Heliyon 10 (2024) e36999

Contents lists available at ScienceDirect

Heliyon

journal homepage: www.cell.com/heliyon

Research article

Paddy seed viability prediction based on feature fusion of color and hyperspectral image with multivariate analysis

Abdullah Al Siam, M. Mirazus Salehin, Md. Shahinur Alam, Sahabuddin Ahamed, Md. Hamidul Islam, Anisur Rahman *

Department of Farm Power and Machinery, Bangladesh Agricultural University, Mymensingh, 2202, Bangladesh

A R T I C L E I N F O

Keywords:
Seed quality
Paddy seed
Viability
Hyperspectral imaging
Machine vision

A B S T R A C T

Seed viability is essential to have a homogeneous plant population. The seed industry cannot adopt traditional procedures for seed viability evaluation since they are destructive, time-consuming, and need chemicals. This study aimed to investigate the potential of combining hyperspectral and color image features to differentiate viable and non-viable paddy seeds. The hyperspectral and color image of the 355 paddy seeds was captured and later used to examine their viability. An image processing algorithm was developed to extract features from color images of paddy seeds and investigated significant differences in the retrieved feature data using variance analysis. The spectra were extracted from the selected region of interest (ROI) of the hyperspectral paddy seed image and averaged. In the next step, the partial least square discrimination analysis (PLS-DA) model was developed to distinguish viable and non-viable paddy seeds. Initially, the PLS-DA model was developed using spectral data with different preprocessing techniques, and the result obtained an accuracy of 88.9 % in the calibration set and 86.1 % in the prediction set using Savitzky-Golay 2nd derivative preprocessed spectra. With the fusion of spectral and significant color image features, the model's accuracy improved to 93.3 % and 90.9 % in the calibration and prediction sets, respectively. Results also showed that the fusion of selected color image features with Savitzky-Golay 2nd derivative preprocessed spectra could achieve higher F1-score, recall, and precision values. The visualization map for the viable and non-viable paddy seeds was also developed utilizing the most effective predictive model. The results demonstrate the possibility of using the fusion of the hyperspectral and color image features to sort seeds according to viability, which may be applied in developing an online seed sorting method.

1. Introduction

Paddy (*Oryza sativa*) or rice is the most important staple crop in Asia and a significant source of food for nearly half of the world's population [1]. In many countries, including Bangladesh, India, and China rice consumption accounts for more than 50 % of the daily caloric intake of the population. Paddy is also a key source of income for millions of smallholder farmers in Asia. According to the United States Department of Agriculture (USDA), following China and India, Bangladesh is the world's third-largest producer of rice [2]. Seed is one of the most critical factors in the production of crops and should be considered at the first stage of crop cultivation, and

* Corresponding author.
E-mail address: anis_fpm@bau.edu.bd (A. Rahman).

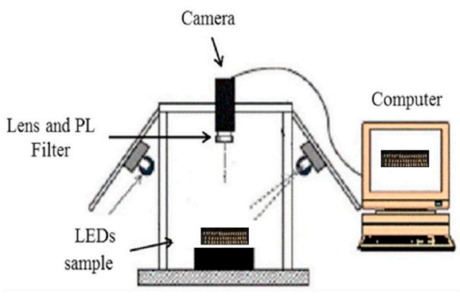
https://doi.org/10.1016/j.heliyon.2024.e36999
Received 26 February 2024; Received in revised form 26 August 2024; Accepted 26 August 2024
Available online 27 August 2024
2405-8440/

need to use good quality paddy seeds [3]. Several factors influence seed quality, including genetic purity, physical purity, moisture content, and viability. A crucial aspect of seed quality is viability, strongly related to germination rate, resilience to biotic and abiotic stress, and plant performance [4], which falls out as storage time increases [5]. Under the International Seed Testing Association (ISTA) guidelines, the conventional techniques for evaluating the viability of paddy seeds include tetrazolium staining [6,7], conductivity tests [8], immunoassay tests, accelerated ageing tests, and germination tests [9–11]. However, the traditional methods of determining paddy seed viability have some drawbacks. Such as, these methods are time-consuming and may take up to several weeks to produce results; they are not always accurate and may produce false positives or false negatives, leading to inaccurate predictions of seed viability and also damage or destroy the seed being tested, making it unsuitable for planting or further testing [12]. This is especially difficult when dealing with rare or valuable seed varieties. Hyperspectral imaging (HSI) and machine vision systems are simple, quick, contact-free, and non-destructive technologies that have been successfully used for crop seeds, including rice [3,13–21], maize [22–26], wheat [27], soybean [28–30], sugar beat [31,32], hazelnut [33], peanut [34], watermelon [35], sunflower [36] and muskmelon [37], etc. In the previous study, Qi et al. [18] applied a technique to detect rice seed vigour using near-infrared hyperspectral imaging and convolutional neural network (CNN) model with different transfer learning. The experimental results showed that the convolutional neural network model with MixStyle transfer knowledge reached an average of 85.11 % accuracy in validation sets. Hong et al. [19] developed a technology to predict the viability of rice seeds using visible–near infrared (VIS–NIR) hyperspectral imaging system and spectral-spatial information modelling, namely CNN, PLS–CNN, and dual branch networks applied for viability prediction, and the result shows that an accuracy and F1 scores of approximately 90 % and 86.49 %, respectively. Jin et al. [20] determined the viability and vigour of naturally-aged rice seeds using near-infrared hyperspectral imaging with machine learning algorithms. The overall results showed that deep learning methods and conventional machine learning methods could predict the viability and vigour of different rice seeds well, and the accuracy of most models was over 85 %. Qi et al. [21] detected the viability of natural ageing seeds using near-infrared hyperspectral imaging, spectral angle mapper generative adversarial network (SAM-GAN) and CNN model with real data modelling, fake data modeling and mixed modeling of real data and fake data. The accuracy of the CNN model established by real data modelling, fake data modeling and mixing real data with fake data generated by SAM-GAN reaches nearly 72.65 %, 74.50 % and 98.71 %, respectively, for the mix of four rice varieties. Most of the above method is mainly based on spectral information, and the image information has not been analyzed. In this study, combining hyperspectral and color image features such as color, morphological, and textural characteristics was used as an indicator of seed viability. This fusion can provide a comprehensive view of the paddy seed's internal and external characteristics. However, to the best of our knowledge, no reports have yet addressed on the feature fusion of color and spectral features form hyperspectral image to examine the viability of paddy seeds. Therefore, this study investigated the possibility of combining color and spectral features form hyperspectral image to discriminate viable and non-viable paddy seeds. The specific objectives of this study were to acquire and extract corresponding spectral data from the color and hyperspectral image feature of the paddy seed; develop and evaluate a multivariate model of classification for paddy seed viability, and finally, develop the chemical image using the most effective model of prediction to assess the paddy seed viability.

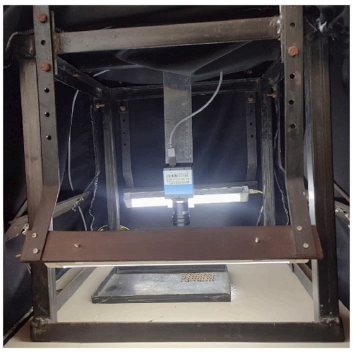
2. Materials and method

2.1. Paddy sample selection

A total of 2 kg of fresh paddy seeds (BRRI *dhan*28) were brought from the local market of Mymensingh City, Bangladesh. The grains were long and slender, with an average length of 7.5–8.5 mm and a width of 2.7–3.0 mm. Three hundred fifty five seeds were selected from the purchased lot, free from disease, cracks, and discoloration, and placed randomly with labels into ten sample trays for hyperspectral and color image acquisition.



(a) Schematic diagram



(b) Experimental setup

Fig. 1. The color image acquisition system for paddy seed. (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

2.2. Image acquisition

2.2.1. Color image acquisition

A camera (DFK 42AU02, Imaging source, Germany) with a lens (Pentax, Tokyo, Japan) was used to acquire the color image of paddy seeds. Four LED lights (4000K, OEM&ODM, China) were positioned in such a way as to provide the best possible visibility of the sample, with each measuring 30 cm in length and having a power rating of 6 W. A sheet of white paper was used to conduct a white balancing before the acquisition of images. A polarizing filter (Edmund Optics, NY, USA) was also used to remove any halation caused by reflected light. The filter was placed in front of the camera lens to eliminate any unwanted reflections that could interfere with the quality of the images. The schematic and experimental setup of the color image acquisition system for paddy seed is shown in Fig. 1(a) and (b), respectively. The sample tray with labeled seeds was positioned beneath the camera on a dark surface for every image acquisition. A 12 cm gap was maintained between the paddy seed and the camera. Once the images were captured, the image was transferred to a computer for further processing.

2.2.2. Hyperspectral image acquisition

A desktop hyperspectral imaging device (HSVIS-12bit-15fps, HYSPIM, Sweden) was used in this study to acquire images of the paddy seeds. The schematic and experimental setup of the hyperspectral image acquisition system for paddy seed is shown in Fig. 2(a) and (b), respectively. The system included a line-scan image spectrograph covering the spectral range of 400–800 nm with a translation stage, a lighting system, and built-in computer software for controlling the camera. The spectral range has 367 bands with a resolution of 2 nm each. Four tungsten-halogen lights were used as light sources to illuminate the paddy seed samples. Before capturing the images, the white balance was adjusted using a sheet of white paper, while the black balance was adjusted using a cover placed in front of the lens. During image acquisition, the translation stage was used to move a sample tray with labeled seeds through the system underneath the camera. The images were captured by scanning the tray line by line. The scanning speed was 0.5 cm/s, and a 25 cm gap was maintained between the camera and the sample. After acquiring the 3-D image data cube, the spectral data were extracted from each paddy seed image's selected region of interest (ROI) using the built-in computer-assisted software and then averaged. The labeled spectral and image data were then transferred to a computer to further develop the multivariate classification model.

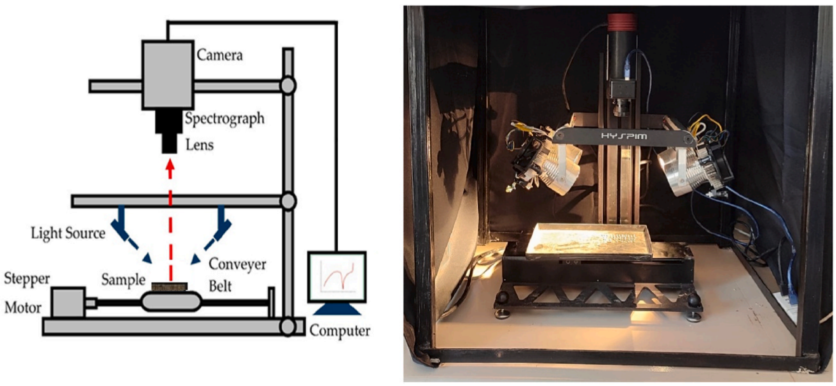
2.3. Paddy seed viability test

After acquiring color and hyperspectral images, the paddy seed samples were subjected to a viability test to confirm their ability to germinate. The testing process followed the International Seed Testing Association's (ISTA) guidelines. For this purpose, petri dishes were used to contain sterile sand moistened to a field capacity. The labeled seeds were placed in the containers and exposed to laboratory conditions with a 25–30 °C temperature range and relative humidity of 70–80 %. The counting of viable seeds was performed fourteen days after the seed set. During the counting process, any abnormal seedlings, dead seeds, and non-germinated seeds were considered non-viable, whereas normal seedlings were considered viable. Based on the seed viability results obtained from the test, the viable and non-viable seed samples were labeled accordingly.

2.4. Data analysis

2.4.1. Color image analysis

The paddy seed color images should be separate from the background. At first, the color images were transformed to HSV color space, and the channel histogram was used to establish the threshold value. The chosen histogram threshold value was then used to construct a masked image, and *imfill* and *bwareaopen* operations were applied to remove all small artifacts. The segmented paddy seed



(a) Schematic diagram (b) Experimental setup

Fig. 2. The hyperspectral image acquisition system for paddy seed.

image was obtained using the resulting masked image. This process was repeated for every paddy seed image. The segmented paddy seed image was then analyzed, and seven color indexes were extracted, which include the average red (R), green (G), and blue colors (B), brightness (BR), and normalized red (NRI), green (NGI), and blue indices (NBI) according to the previous study [3]. For each segmented paddy seed image, nine morphological features were obtained using an ellipse-based method [38,39], which include area, perimeter, major and minor axis length, orientation, eccentricity, solidity, extent, and equivalent diameter. The surface texture of the seed is a significant characteristic to consider in image analysis, and many researchers use a method called grey-level co-occurrence matrix (GLCM) to measure it [39,40]. In the proposed study, four textural features, which include contrast, correlation, energy, and homogeneity, were derived from the GLCM.

2.4.2. Statistical analysis of features

The ANOVA (Analysis of Variance) is a statistical technique used to test for variations between the means of two or more groups that are statistically significant. In this study, ANOVA was used to analyze the features of each paddy seed and determine if there were any significant differences between the viable and non-viable groups. The ANOVA test calculates the F-statistic, which quantifies the ratio of the between-group variance to the within-group variance. A high F-statistic indicates that there are significant differences between the groups, while a low F-statistic indicates that the differences are not significant. The extracted color, morphology, and texture features were imported into Microsoft Excel 2016 and categorized into viable and non-viable for each feature based on the viability test. Finally, ANOVA was performed using Microsoft Excel 2016 to analyze the features of each paddy seed and determine if there were any significant differences between them.

2.4.3. Spectral data analysis

Spectral analysis is a technique used to obtain information about a sample's chemical and physical properties based on its spectral signature. In the case of paddy seed samples, spectral analysis can provide valuable information about their moisture, protein, and starch content, among other characteristics. By analyzing the spectra, we can determine the viability of the seeds based on their chemical and physical properties. During the acquisition of spectra, noise from many physical elements, such as sensor sensitivity, light source, ambient temperature, and electric field noise, frequently affects the spectra data collected from the spectroscopic devices. Hence, the preprocessing of spectral data using mathematical analyses is essential to extract relevant information from the sample and eliminate undesirable spectral data fluctuations. The preprocessing techniques are used to correct the spectrum, enhance the spectral signal, and minimize unwanted information, including baseline shifts, path-length variations, scatter variations, and background noise [41,42]. In this study, the averaged spectra were independently preprocessed using different methods like normalization, multiplicative scatter correlation (MSC), standard normal variable (SNV), and Savitzky-Golay derivatives to remove obtrusive noise from the original data and make the spectral data more suitable for analysis.

2.4.4. Multivariate classification model

Spectral data can be very complex due to a large number of variables, making it challenging to interpret the data effectively. Each variable may represent a specific trait of the sample, or it may not correspond to anything at all. Therefore, analyzing such complex data requires using multivariate data analysis tools. The most popular and practical multivariate classification model to analyze spectral data is partial least squares regression/discriminant analysis (PLSR/PLS-DA), principal component analysis (PCA), support vector machine (SVM), linear discriminant analysis (LDA), artificial neural network (ANN), etc. This study used the PLS-DA multivariate classification technique to distinguish viable paddy seeds from non-viable paddy seeds.

The partial least square-discriminant analysis (PLS-DA) is an algorithm of classification that is useful for both predictive and descriptive modeling and for selecting relevant variables. It seeks to find the latent variables that explain the maximum covariance between the predictors (spectra/features) and the response (class labels) variables. These variables are then used to construct a linear discriminant function that separates the classes in the data in the most optimal way. The PLS-DA was utilized to classify paddy seeds according to their viability. The response variable Y in PLS-DA is a collection of binary variables related to the sample's categories or classes. The PLS-DA is stated as follows:

$$Y = Xb + E \tag{1}$$

Where X is a matrix of order $n \times p$ that contains the image feature of every class; b is the coefficient of regression, and E is the term for error. The spectral data of viable and non-viable seeds were put in a matrix X and an artificial value representing class was present in the Y matrix for this study's creation of the PLS-DA model, as shown below:

$$Y = \begin{cases} 1 = \textit{sample belongs to viable group} \\ 3 = \textit{sample belongs to non}-\textit{viable group} \end{cases} \tag{2}$$

A baseline of ±1.5 was used for each group to appropriately identify the samples. Samples within the range of ±1.5 from any group were regarded as belonging to that group. In this study, the classification model was developed using a dataset randomly divided into two groups: a training set of 60 % of the data and a prediction set of 40 %. During the calibration model development, the approach of 10-fold cross-validation was used. Several factors were taken into account when assessing the classification model's performance, including the number of samples in the external prediction set that the model correctly classified as belonging to the modeled category and accepted, the number of samples that the model correctly classified as not belonging to the modeled category and rejected, F1-score, precision, recall and overall accuracy. The F1-score was a measure of prediction accuracy that was calculated from the precision and recall, where the precision was the number of true positive results divided by the number of all positive results, including those not identified correctly, and the recall was the number of true positive results divided by the number of all samples that should have been identified as positive. The overall accuracy was calculated as the percentage of actual results from all instances examined in the study. A higher accuracy value indicates a better-performing model. The MATLAB software [43] was used for image processing, feature extraction, and creation of the classification model.

2.4.5. Viability prediction and visualization

In this work, the beta coefficient of the classification model for the combined hyperspectral and selected features from the color image was used to develop visual maps for the differentiation of the viable and non-viable paddy seeds. The HSI at the optimal wavelengths was unfolded into a 2-D matrix and multiplied with the regression coefficients together with the summation of the color characteristics coefficient acquired from the fusion classification model to produce maps for the samples. After that, the generated vector was folded back to the 2-D image, on which a median filter of order 5 × 5 was applied to improve the visual representation. All the image processing steps involved in the visualization goals were performed using a program developed with MATLAB software [43].

3. Result and discussion

3.1. Characteristics of spectral profiles

Fig. 3 shows the average relative reflectance spectra at the spectral range (400–800 nm) using 368 bands (variables). The graph indicates that viable and non-viable seeds exhibit identical spectral patterns, with viable seeds having higher reflectance than non-viable seeds.

At wavelengths between 600 and 800 nm, the average values of spectra of viable seeds were higher. As seen from the second derivative spectra, there is a peak at the wavelengths 600 nm. This wavelength was mainly related to the color of the paddy seeds. However, for 2nd derivatives of average reflectance spectra, the spectral differences among the viable and non-viable seeds were not significant.

3.2. Statistics of measured samples and feature selection

Of the 355 seeds, 221 were found to be viable in this experiment, while the remaining 134 were non-viable, resulting in an overall viability percentage of 60.6 %. In this study, the ANOVA: Single factor test was used to investigate whether there was a noteworthy variation in the image features between viable and non-viable seeds.

To conduct the ANOVA test, the F-value was calculated, which measures the ratio of the variance between the groups and the variance within the groups. The F-value was then compared to the critical value of F at a level of significance of 5 %, which is the threshold value above which the differences between groups are considered statistically significant. The critical value of F at a 5 % level of significance was determined to be 3.861. The ANOVA test results are presented in Table 1, which shows the F values for each of the 20 features extracted from the paddy seed images. The test results showed that for 8 out of the 20 features, the F value exceeded the critical value of $F_{critical}$. This indicates that these eight features significantly differed between the viable and non-viable paddy seeds. These significant features were used in further analysis and model development for seed viability classification.

3.3. Classification model

The study collected 355 spectra from paddy seeds to develop a general model for predicting seed viability. However, due to inadequate spectral resolution, 03 of these spectra were not considered for analysis. This exclusion was likely necessary to ensure that the data analyzed was of sufficient quality and to prevent any inaccurate results that may have arisen from low-quality spectra. Therefore, the remaining 352 spectra were used in developing the classification model for predicting seed viability.

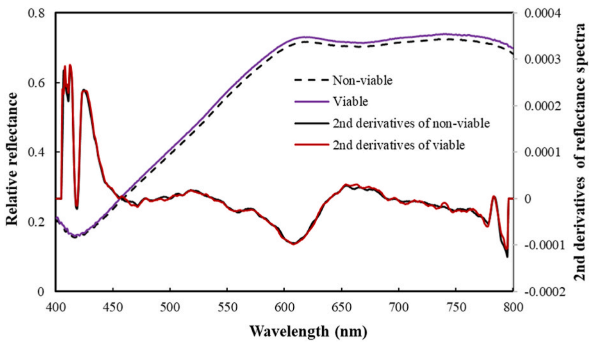


Fig. 3. The average relative reflectance spectra with second derivatives for paddy seed.

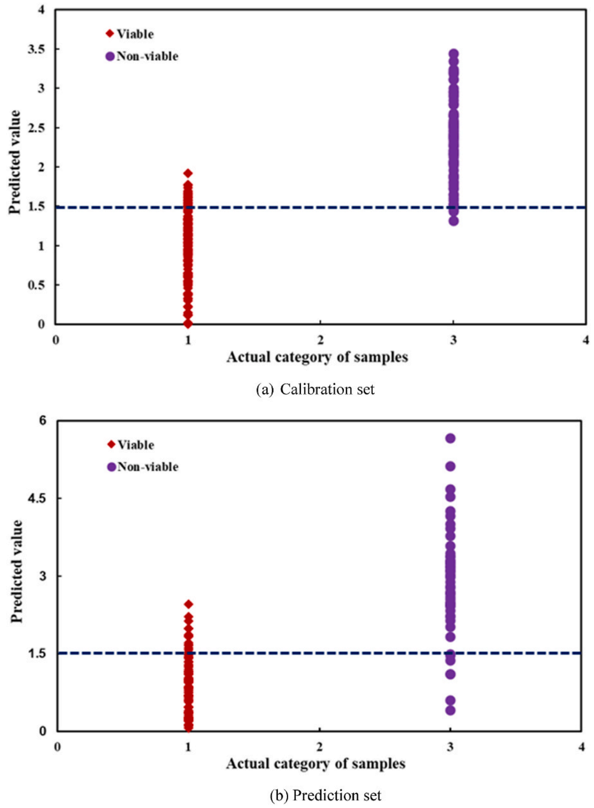


Fig. 4. Classification result of the developed model with Savitzsky-Golay 2nd derivative preprocessed spectra.

Table 1
Color image features and their F-values.

Feature	*F-value*	Feature	*F-value*
Area	**4.912**	**Correlation**	**4.305**
Major Axis Length	2.309	Energy	0.496
Minor Axis Length	**3.910**	Homogeneity	0.157
Eccentricity	0.001	Red	2.828
Orientation	0.054	**Green**	**3.944**
Equivalent Diameter	**4.997**	**Blue**	**3.965**
Solidity	0.511	**Brightness**	**3.915**
Extent	0.287	NRI	2.051
Perimeter	**3.907**	NGI	1.290
Contrast	0.009	NBI	1.829

Table 2
Results of calibration model using spectra with different preprocessing techniques.

Preprocessing methods	Viable seeds (128)		Non-viable seeds (80)		F1-Score (%)	Recall (%)	Precision (%)	Overall accuracy (%)
	Correct	Incorrect	Correct	Incorrect				
Min normalization	98	30	75	5	84.9	95.1	76.6	83.2
Max normalization	88	40	71	9	78.2	90.7	68.7	76.4
Range normalization	90	38	72	8	79.7	91.8	70.3	77.9
MSC	101	27	76	4	86.7	96.2	78.9	85.1
SNV	98	30	74	6	84.5	94.2	76.6	82.7
S–G 1st derivatives	97	31	73	7	83.6	93.3	75.8	81.7
S-G 2nd derivatives	**108**	**20**	**77**	**3**	**90.4**	**97.3**	**84.4**	**88.9**
Raw	91	37	71	9	79.8	91.0	71.1	77.9

3.3.1. Classification model using spectra with different preprocessing techniques

Calibration and prediction models for the different preprocessed spectra were developed by using the PLS-DA method. After preprocessing techniques, the PLS-DA model produced satisfactory accuracies that were adequate, indicating a distinction between viable and non-viable seeds.

The calibration accuracy for paddy seed germinability was 76.4 %–88.9 % (Table 2) and it is clear that the greatest accuracy was achieved by the Savitzsky-Golay 2nd derivative among all the methods, with a value of 88.9 % in the calibration model, surpassing the other techniques used in the study. The F1-score, recall, and precision for the Savitzsky-Golay 2nd derivative preprocessed spectra-based model were 90.4 %, 97.3 %, and 84.4 %, respectively. The Savitzsky-Golay 2nd preprocessed spectra-based model shows a higher accuracy of calibration model out of all (Table 2) for classifying the viable and non-viable paddy seeds. When the calibrated model was applied to the prediction set, the results were presented in Table 3 with an accuracy range of 75.4 %–86.1 %, and the Savitzsky-Golay 2nd derivative preprocessed spectra provided F1-score, recall, precision, and accuracy of 88.3 %, 92.7 %, 84.4 %, and 86.1 %, respectively. Using the Savitzsky-Golay 2nd derivative preprocessing technique made it simpler to discern between viable and non-viable paddy seeds by correcting the spectral baseline effect. In other words, the approach improves classification accuracy by removing the deviation among the baseline spectra from different seeds.

Fig. 4 (a) and (b) depict the calibration and prediction classification model results using Savitzsky-Golay 2nd derivative preprocessed spectra, respectively. The graph shows that the viable and non-viable paddy seeds can be differentiated accurately. However, some seeds are misclassified because of their spectra signature. The fact that the points for the viable and non-viable seeds are clustered independently shows that the PLS-DA model can correctly categorize the seeds into their respective groups.

3.3.2. Spectra-based classification model with best preprocessing method and selected color image feature

Using the best preprocessing method and selected color image features that were identified through the ANOVA test, the classification model was developed to enhance its accuracy of the classification model. The results of the developed model are presented in Table 4. The F1-score, recall, precision, and accuracy were found to be higher at 82.2 %, 93.9 %, 80.8 %, and 93.3 % for the calibration dataset, respectively, and 80.9 %, 91.7 %, 79.1 %, and 90.9 % for prediction dataset respectively, compared with the result obtained using only preprocessed spectra. These results suggest that the fusion of spectral data and the selected color image features have improved the classification model's accuracy. This model showed the better results that were reported for the seed viability using near-infrared hyperspectral imaging combined with the CNN model with an accuracy of 72.65 % for the mix of four varieties rice seeds [21]. The selected color image features likely added additional information that helped to further distinguish between viable and non-viable paddy seeds. The higher overall accuracy in the calibration and prediction datasets indicates that the classification model performs more accurately than before, which is essential for practical applications.

The result of the calibration and prediction classification model with the best preprocessing method and the selected color image feature is shown in Fig. 5 (a) and 5 (b), respectively. Based on these results, the fusion of hyperspectral with color image model has a higher overall prediction performance due to the additional information that color, morphology, and textural feature helped to further distinguish between viable and non-viable paddy seed. Therefore, it can be advantageous to use the hyperspectral image with color image feature information rather than hyperspectral information to differentiate viable and non-viable paddy seeds.

3.4. Visualization map for the viable and non-viable paddy seed

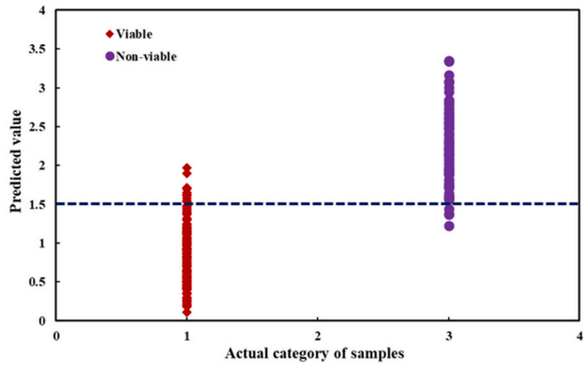
Since each pixel in the HSI corresponds to a spectrum, it is possible to see the chemical components of a sample by looking at the spectrum of individual pixels. In this study, the beta coefficients obtained from the classification model developed using Savitzsky-Golay 2nd derivatives preprocessed spectra with the selected color image feature were applied to each pixel in an image to classify the paddy seed based on viable and non-viable attributes. Fig. 6 (a) shows the original hyperspectral image, and Fig. 6 (b) shows the visualization map of paddy seed based on viability, and it observed that the high intensity of color represents that the seed is more non-viable, which corresponds with the spectra difference between the viable and non-viable paddy seeds. Also, some viable seeds were classified as non-viable, and some misclassified. The distribution maps obtained from the current study show the benefits of HSI with selected features from color images, and the results produced cannot be obtained with either conventional imaging or traditional spectroscopy methods alone.

Table 3
Results of prediction model using spectra with different preprocessing techniques.

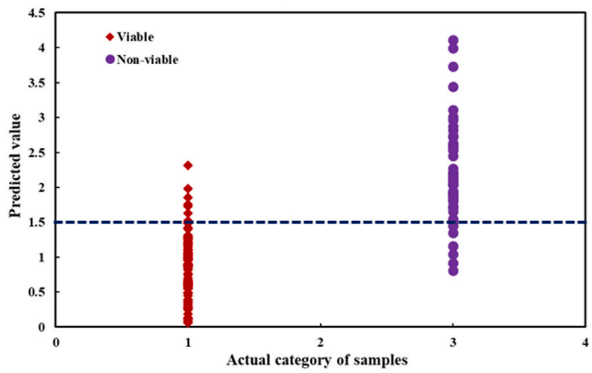
Preprocessing methods	Viable seeds (90)		Non-viable seeds (54)		F1-Score (%)	Recall (%)	Precision (%)	Overall accuracy (%)
	Correct	Incorrect	Correct	Incorrect				
Min normalization	70	20	42	12	81.4	85.4	77.8	77.9
Max normalization	70	20	38	16	79.5	81.4	77.8	75.0
Range normalization	71	19	39	15	80.6	82.6	79.0	75.4
MSC	69	21	43	11	81.1	86.3	76.7	77.8
SNV	70	20	43	11	81.8	86.4	77.9	78.5
S–G 1st derivatives	69	21	41	13	80.2	84.1	76.7	76.4
S-G 2nd derivatives	**76**	**14**	**48**	**6**	**88.3**	**92.7**	**84.4**	**86.1**
Raw	65	25	41	13	77.4	83.3	72.2	73.6

Table 4
Results of the developed model using spectra and best preprocessing with the selected feature.

Data Set	Viable seeds		Non-viable seeds		F1-Score (%)	Recall (%)	Precision (%)	Overall accuracy (%)
	Correct	Incorrect	Correct	Incorrect				
Calibration	118	10	76	4	82.2	93.9	80.7	93.3
Prediction	83	7	48	6	80.9	91.7	79.1	90.9

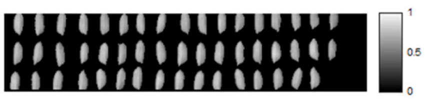


(a) Calibration set

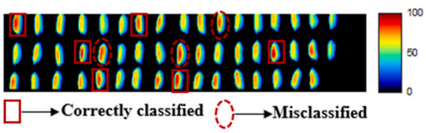


(b) Prediction set

Fig. 5. Classification result of the developed model with the best preprocessing method and selected color image feature. (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)



(a) The original hyperspectral image



(b) Sample visualization where a higher red pixel intensity denotes non-viable seeds

Fig. 6. Visualization map for the viable and non-viable paddy seed.

4. Conclusion

The study successfully demonstrated the feasibility of using hyperspectral image and selected color image features combined with a classification model (PLS-DA) to discriminate viable and non-viable paddy seeds. The classification model based on spectra with the different preprocessing techniques was developed, and the best spectra preprocessing technique was selected. The highest accuracy of

86.1 % was obtained using Savitzky-Golay 2nd derivative preprocessed spectra. Additionally, the fusion of selected color image features (such as color, morphological and textural) with spectra improved the accuracy of the classification model. The model was able to classify the seeds with an accuracy of 93.3 % in the calibration set and 90.9 % in the prediction set. Also, the model has showed a high F1-score, recall, and precision using the fusion of selected color image features with Savitzky-Golay 2nd derivatives preprocessed spectra. Finally, a visualization map was created to transfer the prediction model to each pixel in the image and determine the viable and non-viable paddy seeds for real-time assessment. This indicates that the model is accurate in predicting the viability of the seeds based on spectral and color image feature information. Further, an automated online system could be developed to make this work more effective and fruitful. Additionally, the paddy varietal purity algorithm could be added, which will measure along with paddy seed viability inspection system.

Funding

This research was funded by Ministry of Science and Technology, Peoples Republic of Bangladesh through the special research grant (BS/37-2021-2022) and University Grants Commission of Bangladesh through the research support grant (CropScience-01/2021-2022).

Ethics approval and consent to participate

Not applicable.

Consent for publication

All authors consent to publish this work.

Data availability statement

The data supporting the findings of this study are not deposited in a publicly available repository. However, the data will be made available upon reasonable request. Interested researchers may contact the corresponding author for access to the data.

CRediT authorship contribution statement

Abdullah Al Siam: Writing – original draft, Methodology, Investigation, Conceptualization. **M. Mirazus Salehin:** Visualization, Methodology, Investigation, Formal analysis. **Md. Shahinur Alam:** Validation, Supervision, Methodology, Conceptualization. **Sahabuddin Ahamed:** Writing – review & editing, Visualization, Validation. **Md. Hamidul Islam:** Writing – review & editing, Validation, Supervision. **Anisur Rahman:** Writing – review & editing, Supervision, Funding acquisition.

Declaration of competing interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests:Anisur Rahman reports financial support was provided by Government of the People's Republic of Bangladesh Ministry of Science and Technology. Anisur Rahman reports financial support was provided by University Grants Commission of Bangladesh. If there are other authors, they declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

The authors were grateful to the Bangladesh Agricultural University Research System (BAURES) for financial management and logistic support during the research work.

References

[1] A.S. Juraimi, M.K. Uddin, M.P. Anwar, M.T.M. Mohamed, M.R. Ismail, A. Man, Sustainable weed management in direct seeded rice culture: a review, Aust. J. Crop. Sci. 7 (2013) 989–1002.

[2] M.A. Al Mamun, S.A. Islam Nihad, M.A. Rouf Sarkar, M.A. Aziz, M.A. Qayum, R. Ahmed, N.M. Farhat Rahman, M.I. Hossain, M.S. Kabir, Growth and trend analysis of area, production and yield of rice: a scenario of rice security in Bangladesh, PLoS One 16 (2021) 1–18, https://doi.org/10.1371/journal.pone.0261128.

[3] N. Ansari, S.S. Ratri, A. Jahan, M. Ashik-E-Rabbani, A. Rahman, Inspection of paddy seed varietal purity using machine vision and multivariate analysis, J. Agric. Food Res. 3 (2021) 100109, https://doi.org/10.1016/j.jafr.2021.100109.

[4] J. Dumont, T. Hirvonen, V. Heikkinen, M. Mistretta, L. Granlund, K. Himanen, L. Fauch, I. Porali, J. Hiltunen, S. Keski-Saari, M. Nygren, E. Oksanen, M. Hauta-Kasari, M. Keinänen, Thermal and hyperspectral imaging for Norway spruce (Picea abies) seeds screening, Comput. Electron. Agric. 116 (2015) 118–124, https://doi.org/10.1016/j.compag.2015.06.010.

[5] Mladen Tatić, S. Balešević-Tubić, V. Đorđević, Z. Nikolić, V. Đukić, M. Vujaković, G. Cvijanovic, Soybean seed viability and changes of fatty acids content as affected by seed aging, Afr. J. Biotechnol. 11 (2012) 10310–10316, https://doi.org/10.5897/ajb11.3505.

[6] S.T. Hosomi, C.C. Custódio, P.T. Seaton, T.R. Marks, N.B. Machado-Neto, Improved assessment of viability and germination of Cattleya (Orchidaceae) seeds following storage, Vitro Cell Dev. Biol. Plant 48 (2012) 127–136, https://doi.org/10.1007/s11627-011-9404-1.

[7] M.H. Olesen, P. Nikneshan, S. Shrestha, A. Tadayyon, L.C. Deleuran, B. Boelt, R. Gislum, Viability prediction of ricinus cummunis L. Seeds using multispectral imaging, Sensors 15 (2015) 4592–4604, https://doi.org/10.3390/s150204592.

[8] M.H. Olesen, N. Shetty, R. Gislum, B. Boelt, Classification of viable and non-viable spinach (Spinacia oleracea L.) seeds by single seed near infrared spectroscopy and extended canonical variates analysis, J. Near Infrared Spectrosc. 19 (2011) 171–180, https://doi.org/10.1255/jnirs.928.

[9] D.J. Merritt, A.J. Martyn, P. Ainsley, R.E. Young, L.U. Seed, M. Thorpe, F.R. Hay, L.E. Commander, N. Shackelford, C.A. Offord, K.W. Dixon, R.J. Probert, A continental-scale study of seed lifespan in experimental storage examining seed, plant, and environmental traits associated with longevity, Biodivers. Conserv. 23 (2014) 1081–1104, https://doi.org/10.1007/s10531-014-0641-6.

[10] S. Shrestha, M. Knapič, U. Žibrat, L.C. Deleuran, R. Gislum, Single seed near-infrared hyperspectral imaging in determining tomato (Solanum lycopersicum L.) seed quality in association with multivariate data analysis, Sensor. Actuator. B Chem. 237 (2016) 1027–1034, https://doi.org/10.1016/j.snb.2016.08.170.

[11] A. Rahman, B.K. Cho, Assessment of seed quality using non-destructive measurement techniques: a review, Seed Sci. Res. 26 (2016) 285–305, https://doi.org/10.1017/S0960258516000234.

[12] M. Huang, Q.G. Wang, Q.B. Zhu, J.W. Qin, G. Huang, Review of seed quality and safety tests using optical sensing technologies, Seed Sci. Technol. 43 (2015) 337–366, https://doi.org/10.15258/sst.2015.43.3.16.

[13] W. Kong, C. Zhang, F. Liu, P. Nie, Y. He, Rice seed cultivar identification using near-infrared hyperspectral imaging and multivariate data analysis, Sensors 13 (2013) 8916–8927, https://doi.org/10.3390/s130708916.

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

[15] M.A. Islam, M.R. Hassan, M. Uddin, M. Shajalal, Germinative paddy seed identification using deep convolutional neural network, Multimed. Tool. Appl. 82 (2023) 39481–39501, https://doi.org/10.1007/s11042-023-14914-z.

[16] Z. Qiu, J. Chen, Y. Zhao, S. Zhu, Y. He, C. Zhang, Variety identification of single rice seed using hyperspectral imaging combined with convolutional neural network, Appl. Sci. 8 (2018) 1–12, https://doi.org/10.3390/app8020212.

[17] X. He, X. Feng, D. Sun, F. Liu, Y. Bao, Y. He, Rapid and nondestructive measurement of rice seed vitality of different years using near-infrared hyperspectral imaging, Molecules 24 (2019) 2227, https://doi.org/10.3390/molecules24122227.

[18] H. Qi, Z. Huang, Z. Sun, Q. Tang, G. Zhao, X. Zhu, C. Zhang, Rice seed vigor detection based on near-infrared hyperspectral imaging and deep transfer learning, Front. Plant Sci. 14 (2023) 1–13, https://doi.org/10.3389/fpls.2023.1283921.

[19] S.J. Hong, T. Yang, S.Y. Kim, E.C. Kim, C.H. Lee, N.I. Nurhisna, S. Kim, S.W. Roh, J. Ryu, G. Kim, Nondestructive prediction of rice seed viability using spectral and spatial information modeling of visible–near infrared hyperspectral images, J. ASABE. 65 (2022) 997–1006, https://doi.org/10.13031/ja.14982.

[20] B. Jin, H. Qi, L. Jia, Q. Tang, L. Gao, Z. Li, G. Zhao, Determination of viability and vigor of naturally-aged rice seeds using hyperspectral imaging with machine learning, Infrared Phys. Technol. 122 (2022) 104097, https://doi.org/10.1016/j.infrared.2022.104097.

[21] H. Qi, Z. Huang, B. Jin, Q. Tang, L. Jia, G. Zhao, D. Cao, Z. Sun, C. Zhang, SAM-GAN: an improved DCGAN for rice seed viability determination using near-infrared hyperspectral imaging, Comput. Electron. Agric. 216 (2024) 108473, https://doi.org/10.1016/j.compag.2023.108473.

[22] X. Zhang, F. Liu, Y. He, X. Li, Application of hyperspectral imaging and chemometric calibrations for variety discrimination of maize seeds, Sensors 12 (2012) 17234–17246, https://doi.org/10.3390/s121217234.

[23] A. Ambrose, L.M. Kandpal, M.S. Kim, W.H. Lee, B.K. Cho, High speed measurement of corn seed viability using hyperspectral imaging, Infrared Phys. Technol. 75 (2016) 173–179, https://doi.org/10.1016/j.infrared.2015.12.008.

[24] S. Yang, Q.B. Zhu, M. Huang, J.W. Qin, 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 (2017) 424–433, https://doi.org/10.1007/s12161-016-0597-0.

[25] C. Wakholi, L.M. Kandpal, H. Lee, H. Bae, E. Park, M.S. Kim, C. Mo, W.H. Lee, B.K. Cho, Rapid assessment of corn seed viability using short wave infrared line-scan hyperspectral imaging and chemometrics, Sensor. Actuator. B Chem. 255 (2018) 498–507, https://doi.org/10.1016/j.snb.2017.08.036.

[26] K. Song, Y. Zhang, T. Shi, D. Yang, Rapid detection of imperfect maize kernels based on spectral and image features fusion, J. Food Meas. Char. 18 (2024) 3277–3286, https://doi.org/10.1007/s11694-024-02402-3.

[27] T. Zhang, W. Wei, B. Zhao, R. Wang, M. Li, L. Yang, J. Wang, Q. Sun, A reliable methodology for determining seed viability by using hyperspectral data from two sides of wheat seeds, Sensors 18 (2018) 813, https://doi.org/10.3390/s18030813.

[28] M. Al-Amery, R.L. Geneve, M.F. Sanches, P.R. Armstrong, E.B. Maghirang, C. Lee, R.D. Vieira, D.F. Hildebrand, Near-infrared spectroscopy used to predict soybean seed germination and vigour, Seed Sci. Res. 28 (2018) 245–252, https://doi.org/10.1017/S0960258518000119.

[29] I. Baek, D. Kusumaningrum, L.M. Kandpal, S. Lohumi, C. Mo, M.S. Kim, B.K. Cho, Rapid measurement of soybean seed viability using Kernel-based multispectral image analysis, Sensors 19 (2019) 271, https://doi.org/10.3390/s19020271.

[30] H. Lee, T.Q. Huy, E. Park, H.-J. Bae, I. Baek, M.S. Kim, C. Mo, B.-K. Cho, Machine vision technique for rapid measurement of soybean seed vigor, J. Biosyst. Eng. 42 (2017) 227–233, https://doi.org/10.5307/JBE.2017.42.3.227.

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

[32] J. Wang, L. Sun, W. Xing, G. Feng, J. Yang, J. Li, W. Li, Sugarbeet seed germination prediction using hyperspectral imaging information fusion, Appl. Spectrosc. 77 (2023) 710–722, https://doi.org/10.1177/00037028231171908.

[33] A. Giraudo, R. Calvini, G. Orlandi, A. Ulrici, F. Geobaldo, F. Savorani, Development of an automated method for the identification of defective hazelnuts based on RGB image analysis and colourgrams, Food Control 94 (2018) 233–240, https://doi.org/10.1016/j.foodcont.2018.07.018.

[34] Z. Zou, J. Chen, W. Wu, J. Luo, T. Long, Q. Wu, Q. Wang, J. Zhen, Y. Zhao, Y. Wang, Y. Chen, M. Zhou, L. Xu, Detection of peanut seed vigor based on hyperspectral imaging and chemometrics, Front. Plant Sci. 14 (2023) 1–13, https://doi.org/10.3389/fpls.2023.1127108.

[35] J. Sun, A. Nirere, K.D. Dusabe, Z. Yuhao, G. Adrien, Rapid and nondestructive watermelon (Citrullus lanatus) seed viability detection based on visible near-infrared hyperspectral imaging technology and machine learning algorithms, J. Food Sci. (2024) 4403–4418, https://doi.org/10.1111/1750-3841.17151.

[36] P. Huang, J. Yuan, P. Yang, F. Xiao, Y. Zhao, Nondestructive detection of sunflower seed vigor and moisture content based on hyperspectral imaging and chemometrics, Foods 13 (9) (2024) 1320, https://doi.org/10.3390/foods13091320.

[37] L.M. Kandpal, S. Lohumi, M.S. Kim, J.S. Kang, B.K. Cho, Near-infrared hyperspectral imaging system coupled with multivariate methods to predict viability and vigor in muskmelon seeds, Sensor. Actuator. B Chem. 229 (2016) 534–544, https://doi.org/10.1016/j.snb.2016.02.015.

[38] Z.Y. Liu, F. Cheng, Y. Bin Ying, X.Q. Rao, Identification of rice seed varieties using neural network, J. Zhejiang Univ. - Sci. 6 (B) (2005) 1095–1100, https://doi.org/10.1631/jzus.2005.B1095.

[39] B. Lurstwut, C. Pornpanomchai, Image analysis based on color, shape and texture for rice seed (Oryza sativa L.) germination evaluation, Agric. Nat. Resour. 51 (2017) 383–389, https://doi.org/10.1016/j.anres.2017.12.002.

[40] A.A. Chaugule, S.N. Mali, Identification of paddy varieties based on novel seed angle features, Comput. Electron. Agric. 123 (2016) 415–422, https://doi.org/10.1016/j.compag.2016.03.012.

[41] M. Vidal, J.M. Amigo, Pre-processing of hyperspectral images. Essential steps before image analysis, Chemometr. Intell. Lab. Syst. 117 (2012) 138–148, https://doi.org/10.1016/j.chemolab.2012.05.009.

[42] P. Geladi, Chemometrics in spectroscopy. Part 1. Classical chemometrics, Spectrochim. Acta Part B At. Spectrosc. 58 (2003) 767–782, https://doi.org/10.1016/S0584-8547(03)00037-5.

[43] The MathWorks Inc., MATLAB Version 9.4.0.813654(R2018a), MathWorks Inc., 2018.