Available online at www.sciencedirect.com

ScienceDirect

journal homepage: www.elsevier.com/locate/issn/15375110

Research Paper

# Recognising weeds in a maize crop using a random forest machine-learning algorithm and near-infrared snapshot mosaic hyperspectral imagery

*Junfeng Gao* [a], *David Nuyttens* [b], *Peter Lootens* [c], *Yong He* [d,*], *Jan G. Pieters* [a]

[a] *Department of Biosystems Engineering, Ghent University, Coupure Links 653, 9000, Gent, Belgium*
[b] *Technology and Food Science Unit, ILVO, Burg. Van Gansberghelaan 115, 9820, Merelbeke, Belgium*
[c] *Plant Sciences Unit, ILVO, Caritasstraat 39, 9090, Melle, Belgium*
[d] *College of Biosystems Engineering and Food Science, Zhejiang University, Yuhangtang Road 866, 310058, Hangzhou, Zhejiang, China*

ARTICLE INFO

*Article history:*
Received 14 August 2017
Received in revised form
2 February 2018
Accepted 14 March 2018
Published online 30 March 2018

*Keywords:*
Snapshot hyperspectral imaging
Machine learning
Plant classification
Hyperparameter tuning
Feature selection
Cross validation

This study explores the potential of a novel hyperspectral snapshot mosaic camera for weed and maize classification. The image processing, feature engineering and machine learning techniques were discussed when developing an optimal classification model for the three kinds of weeds and maize. A total set of 185 spectral features including reflectance and vegetation index features was constructed. Subsequently, the principal component analysis was used to reduce the redundancy of the constructed features, and the first 5 principal components, explaining over 95% variance ratio, were kept for further analysis. Furthermore, random forests as one of machine learning techniques were built for developing the classifier with three different combinations of features. Accuracy-oriented feature reduction was performed when choosing the optimal number of features for building the classification model. Moreover, hyperparameter tuning was explored for the optimal selection of random forest model. The results showed that the optimal random forest model with 30 important spectral features can achieve a mean correct classification rate of 1.0, 0.789, 0.691 and 0.752 for *Zea mays*, *Convolvulus arvensis*, *Rumex* and *Cirsium arvense*, respectively. The McNemar test showed an overall better performance of the optimal random forest model at the 0.05 significance level compared to the k-nearest neighbours (KNN) model.



* *Corresponding author*. Fax: +86 571 88982143.
E-mail address: yhe@zju.edu.cn (Y. He).
https://doi.org/10.1016/j.biosystemseng.2018.03.006

**Nomenclature**

| | |
|---|---|
| $d$ | Dimension of features in the dataset |
| $E_{ii}$ | Diagonal value of confusion matrix |
| $F_1$ | A weighted average of precision and recall |
| $\alpha$ | One sample of the dataset |
| $m$ | The number of trees to build Random Forests |
| $M$ | Eigenvector Matrix |
| $n$ | The number of selected features to build Random Forests |
| $q_{01}$ | Number of samples misclassified by KNN but not by RF |
| $q_{10}$ | Number of samples misclassified by RF but not by KNN |
| $\mu$ | Eigenvalues |
| $\beta$ | Eigenvector |
| $R_{calibrated}(\lambda)$ | Calibrated reflectance at wavelength $\lambda$ |
| $Raw(\lambda)$ | Uncalibrated digital number of pixel at wavelength $\lambda$ |
| $W(\lambda)$ | The digital number of calibration panel at wavelength $\lambda$ |
| $X_i$ | A bootstrap subset |
| $\Psi$ | Statistical result from the McNemar test |

*Abbreviations*

| | |
|---|---|
| CV | Cross Validation |
| DC | Dark Current value |
| KNN | K-Nearest Neighbour |
| NDVI | Normalised Difference Vegetation Index |
| NIR | Near Infrared |
| OOB | out of bag samples |
| RF | Random Forests |
| ROI | Region of Interest |
| RVI | Ratio Vegetation Index |
| SSWM | Site-Specific Weed Management |
| VIs | Vegetation Indices |
| VB | Visual band |

## 1. Introduction

Maize (*Zea mays*), one of the main cereals for food, forage and processed industrial products, is widely grown worldwide and a greater amount of maize is produced every year than any other grain (Ostrý, Malíř, & Pfohl-Leszkowicz, 2015). Although the maize yield increased to 1080 million tonnes in 2016 according to the statistics of the Food and Agriculture Organization of United Nations (FAO),[1] the quality of this crop still faced many problems such as weed infestation, animal pests and pathogens (Oerke, 2006). Weeds are one of the most important factors to limit maize production. They cause significant yield losses worldwide with an average of 29.2% if no weed control is applied (Dogan, Ünay, Boz, & Albay, 2004; Oerke & Steiner, 1996). Generally, most fields are infested with multiple weeds. For maize fields, *Convolvulus arvensis* and *Cirsium arvense* are the common weeds in central and western Europe (Meissle et al., 2010). In some certain circumstances, *Rumex* is also germinated among maize seedlings due to the easy propagation of its seeds. Besides, they are all perennial dicotyledons, which are suitable to control using chemical or mechanical ways (Macías, Castellano, & Molinillo, 2000; Zhang, 2003). The common weed management methods include prevention and cultural, mechanical, biological and chemical approaches (Harker & O'Donovan, 2013). Chemical methods such as spraying effective herbicides are the dominant management techniques for weed control in modern agriculture (Harker & O'Donovan, 2013). In most weed control methods, it is generally accepted to be most effective to control weeds in their early growth stage (López-Granados, 2011). Especially for maize crop, it is difficult to spray in practices in late growth stages due to the height of maize plants.

Under natural growing conditions, weeds are generally distributed in small patches, but farmers often uniformly spray herbicide in their fields, which is not in agreement with sustainable agriculture development and increases the cost of crop production. Site-specific weed management (SSWM), a precision agriculture approach, refers to the spatially variable application of weed control strategies for achieving the minimisation of herbicide usage (Shaw, 2005). It is useful in monitoring and managing weed patches at early growth stages (Shaner & Beckie, 2014). However, one of the main technical challenges of implementation lies in weed detection or classification (Shaner & Beckie, 2014; Slaughter, Giles, & Downey, 2008).

Currently, most weed detection studies can be classified into two groups. The first group utilises geometric differences for identification, such as leaf shape, texture, crop location. The second group differentiates weeds from crops using spectral reflectance characteristics (Slaughter et al., 2008; Thompson, Stafford, & Miller, 1991). Based on the two principles, various sensors, both imaging and non-imaging ones, have been applied in the investigation of weed detection in recent years. RGB cameras are widely applied for weed detection due to their general availability and low cost (Romeo et al., 2013; Tellaeche, Pajares, Burgos-Artizzu, & Ribeiro, 2011; Torres-Sánchez, López-Granados, De Castro, & Peña-Barragán, 2013; Gao et al., 2018). However, RGB cameras provide only limited spectral information as they only record information using three broad bands. To obtain more spectral information, a hyperspectral camera was introduced in classification applications (Gao, Li, Zhu, & He, 2013). Hyperspectral imaging sensors often involve more and narrower bands to gain detailed spectral information. Every pixel from hyperspectral images has complete spectrum information which has been used for a variety of applications in agriculture (Thenkabail et al., 2013). For example, the applications of line-scanning hyperspectral imagery for weed species recognition were presented by Okamoto, Murata, Kataoka, and Hata (2007) and Pantazi, Moshou, and Bravo (2016). Wendel and Underwood (2016) also developed a self-supervised training data generation and weed detection system for vegetable fields. However, these systems, based on line-scanning hyperspectral sensors, are negatively affected by the rapid motion of platforms or objects because of the need to scan image. A snapshot hyperspectral system, without scanning,

[1] FAOSTAT data website, http://www.fao.org/faostat/en/#home; accessed 21 January 2018.

enables both spatial and spectral information to be captured simultaneously during a single integration time of a detector array. The video-rate hyperspectral datacubes provided by this system also ensure their high efficiency for monitoring crops. Besides, this system can be developed into a compact device which is appropriate for drone-based remote sensing in agriculture (Ishida et al., 2018).

Random Forests (RF) algorithm is one of the popular and powerful machine-learning techniques (Breiman, 2001). It builds many decision trees that are aggregated to compute a classification by means of a majority vote of the classifier ensemble. Each decision tree is constructed by randomly selecting a subset of features and using a different bootstrap sample from original data, which can reduce the effects of overfitting and improve generalisation (Peters et al., 2007). Moreover, RF can provide feature importance ranking, which is valuable for feature selection. There is a substantial body of research covering RF applications in the remote sensing community (Belgiu & Drăgu, 2016), such as land-cover classification (Nitze, Barrett, & Cawkwell, 2015), tree species mapping (Ghosh, Fassnacht, Joshi, & Koch, 2014), and vegetation classification (Juel, Groom, Svenning, & Ejrnæs, 2015).

In this study, a snapshot mosaic hyperspectral imaging sensor was applied to weed and maize classification. We propose an approach to process these snapshot hyperspectral images to obtain the spectral reflectance of a region of interest (ROI). The specific objectives of this research are (1) to explore the feasibility of near infrared (NIR) snapshot mosaic hyperspectral camera in weed and maize classification; (2) to determine the relevant spectral wavelengths and important features for classification; (3) to provide optimal parameters for building a RF model.

## 2. Materials and methods

### 2.1. Sample preparation and data collection

Seeds of *C. arvensis* (2) and *Z. mays* (12) were sown in pots (10) that were filled with sandy soil. The image recording, with snapshot mosaic hyperspectral camera in the plant laboratory of the Institute for Agricultural and Fisheries Research (ILVO) in Belgium, started when the *C. arvensis* and *Z. mays* plants emerged and had 2 unfolded leaves. *Rumex* (7) and *C. arvense* (7) plants were taken from the experimental field of ILVO and then grown in pots (10) for image recording. Image recording was finished by April 18th, 2016. During this time, all the plants were imaged twice a week (every Monday and Friday) and the image data collection took about 1 h on each occasion. The plants were returned to the glasshouse after each image data collection. Table 1 presents further details of the image data collection. ROIs (Garcia-Ruiz, Wulfsohn, & Rasmussen, 2015) were used as samples in our dataset. They were derived from polygon areas which were constructed by manually drawing several points in the edge of leaves in the collected hyperspectral images. This procedure was implemented with *roipoly* function in Matlab 2016. The total number of leaf ROIs for *C. arvensis*, *Rumex*, *C. arvense* and *Z. mays* were 79, 80, 80 and 84, respectively.

**Table 1 – The further description of dataset.**

| Plant species | Starting date | Finish date | Images used | Selected ROIs |
|---|---|---|---|---|
| *Convolvulus arvensis*[a] | 29/02/2016 | 18/04/2016 | 24 | 79 |
| *Rumex* | 21/03/2016 | 18/04/2016 | 24 | 80 |
| *Cirsium arvense* | 21/03/2016 | 18/04/2016 | 24 | 80 |
| *Zea mays* | 04/04/2016 | 18/04/2016 | 25 | 84 |

[a] The plants were fertilised 0.19 g Nitrogen for each plant in the date of 29/03/2016.

### 2.2. Snapshot mosaic hyperspectral imaging sensor and experiment

The snapshot mosaic hyperspectral cameras are compact and lightweight (Fig. 1), with a total mass less than 500 g including the body frame, two filters (Edmund Optics, USA), two camera heads with sensors (CMOSIS CMV2000 based) from imec company in Belgium, an embedded chip system (3D-ONE, Netherlands), storage and power distribution unit. Only the NIR camera, with 25 bands (601 nm, 605 nm, 614 nm, 627 nm, 636 nm, 644 nm, 652 nm, 660 nm, 669 nm, 677 nm, 684 nm, 698 nm, 724 nm, 738 nm, 750 nm, 764 nm, 776 nm, 790 nm, 802 nm, 814 nm, 833 nm, 845 nm, 855 nm, 866 nm, 871 nm) displayed as 5 × 5 mosaic array, was used in our study and the

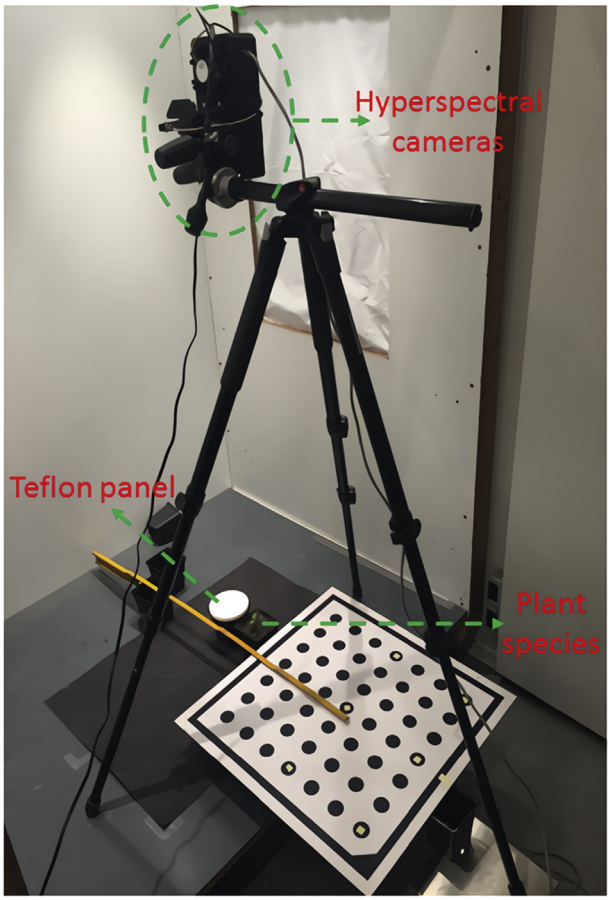


**Fig. 1 – The setup of experiment.**

spatial resolution of each spectral band image was 403 × 216 pixel. During the image data collection, the camera was mounted on top of the tripod and then placed in a cabinet (1.2 m × 1.0 m × 2.0 m). The distance between the camera lens and the pots was 1.0 m. Four halogen lamps (50w × 36°, 12 V) from OSRAM company were deployed in the four corners of the cabinet ceiling for providing homogeneous light. The exposure time in every image recording was set to be 0.099 s. In our study, the main processing steps for weed and maize crop classification are depicted in Fig. 2. In the image processing procedure, the raw image is composed of 5 × 5 single band sub-images. Subsequently, the 25 sub-images were cropped and the averaged digital numbers of the selected ROIs were calculated. Through the calibration equation, the calibrated reflectance of the ROIs for every single band was finally obtained.

### 2.3. Reflectance calibration

The raw snapshot hyperspectral images were calibrated with the white Teflon panel reference by Eq. (1).

$$R_{Calibrated}(\lambda) = \frac{Raw(\lambda) - DC}{W(\lambda) - DC} \times 100\% \quad (1)$$

where $R_{Calibrated}$ and Raw are the calibrated reflectance and raw hyperspectral images respectively, $\lambda$ is the wavelength of camera, $W(\lambda)$ is averaged intensity value of the white Teflon panel reference. $DC$ is dark current value of the camera.

### 2.4. Vegetation indices

Vegetation indices (VIs) are related to many properties of plants and are frequently used to classify plant species, identify the health status of plants and estimate green biomass and crop yield (Viña, Gitelson, Nguy-Robertson, & Peng, 2011). The normalised difference vegetation index (NDVI) and ratio vegetation index (RVI) were calculated by Eqs. (2) and (3), respectively. In this paper, there are 80 NDVIs and 80 RVIs constructed features in total. The range of NDVI is between 0 and 1. No pre-treatments were applied in the calculation of NDVI values. While for the values of the RVI features, they were scaled from 0 to 1 by Eq. (4).

$$NDVI(\gamma_1, \gamma_2) = \frac{NIR(\gamma_1) - VB(\gamma_2)}{NIR(\gamma_1) + VB(\gamma_2)} \quad (2)$$

$$RVI(\gamma_1, \gamma_2) = \frac{NIR(\gamma_1)}{VB(\gamma_2)} \quad (3)$$

where $\gamma_1$ represents one NIR band from 724 nm, 738 nm, 750 nm, 764 nm, 776 nm, 790 nm, 802 nm, 814 nm and $\gamma_2$ represents one visual band (VB) from 601 nm, 605 nm, 614 nm, 627 nm, 636 nm, 644 nm, 652 nm, 660 nm, 669 nm, 677 nm.

$$RVI_{Scaled} = \frac{RVI - RVI_{Min}}{RVI_{Max} - RVI_{Min}} \quad (4)$$

### 2.5. Principal component analysis and classification models

Principal component analysis (PCA) is a popular method for dimensionality reduction, feature extraction and data compression (Chamundeeswari, Singh, & Singh, 2009). It can be defined as the orthogonal transformation of raw data into a set of values of linearly uncorrelated variables, known as principal components, in lower dimensionality. The first few

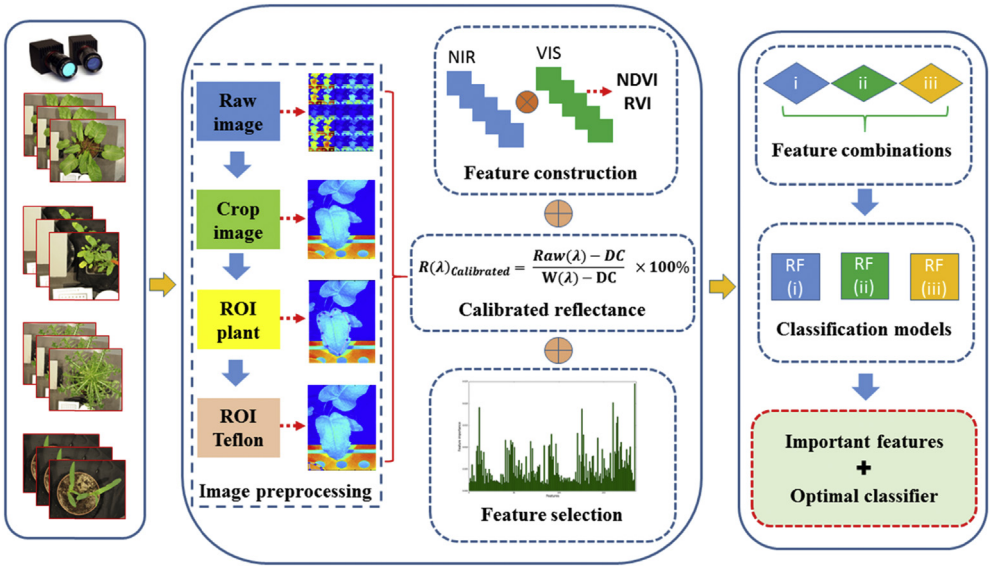


**Fig. 2 – Key steps of weed and crop recognition by snapshot hyperspectral imaging. There are three main steps for weed and crop recognition. (1) Weeds and crop were imaged by NIR snapshot hyperspectral camera in laboratory; (2) Through image preprocessing and calibration formula, the reflectance was obtained as band features, then NDVI and RVI features were constructed by reflectance in VB and NIR regions. After feature construction, feature selection algorithms were applied for extracting distinctive features. (3) Different types of features were combined as model input for random forests (RF), and the goal of this research is to bulid a RF model with optimal hyperparameters and important features for weed and crop recognition.**

principal components generally have the largest variances which could explain most relevant information in the raw data. The algorithm is summarised in the following steps (Wold, Esbensen, & Geladi, 1987):

(1) Standardise the data set ($d$-dimensional features) without considering class labels;
(2) Compute the covariance matrix of the whole data set;
(3) Compute eigenvectors ($\beta_1$, $\beta_2$, …, $\beta_d$) and corresponding eigenvalues ($\mu_1$, $\mu_2$, …, $\mu_d$);
(4) Sort and select top $k$ eigenvectors to form a $d \times k$ dimensional matrix M;
(5) Transform the samples into the new subspace by Eq (5);

$$y = M^T \times \alpha \tag{5}$$

where $\alpha$ represents one sample ($d \times 1$ dimension) in the raw data set, y is the transformed sample ($k \times 1$ dimension) in the new subspace.

In RF, about one-third of the data is left out of the bootstrap sample and not used in the construction of the decision tree. This remaining data, also called out-of-bag samples (OOB), can be used to evaluate the OOB errors as well as to determine the importance of features. The algorithm for aggregating a random forest of $m$ decision trees goes as follows:

(1) for $l = 1$ to $m$:
  (i) randomly choose a bootstrap subset $X_i$ which contains two thirds of instances in the original data set;
  (ii) use $X_i$ to build an unpruned decision tree with randomly selected $n$ features from all the features;
(2) predict new samples based on the majority vote of the ensemble of m decision trees.

The number of trees ($m$) and the number of randomly selected features ($n$) to split the tree nodes are two hyperparameters which need to be optimised for obtaining a minimal random forest error. RF also provides valuable information for estimating the importance of a feature by calculating how much the OOB error increases when OOB data for that feature are permuted while all other features are left unchanged.

K-nearest neighbours (KNN) algorithm is a non-parametric approach which can be used for classification and regression (Altman, 1992). The input consists of the K closest training examples in the feature space. In this study, the output is predicted by the majority vote of its neighbours with Euclidean distance. K was assigned to be 5 in the classification model. Python was used to implement PCA algorithm and to develop the two classification models (KNN and RF).

### 2.6. Metrics for classification model evaluation

When referring to the performance of a classification model, the ability of a model to correctly predict or separate classes is emphasised. The confusion matrix (Stehman, 1997) gives a full description of errors made by classifiers. In this matrix, precision, recall and F1 values can also be calculated as performance metrics for model evaluation. Precision is a measure of prediction relevancy defined by Eq. (6). Recall, also called sensitivity, is a measure of the capability of a classifier to select instances of a certain class from dataset and corresponds to the true positive rate (Eq. (7)). $F_1$ score, calculated by Eq. (8), is interpreted as a weighted average of precision and recall. It is the harmonic mean of precision and recall. The best value of $F_1$ is 1 and the worst value is 0. In addition, another common statistic for reporting the performance of a multi-classification model is model accuracy (Sokolova & Lapalme, 2009). Model accuracy is the overall correctness of the model and is calculated as the sum of correct samples divided by the total number of samples.

$$\text{Precision}_i = \frac{E_{ii}}{\Sigma_j E_{ji}} \tag{6}$$

$$\text{Recall}_i = \frac{E_{ii}}{\Sigma_j E_{ij}} \tag{7}$$

$$F(i)_1 = 2*\frac{\text{Precision}_i * \text{Recall}_i}{\text{Precision}_i + \text{Recall}_i} \tag{8}$$

In the confusion matrix, where $E_{ii}$ represents diagonal elements of the i-th class, $\Sigma_j E_{ji}$ represents the sum of i-th class predicted by classification model. $\Sigma_j E_{ij}$ represents the total of true values of the i-th class.

### 2.7. Statistical model comparison

The McNemar test (Everitt, 1992, p. 164) was employed for comparing the performance of RF and KNN model. The predictions made by the two models with the same features in the data set were compared with the true values and used to construct the $2 \times 2$ contingency table (Everitt, 1992, p. 164). Under the null hypothesis, the two models should have the same classification error rate. McNemar's test is based on a $\chi^2$-test for goodness-of-fit that compares the distribution of counts expected under the null hypothesis to the observed counts. The following statistic is approximately distributed $\chi^2$ with 1° of freedom (Dietterich, 1998).

$$\Psi = \frac{\left(|q_{01} - q_{10}| - 1\right)^2}{q_{01} + q_{10}} \tag{9}$$

In this paper, where $q_{01}$ is the number of samples misclassified by KNN but not by RF, $q_{10}$ is the number of samples misclassified by RF but not by KNN. If the null hypothesis is correct, then the probability that this quantity is greater than $\chi^2_{1,0.95} = 3.841$ is less than 0.05.

## 3. Results

### 3.1. Principal component analysis for constructed features

To reduce redundancy, the constructed 80 NDVI and 80 RVI features were subjected to PCA separately. Figure 3 shows the

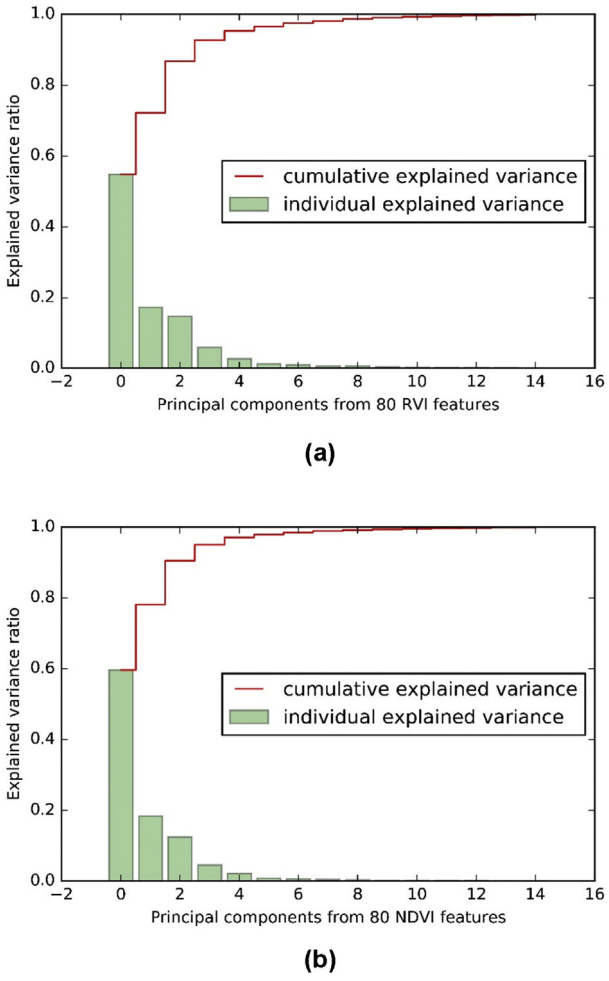


**Fig. 3 – The explained variance ratio of principal components. (a) the principal components derived from 80 RVI features; (b) the principal components derived from 80 NDVI features.**

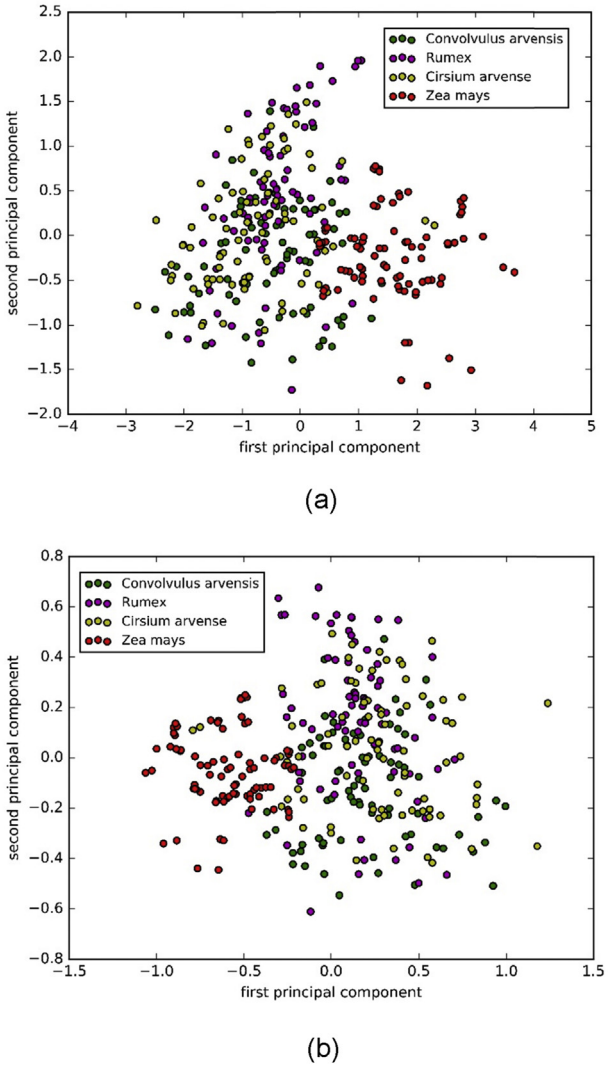


**Fig. 4 – The distribution of samples in first two principal components. (a) the principal components from 80 RVI features; (b) the principal components from 80 NDVI features.**

results of principal components and their explained variance ratio. The sum of the first five principal components of RVI features (95.34%) and NDVI features (97.29%) have all explained over 95% original variations. Their first principal components constitute 54.83% and 59.66% variance ratio, respectively. The distribution of all samples in the first two principal components is illustrated in Fig. 4. It can be seen that the three kinds of weeds mix with each other and distribute irregularly. While the distribution of maize is much better, they tend to focus on one side of the weeds, though several weed samples of *C. arvense* mix among them. This pattern indicates that the maize perhaps could obtain better classification results than the weeds.

### 3.2. *Random forests model with all features*

Cross validation (CV) was employed for evaluation of the RF. The original data were split into 5 folds, using the folds one by one for testing and the remaining folds as training set (Fig. 5). In this way, each sample in the whole data can be tested once. The optimal value of *n* for building the RF is assumed to be around the square root of the total number of features (185). Therefore, the value of *n* is set between 5 and 20. The range of *m* is between 100 and 300. A grid search approach (Hsu, Chang, & Lin, 2008) was used to search for the optimal parameters (*m*, *n*) for building RF based on overall accuracy in the nested CV. Figure 6 gives the overall accuracy for each grid point. The best parameters are marked as a red point in Fig. 6b. When *m* equals 231 and n equals 8, the RF model achieves the maximum overall correct classification rate (0.827). These two optimised parameter values were then used in building random forests in the full dataset with 5-fold CV. The results of each RF are shown in Table 2. The mean overall accuracy in all the folds is 0.808 which is slightly lower than that in the nested CV (0.827), and the standard deviation is 0.050. Figure 7 is the confusion matrix of the RF model using 5-fold CV. It can be seen that all maize (*Z. mays*) samples were recognised in the RF model. But for the weed samples, they were confused with each other. Weed1 (*C. arvensis*) was better classified than weed2 (*Rumex*) and weed3 (*C. arvense*). Their classification accuracies are 0.785, 0.663 and 0.713, respectively.

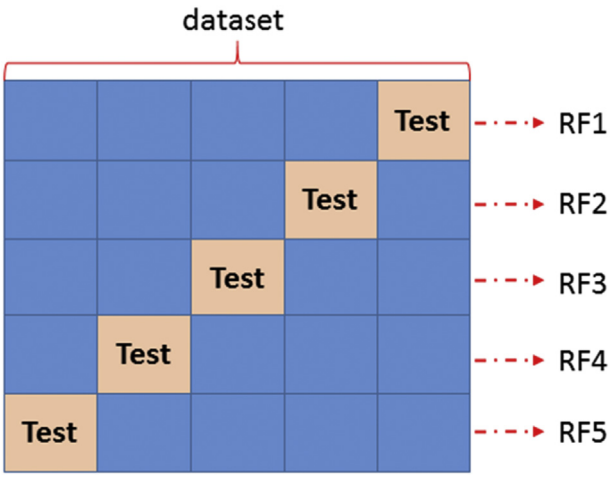


**Fig. 5 – 5-Fold CV for random forests.**

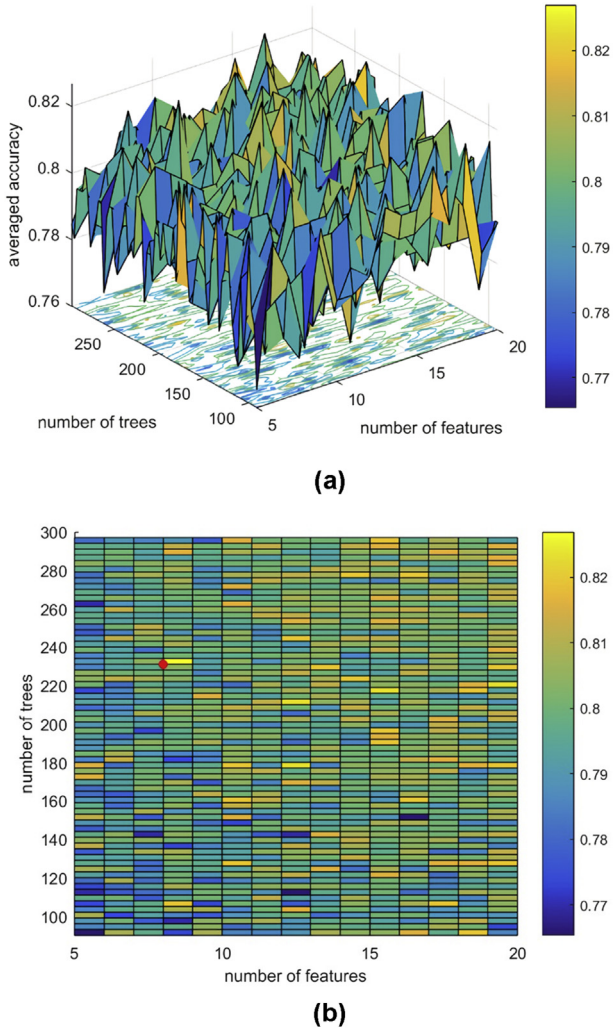


**Fig. 6 – The distribution of overall recognition rate in the grids. (a) 3D surface curve of overall recognition rate; (b) Performance map of overall recognition rate in each grid point, the best parameters are marked in red point.**

**Table 2 – The recognition rates of the 5-fold CV in the whole dataset.**

| Model | RF1 | RF2 | RF3 | RF4 | RF5 |
|---|---|---|---|---|---|
| Recognition rate | 0.877 | 0.754 | 0.846 | 0.813 | 0.750 |

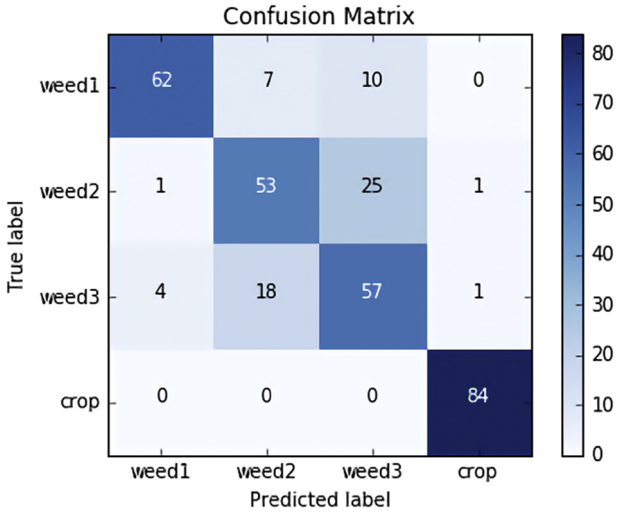


**Fig. 7 – The confusion matrix of all samples. Weed1, weed2, weed3 and crop represent *Convolvulus arvensis*, *Rumex*, *Cirsium arvense and Zea mays* plant species.**

### 3.3. *Feature importance and selection*

After feature construction, 185 features were obtained in total, but not all of them are equally distinctive to discriminate between weeds and maize. RF allows the importance of every feature to be evaluated, based on OOB errors. The importance score of each feature is displayed in Fig. 8. Based on these importance scores, the features were ranked, and accuracy-oriented feature reduction (Loosvelt, Peters, Skriver, De Baets, & Verhoest, 2012) was performed in a CV loop to select the optimal number of features. Specifically, the most important feature was first used to build the model, and then the ranked features were added one by one to build the models, respectively. This procedure was repeated until the least important feature used to build the model. Figure 9 presents the overall accuracy of every model with the number of selected features. The overall classification rate achieves 0.52 just using the first most important feature (871 nm). Then this rate sharply increases to around 0.77 when the next 10 most important features were used. Afterwards, the overall classification rate fluctuates at 0.78 and does not improve much with further added features. Finally, based on the maximum of accumulated accuracy, the top 30 important features (Table 3) were selected for building the model.

### 3.4. *Classification results and model comparison*

The three different combinations of features, (*i*) 25 bands and 10 PCA features, (*ii*) 30 important features selected by RF and

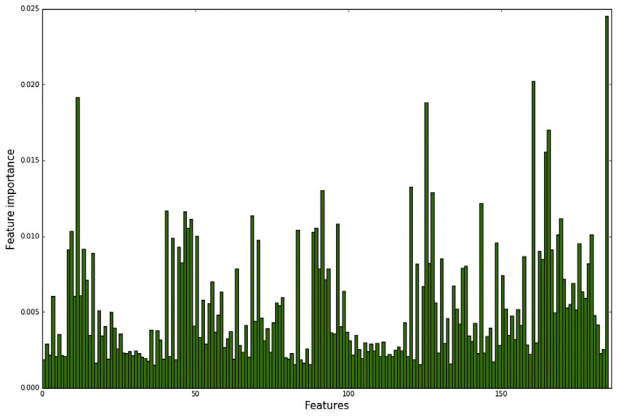


**Fig. 8 – The importance score of each feature.**

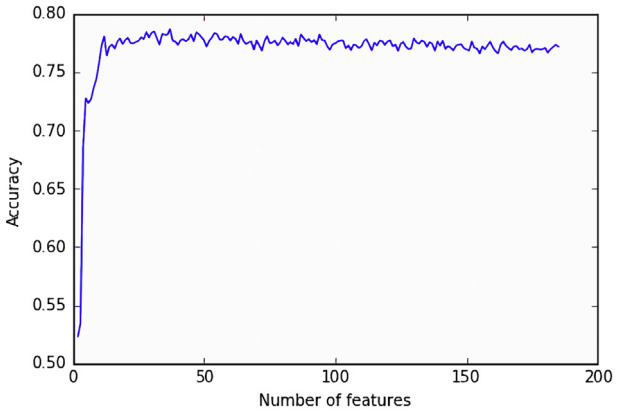


**Fig. 9 – The overall recognition rate of the model with features selected based on their importance.**

(*iii*) 185 whole features, were tested by building optimal RF for investigating their influence in overall classification rate. When tuning hyperparameters for modelling with (*i*) and (*ii*) features, their range of n value was set from 2 to 10. The rest of the tuning procedures was the same as when modelling with (*iii*) features, which was discussed above. Table 4 gives the optimal hyperparameters of these three RF models. Running the RF model three times with these parameters, the mean value and standard deviation of model metrics for each plant species are compared in Table 5. The recalls of *Z. mays* are always 1.0 in three different kinds of feature combinations, and their high values of precision also indicate that most predicted *Z. mays* are from original true labels. Compared to the features from (*i*) and (*iii*), the RF with (*ii*) performs better with the optimal parameters ($m = 234$, $n = 5$).

**Table 4 – Optimal hyperparameters for RF with different feature combinations.**

| Hyperparameters | (i) Bands + PCA | (ii) Important features | (iii) All features |
|---|---|---|---|
| Number of trees (m) | 219 | 234 | 231 |
| Number of features (n) | 6 | 5 | 8 |

Its classification rates for *C. arvensis*, *Rumex*, *C. arvense* and *Z. mays* are 0.789, 0.691, 0.752 and 1.0, respectively. Moreover, the optimal of RF was compared to the KNN model by the McNemar analysis. The result shows that $q_{01} = 30$, and $q_{10} = 15$. The value of statistic of $\Psi$ was calculated as 4.356 (P-Value $= 0.037 < 0.05$). Thus, it is concluded that the two models have significantly different performances at the 0.05 significance level and the optimal RF with the 30 important features had a better prediction performance compared to the KNN model ($q_{01} > q_{10}$).

## 4. Discussion

This paper presents a weed/maize classification system with a novel approach using snapshot hyperspectral NIR camera sensor. In respect of feature selection, only 8 single band features are included in the top 30 features (Table 3), but one of them, 871 nm, was evaluated as the most important feature (0.025). The other important features are mostly from NDVI and RVI features indicating that VIs are very significant features in the discrimination of vegetation species. This is consistent with the conclusions of the earlier studies (Jurado-Expósito, López-Granados, Atenciano, García-Torres, & González-Andújar, 2003; Smith & Blackshaw, 2003; Vrindts, De Baerdemaeker, & Ramon, 2002). Based on the definitions of VIs (Pearson & Miller, 1972), the differences between plant species in single wavelength spectrum are accentuated for classification compared with single band reflectance features.

The reflectance of vegetation is governed by the concentration and distribution of biochemical constituents and internal structure as well as leaf surface properties (Peñuelas & Filella, 1998). The combination of chlorophyll absorption and strong scattering of the light by the leaf internal cellular structure affects the red-edge (680–760 nm) reflectance of plant leaves and canopies (Ray, Murray, Chehbouni, & Njoki, 1993). In our study, the red-edge wavelengths such as 677 nm, 724 nm, 738 nm and 764 nm and near infrared bands of 871 nm, 776 nm are frequently present in the 30 important features. This finding is quite important for SSWM, whose priority task is to recognise weeds and crop (López-Granados

**Table 3 – The 30 most important features selected by the RF.**

| RVI features | NDVI features | Band reflectance |
|---|---|---|
| RVI [738, 605], RVI [776, 601], RVI [776, 652], RVI [802, 669], RVI [776, 669], RVI [776, 660], RVI [724, 677], RVI [738, 627], RVI [776, 636], RVI [814, 601], RVI [776, 614], RVI [790, 601] | NDVI [776, 644], NDVI [738, 605], NDVI [776, 601], NDVI [776, 660], NDVI [802, 627], NDVI [738, 652], NDVI [724, 677], NDVI [724, 627], NDVI [802, 669], NDVI [724, 669] | 871 nm, 814 nm, 764 nm, 644 nm, 636 nm, 601 nm, 669 nm, 677 nm |

**Table 5 – Metrics for RF with three different combinations of features.**

| Plant species | Metrics | (i) Bands + PCA | (ii) 30 important features | (iii) All features |
|---|---|---|---|---|
| *Convolvulus arvensis* | Precision | 0.938 ± 0.014 | 0.959 ± 0.008 | 0.944 ± 0.016 |
| | Recall | 0.755 ± 0.007 | 0.789 ± 0.007 | 0.781 ± 0.008 |
| | F1 | 0.836 ± 0.003 | 0.866 ± 0.004 | 0.854 ± 0.006 |
| *Rumex* | Precision | 0.589 ± 0.016 | 0.703 ± 0.019 | 0.682 ± 0.005 |
| | Recall | 0.617 ± 0.014 | 0.691 ± 0.029 | 0.654 ± 0.015 |
| | F1 | 0.603 ± 0.015 | 0.697 ± 0.024 | 0.670 ± 0.012 |
| *Cirsium arvense* | Precision | 0.617 ± 0.015 | 0.659 ± 0.013 | 0.620 ± 0.005 |
| | Recall | 0.658 ± 0.019 | 0.752 ± 0.015 | 0.742 ± 0.026 |
| | F1 | 0.637 ± 0.017 | 0.698 ± 0.011 | 0.676 ± 0.011 |
| *Zea mays* | Precision | 0.930 ± 0.006 | 0.940 ± 0.006 | 0.984 ± 0.007 |
| | Recall | 1 ± 0 | 1 ± 0 | 1 ± 0 |
| | F1 | 0.963 ± 0.003 | 0.969 ± 0.003 | 1.992 ± 0.003 |

et al., 2016). The RF model with the features selected by PCA performs even worse than no feature selection. A possible reason is ignoring the remaining less significant components. All three RF models show that it is much more difficult to recognise *Rumex* (weed2) and *C. arvense* (weed3). Especially for *Rumex*, almost one third of the plants were predicted as being *C. arvense*, indicating that these two plants may have similar spectral features in certain wavelengths. In our experiment, the *Rumex* and *C. arvense* plants were from the field rather than having been raised in pots. At the start of image data collection, the changes of environmental conditions might have resulted in some physiological stress (Tardieu, 2013), consequently similar spectral response of these two species in certain wavelengths. As well as spectral features for weed and crop classification, other features like shape and texture features are suggested to be explored for use to boost the accuracy of classification. Moreover, features like local vegetation colour descriptors or edge region features were also discussed by Kazmi, Garcia-Ruiz, Nielsen, Rasmussen, and Andersen (2015a, 2015b). From the perspective of agronomy, crop rows can assist online weed detection and management (Jones, Gée, & Truchetet, 2009; Tang, Chen, Miao, & Wang, 2016), but one of the limitations is that intra-row weeds could not be detected by this approach.

Kazmi et al. (2015b) also discussed KNN models for creeping thistle detection in sugar beet fields. It is difficult to directly compare our RF classifier to other models in the literature due to the variety of databases and plant species. In our case, the McNemar statistical analysis shows that the RF with 30 important features can achieve a better prediction performance than the KNN classification model. Generally, the choice for selecting a classification model strongly depends on specific tasks and objects, but some criteria should be considered such as number of features, number of samples, data types and research purposes (Behmann, Mahlein, Rumpf, Römer, & Plümer, 2014). Random Forests is suggested if information in data distribution is weak and unfamiliar (Breiman, 2001). Without the consideration of weed species, the classifier turns into a binary classification. Compared with multi-classification, the binary classification of weed and crop tends to obtain higher classification rates (Pantazi et al., 2016). One-class support vector machine (Choi, 2009) is suggested to be explored. It meets the real situation well, generally there is only one crop and multiple weed species in fields when considering the implementation of non-selective herbicide spraying.

Generally, line scan hyperspectral imaging has hundreds of spectral bands for exploitation. In our paper, we constructed 160 VIs from 25 spectral bands in the snapshot hyperspectral camera and then used random forests to evaluate every spectral feature importance for weed and maize classification. The results demonstrated that it is possible to recognise the three kinds of weeds and maize crop in the laboratory using the NIR snapshot mosaic hyperspectral images. However, it is important to note that there are still some limitations or considerations in order to transfer the value of the proposed method from the laboratory to real-world conditions. For example, this camera can be deployed under controlled environment (e.g. using a curtain or shelter to block natural light) if a ground-based vehicle is available (Zhang, Staab, Slaughter, Giles, & Downey, 2012). Practical settings like lens aperture, exposure, camera height and platform speed also need to be explored for obtaining high quality image data. For natural light environment, the effect of illumination compensation for hyperspectral imaging has been discussed by Wendel and Underwood (2017). The dataset used in our paper was relatively limited since our main purpose was to demonstrate the principles to recognise weeds and crop with a snapshot mosaic hyperspectral camera. Therefore, a further test is required to confirm and demonstrate robustness of the proposed classifier to changing conditions and to larger datasets compared to static data collection in laboratory.

This study will need to be followed by field experiments. The averaged spectral reflectance, not just from ROI but from whole plants, should be segmented and extracted from soil background automatically. Overall, this study showed that the use of snapshot hyperspectral data could differentiate three kinds of weeds and maize in laboratory scale. Additionally, the important features selected by random forests can be considered as the significant spectral signatures for *C. arvensis*, *Rumex*, *C. arvense* and *Z. mays* classification. The proposed approach can be further supported through other snapshot hyperspectral sensor applications like scouting early growth stage field weeds using unmanned aerial vehicles or specialised designed field vehicles.

## 5. Conclusions

In this research, the possibility of using a snapshot mosaic hyperspectral camera was evaluated for weed and maize classification. RF was tested to build classifiers with different spectral feature combinations. After feature construction, the 30 important features were selected by the accuracy-oriented feature reduction procedure based on their importance scores. It is shown that crop (*Z. mays*) can be recognised with a very high precision (94%) and recall (100%). The precision values for the three kinds of weeds, *C. arvensis*, *Rumex* and *C. arvense*, are 95.9%, 70.3%, and 65.9%, respectively. Additionally, vegetation indices are effective approaches to build significant features for the classification of weeds and crop. In particular, bands near the red edge appear frequently in the 30 important features. The findings could support further applications of this camera in the field for implementing SSWM.

## Acknowledgements

The study had been made possible by the Special Research Fund (BOF) of the Ghent University (No. 01SC3616). This study was also supported by the Technology and Food Science Unit of the Institute of Agriculture and Fisheries Research (ILVO), Belgium. The authors wish to express their gratitude to Koen Mertens, Bernard De Baets, and Wouter Saeys, also to Filip De Brouwer for his assistances in the experiment. Besides, we also thank Yingwang Gao for his critical reading in this paper.

## REFERENCES

Altman, N. S. (1992). An introduction to kernel and nearest-neighbor nonparametric regression. *American Statistician, 46*(3), 175–185. https://doi.org/10.1080/00031305.1992.10475879.

Behmann, J., Mahlein, A.-K., Rumpf, T., Römer, C., & Plümer, L. (2014). A review of advanced machine learning methods for the detection of biotic stress in precision crop protection. *Precision Agriculture*, 1–22. https://doi.org/10.1007/s11119-014-9372-7.

Belgiu, M., & Drăgu, L. (2016). Random forest in remote sensing: A review of applications and future directions. *ISPRS Journal of Photogrammetry and Remote Sensing, 114*, 24–31. https://doi.org/10.1016/j.isprsjprs.2016.01.011.

Breiman, L. (2001). Random forests. *Machine Learning, 45*(1), 5–32. https://doi.org/10.1023/A:1010933404324.

Chamundeeswari, V. V., Singh, D., & Singh, K. (2009). An analysis of texture measures in PCA-based unsupervised classification of SAR images. *IEEE Geoscience and Remote Sensing Letters, 6*(2), 214–218.

Choi, Y. S. (2009). Least squares one-class support vector machine. *Pattern Recognition Letters, 30*(13), 1236–1240. https://doi.org/10.1016/j.patrec.2009.05.007.

Dietterich, T. G. (1998). Approximate statistical tests for comparing supervised classi cation learning algorithms. *Neural Computation, 10*, 1895–1923. https://doi.org/10.1162/089976698300017197.

Dogan, M. N., Ünay, A., Boz, Ö., & Albay, F. (2004). Determination of optimum weed control timing in Maize ( Zea mays L.). *Turkish Journal of Agriculture and Forestry, 28*(5), 349–354.

Everitt, B. S. (1992). *The analysis of contingency tables. Monographs on statistics and applied probability, 45*(1990). https://doi.org/10.1002/bimj.4710350708.

Gao, J., Li, X., Zhu, F., & He, Y. (2013). Application of hyperspectral imaging technology to discriminate different geographical origins of Jatropha curcas L. seeds. *Computers and Electronics in Agriculture, 99*, 186–193. https://doi.org/10.1016/j.compag.2013.09.011.

Gao, J., Liao, W., Nuyttens, D., Lootens, P., Vangeyte, J., Pižurica, A., et al. (2018). Fusion of pixel and object-based features for weed mapping using unmanned aerial vehicle imagery. *International Journal of Applied Earth Observation and Geoinformation, 67*, 43–53. https://doi.org/10.1016/j.jag.2017.12.012.

Garcia-Ruiz, F. J., Wulfsohn, D., & Rasmussen, J. (2015). Sugar beet (Beta vulgaris L.) and thistle (Cirsium arvensis L.) discrimination based on field spectral data. *Biosystems Engineering, 139*, 1–15. https://doi.org/10.1016/j.biosystemseng.2015.07.012.

Ghosh, A., Fassnacht, F. E., Joshi, P. K., & Koch, B. (2014). A framework for mapping tree species combining hyperspectral and LiDAR data: Role of selected classifiers and sensor across three spatial scales. *International Journal of Applied Earth Observation and Geoinformation, 26*(1), 49–63. https://doi.org/10.1016/j.jag.2013.05.017.

Harker, K. N., & O'Donovan, J. T. (2013). Recent weed control, weed management, and integrated weed Management. *Weed Technology, 27*(1), 1–11. https://doi.org/10.1614/WT-D-12-00109.1.

Hsu, C., Chang, C., & Lin, C. (2008). A practical guide to support vector classification. *BJU International, 101*(1), 1396–1400. https://doi.org/10.1177/02632760022050997.

Ishida, T., Kurihara, J., Viray, F. A., Namuco, S. B., Paringit, E. C., Perez, G. J., et al. (2018). A novel approach for vegetation classification using UAV-based hyperspectral imaging. *Computers and Electronics in Agriculture, 144*, 80–85. https://doi.org/10.1016/j.compag.2017.11.027.

Jones, G., Gée, C., & Truchetet, F. (2009). Modelling agronomic images for weed detection and comparison of crop/weed discrimination algorithm performance. *Precision Agriculture, 10*(1), 1–15. https://doi.org/10.1007/s11119-008-9086-9.

Juel, A., Groom, G. B., Svenning, J. C., & Ejrnæs, R. (2015). Spatial application of Random Forest models for fine-scale coastal vegetation classification using object based analysis of aerial orthophoto and DEM data. *International Journal of Applied Earth Observation and Geoinformation, 42*, 106–114. https://doi.org/10.1016/j.jag.2015.05.008.

Jurado-Expósito, M., López-Granados, F., Atenciano, S., García-Torres, L., & González-Andújar, J. L. (2003). Discrimination of weed seedlings, wheat (Triticum aestivum) stubble and sunflower (Helianthus annuus) by near-infrared reflectance spectroscopy (NIRS). *Crop Protection, 22*(10), 1177–1180. https://doi.org/10.1016/S0261-2194(03)00159-5.

Kazmi, W., Garcia-Ruiz, F. J., Nielsen, J., Rasmussen, J., & Andersen, H. J. (2015a). Detecting creeping thistle in sugar beet fields using vegetation indices. *Computers and Electronics in Agriculture, 112*, 10–19. https://doi.org/10.1016/j.compag.2015.01.008.

Kazmi, W., Garcia-Ruiz, F. J., Nielsen, J., Rasmussen, J., & Andersen, H. J. (2015b). Exploiting affine invariant regions and leaf edge shapes for weed detection. *Computers and Electronics in Agriculture, 118*, 290–299. https://doi.org/10.1016/j.compag.2015.08.023.

Loosvelt, L., Peters, J., Skriver, H., De Baets, B., & Verhoest, N. E. C. (2012). Impact of reducing polarimetric SAR input on the uncertainty of crop classifications based on the random forests algorithm. *IEEE Transactions on Geoscience and Remote Sensing, 50*(10 PART2), 4185–4200. https://doi.org/10.1109/TGRS.2012.2189012.

López-Granados, F. (2011). Weed detection for site-specific weed management: Mapping and real-time approaches. *Weed Research, 51*, 1–11. https://doi.org/10.1111/j.1365-3180.2010.00829.x.

López-Granados, F., Torres-Sánchez, J., Serrano-Pérez, A., de Castro, A. I., Mesas-Carrascosa, F. J., & Peña, J. M. (2016). Early season weed mapping in sunflower using UAV technology: Variability of herbicide treatment maps against weed thresholds. *Precision Agriculture, 17*(2), 183–199. https://doi.org/10.1007/s11119-015-9415-8.

Macías, F. A., Castellano, D., & Molinillo, J. M. G. (2000). Search for a standard phytotoxic bioassay for allelochemicals. Selection of standard target species. *Journal of Agricultural and Food Chemistry, 48*(6), 2512–2521. https://doi.org/10.1021/jf9903051.

Meissle, M., Mouron, P., Musa, T., Bigler, F., Pons, X., Vasileiadis, V. P., et al. (2010). Pests, pesticide use and alternative options in European maize production: Current status and future prospects. *Journal of Applied Entomology, 134*(5), 357–375. https://doi.org/10.1111/j.1439-0418.2009.01491.x.

Nitze, I., Barrett, B., & Cawkwell, F. (2015). Temporal optimisation of image acquisition for land cover classification with random forest and MODIS time-series. *International Journal of Applied Earth Observation and Geoinformation, 34*(1), 136–146. https://doi.org/10.1016/j.jag.2014.08.001.

Oerke, E. C. (2006). Crop losses to pests. *The Journal of Agricultural Science, 144*(01), 31. https://doi.org/10.1017/S0021859605005708.

Oerke, E. C., & Steiner, U. (1996). Estimation of yield losses in maize cultivation: Yield losses and crop protection-the growing situation for the economically most important crops. *German Phytomedical Society Series, 6*, 63–79.

Okamoto, H., Murata, T., Kataoka, T., & Hata, S.-I. (2007). Plant classification for weed detection using hyperspectral imaging with wavelet analysis: Research paper. *Weed Biology and Management, 7*(1), 31–37. https://doi.org/10.1111/j.1445-6664.2006.00234.x.

Ostrý, V., Malíř, F., & Pfohl-Leszkowicz, A. (2015). Comparative data concerning aflatoxin contents in bt maize and non-Bt isogenic maize in relation to human and animal health-a review. *Acta Veterinaria Brno, 84*(1), 47–53. https://doi.org/10.2754/avb201585010047.

Pantazi, X. E., Moshou, D., & Bravo, C. (2016). Active learning system for weed species recognition based on hyperspectral sensing. *Biosystems Engineering*, 1–10. https://doi.org/10.1016/j.biosystemseng.2016.01.014.

Pearson, R. L., & Miller, L. D. (1972). Remote mapping of standing crop biomass for estimation of the productivity of the shortgrass prairie. *Remote Sensing of Environment, VIII*. -1, 1355. Retrieved from: http://adsabs.harvard.edu/abs/1972rse.conf.1355P\npapers2://publication/uuid/4ABE1403-D3C3-40B7-B0A7-6C2C1BFDA6E1.

Peñuelas, J., & Filella, L. (1998). Visible and near-infrared reflectance techniques for diagnosing plant physiological status. *Trends in Plant Science, 3*(4), 151–156. https://doi.org/10.1016/S1360-1385(98)01213-8.

Peters, J., De Baets, B., Verhoest, N. E. C., Samson, R., Degroeve, S., De Becker, P., et al. (2007). Random forests as a tool for ecohydrological distribution modelling. *Ecological Modelling, 207*(2–4), 304–318. https://doi.org/10.1016/j.ecolmodel.2007.05.011.

Ray, T. W., Murray, B. C., Chehbouni, A., & Njoki, E. (1993). The red edge in arid region vegetation: 340-10160 nm spectra. In *AVIRIS airborne geoscience workshop* (pp. 149–152).

Romeo, J., Guerrero, J., Montalvo, M., Emmi, L., Guijarro, M., Gonzalez-de-Santos, P., et al. (2013). Camera sensor arrangement for crop/weed detection accuracy in agronomic images. *Sensors, 13*(4), 4348–4366. https://doi.org/10.3390/s130404348.

Shaner, D. L., & Beckie, H. J. (2014). The future for weed control and technology. *Pest Management Science, 70*(9), 1329–1339. https://doi.org/10.1002/ps.3706.

Shaw, D. R. (2005). Remote sensing and site-specific weed management. *Frontiers in Ecology and the Environment, 3*(10), 526–532. https://doi.org/10.1890/1540-9295(2005)003[0526:RSASWM]2.0.CO;2.

Slaughter, D. C., Giles, D. K., & Downey, D. (2008). Autonomous robotic weed control systems: A review. *Computers and Electronics in Agriculture, 61*(1), 63–78. https://doi.org/10.1016/j.compag.2007.05.008.

Smith, A. M., & Blackshaw, R. E. (2003). Weed-crop discrimination using remote sensing: A detached leaf experiment. *Weed Technology, 17*(4), 811–820. https://doi.org/10.1614/WT02-179.

Sokolova, M., & Lapalme, G. (2009). A systematic analysis of performance measures for classification tasks. *Information Processing and Management, 45*(4), 427–437. https://doi.org/10.1016/j.ipm.2009.03.002.

Stehman, S. V. (1997). Selecting and interpreting measures of thematic classification accuracy. *Remote Sensing of Environment, 62*(1), 77–89. https://doi.org/10.1016/S0034-4257(97)00083-7.

Tang, J.-L., Chen, X.-Q., Miao, R.-H., & Wang, D. (2016). Weed detection using image processing under different illumination for site-specific areas spraying. *Computers and Electronics in Agriculture, 122*, 103–111. https://doi.org/10.1016/j.compag.2015.12.016.

Tardieu, F. (2013). Plant response to environmental conditions: Assessing potential production, water demand, and negative effects of water deficit. *Frontiers in Physiology, 4*, 1–11. https://doi.org/10.3389/fphys.2013.00017.

Tellaeche, A., Pajares, G., Burgos-Artizzu, X. P., & Ribeiro, A. (2011). A computer vision approach for weeds identification through Support Vector Machines. *Applied Soft Computing, 11*(1), 908–915. https://doi.org/10.1016/j.asoc.2010.01.011.

Thenkabail, P. S., Mariotto, I., Gumma, M. K., Middleton, E. M., Landis, D. R., & Huemmrich, K. F. (2013). Selection of hyperspectral narrowbands (hnbs) and composition of hyperspectral twoband vegetation indices (HVIS) for biophysical characterization and discrimination of crop types using field reflectance and hyperion/EO-1 data. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 6*(2), 427–439. https://doi.org/10.1109/JSTARS.2013.2252601.

Thompson, J. F., Stafford, J. V., & Miller, P. C. H. (1991). Potential for automatic weed detection and selective herbicide application. *Crop Protection, 10*(4), 254–259. https://doi.org/10.1016/0261-2194(91)90002-9.

Torres-Sánchez, J., López-Granados, F., De Castro, A. I., & Peña-Barragán, J. M. (2013). Configuration and specifications of an Unmanned Aerial Vehicle (UAV) for early site specific weed management. *PLoS One, 8*(3), e58210. https://doi.org/10.1371/journal.pone.0058210.

Viña, A., Gitelson, A. A., Nguy-Robertson, A. L., & Peng, Y. (2011). Comparison of different vegetation indices for the remote assessment of green leaf area index of crops. *Remote Sensing of Environment, 115*(12), 3468–3478. https://doi.org/10.1016/j.rse.2011.08.010.

Vrindts, E., De Baerdemaeker, J., & Ramon, H. (2002). Weed detection using canopy reflection. *Precision Agriculture, 3*, 63–80. https://doi.org/10.1023/A:1013326304427.

Wendel, A., & Underwood, J. (2016). Self-supervised weed detection in vegetable crops using ground based hyperspectral imaging. In *2016 IEEE international conference on robotics and automation (ICRA)* (pp. 5128–5135). https://doi.org/10.1109/ICRA.2016.7487717.

Wendel, A., & Underwood, J. (2017). Illumination compensation in ground based hyperspectral imaging. *ISPRS Journal of Photogrammetry and Remote Sensing, 129*, 162–178. https://doi.org/10.1016/j.isprsjprs.2017.04.010.

Wold, S., Esbensen, K., & Geladi, P. (1987). Principal component analysis. *Chemometrics and Intelligent Laboratory Systems, 2*(1–3), 37–52. https://doi.org/10.1016/0169-7439(87)80084-9.

Zhang, Z. P. (2003). Development of chemical weed control and integrated weed management in China. *Weed Biology and Management, 3*(4), 197–203. https://doi.org/10.1046/j.1444-6162.2003.00105.x.

Zhang, Y., Staab, E. S., Slaughter, D. C., Giles, D. K., & Downey, D. (2012). Automated weed control in organic row crops using hyperspectral species identification and thermal micro-dosing. *Crop Protection, 41*, 96–105. https://doi.org/10.1016/j.cropro.2012.05.007.