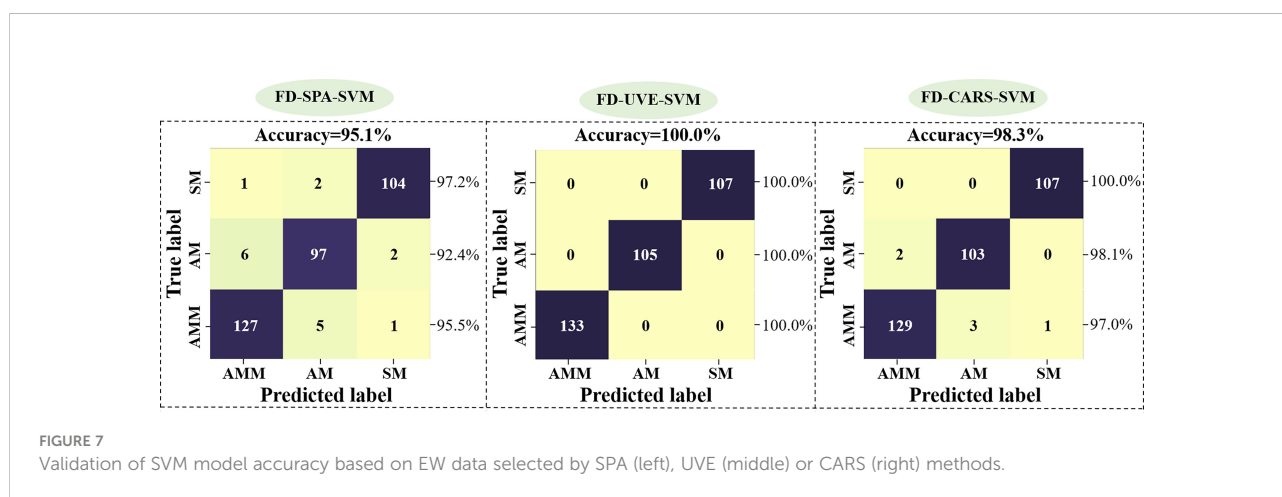
Based on the above results showing that SVM classification provided the highest accuracy with machine vision imaging data, while the FD-UVE-SVM model showed the highest accuracy with HSI-based data, we next visually examined a set of 90 seeds, including 30 AMM, 30 AM, and 30 SM, to manually verify the classifications made by these two models (Figure 8). The SVM model based on machine vision image data (Figure 8A) could correctly predict the 30 SM seeds, while 7 of the 30 AMM seeds were incorrectly predicted as AM seeds, and 4 of the 30 AM seeds were incorrectly predicted as AMM seeds (i.e., an overall accuracy of 87.8%), indicating sub-optimal discriminatory power between AM and AMM seeds. By contrast, FD-UVE-SVM modelling of HSI data correctly predicted the types of all 90 seeds (i.e., 100.0% accuracy), indicating that hyperspectral data contained sufficient information to accurately discriminate between AMM, AM, and

SM seeds, supporting its application in routine analysis required for seed sorting and quality control in production settings.

4 Discussion

In this study, HSI data outperformed machine vision image data in SVM-based models for distinguishing AMM, AM and SM seeds, which aligns well with studies comparing these approaches for classifying kernels of rice (Fabiyi et al., 2020) and maize (Tu et al., 2022), as well as comparing chlorophyll content in sorghum leaves (Zhang et al., 2022a). It is likely that machine vision data resulted in lower accuracy in classifying AMM and AM seeds because the features extracted by machine vision were purely morphological phenotypes. Since AMM and AM belong to the same species, they share highly similar seed morphology. Thus, other phenotypic data, such seed traits related to the internal accumulation or deposition of specific metabolites affecting light reflection or diffraction captured by HSI data can provide more informative features for discriminating among subspecies. Although both methods are non-destructive and high-throughput tests that enable screening of intact seeds, each technology is accompanied by advantages and disadvantages. The main advantages of machine vision over HSI are the relatively low instrument cost and faster image acquisition, which are linked to its main disadvantage of capturing strictly morphological information that cannot account for many internal, physiological seed traits. By contrast, HSI data includes hundreds or thousands of spectral bands, and therefore contains more information allowing more robust discrimination among samples. However, it should be pointed out that the high dimension of spectral data limits the calculation speed and processing time to some extent. Moreover, HSI equipment is relatively expensive and the operating cost is high.

In HSI data analysis, FD pre-processing achieved superior results to MSC and SNV, which aligns well with studies for classifying kernels of sugar beet (Yang et al., 2018) and the
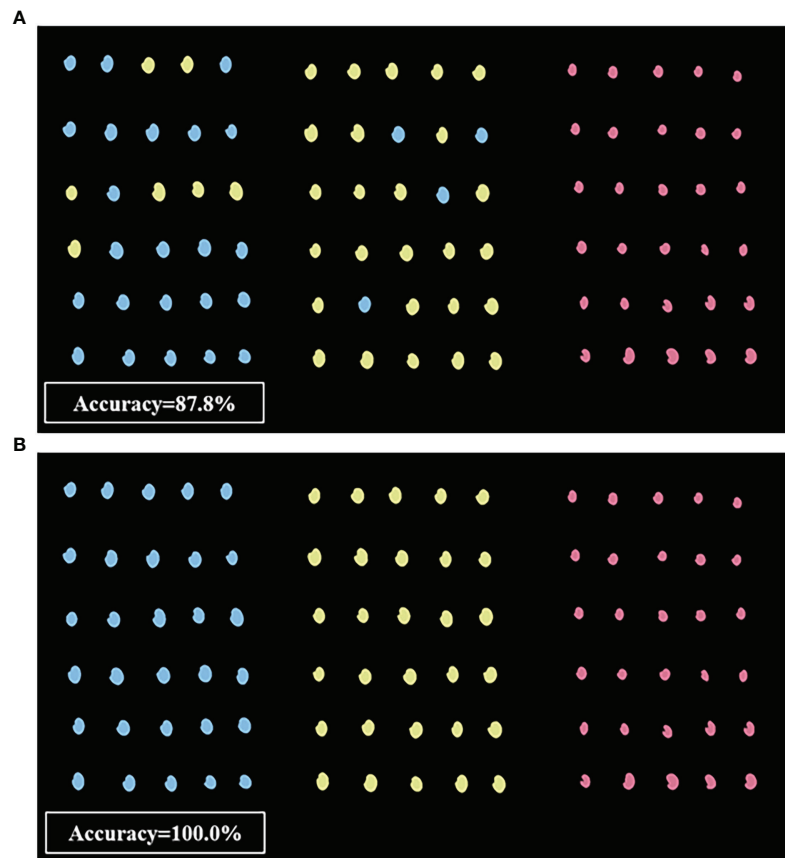

FIGURE 7
Validation of SVM model accuracy based on EW data selected by SPA (left), UVE (middle) or CARS (right) methods.

FIGURE 8
Visualization results based on **(A)** machine vision optimal model and **(B)** HSI optimal model (from left to right: AMM, AM, and SM seeds).

detection of germination rates in sorghum-sudan grass seed (Hui et al., 2022). It is likely that due to FD eliminates baseline drift in AMM, AM and SM seed spectral data and improves the spectral band characteristics and spectral resolution, thus FD pre-processing achieves the highest accuracy (Qu and Liu, 2017; Li et al., 2021). However, it is worth noting that spectral derivatives can also increase the noise level and reduce the spectral signal-to-noise ratio, which is detrimental to modeling. The higher the derivative order, the more serious the degradation of the signal-to-noise ratio. Therefore, in spectral analysis, only FD or second derivative (SD) is generally used for spectra.

Selecting a subset of EW features by SPA or UVE or CARS to construct the model can dramatically reduce processing time (Gao et al., 2013). SPA-based screening of denoised EWs captures a portion of bands in the NIR wavelength range (>780nm), whereas machine vision technology collects wavelengths largely in the VIS range. These additional NIR EWs may explain why HSI technology can provide higher accuracy than machine vision in distinguishing AMM, AM, and SM seeds.

Although HSI technology has enormous potential for discrimination among highly (visually) similar seeds, sample sizes in this study, including nine batches of AMM, seven batches of AM, and seven batches of SM, were not adequately large for robust statistical verification. Follow-up research will necessarily include larger sample sets improve the generalizability and accuracy of the classification model. In addition, in order to maintain model validity when testing seeds from different origins and growing seasons, the model can be updated using a method established for updating a maize seed detection model (Guo et al., 2017; Tu et al., 2022).

5 Conclusion

We compared machine vision and HSI image data to classify AMM, AM, and SM seeds, which led to the following specific conclusions:

1) SVM-based models of machine vision image data to distinguish AMM, AM, and SM seeds indicated that

AM/AMM seeds could be distinguished from SM seeds at >99.0%, but could not well-distinguish between AMM and AM seeds.

2) The application of a FD-UVE-SVM model to HSI data resulted in 100.0% accuracy, thus validating the SVM classification model as the best suited for distinguishing SM, AM, and AMM seeds.

3) Verification of model accuracy based on machine vision and HSI data from a 90-seed verification set indicated that predictive accuracy was 100.0% with HSI data, demonstrating the efficiency, reliability, and simplicity of this model, and importantly, revealing that HSI is more suitable for discriminating among AMM, AM, and SM seeds.

In general, this study used machine vision and HSI technology to classify AMM, AM, and SM seeds. FD-UVE-SVM modeling of HSI data can be used to accurately distinguish AMM, AM, and SM. This strategy can be adapted for routine analyses in production facilities. These advances can in turn increase the economic benefits of Astragalus seeds.

Data availability statement

The original contributions presented in the study are included in the article/Supplementary Material. Further inquiries can be directed to the corresponding authors.

Author contributions

QS, XD, and YX designed the experiments. YX collected the RGB, HSI images, analyzed the data, and wrote the manuscript. WW, YC, TZ, KT, YH, and HC provided comments and suggestions for the manuscript and checked the writing. All authors contributed to the article and approved the submitted version.

Funding

Conflict of interest

Author HC was employed by Hengde Materia Medica Beijing Agricultural Technology Co., Ltd.

The remaining 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.

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpls.2022.1031849/full#supplementary-material

References

Awanthi, M. G. G., Jinendra, B. M. S., Navaratne, S. B., and Navaratne, C. M. (2019). Adaptation of visible and short wave near infrared (VIS-SW-NIR) common PLS model for quantifying paddy hardness. *J. Cereal Sci.* 89, 102795. doi: 10.1016/j.jcs.2019.102795

Bai, C., Hao, J., Hao, X., and Feng, M. (2018). Preparation of astragalus membranaceus lectin and evaluation of its biological function. *BioMed. Rep.* 9, 345–349. doi: 10.3892/br.2018.1132

Caporaso, N., Whitworth, M. B., Grebby, S., and Fisk, I. D. (2018). Non-destructive analysis of sucrose, caffeine and trigonelline on single green coffee beans by hyperspectral imaging. *Food Res. Int.* 106, 193–203. doi: 10.1016/j.foodres.2017.12.031

Cen, H., and He, Y. (2007). Theory and application of near infrared reflectance spectroscopy in determination of food quality. *Trends Food Sci. Technol.* 18, 72–83. doi: 10.1016/j.tifs.2006.09.003

Cheng, C., Liu, J., Wang, H., and Xiong, W. (2010). Infrared spectroscopic studies of Chinese medicines. *Appl. Spectrosc Rev.* 45, 165–178. doi: 10.1080/05704920903574256

Chen, Z., Liu, L., Gao, C., Chen, W., Vong, C. T., Yao, P., et al. (2020). Astragali radix (Huangqi): A promising edible immunomodulatory herbal medicine. *J. Ethnopharmacol* 258, 112895. doi: 10.1016/j.jep.2020.112895

Choudhary, R., Mahesh, S., Paliwal, J., and Jayas, D. S. (2009). Identification of wheat classes using wavelet features from near infrared hyperspectral images of bulk samples. *Biosyst. Eng.* 102, 115–127. doi: 10.1016/j.biosystemseng.2008.09.028

Duan, L., Chen, T., Li, M., Chen, M., Zhou, Y., Cui, G., et al. (2012). Use of the metabolomics approach to characterize Chinese medicinal material huangqi. *Mol. Plant* 5, 376–386. doi: 10.1093/mp/ssr093

Esquerre, C., Gowen, A. A., Burger, J., Downey, G., and O'Donnell, C. P. (2012). Suppressing sample morphology effects in near infrared spectral imaging using

chemometric data pre-treatments. *Chemometr. Intell. Lab. Syst.* 117, 129–137. doi: 10.1016/j.chemolab.2012.02.006

Fabiyi, S. D., Vu, H., Tachtatzis, C., Murray, P., Harle, D., Dao, T. K., et al. (2020). Varietal classification of rice seeds using RGB and hyperspectral images. *IEEE Access* 8, 22493–22505. doi: 10.1109/ACCESS.2020.2969847

Galvão, R. K. H., Araújo, M. C. U., Fragoso, W. D., Silva, E. C., José, G. E., Soares, S. F. C., et al. (2008). A variable elimination method to improve the parsimony of MLR models using the successive projections algorithm. *Chemometr. Intell. Lab. Syst.* 92, 83–91. doi: 10.1016/j.chemolab.2007.12.004

Gao, J., Li, X., Zhu, F., and He, Y. (2013). Application of hyperspectral imaging technology to discriminate different geographical origins of jatropha curcas l. seeds. *Comput. Electron Agric.* 99, 186–193. doi: 10.1016/j.compag.2013.09.011

Guo, D., Zhu, Q., Huang, M., Guo, Y., and Qin, J. (2017). Model updating for the classification of different varieties of maize seeds from different years by hyperspectral imaging coupled with a pre-labeling method. *Comput. Electron Agric.* 142, 1–8. doi: 10.1016/j.compag.2017.08.015

Hong, P. T. T., Hai, T. T. T., Lan, L. T., Hoang, V. T., Hai, V., and Nguyen, T. T. (2016). "Comparative study on vision based rice seed varieties identification", in *2015 Seventh International Conference on Knowledge and Systems Engineering (KSE)*. 377–382. doi: 10.1109/KSE.2015.46

Huang, K., and Cheng, J. (2017). A novel auto-sorting system for Chinese cabbage seeds. *Sensors (Switzerland)* 17, 886. doi: 10.3390/s17040886

Hui, Y., Wang, D., Tang, X., Peng, Y., Wang, H., Zhang, H., et al. (2022). Detection of sorghum-Sudan grass seed germination rate based on near infrared spectroscopy. *Spectrosc. Spectral Anal.* 42, 423–427. doi: 10.3964/j.issn.1000-0593 (2022)02-0423-05

Kujawa, S., Nowakowski, K., Tomczak, R. J., Dach, J., Boniecki, P., Weres, J., et al. (2014). Neural image analysis for maturity classification of sewage sludge composted with maize straw. *Comput. Electron Agric.* 109, 302–310. doi: 10.1016/j.compag.2014.08.014

Li, J., Huang, W., Zhao, C., and Zhang, B. (2013). A comparative study for the quantitative determination of soluble solids content, pH and firmness of pears by Vis/NIR spectroscopy. *J. Food Eng.* 116, 324–332. doi: 10.1016/j.jfoodeng.2012.11.007

Li, X., Jiang, H., Jiang, X., Shi, M., and Fielke, J. M. (2021). Identification of geographical origin of Chinese chestnuts using hyperspectral imaging with 1D-CNN algorithm. *Agriculture* 11, 1274. doi: 10.3390/agriculture

Liu, Q., Wang, Z., Long, Y., Zhang, C., Fan, S., and Huang, W. (2022). Variety classification of coated maize seeds based on raman hyperspectral imaging. *Spectrochim Acta A Mol. Biomol Spectrosc* 270, 120772. doi: 10.1016/j.saa.2021.120772

Nansen, C., Zhao, G., Dakin, N., Zhao, C., and Turner, S. R. (2015). Using hyperspectral imaging to determine germination of native Australian plant seeds. *J. Photochem. Photobiol. B* 145, 19–24. doi: 10.1016/j.jphotobiol.2015.02.015

Nazari, L., Shaker, M., Karimi, A., and Ropelewska, E. (2021). Identification of sorghum genotypes using a machine vision system. *J. Food Process Eng.* 44, e13673. doi: 10.1111/jfpe.13673

Nie, P., Zhang, J., Feng, X., Yu, C., and He, Y. (2019). Classification of hybrid seeds using near-infrared hyperspectral imaging technology combined with deep learning. *Sens Actuators B Chem.* 296, 126630. doi: 10.1016/j.snb.2019.126630

Paoletti, M. E., Haut, J. M., Plaza, J., and Plaza, A. (2019). Deep learning classifiers for hyperspectral imaging: A review. *ISPRS J. Photogramm. Remote Sens.* 158, 279–317. doi: 10.1016/j.isprsjprs.2019.09.006

Patel, K. K., Kar, A., Jha, S. N., and Khan, M. A. (2012). Machine vision system: a tool for quality inspection of food and agricultural products. *J. Food Sci. Technol.* 49, 123–141. doi: 10.1007/s13197-011-0321-4

Put, R., Daszykowski, M., Baczek, T., and vander Heyden, Y. (2006). Retention prediction of peptides based on uninformative variable elimination by partial least squares. *J. Proteome Res.* 5, 1618–1625. doi: 10.1021/pr0600430

Qu, Y., and Liu, Z. (2017). Dimensionality reduction and derivative spectral feature optimization for hyperspectral target recognition. *Optik (Stuttg)* 130, 1349–1357. doi: 10.1016/j.ijleo.2016.11.143

Reid, L. M., O'Donnell, C. P., and Downey, G. (2006). Recent technological advances for the determination of food authenticity. *Trends Food Sci. Technol.* 17, 344–353. doi: 10.1016/j.tifs.2006.01.006

Ruslan, R., Khairunniza-Bejo, S., Jahari, M., and Ibrahim, M. F. (2022). Weedy rice classification using image processing and a machine learning approach. *Agric. (Switzerland)* 12, 645. doi: 10.3390/agriculture12050645

Savakar, D. G., and Anami, B. S. (2009). Recognition and classification of food grains, fruits and flowers using machine vision. *Int. J. Food Eng.* 5, 14. doi: 10.2202/1556-3758.1673

Silalahi, D. D., Midi, H., Arasan, J., Mustafa, M. S., and Caliman, J. P. (2018). Robust generalized multiplicative scatter correction algorithm on pretreatment of

near infrared spectral data. *Vib Spectrosc* 97, 55–65. doi: 10.1016/j.vibspec.2018.05.002

Sun, J., Jiang, S., Mao, H., Wu, X., and Li, Q. (2016). Classification of black beans using visible and near infrared hyperspectral imaging. *Int. J. Food Prop* 19, 1687–1695. doi: 10.1080/10942912.2015.1055760

Sun, J., Lu, X., Mao, H., Jin, X., and Wu, X. (2017a). A method for rapid identification of rice origin by hyperspectral imaging technology. *J. Food Process Eng.* 40, e12297. doi: 10.1111/jfpe.12297

Sun, J., Zhang, L., Zhou, X., Yao, K., Tian, Y., and Nirere, A. (2021). A method of information fusion for identification of rice seed varieties based on hyperspectral imaging technology. *J. Food Process Eng.* 44, e13797. doi: 10.1111/jfpe.13797

Sun, Y., Liu, Y., Yu, H., Xie, A., Li, X., Yin, Y., et al. (2017b). Non-destructive prediction of moisture content and freezable water content of purple-fleshed sweet potato slices during drying process using hyperspectral imaging technique. *Food Anal. Methods* 10, 1535–1546. doi: 10.1007/s12161-016-0722-0

Tu, K., Wen, S., Cheng, Y., Xu, Y., Pan, T., Hou, H., et al. (2022). A model for genuineness detection in genetically and phenotypically similar maize variety seeds based on hyperspectral imaging and machine learning. *Plant Methods* 18, 81. doi: 10.1186/s13007-022-00918-7

Tu, K., Wen, S., Cheng, Y., Zhang, T., Pan, T., Wang, J., et al. (2021). A non-destructive and highly efficient model for detecting the genuineness of maize variety 'JINGKE 968' using machine vision combined with deep learning. *Comput. Electron Agric.* 182, 106002. doi: 10.1016/j.compag.2021.106002

Wakholi, C., Kandpal, L. M., Lee, H., Bae, H., Park, E., Kim, M. S., et al. (2018). Rapid assessment of corn seed viability using short wave infrared line-scan hyperspectral imaging and chemometrics. *Sens Actuators B Chem.* 255, 498–507. doi: 10.1016/j.snb.2017.08.036

Wang, E., and Liu, M. (1996). A comparative biology study on the two medicinal astragalus. *Bull. Bot. Res.* 16, 87–93.

Wang, Z., Chen, J., Fan, Y., Cheng, Y., Wu, X., Zhang, J., et al. (2020). Evaluating photosynthetic pigment contents of maize using UVE-PLS based on continuous wavelet transform. *Comput. Electron Agric.* 169, 105160. doi: 10.1016/j.compag.2019.105160

Wang, H., Song, C., Sha, M., Liu, J., Li, L., and Zhang, Z. (2018). Discrimination of medicine radix astragali from different geographic origins using multiple spectroscopies combined with data fusion methods. *J. Appl. Spectrosc* 85, 313–319. doi: 10.1007/s10812-018-0650-4

Wang, J., Sun, L., Feng, G., Bai, H., Yang, J., Gai, Z., et al. (2022). Intelligent detection of hard seeds of snap bean based on hyperspectral imaging. *Spectrochim Acta A Mol. Biomol Spectrosc* 275, 121169. doi: 10.1016/j.saa.2022.121169

Wang, J., Zhang, H., and Jin, X. (2005). Morphologic characteristics and identification methods of seeds in astragalus membranaceus var. mongholicus and a. membranaceus. *Chin. Trad. Herb. Drugs* 36, 117–120.

Wang, L., Sun, D., Pu, H., and Zhu, Z. (2016). Application of hyperspectral imaging to discriminate the variety of maize seeds. *Food Anal. Methods* 9, 225–234. doi: 10.1007/s12161-015-0160-4

Wu, Y., Peng, S., Xie, Q., Han, Q., Zhang, G., and Sun, H. (2019). An improved weighted multiplicative scatter correction algorithm with the use of variable selection: Application to near-infrared spectra. *Chemometr. Intell. Lab. Syst.* 185, 114–121. doi: 10.1016/j.chemolab.2019.01.005

Xiao, Q., Bai, X., Gao, P., and He, Y. (2020). Application of convolutional neural network-based feature extraction and data fusion for geographical origin identification of radix astragali by visible/short-wave near-infrared and near infrared hyperspectral imaging. *Sensors (Switzerland)* 20, 1–14. doi: 10.3390/s20174940

Xu, Y., Tu, K., Cheng, Y., Hou, H., Cao, H., Dong, X., et al. (2021). Application of digital image analysis to the prediction of chlorophyll content in astragalus seeds. *Appl. Sci. (Switzerland)* 11, 8744. doi: 10.3390/app11188744

Yang, X., Hong, H., You, Z., and Cheng, F. (2015a). Spectral and image integrated analysis of hyperspectral data for waxy corn seed variety classification. *Sensors (Switzerland)* 15, 15578–15594. doi: 10.3390/s150715578

Yang, Y., Sun, D., Pu, H., Wang, N., and Zhu, Z. (2015b). Rapid detection of anthocyanin content in lychee pericarp during storage using hyperspectral imaging coupled with model fusion. *Postharvest Biol. Technol.* 103, 55–65. doi: 10.1016/j.postharvbio.2015.02.008

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

Yang, R., Tian, H., and Kan, J. (2018). Classification of sugar beets based on hyperspectral and extreme learning machine methods. *Appl. Eng. Agric.* 34, 891–897. doi: 10.13031/aea.12903

Yang, Z., Yang, Y., and E, X., (2020). Research progress on the difference between astragalus mongholicus and astragalus membranaceus. *J. Chin. Med. Mater.* 43, 1261–1265. doi: 10.13863/j.issn1001-4454.2020.05.044

Yang, J., Yin, C., Miao, X., Meng, X., Liu, Z., and Hu, L. (2021b). Rapid discrimination of adulteration in *radix astragali* combining diffuse reflectance mid-infrared fourier transform spectroscopy with chemometrics. *Spectrochim Acta A Mol. Biomol Spectrosc* 248, 119251. doi: 10.1016/j.saa.2020.119251

Yang, S., Zhu, Q., Huang, M., and Qin, J. (2017). Hyperspectral image-based variety discrimination of maize seeds by using a multi-model strategy coupled with unsupervised joint skewness-based wavelength selection algorithm. *Food Anal. Methods* 10, 424–433. doi: 10.1007/s12161-016-0597-0

Yan, C., Nie, F., Dong, L., and Li, L. (2005). Identification of two astragalus seeds by scanning electron microscope and ultraviolet spectrum. *LISHIZHEN Med. Materia Med. Res.* 16, 720–721.

Yan, L., Wan, T., Zhang, Z., Wang, X., and Sun, W. (2001). Analysis on botanical characters of astragalus membranacens bunge. and a. membranacens var. mongolicus (Bunge) hsiao. *J. Inner Mongolia Agric. Univ.* 22, 71–77.

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. *Spectrochim Acta A Mol. Biomol Spectrosc* 239, 118488. doi: 10.1016/j.saa.2020.118488

Zhang, J., Feng, X., Liu, X., and He, Y. (2018a). Identification of hybrid okra seeds based on near-infrared hyperspectral imaging technology. *Appl. Sci. (Switzerland)* 8, 1793. doi: 10.3390/app8101793

Zhang, H., Ge, Y., Xie, X., Atefi, A., Wijewardane, N. K., and Thapa, S. (2022a). High throughput analysis of leaf chlorophyll content in sorghum using RGB, hyperspectral, and fluorescence imaging and sensor fusion. *Plant Methods* 18, 60. doi: 10.1186/s13007-022-00892-0

Zhang, H., Hou, Q., Luo, B., Tu, K., Zhao, C., and Sun, Q. (2022b). Detection of seed purity of hybrid wheat using reflectance and transmittance hyperspectral imaging technology. *Front. Plant Sci.* 13, 1015891. doi: 10.3389/fpls.2022.1015891

Zhang, R., Li, C., Zhang, M., and Rodgers, J. (2016). Shortwave infrared hyperspectral reflectance imaging for cotton foreign matter classification. *Comput. Electron Agric.* 127, 260–270. doi: 10.1016/j.compag.2016.06.023

Zhang, T., Wei, W., Zhao, B., Wang, R., Li, M., Yang, L., et al. (2018b). A reliable methodology for determining seed viability by using hyperspectral data from two sides of wheat seeds. *Sensors (Switzerland)* 18, 813. doi: 10.3390/s18030813

Zhang, C., Yang, X., Wei, J., Chen, N., Xu, J., Bi, Y., et al. (2021). Ethnopharmacology, phytochemistry, pharmacology, toxicology and clinical applications of radix astragali. *Chin. J. Integr. Med.* 27, 229–240. doi: 10.1007/s11655-019-3032-8

Zhang, H., Zhang, L., Zhou, X., and Chen, Z. (2009). Comparative study on growth and development of *astragalus mongholicus* and *astragalus membranaceus*. *China J. Chin. Materia Med.* 34, 1591–1592.

Zhao, Y., Zhang, C., Zhu, S., Gao, P., Feng, L., and He, Y. (2018). Non-destructive and rapid variety discrimination and visualization of single grape seed using near-infrared hyperspectral imaging technique and multivariate analysis. *Molecules* 23, 1352. doi: 10.3390/molecules23061352

Zheng, S., Shang, X., Zeng, Y., and Wang, J. (2019). Research on identification of two plant origin of *astragali radix* based on specific molecular marker technology. *Modern Chin. Med.* 21, 307–311. doi: 10.13313/j.issn.1673-4890.20181117001

Zhou, Q., Huang, W., Fan, S., Zhao, F., Liang, D., and Tian, X. (2020a). Non-destructive discrimination of the variety of sweet maize seeds based on hyperspectral image coupled with wavelength selection algorithm. *Infrared Phys. Technol.* 109, 103418. doi: 10.1016/j.infrared.2020.103418

Zhou, S., Sun, L., Xing, W., Feng, G., Ji, Y., Yang, J., et al. (2020b). Hyperspectral imaging of beet seed germination prediction. *Infrared Phys. Technol.* 108, 103363. doi: 10.1016/j.infrared.2020.103363