MULTICLASS CLASSIFICATION ON SOYBEAN AND WEED SPECIES USING A CUSTOMIZED GREENHOUSE ROBOTIC AND HYPERSPECTRAL COMBINATION SYSTEM

Mohammed Raju Ahmed[1], Billy Ram[1], Cengiz Koparan[1], Kirk Howatt[2], Yu Zhang[1], Xin Sun[1,*]

[1] Department of Agricultural and Biosystems Engineering, North Dakota State University, Fargo, USA.
[2] Department of Plant Sciences, North Dakota University System, Fargo, USA.
[*] Correspondence: xin.sun@ndsu.edu

HIGHLIGHTS

- A novel combination of robotic and hyperspectral camera systems was customized to collect data in a greenhouse environment.
- Five popular weed species and crop (soybean) data were collected and analyzed in this research.
- Seven different types of hyperspectral data preprocessing methods were tested to obtain the best performance of the model.

ABSTRACT. *Soybean production is greatly affected by different types of weeds such as horseweed, kochia, ragweed, redroot pigweed, and waterhemp in the midwestern region of the United States. Identification of the soybean plants and the weeds is crucial to controlling the weeds in precision agriculture. The objective of this study was to classify soybean plants and five weed species where a hyperspectral imaging camera with a spectral range of 400 to 1000 nm was used to acquire the images. To acquire the HSI images, a customized robotic hyperspectral data collection scanning platform was developed and used in the greenhouse. A total of 983 hyperspectral data cubes were captured from the greenhouse environment (n = 252, soybean; n = 731, weeds). Spectral information was extracted from the collected images and then a classification model was developed by applying partial least squares discriminant analysis (PLS-DA). To construct the calibration and validation data set, the images were each divided into 70% and 30% ratios for model training and testing, respectively. Seven types of data preprocessing techniques, including mean normalization, maximum normalization, range normalization, multiplicative scatter correction (MSC), standard normal variate (SNV), Savitzky–Golay first derivatives, and Savitzky–Golay second derivatives, were explored individually and accepted as the best preprocessing method based on the highest performance in calibration and validation results. The results showed maximum validation model performance was found at 86.2% by applying the Savitzky–Golay second derivatives preprocessing method for multiclass and 89.4% for binary class. The most important wavelength information was evaluated from the beta coefficient developed using the same preprocessing method. Finally, chemical images were generated using a best-performer model to identify the soybean plants from weeds. The generated images showed a significant difference in chemical composition between soybean and weed plants at 443 nm, 633 nm, 743 nm, and 968 nm. The correlation between these peaks and the chemical components of the plants involves α-carotenoid, chlorophyll, and moisture, respectively. This study shows promising results for the application of HSI in weed control systems for soybeans and relevant weed identification in precision agriculture applications.*

Keywords. *Greenhouse robotic system, Hyperspectral imaging, Precision agriculture, Weed identification.*

Soybeans are considered a major crop all over the world. In 2020, global soybean production was 339 million metric tons (Anon, 2020). It has multiple uses for humans, such as vegetable oil extracted from the soybean, dairy products (soya milk), and bean curd (tofu); it is also an important food source for animals. As a result, maximum soybean productivity is critical not only for the agricultural economic context but also for the nutrient-rich food supply and security it provides for humans and animals. Weeds are recognized as unwanted plants in the agricultural crop fields that grow spontaneously with the economic crops (dos Santos Ferreira et al., 2017). It causes multiple negative effects on crops, such as rivalry for water, light, and soil nutrients; difficulty in harvesting operations; and increased disease and pest risk (Rizzardi and Fleck, 2004). As an overall consequence of these effects, the production and quality of the crops decreased, which caused a substantial loss to farmers and impacted the economy of the

Submitted for review on 30 March 2022 as manuscript number ITSC 15131; approved for publication as a Research Article by Associate Editor Dr. Chenghai Yang and Community Editor Dr. Naiqian Zhang of the Information Technology, Sensors, & Control Systems Community of ASABE on 1 August 2022.

  ISSN 2769-3295 https://doi.org/10.13031/ja.15131

country (Voll et al., 2005). Hence, weeds are identified as the number one problem in all major soybean producing countries (Vivian et al., 2013). According to estimates, 37% of soybean production losses occur due to the effects of weeds, while diseases and pests account for 22% of losses (Oerke and Dehne, 2004). In the northern U.S., the most common weeds that typically emerge in soybean fields are waterhemp, redroot pigweed, horseweed, kochia, and ragweed (Clay, 2013). The weeds are found in the field at or soon after soybean planting. Weed control management is crucial to ensure the maximum and highest quality soybean production.

Different types of weed control practices are used, such as chemical and/or herbicide sprays, biological practices where other plants and living things are put to use, and physical methods such as using cultivators and removing weeds manually (Suzuki, Okamoto et al., 2008). The last two practices require intense amounts of work, hard labor, and a lengthy time commitment. Therefore, using herbicides has become the most common method for weed control. Using high volumes of herbicides comes with a number of consequences, such as making the environment hostile and causing health hazards for mankind and animals. Thus, Site-Specific Weed Management (SSWM) is essential and involves identifying areas of weed density to apply herbicide (Eddy et al., 2014). This operation reduces herbicide requirements by up to 70%, which is both cost-effective and more environmentally friendly (Mortensen et al., 1995). For the implementation of SSWM and an automated weeding system, the identification of crops and weeds is the primary task where hyperspectral imaging (HSI) can be applied as a potential option.

Hyperspectral images contain information from tens or hundreds of wavebands with different resolution values (ElMasry and Sun, 2010). It is a powerful technique that integrates both the spectral and spatial information of an object. Image data from a color camera (RGB imaging) has high spatial resolution with only three wavebands; in the HSI technique, hyperspectral image data can be perceived as a three-dimensional datacube containing more wavebands while the spatial resolution is low (Chao et al., 2001). This new dimension is advantageous because it is useful for inquiring into the chemical information of the sample. In addition, hyperspectral systems produce data with very high spectral resolution compare to the traditional multispectral systems that are intended to improve the discrimination of samples (Gray et al., 2009; Narumalani et al., 2009). In recent studies, HSI has shown promise as a tool for crops and weed classification, such as two varieties of cabbage and five types of weed identification (Wei et al., 2015), wheat and canary grass classification (Kaur et al., 2015), cotton and weed detection (Alchanatis et al., 2005), weeds recognition in maize crops (Gao et al., 2018), and classification of two weed spices (redroot pigweed and wild oat) and three crop species (field pea, canola, and spring wheat) (Eddy et al., 2014). A high accuracy (87%) was obtained for classifying morningglory (*Ipomoea lacunosa*) in soybean (*Glycine max*) using HSI and linear discriminant analysis (Koger et al., 2003). HSI was applied for weed detection in soybean fields where the validation result was found to be more than 90% using linear discriminant analysis (LDA) (Suzuki et al., 2008). Later, for differentiating soybean and six weed species, HSI combined with principal component analysis (PCA) and LDA were applied, where LDA obtained 80% accuracy (Gray et al., 2009). More studies were found where HSI was successfully applied to classify weeds and soybean plants (Su, 2020; Fletcher, 2016; Fletcher and Reddy, 2016; Gray et al., 2008; Gibson et al., 2004). Overall, HSI was successfully applied for weed management and achieved ground-level plant species discrimination.

The main objective of this study was to evaluate the potential of using HSI where a customized greenhouse robotic platform was used for image acquisition of soybean plants and five types of common weeds, namely horseweed, kochia, ragweed, redroot pigweed, and waterhemp found in the northern area of the USA. Multivariate data analysis was performed to analyze the HSI data, and several preprocessing techniques were probed to acquire the highest result. The data was collected from a typical greenhouse condition in order to develop a general model that can be applied in the real world; the customized robotic platform was meant to be applied to real world issues as well. The findings of this research will be beneficial in developing a smart weed management system.

Materials and Methods

Soybean and Weed Sample Preparation

The soybean seeds were collected from the local market in Fargo, North Dakota, USA. Five species of weed seeds (horseweed, kochia, ragweed, redroot pigweed, and waterhemp) were collected from the Waldron greenhouse at North Dakota State University. Plastic pots (3.5(L) × 3.5(W) × 5(H) inches) were used to plant the seeds, and 2 to 3 of the same variety of seeds were planted in each pot filled with soil purchased from the local market. Later, the pots were placed on a plastic pot holder tray (21(L) × 10.5(W) ×3.5 (H) inches) where each tray contained 18 pots. Then, all the pot holders were kept on the tables inside a greenhouse in a random mixing condition (Waldron greenhouse at North Dakota State University) during the spring of 2021. A total of 150 pots were used for soybeans and 100 pots for individual weed species. Normal watering was conducted to reduce the dryness of the soil. The light condition was followed by the natural day and nighttime. After 21 days of planting, when the plant heights were found 10 to 12 cm on average, image acquisition was started. Weed management is most important during the young age of the crop since early-stage plants suffer for water and nutrients as they compete with the weeds. To evaluate the importance of the statement, the images were collected when the soybean plant was at a young age. Figure 1 shows the crop and weed samples used for this study.

Hyperspectral Image Acquisition

A lab-based visible-near-infrared (Vis-NIR) hyperspectral system was used in reflectance mode for image acquisition. The wavelength range of the HSI sensor was 400 to 1000 nm. The main parts of the system were: (1) the camera

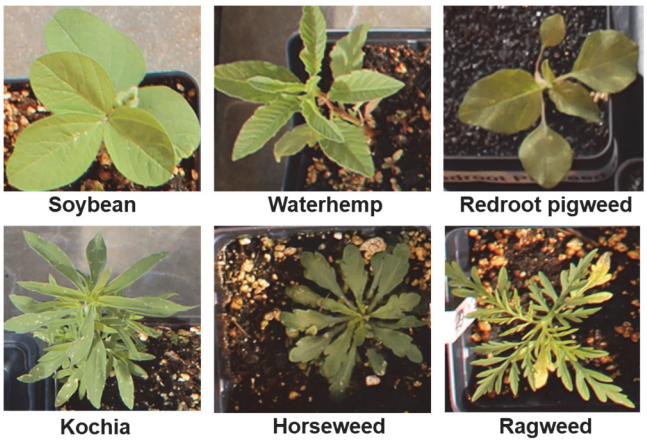


Figure 1. Soybean plant and five weed species.

(Specim FX 10e, Middleton Spectral Vision, Middleton, WI, USA); (2) the illumination system (4 halogen bulbs, 25 watt each); (3) the customized Farmbot frame (model: Genesis v1.6, *FarmBot Inc, San Luis Obispo, CA, USA*) for holding the camera for scanning; and (4) the Lumo Scanner software, which allows the user to set up, adjust, and acquire data for the supported Specim spectral camera and the customized HSI camera mount to set the scanning speed. Figure 2 illustrates the experiment system. The outside lighting conditions greatly affect the HSI imaging quality. In the greenhouse, natural sunlight and the greenhouse illumination system were available; therefore, the lighting conditions were too scattered and uneven. To protect the spectral information of the sample from the outside light effect, the HSI camera and the image acquisition area was covered by a light-insulating cover (fig. 2). During image acquisition, the halogen lights attached to the system produce enough light to capture the proper image without interference from external lights.

The original platform, including the Farmbot component, was designed as a data repository and has a computer numerical control (CNC) function controlled by software (fig. 3a). For collecting image data in this research, the frame of the robot was first customized based on the previous work from the group (fig. 3b) (Costa et al., 2022), and then updated with the new version, which included lighting system improvement and cover design for better image quality (fig. 2). In general, the HSI camera captures the image by line scanning, and during the scan, the camera remains static while the sample moves under the Field of View (FOV) of the camera. However, in this study, samples remained on the table and were difficult to move during image acquisition. Therefore, the robotic platform was customized to move the camera and scan the samples. This procedure also helps to scan a large area with multiple samples in an image. Figure 3 shows the customization of the robotic platform.

As each pot contains 2 to 3 plants, a total of 252 soybean plants were scanned from 150 pots. Weeds were also scanned following the same procedure and found to be horseweed-149, kochia-156, ragweed-151, redroot pigweed-118, and waterhemp-157, for a total of 731 samples. However, some of the soybeans and weeds found had not germinated. As a result, the overall sample number count was adjusted to account for these plants. To aid the imaging process, all images were subjected to a comprehensive set of image interrogation and analysis tools. A total of 1950 scans on 224 bands were performed to complete each image. Considering the signal-to-noise ratio, the conveyor platform speed was set to 8 mm/s and the distance between the sample and the camera lens was 25 cm. Due to the plant height variations, some of the samples became slightly blurry even after applying the proper focusing procedure. The frame rate was set at 50 Hz with an exposure time of 10 ms. The binning value of 2 was applied for spectral and spatial during scanning. For system calibration, a Teflon bar (20 × 2 cm) with ~99% reflectance (white) area was first placed on the

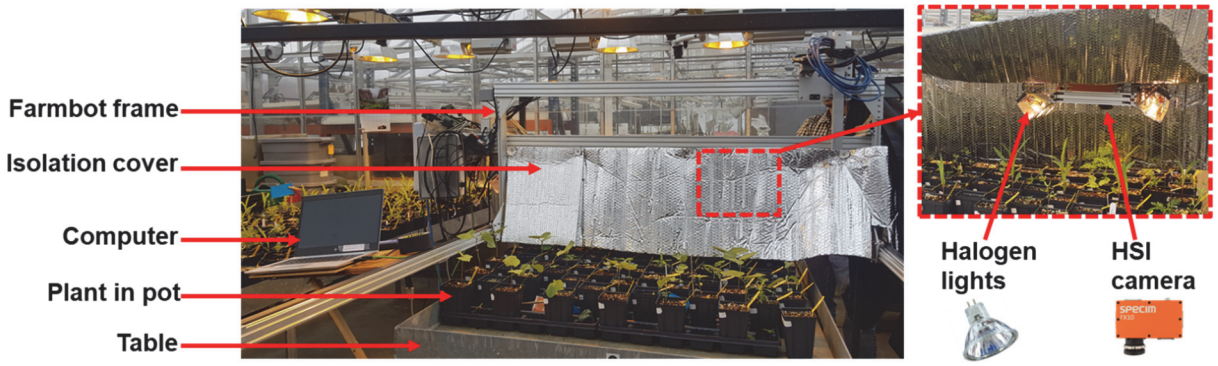


Figure 2. Demonstrations of HSI system used for soybean and weeds in greenhouse environment.

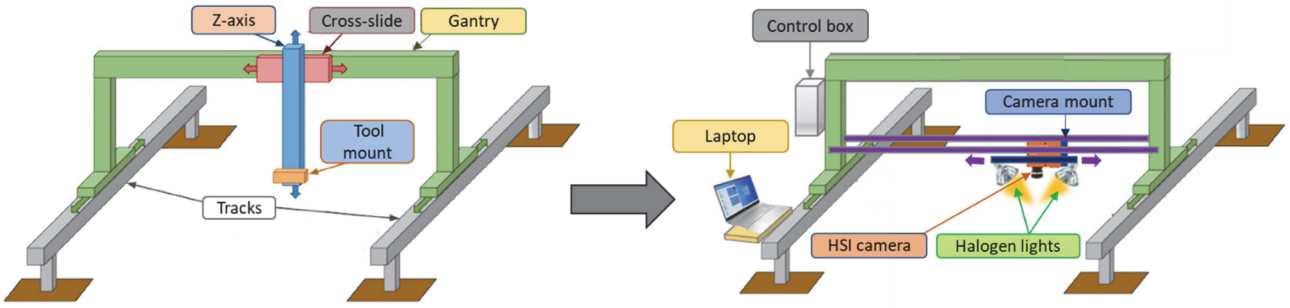


Figure 3. (a) original and (b) customized schematic diagram of Farmbot.

scanning bed and captured as a calibration image. The following equation was used to calibrate the hyperspectral images:

$$I = \frac{I_0 - D}{W - D} \quad (1)$$

where

I = the calibrated image
I_0 = the raw hyperspectral image
W = mean value of the white reference
D = mean value of the dark reference.

Spectral Data Extraction

Due to illumination light heat and ambient environmental conditions, some random spatial noises were generated during image acquisition. A median filter (window size: 3 × 3) was applied to each band image to reduce the noise and produce a clean image. Later, the region of interest (ROI) was manually selected using the 933 nm band image and spectral information was extracted from the respective samples' HSI images (Arngren et al., 2011). The mean of the collected spectral data was saved in a plain text format (.txt) according to the sample's types. Later, the similar types of sample data were accumulated and converted into an XLSX file for classification operation.

Preprocessing of the Raw Data

Preprocessing is an important task to reduce the effect of asymmetries in the spectral data, which may occur due to noise, light scattering, sample texture, and ambient temperature. Preprocessing improves the data quality to enhance the model performance. In this study, eight different types of preprocessing were applied to the spectral data: mean normalization, maximum normalization, range normalization, multiplicative scatter correction (MSC), standard normal variate (SNV), Savitzky–Golay first derivatives, Savitzky–Golay second derivatives, and data smoothing. Normalization preprocessing fits the spectral data into a similar range, while SNV and MSC remove slope differences and unwanted scatter effects. Savitzky–Golay derivatives improved the spectral resolution by removing the baseline offsets and overlapping peaks. Sometimes, the noise arising from the instruments can be resolved by using a data smoothing filter (Wakholi et al., 2018; Lasch, 2012; Vidal and Amigo, 2012; Rinnan et al., 2009; Maleki et al., 2007).

Model Development

The objective was to develop a model that could classify individual types of crops and weeds. To achieve the aim, partial least squares discriminant analysis (PLS-DA) was employed in this study. PLS-DA has proven to be a very versatile method for multivariate data analysis, and the number of applications is steadily increasing in research fields like machine learning and chemometrics (Mehmood et al., 2012). It is a supervised method developed to address the problem of making accurate predictions in multivariate problems (Martens and Naes, 1992). PLS-DA was implemented as a multivariate analysis and regression method to determine the linear models of prediction between the spectral data (X – *matrix,* $N_{samples} \times K_{wavelengths}$) and the values of the parameters obtained from the reference measurement (Y – *matrix,* $N_{samples} \times 1$). The general expression of PLS-DA is:

$$Y = X \times b + E \quad (2)$$

where

X = n × p that holds the spectral values of each class
b = regression coefficient
E = the error term.

The linear relationship between X and Y is predicted using equations (3) and (4).

$$X = TP^T + E_X \quad (3)$$

$$Y = UQ^T + E_Y \quad (4)$$

where Y is the matrix of dependent variables, and X is the $n \times p$ matrix of independent variables corresponding to the spectral variables for each hyperspectral measurement. The matrix X decomposes into the loading matrix P, score matrix T, and error matrix E_X. The matrix Y decomposes into the loading matrix Q, score matrix U, and error matrix E_y. Furthermore, the entire X and Y matrix data were divided into calibration and validation sets, which consisted of 70% of the data for calibration and 30% for validation (Uyeh et al., 2021; Faqeerzada et al., 2020; Ryan and Ali, 2016).

Chemical Image Visualization

From the studied data preprocessing methods, the accepted method was selected based on the highest accuracy. Afterwards, the chemical image was generated for the visualization of soybean plants and weeds by the multiplication of the beta coefficients and the original masked HSI image.

The hyperspectral image was unfolded into a two-dimensional (2D) matrix and then multiplied by the regression (beta) coefficient obtained from the best calibration model and applied to the selected wavelengths. The resultant vector was then folded back to form a 2D image. A median filter of 3 × 3 was applied to the 2D image to enhance the image quality for visual display. The difference between the predicted attributes within one sample and those from other sources can be visualized from the generated 2D images. Except for HSI image acquisition, all of the operations were executed by using MATLAB (2020a, MathWorks, Natick, MA, USA) based on a custom-built algorithm.

Results and Discussion

Spectral Data Explanation for Crop and Weed HSI Images

Raw spectra were collected from each plant from 397 to 1004 nm (224 wavebands). Using the original sample spectra from each individual group, the mean spectra were generated and presented in figure 4a. The spectral pattern was found to be different in soybean compared to the weed group, where soybean plant spectra showed a high intensity pick from the 700 to 950 nm range. However, another high intensity was observed between 530 and 580 nm in the original spectra. After applying different preprocessing techniques to the raw data set, Savitzky–Golay second derivatives showed a clearer spectral difference in soybean (fig. 4b). The spikes were observed almost in the same region as the original spectra, but the picks were sharp and clear. The spectral reflectance is highly corelated to α-carotenoid, anthocyanin, chlorophyll, and moisture (Su, 2020).

Multiclass PLS-DA Model Results for Crop and Weed Classification

Considering the six classes of the samples, PLS-DA was performed to develop the model. PLS-DA is a supervised classifier which is suitable for this study because the predictors had more variables than the observations and a high level of correlation was present among the original predictors (Wold et al., 2001). Before applying the PLS-DA, an unsupervised method called principal component analysis (PCA) was tested. PCA performed poorly, and no cluster group was observed in the PCA result plot. To identify the outliers in the data set, Hoteling's T^2 ellipse and Q statistic were used, where Hoteling's T^2 measures the variation in each sample within the model and the Q statistic evaluates how each sample conforms to the model (Yasmin et al., 2019). No outliers were detected during this operation.

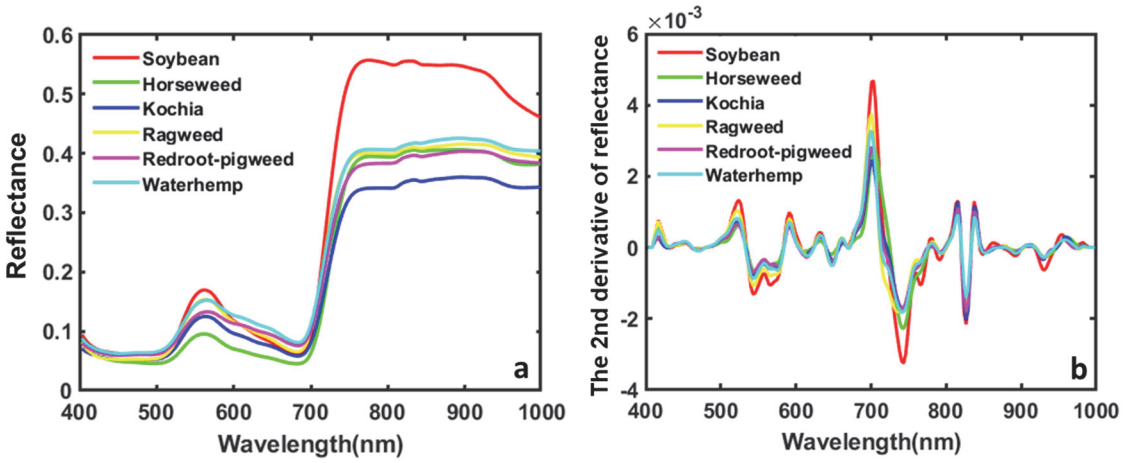
A total of seven types of preprocessing were tested to obtain the highest accuracy. After applying all the preprocessing methods, the classification results of PLS-DA are presented in table 1. Savitzky–Golay second derivatives performed the best, producing the highest overall accuracy of 86.2% using ten latent variables, as Savitzky–Golay pre-processed spectra always give more accurate spectral information (Kandpal et al., 2015). Latent variables were calculated from the prediction set of the classification model considering the lower error rate, which is a common practice to evaluate the classification model performance (Ballabio and Consonni, 2013).

Figure 5a shows the PLS-DA result using the Savitzky-Golay 2nd derivation preprocessing method with a good prediction accuracy of 86.2% using 10 latent variables. Some of the samples were misclassified as soybean plants, especially from the horseweed and ragweed groups. This may occur due to the temperature difference during spectral acquisition, vibration of the sample, light scattering, and chemical

Table 1. Multiclass PLS-DA model classification results applying different preprocessing methods.

Preprocessing methods	Calibration accuracy (%)	Prediction accuracy (%)	LVs[a]
Mean normalization	88.2	82.6	15
Max normalization	81.5	77.4	15
Range normalization	82.4	77.8	15
MSC[b]	81.8	75.1	15
SNV[c]	80.3	76.2	15
Savitzky–Golay 1st[d]	86.4	81.6	9
Savitzky–Golay 2nd[e]	91.2	86.2	10
Raw[f]	74.7	68.4	12

[a] LVs: Latent variables
[b] MSC: Multiplicative Scatter Correction
[c] SNV: Standard Normal Variate
[d] Savitzky-Golay 1st derivation
[e] Savitzky-Golay 2nd derivation
[f] Raw: Raw data model



(a) (b)

Figure 4. (a) raw mean spectra of soybean and weeds and (b) mean spectral after applying Savitzky–Golay second derivatives.

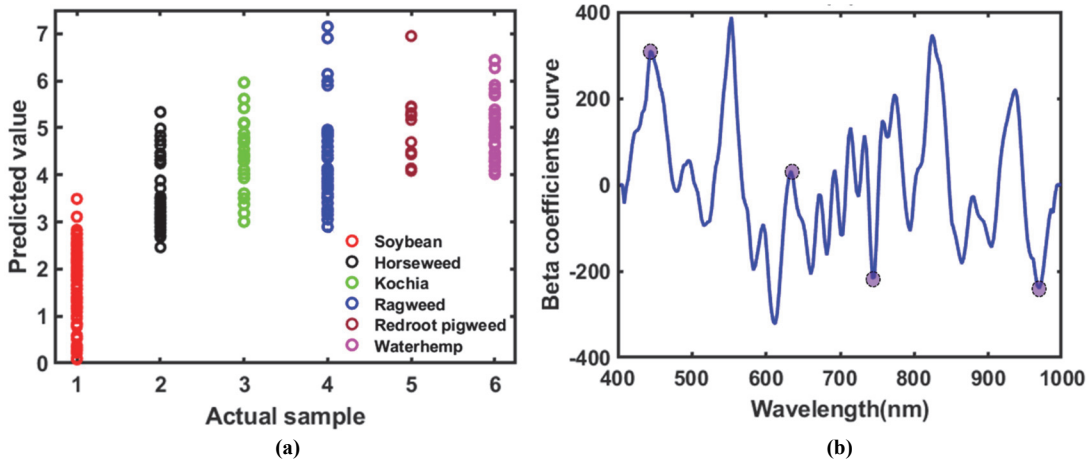


Figure 5. (a) PLS-DA classification results for soybean and five weed species applying Savitzky-Golay 2nd derivation preprocessing method. (b) Beta coefficient curve derived from the PLS-DA model using Savitzky–Golay second derivatives preprocessing methodology. Significant wavelengths for distinguishing soybean plants and weeds are marked with violet circles.

composition similarities. The beta coefficient curve provides significant information related to the particular peak difference and wavelength selection, which is important to minimize the excessive information from the HSI images and optimize the final classification system (Cheng and Sun, 2015). Overtones and combinations of fundamental vibrations of molecules comprised of -OH, -NH, and -CH groups are characterized in the near-infrared (NIR) spectrum. These would absorb based on the chemical composition of the samples (Workman, Jr. and Weyer 2007).

The spectral reflectance is also affected by the plant cell structure and the physical form of the plant surface (Vogelmann, 1989). The highest absolute values of the beta coefficient were observed at 443 nm and 743 nm (α-carotenoid), 633 nm (chlorophyll), and 968 nm (moisture), which are displayed in figure 5b. The correlation between these peaks and chemical components of the plants is α-carotenoid, chlorophyll, moisture (Zwiggelaar, 1998; Gausman et al., 1981; Su, 2020). Soybeans and weeds contain these ingredients; however, their concentration varies greatly among different plants, which causes the vibration amplitude of spectral absorption to be different. At a young age, the soybean leaves are bigger in size and softer compared to the weed leaves. The moisture content and chlorophyll content might be higher and due to these differences, peaks were found at 633 nm and 968 nm which are particularly responsible for C-H and O-H overtones corelate with chlorophyll and moisture for the used samples. Several studies have identified important spectral wavebands ranging from 350 to 2500 nm for the spectral characterization of plant and weed species (Eddy et al., 2014; Thenkabail et al., 2004; Lewis, 2002) and the observed spectral bands were found to be similar to the previous studies. The model used only 10 latent variables to achieve the best result, which is advantageous because using a limited number of latent variables reduces calculation time, which is important when applying the model in the real world.

Binary PLS-DA Model Results for Crop and Weed Classification

In the multiclass classification, a significant difference was observed between the soybean plants and the weeds. Therefore, a binary classification was developed which is potential to discriminate the soybean plants and any weeds from the data set. As a result, the model requires fewer calculations, making the operation faster, which is advantageous when transferring the model to an online hyperspectral system. The preprocessing methods remained same for binary class observation and among them Savitzky-Golay 2nd derivation preprocessing method performed the best with 89.4% accuracy. Figure 6 shows the summary of the binary class model results. A significant intensity difference was found for soybean and weeds both for raw (fig. 6a) and preprocessed data sets (fig. 6b) which is similar to the multiclass classification. In the classification figure 6c showed some of the soybean and weeds were misclassified. Beta coefficient curve (fig. 6d) indicated most two important wavebands at 443 nm and 743 nm, and 968 nm which are correlated to α-carotenoid and chlorophyll. The results of used preprocessing methods are recorded in table 2. The higher accuracy was obtained by developing binary class PLS-DA model.

Chemical Contents Analysis Results for Crop and Weed HSI Image

Every pixel of a hyperspectral image contains a unique spectrum. Therefore, soybeans and weeds can be identified by the unique spectrum of any pixel in the sample. To generate the chemical image of the used samples, all spatial pixels of the hyperspectral image were considered as spectral data was extracted using the whole plants defined as a region of interest (ROI).

The images were developed by multiplying the obtained beta coefficient (regression coefficient) from the binary class PLS-DA model with the spectra of each pixel in the image. The chemical mapping image offers rapid and easy access to the spatial distributions in which the relative intensities are indicated by the color bar. In the chemical image (fig. 7b), the soybean plants showed high intensity (yellow color) compared to weeds (blue color). A binary classification model was generated where the soybean plants are clearly visible compared to the weeds (fig. 7c). These acquired distribution maps validate the benefits of HSI in analyzing heterogeneous samples like crops and weeds. However, in some

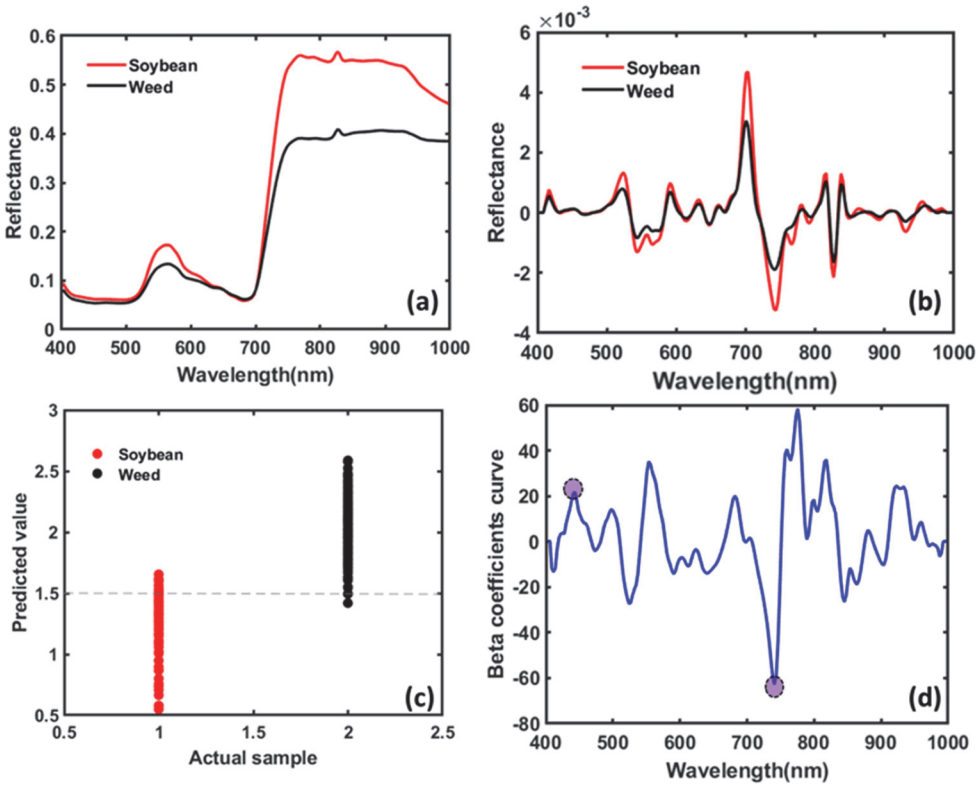


Figure 6. (a) raw mean spectra of the soybean and weeds, (b) mean spectral after applying Savitzky–Golay second derivatives, (c) binary classification results for soybean and weed species, and (d) beta coefficient curve derived from the PLS-DA model. Significant wavelengths for distinguishing soybean plants and weeds are marked with violet circles.

Table 2. Binary PLS-DA model classification results applying different preprocessing methods.

Preprocessing methods	Calibration accuracy (%)	Prediction accuracy (%)	LVs[a]
Mean normalization	90.3	84.4	14
Max normalization	83.7	81.4	15
Range normalization	81.5	77.7	15
MSC[b]	88.6	84.1	13
SNV[c]	81.2	75.8	15
Savitzky–Golay 1st [d]	88.7	83.3	10
Savitzky–Golay 2nd [e]	93.4	89.4	10
Raw[f]	77.6	71.3	15

[a] LVs: Latent variables
[b] MSC: Multiplicative Scatter Correction
[c] SNV: Standard Normal Variate
[d] Savitzky-Golay 1st derivation
[e] Savitzky-Golay 2nd derivation
[f] Raw: Raw data model

parts of the soybean plants, blue color was observed to have an effect on the model preference, which is also observed in the binary image. The possible reason for this misclassification might be the chlorophyll and moisture content similarities between the soybean and weed plants.

The data collection was conducted in a natural greenhouse condition, which was beneficial for developing a general model. The customized robotic platform is able to cover a substantial area during the image scanning by the HSI camera compared to a lab-based scanner. Therefore, a significant number of plants and weeds were captured in a single image, which made the image acquisition procedure faster and less laborious. For the real-world application, this feature is considered advantageous for scanning large areas. Multispectral imaging was used for soybean and two types of weed (giant foxtail and velvetleaf) detection by Gibson et al. (2004) and obtained 83% accuracy. In this experiment, an additional number of functional wavebands were used to investigate more spectral information and performance increased by 4% where the classification group increased. Another study was conducted using a portable hyperspectral imaging system (wavelength range: 360 to 1010 nm) for soybean and four types of weeds in Japan (Suzuki et al., 2008), where 30 samples were used and 15 wavebands were selected for model development. In this study, the sample number and waveband number were significantly increased to build a robust model. The image quality was also improved by using a static, customized robotic platform. A study was performed for maize plants and weed detection where only one-class classification was tested (Pantazi et al., 2016). Zhang et al. (2012) applied HSI for weed mapping in the tomato field and the spectral range was used from 384 nm to 810 nm. In this study, both multiclass and one-class classification were tested with a higher spectral range that was advantageous for revealing more important spectral information.

The presented algorithm was developed to build the model for identifying soybean plants and weeds. However, for online use of the system, an additional algorithm development is the next step of this research. The speed of the image processing and decision-making time of the model is still rather low due to the hyperspectral image size and image features. To overcome this situation, a high-performing

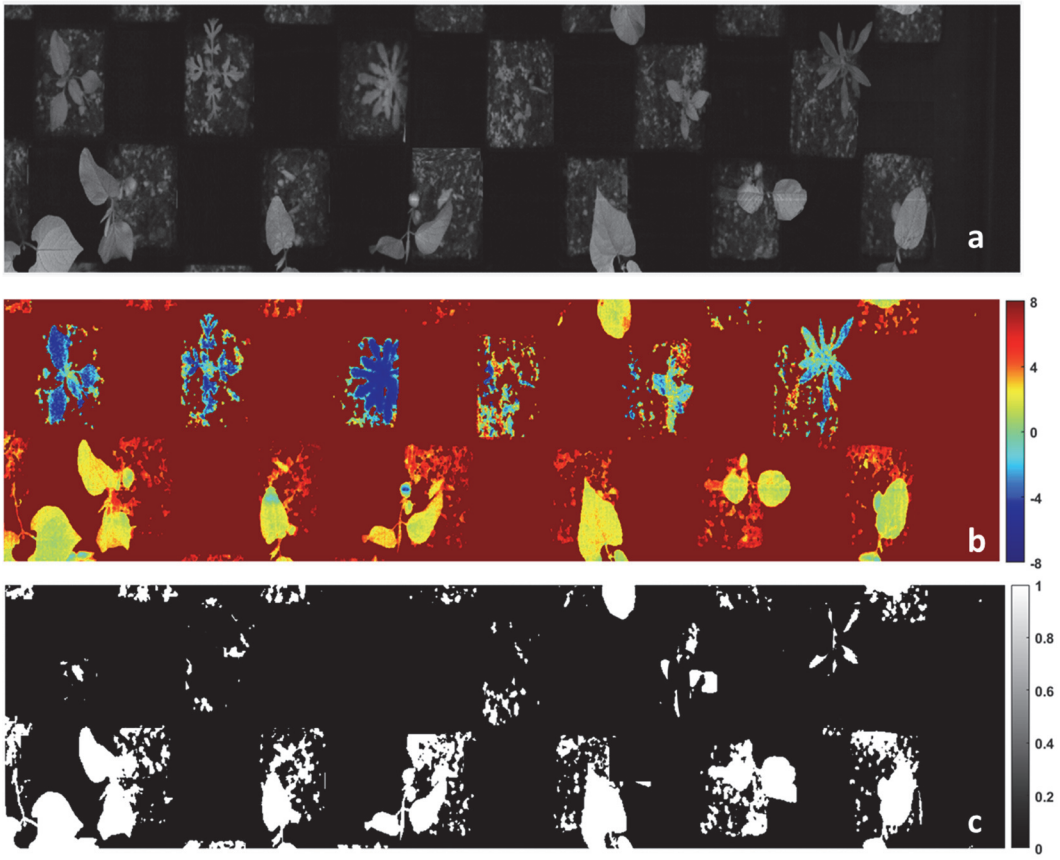


Figure 7. Generated chemical image to distinguish soybean plants (yellow) and weeds (blue) using Vis-NIR hyperspectral images from beta coefficient curve. (a) Raw image, (b) chemical image, and (c) binary image.

computer can be used, which could reduce the processing time significantly. The customized robotic platform was designed to be used in the greenhouse and in the field for image acquisition. Therefore, model performance can be validated with real-field images. The results of this study showed the potential use of HSI for crop and weed identification, which can be used to develop an online hyperspectral imaging system for weed management in the domain of precision agriculture.

Conclusions

Soybean is an important crop all over the world. Therefore, weed management is crucial to produce the maximum production. In this study, HSI was used to identify soybean plants among five types of weeds where a semi-automatic robotic platform was used for image acquisition. In developing a multiclass PLS-DA model combined with Savitzky–Golay second derivatives, the highest accuracy was found to be 86.2%, and a higher accuracy (89.4%) was obtained for the binary class. The best wavebands were observed from the beta coefficient from 443 to 968 nm. The result of this study showed a potential solution to detect crops and weeds, and the technique can be used for the automation of weed management for further study in the farm fields.

Acknowledgments

This material is based upon work partially supported by the U.S. Department of Agriculture, agreement number 58-6064-8-023. Any opinions, finding, conclusions or recommendations expressed in this publication are those of the author(s) and do not necessarily reflect the view of the U.S. Department of Agriculture. This work is/was supported by the USDA National Institute of Food and Agriculture, Hatch project number ND01487.

References

Alchanatis, V., Ridel, L., Hetzroni, A., & Yaroslavsky, L. (2005). Weed detection in multi-spectral images of cotton fields. *Comput. Electron. Agric., 47*(3), 243-260. https://doi.org/10.1016/j.compag.2004.11.019

Arngren, M., Hansen, P. W., Eriksen, B., Larsen, J., & Larsen, R. (2011). Analysis of pregerminated barley using hyperspectral image analysis. *J. Agric. Food. Chem., 59*(21), 11385-11394. https://doi.org/10.1021/jf202122y

Ballabio, D., & Consonni, V. (2013). Classification tools in chemistry. Part 1: Linear models. PLS-DA. *Anal. Methods, 5*(16), 3790-3798. https://doi.org/10.1039/C3AY40582F

Chao, K., Chen, Y. R., Hruschka, W. R., & Park, B. (2001). Chicken heart disease characterization by multi-spectral imaging. *Appl. Eng. Agric., 17*(1), 99-106. https://doi.org/10.13031/2013.1926

Cheng, J.-H., & Sun, D.-W. (2015). Rapid quantification analysis and visualization of escherichia coli loads in grass carp fish flesh by hyperspectral imaging method. *Food Bioprocess Technol., 8*(5), 951-959. https://doi.org/10.1007/s11947-014-1457-9

Clay, S. A. (2013). Common broadleaf weeds of South Dakota. In D. E. Clay, C. G. Carlson, S. A. Clay, L. Wagner, D. Deneke, & C. Hay (Eds.), *iGrow Soybeans: Best management practices for soybean production.* Brookings, SD: South Dakota State University, SDSU Extension.

Costa C, Zhang Y, Howatt K, Ram B, Stenger J, Nowatzki J, Bajwa S, Sun X. Palmer amaranth (Amaranthus palmeri S. Watson) and soybean (Glycine max L.) classification in Greenhouse using hyperspectral imaging and Chemometrics Methods. Journal of the ASABE. 2022;65(1):179-88. doi: 10.13031/ja.14321

Dos Santos Ferreira, A., Matte Freitas, D., Gonçalves da Silva, G., Pistori, H., & Theophilo Folhes, M. (2017). Weed detection in soybean crops using ConvNets. *Comput. Electron. Agric., 143*, 314-324. https://doi.org/10.1016/j.compag.2017.10.027

Eddy, P. R., Smith, A. M., Hill, B. D., Peddle, D. R., Coburn, C. A., & Blackshaw, R. E. (2014). Weed and crop discrimination using hyperspectral image data and reduced bandsets. *Canadian J. Remote Sens., 39*(6), 481-490. https://doi.org/10.5589/m14-001

ElMasry, G., & Sun, D.-W. (2010). Chapter 1: Principles of hyperspectral imaging technology. In D.-W. Sun (Ed.), *Hyperspectral imaging for food quality analysis and control* (pp. 3-43). San Diego, CA: Academic Press. https://doi.org/10.1016/B978-0-12-374753-2.10001-2

Faqeerzada, M. A., Rahman, A., Kim, G., Park, E., Joshi, R., Lohumi, S., & Cho, B.-K. (2020). Prediction of moisture contents in green peppers using hyperspectral imaging based on a polarized lighting system. *Korean J. Agric. Sci., 47*(4), 995-1010. https://doi.org/10.7744/kjoas.20200083

Fletcher, R. S. (2016). Using vegetation indices as input into random forest for soybean and weed classification. *Am. J. Plant Sci., 7*(15), 2186-2198. https://doi.org/10.4236/ajps.2016.715193

Fletcher, R. S., & Reddy, K. N. (2016). Random forest and leaf multispectral reflectance data to differentiate three soybean varieties from two pigweeds. *Comput. Electron. Agric., 128*, 199-206. https://doi.org/10.1016/j.compag.2016.09.004

Gao, J., Nuyttens, D., Lootens, P., He, Y., & Pieters, J. G. (2018). Recognising weeds in a maize crop using a random forest machine-learning algorithm and near-infrared snapshot mosaic hyperspectral imagery. *Biosyst. Eng., 170*, 39-50. https://doi.org/10.1016/j.biosystemseng.2018.03.006

Gausman, H. W., Menges, R. M., Richardson, A. J., Walter, H., Rodriguez, R. R., & Tamez, S. (1981). Optical parameters of leaves of seven weed species. *Weed Sci., 29*(1), 24-26. https://doi.org/10.1017/S0043174500025777

Gibson, K. D., Dirks, R., Medlin, C. R., & Johnston, L. (2004). Detection of weed species in soybean using multispectral digital images. *Weed Technol., 18*(3), 742-749. https://doi.org/10.1614/WT-03-170R1

Gray, C. J., Shaw, D. R., & Bruce, L. M. (2009). Utility of hyperspectral reflectance for differentiating soybean (*Glycine max*) and six weed species. *Weed Technol., 23*(1), 108-119. https://doi.org/10.1614/WT-07-117.1

Gray, C. J., Shaw, D. R., Gerard, P. D., & Bruce, L. M. (2008). Utility of multispectral imagery for soybean and weed species differentiation. *Weed Technol., 22*(4), 713-718. https://doi.org/10.1614/WT-07-116.1

Kandpal, L. M., Lee, S., Kim, M. S., Bae, H., & Cho, B.-K. (2015). Short wave infrared (SWIR) hyperspectral imaging technique for examination of aflatoxin B1 (AFB1) on corn kernels. *Food Control, 51*, 171-176. https://doi.org/10.1016/j.foodcont.2014.11.020

Kaur, R., Mahey, R. K., & Kingra, P. K. (2015). Multispectral remote sensing to distinguish the little seed canary grass (*Phalaris minor*) from wheat crop under field conditions for environmental sustainability and precision weed management. In A. K. Singh, J. C. Dagar, A. Arunachalam, R. Gopichandran, & K. N. Shelat (Eds.), *Climate change modelling, planning and policy for agriculture* (pp. 57-66). New Delhi: Springer. https://doi.org/10.1007/978-81-322-2157-9_7

Koger, C. H., Bruce, L. M., Shaw, D. R., & Reddy, K. N. (2003). Wavelet analysis of hyperspectral reflectance data for detecting pitted morningglory (*Ipomoea lacunosa*) in soybean (*Glycine max*). *Remote Sens. Environ., 86*(1), 108-119. https://doi.org/10.1016/S0034-4257(03)00071-3

Lasch, P. (2012). Spectral preprocessing for biomedical vibrational spectroscopy and microspectroscopic imaging. *Chemom. Intell. Lab. Syst., 117*, 100-114. https://doi.org/10.1016/j.chemolab.2012.03.011

Lewis, M. (2002). Spectral characterization of Australian arid zone plants. *Canadian J. Remote Sens., 28*(2), 219-230. https://doi.org/10.5589/m02-023

Maleki, M. R., Mouazen, A. M., Ramon, H., & De Baerdemaeker, J. (2007). Multiplicative scatter correction during on-line measurement with near infrared spectroscopy. *Biosyst. Eng., 96*(3), 427-433. https://doi.org/10.1016/j.biosystemseng.2006.11.014

Martens, H., & Naes, T. (1992). *Multivariate calibration.* John Wiley & Sons.

Mehmood, T., Liland, K. H., Snipen, L., & Sæbø, S. (2012). A review of variable selection methods in Partial Least Squares Regression. *Chemom. Intell. Lab. Syst., 118*, 62-69. https://doi.org/10.1016/j.chemolab.2012.07.010

Mortensen, D. A., Johnson, G. A., Wyse, D. Y., & Martin, A. R. (1995). Managing spatially variable weed populations. In P. C. Robert, R. H. Rust, & W. E. Larson (Eds.), *Site-specific management for agricultural systems* (pp. 395-415). John Wiley & Sons. https://doi.org/10.2134/1995.site-specificmanagement.c27

Narumalani, S., Mishra, D. R., Wilson, R., Reece, P., & Kohler, A. (2009). Detecting and mapping four invasive species along the floodplain of North Platte River, Nebraska. *Weed Technol., 23*(1), 99-107. https://doi.org/10.1614/WT-08-007.1

Oerke, E. C., & Dehne, H. W. (2004). Safeguarding production—losses in major crops and the role of crop protection. *Crop Prot., 23*(4), 275-285. https://doi.org/10.1016/j.cropro.2003.10.001

Pantazi, X.-E., Moshou, D., & Bravo, C. (2016). Active learning system for weed species recognition based on hyperspectral sensing. *Biosyst. Eng., 146*, 193-202. https://doi.org/10.1016/j.biosystemseng.2016.01.014

Rinnan, Å., Berg, F. v., & Engelsen, S. B. (2009). Review of the most common preprocessing techniques for near-infrared spectra. *TrAC, Trends Anal. Chem., 28*(10), 1201-1222. https://doi.org/10.1016/j.trac.2009.07.007

Rizzardi, M. A., & Fleck, N. G. (2004). Métodos de quantificação da cobertura foliar da infestação de plantas daninhas e da cultura da soja. *Ciência Rural, 34*(1), 13-18. https://doi.org/10.1590/S0103-84782004000100003

Ryan, K., & Ali, K. (2016). Application of a partial least-squares regression model to retrieve chlorophyll-a concentrations in coastal waters using hyper-spectral data. *Ocean Sci. J., 51*(2), 209-221. https://doi.org/10.1007/s12601-016-0018-8

SoyStats. (2022). International: World Soybean Production. Retrieved from http://soystats.com/international-world-soybean-production/

Su, W.-H. (2020). Advanced machine learning in point spectroscopy, RGB- and hyperspectral-imaging for automatic

discriminations of crops and weeds: A review. *Smart Cities, 3*(3), 767-792. https://doi.org/10.3390/smartcities3030039

Suzuki, Y., Okamoto, H., & Kataoka, T. (2008). Image segmentation between crop and weed using hyperspectral imaging for weed detection in soybean field. *Environ. Control. Biol., 46*(3), 163-173. https://doi.org/10.2525/ecb.46.163

Thenkabail, P. S., Enclona, E. A., Ashton, M. S., & Van Der Meer, B. (2004). Accuracy assessments of hyperspectral waveband performance for vegetation analysis applications. *Remote Sens. Environ., 91*(3-4), 354-376. https://doi.org/10.1016/j.rse.2004.03.013

Uyeh, D. D., Kim, J., Lohumi, S., Park, T., Cho, B.-K., Woo, S.,... Ha, Y. (2021). Rapid and non-destructive monitoring of moisture content in livestock feed using a global hyperspectral model. *Animals, 11*(5), 1-17. https://doi.org/10.3390/ani11051299

Vidal, M., & Amigo, J. M. (2012). Preprocessing of hyperspectral images. Essential steps before image analysis. *Chemom. Intell. Lab. Syst., 117*, 138-148. https://doi.org/10.1016/j.chemolab.2012.05.009

Vivian, R., Reis, A., Kálnay, P. A., Vargas, L., Ferreira, A. C., & Mariani, F. (2013). Weed management in soybean—issues and practices. In H. A. El-Shemy (Ed.), *Soybean-pest resistance* (pp. 50-84). https://doi.org/10.5772/54595

Vogelmann, T. C. (1989). Penetration of light into plants. *Photochem. Photobiol., 50*(6), 895-902. https://doi.org/10.1111/j.1751-1097.1989.tb02919.x

Voll, E., Gazziero, D. L., Brighenti, A. M., Adegas, F. S., Gaudêncio, C. d., & Voll, C. E. (2005). *A dinâmica das plantas daninhas e práticas de manejo* (Vol. 260). Embrapa Soja Londrina.

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

Wei, D., Huang, Y., Chunjiang, Z., & Xiu, W. (2015). Identification of seedling cabbages and weeds using hyperspectral imaging. *Int. J. Agric. Biol. Eng., 8*(5), 65-72. https://doi.org/10.3965/j.ijabe.20150805.1492

Wold, S., Sjöström, M., & Eriksson, L. (2001). PLS-regression: A basic tool of chemometrics. *Chemom. Intell. Lab. Syst., 58*(2), 109-130. https://doi.org/10.1016/S0169-7439(01)00155-1

Workman Jr., J., & Weyer, L. (2007). *Practical guide to interpretive near-infrared spectroscopy* (1st ed.). Boca Raton: CRC press. https://doi.org/10.1201/9781420018318

Yasmin, J., Raju Ahmed, M., Lohumi, S., Wakholi, C., Kim, M. S., & Cho, B.-K. (2019). Classification method for viability screening of naturally aged watermelon seeds using FT-NIR spectroscopy. *Sensors, 19*(5), 1190. https://doi.org/10.3390/s19051190

Zhang, Y., Slaughter, D. C., & Staab, E. S. (2012). Robust hyperspectral vision-based classification for multi-season weed mapping. *ISPRS J. Photogramm. Remote Sens., 69*, 65-73. https://doi.org/10.1016/j.isprsjprs.2012.02.006

Zwiggelaar, R. (1998). A review of spectral properties of plants and their potential use for crop/weed discrimination in row-crops. *Crop Prot., 17*(3), 189-206. https://doi.org/10.1016/S0261-2194(98)00009-X