Spectral Analysis and Mapping of Blackgrass Weed by Leveraging Machine Learning and UAV Multispectral Imagery

Jinya Su[a,1,*], Dewei Yi[b,1], Matthew Coombes[c], Cunjia Liu[c], Xiaojun Zhai[a], Klaus McDonald-Maier[a], Wen-Hua Chen[c]

[a]*School of Computer Science and Electronic Engineering, University of Essex, Colchester, CO4 3SQ, U.K.*
[b]*Department of Computing Science, University of Aberdeen, Aberdeen, AB24 3FX, U.K.*
[c]*Department of Aeronautical and Automotive Engineering, Loughborough University, Loughborough LE11 3TU, U.K.*

Abstract

Accurate weed mapping is a prerequisite for site-specific weed management to enable sustainable agriculture. This work aims to analyse (spectrally) and mapping blackgrass weed in wheat fields by integrating Unmanned Aerial Vehicle (UAV), multispectral imagery and machine learning techniques. 18 widely-used Spectral Indices (SIs) are generated from 5 raw spectral bands. Then various feature selection algorithms are adopted to improve model simplicity and empirical interpretability. Random Forest classifier with Bayesian hyperparameter optimization is preferred as the classification algorithm. Image spatial information is also incorporated into the classification map by Guided Filter. The developed framework is illustrated with an experimentation case in a naturally blackgrass infected wheat field in Nottinghamshire, United Kingdom, where multispectral images were captured by RedEdge on-board DJI S-1000 at an altitude of 20m with a ground spatial resolution of 1.16 cm/pixel. Experimental results show that: (i) a good result (an average precision, recall and accuracy of 93.8%, 93.8%, 93.0%) is achieved by the developed system; (ii) the most discriminating SI is triangular greenness index (TGI) composed of Green-NIR, while wrapper feature selection can not only reduce feature number but also achieve a better result than using all 23 features; (iii) spatial information from Guided filter also helps improve the classification performance and reduce noises.

Keywords: Blackgrass weed; Guided filter; Random Forest; Spectral Index (SI); Unmanned Aerial Vehicle (UAV).

1. Introduction

An increasing world population (9 billion by 2050) is placing an unprecedented demand (a 70% increase) on contemporary agriculture. This global grand challenge is even severer in consideration of the scarcity of the arable land and natural farming resources, and the societal demand for shrinking agriculture's environmental footprint [1]. Weeds, aggressively competing with crops for water, nutrients and sunlight, are responsible for an approximate 35% reduction in potential global crop yields [2]. Improved weed monitoring can help reduce agricultural use of chemicals

*Corresponding author: School of Computer Science and Electronic Engineering, University of Essex, Colchester, CO4 3SQ, U.K. (J.Su)
Email addresses: `j.su@essex.ac.uk` (Jinya Su), `dewei.yi@abdn.ac.uk` (Dewei Yi), `M.J.Coombes@lboro.ac.uk` (Matthew Coombes), `C.Liu5@lboro.ac.uk` (Cunjia Liu), `xzhai@essex.ac.uk` (Xiaojun Zhai), `kdm@essex.ac.uk` (Klaus McDonald-Maier), `W.Chen@lboro.ac.uk` (Wen-Hua Chen)

[1]Jinya Su and Dewei Yi are joint first authors contributing to this work equally.

(herbicides in particular) and thus contribute to an increased agriculture sustainability.

Various weed management strategies are developed in the literature and practice according to various weed growth stages. Pre-sowing or pre-emergence herbicides can be applied before the emergence of weeds. Post-emergence herbicides, tillage, hand weeding and weed surfacing are common practices after weeds' emergence. However, conventional weed management strategies are to broadcast over the entire field irrespective of weeds' spatial information. This uniform strategy results in economical (a high cost due to overdose), environmental (severer ground water contamination) and social (herbicide residues in agriculture products) risks, and also exacerbates the problem of herbicide resistance, since weeds are usually patchily distributed within fields leading to many weed-free areas [3].

To tackle these challenges, there is a trend to adopt Site-Specific Weed Management (SSWM) strategy according to weed spatial distribution [4]. In this approach, weed mapping at early stages is desirable for timely herbicide applications. However, many challenges still exist in early (seedling) weed mapping due to various reasons including the similarity of spectral reflectance [2]. Therefore, late-season weed mapping, although not common yet, may provide an alternative in practice. This is because for certain weeds (e.g. blackgrass) weed infestation is relatively stable from year to year [5], as a result, late-season weed mapping can be used for the treatment of the subsequent year by using pre-sowing or pre-emergence herbicides. Weed mapping also plays a vital role in assessing the effectiveness of various herbicide treatments.

Blackgrass, an annual grass native to Eurasia, is a major weed of cereal crops in the U.K., especially for autumn sown crops including winter wheat. This weed has significantly adverse effects on crop yields and is prevalent in northern Europe; it is reported that about 70% of fields in the UK are infected by blackgrass [6]. To make it worse, blackgrass has gradually developed herbicide resistance, especially to post-emergence herbicides. Consequently, pre-emergence herbicides have become the main means of chemical control [7]. In addition, according to their effectiveness, practical non-chemical control strategies of blackgrass weeds in wheat fields include spring cropping, pre-sowing plowing, delayed autumn drilling, higher seed rates and competitive cultivars [7].

It is evident that accurate, timely and high-resolution weed maps (including blackgrass weed mapping) are key for SSWM practices [3]. Weed mapping by ground sampling is expensive, time-consuming and not suitable for large-scale applications [8]. Remote sensing of weed canopies is drawing increasing research interests, which is mainly enabled by the rapid development of spectroscopic instruments, the advent of easily-accessed and flexible moving platforms such as Unmanned Aerial Vehicle (UAV) and recent advancements in machine/deep learning algorithms. The underlying rationale is that various materials (land covers) in fields (crop, weed and bare soil) usually exhibit different spectral reflectance values due to their varied physical and chemical characteristics, which can be captured by spectroscopic instruments and subsequently be learnt by machine learning algorithms [9], [10].

There are several studies on weed mapping in crop fields, where the sensors range from low-cost multispectral cameras [11, 12] to high-cost spectrometers, from low to high spatial resolution, and from ground-based (harvesters, tractors) [13, 14] to airborne-based [15, 16]. UAV remote sensing with a user-defined spatial-temporal resolution, a low cost and high flexibility has become an important remote sensing approach [4]. For example, in [17], the problem of broad-leaf and grass weed detection in wide-row herbaceous crops are considered by using UAV visible imagery and neural network model. Maximum likelihood and Support Vector Machine (SVM) are compared in [18] for weed spatial distribution in onion field. In [5], piloted airborne multispectral imagery (MSI) with a resolution of 25 cm/pixel is adopted for cruciferous weed mapping among cereal crops by VISible-NIR (VIS-NIR) derived Spectral Indices (SIs). To alleviate the problem of crop/weed spectral similarity at early growth stages, crop row spatial information is exploited in [2]. Spatial (crop row) and spectral characteristics are also exploited in [19] for weed mapping in maize fields by six-band MSI. Visual features and geometric information of detected vegetation are also employed to classify crops and weeds in [20] by using RGB-NIR aerial image at an altitude of 3 m.

Recently, there is also a trend for weed mapping by deep learning approaches [13, 14, 15, 21]. For instance, the problem of crop/weed semantic segmentation is considered by using images collected by agricultural robots in [13, 14]. Deep neural network (e.g. SegNet) was adopted in pioneering work [15, 21] for semantic weed classification from sugar beet by aerial MSI collected by a small UAV at an height of 2m. It is noted, however, that a large amount of labelled data is usually required in deep learning approaches for images (ideally) with a relatively high spatial resolution [22]. Deep learning approaches are not applied in the study considering that: (1) only a limited number of pixels are labelled for the MSI due to the challenges in labelling aerial images; (2) the labelled pixels are sparely distributed in the field of interest (see, Fig 5); (3) the image resolution is far lower than the ones collected by agricultural robots or UAV at a very low height [13, 14, 15], since the drone camera in this study can not be operated at a very low altitude (due to the image calibration and stitching issue at an altitude lower than 10 m); (4) as described in Section 2.1, the labelling approach is not suitable for deep learning approaches, either. The following observations are drawn for the research in terms of research motivation, gaps and challenges:

(1) There is an urgent need for an automatic remote sensing based weed mapping to enable SSWM at field scales;

(2) In-season weed mapping is significant and possesses new research challenges over early-season mapping due to

the lack of crop row information in wheat fields;

(3) There are studies for weed mapping by using spectral and/or spatial information, little work is available on systematically selecting appropriate SIs or their optimal combination for a simple but effective classifier;

(4) To date, the use of UAV, MSI and machine learning for blackgrass weed mapping has not yet been evaluated.

The main contribution of this study lies in the initial development of an automatic weed mapping framework by integrating five-band MSI, low-altitude UAV platform and machine learning algorithms, and its real-life validation in a naturally blackgrass infected wheat field. The specific objectives are to:

(i) Determine whether or not blackgrass weeds can be discriminated from wheat by applying spectral analysis and classification algorithm to aerial MSI;

(ii) Identify the most discriminating spectral bands, SIs and their optimal combination by using feature generation and selection algorithms;

(iii) Exploit both spectral and spatial information for an accurate blackgrass probabilistic map;

(iv) Initially experimentally validate the system in a naturally blackgrass infested wheat field.

2. Materials

2.1. Study field

Experiments were carried out in a naturally blackgrass (Alopecurus myosuroides) infected wheat (Triticum aestivum L.) field (GPS coordinate, latitude: $53^o02'45''N$, longitude: $0^o45'29''W$, altitude: 14m a.s.l.) of Newark, Nottinghamshire, U.K. (please refer to Fig 1 for the location of the test field). Data collection was done on 05/Jun/2018 (mid-day of a sunny day), when wheat and blackgrass weeds are in the stages of full ear emergence and early seed shedding. The late-season imaging is mainly due to the significant challenges in blackgrass groundtruth labelling at an early stage in UAV imagery, especially for a naturally blackgrass infected wheat field in this study. As shown in the left bottom plot of Fig 1, red rope (treated as background pixels in labelling) has been used to help agronomists label the aerial image after acquisition and pre-processing. This will also prohibit the application of deep learning approaches, since this information will also be learnt automatically in deep learning framework. It is noted that red rope is to help agronomists build the knowledge of blackgrass weed in the aerial images. Upon knowing its characteristics, a

GPS location: 53° 02'45.0"N, 0° 45'29.0"W

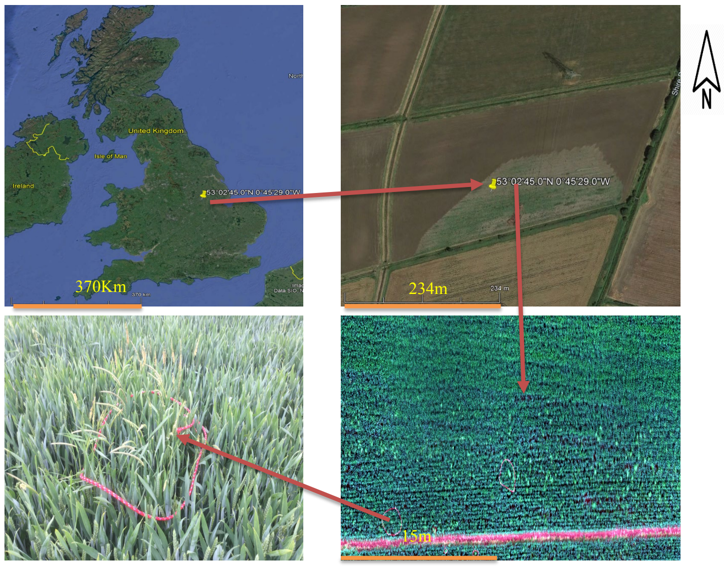

Figure 1: Test wheat field for blackgrass weed mapping including geographic location (Google earth engine), satellite image of the field of interest, false-color UAV image and ground images of blackgrass weed.

large number of blackgrass samples (within and beyond the red rope regions) are manually labelled for the purpose of spectral analysis and model construction. It is also noted that the purple line at the bottom right of Fig 1 is a field path (background pixels in labelling).

2.2. MSI image acquisition

In this study, multispectral camera is adopted for blackgrass weed mapping. This is because compared to hyperspectral camera [23], multispectral camera is lightweight, with a low cost, and of high spatial resolution and so applicable to large areas of interest. While compared to RGB camera [24], multispectral camera possesses additional spectral bands and is also less affected by environmental variations due to the availability of reflectance calibration panel. To obtain aerial MSI, S1000 Octocopter (DJI, Shenzhen, China) and RedEdge multispectral camera (MicaSense, Seattle, WA, USA) are adopted, where the developed system is displayed in Fig 2. The specification of DJI S1000 is referred to [9, 25]. RedEdge camera is a light-weight (135g), small ($5.9cm \times 4.1cm \times 3.0cm$), high-resolution ($1280 \times 960$ pixels) camera, which can capture five narrow spectral bands with GPS information. The five bands (wavelength/bandwidth) include Blue (475/20 nm), Green (560/20 nm), Red (668/10 nm), RedEdge (717/10 nm) and NIR (840/40 nm).

During the flight, a gimbal is adopted to fix the camera pointing vertically downwards so that the the adverse effects of UAV motion/vibration on image quality are attenuated. The flight altitude is set to 20m above ground with

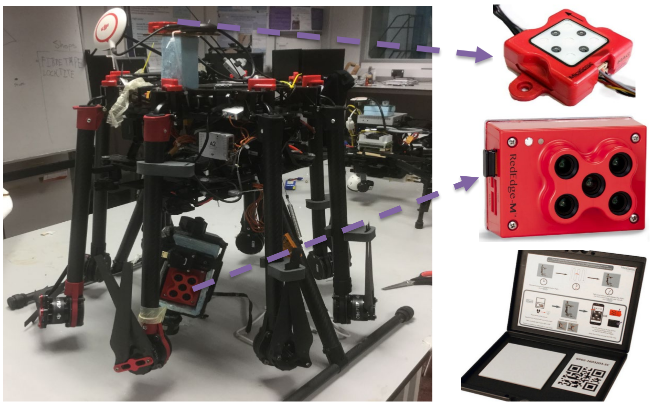

Figure 2: UAV-Camera system: DJI S1000 (left), Downwelling light sensor (top), RedEdge camera (middle) and calibration panel (bottom).

a ground spatial resolution of 1.16cm/pixel in the orthomosaic image. DJI Ground Station 4.0 is used to plan, control and monitor the UAV flight. The flight path, UAV forward velocity (1m/sec.) and camera triggering are set in the Pix4DCapture App to make sure that an overlap and sidelap up to 75% is guaranteed for the purpose of accurate image stitching. A total number of 460 x 5 images are obtained in the field trial, where the total covered area after image stitching is about 0.004 square km. A selected portion of the stiched image with sever blackgrass weed infection is selected for the spectral analysis and model test.

For image reflectance calibration, both Downwelling Light Sensor (DLS) and MicaSense's Calibrated Reflectance Panel (CRP) are adopted. DLS is to record data on the amount of light from the sky, which is useful for situations where ambient light conditions are changing during a flight. While CRP with known reflectance values (i.e. 0.57, 0.57, 0.56, 0.51 and 0.55 for Blue, Green, Red, NIR and RedEdge) is for absolute reflectance calculation. Before and after each flight, an image of the CRP is taken at about 1m without shadow, which will be used for reflectance calibration in Pix4DMapper software. Images captured with RedEdge-M conforms to standard formats (TIFF) with standard metadata such as GPS information (latitude/longitude/altitude and date/time), attitude data (pitch, roll and heading angles) and camera information (e.g. exposure time, ISO speed, black level).

2.3. Image preprocessing by Pix4DMapper

Professional photogrammetry software is then required to align the bands, calibrate the images and create georeferenced reflectance maps, based on which various SVIs can be calculated. These tasks are conducted by Pix4DMapper software of version 4.3.31 (Pix4D SA, Switzerland), where more detailed information is referred to [25]. To make it more intuitive, the overall workflow is displayed in Fig 3, which include UAV-Camera system, flight path planning in Pix4DCapture, image pre-processing in Pix4DMapper. The outputs of image preprocessing process are spectral bands and SVI GeoTIFF images of the whole site.

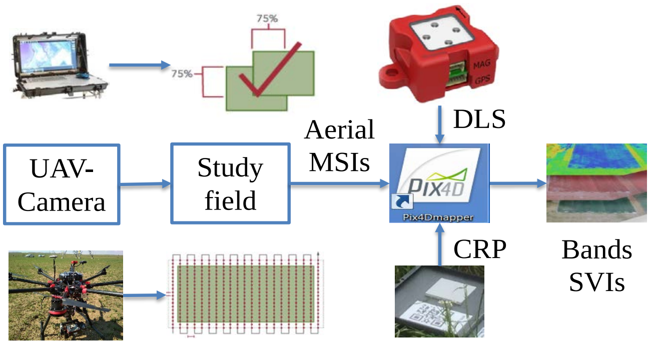

Figure 3: Workflow of MSI image acquisition and preprocessing.

3. Methods

This section introduces the methods for blackgrass weed mapping. The task of automatic blackgrass weed mapping is formulated as a supervised classification problem. The developed framework seamlessly integrates a number of advanced techniques: feature generation to enhance feature discrimination ability, feature selection for dimension reduction, Random Forest (RF) for classification, and Guided Filter for spatial information enhancement. The overall framework is displayed in Fig 4.

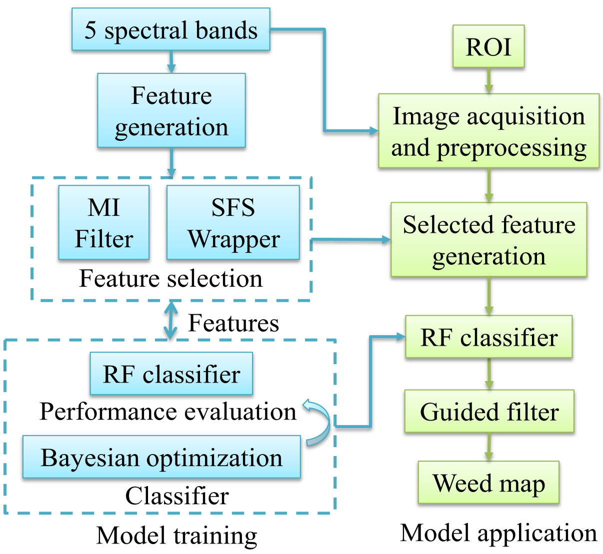

Figure 4: The developed system for blackgrass weed mapping in wheat fields: model training (left) and model application to RoI (right).

The system in Fig 4 consists of two stages: model training and model application. In training step, 5 raw spectral bands are first manipulated to generate 18 SI features. Then feature selection is used to reduce feature dimensionality and improve empirical interpretability, where both filter (Mutual Information (MI) ranking) and wrapper (Sequential

Forward Selection (SFS)) feature selection approaches are considered [26] to identify the better one. The selected features are to train a classification model, where RF classifier with Bayesian hyperparameter optimization is adopted. In applying the model to the Region of Interest (RoI), the trained RF classifier is applied to the selected spectral features; then the classification probability maps are further processed by guided filter so that spatial information can be incorporated into the final weed map.

3.1. Aerial imagery with groundtruth labelling

The aerial imagery is firstly introduced including its groundtruth labelling. The raw aerial images were first processed by the procedures in Section 2.3 for spectral band reflectance and SI map generation. The RGB composite with image adjustment for intensity enhancement is displayed in Fig 5. Random parts of the image are also manually (via zoom in and zoom out) labelled by "Image Labeller" in Matlab. In particular, 125164 out of 3154051 pixels are labelled into three classes: Blackgrass (53635), Wheat (47133) and Background (24396).

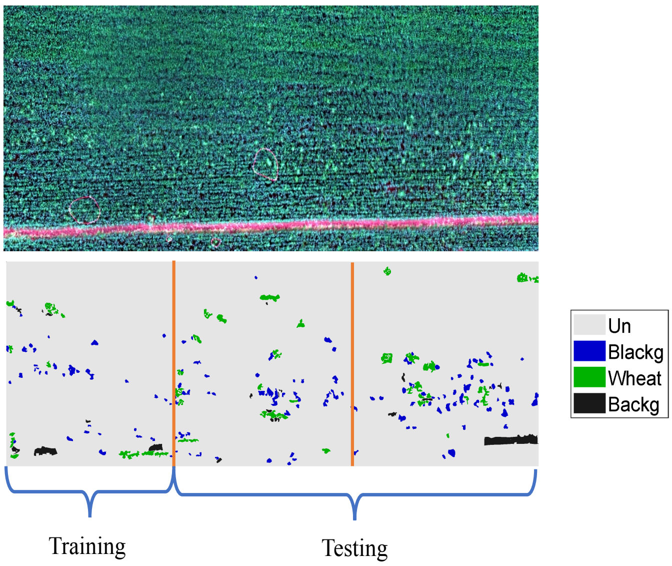

Figure 5: Upper: RGB composite of the RoI with image intensity enhancement; Lower: labelled classes by visual inspection in Matlab "Image Labeller" including Un being unlabelled region. Pixel resolution: 1.16cm/pixel, image size: 3000x1050.

3.2. Feature generation

In order to maximally represent image characteristics, 18 widely-used SIs are generated as the potential features in addition to 5 raw spectral bands. SI refers to a mathematical expression combining the surface reflectance at two or three spectral bands in order to enhance spectral differences of various objects. SI is a common technique in precision agriculture and has been widely applied in a number of areas such as weed mapping [2], yellow rust monitoring [9]. Following preceding work [9], 18 widely-used SIs for RedEdge multispectral camera are generated, which are summarized in Table 1.

3.3. Feature selection

Feature Selection (FS) is to select a subset of features for model construction, which is able to simplify the model, reduce the training time, avoid the curse of dimensionality and enhance generalization by reducing the chance of overfitting. Suppose the complete feature set is F_s with d being its feature number, selecting the best subset $S_s \subseteq F_s$ results in possible $2^d - 1$ combinations, indicating exhaustive search is impossible due to the NP-hard problem. Practical methods usually follow certain heuristics. According to various metrics, the existing FS algorithms can be broadly categorised into three classes including filters, wrappers and embedded methods. Filters rely on proxy measures (MI, Pearson correlation) to rank features, which is independent of the classifier and also computationally efficient. Wrappers rely on predictive models to evaluate feature subsets and usually provide the best performance but with a higher computation load. While Embedded methods perform feature selection as part of the model construction process and are usually constrained to certain classifiers.

In this work, to make the classification model simple but effective, both filter and wrapper approaches are considered. In particular, MI between features and class label is adopted as the evaluation metric of filter approach, where the number of selected features is determined by the performance (Out-Of-Bag (OOB) error) of RF classifier. Moreover, Sequential Forward Selection (SFS) is also adopted as the wrapper strategy to identify the best feature combination, where RF is adopted to determine whether a feature should be included or not. The framework for MI filter and SFS wrapper is displayed in Fig 6. Their principles are briefly introduced in the following subsections.

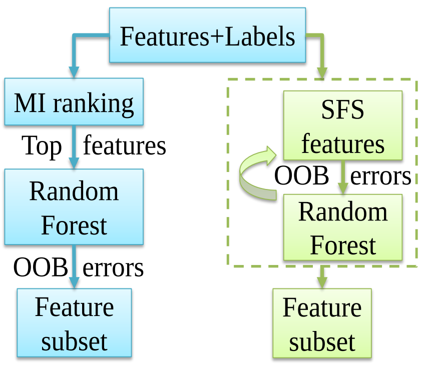

Figure 6: Feature selection: MI filter (left) and SFS wrapper (right), where OOB errors is adopted to evaluate the RF performance.

Table 1: SVIs adopted in this study.

Category (No.)	Full name	ABBRE.	Band	Formula
VIS(4)	Nitrogen Reflectance Index	NRI	Green-Red	$(R_g - R_r)/(R_g + R_r)$
	Greenness Index	GI	Green-Red	R_g/R_r
	Green Leaf Index	GLI	Blue-Green-Red	$(2R_g - R_r - R_b)/(2R_g + R_r + R_b)$
	Triangular Greenness Index	TGI	Blue-Green-Red	$-0.5(\lambda_r - \lambda_b)(R_r - R_g) - (\lambda_r - \lambda_g)(R_r - R_b)$
Green-RE(1)	Anthocyanin Reflectance Index	ARI	Green-RE	$R_g^{-1} - R_{re}^{-1}$
Green-NIR(3)	Green NDVI	GNDVI	Green-NIR	$(R_{nir} - R_g)/(R_{nir} + R_g)$
	Triangular Vegetation Index	TVI	Green-NIR	$0.5[120(R_{nir} - R_g) - 200(R_{nir} - R_g)]$
	ChlorophyII Index-Green	CIG	Green-NIR	$R_{nir}/R_g - 1$
Red-NIR(4)	Normalized Difference Vegetation Index	NDVI	Red-NIR	$(R_{nir} - R_r)/(R_{nir} + R_r)$
	Soil Adjusted Vegetation Index	SAVI	Red-NIR	$1.5(R_{nir} - R_r)/(R_{nir} + R_r + 0.5)$
	Ratio Vegetation Index	RVI	Red-NIR	R_{nir}/R_r
	Optimized Soil Adjusted Vegetation Index	OSAVI	Red-NIR	$1.16(R_{nir} - R_r)/(R_{nir} + R_r + 0.16)$
RE-NIR(2)	ChlorophyII Index-RedEdge	CIRE	RE-NIR	$R_{nir}/R_{re} - 1$
	Normalized Difference RedEdge Index	NDREI	RE-NIR	$(R_{nir} - R_{re})/(R_{nir} + R_{re})$
VIS-RE-NIR(4)	Enhanced Vegetation Index	EVI	Blue-Red-NIR	$2.5(R_{nir} - R_r)/(R_{nir} + 6R_r - 7.5R_b + 1)$
	Transformed ChlorophyII Absorption and Reflectance Index	TCARI	Green-Red-RE	$3[(R_{re} - R_r) - 0.2 * (R_{re} - R_g)(R_{re}/R_r)]$
	ChlorophyII Vegetation Index	CVI	Green-Red-NIR	$R_{nir}R_r/R_g^2$
	Simplified Canopy ChlorophyII Content Index	SCCCI	Red-RE-NIR	NDREI/NDVI

3.3.1. MI filter

There are various feature scoring algorithms. According to various criteria to evaluate the relationship between features and class label, these algorithms include MI, Fisher score, Minimum Redundancy and Maximum Relevance (MRMR) and ReliefF based ones. In comparison with other approaches, MI is not only simple but also can reflect the statistical dependency between two random variables, and therefore is adopted. MI is usually defined for two discrete random variables (feature quantization is adopted to discretize continuous variables into discrete bins) and a higher value means a higher statistical relevance. MI value for random variables Y and Z is defined by

$$MI(Y,Z) = \sum\nolimits_{y\in Y}\sum\nolimits_{z\in Z} P(y,z)log\left(\frac{P(y,z)}{P(y)P(z)}\right), \tag{1}$$

where $P(y), P(z)$ and $P(y,z)$ represent marginal probability distribution of Y and Z, and the joint probability distribution of Y and Z. Then the top features identified by MI ranking can be selected as the features for RF classifier construction so that the optimal feature number can be determined.

3.3.2. Sequential Forward Selection (SFS)

SFS is a typical search strategy for wrappers feature selection. In this approach, features are added sequentially into the feature set, where the evaluation metric for adding a feature is defined as the OOB error of RF. To avoid distracting readers from the main contributions, SFS with RF is summarized in Algorithm 1 (see, Appendices).

3.4. Random forest classifier

The task of classification can be achieved by many algorithms such as classification tree, K-Nearest Neighbours, discriminant analysis, Support Vector Machine [27]. RF is preferred for the task with a small number of labelled data; because RF achieves good performance in term of accuracy and robustness while with a relatively low computation cost. In addition, it can not only return the class label but also the probability in the range of [0,1]. RF is an ensemble learning method, where a number of decision trees are trained (by bootstrap sampling) with the final class output being the mode of individual decision trees. RF, in comparison to individual decision tree, can effectively avoid overfitting and improve robustness. In order to improve its performance, its hyperparameters are automatically tuned by Bayesian optimization. RF classifier with Bayesian optimization is summarized in Algorithm 2 (see, Appendices).

By applying the trained RF to the RoI, one can obtain the initial pixel-wise classification maps

$$P = \{P_1, \cdots, P_C\} \text{ with } P_k = \{p_k^1, \cdots, p_k^n\} \ , \tag{2}$$

where $P_k, k = 1, \cdots, C$ denotes the probability map for class k with C being the class number; $p_k^i \in [0, 1]$ denotes the probability that pixel i belongs to class k.

3.5. Guided filer for spatial information

In real-life applications, pixels in a local region are more likely to share the same class label. This type of spatial information should also be considered. Inspired by the hyperspectral image classification in [28], spatial filtering is adopted to regularize the initial spectral classification maps. Guided filter in [29] is preferred due to its fine properties in enhancing the smoothness of local areas and preserving the edge information of the reference image.

Guided filter assumes a local linear model between guidance image I (input image itself or an reference image) and filter output Q in a local window w_k centred at pixel k:

$$Q_i = a_k I_i + b_k, \forall i \in w_k, \tag{3}$$

where i denotes pixel index, (a_k, b_k) are some unknown linear coefficients being constant in w_k. To determine the coefficients, a cost function is defined, which minimizes the differences between output image Q and input image P,

$$E(a_k, b_k) = \sum_{i \in w_k} [(a_k I_i + b_k - P_i)^2 + \epsilon a_k^2], \tag{4}$$

where ϵ is a regularization parameter preventing a_k being too large. The model (3) and the approximation solution of (4) ensure that $\bigtriangledown Q \approx \bar{a} \bigtriangledown I$ with $\bar{a}$ being the approximation solution of a. Therefore, the edge information in the reference image I can be preserved in filtering output image Q.

In this work, various three-band images including RGB image, the first three principal components of PCA analysis, and the first three SIs by SFS feature selection algorithm are adopted as the reference image for the Guided filter. Then (3) with the solution of (4) is adopted to regularize the initial probabilistic maps $P_i (i = 1, \cdots, C)$ in (2), given by

$$\hat{P}_i = \text{Guided filter}(I, P_i). \tag{5}$$

After all C initial probabilistic maps are processed by Guided filter, the final class label of pixel i is determined by the maximum of filtered maps $\hat{P}_i (i = 1, \cdots, C)$. The overall procedure is summarized in Algorithm 3 (see, Appendices).

3.6. Performance evaluation

Various metrics are adopted for performance evaluation. OOB error is adopted to measure the prediction error of RF classifier to avoid overfitting in feature selection algorithms and hyperparameter optimization. OOB error denotes the mean prediction error on each training sample x_i, using only the trees that did not have x_i in their bootstrap aggregating (or bagging) samples. It is reported that OOB error helps avoid the need for an independent validation dataset. In addition, other popular metrics are also adopted wherever necessary such as accuracy, precision and recall. These metrics rely on a number of definitions including True Positive (TP), False Positive (FP), True Negative (TN) and False Negative (FN) [9]. Accuracy, a good measure for symmetric datasets, can then be defined by

$$Accuracy = \sum_i TP/All, with\ All = TP + FP + TN + FN. \tag{6}$$

Precision and Recall, effective measures for data with uneven distributions, can also be defined for a specific class,

$$\begin{array}{c} Precision = TP/(TP + FP) \\ Recall = TP/(TP + FN) \end{array}. \tag{7}$$

Their mean values for various land cover classes can then be calculated.

4. Results

4.1. Spectral analysis

Spectral analysis is first conducted, where the mean reflectance values (along with 0.5-σ area) of five original bands and 18 SVIs for various land cover classes including Blackgrass (Blackg), Wheat and Background (Backg) are displayed in Figs 7 and 8. Moreover, the correlation analysis between different spectral bands and SVIs is also performed. The correlation map is displayed in Fig 9, where a brighter pixel means a higher correlation value.

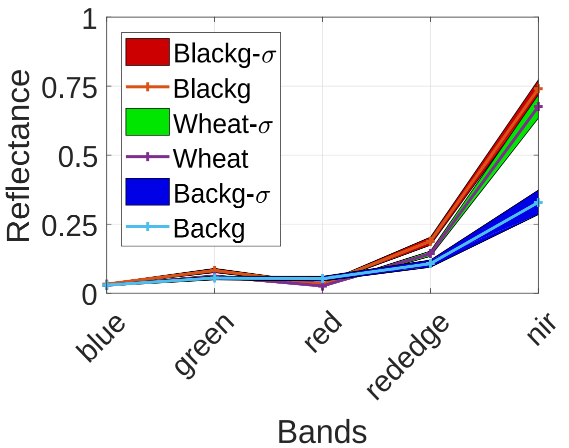

Figure 7: Mean reflectance and 0.5-σ range of five spectral bands for three land cover classes in the wheat field.

The following observations can be drawn from Figs 7, 8 and 9. As seen in Fig. 7, various materials have distinct spectral reflectance values (curves), which can be learnt by machine learning algorithms for classification. As seen

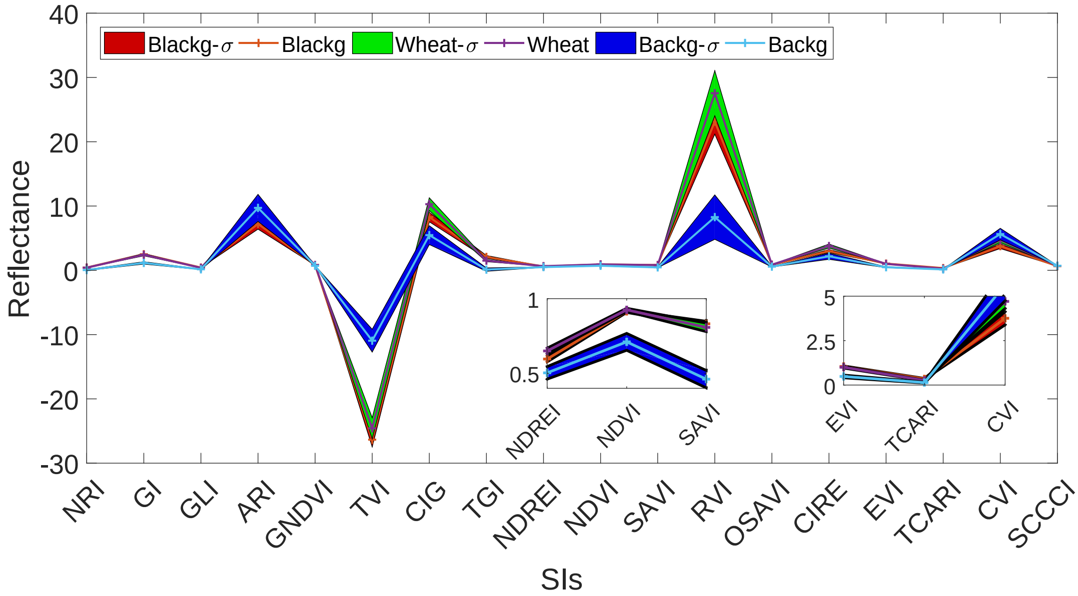

Figure 8: Mean reflectance and 0.5-σ range of 15 SVIs for three land cover classes in the wheat field.

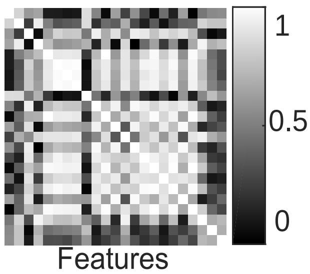

Figure 9: Correlation map of spectral bands & SVIs, where greyscale represents the level of correlation with black (0) and white (1).

in Figs. 7 and 8, reflectance differences are distinct in various spectral bands/SIs for various materials, so various spectral bands/SIs have various discriminating abilities. As seen in Fig 9, many features have high correlation values (brighter), implying a high feature redundancy.

These observations on the one hand show the rationale of using classification for blackgrass weed mapping and on the other hand imply that feature selection is critical for a simple but effective classification model. The latter problem is even severer for the scenario with a limited number of labelled data in this study.

4.2. Feature selection

Considering that feature selection is generally time-consuming due to the large number of feature combinations (and consequently different machine learning models to be trained), without loss of generality, only 15% of the randomly labelled data is used in feature selection. In MI filter, the continuous variables are discretized into 10 discrete bins. "TreeBagger" in Matlab with default parameters and 50 trees is first adopted to obtain the OOB error. Their results are displayed in Fig 10, where vertical axis denotes the average OOB error and horizontal axis represents the features that are firstly selected by the algorithms. To demonstrate the discriminating ability of top (TGI) and bottom (blue) features selected by MI filter, their probability histograms are displayed in Fig 11.

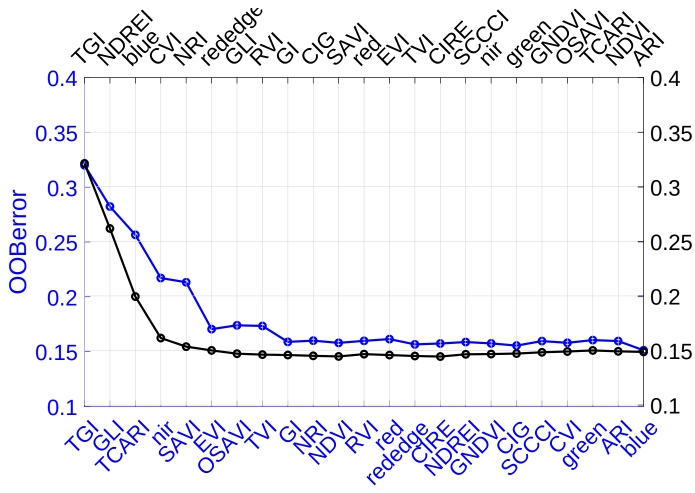

Figure 10: Results of MI (blue) and SFS (dark) for feature selection.

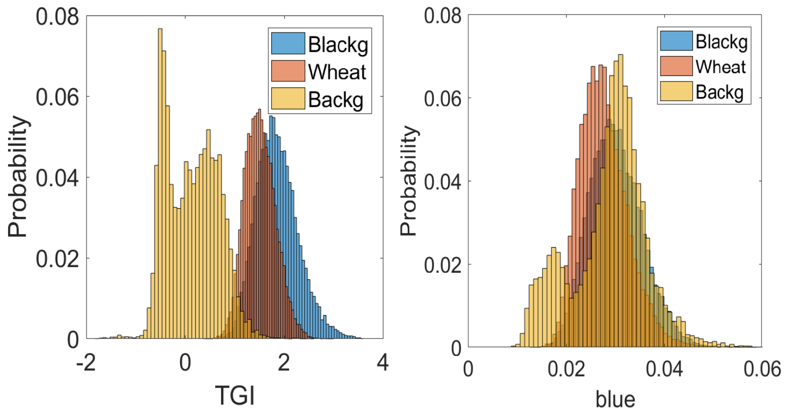

Figure 11: Histogram of best feature (TGI) and worst feature (blue) features by MI.

The following observations can be drawn from Figs 10, 11, and experimental comparisons. First, the best feature by MI filter and SFS wrapper is the same, that is TGI. Second, the best performance of MI filter occurs when all 23 features are selected, where the OOB error is 15.09%; while the best performance of SFS wrapper occurs when selected 12 features are adopted, where the OOB error is 14.43%. The OOB error of SFS wrapper is 14.68% when selected 8 features are adopted, which is very close to the best one and better than MI filter. So SFS wrapper generates a simple, yet effective model. Third, although Blue band has a very low MI value with class label (Fig 10) and a low discriminating ability alone (Fig 11), its combination with other features may result in good performance (Fig 10).

This implies that the MI filter is useful in assessing individual features, while SFS wrapper excels in identifying the best feature combination. Finally, the training time of SFS wrapper is 12 times of MI filter, since (1+23)*23/2=276 RF classifiers are built in SFS wrapper compared to 23 classifiers in MI filter. Therefore, the first 8 features by SFS wrapper are adopted for model construction. In the following subsection, the parameters of RF classifier are optimized to enhance its performance by Bayesian optimization.

4.3. Algorithm verification

Bayesian optimization is adopted to tune the hyperparameters of RF classifier, where two key parameters are considered: minLS and numPTS. minLS is to specify trees' depth/complexity and numPTS controls the number of predictors to sample at each node in tree growing. $minLS \in [1, maxMinLS]$ is selected with maxMinLS being 18 and $numPTS$ in $[1, numF]$ with $numF$ being the feature number. Tree number is chosen 100, 'AcquisitionFunctionName' is chosen 'expected-improvement-plus'. Then two algorithm verification approaches are adopted as below.

4.3.1. Random split

In random split test, the random 15% of the labelled dataset is for algorithm training and the remaining random 85% is for algorithm testing, where the confusion matrix (calculated on the validation set) is shown in left plot of Fig 12. In left plot of Fig 12, the target and output class denote ground truth and predicted class. The diagonal cells

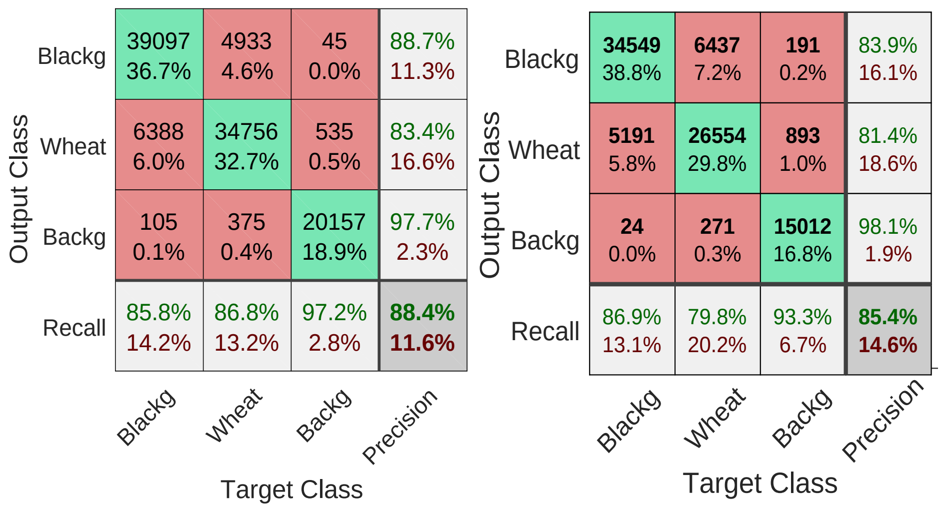

Figure 12: Confusion matrix for random split test (left) and spatial split test (right).

in green show the number/percentage of correct classification; the off-diagonal cells are where the misclassification occurs. For Blackg class, 39097 in green is TP and 4933 + 45 in red is FP, 6388 + 105 in red is FN. So precision for Blackg class is 39097/(39097+4933+45) = 88.7%, while recall for Blackg class is 39097/(39097+6388+105) = 85.8%.

Similarly, precision and recall for Wheat and Backg classes are 83.4%, 86.8%, and 97.7%, 97.2%. As a result, the mean precision and recall for RF classifier are 89.9% and 89.9%. The cell at bottom right displays the overall accuracy (88.4%). Precision and recall for background class are very high, since its reflectance values are significantly different from other classes and can be easily classified.

4.3.2. Spatial split

Different from random split test, in spatial split test, the ROI in Fig 5 is vertically split into three equal parts, where the labelled pixels in the leftmost part is for algorithm training and the remaining two parts are for algorithm testing. The confusion matrix is shown in the right plot of Fig 12. One can see that the performance of spatial split is slightly worse than the random split, this is mainly due to the lack of diversity in spatial split test. This implies that in real-world applications, more diverse datasets from different locations are desirable in order to improve algorithm generalization.

4.4. Application to RoI

The optimized RF classifier with selected features is then applied to the whole RoI in Section 3.1. The classification results by only spectral feature are shown in Fig 13 (a). The Guided filter in Section 3.5 is also applied to the initial probabilistic maps to incorporate spatial information in reference images. Three reference images are tested: RGB composite (with image enhancement), the top three principal components (the values in each PC are transformed into [0, 1]) by PCA (PCA-Top3) and the top three SIs (the values of each SI are also transformed into [0, 1]) identified by SFS wrapper (SFS-Top3). Their comparative results are shown in Fig. 13 (b), (c) and (d). Their performance is also summarized in Table 2.

Table 2: Performance comparisons for various reference images.

Approach	precision	recall	accuracy
Spectral	89.9%	89.9%	88.4%
RGB	93.4%	93.5%	92.5%
PCA-Top3	93.9%	94.0%	93.1%
SFS-Top3	93.8%	93.8%	93.0%

The following observations are drawn from Fig 13, and Table 2. First, it follows from Fig 13 that pixel-wise classification by only spectral feature may result in random noises and including spatial information by Guided filter

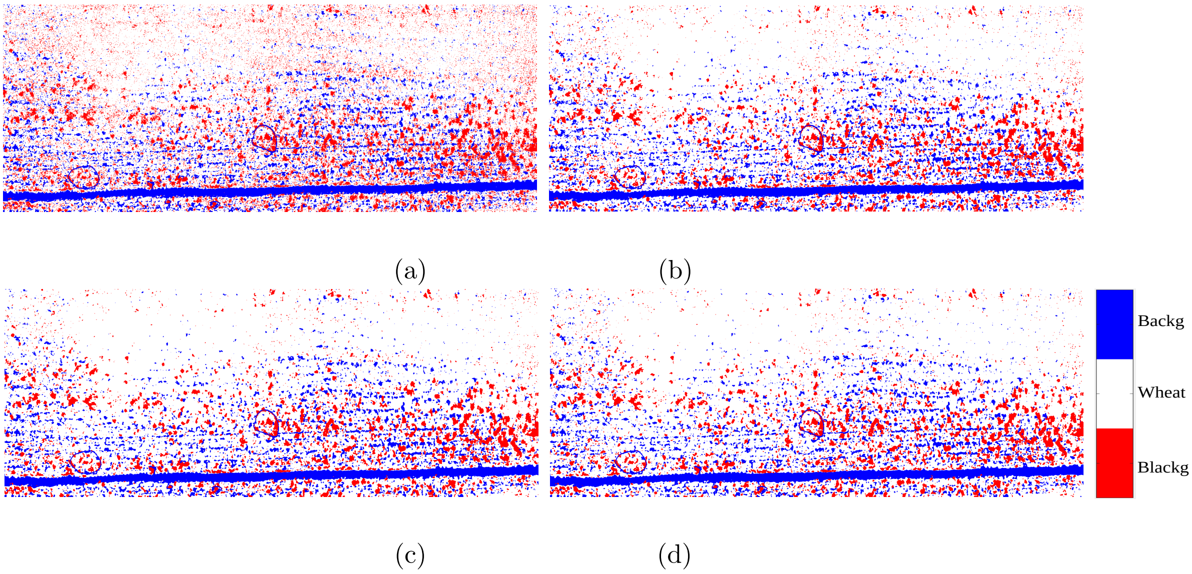

Figure 13: Classification map by using only spectral feature (a); Classification results by Guided filter regularization: RGB (b), SFS-Top3 (c) and PCA-Top3 (d).

(Fig 13) can improve the result by reducing the noises while preserving the edge information in reference image. Secondly, all Guided filter approaches (Table 2) outperform the purely spectral based ones; PCA and SFS have similar performance, slightly better than RGB based one. As a by-product, the probabilistic blackgrass weed map is displayed in Fig 14, which can be used for weed management in subsequent year.

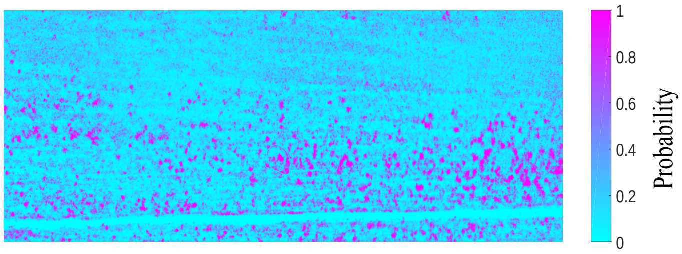

Figure 14: Blackgrass weed probabilistic map with Guided filter regularization by SFS-Top3 reference image.

5. Discussions

Site-specific weed management (SSWM) is paramount for sustainable agriculture (i.e. generating more and better outputs with less inputs while with decreased environmental footprint) in order to meet the world's future food security and sustainability needs. Instead of ground field sampling, remote/proximal sensing is drawing increasing research interests due to its potential for large scale applications with less human involvement.

UAV remote sensing for precision agriculture is still a developing technology and has initially been applied to a number of areas such as disease, drought and nitrogen monitoring [4, 9] due to its user-defined spatial, spectral and temporal resolutions and suitability for application at farmland scales. Even for weed mapping, there are already a number of studies [2, 5, 15, 16, 19, 20, 30]. It is noted, however, that early (seedling) weed detection [2, 19] is only possible for certain wheat/weed combinations by using very high-resolution images or hyperspectral images [23]. Unfortunately, wheat is monocotyledonous crop, which makes it extremely challenging for early (seedling) blackgrass mapping by using UAVs due to various reasons such as wheat/blackgrass spectral similarity, low spatial resolution [16], challenges in ground truth labelling. Therefore, this work (the first one in using UAV multispectral remote sensing for blackgrass weed mapping) is focused on late-season weed mapping, which is still very useful in a number of SSWM practices such as designing SSWM for subsequent years, applying in-season post-emergence herbicides, assessing the effectiveness of herbicide applications [16, 6]

Different from conventional studies [2, 5] for weed mapping by using MSI, where only a very limited number of spectral vegetation indices are adopted for model learning, this work first generates a relatively large number of spectral features to enhance feature discriminating abilities [9]. On this basis, advanced feature selection algorithms are further adopted to reduce weak features and identify the best (reduced-order) feature combination. With selected features, RF classifier is adopted for classification task in this work (i.e. learning from a limited number of labelled data) due to its fine properties in term of accuracy and robustness while with a relatively low computation cost. The hyperparameters of RF are further optimized by using Bayesian optimization to guarantee better performance.

Another novelty of this work is to adopt Guided filter [29] to regularize the probabilistic maps so that spatial information can be incorporated for better classification performance [28]. To identify a suitable reference image for Guided filter, different three-band images are extracted from 5 spectral bands and 18 spectral vegetation indices by using RGB composite, PCA analysis and SFS Wrapper feature selection. Finally, the blackgrass weed mapping system by using aerial MSI is applied to ROI with a promising result in term of Precision, Recall and Accuracy.

Regarding image segmentation algorithm, this study only considered the random forest algorithm with hyperparameter optimization. Interested readers are referred to the recent study [27], which compares and assesses different Geographic Object-Based Image Analysis (GEOBIA) and machine learning algorithms using UAV multispectral imagery via a case study in a citrus orchard and an onion crop. More importantly, different from this study, operational interest and aspects such as requested time and computing resources as well as the expertise needed to implement them are also available in [27] for a practical application.

Regarding the multispectral sensor in this study, for the time being, the UAV has to be operated at an altitude of about 20 meters so that centimetre-resolution image can be obtained for the purpose of a good image segmentation. However, flight height of 30-50 meters and more would be more desirable to cover a relatively large field of interest in a short time. Therefore, new multispectral camera of high geometrical and spectral resolution would be more desirable for practical implementation.

Though the experimental results are very positive, there is still much room for further development in terms of data, algorithm and experimental verification. Several aspects are as below:

(1) Only spectral and spatial features are considered in this study; temporal information should also be investigated to achieve an early and more reliable weed mapping;

(2) With the advent of a large volume of labelled (high-resolution) images for precision agriculture applications, deep features rather than hand-crafted features may be considered by using end-to-end deep neural network.

(3) More aerial datasets collected under various conditions (e.g. time, weather, UAV altitude, wheat varieties) will be used to enhance the robustness and improve the generalization of the developed framework.

6. Conclusions

This work exploits the potentials of five-band multispectral camera, small airborne platform and machine learning algorithms (e.g. feature generation and selection, Bayesian parameter optimization, and spatial information enhancement by Guided filter) for the automatic mapping of blackgrass weeds in winter wheat. A blackgrass weed mapping system is initially developed by learning from spectral-spatial features of labelled aerial MSI. The system is initially validated on a naturally blackgrass infected wheat field, where aerial MSIs were collected at an altitude of 20m with a ground spatial resolution of 1.16cm/pixel. Comparative experimental results show that the developed system can achieve a satisfying classification result with an average precision, recall and accuracy of 93.8%, 93.8% and 93.0% when wheat and blackgrass weeds are in the stages of full ear emergence and early seed shedding. The SI with the best discriminating ability is TGI. It is also discovered that wrapper feature selection can substantially reduce feature dimension from original 23 to 8 while achieving better performance than using all 23 features. Spatial information from Guided filter is proved to be effective in helping attenuate the noises of pixel-wise spectral classification and improve classification performance.

Appendices

Algorithm 1: SFS with Random Forest

(a) Start with an empty set $Y_0 = \emptyset, k = 0$ with full feature set $Y = \{y_1, \cdots, y_d\}$;

(b) Select the next best feature x^+ by

$$x^+ = \underset{x \in (Y - Y_k)}{\arg\min}\ OOBErr(Y_k + x),$$

where OOBErr(Z) denotes OOB error of the Random Forest classifier trained by using feature set Z;

(c) Update $Y_{k+1} = Y_k + x^+$ with $k = k + 1$;

(d) Repeat Steps (b) (c) until termination rules (desired feature number or OOBErr increment) are satisfied.

Algorithm 2: Random Forest with Bayesian optimization

(a) Set tree number and stopping rules;

(b) Choose hyperparameters as $\lambda \in \Omega$, then train classifier with λ. Define objective function as the mean of OOBErr

$$\lambda_{opt} = \underset{\lambda \in \Omega}{\arg\min}\, OOBErr(\lambda);$$

(c) Sequentially perform: 1. fitting a Gaussian process for data points $\{\lambda_i, OOBErr(\lambda_i)\}$ with new data point updating; 2. identifying new point for evaluation by maximizing the acquisition function;

(d) Terminate iteration when stopping criterion are satisfied.

Algorithm 3: Guided filter for image regularization

(a) Given an initial pixel-wise probabilistic maps $P = \{P_1, \cdots, P_C\}$ with only spectral information;

(b) Process the initial probabilistic maps by Guided filter, resulting in filtered maps $\hat{P} = \{\hat{P}_1, \cdots, \hat{P}_C\}$;

(c) Obtain the class label for pixel j by the maximum of the posterior maps, given by

$$[M_i] = \underset{j \in [1, C]}{\arg\max}\{\hat{P}_1, \cdots, \hat{P}_C\}.$$

Acknowledgements

This work was supported by Science and Technology Facilities Council (STFC) with grant numbers ST/N006852/1 and ST/V00137X/1.

References

[1] J. A. Foley, N. Ramankutty, K. A. Brauman, E. S. Cassidy, J. S. Gerber, M. Johnston, N. D. Mueller, C. OConnell, D. K. Ray, P. C. West, *et al.*, "Solutions for a cultivated planet," *Nature*, vol. 478, no. 7369, p. 337, 2011.

[2] M. Pérez-Ortiz, J. Peña, P. A. Gutiérrez, J. Torres-Sánchez, C. Hervás-Martínez, and F. López-Granados, "A semi-supervised system for weed mapping in sunflower crops using unmanned aerial vehicles and a crop row detection method," *Applied Soft Computing*, vol. 37, pp. 533–544, 2015.

[3] J. M. Peña-barragán, F. López-granados, M. Jurado-expóstto, and L. García-torres, "Mapping ridolfia segetum patches in sunflower crop using remote sensing," *Weed Research*, vol. 47, no. 2, pp. 164–172, 2007.

[4] C. Zhang and J. M. Kovacs, "The application of small unmanned aerial systems for precision agriculture: a review," *Precision agriculture*, vol. 13, no. 6, pp. 693–712, 2012.

[5] A. I. de Castro, M. Jurado-Expósito, J. M. Peña-Barragán, and F. López-Granados, "Airborne multi-spectral imagery for mapping cruciferous weeds in cereal and legume crops," *Precision Agriculture*, vol. 13, no. 3, pp. 302–321, 2012.

[6] J. Lambert, H. Hicks, D. Childs, and R. Freckleton, "Evaluating the potential of unmanned aerial systems for mapping weeds at field scales: a case study with alopecurus myosuroides," *Weed research*, vol. 58, no. 1, pp. 35–45, 2018.

[7] P. Lutman, S. Moss, S. Cook, and S. Welham, "A review of the effects of crop agronomy on the management of a lopecurus myosuroides," *Weed Research*, vol. 53, no. 5, pp. 299–313, 2013.

[8] L. Rew and R. Cousens, "Spatial distribution of weeds in arable crops: are current sampling and analytical methods appropriate?," *Weed Research*, vol. 41, no. 1, pp. 1–18, 2001.

[9] J. Su, C. Liu, M. Coombes, X. Hu, C. Wang, X. Xu, Q. Li, L. Guo, and W.-H. Chen, "Wheat yellow rust monitoring by learning from multispectral uav aerial imagery," *Computers and electronics in agriculture*, vol. 155, pp. 157–166, 2018.

[10] Z. Mi, X. Zhang, J. Su, D. Han, and B. Su, "Wheat stripe rust grading by deep learning with attention mechanism and images from mobile devices," *Frontiers in Plant Science*, vol. 11, 2020.

[11] G. Modica, G. Messina, G. De Luca, V. Fiozzo, and S. Praticò, "Monitoring the vegetation vigor in heterogeneous citrus and olive orchards. a multiscale object-based approach to extract trees crowns from uav multispectral imagery," *Computers and Electronics in Agriculture*, vol. 175, p. 105500, 2020.

[12] N. Wilke, B. Siegmann, J. A. Postma, O. Muller, V. Krieger, R. Pude, and U. Rascher, "Assessment of plant density for barley and wheat using uav multispectral imagery for high-throughput field phenotyping," *Computers and Electronics in Agriculture*, vol. 189, p. 106380, 2021.

[13] S. Haug, A. Michaels, P. Biber, and J. Ostermann, "Plant classification system for crop/weed discrimination without segmentation," in *IEEE winter conference on applications of computer vision*, pp. 1142–1149, IEEE, 2014.

[14] A. Milioto, P. Lottes, and C. Stachniss, "Real-time semantic segmentation of crop and weed for precision agriculture robots leveraging background knowledge in cnns," in *2018 IEEE International Conference on Robotics and Automation (ICRA)*, pp. 2229–2235, IEEE, 2018.

[15] I. Sa, Z. Chen, M. Popović, R. Khanna, F. Liebisch, J. Nieto, and R. Siegwart, "weednet: Dense semantic weed classification using multispectral images and mav for smart farming," *IEEE Robotics and Automation Letters*, vol. 3, no. 1, pp. 588–595, 2018.

[16] F. López-granados, "Weed detection for site-specific weed management: mapping and real-time approaches," *Weed Research*, vol. 51, no. 1, pp. 1–11, 2011.

[17] J. Torres-Sánchez, F. J. Mesas-Carrascosa, F. M. Jiménez-Brenes, A. I. de Castro, and F. López-Granados, "Early detection of broad-leaved and grass weeds in wide row crops using artificial neural networks and uav imagery," *Agronomy*, vol. 11, no. 4, p. 749, 2021.

[18] G. Rozenberg, R. Kent, and L. Blank, "Consumer-grade uav utilized for detecting and analyzing late-season weed spatial distribution patterns in commercial onion fields," *Precision Agriculture*, pp. 1–16, 2021.

[19] J. M. Pena, J. Torres-Sánchez, A. I. de Castro, M. Kelly, and F. López-Granados, "Weed mapping in early-season maize fields using object-based analysis of unmanned aerial vehicle (uav) images," *PloS one*, vol. 8, no. 10, p. e77151, 2013.

[20] P. Lottes, R. Khanna, J. Pfeifer, R. Siegwart, and C. Stachniss, "Uav-based crop and weed classification for smart farming," in *Robotics and Automation (ICRA), 2017 IEEE International Conference on*, pp. 3024–3031, IEEE, 2017.

[21] I. Sa, M. Popović, R. Khanna, Z. Chen, P. Lottes, F. Liebisch, J. Nieto, C. Stachniss, A. Walter, and R. Siegwart, "Weedmap: a large-scale semantic weed mapping framework using aerial multispectral imaging and deep neural network for precision farming," *Remote Sensing*, vol. 10, no. 9, p. 1423, 2018.

[22] J. Su, D. Yi, B. Su, Z. Mi, C. Liu, X. Hu, X. Xu, L. Guo, and W.-H. Chen, "Aerial visual perception in smart farming: Field study of wheat yellow rust monitoring," *IEEE Transactions on Industrial Informatics*, vol. 17, no. 3, pp. 2242–2249, 2021.

[23] F. López-Granados, M. Jurado-Expósito, J. M. Peña-Barragán, and L. García-Torres, "Using remote sensing for identification of late-season grass weed patches in wheat," *Weed Science*, vol. 54, no. 2, pp. 346–353, 2006.

[24] J. Su, M. Coombes, C. Liu, Y. Zhu, X. Song, S. Fang, L. Guo, and W.-H. Chen, "Machine learning-based crop drought mapping system by uav remote sensing rgb imagery," *Unmanned systems*, vol. 8, no. 01, pp. 71–83, 2020.

[25] J. Su, C. Liu, X. Hu, X. Xu, L. Guo, and W.-H. Chen, "Spatio-temporal monitoring of wheat yellow rust using uav multispectral imagery," *Computers and electronics in agriculture*, vol. 167, p. 105035, 2019.

[26] S. Song, Y. Gong, Y. Zhang, G. Huang, and G.-B. Huang, "Dimension reduction by minimum error minimax probability machine," *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, vol. 47, no. 1, pp. 58–69, 2016.

[27] G. Modica, G. De Luca, G. Messina, and S. Praticò, "Comparison and assessment of different object-based classifications using machine learning algorithms and uavs multispectral imagery: a case study in a citrus orchard and an onion crop," *European Journal of Remote Sensing*, vol. 54, no. 1, pp. 431–460, 2021.

[28] X. Kang, S. Li, and J. A. Benediktsson, "Spectral–spatial hyperspectral image classification with edge-preserving filtering," *IEEE transactions on geoscience and remote sensing*, vol. 52, no. 5, pp. 2666–2677, 2014.

[29] K. He, J. Sun, and X. Tang, "Guided image filtering," in *European conference on computer vision*, pp. 1–14, Springer, 2010.

[30] E. R. Hunt Jr, P. C. Doraiswamy, J. E. McMurtrey, C. S. Daughtry, E. M. Perry, and B. Akhmedov, "A visible band index for remote sensing leaf chlorophyll content at the canopy scale," *International Journal of Applied Earth Observation and Geoinformation*, vol. 21, pp. 103–112, 2013.