# Multiclass Classification on Soybean and Weed Species using A Novel Customized Greenhouse Robotic and Hyperspectral Combination System

**Mohammed Raju Ahmed[a], Billy Graham Ram[a], Cengiz Koparan[a], Kirk Howatt[b], Yu Zhang[a] Xin Sun[a*]**

[a]*Department of Agricultural and Biosystems Engineering, North Dakota State University, Fargo, ND 58108, USA.*

[b]*Department of Plant Sciences, North Dakota State University, PO Box 6050, Fargo, ND 58108-6050, USA.*

**Corresponding authors: E-mail address: xin.sun@ndsu.edu*

## ABSTRACT

Soybean production is greatly affected by different types of weeds such as horseweed, kochia, ragweed, redroot pigweed, and waterhemp in the Midwest region of U.S. Identification of the soybean plants and the weeds are crucial to control the weed in precision agriculture. The objective of this study was to classify soybean plants and 5 weed species where a hyperspectral imaging camera with a spectral range from 400-1000 nm was used to acquire the images. To acquire the HSI images, a customize 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 regression (PLSR) analysis. To construct the calibration and validation data set, the images were divided into 70% and 30% ratio for model training and testing, respectively. Eight types of data preprocessing techniques including 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 were explored individually and accepted the best preprocessing method based on the highest performance on calibration and validation results. The results showed maximum validation model performance was found 86.2% by applying Savitzky–Golay second derivatives preprocessing method. The most important wavelength information was evaluated from beta coefficient developed using the same preprocessing method. Finally, chemical images were generated using 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, 553 nm, 633 nm, 743 nm, and 968 nm. The correlation between these peaks and chemical components of the plants are α-caroteniod, anthocyanin, chlorophyll, and moisture respectively. This study shows promising results for the application of HSI in weed control system for soybean and relevant weeds identification in precision agriculture applications.

Keywords: greenhouse robotic system, hyperspectral imaging, precision agriculture, weed identification

## 1. Introduction

Soybean is considered as one of the major crops all over the world. The annual production of soybean was 339 million metric tons in 2020 globally (*International: World Soybean Production*, 2020). It has multiple use to the human such as vegetable oil extracted from the soybean, dairy products (soya milk), and bean curd (tofu) and animal consumptions. Therefore, maximum productivity of soybean is important not only for the agricultural economy context but also nutrient food supply and security to the mankind 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 increase the diseases and pest risk (Rizzardi & Fleck, 2004). As an overall consequence of these effects, the production and quality of the crops decreased that executes a substantial loss to the farmers and also impact on the country economy (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% soybean production occurs due to

weeds alone, while diseases and pests are accounted for 22% losses (Oerke & Dehne, 2004). In the northern U.S., the most common weeds that typically emerge in the soybean fields are waterhemp, redroot pigweed, horseweed, kochia, and ragweed (Clay et al., 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.

Weed control is practiced by means of chemical/ herbicides spray, biological where plants and other living things are used, and physical methods such as using cultivator and manually weed removing (Suzuki et al., 2008). The last two practices require intense amount work, hard labor, and long time. Therefore, using herbicides become the most common method for weed control. Using high volume of herbicides come with a number of consequences such as environment hostile and health hazards for mankind and animals. Thus, Site-Specific Weed Management (SSWM) is essential that involves identifying areas of weed density to apply herbicide (Eddy et al., 2014). This operation reduces herbicides requirements up to 70% which is cost effective and environment stewardship (Mortensen et al., 1995). For the implementation of SSWM and automated weeding system, 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 several wavebands with different resolution values (ElMasry & Sun, 2010). It is a powerful technique that integrates both the spectral and spatial information of the object. Image data from color camera (RGB imaging) is considered two-dimensional; in HSI technique, one new dimension called 'spectrum' is added and the hyperspectral image data can be perceived as a three-dimensional datacube (Chao et al., 2001). This new dimension is advantageous to inquire 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 tool for crops and weed classification such as two varieties of cabbages and five types of weeds identification (Wei et al., 2015), wheat and canary grass classification (Kaur et al., 2015), cotton and weed detection

(Alchanatis et al., 2005), weeds recognizing in maize crops (Gao et al., 2018) and classification on 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 to classify 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 field where the validation result was found more than 90% using linear discriminant analysis (LDA) (Suzuki et al., 2008). Later, for differentiating soybean and six weed species, HSI combine 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 & Fletcher, 2016; Fletcher & 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 with a customized greenhouse robotic platform for image acquisition and classification of soybean plants and five types of common weeds namely horseweed, kochia, ragweed, redroot pigweed, and waterhemp found in the northern area of 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 to develop a general model which can be applicable for real world application. The findings of this research will be beneficial to develop a smart weed management system eventually.

## 2. Materials and Methods

### 2.1. Soybean and weeds 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 Waldron greenhouse at North Dakota State University. Plastic pots (3.5(L)×3.5(W)×5(H) inch) were used to plant the seeds and 2/3 same verity 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) inch) 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. Total 150 pots were used for soybean 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 10-12 cm in average, image acquisition was started. **Fig. 1** shows the crop and weed samples used for this study.

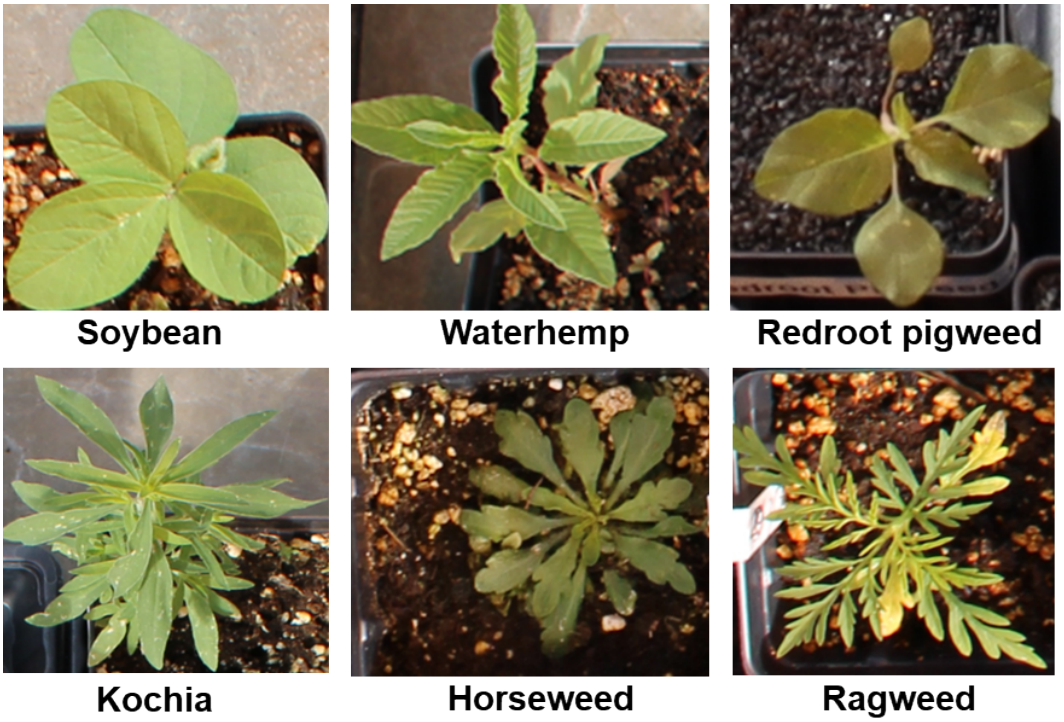



**Fig. 1** Soybean plants and five weed species

2.2. 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 400–1000 nm. The main parts of the system were: (1) the camera (Specim FX 10e, Middleton Spectral Vision, Middleton, WI, USA), (2) 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. **Fig. 2** illustrates the experiment system.

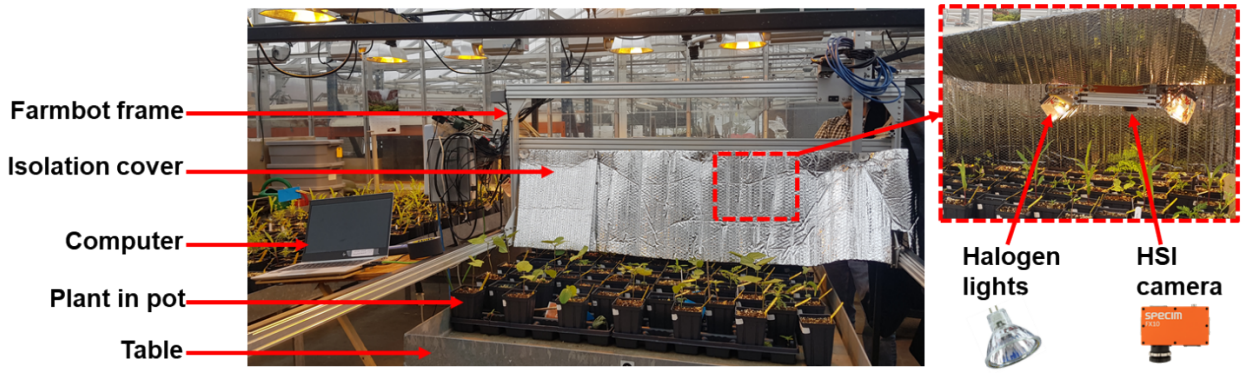



**Fig. 2** Demonstrations of the HSI system used for soybean and weeds in greenhouse environment.

The original Farmbot was designed for data repository of a farm. It is an open source precision agriculture computer numerical control (CNC) farming project. It is a cartesian coordinate robot farming machine with a software to control the robot. For collecting image data in the greenhouse, the frame of the robot was customized where a newly build frame was set with attaching a hyperspectral camera and halogen lights for illumination of the samples. In general, when HSI image is scanned, the camera remains fixed while sample keep moving and image is saved by line scanning. However, in this study, samples were remained on the table and strenuous 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. **Fig. 3** shows the customization of the robotic platform.

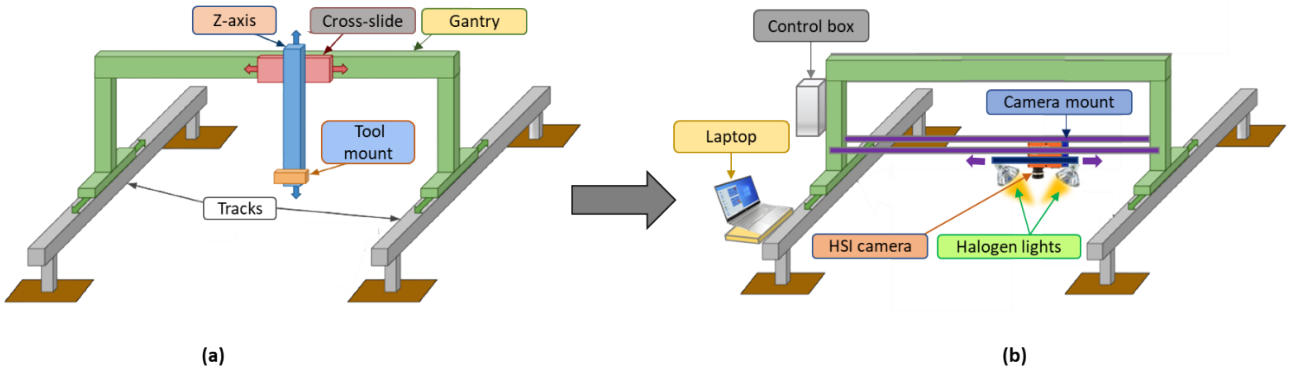


**Fig. 3** Schematic diagram of the robotic platform that was used in this experiment. **(a)** original version **(b)** customized version.

As each pot contains 2/3 plants, thus total 252 soybean plants were scanned from 150 pots. Weeds were also scanned following the same procedure and found horseweed-149, kochia-156, ragweed-151, redroot pigweed-118, waterhemp-157, in total 731 samples. However, some of the soybeans and weeds found not germinated. Therefore, sample number made alteration. All images were subjected to a complete set of image interrogation and analysis tools to aid the imaging process. Total 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 samples became slightly blur even after applying proper focusing procedure. The frame rate was set 50 Hz with exposure time 10 ms. The binning value 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 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 is the calibrated image, $I_0$ is the raw hyperspectral image, W represents a mean value of the white reference, and D represents a mean value of the dark reference.

## 2.3. Spectral data extraction

Due to illumination lights heat and ambient environmental conditions, some random spatial noises were generated during image acquisition. A median filter (window size: 3 × 3) was applied on each band image to reduce the noise and produce 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 txt format according to

the sample's types. Later, the similar types of samples data were accumulated and converted to xlsx file for classification operation.

## 2.4. Preprocessing of the raw data

Pre-processing 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. Pre-processing improves the data quality to enhance the model performance. In this study, eight different types of pre-processing 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 pre-processing fits the spectral data in 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 & Amigo, 2012; Rinnan et al., 2009; Maleki et al., 2007).

## 2.5. Model development

The objective was to develop a model that can classify individual types of crop and weeds. To achieve the aim, partial least squares regression (PLSR) analysis was employed in this study. PLSR 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 specifically established to address the problem of making good predictions in multivariate problems (Martens & Naes, 1992). PLSR was implemented as 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 linear relationship between X and Y is predicted using Equations (2) and (3).

$$X = TP^T + E_X \quad \text{..... (2)}$$

$$Y = UQ^T + E_Y \quad \text{..... (3)}$$

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 & Ali, 2016).

The efficiency of PLSR model was evaluated based on the calibration accuracy ($R^2C$), root-mean square errors of calibration (RMSEC), prediction accuracy ($R^2P$), and root-mean square errors of prediction (RMSEP). The coefficient of determination ($R^2$) was calculated using the equation (4). Inappropriate numbers of latent variables selection can cause under- or over-fitting, leading to suppression of spectral information, incorrect model interpretation, and spectral noise in the regression model. Consequently, the optimal number of latent variables was selected based on the lowest value of predicted root mean square error (RMSE) (equation 5). The RMSE value measures the accuracy of the model prediction and evaluate model performance in regression analysis (Yasmin et al., 2019).

$$R^2 = 1 - \frac{\sum_{i=1}^{z}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{z}(y_i - \bar{y}_i)^2} \quad \text{.....(4)}$$

$$RMSE = \sqrt{\frac{1}{z}\sum_{i=1}^{z}\left(y_i - \hat{y}_i\right)^2} \quad \text{.....(5)}$$

2.6. Chemical image visualization

From the studied data pre-processing 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 for enhancing image quality for visual display. The difference in 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 the operations were executed by using MATLAB (2020a, MathWorks, Natick, MA, USA) based on a custom-built algorithm.

## 3. Result and Discussion

### 3.1. Spectral data explanation for crop and weed HSI images

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

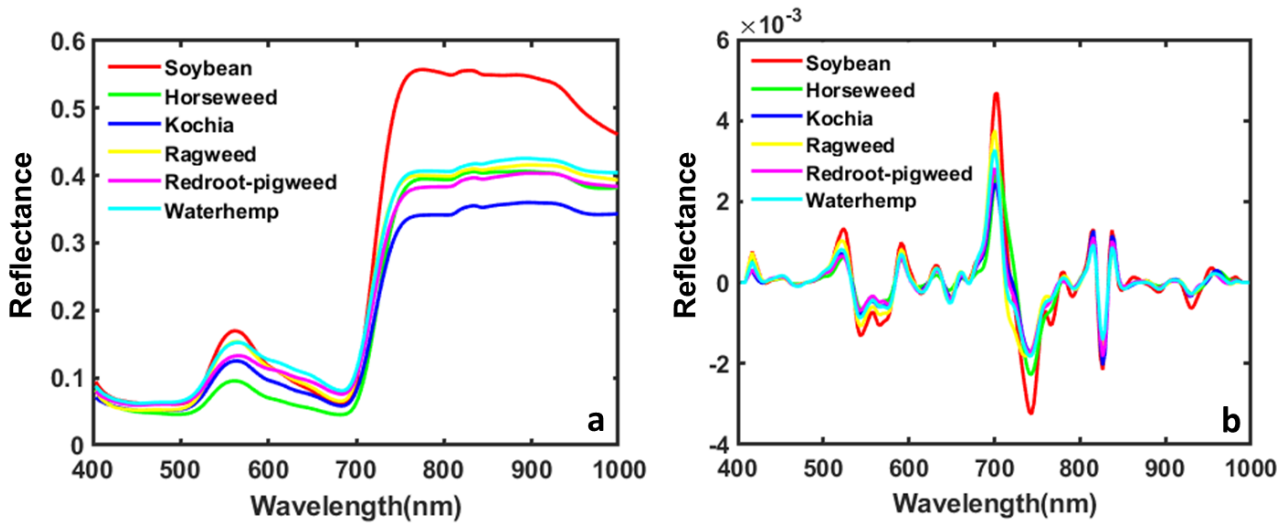



**Fig. 4** Raw mean spectra of the soybean and weeds (a) and mean spectral after applying Savitzky–Golay second derivatives (b).

### 3.2. PLSR model results for crop and weed classification

Considering the six class of the samples, PLSR multivariate data analysis were performed to develop the model. PLSR 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 PLSR, 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 identity 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 Q statistic evaluates that how each sample conforms the model (Jannat Yasmin et al., 2019). No outliers were detected during this operation.

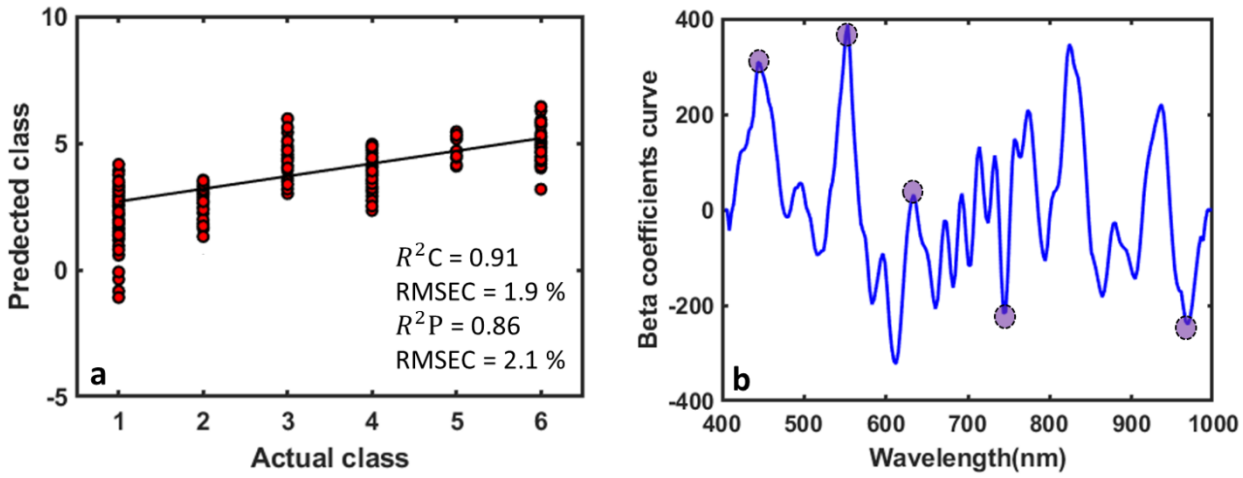
Total seven types of preprocessing were tested to obtain the highest accuracy. After applying all the preprocessing methods, the classification results of PLSR are presented in **Table 1**. The Savitzky–Golay second derivatives performed the best, producing the highest overall accuracy of 86.2% using 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 & Consonni, 2013). The least RMSEP value (2.1%) was also obtained using Savitzky–Golay second derivatives.

**Table 1.** PLSR model classification results applying different pre-processing methods.

| **Preprocessing methods** | **Calibration accuracy ($R^2C$ [a]) %** | **RMSEC [b] %** | **Prediction accuracy ($R^2P$ [c]) %** | **RMSEP [d] %** | **LVs [e]** |
|---|---|---|---|---|---|
| Mean normalization | 88.2 | 2.9 | 82.6 | 3.4 | 15 |
| Max normalization | 81.5 | 3.4 | 77.4 | 4.1 | 15 |
| Range normalization | 82.4 | 4.2 | 77.8 | 4.8 | 15 |
| MSC [f] | 81.8 | 3.7 | 75.1 | 4.3 | 15 |
| SNV [g] | 80.3 | 2.4 | 76.2 | 3.2 | 15 |
| Savitzky–Golay 1st [h] | 86.4 | 2.1 | 81.6 | 3.1 | 9 |
| Savitzky–Golay 2nd [i] | 91.2 | 1.9 | 86.2 | 2.1 | 10 |
| Raw [j] | 74.7 | 4.7 | 68.4 | 5.6 | 12 |

[a] $R^2C$: Coefficient of determination for calibration; [b] RMSEC: root-mean square error of calibration; [c] $R^2P$: Coefficient of determination for prediction; [d] RMSEP: root-mean square error of prediction; [e] LVs: Latent variables; Preprocessing methods ([f] MSC: Multiplicative Scatter Correction; [g] SNV: Standard Normal Variate; [h] Savitzky-Golay 1st derivation; [i] Savitzky-Golay 2nd derivation and [j] Raw: Raw data model.




(a) (b)

**Fig. 5** (a) PLSR classification results for soybean and five weed species (1- Soybean, 2- Horseweed, 3- Kochia, 4- Ragweed, 5- Redroot pigweed, and 6- Waterhemp) applying Savitzky-Golay 2nd derivation preprocessing method. (b) Beta coefficient curve derived from the PLS-DA model using Savitzky–Golay second derivatives pre-processing methodology. Significant wavelengths for distinguishing soybean plants and weeds are marked with violet circles.

**Fig. 5 (a)** shows the PLSR result using Savitzky-Golay 2nd derivation preprocessing method with a good prediction accuracy ($R^2$) over 85% where most of the predicted values fall near the line of best fit. Some of the samples were observed far from the best fit line which may occur may be for the temperature difference during spectral acquisition or vibration of the sample. 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 & Sun, 2015). Overtones and combinations of fundamental vibrations of molecules comprising -OH, -NH, and -CH groups are characterized in the near-infrared (NIR) spectrum. These would absorb based on the chemical component of the samples (Workman, Jr. & 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, 553 nm, 633 nm, 743

nm, and 968 nm (**Fig. 5 (b)**). The correlation between these peaks and chemical components of the plants are α-caroteniod, anthocyanin, chlorophyll, and moisture respectively (Zwiggelaar, 1998; Gausman et al., 1981; Su, 2020). Soybean and weeds leave content these ingredients; however, their concentration vary greatly among different plants that causes the vibration amplitude of spectral absorption to be different. Soybean leaves are bigger in size compared to the weed leaves; therefore, the moisture content and chlorophyll content might be higher and due to these difference peaks were found in 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 bands range from 350 to 2500 nm for spectral characterization of plant and weed species (Eddy et al., 2014; Thenkabail et al., 2004; Lewis, 2002) and the observed spectral bands were found similar with the previous studies.

### 3.3. Chemical contents analysis results for crop and weed HSI image

Hyperspectral image contents a unique spectrum in every pixel. Therefore, soybean and weeds can be identified by their 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 region of interest (ROI).

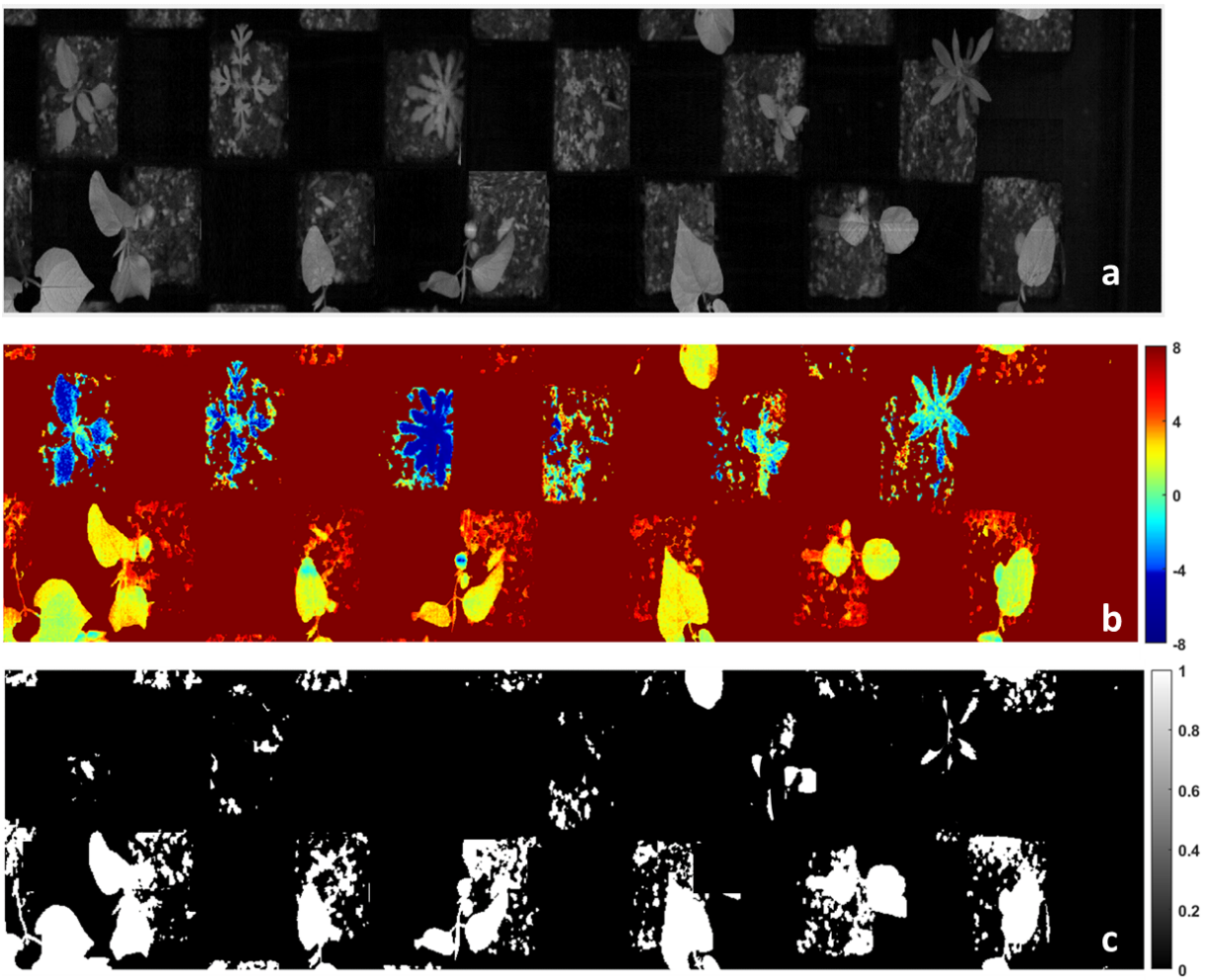



**Fig. 6** Generated chemical image to distinguish soybean plants (yellow) and weeds (blue) using Vis-NIR hyperspectral images of 633 nm. (a) Raw image, (b) chemical image, and (c) binary image.

The images were developed by multiplying the obtained beta coefficient (regression coefficient) from the PLSR 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. 6, row 2)**, 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. 6, row 3)**. These acquired distribution maps validate the benefits of HSI in analyzing heterogeneous samples like crops and weeds. However, in some part on the soybean plants, blue color was observed that effect on the model preference which is also observed in the binary image. The

283 possible reason of this misclassification might be the chlorophyll and moisture content similarities both the
284 soybean and weed plants.

285 The data collection was conducted in a natural greenhouse condition which was beneficial to develop a
286 general model. The customized robotic platform is able to cover a substantial area during image scanning
287 by the HSI camera compared to lab-based scanner. Therefore, a significant number of plants and weeds
288 were captured in a single image which made the image acquisition procedure faster and less laborious. For
289 the real-world application, this feature is considered as advantageous to scan large size area. Similar study
290 was conducted for maize and weeds detection (Pantazi et al., 2016) using 200 wavebands (range 435-834
291 nm). In this experiment, additional number of wavebands were used that was functional to investigate more
292 spectral information. The presented algorithm was developed to build the model for identifying Soybean
293 plants and weeds; however, for online use of the system, an additional algorithm development is the further
294 strategy of this research. The speed of the image processing and decision-making time of the model is still
295 rather low due to the hyperspectral image size and image features. To overcome this situation, a high
296 performing computer can be used which could reduce the processing time significantly. The result of this
297 study showed the potential use of HSI for crop and weed identification.

298 **4. Conclusions**

299 Soybean is an important crop for all over the world, therefore weed management is crucial to produce
300 the maximum production. In this study, HSI was used to identify the soybean plants among five types of
301 weeds where a semi-automatic robotic platform was used for image acquisition. Developing a PLSR model
302 combine with Savitzky–Golay second derivatives the highest accuracy was found 85.6%. The best
303 wavebands were observed from the beta coefficient from 443 nm to 968 nm. The result of this study showed
304 a potential solution to detect crop and weeds and the technique can be used for the automation of weed
305 management for the further study.

## 5. Acknowledgement

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. *Computers and Electronics in Agriculture*, *47*(3), 243–260. https://doi.org/10.1016/J.COMPAG.2004.11.019

Arngren, M., Waaben Hansen, P., Eriksen, B., Larsen, J., & Larsen, R. (2011). Analysis of Pregerminated Barley Using Hyperspectral Image Analysis. *Journal of Agricultural and Food Chemistry*, *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. In *Analytical Methods* (Vol. 5, Issue 16, pp. 3790–3798). Royal Society of Chemistry. 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. *Applied Engineering in Agriculture*, *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 and Bioprocess Technology*, *8*(5), 951–959. https://doi.org/10.1007/s11947-014-1457-9

Clay, D., Carlson, C., Clay, S., Wagner, L., Deneke, D., & Hay, C. (2013). Common Broadleaf Weeds of

South Dakota. In *iGrow Soybeans: Best Management Practices for Soybean Production* (pp. 241–266). SDSU Extension.

dos Santos Ferreira, A., Matte Freitas, D., Gonçalves da Silva, G., Pistori, H., & Theophilo Folhes, M. (2017). Weed detection in soybean crops using ConvNets. *Computers and Electronics in Agriculture*, *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 Journal of Remote Sensing*, *39*(6), 481–490. https://doi.org/10.5589/M14-001

ElMasry, G., & Sun, D. W. (2010). Principles of Hyperspectral Imaging Technology. In D.-W. Sun (Ed.), *Hyperspectral Imaging for Food Quality Analysis and Control* (pp. 3–43). Academic Press. https://doi.org/10.1016/B978-0-12-374753-2.10001-2

Faqeerzada, M., 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 Journal of Agricultural Science*, *47*(4), 995–1010. https://doi.org/10.7744/kjoas.20200083

*FarmBot, Open-Source CNC Farming*. (n.d.). Retrieved December 6, 2021, from https://farm.bot/

Fletcher, R. S., & Fletcher, R. S. (2016). Using Vegetation Indices as Input into Random Forest for Soybean and Weed Classification. *American Journal of Plant Sciences*, *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. *Computers and Electronics in Agriculture*, *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. *Biosystems Engineering*, *170*, 39–50. https://doi.org/10.1016/J.BIOSYSTEMSENG.2018.03.006

Gausman, H. W., Menges, R. M., Richardson, A. J., Water, H., Rodriguez, R. R., & Tamez, S. (1981). Optical Parameters of Leaves of Seven Weed Species. *Weed Science*, *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 Technology*, *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 Technology*, *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 Technology*, *22*(4), 713–718. https://doi.org/10.1614/WT-07-116.1

*International: World Soybean Production*. (2020). http://soystats.com/international-world-soybean-production/

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 *Climate Change Modelling, Planning and*

*Policy for Agriculture* (pp. 57–66). Springer, New Delhi. 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 Sensing of Environment*, *86*(1), 108–119. https://doi.org/10.1016/S0034-4257(03)00071-3

Lasch, P. (2012). Spectral pre-processing for biomedical vibrational spectroscopy and microspectroscopic imaging. *Chemometrics and Intelligent Laboratory Systems*, *117*, 100–114. https://doi.org/10.1016/j.chemolab.2012.03.011

Lewis, M. (2002). Spectral characterization of Australian arid zone plants. *Canadian Journal of Remote Sensing*, *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. *Biosystems Engineering*, *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. *Chemometrics and Intelligent Laboratory Systems*, *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, Ltd. 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 Technology*, *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 Protection*, *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. *Biosystems Engineering*, *146*, 193–202. https://doi.org/10.1016/j.biosystemseng.2016.01.014

Rinnan, Å., Berg, F. van den, & Engelsen, S. B. (2009). Review of the most common pre-processing techniques for near-infrared spectra. In *TrAC - Trends in Analytical Chemistry* (Vol. 28, Issue 10, pp. 1201–1222). Elsevier. 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 das plantas daninhas e da cultura de soja. *Ciência Rural*, *34*(1), 13–18. https://www.scopus.com/record/display.uri?eid=2-s2.0-33646739202&origin=inward

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 Science Journal*, *51*(2), 209–221. https://doi.org/10.1007/s12601-016-0018-8

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. *Environmental Control in Biology*, *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 Sensing of*

*Environment*, *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., Lee, W. 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). Pre-processing of hyperspectral images. Essential steps before image analysis. *Chemometrics and Intelligent Laboratory Systems*, *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. C., & Mariani, F. (2013). Weed management in soybean—issues and practices. In Hany El-Shemy (Ed.), *Soybean:Pest Resistance* (pp. 50–84). Intechopen.

Vogelmann, T. C. (1989). Penetration of light into plants. *Photochemistry and Photobiology*, *50*(6), 895–902. https://doi.org/10.1111/J.1751-1097.1989.TB02919.X

Voll, E., Gazziero, D., Brighenti, A., & Adegas, F. (2005). *A dinâmica das plantas daninhas e práticas de manejo*. Embrapa Soja. http://livimagens.sct.embrapa.br/amostras/00079260.pdf

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

Wei, D., Huang, Y., Chunjiang, Z., Xiu, W., B, H. Y., & J, Z. C. (2015). Identification of seedling cabbages and weeds using hyperspectral imaging. *International Journal of Agricultural and Biological Engineering*, *8*(5), 65–72. https://doi.org/10.25165/IJABE.V8I5.1492

Wold, S., Sjöström, M., & Eriksson, L. (2001). PLS-regression: a basic tool of chemometrics. *Chemometrics and Intelligent Laboratory Systems*, *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. In *CRC press* (1st ed.). CRC Press. https://doi.org/10.1201/9781420018318

Yasmin, J., Ahmed, M. R., Lohumi, S., Wakholi, C., Lee, H., Mo, C., & Cho, B.-K. (2019). Rapid authentication measurement of cinnamon powder using FT-NIR and FT-IR spectroscopic techniques. *Quality Assurance and Safety of Crops and Foods*, *11*(3). https://doi.org/10.3920/QAS2018.1318

Yasmin, Jannat, Ahmed, M. R., 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 (Switzerland)*, *19*(5). https://doi.org/10.3390/s19051190

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