Research Article

Received: 14 October 2021 Revised: 12 January 2022 Accepted article published: 7 March 2022 Published online in Wiley Online Library: 23 March 2022

(wileyonlinelibrary.com) DOI 10.1002/ps.6864

# Hyperspectral plant sensing for differentiating glyphosate-resistant and glyphosate-susceptible johnsongrass through machine learning algorithms

**Yanbo Huang,[a*] Xiaohu Zhao,[b] Zeng Pan,[b] Krishna N. Reddy[c] and Jingcheng Zhang[b]**

## Abstract

**BACKGROUND: Johnsongrass (*Sorghum halepense*) is one of the weeds that evolves resistance to glyphosate [*N*-(phosphonomethyl)-glycine], the most widely used herbicide, and the weed may cause agronomic troublesome in the southern USA. This paper reports a study on developing a hyperspectral plant sensing approach to explore the spectral features of glyphosate-resistant (GR) and glyphosate-sensitive (GS) plants to evaluate this approach using machine learning algorithms to differentiate between GR and GS plants.**

**RESULTS: On average, GR plants have higher spectral reflectance compared with GS plants. The sensitive spectral bands were optimally selected using the successive projections algorithm respectively wrapped with the machine learning algorithms of k-nearest neighbors (KNN), random forest (RF), and support vector machine (SVM) with Fisher linear discriminant analysis (FLDA) to classify between GS and GS plants. At 3 weeks after transplanting (WAT) KNN and SVM could not acceptably classify the GR and GS plants but they improved significantly with the stages to have their overall accuracies reaching 73% and 77%, respectively, at 5 WAT. RF and FLDA had a better ability to classify the plants at 3 WAT but RF was low in accuracy at 2 WAT while FLDA dropped accuracy to 50% at 4 WAT from 57% at 3 WAT and raised it to 73% at 5 WAT.**

**CONCLUSIONS: Previous studies were conducted developing the hyperspectral imaging approach to differentiate GR Palmer amaranth from GS Palmer amaranth and GR Italian ryegrass from GS Italian ryegrass with classification accuracies of 90% and 80%, respectively. This study demonstrated that the hyperspectral plant sensing approach could be developed to differentiate GR johnsongrass from glyphosate-sensitive GS johnsongrass with the highest classification accuracy of 77%. The comparison with our previous studies indicated that the similar hyperspectral approach could be used and transferred from classification across different GR and GS weed biotypes, such as Palmer amaranth, Italian ryegrass and johnsongrass, so it is highly possible for classification of more other GR and GS weed biotypes as well. On the basis of classic pattern recognition approaches the process of plant classification can be enhanced by modeling using machine learning algorithms.**


**Keywords:** glyphosate-resistant weed; johnsongrass; hyperspectral plant sensing; machine learning

## 1 INTRODUCTION

Various herbicides have been developed and used to control weeds in in crop fields to enhance crop productivity. Among the herbicides glyphosates, [*N*-(phosphonomethyl)-glycine] is the one which has been most widely used, especially intensive use in increasingly adopted genetically engineered glyphosate-resistant (GR) cropping systems in recent years[1] to ensure weed control and crop production at the same time. Based on the most recent available report, genetically engineered herbicide-tolerant crops account for about 56% of global glyphosate use, and in the USA no pesticide has come remotely close to such intensive and widespread use,[2] maybe because the nonselective feature of glyphosate to weeds and genetically engineered glyphosate-tolerant crops, such as corn, soybean and cotton, dominate genetically engineered herbicide-tolerant cropping systems. Such intense and widespread use of glyphosate has exerted a high selection pressure on weed populations and has resulted in the

* *Correspondence to: Y Huang, US Department of Agriculture, Agricultural Research Service, Genetics and Sustainable Agriculture Research Unit, Mississippi State, MI, USA. E-mail: yanbo.huang@usda.gov*

a *US Department of Agriculture, Agricultural Research Service, Genetics and Sustainable Agriculture Research Unit, Mississippi State, MS, USA*

b *Hangzhou Dianzi University, Hangzhou, China*

c *US Department of Agriculture, Agricultural Research Service, Crop Production Systems Research Unit, Stoneville, MS, USA*

evolution of GR weeds in crop fields. Based on the data extracted from the International Herbicide-Resistant Weed Database,[3] globally there are 51 GR weed biotypes that have been consecutively found since 1996. For the same time period in the USA 16 GR weed biotypes have been found consecutively. In Mississippi nine GR weed biotypes have been found so far.

Johnsongrass (*Sorghum halepense*) is a GR weed that was first reported in Argentina in 2005 in a soybean field, followed by more cases in Argentina in other soybean fields in 2015 and in Australia on fallow land in Queensland in 2019.[3] In the USA, GR johnsongrass was first reported in soybean fields in Arkansas in 2007 and in soybean fields in Mississippi and Louisiana, respectively, in 2008 and 2010.[3] Johnsongrass is a monocot weed in the Poaceae family. This biotype occurs in crop fields, pastures, abandoned fields, rights-of-way, forest edges and along streambanks. It thrives in open, disturbed, rich, bottom ground, particularly in cultivated fields. As mentioned above, johnsongrass was found to eb resistant to glyphosate in Argentina and the USA, including Mississippi. To effectively manage and control GR johnsongrass and other GR weeds in crop fields, it is necessary to identify GR and glyphosate-susceptible (GS) johnsongrass, and map the distribution of the GR and GS johnsongrass patches and clusters over crop fields.

Traditionally, detection of GR and GS plants is destructive and the plants are assessed by determining the plant physiological and biochemical changes before and after glyphosate treatment on the whole plant, single leaves or leaf disks.[4,5] With glyphosate treatment, GS plants will show symptoms of injury with time after treatment while GR plants will have no or many fewer symptoms of injury depending on the resistance mechanism and growing conditions. This traditional method is regularly used by weed scientists and is effective and accurate for conducting plant physiological studies on GR and GS weeds. However, this method is destructive, tedious and labor intensive.

Glyphosate-treated plants typically indicate changes in physiological, biochemical and cell structures.[6] These changes can be interpreted as plant stress, which can be detected by analysis of measured plant optical responses from spectral characteristics. Hyperspectral analysis has been developed for plant sensing using sensors that subdivide the visible to infrared electromagnetic spectrum into numbers of wavelength bins to allow detection of subtle changes in plant spectra.[7,8] This technology has been applied in agricultural studies, notably identification of plant stress caused by drought, nutrient deficiencies, pest infestations, and herbicide applications.[9,10,11–13]

To develop a rapid, nondestructive method for detection of GR and GS weeds, Reddy *et al.*[14] performed a research study to differentiate GR Palmer amaranth (*Amaranthus palmeri S.Wats.*) from GS Palmer amaranth using hyperspectral imaging technology. The experiments were designed and conducted for forward selection to select the bands, and analyzed to evaluate three different plant sample populations, using Fisher's linear discriminant analysis to reduce dimensionality and maximum likelihood to classify plants. The results showed that GR and GS Palmer amaranth had the best separability in the 400–500 nm, 650–690 nm, 730–740 nm and 800–900 nm spectral regions, which could be consistently differentiated with accuracies over 90% without glyphosate treatment. Huang *et al.*[15] further developed the approach of hyperspectral imaging to rapid sensing of Italian ryegrass (*Lolium perenne ssp. multiflorum*) plants to determine GR plants from GS ones. With the hyperspectral images, sensitive spectral bands were determined with the reduced dimensionality of the available spectral bands and the 15 most sensitive wavelength bands were selected. Then, the maximum likelihood classification was conducted for plant sample differentiation of GR Italian ryegrass from GS Italian ryegrass, with the overall classification accuracy between 75% and 80%. Conventional data analysis and pattern recognition algorithms provided tools for effective data classification and modeling of hyperspectral plant sensing.[11,16] Machine learning algorithms and approaches have been developed for much improved data classification and modeling of hyperspectral plant sensing.[16–20] In recent years machine learning has been developed into the stage of deep learning.[21–23] In the domain of weed science Ferreira *et al.*[24] proposed and developed the use of deep learning for the detection of weeds in soybean crops. Yu *et al.*[25] used deep learning convolutional neural networks for detection of dandelion (*Taraxacum officinale* Web.), ground ivy (*Glechoma hederacea* L.) and spotted spurge (*Euphorbia maculata* L.) growing in perennial ryegrass. However, deep learning typically requires a large amount of data and has issues of high computational complexity and poor interpretation. Nevertheless, although regular machine learning algorithms still have relatively higher complexity, they basically can be effectively used as statistical regression and analysis. Zhang *et al.*[13] developed the machine learning algorithms of Naive Bayes, Random Forest and Support Vector Machine to evaluate the imaged hyperspectral response of soybean plants to different dicamba rates and develop appropriate spectral features and models for assessing the crop damage from dicamba.

With our success in previous research with GR and GS Palmer amaranth and Italian ryegrass differentiations through hyperspectral imaging, we further investigated and tested if a similar spectral approach could be transferred to other GR weed biotypes through machine learning algorithms. Therefore, the objective of this research was to develop a hyperspectral plant sensing method using a spectroradiometer with analysis from machine learning algorithms for rapid, nondestructive measurement, processing and analysis of spectral data for differentiation between GR and GS johnsongrass.

## 2 MATERIALS AND METHODS

### 2.1 Experiment setup

GR and GS johnsongrass biotypes were raised from seeds and used in the study. Seeds were planted in middle of November in 2016 in the research facilities of the US Department of Agriculture, Agricultural Research Unit, Crop Production Systems Research Unit at Stoneville, Mississippi (latitude 33.445062°, longitude 90.869967°). The planting was conducted in a controlled-environment chamber at 1 cm depth in 50 × 20 × 6 cm plastic trays with holes that contained potting mix. Two weeks after emergence, johnsongrass plants were transplanted into 10 × 10 × 10 cm pots with the potting mix in the greenhouse. Plants were fertilized with a nutrient solution and subirrigated as needed. The greenhouse was maintained at 28/22 ± 3 °C day/night temperature with natural light supplemented by sodium vapor lamps to provide a 12 h photoperiod. GR and GS plants, both at 30–40 cm tall, were used for reflectance measurements. About 50 pots of johnsongrass were planted for GR and GS biotypes, respectively, and then 36 pots of each biotype were selected for the experiment for the analysis of the research. After the experiment the plants were sprayed with Roundup WeatherMax® (Monsanto Co., St Louis, MO, USA), which is glyphosate solution prepared using glyphosate potassium salt 540 g ae $L^{-1}$



  

sprayed at a rate of 866 g ae $ha^{-1}$ to confirm the truth of the GR and GS ones.

### 2.2 Plant hyperspectral sensing

At 3 weeks after transplanting in pots (WAT), 4 WAT and 5 WAT, each pot was moved outside the greenhouse and measured under natural sunlight with an ASD handheld portable spectroradiometer (ASD Inc., Boulder, CO, USA) (Fig. 1). The hyperspectral measurements were conducted on three leaves from each pot of plants between 11:00 AM to 1:00 PM on clear sunny days. The average of the hyperspectral data from the three-leaf measurement was used as the representative hyperspectral measurement for that pot. From the measurements the ASD spectroradiometer provided visible–near infrared (VNIR) spectra of johnsongrass plants in the range of 325–1075 nm with 25° field of view. The distance between the sensor probe and the target leaves was kept at 10 cm. For radiometric calibration, a 0.3 m × 0.3 m Spectrolon® white reference target with 99% nominal reflectance (Labsphere, North Sutton, NH, USA) was measured once for every hyperspectral measurements of 10 plants under stable sunlight. Before taking measurements each day, the spectroradiometer was optimized with a white reference measurement to avoid data saturation and a dark current measurement to reduce data noises. To remove vegetation interference from the grass field in the background a sheet of black felt was used to cover the backs of the plant pots (Fig. 1). Black felt has a very low spectral reflectance that is significantly different from the spectral reflectance of the plants. Although the wavelength range of the hyperspectral data was 325–1075 nm, the beginning and the end of the actual spectra were quite noisy. Therefore, for data analysis hyperspectral data between 400 and 900 nm were extracted and used as the representative spectra of each measurement. Figure 2 shows the average spectra of GR and GS plants and the nonreflective black felt. The two reflectance curves have similar shapes. The GR curve had elevated green and NIR values compared with the GS curve, indicating the difference in plant pigment and structure.

### 2.3 Data analysis

In data analysis, the averaged spectral band data at each WAT for GR and GS biotypes were randomly divided into 70% for a training data set and 30% for the testing data set so that there were 50 (25 plant samples × 2 biotypes) spectrum data points for model training and 22 (11 plant samples × 2 biotypes) for model testing. The spectral data were standardized using a $Z$ score transform to evaluate the value's relationship to the mean of a group of values in terms of standard deviations:

$$Z_{score,g,i,j} = \frac{x_{g,i,j} - \overline{x}_{g,j}}{s_{g,j}}, \quad (1)$$

where $Z_{score,s,i,j}$ is the $Z$ score of GR or GS biotypes ($g$) for the $i$th spectrum at the $j$th wavelength, $x_{g,i,j}$ is the spectral value of GR or GS biotypes ($g$) for the $i$th spectrum at the $j$th wavelength, $\overline{x}_{g,j}$ is the mean value of spectral values of GR or GS biotypes ($g$) for at the $j$th wavelength, and $s_{g,j}$ is the standard deviation of spectral values of GR or GS biotypes ($g$) at the $j$th wavelength.

#### 2.3.1 *Spectral band selection*

For differentiating GR and GS johnsongrass, sensitive bands of hyperspectra were identified as being associated with different plant biotypes. The conventional approach to spectral band selection is through correlation analysis between leaf biological responses and spectral features and then selection of the most statistically significant spectral bands[26] or direct choice of the maximum absorption wavelengths.[27] However, there exists high collinearity among the hyperspectral bands, and the usual band selection methods that treat each band as an individual variable tend to generate bands with relatively high information redundancy. To resolve this issue, the successive projections algorithm (SPA) was used in this study for sensitive band selection by conducting a variable projection and matrix analysis to obtain a set of wavelengths with the least collinearity between each other.

SPA is a variable-selection method used to select optimal variables with minimal redundancy with designated models. It has been used to determine effective wavelengths with minimal redundancy.[28–31] SPA performs a projection operation of variable vectors to select uniquely effective variables among those with collinearity. To implement SPA with a model, such as the models based on machine learning, the variable data are formed into a $n \times m$ matrix $\boldsymbol{X} = (\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_m)$, where each vector $\boldsymbol{x}_i\ (i = 1, 2, \ldots, m)$ occupies a column in the matrix to represent a variable with sample size $n$. At the beginning of the algorithm implementation the projections are operated on the $\boldsymbol{X}$ matrix and produce $m$ chains of $S$ variables, where $S = min\{n-1, m\}$ is the maximum number of variables to be selected. Each variable in a chain is selected to have the least collinearity with the

**Figure 1.** Johnsongrass plant leaf hyperspectral measurement.

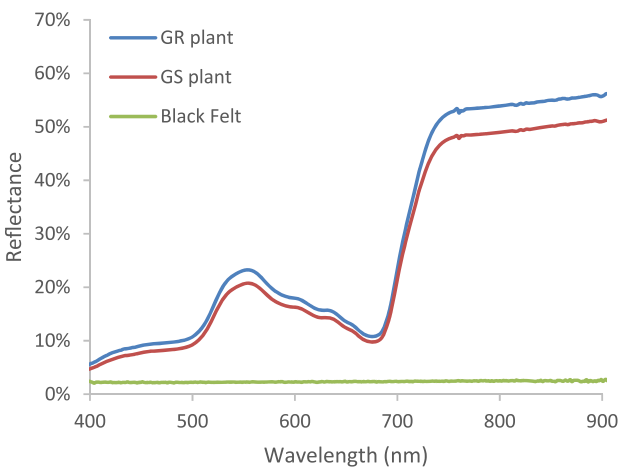


**Figure 2.** Average johnsongrass plant hyperspectral reflectance and black felt reflectance.



  

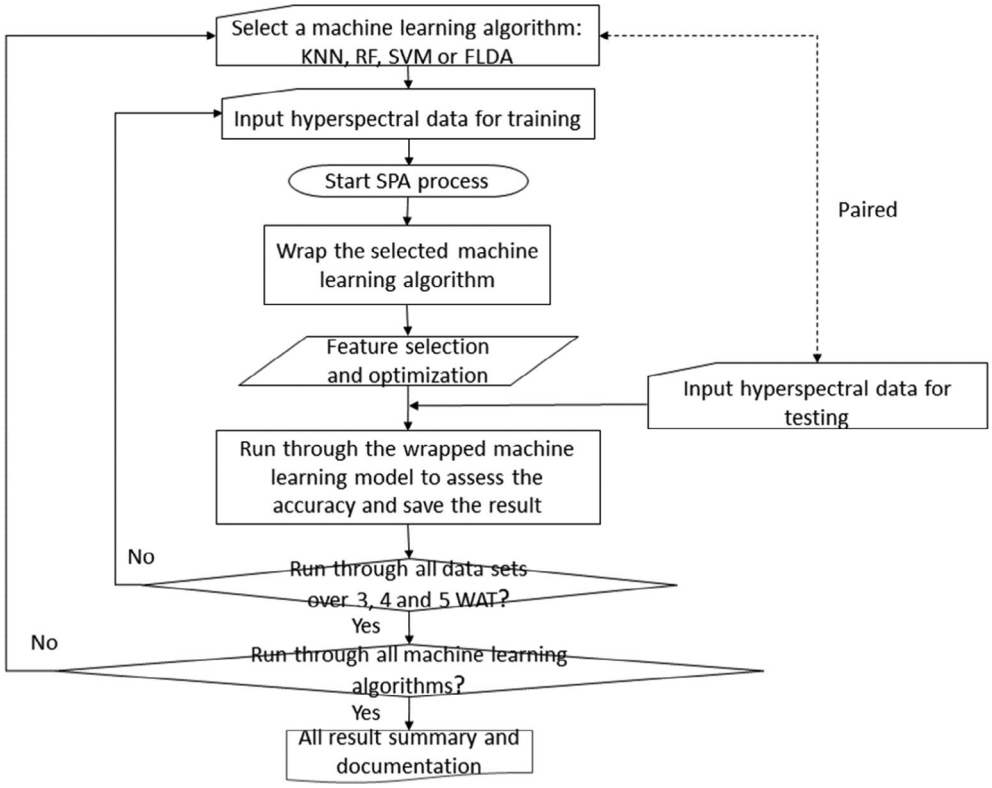


**Figure 3.** Workflow of the research project.

**Table 1.** Initial spectral band selection accuracies from the hyperspectral data measured at 3 WAT wrapped with different machine learning algorithms

| | Number of band groups | Overall accuracy |
|---|---|---|
| KNN | 12 | 36% |
| RF | 30 | 64% |
| SVM | 501 | 36% |
| FLDA | 6 | 91% |

**Table 2.** Initial spectral band selection from the hyperspectral data measured at 4 WAT

| | Number of band groups | Overall accuracy |
|---|---|---|
| KNN | 27 | 60% |
| RF | 15 | 64% |
| SVM | 501 | 50% |
| FLDA | 9 | 55% |

**Table 3.** Initial spectral band selection from the hyperspectral data measured at 5 WAT

| | Number of band groups | Overall accuracy |
|---|---|---|
| KNN | 183 | 77% |
| RF | 3 | 68% |
| SVM | 501 | 77% |
| FLDA | 7 | 100% |

previous ones. Each chain is constructed to start from one of the variables $\boldsymbol{x}_i$ $(i=1,2,\ldots,m)$.

Paiva *et al.*[32] released the code of the SPA algorithm in 2012 based on their original work in 2001.[28] For this study we programed the algorithm based on the code with $m=36$ and $n=501$ so $S=36$.

### 2.3.2 *Data classification through machine learning*

Usually feature selection in pattern recognition is conducted independently from the classification process, which is called the filter method of feature selection.[33] However, the filter method of feature selection does not consider the purpose of the classifier and the relevant feature subset will not reflect the classifier's specific characteristics.[34] The other method of feature selection is called the wrapper method, and it can be used to wrap a classifier and feature selection together to evaluate the performance of the iteratively selected features according to the accuracy of the classifier.[35] The wrapper method essentially optimizes the classifier performance with selected features. Although the wrapper method typically has a higher computational complexity than the filter method due to iterative learning and cross-validation, the implementation of the processes is acceptable for regular computers in most cases for integrated operation of feature selection and classification.

To differentiate the GR johnsongrass from the GS biotypes in this study, the SPA spectral feature selection algorithm was wrapped with the commonly used machine learning algorithms, including k-nearest neighbors (KNN), random forest (RF), and support vector machine (SVM), to compare them with conventional Fisher linear discriminate analysis (FLDA). These machine learning algorithms are representative classifiers based on different



  

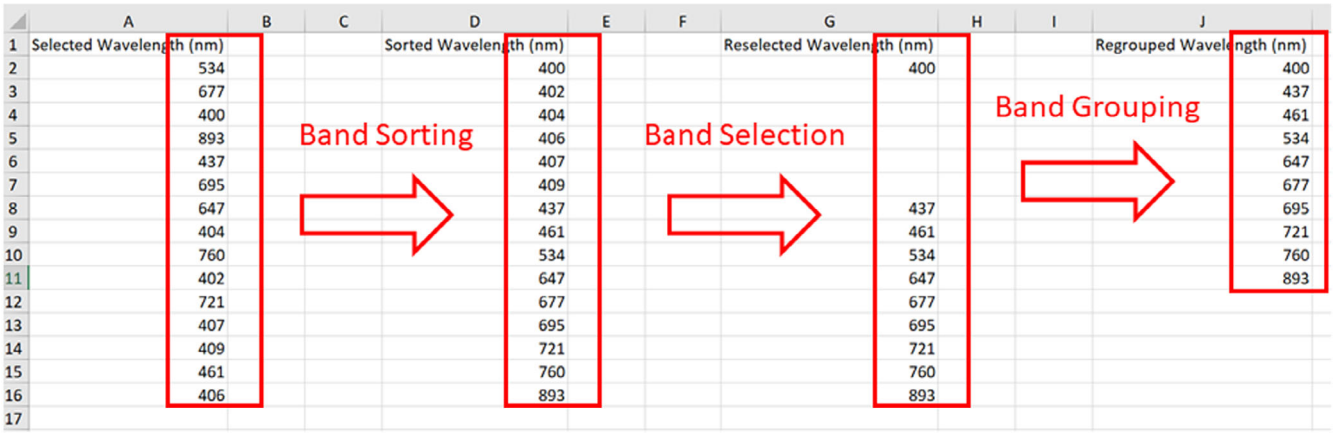


**Figure 4.** Example process of band selection optimization of FLDA at 3 WAT.

**Table 4.** Optimized selected spectral bands from the hyperspectral data measured at 3 WAT and machine learning wrapping accuracies

| | Selected wavelengths (nm) | Overall accuracy |
|---|---|---|
| KNN | 400, 417, 435, 464, 509, 555, 647, 677, 698, 721, 760,893 | 32% |
| RF | 400, 412, 437, 510, 555, 647, 677, 688, 705, 760, 893 | 59% |
| SVM | 400, 421, 463, 520, 606, 647, 677, 694, 713, 760 893 | 36% |
| FLDA | 400, 437, 461, 534, 647, 677, 695, 721, 760, 893 | 59% |

**Table 5.** Final spectral band selection from the hyperspectral data measured at 4 WAT and machine learning wrapping accuracies

| | Selected wavelengths (nm) | Overall accuracy |
|---|---|---|
| KNN | 400, 412, 420, 453, 556, 678, 696, 762, 873, 899 | 50% |
| RF | 400, 419, 440, 556, 678, 696, 728, 762, 873, 899 | 54% |
| SVM | 400, 412, 440, 542, 642, 678, 719, 750, 762, 899 | 50% |
| FLDA | 400, 412, 420, 453, 556, 678, 696, 762, 873, 899 | 50% |

pattern recognition principles and are widely used in the detection of crop stress in crop production processes. KNN is a neural network algorithm with low complexity. It classifies an input data point based on the distance between the data point and its k nearest data points in the training dataset.[36,37] RF is a bagging ensemble algorithm that consists of a group of decision trees each of which is not necessarily related.[38–40] SVM uses kernel functions and hyperplanes as a decision boundary to separate data into classes in a n-dimensional spaces. Multiclass SVMs are usually implemented by combining several two-class SVM classifiers.[41,42] FLDA is based on a linear projection from the original dimensional space to a low dimensional space by maximizing the between-class scatter and minimizing the within-class scatter.[43]

#### 2.3.3 *Data classification assessment*

The confusion matrices were generated with the results of classification of the testing dataset to visualize the performance of the classifiers of the machine learning algorithms. With each confusion matrix an overall accuracy (OA) was calculated to assess the performance of the corresponding classifier as:

$$OA = \frac{N_c}{N_c + N_m} \times 100\% \quad (2)$$

where $N_c$ is the number of samples which are correctly classified and $N_m$ is the number of samples which are incorrectly classified.

#### 2.3.4 *Data analysis computing*

The SPA spectral band selection, machine learning plant classification, and classification assessment were programmed and computed using MATLAB R2009a software (MathWorks Inc., Natick, MA, USA). Figure 3 shows the workflow of the research project.

## 3 RESULTS

### 3.1 SPA spectral band selection

Initially, 15 bands were selected, and the results are shown in Tables 1–3 for the data at three different WATs.

By inspecting the initial results of band selection, it was found that most results for each model are not unique and some of the 15 selected bands are too close to be differentiated. The band selection procedure was therefore optimized by further band selection based on sorting the selected bands to pick the shortest wavelength from the wavebands apart within 10 nm. Figure 4 shows an example process of band selection optimization of FLDA at 3 WAT and other cases of the algorithms at different WATs were implemented similarly to this process. The results for the optimized band selection are shown in Tables 4–6 for the data at different WATS.

### 3.2 Machine learning plant classification

Tables 7–9 show the results of GR and GS johnsongrass classification using different machine learning classifiers at three different WATs.



  

**Table 6.** Final spectral band selection from the hyperspectral data measured at 5 WAT and machine learning wrapping accuracies

| | Selected wavelengths (nm) | Overall accuracy |
|---|---|---|
| KNN | 400, 437, 461, 534, 647, 677, 695, 721, 760, 893 | 73% |
| RF | 400, 414, 453, 520, 558, 649, 681, 701, 717, 759, 899 | 55% |
| SVM | 400, 414, 450, 512, 558, 649, 681, 701, 723, 759, 899 | 77% |
| FLDA | 400, 414, 453, 513, 558, 647, 681, 700, 723, 760, 813, 900 | 73% |

**Table 7.** Confusion matrices of GR and GS johnsongrass classifications based on the spectral data at 3 WAT

| Calculated class | | | |
|---|---|---|---|
| Actual class | GR | KNN –GS | Accuracy |
| GR | 6 | 5 | 54% |
| GS | 10 | 1 | 9% |
| Overall accuracy | | | 32% |
| Actual class | GR | RF –GS | Accuracy |
| GR | 7 | 4 | 64% |
| GS | 5 | 6 | 54% |
| Overall accuracy | | | 59% |
| Actual class | GR | SVM –GS | Accuracy |
| GR | 3 | 8 | 27% |
| GS | 6 | 5 | 55% |
| Overall accuracy | | | 36% |
| Actual class | GR | FLDA –GS | Accuracy |
| GR | 6 | 5 | 55% |
| GS | 4 | 7 | 64% |
| Overall accuracy | | | 59% |

## 4 DISCUSSION

Figure 5 shows the OAs of the four machine learning algorithms wrapped with the SPA spectral band selection. It can be seen that among the algorithms, KNN and SVM could not acceptably classify the GR and GS plants at the early stage (3 WAT) but they improved significantly with the stages and their OAs reached 73% and 77% (5 WAT), respectively, while 77% OA of SVM is the highest overall accuracy among all classification algorithms. RF and FLDA were better able to classify the plants at the early stage but RF had reduced accuracy with the stages, which might be because of the limited extrapolation ability of the model, while FLDA had reduced accuracy at 4 WAT (50%) from 57% at 3 WAT and increased to 73% at 5 WAT, which is considered inconsistent across the stages. Overall, KNN and SVM are preferred with consistent accuracy improvements with the stages, especially SVM reached the highest accuracy 77% at 5 WAT among other algorithms.

**Table 8.** Confusion matrices of GR and GS johnsongrass classifications based on the spectral data at 4 WAT

| Calculated class | | | |
|---|---|---|---|
| Actual class | GR | KNN –GS | Accuracy |
| GR | 7 | 4 | 64% |
| GS | 7 | 4 | 36% |
| Overall accuracy | | | 50% |
| Actual class | GR | RF –GS | Accuracy |
| GR | 6 | 5 | 55% |
| GS | 5 | 6 | 55% |
| Overall accuracy | | | 55% |
| Actual class | GR | SVM –GS | Accuracy |
| GR | 2 | 9 | 18% |
| GS | 2 | 9 | 82% |
| Overall accuracy | | | 50% |
| Actual class | GR | FLDA –GS | Accuracy |
| GR | 8 | 3 | 73% |
| GS | 8 | 3 | 27% |
| Overall accuracy | | | 50% |

**Table 9.** Confusion matrices of GR and GS johnsongrass classifications based on the spectral data at 5 WAT

| Calculated class | | | |
|---|---|---|---|
| Actual class | GR | KNN –GS | Accuracy |
| GR | 10 | 1 | 91% |
| GS | 5 | 6 | 55% |
| Overall accuracy | | | 73% |
| Actual class | GR | RF -GS | Accuracy |
| GR | 9 | 2 | 82% |
| GS | 8 | 3 | 27% |
| Overall accuracy | | | 55% |
| Actual class | GR | SVM –GS | Accuracy |
| GR | 10 | 1 | 91% |
| GS | 4 | 7 | 64% |
| Overall accuracy | | | 77% |
| Actual class | GR | FLDA –GS | Accuracy |
| GR | 9 | 2 | 82% |
| GS | 4 | 7 | 64% |
| Overall accuracy | | | 73% |

Previous studies have been conducted to develop the hyperspectral imaging approach to differentiation of GR Palmer amaranth from GS Palmer amaranth and 90% classification accuracy was achieved.[14] A study has also been conducted to develop the hyperspectral imaging approach to differentiation of GR Italian ryegrass from GS Italian ryegrass with the highest classification accuracy of 80%.[15] This study developed a similar hyperspectral



  

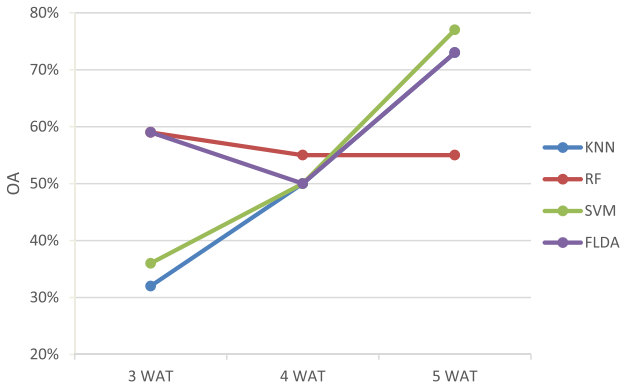


**Figure 5.** Overall accuracies of all four machine learning algorithms wrapped SPA with WATs.

plant sensing approach to differentiation of GR johnsongrass from GS johnsongrass and the highest classification accuracy of 77% was achieved. These results indicate that the hyperspectral approach used for classification of these GR and GS weed biotypes, such as Palmer amaranth, Italian ryegrass and johnsongrass, could be transferred to classification of other GR and GS weed biotypes. The process of plant classification can be conducted by classic pattern recognition approaches and enhanced by machine learning algorithms. The optical spectral approaches can therefore be an alternative, which does not require glyphosate application, to the conventional approach to determine GR and GS weeds, which does need application of glyphosate.

This study indicates that hyperspectral remote sensing is promising in potentially differentiating various GR and GS weed biotypes for precision weed management. In general GR and GS weed plants look very alike and visually it is difficult to distinguish them. The mechanisms involved in glyphosate resistance in GR weeds may affect its leaf chemical composition while differences in chemical composition in turn could affect light absorption patterns in GR and GS plants. This is the basis for using hyperspectral remote sensing to differentiae hyperspectral reflectance properties between GR and GS plants. As demonstrated in this study, in fields or when using a drone-based system, hyperspectral sensitive bands can be similarly selected, and the sensors can be designed based on the selected bands and used on a ground-based system or drone. Machine learning is a tool to enhance data modeling and analysis. With machine learning the complex relations in data variables such as spectral bands, vegetation indices, and GR and GS class labels can be captured. Although each weed biotype may have a particular spectral signature, there are uncertainties in data analysis and modeling, which mainly come from the variation of the data collected. For example, the image quality under different weather conditions may affect the availability of the quality spectral data of the target plants. Data collection, especially for early detection of HR weeds in fields, should therefore ensure sufficient sample size with repeatability over different conditions. The drone should be flown at an optimal altitude to acquire quality images over the fields. Overall, hyperspectral remote sensing has great potential for early detection of GR weed biotypes in fields using ground-based or drone-based systems. For these applications the optimal settings of the systems have to be determined to adapt to different environmental conditions, and the quality and coverage of the data are fundamental to the success of the applications.

## ACKNOWLEDGEMENTS

The authors thank Ryan Poe and Howard Brand for their technical assistance in spectral measurement of the plants, and the Mississippi Soybean Promotion Board for funding.

## DATA AVAILABILITY STATEMENT

Research data are not shared.

## REFERENCES

1 Duke SO, Lydon J, Koskinen WC, Moorman TB, Chaney RL and Hammerschmidt R, Glyphosate effects on plant mineral nutrition, crop rhizophere microbiota, and plant disease in glyphosate-resistant crops. *J Agric Food Chem* **60**:10375–10397 (2012). https://doi.org/10.1021/jf302436u.

2 Benbrook CM, Trends in glyphosate herbicide use in the United States and globally. *Environ Sci Eur* **28**:1–15 (2016). https://doi.org/10.1186/s12302-016-0070-0.

3 Heap I & Liebl R Herbicide-Resistant Weeds by Site of Action. Available: http://www.weedscience.org/Pages/SOASummary.aspx. International Survey of Herbicide Resistant Weeds [6 January 2021].

4 Koger CH and Reddy KN, Role of absorption and translocation in the mechanism of glyphosate resistance in horseweed (Conyza canadensis). *Weed Sci* **53**:84–89 (2005). https://doi.org/10.1614/WS-04-102R.

5 Shaner DL, Nadler-Hassar T, Henry WB and Koger CH, A rapid in vivo shikimate accumulation assay with excised leaf discs. *Weed Sci* **53**:769–774 (2005). https://doi.org/10.1614/WS-05-009R.1.

6 Silva FB, Costa AC, Alves RRP and Megguer CA, Chlorophyll fluorescence as an indicator of cellular damage by glyphosate herbicide in *Raphanus sativus* L. plants. *Am J Plant Sci* **5**:2509–2519 (2014). https://doi.org/10.4236/ajps.2014.516265.

7 Blackburn GA, Hyperspectral remote sensing of plant pigments. *J Exp Bot* **58**:855–867 (2007). https://doi.org/10.1093/jxb/erl123.

8 Thenkabail PS, Gumma MK, Teluguntla P and Mohammed IA, Hyperspectral remote sensing of vegetation and agricultural crops. *Photogramm Eng Remote Sens* **80**:697–710 (2014).

9 Mahlein AK, Steiner U, Hillnhutter C, Dehne HW and Oerke EC, Hyperspectral imaging for small-scale analysis of symptoms caused by different sugar beet diseases. *Plant Methods* **8**:3 (2012). https://doi.org/10.1186/1746-4811-8-3.

10 Ashourloo D, Mobasheri M and Huete A, Developing two spectral disease indices for detection of wheat leaf rust (puccinia triticina ). *Remote Sens (Basel)* **6**:4723–4740 (2014). https://doi.org/10.3390/rs6064723.

11 Lowe A, Harrison N and French AP, Hyperspectral image analysis techniques for the detection and classification of the early onset of plant disease and stress. *Plant Methods* **13**:80 (2017). https://doi.org/10.1186/s13007-017-0233-z.

12 Huang Y, Yuan L, Reddy KN and Zhang J, In-situ plant hyperspectral sensing for early detection of soybean injury from dicamba. *Biosyst Eng* **149**:51–59 (2016). https://doi.org/10.1016/j.biosystemseng.2016.06.013.

13 Zhang J, Huang Y, Reddy KN and Wang B, Assessing crop damage from dicamba on non-dicamba-tolerant soybean by hyperspectral imaging through machine learning. *Pest Manag Sci* **75**:3260–3272 (2019). https://doi.org/10.1002/ps.5448.

14 Reddy KN, Huang Y, Lee MA, Nandula VK, Fletcher RS, Thomson SJ *et al.*, Glyphosate-resistant and -susceptible palmer amaranth (Amaranthus palmeri S. Wats.): hyperspectral reflectance properties of plants and potential for classification. *Pest Manag Sci* **70**:1910–1917 (2014). https://doi.org/10.1002/ps.3755.

15 Huang Y, Lee MA, Nandula VK and Reddy KN, Hyperspectral imaging for differentiating glyphosate-resistant and glyphosate-susceptible Italian ryegrass. *Am J Plant Sci* **9**:1467–1477 (2018). https://doi.org/10.4236/ajps.2018.97107.

16 Liaghat S, Ehsani R, Mansor S, Shafri HZM, Meon S, Sankaran S *et al.*, Early detection of basal stem rot disease (Ganoderma) in oil palms based on hyperspectral reflectance data using pattern recognition algorithms. *Int J Remote Sens* **35**:3427–3439 (2014). https://doi.org/10.1080/01431161.2014.903353.



  

17 Huang Y, Advances in artificial neural networks - methodological development and application. *Algorithms* **2**:973–1007 (2009). https://doi.org/10.3390/algor2030973.
18 Huang Y, Lan Y, Thomson SJ, Fang A, Hoffmann WC and Lacey R, Development of soft computing and applications in agricultural and biological engineering. *Comput Electron Agric* **71**:107–127 (2010). https://doi.org/10.1016/j.compag.2010.01.001.
19 Golhani K, Balasundram SK, Vadamalai G and Pradhan B, A review of neural networks in plant disease detection using hyperspectral data. *Informa Process Agricult* **5**:354–371 (2018). https://doi.org/10.1016/j.inpa.2018.05.002.
20 Huang Y, Research status and applications of nature-inspired algorithms for agri-food production. *Int J Agricult Biol Eng* **13**:1–9 (2020). https://doi.org/10.25165/j.ijabe.20201304.5501.
21 Ongsulee P, Artificial intelligence, machine learning and deep learning. Proceedings of the 15th International Conference on ICT and Knowledge Engineering (ICT&KE), pp. 1–6 (2017). https://doi.org/10.1109/ICTKE.2017.8259629
22 Zhang L, Tan J, Han D and Zhu H, From machine learning to deep learning: progress in machine intelligence for rational drug discovery. *Drug Discov Today* **22**:1680–1685 (2017). https://doi.org/10.1016/j.drudis.2017.08.010.
23 Xin Y, Kong L, Liu Z, Chen Y, Li Y, Zhu H *et al*., Machine learning and deep learning methods for cybersecurity. *IEEE Access* **6**:35365–35381 (2018). https://doi.org/10.1109/ACCESS.2018.2836950.
24 Ferreira AS, Freitas DM, da Silva GG, Pistori H and Folhes MT, Weed detection in soybean crops using ConvNets. *Comput Electron Agricult* **143**:314–324 (2017). https://doi.org/10.1016/j.compag.2017.10.027.
25 Yu J, Schumann AW, Cao Z, Sharpe SM and Boyd NS, Weed detection in perennial ryegrass with deep learning convolutional neural network. *Front Plant Sci* **10**:1422 (2019).
26 Cheng T, Rivard B and Sanchez-Azofeifa A, Spectroscopic determination of leaf water content using continuous wavelet analysis. *Remote Sens Environ* **115**:659–670 (2011). https://doi.org/10.1016/j.rse.2010.11.001.
27 Le Maire G, François C and Dufrene E, Towards universal broad leaf chlorophyll indices using PROSPECT simulated database and hyperspectral reflectance measurements. *Remote Sens Environ* **89**:1–28 (2004). https://doi.org/10.1016/j.rse.2003.09.004.
28 Araújo MCU, Saldanha TCB, Galvão RKH, Yoneyama T, Chame HC and Visani V, The successive projections algorithm for variable selection in spectroscopic multicomponent analysis. *Chemometr Intell Lab* **57**:65–73 (2001). https://doi.org/10.1016/S0169-7439(01)00119-8.
29 Liu F, Jiang Y and He Y, Variable selection in visible/near infrared spectra for linear and nonlinear calibration: a case study to determine soluble solids content of beer. *Anal Chim Acta* **635**:45–52 (2009). https://doi.org/10.1016/j.aca.2009.01.017.
30 Xie C, Shao Y, Li X and He L, Detection of early blight and late blight diseases on tomato leaves using hyperspectral imaging. *Sci Rep* **5**:16564 (2015). https://doi.org/10.1038/srep16564.
31 Milanez K, Nóbrega T, Nascimento D, Galvão R and Pontes M, Selection of robust variables for transfer of classification models employing the successive projections algorithm. *Anal Chim Acta* **984**:76–85 (2017). https://doi.org/10.1016/j.aca.2017.07.037.
32 Paiva HM, Soares SFC, Galvão RKH and Araújo MCU, A graphical user interface for variable selection employing the successive projections algorithm. *Chemom Intel Lab Syst* **118**:260–266 (2012). https://doi.org/10.1016/j.chemolab.2012.05.014.
33 Wang S and Zhu W, Sparse graph embedding unsupervised feature selection. *IEEE Trans Syst Man Cybern: Systems* **48**:329–341 (2016). https://doi.org/10.1109/TSMC.2016.2605132.
34 Chrysostomou K, Wrapper feature selection. In: Encyclopedia of Data Warehousing and Mining, Second ed. IGI Global, pp. 2103–2108. (2009).
35 Freeman C, Kulic D and Basir O, Feature-selected tree-based classification. *IEEE Trans Cybern* **43**:1990–2004 (2013). https://doi.org/10.1109/TSMCB.2012.2237394.
36 Cover T and Hart P, Nearest neighbor pattern classification. *IEEE Trans Inform Theory* **13**:21–27 (1967). https://doi.org/10.1109/TIT.1967.1053964.
37 Alimjan G, Sun T, Liang Y, Jumahun H and Guan Y, A new technique for remote sensing image classification based on combinatorial algorithm of SVM and KNN. *Int J Pattern Recognit Artif Intell* **32**:1859012 (2017). https://doi.org/10.1142/S0218001418590127.
38 Strobl C, Boulesteix A, Kneib T, Augustin T and Zeileis A, Conditional variable importance for random forests. *BMC Bioinform* **9**:1–11 (2008). https://doi.org/10.1186/1471-2105-9-307.
39 Belgiu M and Drăguț L, Random forest in remote sensing: a review of applications and future directions. *ISPRS J Photogramm Remote Sens* **114**:24–31 (2016). https://doi.org/10.1016/j.isprsjprs.2016.01.011.
40 Gregorutti B, Michel B and Saint-Pierre P, Correlation and variable importance in random forests. *Stat Comput* **27**:659–678 (2017). https://doi.org/10.1007/s11222-016-9646-1.
41 Liu Y and Zheng Y, One-against-all multi-class SVM classification using reliability measures, in *Proceedings of 2005 IEEE International Joint Conference on Neural Networks*, IEEE, Piscataway, NJ, USA, pp. 849–854 (2005). https://doi.org/10.1109/IJCNN.2005.1555963
42 Calderón MR, Navas-Cortés JA and Zarco-Tejada PJ, Early detection and quantification of verticillium wilt in olive using hyperspectral and thermal imagery over large areas. *Remote Sens (Basel)* **7**:5584–5610 (2015). https://doi.org/10.3390/rs70505584.
43 Liu Q, Huang R, Lu H and Ma S, Face recognition using kernel-based fisher discriminant analysis. *Proc Fifth IEEE Int Conf Autom Face Gesture Recognit* **1**:197–201 (2002). https://doi.org/10.1109/AFGR.2002.1004157.



  