Black-grass (*Alopecurus myosuroides*) in cereal multispectral detection by UAV

Jonathan Cox[1] , Xiaodong Li[2], Charles Fox[3] and Shaun Coutts[4]

[1]Postdoctoral Research Associate, School of Computer Science, University of Lincoln, Brayford Pool, Lincoln LN6 7TS, UK; and ARWAC Ltd., Laughton, Sleaford NG34 0HE, UK; [2]Research Associate, Lincoln Institute for Agricultural Technology, University of Lincoln, Riseholme Campus, Lincoln LN2 2LF, UK; and ARWAC Ltd., Laughton, Sleaford NG34 0HE, UK; [3]Senior Lecturer in Computer Science, School of Computer Science, University of Lincoln, Brayford Pool, Lincoln LN6 7TS, UK and [4]Associate Professor, Lincoln Institute for Agricultural Technology, University of Lincoln, Riseholme Campus, Lincoln LN2 2LF, UK

Abstract

Site-specific weed management (on the scale of a few meters or less) has the potential to greatly reduce pesticide use and its associated environmental and economic costs. A prerequisite for site-specific weed management is the availability of accurate maps of the weed population that can be generated quickly and cheaply. Improvements and cost reductions in unmanned aerial vehicles (UAVs) and camera technology mean these tools are now readily available for agricultural use. We used UAVs to collect aerial images captured in both RGB and multispectral formats of 12 cereal fields (wheat [*Triticum aestivum* L.] and barley [*Hordeum vulgare* L.]) across eastern England. These data were used to train machine learning models to generate prediction maps of locations of black-grass (*Alopecurus myosuroides* Huds.), a prolific weed in UK cereal fields. We tested machine learning and data set resampling methods to obtain the most accurate system for predicting the presence and absence of weeds in new out-of-sample fields. The accuracy of the system in predicting the absence of *A. myosuroides* is 69% and its presence above 5 g in weight with 77% accuracy in new out-of-sample fields. This system generates prediction maps that can be used by either agricultural machinery or autonomous robotic platforms for precision weed management. Improvements to the accuracy can be made by increasing the number of fields and samples in the data set and the length of time over which data are collected to gather data across the entire growing season.

Introduction

The use of unmanned aerial vehicles (UAVs) and machine learning is increasingly important in precision agriculture, particularly for weed management (Boursianis et al. 2020; Fernández-Quintanilla et al. 2018; Shaner and Beckie 2014; Tsouros et al. 2019). Herbicides are essential for weed management but incur economic and environmental costs that are compounded by the traditional approach of tractor-based bulk application to entire fields, even in weed-free areas. A more efficient method is site-specific weed management (Christensen et al. 2009; López-Granados 2011), which involves creating customized weed maps for each field. Then, herbicide application by small ground robots can focus on areas with weeds, saving time and reducing herbicide use. While cameras on ground robots may also detect weeds, they are slower and require initial travel across the entire field to detect and remove weeds (Binch et al. 2018; Binch and Fox 2017; Hall et al. 2017; Lottes et al. 2016; Milioto et al. 2018; Sheikh et al. 2020; Wendel and Underwood 2016; Wu et al. 2019). Traditionally, the most cost-effective methods to capture aerial images for weed detection and mapping (especially at scale) have been light aircraft or satellites (Brown and Noble 2005; Christensen et al. 2009; Lamb and Brown 2001), but they struggle to detect small patches of weeds due to relatively low spatial resolution. However, recent UAV and camera technology has improved and become cheaper, allowing UAVs to capture higher spatial resolution images for weed mapping (Boursianis et al. 2020; Huang et al. 2018; López-Granados 2011). This has led to a growing commercial sector of UAV companies specializing in aerial imagery for agricultural purposes (Drone AG 2023; Droneflight 2023; Drone Photography Services 2023; iRed 2023; Skeye Train 2023; SkyCam 2023).

We aim to investigate the accuracy of machine learning methods when applied to UAV RGB and multispectral aerial images for predicting the presence of black-grass (*Alopecurus myosuroides* Huds.) in fields that were not in the training data. Accurate predictions for out-of-sample fields are vitally important to the widespread deployment of UAV imagery in weed detection, as collecting ground-truth samples for every new field is not feasible. Previous research has predicted *A. myosuroides* density at a lower resolution (20 m by 20 m) with models that performed poorly on out-of-sample fields (Lambert et al. 2018, 2019), but out-of-sample weed detection may still be viable in other crops (Kim et al. 2019). We explore methods to

improve the prediction resolution to approximately 1 m. To achieve this, ground-truth samples are collected within 1 m by 1 m quadrats, and high-resolution cameras with pixel sizes ranging from 22 to 27 mm are used for the aerial imagery.

Related Work

UAVs are becoming increasingly popular in agriculture as an inexpensive, agile, and time-efficient platform for a wide range of applications (Kim et al. 2019; Maddikunta et al. 2021; Tsouros et al. 2019). Weed detection using UAVs and machine learning methods has been developed across many different crop and weed species (Lambert et al. 2018, 2019; Lottes et al. 2017; Mohidem et al. 2021; Popović et al. 2017; Su et al. 2022); to date two main categories of classification have been studied: pixel-based and object-based classification. Generally, studies using pixel-based classification have captured images at higher altitudes, around 20 m and higher, and studies using object-based classification at lower altitudes, as these images have the resolution required to recognize small objects (e.g., leaves or whole plants). Rozenberg et al. (2021) investigated both pixel- and object- based classification using two different classifiers in onion (*Allium cepa* L.) fields. They found that both pixel and object methods produced very similar results, with average overall accuracies between 94% and 96%. However, the pixel method had better performance with lower-resolution images than the object method.

Other studies have focused on detecting specific weed species such as *A. myosuroides* in wheat (*Triticum aestivum* L.). Lambert et al. (2018) used a pixel-based random forest method across 18 fields using RGB and red-edge (RE) cameras, achieving accuracies of 61% to 87%. They expanded their work by using near-infrared (NIR) cameras, vegetation indices, and a convolutional neural network, resulting in an area under the curve of 0.825 (Lambert et al. 2019). Su et al. (2022) used a pixel-based random forest classifier on RGB, NIR, and RE images, as well as 18 vegetation indices, achieving an accuracy of 93%, but only used one field for sampling and testing.

In this study, we investigate the potential for using UAVs to capture images and machine learning methods to create weed maps. We focus on UAV systems that are readily available for hire by agricultural users and well-known machine learning libraries that can be used for weed distribution analysis. Our first aim was to investigate whether images captured by UAVs are capable of detecting weeds, and if so, the size of weeds that can be detected. To do this, we captured aerial images of twelve wheat and barley (*Hordeum vulgare* L.) fields and created ground-truth data sets of *A. myosuroides* weights. Second, we wanted to determine whether the use of a multispectral camera improves the detection of the weeds or whether the use of a cheaper RGB camera would produce accurate results. To do this, we captured both RGB and multispectral images of the fields and compared them. Third, to establish which machine learning method produces the most accurate results, we evaluated four machine learning methods frequently used for this task. Finally, to investigate whether these models could be transferred across fields, we removed one field from the data set and trained the machine learning applications on the remaining eleven fields to detect weeds in new, previously unobserved fields.

Materials and Methods

To collect the image and ground-truth data, twelve cereal fields with *A. myosuroides* were selected in Bourne, Lincolnshire, and Peterborough, Cambridgeshire, UK. The aerial images and ground-truth data were collected from July 27 to 31, 2020. The fields included six winter wheat, one spring wheat, and five spring barley, with varying sizes ranging from 10 to 40 ha. The spring wheat and spring barley fields were at growth stages GS85 to GS89 and the winter wheat fields were GS90 to GS95, near to being harvested.

Aerial Images

Aerial images were taken by a commercial UAV imagery company (iRed, Emsworth, UK) (iRed 2023) using a MicaSense RedEdge3 Multi-spectral camera. The camera was calibrated, and all bands were stitched and orthorectified using ground control points by iRed, so that the processed files were the starting point for our analysis. The images had pixel sizes ranging from 22 to 27 mm for the RGB images and from 57 to 76 mm for the normalized difference vegetation index (NDVI) and normalized difference red-edge index (NDRE). Images were collected flying at a height of 75 to 100 m and captured five bands: blue (455 to 495 nm), green (540 to 580 nm), red (658 to 678 nm), RE (707 to 727 nm), and NIR (800 to 880 nm). The last two bands are used to calculate the NDVI (Equation 1) and NDRE (Equation 2) (Gitelson et al. 2002; Torres-Sánchez et al. 2014). These are ratios of reflected light over incoming light to detect vegetation and are based on the fact that plant leaves strongly reflect NIR and RE.

$$NDVI = \frac{NIR - Red}{NIR + Red} \qquad [1]$$

$$NDRE = \frac{NIR - RE}{NIR + RE} \qquad [2]$$

Their values range from −1 corresponding to water, 0 for rock and sand, and 0.2 to 1 for vegetation. RGB sensors are much cheaper with higher resolution than NIR or RE sensors (note the resolution difference in our images). Thus, it would be beneficial to train classifiers on RGB and indices that only use RGB data; a commonly used index is the visible atmospherically resistant index (VARI) (Equation 3) (Gitelson et al. 2002), which only needs RGB data:

$$VARI = \frac{Green - Red}{Green + Red - Blue} \qquad [3]$$

Its values range from −1 to 1, as with NDVI and NDRE. Figure 1 shows the color aerial image and the three indices of one field sampled.

In the field, the most obvious difference between the crop and the weed was the level of senescence, with the crops exhibiting greater senescence than the weed. For this reason, we focused on indices that are designed to pick up living photosynthetic material (i.e., green leaves) and that are obtainable with widely available, cheaper, drone-mounted sensors (i.e., five-band multispectral imagery), and that are simple to calculate with only spectral data. Even with these restrictions, there are several other indices we could have used (e.g., Triangular Greenness Index). However, as all these indices focus on the relative amount of green reflectance in different ways, they tend to correlate with each other to varying degrees.
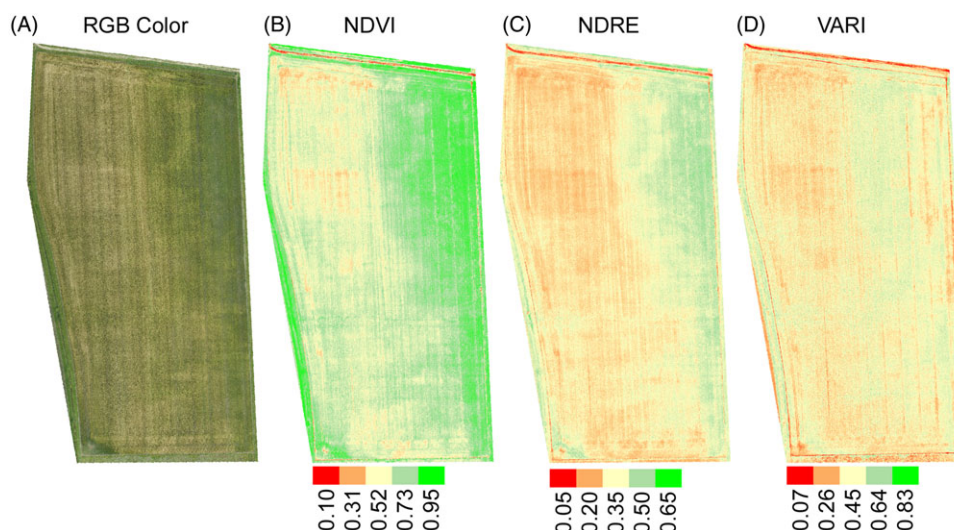
Figure 1. Aerial image and indices of Field 1: (A) RGB color image, (B) normalized difference vegetation (NDVI), (C) normalized difference red-edge (NDRE), and (D) visible atmospherically resistant index (VARI).

Ground-Truth Data Collection

Ground-truth data were obtained by field walking through the twelve fields and collecting all *A. myosuroides* seed heads and stalks within 1 m by 1 m quadrats. Quadrats were placed wherever *A. myosuroides* was encountered, and in cases where there was a large dense patch, we randomly placed the quadrat in two to four locations within the patch, depending on its size. In addition, we sampled no-weed quadrats where we visually verified no weeds were present. The center coordinates of the quadrat were recorded using a Leica CS20 and GS07 RTK GNSS unit (Leica Geosystems 2023) with an accuracy of 10 to 20 mm. The harvested samples were then dried and weighed to create the ground-truth data set, which includes the coordinate at the center of the quadrat and the sample weight. There are 406 *A. myosuroides* and 649 no-weed quadrats, a total of 1,055 quadrats sampled across the twelve fields. Figure 2 shows the distribution of *A. myosuroides* sample weights.

Data Set Definition

Data sets were created by matching up the ground-truth sample coordinates with the coordinates of the orthorectified aerial images. To increase the resolution of the NIR and RE images, the pixels were split up, the new pixels were assigned the same values as the original pixel, and all images were upscaled to the same resolution as the RGB images. For each 1 m by 1 m quadrat, the RGB and multispectral pixels were extracted from the aerial images centered around the GPS location, and the indices were calculated for each pixel. The data set contains the red, green, blue, NDVI, NDRE, and VARI values for each pixel within each 1 m by 1 m quadrat. Each pixel was labeled with the sample weight or zero where the quadrat contained no weeds (i.e., all pixels in the quadrat are assigned the same *A. myosuroides* weight). The data sets consist of pixel and weight data from both wheat and barley fields. Training the classifiers on a data set containing both crop types has the potential to enhance the overall applicability of the classification method, allowing it to develop a more robust understanding of the relevant features and patterns necessary for accurate weed identification and classification.
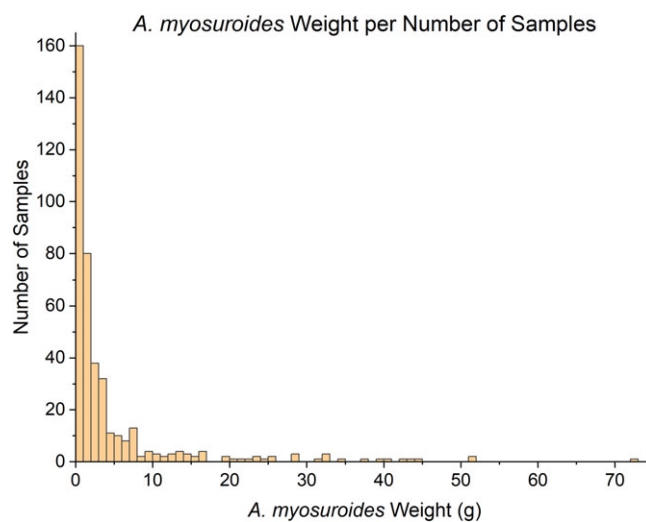

Figure 2. Histogram of *Alopecurus myosuroides* sample weights harvested. This graph excludes the 649 samples with no (0 g) *Alopecurus myosuroides*.

To investigate the sizes of *A. myosuroides* that classifiers can detect, we created four data sets, three with two classes and one with four classes. The aim is to determine whether the classifiers can simply distinguish between a threshold of plant weights or differentiate between several classes (C1, C2, etc.) of weights at the same time:

- two classes: threshold at 0 g, C1 = 0 g, C2 > 0 g
- two classes: threshold at 3 g, C1 $\leq$ 3 g, C2 > 3 g
- two classes: threshold at 5 g, C1 $\leq$ 5 g, C2 > 5 g
- four classes: C1 = 0 g, 0 g < C2 < 5 g, 5 g $\leq$ C3 < 10 g, C4 $\geq$ 10 g

The machine learning classifiers are predicting the probability that a pixel comes from a quadrat in a given weed density class. We tested several other class structures and thresholds. The results showed the same pattern and the conclusions were unaffected. In this paper, we present the most relevant results.

Machine Learning Classification

Four classifiers from Scikit.learn (scikit-learn 2023a) were trained and tested on the data; Random Forest, support vector machine (SVM), gradient boosting, and multilayer perceptron (MLP). They estimate P(quadrat | pixel data), the probability that a quadrat centered on the pixel would be labeled as a weed or not a weed, given the data from the pixel only. For each field the classifiers are run on ground-truth samples for testing and then across the whole aerial image, classifying every pixel to generate weed prediction maps of each field. The random forest and SVM classifiers performed very poorly, and thus we present here only the more relevant data from the gradient-boosting and MLP classifiers. The features the classifiers were trained on were the red, green, blue, NDVI, NDRE, and VARI pixel data. The training data were randomly split, with 10% used for training and 90% for testing. The parameters used for the gradient-boosting classifier were: n estimators = 500, max depth = 4, minimum samples split = 5, and learning rate = 0.01 (scikit-learn 2023b). For the MLP classifier, the parameters used were: activation identity and learning rate invscaling (scikit-learn 2023c).

The data sets used to train and test the machine learning classifiers contained samples from eleven of the twelve fields. Samples from the excluded field were used to test the classifiers and generate the performance metrics. This process was repeated with each field excluded in turn to test the classifiers on unseen fields. The data sets are unbalanced, as overall, 62% of the data is empty quadrats and 38% contains *A. myosuroides*. To improve the classification accuracy, the data are resampled to balance the number of data points in each class. Several resampling methods were tried, including random oversampling, which randomly duplicates samples in the minority classes, and random under-sampling which randomly deletes samples in the majority class. Two methods of synthetic minority oversampling technique (SMOTE) (Batista et al. 2004; Chawla et al. 2002) were also tried to rebalance the data set. SMOTE-ENN combines oversampling (SMOTE) with edited nearest neighbors (ENN) (Wilson 1972); this method generates new samples where some may overlap between classes, and ENN locates and removes misclassified samples by comparing the samples to their nearest neighbors. SMOTE-Tomek also involves SMOTE oversampling, but combines with Tomek links (Tomek 1976). Tomek links finds pairs of samples in opposite classes which are their nearest neighbor. The majority instance of the pair is removed, which removes unwanted overlap and more clearly defines a border between the classes. For both resampling methods, only the minority classes were resampled, and the default parameters were used as detailed in scikit-learn (2023d, 2023e).

Performance Metrics

To investigate the performance of classification, four metrics were calculated: accuracy, balanced accuracy (BA), Cohen's kappa, and Matthews correlation coefficient (MCC). Overall accuracy represents the proportion of all correct predictions, where TP is true positive (correctly classified weeds), TN is true negative (correctly classified no weeds), FP is false positive (incorrectly classified weeds), and FN is false negative (incorrectly classified no weeds). However, accuracy is a poor measure when the testing data set is unbalanced. Consider the extreme case where 99% of samples are positive. A classifier can achieve 99% accuracy by always predicting positive cases, but such a classifier has no predictive skill. We use three other performance measures that attempt to account for class

imbalance in different ways. BA (Equation 4) is the average of the proportion of positive and negative samples that are correctly classified.

$$\frac{1}{2}\left(\frac{TP}{TP + FN} + \frac{TN}{TN + FP}\right) \quad [4]$$

Cohen's kappa, κ (Equation 5) (Cohen 1960) compares observed accuracy, p_0, with expected accuracy, p_e.

$$\kappa = \frac{p_0 - p_e}{1 - p_e} \quad [5]$$

κ ranges between −1 and 1, where 1 is perfect agreement between observed and expected accuracy, between 0 and 1 there is some agreement, and at less than 0 there is no agreement and the classification is worse than random guessing. Finally, we use a balance-corrected measure of the correlation between the observed data and the predictions, MCC (Equation 6) (Boughorbel et al. 2017; Matthews 1975).

$$MCC = \frac{TP \times TN - FP \times FN}{\sqrt{(TP + FP)(TP + FN)(TN + FP)(TN + FN)}} \quad [6]$$

MCC ranges between −1 and 1, where 1 is perfect correlation, 0 is no better than random guessing, and −1 is an inverse prediction (total disagreement between truth and prediction). We are interested in how well our classifiers work on unseen fields, as this is the most likely use case for aerial images, where a pretrained classifier is applied to images from a new field for which there are no ground data. Thus, we calculate TP, TN, FN, and FP on the excluded fields, and where a single value for a classifier and resampling method is given, it is the total performance across all twelve fields when excluded from the training set.

Results and Discussion

Our aims were: first, to determine the size of weeds that can be accurately detected and whether the classifiers can distinguish between several weed size classes; second, to investigate the accuracy of the classification when using multispectral images as compared to RGB color images; third, to evaluate the performance of machine learning classifiers in detecting *A. myosuroides* in cereal fields from images captured using a UAV; and finally to investigate if these models could be transferred across fields. Accurate predictions for weed detection are important for the widespread deployment of UAV imagery for precision weed detection and management.

The results of the best-performing classification methods tested are shown in Tables 1 and 2. The tables illustrate the different classifiers, data sets, resampling methods and image sets used for each test, with data from all twelve fields where samples from the field being classified were removed in turn from the data set.

In our tests, the MLP classifier consistently outperformed gradient boosting across all data and image sets. Additionally, MLP classification required considerably less time to process the data, taking only half to a twentieth of the time compared with gradient boosting. Our performance results indicate that this method of classification is effective at identifying larger *A. myosuroides* plants (larger than 3 g or 5 g). The accuracy for the two-classes threshold at 3 g ranges between 64% and 67%, and with the threshold at 5 g, the accuracies improve to 64% to 72%. However, the classifiers

Table 1. Performance of classifiers for gradient boosting and the multilayer perceptron (MLP) under different resampling strategies and data sets for the color, normalized difference vegetation (NDVI), and normalized difference red-edge (NDRE) images.

Classifier	Data set	Resampling method[a]	Accuracy	Balanced accuracy	MCC[b]	Cohen's kappa
Gradient boosting	Two classes: 0 g	SMOTE-Tomek	0.5419	0.5033	0.0068	0.0067
	Two classes: 3 g	SMOTE-ENN	0.6425	0.6516	0.2025	0.1546
	Two classes: 5 g	SMOTE-Tomek	0.6874	0.6927	0.2218	0.1558
	4 class	Random oversampling	0.3225	0.3572	0.0230	0.0210
MLP	Two classes: 0 g	Random oversampling	0.6139	0.4957	−0.0188	0.0340
	Two classes: 3 g	SMOTE-Tomek	0.6582	0.6798	0.2407	0.1835
	Two classes: 5 g	SMOTE-Tomek	0.6989	0.7213	0.2553	0.1804
	Four classes	Random undersampling	0.3375	0.3855	0.0706	0.0630

[a]SMOTE-ENN, SMOTE-Tomek: resampling techniques that generate or remove samples in the data set to balance the classes.
[b]MCC, Matthews correlation coefficient.

Table 2. Performance of classifiers for gradient boosting and the multilayer perceptron (MLP) using the color and visible atmospherically resistant index (VARI) images.

Classifier	Data set	Resampling method[a]	Accuracy	Balanced accuracy	MCC[b]	Cohen's kappa
Gradient boosting	Two classes: 0 g	SMOTE-ENN	0.5469	0.5213	−0.0138	0.0419
	Two classes: 3 g	SMOTE-ENN	0.6656	0.6587	0.2147	0.1685
	Two classes: 5 g	SMOTE-ENN	0.6447	0.6928	0.2157	0.1398
	Four classes	SMOTE-Tomek	0.3183	0.3703	0.0534	0.0470
MLP	Two classes: 0 g	SMOTE-ENN	0.5443	0.5150	0.1655	0.0298
	Two classes: 3 g	SMOTE-ENN	0.6852	0.6654	0.2269	0.1833
	Two classes: 5 g	SMOTE-ENN	0.6853	0.7262	0.2580	0.1772
	Four classes	SMOTE-Tomek	0.3316	0.3712	0.0641	0.0560

[a]SMOTE-ENN, SMOTE-Tomek: resampling techniques that generates or removes samples in the data set to balance the classes.
[b]MCC, Matthews correlation coefficient.

Table 3. Confusion matrix and metrics of the multilayer perceptron (MLP) with SMOTE-Tomek resampling, a technique that generates or removes samples in the data set to balance the classes, of the data set with four classes of *Alopecurus myosuroides* weight using the color, normalized difference vegetation (NDVI), and normalized difference red-edge (NDRE) images.

		Predicted class			
		1 (0 g)	2 (0 g–5 g)	3 (5 g–10 g)	4 (≥10 g)
Actual class	1 (0 g)	24.05%	36.01%	15.67%	24.27%
	2 (0 g–5 g)	29.03%	39.05%	13.02%	18.90%
	3 (5 g–10 g)	15.43%	17.85%	26.25%	40.48%
	4 (≥10 g)	4.66%	3.54%	28.89%	62.91%
Accuracy					0.3316
Balanced accuracy					0.3712
MCC[a]					0.0641
Cohen's kappa					0.0560

[a]MCC, Matthews correlation coefficient.

encountered difficulty distinguishing between no–*A. myosuroides* quadrats and quadrats with very small *A. myosuroides* plants. With the threshold at 0 g across both the classifiers, resampling methods, and image sets, the accuracies are around 50%, which is no better than guessing. Increasing the threshold from 0 g to 5 g produces a large increase in the performance of the classifier, accuracy increases by 25% and BA by 40%. The classifiers also have difficulty with the four-classes data set. Table 3 shows this result, which has trouble separating no–*A. myosuroides* samples and small *A. myosuroides* plants, class 1 and class 2, but have more success with larger plants, class 4. As shown in Figure 2, a majority of *A. myosuroides* samples collected are small plants (<3 g). The classifier has difficulty with these, because the area within a 1 m by 1 m quadrat that contains a small *A. myosuroides* plant (e.g., <3 g) looks almost identical to a 1 m by 1 m quadrat with no–*A. myosuroides*. These small *A. myosuroides* plants only occupy a few pixels within the quadrat, with the rest of the pixels being soil and

crop. This can cause the classifiers to try and discriminate on features that are not only the weeds but also the surrounding area of soil and crop. Larger weeds are going to be easier to detect, as there are more pixels in the quadrat that have features that relate only to the weed, and not the crop or soil. That is, quadrats with larger weeds are more distinct from the background crop than quadrats with smaller weeds, and thus the classification task is easier.

Table 4 shows two of the best-performing results from the classifiers (Tables 1 and 2). The table shows the results of the true and false rates and the three metrics calculated comparing the RGB, NDVI, and NDRE images with only RGB and RGB-derived VARI. The two results are: the MLP classifier with SMOTE-Tomek resampling on the two-classes 5 g data set using the RGB images, NDRE, and NDVI; and the MLP classifier with SMOTE-ENN resampling on the two-classes 5 g data set using the RGB images and VARI. Using this method, the classifier effectively ignores any

Table 4. The true positive (TP), false positive (FP), true negative (TN), false negative (FN) rates (P is *Alopecurus myosuroides* ≥ 5 g; N is no *Alopecurus myosuroides* <5 g) and metrics (accuracy, balanced accuracy [BA], Matthews correlation coefficient [MCC], and Cohen's kappa) of two results from the classifiers of all twelve fields, multilayer perceptron (MLP) with SMOTE-ENN resampling, a resampling technique that generates or removes samples in the data set to balance the classes, of two classes threshold at 5 g comparing the RGB, NDVI and NDRE images and only using the color images and visible atmospherically resistant index (VARI).

	RGB, NDVI[a], and NDRE[b]	RGB and VARI
TP	74.79%	77.51%
FP	30.54%	32.26%
TN	69.46%	67.74%
FN	25.21%	22.49%
Accuracy	0.6989	0.6853
BA	0.7213	0.7262
MCC	0.2553	0.2580
Cohen's kappa	0.1804	0.1772

[a]NDVI, normalized difference vegetation index.
[b]NDRE, normalized difference red-edge index.

A. myosuroides less than 5 g and treats the area as having no *A. myosuroides*; however, this produces the best-performing results of the metrics and correctly predicts the absence of *A. myosuroides* (class 1) 69.46% and the presence of *A. myosuroides* above 5 g (class 2) 74.79% when using the RGB images, NDRE, and NDVI. A surprising result is (b), which only uses the RGB images and the RGB-derived VARI; the performance of this index is almost the same as (a). Using the RGB and VARI images (b), the classifier correctly predicts the absence of *A. myosuroides* (class 1) 67.74% and the presence of *A. myosuroides* above 5 g (class 2) 77.51%.

Across all twelve fields, the classifiers performance metrics are high; however, the performance varies when looking at individual fields. Figure 3 and Table 5 show the accuracy and true and false rates of the twelve fields individually using the best-performing classifier for each data set. (We do not show BA per field because this estimator is not meaningful for small class sample sizes.)

As expected from the overall metrics (Table 1) the results for the two-classes threshold at 0 g (Figure 3A) are poor with accuracies around or below 50%. As the threshold is increased to 3 g (Figure 3B), and 5 g (Table 5), the accuracy and true and false rates vary but are on the whole higher than those for the 0 g data set, with most field accuracies greater than 70%. With the latter two sets of results, the reason the accuracy varies so much is that very few fields contained an abundance of *A. myosuroides* and amounts above the thresholds of 3 g and 5 g to test the classifiers.

Fields 3 and 6 in Table 5 have a reasonable balance of ground-truth data, representing most of the *A. myosuroides* sample data overall. When these fields are being classified and their data are removed from the training data, most of the *A. myosuroides* data is removed from the training data. As a result, the resampling methods have a very small number of samples to generate from. The classifier also has high accuracy at predicting no–*A. myosuroides* areas across most of the fields, apart from fields 3, 6, and 9 (Table 5), because the colors of the images in these three fields are darker and browner, whereas the other nine fields are lighter and greener; the training data contain mostly lighter areas for no *A. myosuroides*, so the classifier confuses these darker areas. The performance of the classifiers exhibits minimal variation based on the crop type, as both wheat and barley fields demonstrate both high and low accuracies. There is no single crop type that consistently exhibits either high or low accuracy across the classifiers.

The main result is that UAV images and machine learning methods are capable of accurately predicting the presence and absence of *A. myosuroides* in out-of-sample fields. Across twelve fields, the classifiers correctly predicted no *A. myosuroides* 69.46% and *A. myosuroides* 77.51%. However, the accuracy for individual fields is low. Overall, the classifier is accurate with out-of-sample new fields, but this could be improved by increasing the number of fields sampled in the data set. Some of the fields sampled contained high densities of *A. myosuroides*, whereas others contained very little or no *A. myosuroides*, and this skews the data set. A wider variety of fields that contain more and various sizes of *A. myosuroides* should be sampled to improve the classification. Lower reliability and accuracy predictions are useful to indicate the areas with and without weeds so weed management can be focused on those areas with predicted weeds. A robotic weed management platform can spend more time in the areas with predicted weeds rather than covering the entire field, which will be mainly weed-free.

The use of RGB-only images and the calculated VARI results in comparably good predictions: no *A. myosuroides* = 67.74% correctly and *A. myosuroides* = 77.51% correctly (Table 4).

The data sets with RGB, NDRE, and NDVI images have five features on which the classifiers are trained: the five pixel values taken from the aerial images. Each image contains RGB colors and the two indices, NDVI and NDRE. The data sets with RGB and VARI images contain four features. Figure 4 shows the importance of these four and five features to the gradient-boosting classifier and how they influence its predictions. The most important being the NDVI and VARI indices followed by green light; as expected, the two indices highlight the *A. myosuroides* and the green of the *A. myosuroides* plant is picked out against the background of the soil and crop. Previous studies of multispectral images on weed detection have reported that NIR light is important for this task (López-Granados et al. 2008; Smith and Blackshaw 2003). The classifiers can be used to generate prediction maps of *A. myosuroides* locations. Figure 5 shows one of the fields surveyed using the two-classes MLP classifier with the threshold at 5 g. Figure 5a is prediction map overlaid on the aerial image; the red areas are where the classifier predicts the location of *A. myosuroides* with a probability greater than 75%, and the locations have been grouped into clusters to more clearly show the areas with weeds. Figure 5A shows the map of the classifier's prediction probability, demonstrating the classifier is confident across most of the field. These maps could be used to remove *A. myosuroides* more efficiently by focusing weed management methods on the highlighted areas rather than the whole field.

We have shown UAV imagery and machine learning can be used to accurately classify and predict the presence and absence of *A. myosuroides*. Across twelve out-of- sample fields, our classifier predicts the absence (69.46%) and presence (77.51%) of *A. myosuroides* correctly, with Cohen's kappa from 0.177 to 0.186. Other studies have higher kappas, 0.22 to 0.32. Direct comparison of metrics is hard, as these studies used visual observation to classify *A. myosuroides* abundance (Lambert et al. 2018, 2019) and larger sample areas. In our study, larger weed plants (>3 g) can be more accurately detected, as smaller plants are harder to distinguish against the background of soil and crops.

Our results show that using only a RGB color camera and deriving the VARI yields accurate results as good as those produced using multispectral cameras. This may be useful where only RGB color images are available and to reduce costs and weight by not using multispectral cameras on UAVs. Of the two machine learning methods shown, MLP was slightly more accurate than

Table 5. Classification accuracy, true positive (TP), false positive (FP), true negative (TN), and false negative (FN) rates and ground-truth data (P is *Alopecurus myosuroides* ≥ 5 g; N is no *Alopecurus myosuroides* < 5 g) of the twelve individual fields using the multilayer perceptron (MLP) classifier with SMOTE-Tomek resampling, a resampling technique that generates or removes samples in the data set to balance the classes, of two-classes threshold at 5 g using the color, normalized difference vegetation (NDVI), and normalized difference red-edge (NDRE) images.

| Field number	Crop	Accuracy	TP	FP	TN	FN	Ground truth (P/N)
1	Spring barley	0.9540	4.18%	95.82%	96.29%	3.71%	1/102
2	Winter wheat	0.8574	0.00%	0.00%	85.74%	14.26%	0/73
3	Winter wheat	0.2717	96.65%	3.35%	1.82%	98.18%	27/74
4	Winter wheat	0.8411	9.15%	90.85%	89.67%	10.33%	6/81
5	Spring barley	0.5884	0.00%	0.00%	58.84%	41.16%	0/62
6	Winter wheat	0.4135	86.13%	13.87%	15.23%	84.77%	35/60
7	Spring barley	0.9604	0.00%	0.00%	96.04%	3.96%	0/62
8	Spring wheat	0.8949	0.00%	100.00%	90.98%	9.02%	1/60
9	Winter wheat	0.1792	85.94%	14.06%	14.57%	85.43%	4/81
10	Spring barley	0.5529	50.61%	49.39%	55.58%	44.42%	6/95
11	Winter wheat	0.9903	0.00%	100.00%	99.97%	0.03%	1/102
12	Spring barley	0.9269	0.00%	100.00%	95.91%	4.09%	0/73
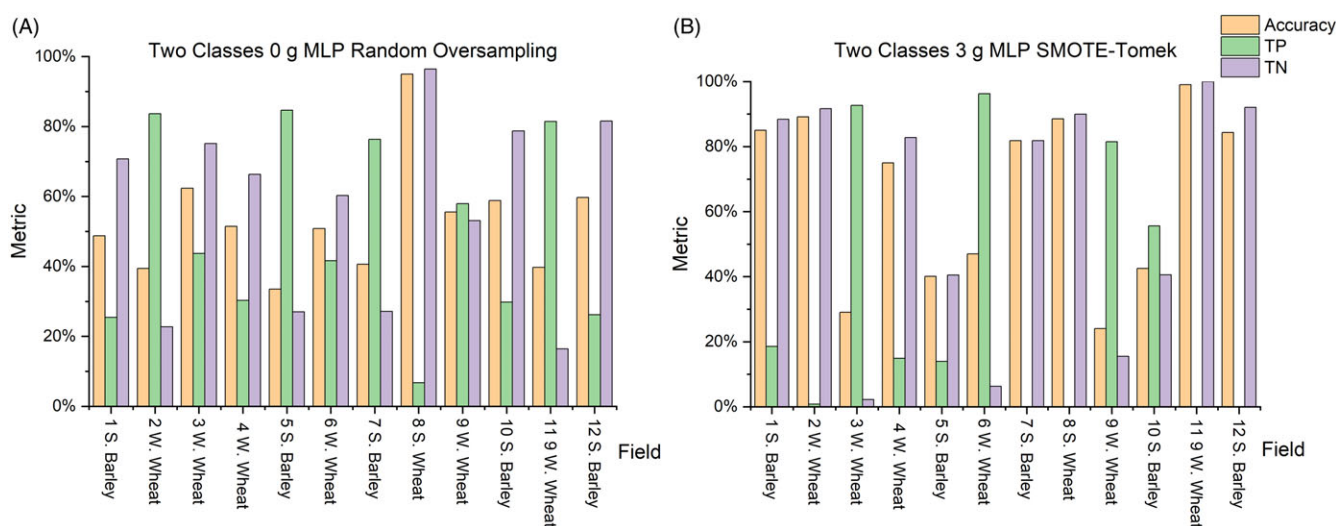
Figure 3. Accuracy and true positive (TP) and true negative (TN) rates for the twelve individual fields sampled using the multilayer perceptron (MLP) classifier with (A) random oversampling of the two classes data set with the threshold at 0 g and (B) SMOTE-Tomek resampling, a resampling technique that generates or removes samples in the data set to balance the classes, of the two-classes data set with the threshold at 3 g, both using the color, normalized difference vegetation (NDVI), and normalized difference red edge (NDRE) images.

gradient boosting at classifying pixels as either a weed or not a weed for both multispectral and RGB images.

The classifier's accuracy is low in individual fields, as in other studies (Lambert et al. 2018, 2019), due to fields not containing abundant and varying amounts and sizes of weed plants. As with all machine learning tasks, the main limitation is the availability of a large amount of relevant, labeled data, in this task, collected across multiple fields at different growth stages of the crops and weeds. The method may work better if samples and images are collected slightly earlier in the growing season (early June to July) when previous studies collected data (Lambert et al. 2018, 2019; Su et al. 2022) or from a longer period of time across the growing season. Like other studies in this area, our classifiers are unlikely to work well outside the context in which they were trained (Lambert et al. 2018, 2019). This is especially true with respect to the growth stage of the crop. The data we trained on were collected late in the season (late July) when the crop was yellower (more senesced) than the weed. As a result, this approach is highly unlikely to work before senescence has begun without additional data and model training. We also targeted fields with known *A. myosuroides* infestations. So most green weeds in these fields were *A. myosuroides*. But this is
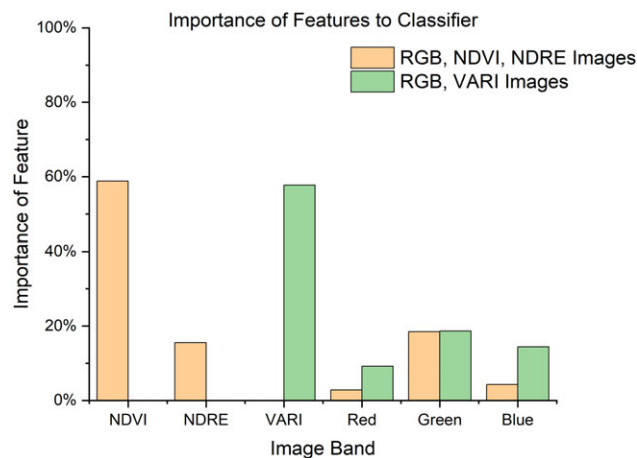

Figure 4. Feature importance of the gradient-boosting classifier with SMOTE-Tomek resampling of the two-classes threshold at 5 g data set using the color (RGB) images, normalized difference red-edge (NDRE), and normalized difference vegetation (NDVI) and SMOTE-ENN resampling, a resampling technique that generates or removes samples in the data set to balance the classes, of the two-classes threshold at 5 g data set using the color images and visible atmospherically resistant index (VARI) averaged across all twelve fields.
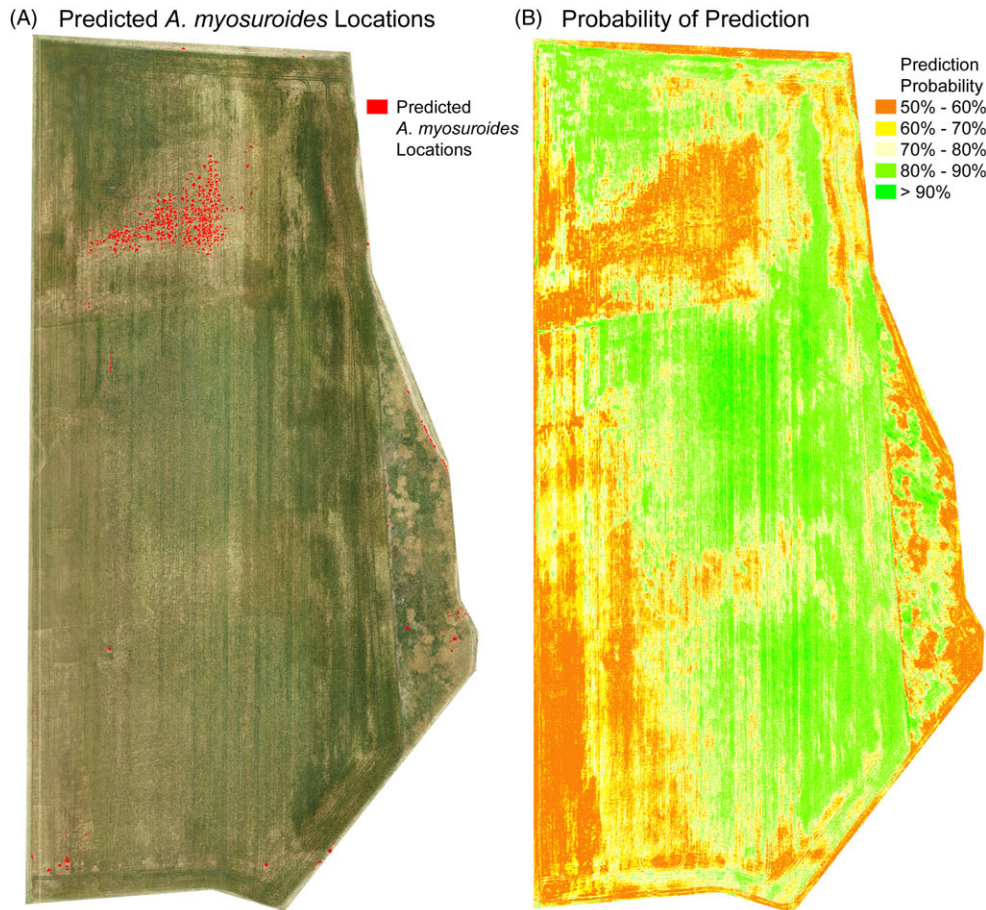
(A) Predicted *A. myosuroides* Locations   (B) Probability of Prediction

Figure 5. Aerial images of field twelve: (A) color image overlaid with the classifier's predicted locations of *Alopecurus myosuroides* (red areas) where its predicted probability is greater than 75% and (B) the classifier's prediction probability of either class 1 or class 2 (C1 ≤ 5 g or C2 > 5 g). The classifier used is the multilayer perceptron (MLP) classifier with SMOTE-Tomek resampling, a resampling technique that generates or removes samples in the data set to balance the classes, with two-classes threshold at 5 g with 1,703 average pixels per quadrat using the color images, normalized difference red-edge (NDRE), and normalized difference vegetation (NDVI) data sets.

not always going to be the case, and in fields with large populations of other weed species, greener areas are more likely to be patches of other species. Collecting data to train classifiers across multiple species would be a valuable, if expensive, project.

However, our classifiers worked successfully even very late in the season just before harvesting. The efficacy of our method in detecting larger weeds highlights its suitability for utilizing UAV imagery and implementing weed management later in the growing season when the weeds are larger. Most common herbicides used to stop *A. myosuroides* are preemergence herbicides, applied before the crop and weeds emerge. As these images are taken at the end of the season, the model presented can be used to predict weed distributions at the end of the growing season. Late-season images can be used to track the progress of weed control programs both in season and over longer time periods. Also, if only large patches require control, weed maps from the previous year can be used for variable herbicide application in the following rotation, because large weed patches do not move much from year to year. But if very high levels of weed control, with even small individual plants targeted, are required, the classifier used would have to have very low FN rates, even for small plants, much lower than the best classifier we present here. Such a classifier could be built but will require higher-quality images (possibly collected from ground vehicles) and much larger and more diverse annotated data sets. Future work could also involve agricultural economics surveys to

determine the best time to capture UAV images, use the classification method, and apply herbicides. There are no major technical barriers to achieving this, but it would require sustained investment over multiple years.

References

Batista GEAPA, Prati RC, Monard MC (2004) A study of the behavior of several methods for balancing machine learning training data. SIGKDD Explor Newsl 6:20–29

Binch A, Cooke N, Fox C (2018) Rumex and Urtica Detection in Grassland by UAV. In Proceedings of the 14th International Conference on Precision Agriculture. Montreal: International Society of Precision Agriculture

Binch A, Fox C (2017) Controlled comparison of machine vision algorithms for Rumex and Urtica detection in grassland. Comput Electron Agric 140: 123–138.

Boughorbel S, Jarray F, El-Anbari M (2017) Optimal classifier for imbalanced data using Matthews correlation coefficient metric. PLoS ONE 12:1–17

Boursianis AD, Papadopoulou MS, Diamantoulakis P, Liopa-Tsakalidi A, Barouchas P, Salahas G, Karagiannidis G, Wan S, Goudos SK (2020) Internet of Things (IoT) and agricultural unmanned aerial vehicles (UAVs) in smart farming: a comprehensive review. Internet of Things 18:100187

Brown RB, Noble SD (2005) Site-specific weed management: sensing requirements—what do we need to see? Weed Sci 53:252–258

Chawla NV, Bowyer KW, Hall LO, Kegelmeyer WP (2002) SMOTE: synthetic minority over-sampling technique. J Artif Intell Res 16:321–357

Christensen S, Søgaard HT, Kudsk P, Nørremark M, Lund I, Nadimi ES, Jørgensen R (2009) Site-specific weed control technologies. Weed Res 49:233–241

Cohen J (1960) A coefficient of agreement for nominal scales. Educ Psychol Meas 20:37–46

Drone AG (2023) Drone Ag: Farmers Who Know Drones—Software/Training/Sensors. https://droneag.farm.Accessed: February 8, 2023

Droneflight (2023) Inspection Services and Survey Services Using Drones. https://droneflight.co.uk. Accessed: February 8, 2023

Drone Photography Services (2023) Aerial Photography and Filming using Drones | Cornwall & the UK. https://dronephotographyservices.co.uk Accessed: February 8, 2023

Fernández-Quintanilla C, Peña JM, Andújar D, Dorado J, Ribeiro A, López-Granados F (2018) Is the current state of the art of weed monitoring suitable for site-specific weed management in arable crops? Weed Res 58:259–272

Gitelson AA, Kaufman YJ, Stark R, Rundquist D (2002) Novel algorithms for remote estimation of vegetation fraction. Remote Sens Environ 80:76–87

Hall D, Dayoub F, Kulk J, McCool C (2017) Towards unsupervised weed scouting for agricultural robotics. Pages 5223–5230 in 2017 IEEE International Conference on Robotics and Automation (ICRA). Singapore: IEEE

Huang H, Deng J, Lan Y, Yang A, Deng X, Zhang L (2018) A fully convolutional network for weed mapping of unmanned aerial vehicle (UAV) imagery. PLoS ONE 13:1–19

iRed (2023) Thermal Imaging, Remote Sensing and Integrated Drone Solutions. https://ired.co.uk.Accessed: February 8, 2023

Kim J, Kim S, Ju C, Son HI (2019) Unmanned aerial vehicles in agriculture: a review of perspective of platform, control, and applications. IEEE Access 7:105100–105115

Lamb DW, Brown RB (2001) PA—precision agriculture: remote-sensing and mapping of weeds in crops. J Agric Eng Res 78:117–125

Lambert JP, Childs DZ, Freckleton RP (2019) Testing the ability of unmanned aerial systems and machine learning to map weeds at subfield scales: a test with the weed *Alopecurus myosuroides* (Huds). Pest Manag Sci 75:2283–2294

Lambert JPT, Hicks HL, Childs DZ, Freckleton RP (2018) Evaluating the potential of unmanned aerial systems for mapping weeds at field scales: a case study with *Alopecurus myosuroides*. Weed Res 58:35–45

Leica Geosystems (2023) Leica GS07. https://leica-geosystems.com/en-gb/products/gnss-systems/smart-antennas/leica-gs07-made-for-you. Accessed: February 8, 2023

López-Granados F (2011) Weed detection for site-specific weed management: mapping and real-time approaches. Weed Res 51:1–11

López-Granados F, Peña-Barragán JM, Jurado-Expósito M, Francisco-Fernández M, Cao R, Alonso-Betanzos A, Fontenla-Romero O (2008) Multispectral classification of grass weeds and wheat (*Triticum durum*) using linear and nonparametric functional discriminant analysis and neural networks. Weed Res 48:28–37

Lottes P, Hoeferlin M, Sander S, Müter M, Schulze P, Stachniss LC (2016) An effective classification system for separating sugar beets and weeds for precision farming applications. Pages 5157–5163 in 2016 IEEE International Conference on Robotics and Automation (ICRA). Stockholm: IEEE

Lottes P, Khanna R, Pfeifer J, Siegwart R, Stachniss C (2017) UAV-based crop and weed classification for smart farming. Pages 3024–3031 in 2017 IEEE International Conference on Robotics and Automation (ICRA). Singapore: IEEE

Maddikunta PKR, Hakak S, Alazab M, Bhattacharya S, Gadekallu TR, Khan WZ, Pham QV (2021) Unmanned aerial vehicles in smart agriculture: applications, requirements, and challenges. IEEE Sensors J 21:17608–17619

Matthews B (1975) Comparison of the predicted and observed secondary structure of T4 phage lysozyme. Biochimi Biophys Acta 405:442–451

Milioto A, Lottes P, Stachniss C (2018) Real-time semantic segmentation of crop and weed for precision agriculture robots leveraging background knowledge in CNNs. Pages 2229–2235 in 2018 IEEE International Conference on Robotics and Automation (ICRA). Brisbane: IEEE

Mohidem NA, Che'Ya NN, Juraimi AS, Fazlil Ilahi WF, Mohd Roslim MH, Sulaiman N, Saberioon M, Mohd Noor N (2021) How can unmanned aerial vehicles be used for detecting weeds in agricultural fields? Agriculture 11:1004

Popović M, Hitz G, Nieto J, Sa I, Siegwart R, Galceran E (2017) Online informative path planning for active classification using UAVs. Pages 5753–5758 in 2017 IEEE International Conference on Robotics and Automation (ICRA). Singapore: IEEE

Rozenberg G, Kent R, Blank L (2021) Consumer-grade UAV utilized for detecting and analyzing late-season weed spatial distribution patterns in commercial onion fields. Precis Agric 22:1317–1332

scikit-learn (2023a) scikit-learn: machine learning in Python. https://scikit-learn.org/stable. Accessed: February 8, 2023

scikit-learn (2023b) sklearn.ensemble. GradientBoostingClassifier. scikit-learn 1.2.2 Documentation. https://scikit-learn.org/stable/modules/generated/sklearn.ensemble. GradientBoostingClassifier.html. Accessed: February 8, 2023

scikit-learn (2023c) sklearn.neuralnetwork. MLPClassifier−scikit−learn1.2.2-documentation. https://scikit-learn.org/stable/modules/generated/sklearn.neural_network.MLPClassifier.html Accessed: February 8, 2023

scikit-learn (2023d) SMOTEENN, Version 0.10.1. https://imbalanced-learn.org/stable/references/generated/imblearn.combine.SMOTEENN.html. Accessed: February 8, 2023

scikit-learn (2023e) SMOTETomek, Version 0.10.1. https://imbalanced-learn.org/stable/references/generated/imblearn.combine.SMOTETomek.html. Accessed: February 8, 2023

Shaner DL, Beckie HJ (2014) The future for weed control and technology. Pest Manag Sci 70:1329–1339

Sheikh R, Milioto A, Lottes P, Stachniss C, Bennewitz M, Schultz T (2020) Gradient and log-based active learning for semantic segmentation of crop and weed for agricultural robots. Pages 1350–1356 in 2020 IEEE International Conference on Robotics and Automation (ICRA). Paris: IEEE

Skeye Train (2023) Skeye Train—The UK's Leading Drone Training Provider. https://skeyetrain.co.uk.Accessed: February 8, 2023

SkyCam (2023) Hire Drone & Pilots in UK for Aerial Activities like Photography, Surveys, Filming & Videography. https://www.theskycam.co.uk Accessed: February 8, 2023

Smith AM, Blackshaw RE (2003) Weed: crop discrimination using remote sensing: a detached leaf experiment. Weed Technol 17:811–820

Su J, Yi D, Coombes M, Liu C, Zhai X, McDonald-Maier K, Chen WH (2022) Spectral analysis and mapping of blackgrass weed by leveraging machine learning and UAV multispectral imagery. Comput Electron Agric 192:106621

Tomek I (1976) Two modifications of CNN. IEEE Transactions on Systems, Man, and Cybernetics SMC-6:769–772

Torres-Sánchez J, Peña J, de Castro A, López-Granados F (2014) Multi-temporal mapping of the vegetation fraction in early-season wheat fields using images from UAV. Comput Electron Agric 103:104–113

Tsouros DC, Bibi S, Sarigiannidis PG (2019) A review on UAV-based applications for precision agriculture. Information 10:349

Wendel A, Underwood J (2016) Self-supervised weed detection in vegetable crops using ground based hyperspectral imaging. Pages 5128–5135 in 2016 IEEE international conference on robotics and automation (ICRA). Stockholm: IEEE

Wilson DL (1972) Asymptotic properties of nearest neighbor rules using edited data. IEEETransactions on Systems, Man, and Cybernetics SMC-2:408–421

Wu X, Aravecchia S, Pradalier C (2019) Design and implementation of computer vision based in-row weeding system. Pages 4218–4224 in 2019 International Conference on Robotics and Automation (ICRA). Montreal: IEEE