Hyperspectral Selection Based Algorithm for Plant Classification

DOI:

[10.1109/IST.2016.7738258](https://doi.org/10.1109/IST.2016.7738258)

Document Version

Accepted author manuscript

[Link to publication record in Manchester Research Explorer](#)

Citation for published version (APA):

Alsuwaidi, A., Veys, C., Hussey, M., Grieve, B., & Yin, H. (2016). Hyperspectral Selection Based Algorithm for Plant Classification. In *IEEE International Conference on Imaging Systems and Techniques* IEEE. <https://doi.org/10.1109/IST.2016.7738258>

Published in:

IEEE International Conference on Imaging Systems and Techniques

Citing this paper

Please note that where the full-text provided on Manchester Research Explorer is the Author Accepted Manuscript or Proof version this may differ from the final Published version. If citing, it is advised that you check and use the publisher's definitive version.

General rights

Copyright and moral rights for the publications made accessible in the Research Explorer are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

Takedown policy

If you believe that this document breaches copyright please refer to the University of Manchester's Takedown Procedures [<http://man.ac.uk/04Y6Bo>] or contact uml.scholarlycommunications@manchester.ac.uk providing relevant details, so we can investigate your claim.

Hyperspectral Selection Based Algorithm for Plant Classification

Ali AlSuwaidi*, Charles Veys, Martyn Hussey, Bruce Grieve, and Hujun Yin*

School of Electrical and Electronic Engineering
The University of Manchester
Manchester, M13 9PL, UK

*E-mail: ali.bghalsuwaidi@postgrad.manchester.ac.uk, hujun.yin@manchester.ac.uk

Abstract—The popularity of using hyperspectral imaging systems in studying and monitoring plant properties and conditions has increased lately. This increase has been driven by both financial and environmental advantages of such systems. Using a nondestructive hyperspectral imaging system improves the breeding process, increases profit, and reduces the usage of herbicide, thus reducing side effects to plants and environment. This paper is concerned with the use of hyperspectral image analysis for differentiating different plant species as well as their conditions. The main contribution of the work lies in the use of feature selection for choosing relevant, discriminant spectral information as the input to the classifier (e.g. SVM), as compared to the use of empirical spectral indices. Two independent hyperspectral datasets, captured by different instrumentations, were used in evaluation. Experimental results show significant improvements in classification accuracy with several feature selection algorithms compared to with the spectral vegetation and disease indices. The study shows that systematically selection of wavelength features can shed light on attributes that differentiate plants and their conditions.

Keywords—feature selection; hyperspectral imaging; spectral indices; support vector machine

I. INTRODUCTION

Hyperspectral imaging is a branch of multivariate imaging [1], started with the discovery of imaging spectrometry in the early 1980s by the NASA Jet Propulsion Laboratory. It has since attracted considerable amount of interest in innovative scientific quests [2]. It is a combination of two technologies: spectroscopy (study of spectral properties) and remote imaging (study of optical properties). Hyperspectral images can be acquired using point, line or plane scanning configuration [3]. The importance of this emerging imaging branch has been rising in the last three decades because it senses a wider range of electromagnetic spectrum as well as gathers a large number of narrow band spectra [3]. It has been utilised in a numbers of applications for instances, medical imaging [4], agricultural monitoring [5, 6], and industrial [7] and chemical processes [8].

Activities in using hyperspectral imaging for studying plant properties and conditions have increased considerably in the last few years due to many financial and environmental benefits [5, 6, 9]. Improving production is an example of financial advantages and reducing herbicide usage is an example of environmental benefits. It has been reported that

identifying weed species help control the use of herbicide process, reduce the resistance of weed species to herbicide, and reduce the pollution effect of herbicide to the environment [9].

Discriminating between weeds and crops is considered as one of the greatest challenges in agriculture [6]. In hyperspectral imaging based approach varying spectral vegetation indices (SVI) have been developed to study plant properties and conditions. For example, four spectral disease indices (SDI) have been proposed and used in order to detect and identify the healthy crops from the unhealthy ones [11]. A good list of these indices can be found in [10].

Recent developments in hyperspectral imaging systems have heightened the need for an efficient and effective analysis method as a large amount of data is generated and it is difficult to analyse the information directly from pixel values. Machine learning is considered as one of the effective analysis tools. However, the main obstacles are finding the relevant information to the problem and dealing with the overfitting issues [12-14]. The former can be alleviated by using feature selection techniques. Feature selection is a process of selecting relevant features to the problem and discards redundant and irrelevant features [12].

Recent years have seen increasing amount of literature on feature selection and its applications. Moreover, there is a large volume of published studies introducing new feature selection algorithms for solving various problems. For instance, a multi objective heuristic algorithm based on analytic hierarchy process has been introduced in clinical medicine to find a discriminatory subset of genes that help diagnose and treat cancer [28]. Another study adopts the existing algorithms to select significant features from combined multi scale and different origin data and use them in biotechnological applications such as strain selection in winemaking [29].

This paper focuses on classifying different crops using feature selection algorithms and comparing the results with that of using the existing spectral indices. The selected relevant features (wavelengths or their combinations) will then be used in the classification stage. Support vector machine (SVM) was used since it is considered as a state-of-the-art classifier [15] and can deal effectively with curse of dimensionality problem [16], hence reducing the risk of overfitting. The classification results have verified that markedly improved discrimination can be achieved.

The reminder of the paper is organised as follows. An overview of the existing spectral indices is given in section II. Various feature selection algorithms are described in section III. Section IV presents the imaging system, datasets and corresponding analysis. The experimental results and discussions are given in section V and followed by the concluding remarks in section VI.

II. SPECTRAL INDICES

A considerable number of studies have been conducted on the use of spectral information in plant classification and several different spectral indices have been proposed. The primary formulations of these indices, developed mainly for measuring chlorophyll content, are the simple ratio (SR) between reference and index reflectance or the normalised difference (ND) between them [24],

$$SR = \frac{R_{ref}}{R_{index}} \quad (1)$$

$$SR = \frac{R_{ref} - R_{index}}{R_{ref} + R_{index}} \quad (2)$$

where R_{ref} represents the reference reflectance and R_{index} represents the index reflectance at a specific wavelength. It should be mentioned that there are few indices which are adapted from the general forms such as soil adjusted vegetation index (SAVI) [25]. A good list of various spectral indices can be found in [10], while the list shown in Table I represents those widely used and also in our experimentation.

The existing indices have been used in many applications such as detection of plant stress [24] and senescing [26]. The spectral signature of normal plant leaf varies in different seasons due to variation in pigment concentration level. That means few wavelengths especially the one associated with the green pigment can be picked and used in simple ratio or normalised difference form to detect stress or senescing level.

TABLE I. LIST OF EMPLOYED INDICES

Index	Formula	Purpose
SR	R_{695}/R_{760}	Plant stress detection
ND	$(R_{750} - R_{705})/(R_{750} + R_{705})$	Plant senescing
MND	$((R_{750} - R_{445}) / ((R_{750} + R_{705} - 2R_{455})))$	Plant pigments estimation
TCARI	$3 \times ((R_{700} - R_{670}) - 2(R_{700} - R_{550})) \times (R_{700}/R_{670})$	Chlorophyll estimation
RTO	$TCARI/OSAVI$	Chlorophyll estimation
PSRI	$(R_{680} - R_{500})/R_{750}$	Plant senescing
HI	$(R_{534} - R_{698})/(R_{534} + R_{698}) - (\frac{1}{2} \times R_{704})$	Disease detection for sugar crops
SBRI	$(R_{570} - R_{513})/(R_{570} + R_{513}) + (\frac{1}{2} \times R_{704})$	Disease detection for sugar crops
CLSI	$(R_{698} - R_{570})/(R_{698} + R_{570}) - R_{734}$	Disease detection for sugar crops
PMI	$(R_{520} - R_{584})/(R_{520} + R_{584}) + R_{724}$	Disease detection for sugar crops

Fig. 1. Red edge. Spectral reflectance of Galant wheat (data clipped before 670 and after 780 nm). Captured on August 21st, 2014 (using UoM hyperspectral imaging system)

It has been reported that the visible spectrum can be used to detect the stressed and senescence leaves since the pigment concentration levels in those condition are reduced by 40% to 50% [26]. Also stress could be detected using the red edge position (REP) [27]. In this case the red edge need to be found $R_{red\ edge} = (R_{670} + R_{780})/2$, which is defined as the maximum slope of the reflectance between red and near infrared (NIR) and then substitute this value in Eqn. (3). Fig. 1 illustrates the red edge of single wheat species.

$$REP = (700 + 40) \times \left[\frac{R_{red\ edge} - R_{700}}{R_{740} - R_{700}} \right] \quad (3)$$

where $R_{red\ edge}$, R_{740} , R_{700} represents the reflectance at calculated red edge, 740 nm, and 700 nm respectively.

Another domain where many studies have been conducted in monitoring plant conditions and properties is weed mapping. One of these studies developed a hyperspectral machine vision system based on three multivariate Bayesian classifiers to separate between tomato and weeds over three seasons [5]. The developed system uses one classifier for each season and there is a global calibration mechanism among all three seasons. The spectral signature over the entire spectrum of their system was used and the reported results showed a good classification level. Another study proposed an automatic system that identifies the weed location using normalised difference vegetation index (NDVI), hue of the entire spectrum, the difference between near infrared (NIR) and red spectrum, and the difference between NIR, red, and hue [6]. The herbicide will only be applied to the weed species, thus reduce the effect to the other species as well as the environment. Moreover, another study was conducted using visible and NIR hyperspectral imaging system to recognise the wheat from different weed species as well as the background (soil) [9]. The recognition technique is based on partial least squares discriminant analysis and the reported result shows a good agreement between the proposed method and field assessment.

In addition, hyperspectral based indices can play a good role in detecting and identifying different disease. There is a study proposing four different disease indices (given in table I)

to discriminate between healthy and unhealthy sugar leaves [11]: i.e. healthy index (HI), sugar beet rust index (SBRI), Cercospora leaf spot index (CLSI), and powdery mildew index (PMI). In this study, the authors considered the combination of single wavelengths and normalised wavelengths difference as the best to develop a disease index. A distance criterion based feature selection algorithm (Relief-F) was used find the best combination for discrimination.

III. FEATURE SELECTION

Data collection process is fundamental to systematic investigation and analysis. The more data collected, the more problem aspects it may cover. However, not all data contains useful or relevant information to the problem investigated. In other words, the collected data can include irrelevant, noise and redundant information. The process of choosing a subset of relevant features and removing irrelevant and redundant one's is called feature selection [12, 17]. In other words, it aims to identify a minimal subset of available features that have the similar performance as the full features (if not better) based on specific evaluation criteria. There are large numbers of studies that suggest how relevant feature x_i can be identified from a feature space $X = \{x_1 \dots x_n\}$ for a specific dataset D . For example, a single feature x_i can be considered as a relevant feature in X if it can be used to distinguish between two different classes $\{c_1, c_2\}$ in the dataset D . The main advantage of using the relevant feature subset is improving the processing time as well as the prediction quality of classification problem, while it is not always the case if all features are used [12].

The process of feature selection can be described in four steps: search organisation, subset evaluation, stopping criteria, and result validation [12, 17, 18]. The first step is responsible for generating different subsets based on certain search strategy. Search strategy includes determining the starting point (forward, backward, or bidirectional) as well as the procedure that describes how feature space is going to be explored (random, sequential, or complete). The second step evaluates the goodness of each subset according to a specific evaluation criterion to estimate the optimal subset. Probability of error, distance, and information are examples of evaluation criteria. The third step defines when the process should be halted (example: reach certain threshold) while the last one determine if the selected subset is the significant one, or stopping criterion is the best to halt the process.

A. Feature Selection Models

Feature selection methods can be broadly categorised, based on evaluation criteria, into three models: filter, wrapper, and embedded [12,13, 17, 18]. In filter model the capability of discriminating different classes depends only on data characteristics. This means the evaluation criteria is independent and it does not require any classification algorithm to generate the optimal subset. The evaluation criteria in this model include distance, information, consistency and dependency while the output of this model is either given as a weighted features or single subset of relevant features. Table II describes the generalised filter algorithm, while Relief-F and correlation based feature selection (CFS) being common examples of this model.

TABLE II. GENERALISED FILTER ALGORITHM

General Filter Algorithm
Requirements: D : data set X : feature space with x_i feature Y : label J : evaluate criteria Feature Selection Process: $S_{initial}$: initialise feature subset (empty or full) Start: S_{new} : add or remove feature to the subset based on $S_{initial}$ Evaluate(S_{new}, D, J) If $J(S_{new})$ is better than $J(S_{initial})$ $S_o = S_{new}$: optimal subset Repeat if stopping criteria == false end

TABLE III. GENERALISED WRAPPER ALGORITHM

General Wrapper Algorithm
Requirements: D : data set X : feature space with x_i feature Y : label CA : classification algorithm Feature Selection Process: $S_{initial}$: initialise feature subset (empty or full) Start: S_{new} : add or remove feature to the subset based on $S_{initial}$ Evaluate(S_{new}, D, CA) If $CA(S_{new})$ performance is better than $CA(S_{initial})$ $S_o = S_{new}$: optimal subset Repeat if stopping criteria == false end

Wrapper model uses a classification algorithm to assess the goodness of the selected features and this is the main difference to the filter model [12,13, 17, 18]. The pseudo code is presented in table III. In this model the selected features are tailored for the classification algorithm and the feature selection process is iterated through the classification process. This means the selected features may not perform well with other classifiers. Sequential forward/backward generation (SFG/SBG) are two examples of this model. The processing time of the forward algorithm is usually less than the backward one. However, the speed of both algorithms or wrapper model in general is slower than the filter model.

The embedded model was introduced to take the advantages of the filter and wrapper models; and it avoids predefined stopping criteria, and can handle large number of data [12,13, 17, 18]. This model uses the data characteristics to determine a group of good subsets, while uses a classification algorithm to find the best subset amongst the good one. A common example of this model is the sparse Bayesian multinomial logistic regression (SBMLR) [19].

In term of implementation [13], the wrapper and embedded models are easy to implement when an existing categorising scheme is used. On the other hand, wrapper might require substantial computational capability especially with large datasets. Whilst, filter model can produce acceptable to good performances in less time compared to the other two models. It should be mentioned that the best model to choose should be

based on the problem and the application. The following subsections describe three feature selection algorithms.

B. Relief-F

A relief is a weighting based algorithm [20, 21] that deals with two-class problems. This algorithm uses the distance between the instances to determine the nearest neighbour from the same (nearest hit) and other (nearest miss) class. The instances are randomly picked and used in the evaluation criterion, this criterion can be represented in (4):

$$J(x_i) = d(x_i - x_{NM}) - d(x_i - x_{NH}) \quad (4)$$

where $J(x_i)$ represents the evaluation criterion, $d(\cdot)$ represents distance function, x_{NM}, x_{NH} , represent the nearest miss and hit respectively. This result of this criterion is used to update the weight vector (initially set to 0) that the highest weight values represent the best feature for class separation:

$$W_{i+1} = W_i - d(x_i - x_{NH}) + d(x_i - x_{NM}) \quad (5)$$

This algorithm was extended to handle multiclass problem and the extended algorithm called Relief-F [21]. The evaluation criterion of this algorithm can be represented by (6):

$$J(x_i) = \sum_{x_j \in NH(C_1)} -d(x_i - x_{NH}) + \sum_{C_r \neq C_1} \frac{P(C_r)}{1 - P(C_1)} \sum_{x_j \in NM(C_r)} d(x_i - x_{NM}) \quad (6)$$

where C_1, C_r represents class 1 and the remaining classes respectively and $P(\cdot)$ denotes the probability. It should be noted that the main difference between this algorithm and the original one is the contribution of the average misses and hits to the evaluation criterion.

C. Correlation Based Feature Selection (CFS)

In this algorithm the correlation between the features and the classes is measured using symmetrical uncertainty (SU) [22]. The reason of this is to minimise the information gain bias introduced to the features and to normalise the value between zero and one. SU can be represented as following equation,

$$SU = 2 \times \left[\frac{I(X, Y)}{H(X) + H(Y)} \right] \quad (7)$$

where $I(X, Y) = H(X) - H(X|Y)$ represents information gain (used as a criterion in our study) and $H(\cdot) = -\sum_i P(\cdot) \log_2(P(\cdot))$ defines the entropy. Entropy measures the level of the uncertainty in data; while information gain gauges the dependency of the data and the output, i.e. amount of information that single variable gives about the other [30]. The output of the correlation is used to evaluate the feature based on the following criterion:

$$Merit_s = \frac{\overline{nr_{cf}}}{\sqrt{n + n(n-1)\overline{r_{ff}}}} \quad (8)$$

where $Merit_s$ represents the heuristic merit of a subset containing n features, $\overline{r_{cf}}$ represents the average feature-class correlation, $\overline{r_{ff}}$ represents the average feature-feature

correlation. It should be mentioned that the search strategy in this algorithm is based on best first search and the process halts if no improvement is achieved after five consecutive iterations.

D. Sequential Forward Selection (SFG)

This is an iterative algorithm, the feature space is explored sequentially, the features are evaluated using a classification algorithm, and relevant features are added to a subset, which is initially an empty set [13]. The process will be terminated if adding new features does not lead to any improvement. The general procedure of this algorithm is similar to the one shown in table III (starting with empty relevant feature subset and add one feature at a time).

In our experiments the described three algorithms were used to select the significant wavelengths. The selected wavelengths of each algorithm were then passed to the classifier, the support vector machine with radial basis function (RBF), to classify the crops. The classification results were then compared with that of using all wavelengths as well as with that of using the indices. Details of the experiments, systems and datasets are described next.

IV. MATERIALS AND METHOD

A. Hyperspectral Imaging System

The datasets were collected from two separate hyperspectral imaging systems of different specifications. One of them is located at the University of Manchester (UoM) while the other at University of Bonn (Bonn). The UoM system operates over both the visible (VIS) and near infrared (NIR) regions with an effective image size of 5184×3456 pixels. This system consists of a Canon camera (EOS1200D) with 24-bits over the RGB channel (8-bits per channel), an adjustable monochromator system (cornerstone 130 1/8m - provided by Newport) that's include broadband bulb, filter wheel, and diffraction grating to separate the spectrum range from 380 to 1100 nm into 5 nm steps. The system operates in a controlled environment (dark box) in order to minimise the effect of the unwanted noise. The Bonn system is a line scanning system that operates over the range of 400 to 1000 nm with 2.8 nm spectral resolution and a spatial resolution of 0.29 mm [11]. the maximum effective line can be achieved with system is 1600 pixels and it is surrounded by six lamps provided by Analytical Spectral Devices Inc., Boulder, USA. More information about the orientation and pre-processing can be found in [11, 23].

B. Datasets

Two datasets with different acquisition dates and exposure times were used for the analysis purpose. The UoM and Bonn datasets was captured by UoM and Bonn hyperspectral imaging systems respectively. The UoM dataset consists of weeds and non-weed species (potato, wheat, and oil seed), while the Bonn dataset contains healthy and unhealthy (Cercospora and rust) sugar species. It should be mentioned that all datasets were prevented from saturation (dynamic range management). Also all datasets were normalised using reference patches; the UoM dataset was spectrally normalised to a reference grey tile (included in the image scene) that has a

Fig. 2. Plant leaves sample. Non-standard RGB representation of (a) UoM steme weed captured on 2015 (b) Bonn healthy sugar capture on 2013

flat reflection coefficient over the entire wavelength range, while the Bonn dataset was spectrally normalised to a white (barium sulphate) reference tile. Fig. 2 illustrates a sample of the UoM and Bonn datasets.

C. Feature Selection vs. Spectral Indices

Three feature selection algorithms (Relief-F, CFS, and SFG) were used to evaluate all wavelengths and estimate the set of relevant wavelengths for each algorithm. In addition information gain (measured in CFS) was used to rank the features in descending order (only the top ones were selected). The relevant wavelength sets were then passed to a SVM classifier (with RBF kernel) to measure the ability of discriminating healthy from unhealthy (classes: healthy, Cercospora, and rust) species in case of Bonn dataset and the weed species from non-weed species (classes: weed, wheat, oilseed, and potato) in case of UoM dataset. The classification results are then compared with the classification result of using all wavelengths as well as the existing spectral indices. It should be mentioned that 50% of the samples were used for the training and the remaining for the testing with 8-fold cross validation.

V. EXPERIMENTAL RESULTS AND DISCUSSION

The experiments investigated the usefulness of feature selection algorithms in plant type and condition classification. The procedure of this experiment was divided into four steps.

First, leaf reflectance extraction; twenty pixels from the leaf area were selected and averaged over the entire wavelength range to get the spectral signature. Averaged reflectance was used to reduce the variation in pixels intensities, i.e. green pixel

TABLE IV. BONN AND UoM DATASETS CLASSIFICATION RESULTS

Technique	Average Classification Rate (%) (Standard Deviation)	
	Bonn dataset	UoM dataset
All Wavelengths	77.43 (4.31)	52.08 (0.41)
SR: R_{695}/R_{760}	70.63 (2.43)	56.25 (3.74)
ND: $(R_{750} - R_{705})/(R_{750} + R_{705})$	73.26 (1.92)	41.67 (6.15)
MND	66.60 (2.29)	66.67 (5.69)
TCARI	59.79 (3.44)	70.83 (5.43)
RTO	61.81 (3.54)	66.67 (5.67)
PSRI	63.89 (2.11)	56.25 (4.88)
HI	64.31 (1.26)	33.33 (4.43)
RI	66.60 (2.21)	27.08 (4.56)
CI	72.57 (2.14)	27.08 (4.53)
PMI	64.86 (1.29)	27.08 (4.74)
Relief-F	89.24 (2.04)	79.63 (4.88)
CFS	88.75 (1.39)	81.48 (4.96)
Info. Gain	87.71 (2.16)	77.08 (4.52)
SFG-SVM	80.56 (1.97)	75.19 (4.40)

values. Second, spectral vegetation and disease indices measurement; the spectral information were used to measure these indices [10, 11]. Third, selecting relevant wavelengths subset and last, comparing the prediction performance. The classification results are presented in table IV, while the selected wavelengths for both datasets is given in table V. All the classification rates were the average of 10 runs.

Several reports and articles have shown the importance of the spectral vegetation indices in speciation and the spectral disease indices in disease detection. Recently many previous studies have noted the importance of feature selection techniques in classification problems. The experimental part of the current study has found that feature selection techniques can play an important role in improving the prediction performance of the classification process. This can be clearly demonstrated in table IV. All feature selection algorithms improve the classification accuracy by at least 3% in case of Bonn dataset and at least 5% in case of UoM dataset. It can be noted also that the separation ability of wrapper model may be affected in case of forward selection [13]. At the beginning the best feature that separates the classes is selected in the first round. In the second round, this feature is combined with another feature and this might limit the separation ability. As seen from table IV, SFG has the minimum rate of improvement in both datasets, possibly due to this problem.

Some overlap of wavelengths in two or more algorithms has been noted and this is clearly indicated in the table V (as bold). That means all feature selection algorithms consider the-

TABLE V. SIGNIFICANT WAVELENGTHS OF THE UoM DATASET

Algorithm	Selected Wavelength (nm)
Relief-F	510 , 515, 520, 525 , 530 , 535, 540, 545, 550, 555, 560, 565 , 570, 575, 580, 585, 590, 595, 600, 605, 610 , 660, 670 , 675
CFS	380 , 430 , 525 , 530 , 565 , 575 , 645 , 670 , 675 , 700, 720, 730
SFG	380 , 425 , 575 , 680
Info. Gain	380 , 385, 390, 395, 400, 405, 410, 415, 420, 425 , 430, 435, 440, 445, 450, 455, 460, 465, 470, 475, 480, 485, 490, 500, 505, 510 , 610 , 615, 620, 625, 630, 635, 640, 645 , 650, 655

-se wavelengths as significant but with different priority (wavelengths sorted in ascent priority). Another important point is that the classification rate of CFS over these data sets is consistently better than most if not all the other feature selection methods (i.e. SFG-SVM, Info. Gain and relief-F). This may attribute to the ability of the former to discard the redundant or irrelevant features compared with the others. In addition, the number of selected wavelengths by CFS is consistently lower than the most.. The wavelengths selected by these feature selection algorithms and those used by the empirical indices share only few common wavelengths (e.g. 550 nm, 670 nm, 680 nm and 700 nm among these algorithms) in the UoM dataset, while (e.g. 513 nm, 570 nm, 584 nm, 698 nm, 704 nm, 724 nm, and 734 nm among these algorithms) in Bonn dataset. This may explain why empirical indices can only achieve limited discrimination. Systematic feature selection and combinations of the selected features can provide significantly enhanced separation ability. The advantage will become even substantial with wider spectra.

The classification rate of control, *Cercospora* and rust indices using non-imaging hyperspectral data has been reported in [11] as 89%, 92%, and 87% respectively using one against all method (class one represents single symptom, while class two combines other symptoms). The accuracy achieved using a validation set (hyperspectral images captured by Bonn system in 2008) was 95.91% for healthy index (HI) and 95.9% for *Cercospora* one (CLSI). In our case, the best classification rate achieved using the same method (one against all – on Bonn dataset captured in 2013) is 96.36%, while the improvement compared to other spectral indices is at least 15%.

VI. CONCLUSION

In this paper, the aim was to assess the usefulness of feature selection algorithms in hyperspectral image data for discrimination of plant types and conditions. The experiments confirm that using only the relevant wavelengths in classification is better than using the whole feature space (mixture of relevant, redundant, and irrelevant features) or the existing spectral indices. The results indicate the significant improvements of the spectral feature selection in plant discriminating. The finding can be used to help in weed management, thus improving the productivity and reducing the effect of the herbicide on the environment. The current study sets a baseline method for plant types and conditions discrimination. It shows a valid approach, though the resulting features may not be applicable to other plant species or different condition not covered than the data used in this work. Future studies will explore combination or fusion feature selection methods for more robust performances, as well as using additional image properties such as textures to best identify plant types and conditions (stressed and diseased).

ACKNOWLEDGMENT

The authors would like to thank Anne-Katrin Mahlein and University of Bonn for provision of the Bonn dataset.

REFERENCES

- [1] H. F. Grahn and P. Geladi, *Techniques and Applications of Hyperspectral Image Analysis*. West Sussex, England: John Wiley & Sons Ltd, 2007.
- [2] A. F. H. Goetz et al., "Imaging spectrometry for earth remote sensing," *Sci.*, vol. 228, no. 4704, pp. 1147-1153, 1985.
- [3] P. L. Geladi, H. F. Grahn, and J. E. Burger, "Multivariate images, hyperspectral imaging: background and equipment," *Techniques and Applications of Hyperspectral Image Analysis*, pp. 1-15: John Wiley & Sons, Ltd, 2007.
- [4] A. B. GH ALSUWALDI, "De-noising hyperspectral images," MSc. dissertation, Dept. Elect. Eng., University of Manchester, Manchester, UK, 2011.
- [5] Y. Zhang, D. C. Slaughter, and E. S. Staab, "Robust hyperspectral vision-based classification for multi-season weed mapping," *ISPRS J. Photogramm. Remote Sens.*, vol. 69, pp. 65-73, 2012.
- [6] H. Liu, C. Saunders, and S. H. Lee, "Development of a proximal machine vision system for off-season weed mapping in broadcast no-tillage fallows," *J. Comput. Sci.*, vol. 9, no. 12, pp. 1803-1821, 2013.
- [7] C. Duchesne, J. J. Liu, and J. F. MacGregor, "Multivariate image analysis in the process industries: A review," *Chemom. Intell. Lab. Syst.*, vol. 117, pp. 116-128, 2012.
- [8] P. Geladi et al., "Image analysis in chemistry. I. Properties of images, greylevel operations, the multivariate image," *Trends Anal. Chem.*, vol. 11, no. 1, pp. 41-53, 1992.
- [9] I. Herrmann et al., "Ground-level hyperspectral imagery for detecting weeds in wheat fields," *Precis. Agric.*, vol. 14, no. 6, pp. 637-659, 2013.
- [10] A. Apan, A. Held, S. Phinn, and J. Markley, "Detecting sugarcane orange rust disease using eo-1 hyperion hyperspectral imagery," *Int. J. Remote Sens.*, vol. 25, no. 2, pp. 489-498, 2004.
- [11] A.-K. Mahlein et al., "Development of spectral indices for detecting and identifying plant diseases," *Remote Sens. Environ.*, vol. 128, pp. 21-30, 2013.
- [12] H. Liu, and H. Motoda, *Feature Selection for Knowledge Discovery and Data Mining*. Kluwer Academic Publishers, 1998.
- [13] I. Guyon, and A. Elisseeff, "An introduction to variable and feature selection," *J. of Mach. Learn. Res.*, vol. 3, pp. 1157-1182, 2003.
- [14] C. Römer et al., "Quantifying moisture and roughness with Support Vector Machines improves spectroscopic soil organic carbon prediction," *J. Plant Nutr. Soil Sci.*, vol. 177, no. 6, pp. 845-847, 2014.
- [15] S. Kulkarni and G. Harman, *An Elementary Introduction to Statistical Learning Theory*. Hoboken, New Jersey: John Wiley & Sons Ltd, 2011.
- [16] G. Camps-Valls and L. Bruzzone, "Kernel-based methods for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 43, no. 6, pp. 1351-1362, June 2005.
- [17] L.C. Molina, L. Belanche, and A. Nebot, "Feature selection algorithms: a survey and experimental evaluation," in *Proceedings 2002 IEEE International Conference on Data Mining (ICDM-03)*, 2002, pp. 306-313.
- [18] H. Liu, and L. Yu, "Toward integrating feature selection algorithms for classification and clustering," *IEEE Trans. Knowl. Data Eng.*, vol. 17, pp. 491-502, 2005.
- [19] G. C. Cawley, N. L. C. Talbot, and M. Girolami, "Sparse multinomial logistic regression via Bayesian l1 regularisation," *Advances in Neural Information Processing Systems*, B. Schölkopf, J. Platt and T. Hofmann, eds., pp. 209-216: MIT Press, 2007.
- [20] K. Kira, and L. A. Rendell, "The feature selection problem: traditional methods and a new algorithm," in *Proceedings of the 10th national conference on Artificial intelligence*, San Jose, California, 1992, pp. 129-134.
- [21] M. Robnik-Šikonja, and I. Kononenko, "Theoretical and empirical analysis of reliefF and reliefF," *Mach. Learn.*, vol. 53, no. 1, pp. 23-69.
- [22] M. A. Hall, and L. A. Smith, "Feature selection for machine learning: comparing a correlation-based filter approach to the wrapper," in *Proceedings of the 12th International Florida Artificial Intelligence Research Society Conference*, 1999, pp. 235-239.
- [23] A.-K. Mahlein, et al., "Hyperspectral imaging for small-scale analysis of symptoms caused by different sugar beet diseases," *Plant Methods*, vol. 8, no. 1, p. 3, 2012.
- [24] D. A. Sims and J. A. Gamon, "Relationships between leaf pigment content and spectral reflectance across a wide range of species, leaf structures and developmental stages," *Remote Sens. Environ.*, vol. 81, no. 2-3, pp. 337-354, 2002.
- [25] G. Rondeaux, M. Steven, and F. Baret, "Optimization of soil-adjusted vegetation indices," *Remote Sens. Environ.*, vol. 55, no. 2, pp. 95-107, 1996.
- [26] A. Gitelson and M. N. Merzlyak, "Spectral reflectance changes associated with autumn senescence of *Aesculus hippocastanum* L. and *Acer platanoides* L. leaves. spectral features and relation to chlorophyll estimation," *J. Plant Physiol.*, vol. 143, no. 3, pp. 286-292, 1994.
- [27] T. P. Dawson and P. J. Curran, "Technical note A new technique for interpolating the reflectance red edge position," *Int. J. Remote Sens.*, vol. 19, no. 11, pp. 2133-2139, Jan. 1998.
- [28] J. Lv et al., "A multi-objective heuristic algorithm for gene expression microarray data classification," *Expert Syst. with Appl.*, vol. 59, pp. 13-19, 2016.
- [29] R. Franco-Duarte et al., "New integrative computational approaches unveil the *Saccharomyces cerevisiae* pheno-metabolomic fermentative profile and allow strain selection for winemaking," *Food Chem.*, vol. 211, pp. 509-520, 2016.
- [30] T. M. Cover, and J. A. Thomas, *Elements of Information Theory* (Wiley Series in Telecommunications and Signal Processing): Wiley-Interscience, 2006.