Smart Agricultural Technology 5 (2023) 100249

Contents lists available at ScienceDirect

Smart Agricultural Technology

journal homepage: www.journals.elsevier.com/smart-agricultural-technology

Machine learning and handcrafted image processing methods for classifying common weeds in corn field

Harsh Pathak [a,b], C. Igathinathane [a,*], Kirk Howatt [d], Zhao Zhang [a,c]

[a] Department of Agricultural and Biosystems Engineering, North Dakota State University, 1221 Albrecht Boulevard, Fargo, ND 58102, USA
[b] Department of Agricultural & Biological Engineering, Purdue University, 225 South University Street, West Lafayette, IN 47907, USA
[c] Key Laboratory of Modern Precision Agriculture System Integration Research, Ministry of Education, China Agricultural University, Beijing, China
[d] Department of Plant Sciences, North Dakota State University, 1420 Bolley Dr, Fargo, ND 58105, USA

ARTICLE INFO

Editor: Stephen Symons

Dataset link: https://doi.org/10.17632/787mkh2kjc.3

Keywords:
Open-source
Weed management
Weed classification
Image processing
Computer vision
Machine learning
Precision agriculture

ABSTRACT

Weed management practices strive to reduce weeds, which compete with crops for nutrients, sunlight, and water and are thus important for maintaining yield. In most weed management practices, the first step is to identify or classify weeds. However, efficient identification and classification of weeds are challenging using conventional manual methods such as field visits. Therefore, in this study, we propose to classify four common weeds in the corn field of North Dakota (common lambsquarters, common purslane, horseweed, and redroot pigweed), also applicable to other regions, using computer vision methods. Weeds were grown in plastic trays under natural conditions from the field soil, and images were collected using an RGB camera. For each weed species, 21 shape features were extracted through a developed ImageJ plugin. The handcrafted simple image processing approach was highly successful in distinguishing common lambsquarters and redroot pigweed from horseweed, while did not perform well with other species combinations and classifying them collectively owing to the similarity of the shapes. However, the three advanced non-parametric machine learning (ML) models, namely, k-nearest neighbor (kNN), random forest (RF), and support vector machine (SVM), resulted in high accuracies with RF outperforming the others. It is recommended that the handcrafted simple image processing algorithm should be tried first, due to its simplicity, for the identification and classification of weeds before resorting to advanced and complex versatile ML modeling approaches. Such tools and methodologies using open-source platforms will be easily accessible and helpful to farmers, producers, and other stakeholders related to crop production.

1. Introduction

Weed pressure causes corn and soybean production to face a yield and financial loss of 50 % and 52 % amounting to 26.7 and 16.2 billion U.S. dollars, respectively [1,2]. Weeds pose a large threat to production as they reduce yield by blocking irrigation canals and by competing against crops for nutrients, water, and sunlight [3,4]. Weeds also harbor pests and release chemicals that are inhibitive to the growth of crops. Corn seeds have the ability to sense whether weeds are growing above the ground by the light that weeds reflect that penetrates the soil. This sense triggers cellular changes that delay germination and hence cause yield losses [5]. In soybean, weed infestation causes delayed maturity, reduced plant height and grain yield, and also affects seed quality. It is, therefore, important to conduct weed management, in which efficient quantification and accurate inspection of weeds play an important role.

Cultural practices such as crop rotation, growing cover crops, and optimizing planting date, planting density, and row spacing were followed to reduce weeds in the field. To handle weeds during the growing season, chemical methods of weed control can be followed. Chemical methods that involve the use of herbicides serve as the most common and highly effective tool for weed management practices to maintain the quality and quantity of crop production [6,7].

Chemical methods, which involve the use of herbicides, are excessively used as a blanket application that can cause water contamination, biodiversity reduction loss in soil moisture and nutrient content, and health hazards to the operator and consumers [6–8]. Furthermore, it gives rise to herbicide-resistance weeds (e.g., Palmer amaranth (*Amaranthus palmeri*), and waterhemp (*Amaranthus tuberculatus*)) [8]. As per the Weed Science Society of America (WSSA), around 43 varieties of glyphosate-resistant (Roundup) weeds exist across the globe [9]. To ad-

* Corresponding author.
E-mail address: Igathinathane.Cannayen@ndsu.edu (C. Igathinathane).

https://doi.org/10.1016/j.atech.2023.100249
Received 16 November 2022; Received in revised form 2 May 2023; Accepted 2 May 2023
Available online 10 May 2023
2772-3755/Published by Elsevier B.V. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

dress the weed issues properly, researchers put forward the site-specific weed management (SSWM) concept. By applying the herbicide exclusively to weeds precisely, SSWM aims to reduce environmental exposure and reduce production costs (less herbicide input). The SSWM concept has been gradually accepted by researchers and farmers.

An alternative to chemical methods is to employ mechanical weeding systems (self-propelled or tractor-driven), which is commonly followed in organic farming. Organic farming contributes to 4 % of the total US food market and is the fastest-growing sector in the US [10]. Mechanical weeding systems can handle emerging weeds during the season. The automated version of mechanical weeding (robots) decreases labor inputs and uses electronic sensors and global positioning system-based guidance along with complex software and hardware technologies to work in non-row crop fields [6]. Automated systems have a high initial price and management costs, poor effectiveness on intra-row weeds, and performance depends on pedoclimatic conditions (especially soil texture and moisture), weed species, and weed growth stage [11]. In general, mechanical weeding systems might cause damage to crops, soil erosion, and nutrient loss.

For the successful implementation of either chemical or mechanical weed control methods, accurate identification of weeds is necessary. Therefore, there is a need for an efficient, robust, reliable, and low-cost method that requires less processing and computational time for weed identification. Nowadays, most weed identification research employs digital image processing and computer vision methods [12]. The most common ways to discriminate weeds are using the visual features that can be further divided into spectral features, spatial contexts, and biological morphology present in the digital image [13].

The major drawback of using the spectral features for weed classification is that the algorithm is not robust and efficient because of three main reasons (1) loading limit of the device to handle multiple cameras on drones, (2) spectral information varies significantly in the natural field environment, and (3) physiological stress affects leaf reflectance properties thereby interfering in classification [8]. In addition, algorithms that use spectral information for classification can underperform if two plants have similar colors [13], and the algorithms tend to be complex and computationally intensive. To overcome this problem, researchers have started using more narrower bands by using the hyperspectral sensors to gain more spectral information of plants [14]. These methods are capable of providing great solutions to weed classification and identification but have certain limitations, such requires intensive camera calibration for different lighting conditions, and high input cost of setup and processing the data. These issues limit its adoption by the farmers. Therefore, the best alternative to hyperspectral sensors would be to use RGB sensors, to make the technology feasible and easily accessible to the farmers [15].

Many algorithms have been developed by using the RGB bands that use spatial context features rather than the spectral information for classifying the weed species. The spatial context features take into account the location or position of the plant for discrimination but fail to consider intra-rows weeds [16,17]. In contrast with the biological morphology, which includes the shape and structure of plants or their parts can help in identifying both inter and intra-row weeds [13]. A shape-based weed detection algorithm was developed to identify weeds from sugar beets, which gave an accuracy of 92.30 ± 2.53 % on the testing dataset using support vector machine (SVM) [18]. The k-nearest neighbor (kNN) was used to classify sugar beets and weeds using textural features and achieved an accuracy of around 91 % [19]. The random forest (RF), also known as an ensemble learning algorithm, is a supervised learning model that comprises a large number of decision trees that avoids over-fitting [20,21]. This RF algorithm has been widely used in agriculture for different purposes, such as disease detection and classification [22], and weed classification [23] because it can handle high dimensionality data, has a fast operation speed, is insensitive to overfitting, and provides good accuracies [24].

Most of the reported research was conducted using proprietary software packages like MATLAB because of easier coding experience from the in-built libraries, and hence, limits its use for farmers [25]. Furthermore, the black-box architecture of algorithms and non open-source codes makes data security a major concern among stakeholders. Therefore, the major objective of the study is to use RGB bands images and open-source software ImageJ [26] and Python and develop a simple image processing-based approach and advanced ML models to classify four common weed species, namely, common lambsquarters (*Chenopodium album*), common purslane (*Portulaca oleracea*), horseweed (*Erigeron canadensis*), and redroot pigweed (*Amaranthus retroflexus*). Using the RGB bands images and open-source software platforms, will make the technology feasible and easily accessible to stakeholders. In addition, because of transparency in the software architecture and open-source codes, the software will enforce data transparency and security and stakeholders will be able to know the background and methodology involved in the process.

In this research, two methods of image processing were developed and compared, namely, simple handcrafted image processing and advanced ML. The simple handcrafted image processing approach involves shape features extraction and plugin development using ImageJ. In advanced methods, the extracted shape features from the simple handcrafted approach were used to train and test three non-parametric ML models (kNN, RF, and SVM). The performance of the methodologies will be assessed through standard performance measures of model development. The methodologies developed are expected to be extended to other weed species as well.

2. Materials and methods

2.1. Test sample materials and data acquisition

Weeds come up naturally in the fields that are specific to the particular location or field. As the weed seeds are already present in the field soil, the weeds will emerge when the conditions are right (e.g., moisture, sunlight, and nutrients). To obtain samples of local weeds in North Dakota, it is just enough to have the field soil as the starting material. Therefore, local weeds grown in eight plastic trays filled with the field soil (Fig. 1A). The field soil samples were collected from the Northern Great Plains Research Laboratory, USDA-ARS, Mandan, ND, USA (46°48′38.8″N, 100°54′48.1″W). To ensure a proper growing environment for weeds, the practices observed were: (i) the trays were exposed to sunlight, (ii) trays were watered once a day, and (iii) two weeks after plant emergence, a dose of “Miracle Grow” solution that had 15-30-15 N-P-K was applied during watering. In this study, the natural environment was used to closely represent field conditions, while controlled greenhouse conditions were reported in several studies [27,28].

A digital single-lens reflex (DSLR) camera (Model Nikon D5100, Nikon Corporation, Tokyo, Japan) was used to capture RGB images of the weeds growing in the tray using the auto mode for data collection. The camera was held perpendicular to the soil surface at a distance of 0.75 m ± 0.25 m. The collected images were preprocessed by cropping them, such that each frame contained only one seedling of weed using ImageJ tools [26].

A dataset of 200 images of each weed species was obtained (total: 4 × 200 = 800). The cropped images were resized to 256 × 256 to have a uniform set of images for further processing (Fig. 1A). The data augmentation methods such as shear, rotation, horizontal and vertical flips, and zoom were also employed to increase the size of the dataset for feeding into the ML classifiers to increase the robustness of the model. In total, sets of 1000 images were created for each of the weed species.

2.2. Image preprocessing

The imagery data collected from the handheld RGB camera was used for classifying the weeds. The overall process flowchart of distinguish-

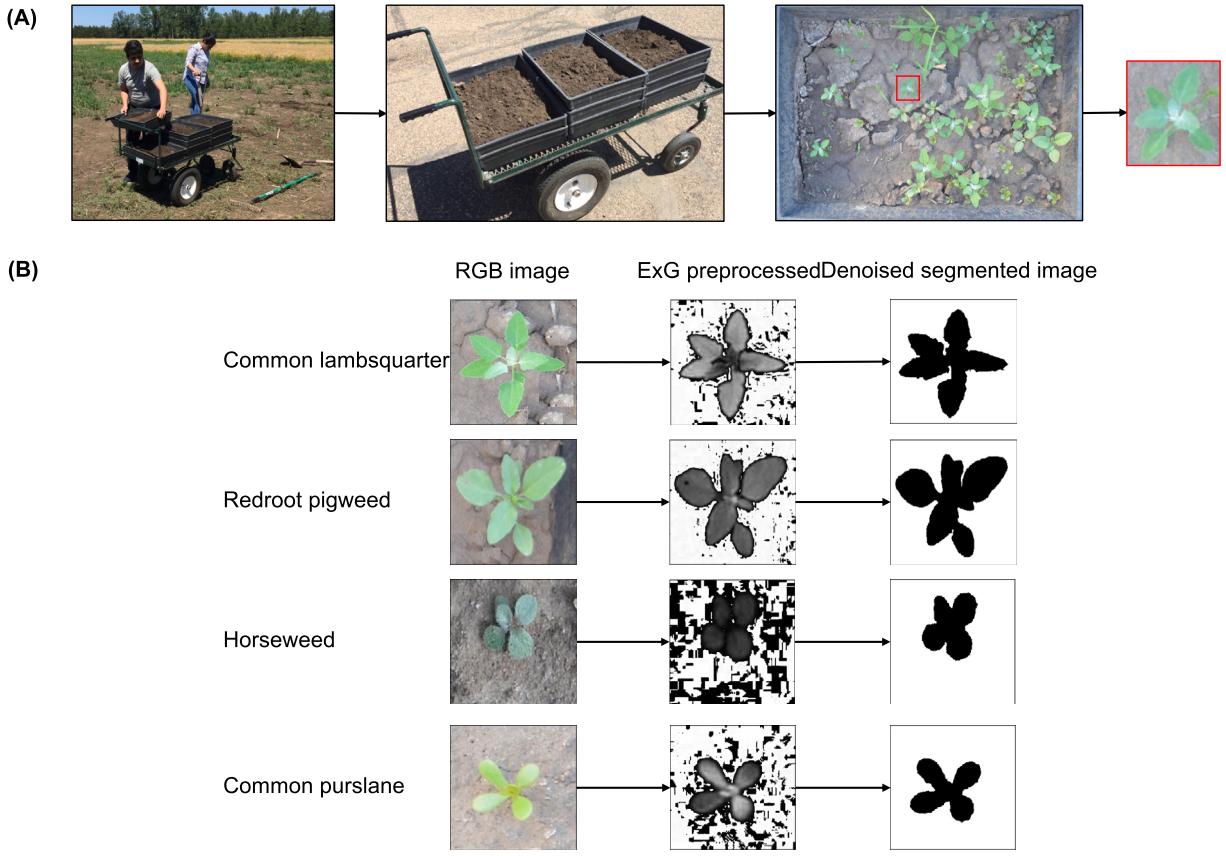


Fig. 1. Field soil collection for growing weeds naturally in plastic trays and image preprocessing. (A) Field soil collected in trays, emerged weeds and whole tray image, and manual clipping of individual plant images from the weeds grown in trays, and (B) Image preprocessing of clipped images using excess green (ExG) preprocessing and segmented denoised binary image.

ing the weed species using shape features with manually clipped images through a direct image processing approach and three ML algorithms showing various stages of image analysis plugin development is presented in Fig. 2. The input image was first segmented by the "Excess Green" (ExG) segmentation method during the preprocessing step and then 21 shape features (standard and derived) were extracted using ImageJ outputs. Shape features were used to create an image analysis weed classification ImageJ plugin. These extracted shape features were further used to train and test the three non-parametric ML algorithms.

After the data acquisition, the next step was to create a binary image that could be further processed to extract shape features for weed identification. Binary image creation involves dividing the input image into two sections namely plant (weed) and background (soil or residue), and this process is known as segmentation. It is an essential step in image processing because inappropriate segmentation can lead to misclassification and decreases the accuracy of the algorithm [29]. The ExG segmentation method [30], a color index-based segmentation method, was used because of its low sensitivity to background errors and lighting conditions [29,31].

$$\text{Object pixel}_{(x,y)} = \begin{cases} \text{Plant, value} = 255 & 2G - B - R \geq \text{Th} \\ \text{Background, value} = 0 & \text{Otherwise} \end{cases}$$

After segmentation, following the logic of the above equation, the "ParticleAnalyzer" ImageJ command (which is coded in the developed plugin) was used to create a binary mask for the area threshold with a lower limit of 1000 pixel^2 to remove noise and other artifacts (Fig. 1B).

2.3. Shape features determination

Geometrical shape features have been widely used for classifying objects present in agricultural imagery. Roundness and area features were used to classify corn and weeds in field imagery [32]. Shape features of digital images were used to identify and count soybean aphids [33]. Several shape features describing the geometric properties of the weed species were extracted from the binary image using the ParticleAnalyzer application programming interface (API). This API provides basic features, such as area, perimeter, feret major and minor axes, bounding box width and height, ellipse major and minor axes, along with standard shape features, such as aspect ratio, roundness, circularity, and solidity. Since the basic shape features, such as area and perimeter have dimensions involved with them, their values change with the resolution of the image and the size of the seedlings due to growth stages. Because of these limitations, these parameters cannot be used directly in classification but can be used to derive other advanced features.

A plugin using ImageJ was developed to derive 21 advanced shape features (non-dimensional) from the binary images (Table 1). Out of the 21 features, the four basic features were obtained directly from the ParticleAnalyzer outputs, while the other advanced shape features were user-coded, based on the parameters set in the ParticleAnalyzer outputs. From the basic features (Table 1), the convex area can be defined as the area of the convex hull (polygon) that wraps the object. The dimensions of major and minor axes were in pixels and corresponded to the orthogonal axes of the fitted ellipse, whose area was equal to the area of the object.

For the advanced shape features (Table 1), the bounding rectangle area is defined as the area of the smallest enclosing rectangle of the object (pixel^2), while the dimension of Feret diameter can be defined as the maximum diameter of the particle (pixel). It is to be noted that all

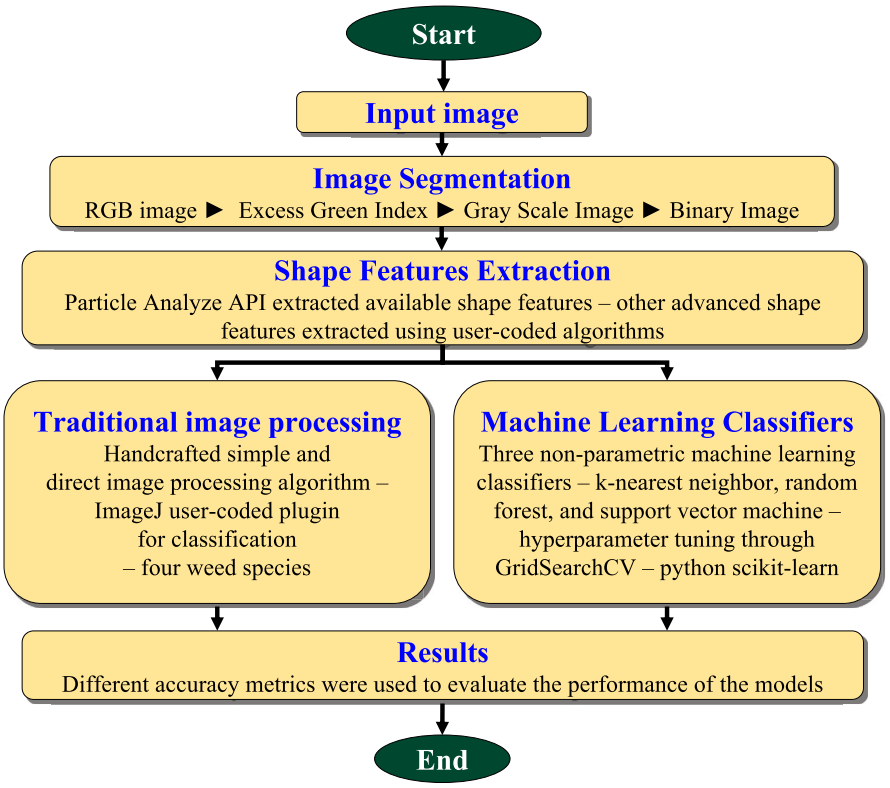


Fig. 2. Overall flowchart of distinguishing four weed species using shape features with handcrafted simple image processing approach and three machine learning algorithms.

Table 1
Basic and advanced shape factors used in the handcrafted image processing and machine learning models development for weeds classification.

Shape features	Shape features
Basic shape featured directly obtained from the ParticleAnalyzer outputs of ImageJ	
$\text{Circularity} = \frac{4\pi \times \text{Area}}{\text{Perimeter}^2}$	$\text{Aspect ratio (AR)} = \frac{\text{Major axis}}{\text{Minor axis}}$
$\text{Roundness} = \frac{4 \times \text{Area}}{\pi \times \text{Major axis}^2}$	$\text{Solidity} = \frac{\text{Area}}{\text{Convex area}}$
Advanced shape features derived from the ParticleAnalyzer outputs of ImageJ	
$\text{Convex area} = \frac{\text{Area}}{\text{Solidity}}$	$\text{Hollowness} = \frac{\text{Convex area} - \text{Area}}{\text{Convex area}}$
$\text{Reverse aspect ratio (RAR)} = \frac{1}{\text{Aspect ratio}}$	$\text{Rectangularity} = \frac{\text{Area}}{\text{Bounding rectangle area}}$
$\text{Feret major axis ratio (FMA)} = \frac{\text{Feret diameter}}{\text{Major axis}}$	$\text{Feret minor axis ratio (FMI)} = \frac{\text{Feret diameter}}{\text{Minor axis}}$
$\text{Convex area Feret ratio (CAF)} = \frac{\text{Convex area}}{\text{Feret diameter}^2}$	$\text{Compactness} = \frac{\text{Area}}{\text{Feret diameter}}$
$\text{Area length ratio (ALR)} = \frac{\text{Area}}{\text{Major axis}^2}$	$\text{Log height width ratio (LHW)} = \log\left[\frac{\text{Height}}{\text{Width}}\right]$
$\text{Elongation} = \frac{\text{Major axis} - \text{Minor axis}}{\text{Major axis} + \text{Minor axis}}$	$\text{Perimeter broadness ratio (PBR)} = \frac{\text{Perimeter}}{2 \times (\text{Width} + \text{Height})}$
$\text{Length perimeter ratio (LPR)} = \frac{\text{Major axis}}{\text{Perimeter}}$	$\text{Modified circularity (MC)} = \frac{4 \times \text{Area}}{\text{Perimeter} \times \text{Feret diameter}}$
$\text{Grum circularity (GC)} = \frac{16 \times \text{Area}^2}{4\pi \times \text{Perimeter} \times \text{Feret diameter}^3}$	$\text{Area compactness (AC)} = \frac{4 \times \text{Area}}{\pi \times \text{Feret diameter}^2}$
$\text{Equivalent perimeter ratio (EPR)} = \frac{\text{Perimeter}}{\text{Equivalent perimeter}}$;	where, $\text{Equivalent perimeter} = 2.0 \times \sqrt{\pi \times \text{Area}}$

Note: Dimensions of areas (Area, Convex area, and Bounding rectangle area) are in pixel^2 and lengths (Perimeter, Major axis, Minor axis, Feret diameter, Width, Height, and Equivalent perimeter) are in pixels.

other shape features derived are dimensionless; however, the "Convex area," having pixel^2 unit, is an intermediate parameter that was also used as a feature in the analysis.

2.4. Handcrafted image processing model

The traditional image processing method (handcrafted approach) uses a set of features (Table 1) and corresponding thresholds for classifying four weeds. The thresholds and features were decided based on density plots and spread factor (SF). The data for each of the features were plotted on the graph, along with the edge density plots, using RStudio scripts, to observe the overlap and identify thresholds for the different features. The features whose density plots overlap the least were used to classify the weeds. To objectively determine the classification, the SF (Eq. (1)) was developed and utilized.

$$\text{Spread factor (SF)} = \frac{\text{Standard deviation} \times 100}{\text{Total spread}} \quad (1)$$

The SF is calculated as a ratio between the standard deviation of the normalized means (mean of feature values) and the normalized total range (spread of feature values) of the distribution of the shape features of the four weed species considered. A lower SF indicates high overlap and higher values indicate better spread and separation of features leading to the establishment of cutoffs leading to accurate classification.

For traditional image processing, 100 weed images of each weed species were used for extracting the shape features, plotting density graph plots, and calculating the SF. An additional 20 images of each weed species were used for testing the classification accuracy of the algorithm and the corresponding accuracies were reported.

2.5. Machine learning methods

The 21 extracted shape features (Table 1) were further used to train and test the three non-parametric ML classifiers, namely, kNN, RF, and SVM, to classify four weed species. The following models were selected because they provide high classification accuracies and have been widely used in agriculture for different applications, such as weed classification [34], disease detection and classification [22], and many more applications. To employ the ML algorithm, the first step is to select the relevant features and then train the model for a particular task. The relevant features were selected using principal component analysis (PCA) and different sets were made to train and test the classifier. Hyperparameters of the classifier were tuned for different sets of data and the corresponding accuracy metrics were evaluated and reported.

2.5.1. Feature selection

Feature selection helps in reducing the dimensionality of high-dimensional patterns [35]. Feeding irrelevant features for training the algorithm decreases its accuracy, increases the computational load, and undue demand on resources. However, the relevant features are not directly known. In this regard, PCA, one of the widely used methods for image processing, pattern recognition, data compression, data mining, ML, and computer vision, is employed [36]. The PCA is a statistical method that helps to determine key features in the high dimensional dataset that explains the differences in observation and reduces the loss of information. All shape features employing PCA were ranked based on their weights. Sets of features representing multiple features in increments of three from 3 to 21 were used to train and test all ML classifiers outlined subsequently. These increments were decided to reduce the computational time and resources.

2.5.2. k-nearest neighbors classifier

The kNN is a simple and popular supervised ML method used for performing classification [37]. This algorithm is also known as a lazy or instance-based learning method because it does not have specialized learning involved other than keeping track of the labeled data. The theory behind kNN is that the algorithm finds k samples based on the distance values that are nearest to the new test samples. The label of each test sample is determined by the majority of votes of its k nearest neighbors [38]. Hence, k is the main hyperparameter in this classifier and needs to be tuned as it directly affects the model's performance.

The GridSearchCV (GSCV) API of the scikit-learn library was used [39] for selecting the optimal value of k for our dataset. Based on a scoring metric, the GSCV provides the best combination of parameters from the list supplied in the parameter grid. A list of k values ranging from 2 to 30 was tested. The choice for p-value was given as 1 or 2, where 1 means Manhattan distance and 2 means Euclidean distance. The resulting combination of parameters was used to train the model for classifying the four weed species with different subsets of features.

2.5.3. Support vector machine classifier

The SVM is a supervised ML algorithm and has been widely used to address classification problems in agriculture [18]. Rather than finding the nearest neighbors, SVM finds a hyperplane in N-dimensional space that can classify distinct classes. The hyperplane maximizes the margins and minimizes the generalization errors. A key parameter in the SVM is the kernel function that transforms the training set of data into a higher dimension space so that a non-linear decision surface could be transformed into a linear surface to facilitate the distinction of data [40]. Apart from the kernel functions, values of gamma (γ) and regularization (C) directly influence the model's accuracy. An exponentially growing sequence of γ and C values is a common approach to finding good parameter values [41]. Therefore, finding the optimal values for γ, and C is necessary. Since there were four weed species, the "one versus rest" feature of the scikit-learn SVM class was used [39].

To find the optimal values of the kernel, γ, C, and GSCV were used. Three different types of kernels, namely "linear," "polynomial," and "gaussian" kernel, were provided along with the list of exponential growing sequences (2^k), where k ranges from [−4, 4]. The resulting parameters were used for training the model with different subsets of features to classify the four weed species.

2.5.4. Random forest classifier

The basic theory behind RF is that each tree provides an outcome for the test data, and the majority vote of all the trees is the label for the test data. Since the accuracy of the model depends on the number of trees, finding the optimal value of the number of trees is necessary [21]. The number of features (max_features) in each split also needs to be investigated because it affects the performance and speed of the algorithm [38].

To obtain the optimal/best set of combinations of the hyperparameters for the RF classifier, GSCV was used. To find the number of trees best suited for classification, a list of integers from 10 to 200 was provided to GSCV. While to find the maximum number of features that each individual tree in the random forest can use for training, two standard types, namely 'sqrt' and 'log2,' were input to GSCV to find the optimal parameters for training the classifier. The parameters provided by the GSCV were used to train the model with different sets of features for performing classification tasks.

2.5.5. Machine learning model evaluation

The dataset was split into two groups, the training group comprised 80 % of the data, and the testing group the remaining 20 %. Before feeding the data into each ML model or performing the feature selection task, feature scaling was performed using scikit-learn StandardScaler API [39]. The performance of the model was evaluated by comparing the predicted labels with the actual labels. The actual labels are true positive (TP), false negative (FN), true negative (TN), and false positive (FP) values. The TP, FN, TN, and FP are illustrated in classifying common lambsquaterts with respect to another weed in Fig. 3. Each classification model was evaluated using testing accuracy (TP/[TP+FP+FN+FP]), precision (TP/[TP+FP]), recall (TP/[TP+FN]), and F1-score (2 × [precision × recall]/[precision + recall]).

3. Results and discussion

3.1. Weed identification with simple image processing handcrafted model

3.1.1. Shape features density distribution

The density distributions of various shape features as overlap plots are presented in Fig. 4. The density plots help visualize the mean (major peaks) and the separation of each weed for the selected features (9) to demonstrate their values distribution. They show clear separated peaks for at least two weeds, or lack thereof, in most cases. The distribution of a couple of low-ranked features, namely aspect ratio (Fig. 4G) showing a left-skewed trend and area compactness (Fig. 4H) showing near-normal based on the visualized distribution.

The density distribution gives clue on whether all or a couple of weeds can be separated and cutoff values determined for classification. Among the displayed shape features, other than aspect ratio and

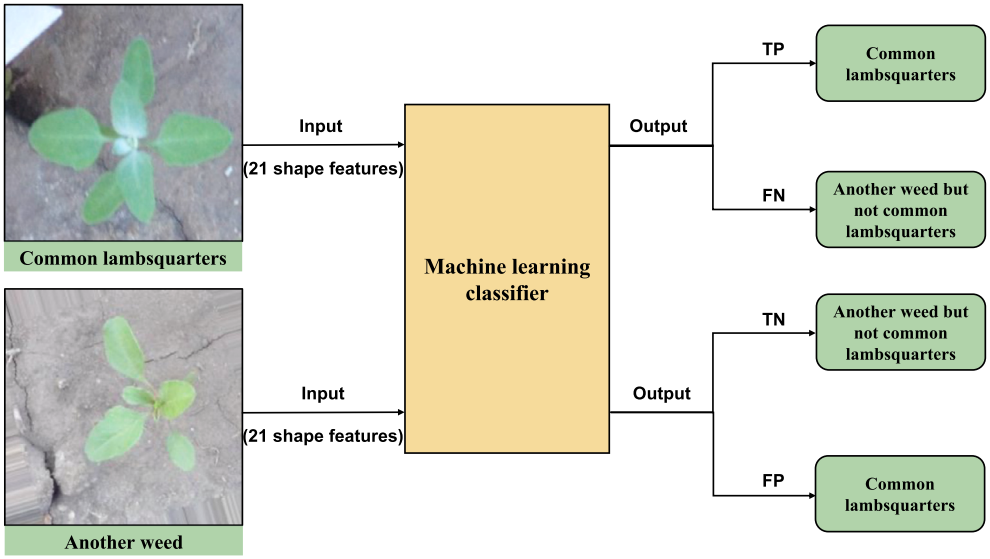


Fig. 3. Classification process involved in the machine learning models depicting four different outcomes using a target and other weed species: true positive (TP), false negative (FN), true negative (TN), and false positive (FP).

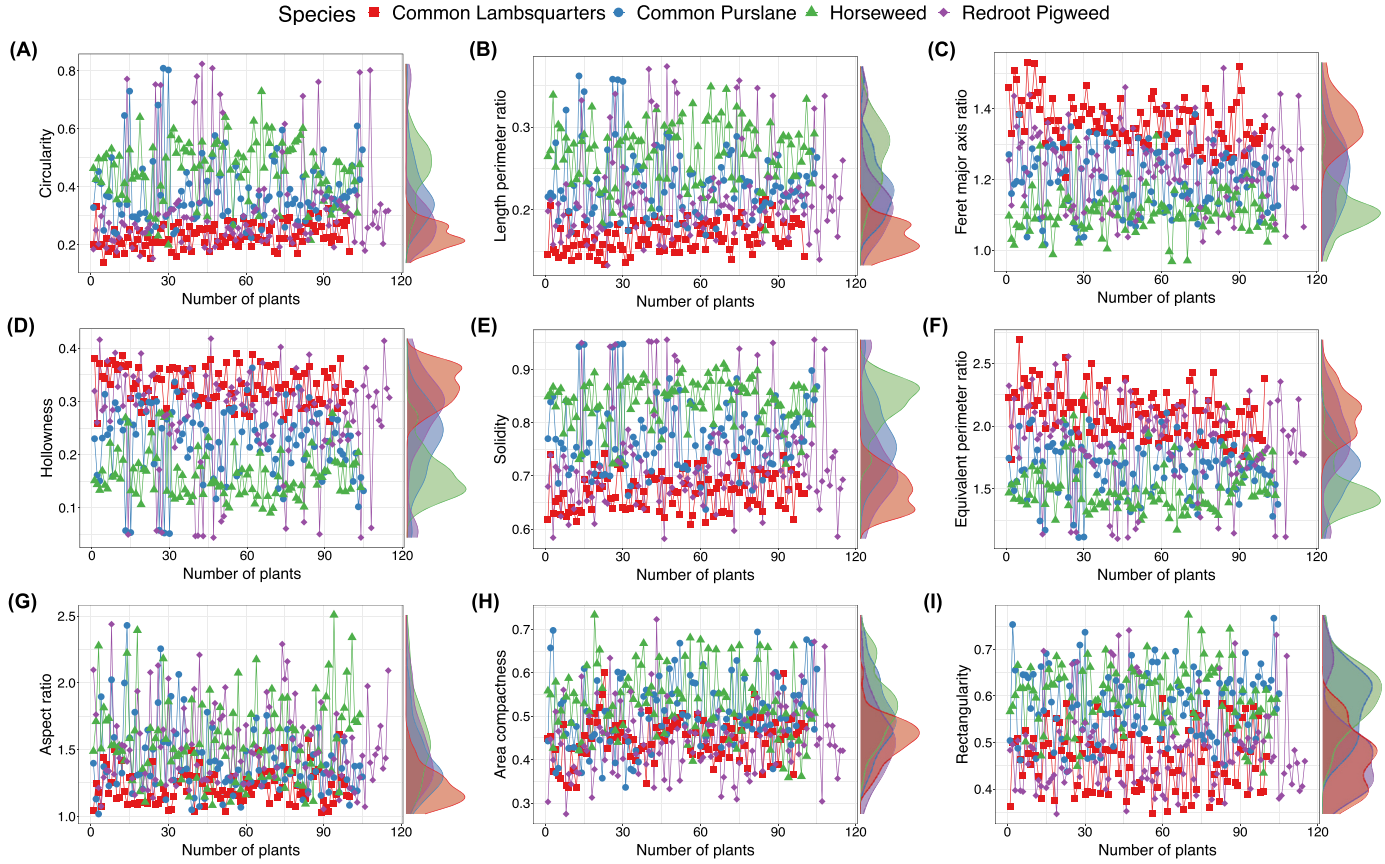


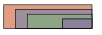
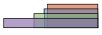
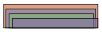
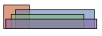
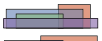
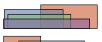
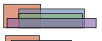
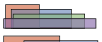
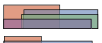
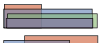
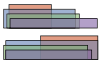
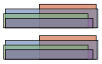
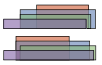
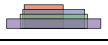
Fig. 4. Variations of selected nine shape features and their density distribution visualization that can be used to classify the four common weed species using handcrafted simple image processing. (A) Circularity, (B) Length perimeter ratio, (C) Feret major axis ratio, (D) Hollowness, (E) Solidity, (F) Equivalent perimeter ratio, (G) Aspect ratio, (H) Area compactness, and (I) Rectangularity.

area compactness the rest shows distinct peaks (Fig. 4). For example, the peaks of common lambsquarters (red) and horseweed (green) are clearly separated in all the selected features except for aspect ratio and area compactness; however, common purslane and redroot pigweed mostly displayed overlaps among the features in general, with rectangularity (Fig. 4I) these can be separated if lambsquarters and horseweed are extracted out, which might also involve further processing.

3.1.2. *Shape features overlap comparison and SF*

Results of the developed SF are presented in Table 2. Test cases with known distribution were also developed to evaluate the SF. The developed SF properly ranks the overlaps of the different species. The test

Table 2
Extracted shape features value ranges, and overlap of values and features overlap ranking for the four common weed species of North Dakota in handcrafted simple image processing model (training set; $n = 100$).

Shape features	Range of features				Overlap bars †	Total spread	SD	Spread factor	Rank
	CL	CP	HW	RP					
Test - No overlap *	0.14–0.20	0.25–0.33	0.43–0.54	0.67–0.82		0.59	0.37	62.79	1
Test - Just touching *	0.01–0.25	0.25–0.50	0.50–0.75	0.75–0.99		1.00	0.33	32.54	2
Test - Overlap lower end *	0.12–0.27	0.14–0.27	0.16–0.27	0.22–0.27		2.93	0.14	4.91	11
Test - Overlap higher end *	0.82–0.97	0.81–0.97	0.78–0.97	0.69–0.97		2.79	0.11	3.79	16
Test - Overlap lower-half *	0.02–0.49	0.04–0.49	0.08–0.49	0.12–0.49		3.62	0.05	1.30	26
Test - Overlap higher-half *	0.65–0.98	0.63–0.98	0.58–0.98	0.53–0.98		3.40	0.06	1.76	25
Test - Overlap most *	0.10–0.98	0.12–0.96	0.15–0.96	0.18–0.98		3.78	0.02	0.57	27
Circularity	0.14–0.33	0.23–0.81	0.20–0.73	0.15–0.82		2.90	0.19	6.57	3 (1)
Compactness	0.18–0.42	0.29–1.03	0.25–0.93	0.20–1.05		2.88	0.19	6.57	4 (2)
Length perimeter ratio	0.13–0.21	0.17–0.36	0.17–0.35	0.13–0.37		2.88	0.19	6.46	5 (3)
Feret major axis ratio	1.21–1.53	1.02–1.38	0.97–1.32	1.02–1.52		2.73	0.17	6.34	6 (4)
Solidity	0.61–0.74	0.64–0.95	0.72–0.91	0.58–0.96		2.66	0.16	6.14	7 (5)
Hollowness	0.26–0.39	0.05–0.36	0.09–0.28	0.04–0.42		2.66	0.16	6.14	8 (6)
Equivalent perimeter ratio	1.73–2.69	1.11–2.11	1.17–2.24	1.10–2.56		2.82	0.17	5.87	9 (7)
Modified circularity	0.22–0.44	0.31–0.70	0.31–0.69	0.24–0.77		2.76	0.16	5.70	10 (8)
Grum circularity	0.08–0.26	0.11-0.43	0.12–0.50	0.07–0.56		2.80	0.14	4.91	12 (9)
Rectangularity	0.35–0.60	0.44–0.77	0.43–0.77	0.35–0.74		3.12	0.14	4.63	13 (10)
Aspect ratio	1.03–1.62	1.02–2.43	1.08–2.51	1.06–2.44		3.23	0.15	4.53	14 (11)
Perimeter broadness ratio	1.02–1.72	0.82–1.45	0.82–1.44	0.82–1.50		2.92	0.13	4.37	15 (12)
Elongation	0.01–0.24	0.01–0.42	0.04–0.43	0.03–0.42		3.38	0.12	3.57	17 (13)
Feret minor axis ratio	1.31–2.11	1.21–2.56	1.27–2.64	1.33–2.98		2.92	0.10	3.56	18 (14)
Area length ratio	0.49–0.77	0.32–0.77	0.31–0.72	0.32–0.74		3.39	0.11	3.30	19 (15)
Roundness	0.62–0.97	0.41–0.98	0.40–0.92	0.41–0.95		3.41	0.10	3.03	20 (16)
Reverse aspect ratio	0.62–0.97	0.41–0.98	0.40–0.92	0.41–0.95		3.41	0.10	3.03	20 (16)
Convex area Feret ratio	0.42–0.64	0.35–0.67	0.35–0.65	0.28–0.63		3.05	0.08	2.67	22 (18)
Area compactness	0.34–0.60	0.34-0.70	0.36–0.73	0.28–0.72		3.18	0.07	2.22	23 (19)
Log height width ratio	−0.15–0.14	−0.16–0.30	−0.18–0.32	−0.28–0.42		2.79	0.05	1.92	24 (20)

Note: CL - common lambsquarters: ■, CP - common purslane: ■, HW - horsweed: ■, and RP - redroot pigweed: ■
Total spread - sum of the normalized range of features with its total width (maximum-mininum) among species, SD - standard deviation of the normalized mid-values of the range of features, Spread factor = (SD / Total spread) × 100, Rank - based on the largest SF and indicates both for test and shape features while that in parenthesis indicates only for shape features (See supplementary material).
* - arbitrary test cases to study overlaps, † - overlap bars stretch to the selected drawing width based on the values of the range of features.

cases illustrate the efficiency of the SF in correctly ranking a distribution with separation (No overlap) than that with ranges without gaps (Just touching) while evaluating different values based on the overlapping ranges of the weeds.

The overlap of the shape feature ranges across different weeds was also visualized readily using overlapped bar plots called “overlap bars” (Table 2). The overlap bars are compact visual aids that were developed from the limits of the ranges. Based on this analysis, the various shape features were evaluated along with known test cases and ranked in ascending order. Test cases of “No overlap” with clear separation of bars among species had the highest SF (62.79; overall ranked 1), with “Just touching” test case having the second highest (32.54; overall ranked 2), while the test case of “Overlap most” with an almost total overlap of coinciding ranges had the lowest SF (0.57; overall ranked 27), and other test and actual data falling in between.

For all the shape features of the four weed species, the overlap bars show overlaps of different degrees and show no clear separations in test cases (No overlap or even Just touching; Table 2). It should be noted that determining cutoff values based on the SF, as a direct method, will serve as the best guess and will produce a classification accuracy proportional to the overlap. The top-ranked, among the shape features, from circularity (rank 1) to modified circularity (rank 8) the SF varied closely (5.70–6.57). The results show that circularity, compactness, and length perimeter ratio were the top three features that can be used for classification. Area compactness and log height-width ratio (ranked 19 and 20) have the lowest SF, due to the lowest standard deviation with the common values of total spread, and therefore, cannot be used for the classification.

It is to be noted that Feret major axis ratio and equivalent perimeter ratio were used to classify common lambsquarters from horseweed, common purslane, and redroot pigweed. Hollowness, which ranks 6 in the Table 2 can be used to classify horseweed from all three weed species. Hollowness and circularity were used to classify common purslane from all other weed species. In addition, rectangularity and length perimeter ratio could be used to classify redroot pigweed from horseweed, common purslane, and common lambsquarters. These features were used to develop a handcrafted model that can be used to classify the four weed species.

Since circularity and compactness are linearly correlated their distributions were similar (Table 2) and were ranked the closer (1 and 2) with closer SF values (6.574 and 6.573). It was seen that though common lambsquarters and horseweed were easily classifiable, most features could not able to separate the common purslane and redroot pigweed. However, rectangularity (ranked 10) could be used to separate them as it shows promising trends, once the common lambsquarters and horseweed data were filtered out through further image processing

as they had some overlaps. This shows that specific shape features will be successful in classifying certain weed species.

Based on the analysis, the developed SF can be used as a parameter to compare the performance of the shape features in classifying the weed species or can be utilized in similar overlap analysis of various features in other applications. It can be observed that a combination of distribution visualization, overlap bars, and SF can be used to visualize, evaluate, and rank continuously varying and overlapping parameters.

3.1.3. Shape features direct approach handcrafted image processing model validation

Twenty images of each weed species were selected to evaluate the performance of the developed model based on the direct approach. The handcrafted model algorithm efficiently classified horseweed and redroot pigweed with an accuracy of 95 % and 80 %, respectively. To correctly identify the horseweed, the hollowness threshold value should be <0.2. For redroot pigweed, the rectangularity threshold should be <0.52 and LPR be >0.16, as well as, circularity be <0.25. Out of the 20 redroot pigweed images, 16 were correctly identified, while two were classified as common lambsquarters and common purslane. In the case of horseweed, the algorithm just misidentified one image as redroot pigweed.

The algorithm insufficiently identified common purslane and common lambsquarters correctly. For common lambsquarters, the FMA value and EPR ratio should be ≥1.2 and 1.82, respectively and LPR be <0.16. To correctly identify common purslane, the FMA, EPR, hollowness, and circularity, should be ≥1.2, 1.82, 0.2, and 0.25, respectively, while the rectangularity threshold should be >0.52. With these conditions, 10 out of 20 common purslanes and 12 out of 20 common lambsquarters were correctly identified, resulting in the accuracy of 50 % and 60 %, respectively. The overall accuracy of the algorithm was 71.25 %. These classification accuracies are quite similar to that found by Søgaard (2005) [42] in classifying three weed species using active shape models. It is to be noted when classifying weeds using shape features, the accuracy of the model classifying the weeds decreases if the resemblance between them is higher [42]. In this present research, seven out of twenty common lambsquarters were identified as redroot pigweed, similarly, seven out of twenty common purslanes were identified as horseweed because of the shape resemblance as a whole plant, although they all have different leaf structures.

Although this direct approach of the handcrafted model using geometrical features is simple, obviously shape resemblance will reduce the classification accuracy, but can be easily incorporated on the mobile application because of the lower computational and time requirements. The classification accuracy of the model was lower because the images used in creating the handcrafted classification model has weed images of different growth stages, which created a some randomness in the feature values. It is expected that deploying these shape features at the same growth stages could increase the classification accuracy of the model. However, to gain more accuracy in classifying the weeds, advanced techniques such as ML should be tested.

3.2. Performance of machine learning models

All the ML classifier models were trained and their hyperparameters were tuned with the different subsets of data, as per the ranks shown in Table 3.

3.2.1. Feature selection

The PCA feature score and the corresponding rank of the 21 features were presented in Table 3. It was found that the results from direct shape features (Table 2) and PCA (Table 3) were relatively similar. Features such as hollowness, FMA, circularity, EPR, and some other features were in the top 10 ranks. Features that were found important in the simple image processing method (Table 2), though the analysis was performed with a smaller dataset of 100 images, relatively coincides with PCA with a dataset of 1000 images.

Table 3
Feature importance score and corresponding rank using principal component analysis of the shape features.

Rank	Features	Feature score × 1000
1	Hollowness	57.0
2	Solidity	56.9
3	Convex area	56.8
4	Feret major axis ratio	43.9
5	Circularity	39.3
6	Compactness	39.3
7	Equivalent perimeter ratio	34.2
8	Modified circularity	33.7
9	Length perimeter ratio	33.0
10	Roundness	32.3
11	Reverse aspect ratio	32.3
12	Area length ratio	32.3
13	Perimeter broadness ratio	30.2
14	Area compactness	28.8
15	Grum circularity	27.8
16	Rectangularity	27.5
17	Convex area feret ratio	26.9
18	Elongation	25.7
19	Log height width ratio	17.5
20	Aspect ratio	10.2
21	Feret minor axis ratio	9.3

3.2.2. k-nearest neighbor classifier

The testing accuracy of the kNN model was the lowest (71.23 %) when just the top three features (Table 3) were used to train the model. The value of k used for training the model was 12, provided by the GSCV class of scikit-learn using three-fold cross-validation. The mean for precision was 0.7225, and recall and F1-score were 0.7175, while the ten-fold cross-validation accuracy of the model was 73.81 %. In contrast, the kNN testing accuracy was maximum (89.55 %), when the top 15 features were used as an input to train the model. The value of hyperparameter k for this model was 3, which was provided by GSCV using the three-fold cross-validation. The average precision and recall of the model were 0.8975, and the F1-score was 0.90. However, the ten-fold cross-validation accuracy of the model was 88.42 %.

3.2.3. Support vector machine classifier

The value of kernel type, γ, and regularization or error bound parameter C were determined through three-fold cross-validation by GSCV. The testing accuracy of the model was minimum (72.39 %) when the model was trained with the top three features (Table 3). The hyperparameter values for the model, namely for γ and C were 16, and the kernel type was Gaussian. The mean value of precision, recall, and F1-score was 0.7375, 0.725, and 0.7325. Furthermore, the ten-fold cross-validation accuracy was 73.06 %.

Although the testing accuracy of the SVM model was maximum (88.74 %) when trained with the first 18 features (Table 3), the model that was trained with all 21 shape features even though it had low testing accuracy of 88.52 %. The model trained with 21 features was selected over the model with 18 features because it has a high F1-score. F1-score was considered over the testing accuracy, since testing accuracy fails to evaluate the performance of ML models when trained with an unbalanced dataset [43]. The optimal parameter values were kernel type as Gaussian, γ as 2^3 (8), and C as 2^2 (4). By using the kernel type as Gaussian and optimal γ and C values, the ten-fold cross-validation accuracy was 89.34 %. While the testing accuracy of the model was 88.52 %. The means for precision, recall, and F1-score were 0.898, 0.895, and 0.898, respectively.

3.2.4. Random forest classifier

The testing accuracy was the minimum (73.89 %) when the model was trained with the top three features (Table 3). The number of trees

Table 4
Performance of the three machine learning models on four different common weed species of North Dakota.

Number of features	Class	k-nearest neighbor			Support vector machine			Random forest		
		Precision	Recall	F1-score	Precision	Recall	F1-score	Precision	Recall	F1-score
3	1	0.82	0.77	0.79	0.85	0.73	0.79	0.86	0.74	0.79
	2	0.63	0.68	0.65	0.66	0.67	0.67	0.68	0.69	0.69
	3	0.80	0.75	0.77	0.80	0.78	0.79	0.80	0.80	0.80
	4	0.64	0.67	0.66	0.64	0.72	0.68	0.66	0.73	0.69
	Average	0.723	0.718	0.718	0.738	0.725	0.733	0.750	0.740	0.743
6	1	0.89	0.89	0.89	0.89	0.84	0.86	0.93	0.93	0.93
	2	0.82	0.83	0.82	0.78	0.79	0.79	0.85	0.83	0.84
	3	0.89	0.89	0.89	0.85	0.88	0.87	0.87	0.91	0.89
	4	0.82	0.82	0.82	0.76	0.78	0.77	0.85	0.83	0.84
	Average	0.855	0.858	0.855	0.820	0.823	0.823	0.875	0.875	0.875
9	1	0.89	0.91	0.90	0.90	0.88	0.89	0.95	0.93	0.94
	2	0.85	0.85	0.85	0.86	0.84	0.85	0.88	0.88	0.88
	3	0.93	0.91	0.92	0.91	0.95	0.93	0.90	0.93	0.92
	4	0.84	0.84	0.84	0.82	0.82	0.82	0.88	0.88	0.88
	Average	0.878	0.878	0.878	0.873	0.873	0.873	0.903*	0.905*	0.905*
12	1	0.92	0.90	0.91	0.90	0.87	0.89	0.92	0.91	0.91
	2	0.81	0.87	0.84	0.87	0.90	0.89	0.85	0.89	0.87
	3	0.94	0.91	0.92	0.95	0.95	0.95	0.90	0.93	0.92
	4	0.86	0.84	0.85	0.84	0.84	0.84	0.89	0.84	0.86
	Average	0.882	0.88	0.88	0.89	0.89	0.892	0.89	0.89	0.89
15	1	0.92	0.90	0.91	0.92	0.87	0.89	0.91	0.90	0.91
	2	0.85	0.90	0.88	0.86	0.88	0.87	0.87	0.88	0.87
	3	0.95	0.92	0.94	0.93	0.95	0.94	0.91	0.93	0.92
	4	0.87	0.87	0.87	0.85	0.84	0.85	0.88	0.86	0.87
	Average	0.898†	0.898†	0.900†	0.890	0.885	0.888	0.893	0.893	0.893
18	1	0.93	0.87	0.90	0.92	0.87	0.89	0.92	0.90	0.91
	2	0.83	0.89	0.86	0.85	0.89	0.87	0.87	0.90	0.88
	3	0.94	0.92	0.93	0.94	0.95	0.94	0.93	0.93	0.93
	4	0.86	0.87	0.87	0.85	0.84	0.85	0.88	0.86	0.87
	Average	0.890	0.888	0.890	0.890	0.888	0.888	0.900	0.898	0.898
21	1	0.83	0.89	0.86	0.92	0.90	0.91	0.90	0.93	0.91
	2	0.87	0.88	0.87	0.86	0.92	0.89	0.91	0.90	0.90
	3	0.92	0.93	0.92	0.94	0.90	0.92	0.91	0.93	0.92
	4	0.88	0.82	0.85	0.87	0.86	0.87	0.91	0.88	0.89
	Average	0.875	0.880	0.875	0.898‡	0.895‡	0.898‡	0.908	0.910	0.905

Note: Classes 1, 2, 3, and 4 refer to common lambsquarters, common purslane, horseweed, and redroot pigweed, respectively. The selected averages with ∗, †, and ‡ show the best performance of the different machine learning classifiers.

and max_features were set to 87 and square root, respectively. The bootstrap setting was intentionally set to 'true' to increase the model performance on the unseen dataset. The mean precision, recall, and F1-score of the model were 0.75, 0.74, and 0.7425. The ten-fold cross-validation accuracy of the model was 75.37 %.

The testing accuracy of the random model classifier was maximum (90.72 %) when trained with all the 21 shape features, while the testing accuracy of the classifier was 90.25 % when trained with the top nine features (Table 3). In both cases, the mean F1-score of the classifier was 0.905. Due to the fewer features involved in training the classifier, the second model was used even though it had a bit lower testing accuracy. The lower number of features reduces the computational resources and time required to train the model. The mean precision and recall of the model were 0.902 and 0.905, respectively. However, the ten-fold cross-validation accuracy of the model was 91.06 %. The number of trees and max_features used to train the classifier was 88 and log2, respectively.

3.2.5. Performance evaluation of ML models

Classification performance results of the three non-parametric ML models (kNN, RF, and SVM) in classifying four weed species are presented in Table 4. The classification accuracy, precision, recall, and F1-score of all three ML models were the minimum when trained with the top three features, due to the fact that ML models require a large dataset to train. Further, increasing the number of features to train the model increased the classification accuracy, precision, recall, and F1-score.

All three ML models had comparable performance, with both average precision and recall in the range of 0.898 to 0.905. The performance of the ML classifiers can also be easily compared using the F1-score, which indicates the classifier's robustness and accuracy. From the plotted bar graph (Fig. 5), the f1-score comparison indicates clearly that RF is the best classifier followed by SVM and kNN.

Out of the different ML models, RF performed best in the present study, followed by SVM and kNN. As observed in the present study, Ashram and Khan (2020) [44] found that RF outperformed SVM in classifying images based on grassy weed densities into three categories. Also, while detecting weeds in a chilli field from unmanned aerial vehicle (UAV) imagery using RF, kNN, and SVM found that RF outperformed the other classifiers, while kNN performed the worst [45]. The RF has high-performance accuracy because the RF classifiers are trained with different sub-samples and subsets of datasets resulting in high generalization and performance on an unseen dataset [46]. While classifying volunteer barley in a field of oilseed rape, SVM performed better than kNN because SVM has complex function classes such as radial functions that can help construct non-linear separation functions [47]. Also, the RF classifier had an accuracy of 95.4 % in classifying common lambsquarters, bindweeds (*Convolvulus arvensis*), and crabgrass (*Digitaria sanguinalis*) and maize (*Zea mays*) in the early season [48], while

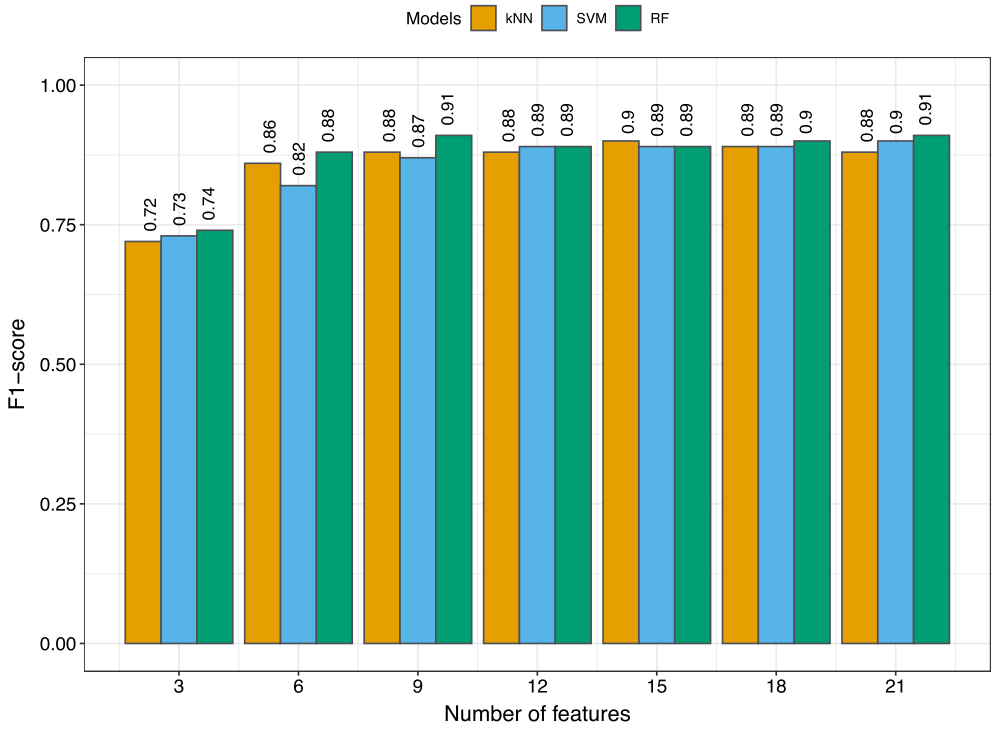


Fig. 5. Comparison of machine learning models based on F1-score. kNN - k-nearest neighbor, RF - random forest, and SVM - support vector machine.

the SVM classifier differentiated maize from weeds with accuracy ranging from 90.19 % to 93.87 % for three years [7].

Overall, the ML models have performed better than the handcrafted direct simple image processing algorithms. The handcrafted image processing algorithms were accurate in identifying two out of four species tested (horseweed and common lambsquarters: 95 %–100 %). While this direct approach was not efficient, with the methodology of shape factors tested, they were simple and direct and could be improved with more samples and other algorithms of shape and texture identification. It should also be noted that all the derived shape factors are also served as intermediate inputs to the advanced ML models. In addition, ML models are more complex than the handcrafted models — which are simpler and do not require rigorous training and testing cycle with cross-validation as followed with ML. It is recommended that if the classification problem supports the use of handcrafted simple image processing algorithms, it is better to resort to them; otherwise, the versatile ML modeling approach should be followed to achieve high accuracy.

3.3. Research limitations and future research directions

In this research, images of four commonly found weed species in North Dakota were used to train the ML models and develop a handcrafted simple image processing model for weed classification. Possible application of the advanced deep learning model was avoided due to the limited dataset of images. Analysis revealed that the handcrafted or the direct and simple image processing model had relatively reduced classification performance because the shape feature threshold values are selected based on the limited dataset. Therefore, the model will require some fine-tuning once new images are added. In this study, both approaches did not consider the overlapping of weeds in the imagery.

In the future, more weed species should be considered, and a large weed imagery dataset should be developed and tested with ML and deep learning models. As the weeds were grown in plastic trays of field soil, the methodology is expected to be applicable in the field, especially during the initial stages; however, field testing should be performed for it is expected to present several challenges (e.g., image background variation, other weeds, lack of dimensions reference). It will be interesting to test the current methodology with field data to check the performance of ML and handcrafted simple image processing models. The usage of shape and textural features in conjunction with image processing and ML models should be explored. These shape features can also be used to distinguish crops from weeds and can help in generating a weed distribution map, which can reduce the use of herbicides. This aspect should be explored further.

4. Conclusion

The developed methodologies that used shape features to classify four different weed species (common lambsquarters, common purslane, horseweed, and redroot pigweed) commonly found in the corn fields using handcrafted simple image processing algorithms and three non-parametric machine learning (ML) models (kNN, SVM, and RF) were successful. The handcrafted simple image processing approach was highly successful with a few weed species in distinguishing each other (common lambsquarters *vs* horseweed; accuracy ≥95 %), while did not perform well with other species combinations (common lambsquarters *vs* redroot pigweed, and horseweed *vs* common purslane; accuracy 25 %–40 %) and classifying them collectively owing to the similarity of the shapes. However, all the ML models resulted in high precision, recall, F1-score, and testing accuracies in classifying all the weed species considered. Out of the different ML models, RF performed best (F1: 0.88–0.94) with nine features, followed by SVM (F1: 0.87–0.91) with 21 features, and kNN (F1: 0.87–0.94) with 15 features.

It is recommended that the handcrafted simple image processing algorithms should be tried first for identification and classification studies, as they are quick and efficient, before resorting to advanced, complex, and versatile ML modeling approaches. Future research should include testing these models in field conditions, increasing the size of the weed imagery dataset, and employing and comparing the deep learning models along with other ML models. The usage of shape and textural features in conjunction with the image processing and ML models that can help prepare weed distribution maps should also be explored.

CRediT authorship contribution statement

H. Pathak: Conceptualization, Methodology, Investigation, Visualization, Writing - original draft, Writing - review & editing. **C. Igathinathane**: Conceptualization, Funding acquisition, Methodology, Project administration, Supervision, Visualization, Writing - original draft, Writing - review & editing. **K. Howatt**: Methodology, Writing - review & editing. **Z. Zhang**: Methodology, Writing - review & editing.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Data used in the study and the data analysis and visualization codes were uploaded to "Mendeley Data" with the citation: Cannayen, Igathinathane; Pathak, Harsh; Howatt, Kirk (2023), "Machine Learning and Handcrafted Image Processing Weed Classification Results Data," Mendeley Data, V3, https://doi.org/10.17632/787mkh2kjc.3.

Acknowledgements

This work was supported by the North Dakota Corn Council, Fund: FAR0030691, and in part by the USDA National Institute of Food and Agriculture, Hatch Project: ND01481, Accession: 1014700. The facilities were provided by USDA-ARS Northern Great Plains Research Laboratory (NGPRL), Mandan, ND. These supports are gratefully acknowledged.

Appendix A. Supplementary material

Supplementary material related to this article can be found online at https://doi.org/10.1016/j.atech.2023.100249.

References

[1] N. Soltani, J.A. Dille, I.C. Burke, W.J. Everman, M.J. VanGessel, V.M. Davis, P.H. Sikkema, Perspectives on potential soybean yield losses from weeds in North America, Weed Technol. 31 (1) (2017) 148–154, https://doi.org/10.1017/wet.2016.2.

[2] N. Soltani, J.A. Dille, I.C. Burke, W.J. Everman, M.J. VanGessel, V.M. Davis, P.H. Sikkema, Potential corn yield losses from weeds in North America, Weed Technol. 30 (4) (2016) 979–984, https://doi.org/10.1614/WT-D-16-00046.1.

[3] C. Pulido, L. Solaque, N. Velasco, Weed recognition by SVM texture feature classification in outdoor vegetable crop images, Ing. Investig. 37 (1) (2017) 68–74, https://doi.org/10.15446/ing.investig.v37n1.54703.

[4] C. Lin, A support vector machine embedded weed identification system, Master's thesis, University of Illinois at Urbana-Champaign, 2010, http://hdl.handle.net/2142/14615.

[5] C. Swanton, What plants sense and say may impact the future of weed control, http://wssa.net/2016/08/what-plants-sense-and-say-may-impact-the-future-of-weed-control, August 10, 2016.

[6] M. Dadashzadeh, Y. Abbaspour-Gilandeh, T. Mesri-Gundoshmian, S. Sabzi, J.L. Hernández-Hernández, M. Hernández-Hernández, J.I. Arribas, Weed classification for site-specific weed management using an automated stereo computer-vision machine-learning system in rice fields, Plants 9 (5) (2020) 559, https://doi.org/10.3390/plants9050559.

[7] Y. Zheng, Q. Zhu, M. Huang, Y. Guo, J. Qin, Maize and weed classification using color indices with support vector data description in outdoor fields, Comput. Electron. Agric. 141 (2017) 215–222, https://doi.org/10.1016/j.compag.2017.07.028.

[8] M. Louargant, G. Jones, R. Faroux, J.-N. Paoli, T. Maillot, C. Gée, S. Villette, Unsupervised classification algorithm for early weed detection in row-crops by combining spatial and spectral information, Remote Sens. 10 (5) (2018) 761, https://doi.org/10.3390/rs10050761.

[9] I. Heap, International herbicide-resistant weed database, http://weedscience.org/Pages/Country.aspx, February 2019.

[10] R.L. Zimdahl, Chapter 11 - Weed management in organic farming systems, in: R.L. Zimdahl (Ed.), Fundamentals of Weed Science, fifth edition, Academic Press, 2018, pp. 337–357.

[11] A. Scavo, G. Mauromicale, Integrated weed management in herbaceous field crops, Agronomy 10 (4) (2020) 466, https://doi.org/10.3390/agronomy10040466.

[12] B. Liu, R. Bruch, Weed detection for selective spraying: a review, Curr. Robot. Rep. 1 (1) (2020) 19–26, https://doi.org/10.1007/s43154-020-00001-w.

[13] A. Wang, W. Zhang, X. Wei, A review on weed detection using ground-based machine vision and image processing techniques, Comput. Electron. Agric. 158 (2019) 226–240, https://doi.org/10.1016/j.compag.2019.02.005.

[14] N. Rai, Y. Zhang, B.G. Ram, L. Schumacher, R.K. Yellavajjala, S. Bajwa, X. Sun, Applications of deep learning in precision weed management: a review, Comput. Electron. Agric. 206 (2023) 107698, https://doi.org/10.1016/j.compag.2023.107698.

[15] H. Pathak, Machine vision methods for evaluating plant stand count and weed classification using open-source platforms, Master's thesis, North Dakota State University, 2021.

[16] R. Sapkota, J. Stenger, M. Ostlie, P. Flores, Towards reducing chemical usage for weed control in agriculture using UAS imagery analysis and computer vision techniques, Sci Rep 13 (2023) 6548, https://doi.org/10.1038/s41598-023-33042-0.

[17] D. Suresh Babu, Plant-stand count and weed identification mapping using unmanned aerial vehicle images, Master's thesis, North Dakota State University, 2018.

[18] A. Bakhshipour, A. Jafari, Evaluation of support vector machine and artificial neural networks in weed detection using shape features, Comput. Electron. Agric. 145 (2018) 153–160, https://doi.org/10.1016/j.compag.2017.12.032.

[19] G. Khurana, N.K. Bawa, Weed detection approach using feature extraction and KNN classification, in: Advances in Electromechanical Technologies, Springer, 2021, pp. 671–679.

[20] L. Breiman, Random forests, Mach. Learn. 45 (1) (2001) 5–32, https://doi.org/10.1023/A:1010933404324.

[21] P. Flores, Z. Zhang, C. Igathinathane, M. Jithin, D. Naik, J. Stenger, J. Ransom, R. Kiran, Distinguishing seedling volunteer corn from soybean through greenhouse color, color-infrared, and fused images using machine and deep learning, Ind. Crop. Prod. 161 (2021) 113223, https://doi.org/10.1016/j.indcrop.2020.113223.

[22] B.J. Samajpati, S.D. Degadwala, Hybrid approach for apple fruit diseases detection and classification using random forest classifier, in: 2016 International Conference on Communication and Signal Processing (ICCSP), IEEE, 2016, pp. 1015–1019.

[23] A.I. De Castro, J. Torres-Sánchez, J.M. Peña, F.M. Jiménez-Brenes, O. Csillik, F. López-Granados, An automatic random forest-OBIA algorithm for early weed mapping between and within crop rows using UAV imagery, Remote Sens. 10 (2) (2018) 285, https://doi.org/10.3390/rs10020285.

[24] M. Belgiu, L. Drăguţ, Random forest in remote sensing: a review of applications and future directions, ISPRS J. Photogramm. Remote Sens. 114 (2016) 24–31, https://doi.org/10.1016/j.isprsjprs.2016.01.011.

[25] N. DeLay, Farm data usage in commercial agriculture, https://ag.purdue.edu/commercialag/home/resource/2020/01/farm-data-usage-in-commercial-agriculture/, 2020.

[26] C.A. Schneider, W.S. Rasband, K.W. Eliceiri, NIH image to ImageJ: 25 years of image analysis, Nat. Methods 9 (7) (2012) 671–675, https://doi.org/10.1038/nmeth.2089.

[27] A. Eide, Y. Zhang, C. Koparan, J. Stenger, M. Ostlie, K. Howatt, S. Bajwa, X. Sun, Image based thermal sensing for glyphosate resistant weed identification in greenhouse conditions, Comput. Electron. Agric. 188 (2021) 106348, https://doi.org/10.1016/j.compag.2021.106348.

[28] D.A. Jamaica-Tenjo, A.E. Puerto-Lara, J.J. Guerrero-Aldana, O.L. García-Navarrete, G.A. Ligarreto-Moreno, Use of multispectral images to evaluate the efficacy of pre-emergent herbicides in peas under greenhouse conditions, Rev. Colomb. Cienc. Hortic. 15 (2) (2021) e11638, http://orcid.org/0000-0003-4331-8716.

[29] E. Hamuda, M. Glavin, E. Jones, A survey of image processing techniques for plant extraction and segmentation in the field, Comput. Electron. Agric. 125 (2016) 184–199, https://doi.org/10.1016/j.compag.2016.04.024.

[30] D.M. Woebbecke, G.E. Meyer, K. Von Bargen, D.A. Mortensen, Color indices for weed identification under various soil, residue, and lighting conditions, Trans. ASAE 38 (1) (1995) 259–269, https://doi.org/10.13031/2013.27838.

[31] M. Basso, E.P. de Freitas, A UAV guidance system using crop row detection and line follower algorithms, J. Intell. Robot. Syst. 97 (3) (2020) 605–621, https://doi.org/10.1007/s10846-019-01006-0.

[32] D.S. Shrestha, B.L. Steward, Shape and size analysis of corn plant canopies for plant population and spacing sensing, Appl. Eng. Agric. 21 (2) (2005) 295–303, https://doi.org/10.13031/2013.18144.

[33] S. Shajahan, S. Sivarajan, M. Maharlooei, S.G. Bajwa, J.P. Harmon, J.F. Nowatzki, C. Igathinathane, Identification and counting of soybean aphids from digital images using shape classification, Trans. ASABE 60 (5) (2017) 1467–1477, https://doi.org/10.13031/trans.12105.

[34] C.A.P. Rojas, L.E.S. Guzman, N.F.V. Toledo, A comparative analysis of weed images classification approaches in vegetables crops, Eng. J. 21 (2) (2017) 81–98, https://doi.org/10.4186/ej.2017.21.2.81.

[35] H. Pathak, C. Igathinathane, Z. Zhang, D. Archer, J. Hendrickson, A review of unmanned aerial vehicle-based methods for plant stand count evaluation in row crops, Comput. Electron. Agric. 198 (2022) 107064, https://doi.org/10.1016/j.compag.2022.107064.

[36] X. Zhang, L. He, J. Zhang, M.D. Whiting, M. Karkee, Q. Zhang, Determination of key canopy parameters for mass mechanical apple harvesting using supervised machine learning and principal component analysis (PCA), Biosyst. Eng. 193 (2020) 247–263, https://doi.org/10.1016/j.biosystemseng.2020.03.006.

[37] Y. Akbulut, A. Sengur, Y. Guo, F. Smarandache, NS-k-NN: neutrosophic set-based k-nearest neighbors classifier, Symmetry 9 (9) (2017) 179, https://doi.org/10.3390/sym9090179.

[38] P. Thanh Noi, M. Kappas, Comparison of random forest, k-nearest neighbor, and support vector machine classifiers for land cover classification using Sentinel-2 imagery, Sensors 18 (1) (2018) 18, https://doi.org/10.3390/s18010018.

[39] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, E. Duchesnay, Scikit-learn: machine learning in Python, J. Mach. Learn. Res. 12 (2011) 2825–2830.

[40] Z. Zhang, P. Flores, C. Igathinathane, D.L. Naik, R. Kiran, J.K. Ransom, Wheat lodging detection from UAS imagery using machine learning algorithms, Remote Sens. 12 (11) (2020) 1838, https://doi.org/10.3390/rs12111838.

[41] C.W. Hsu, C.C. Chang, C.J. Lin, A practical guide to support vector classification, Taipei, Taiwan, https://www.csie.ntu.edu.tw/~cjlin/papers/guide/guide.pdf, 2003.

[42] H.T. Søgaard, Weed classification by active shape models, Biosyst. Eng. 91 (3) (2005) 271–281, https://doi.org/10.1016/j.biosystemseng.2005.04.011.

[43] A. Ahmad, D. Saraswat, V. Aggarwal, A. Etienne, B. Hancock, Performance of deep learning models for classifying and detecting common weeds in corn and soybean production systems, Comput. Electron. Agric. 184 (2021) 106081, https://doi.org/10.1016/j.compag.2021.106081.

[44] T. Ashraf, Y.N. Khan, Weed density classification in rice crop using computer vision, Comput. Electron. Agric. 175 (2020) 105590, https://doi.org/10.1016/j.compag.2020.105590.

[45] N. Islam, M.M. Rashid, S. Wibowo, C.-Y. Xu, A. Morshed, S.A. Wasimi, S. Moore, S.M. Rahman, Early weed detection using image processing and machine learning techniques in an Australian chilli farm, Agriculture 11 (5) (2021) 387, https://doi.org/10.3390/agriculture11050387.

[46] O. Hassanijalilian, C. Igathinathane, S. Bajwa, J. Nowatzki, Rating iron deficiency in soybean using image processing and decision-tree based models, Remote Sens. 12 (24) (2020) 4143, https://doi.org/10.3390/rs12244143.

[47] M. Weis, T. Rumpf, R. Gerhards, L. Plümer, Comparison of different classification algorithms for weed detection from images based on shape parameters, Bornim. Agrartech. Ber. 69 (2009) 53–64, http://www2.atb-potsdam.de/cigr-imageanalysis/images/06_115_Weis_Vortrag%20_2_.pdf.

[48] J. Gao, W. Liao, D. Nuyttens, P. Lootens, J. Vangeyte, A. Pižurica, Y. He, J.G. Pieters, Fusion of pixel and object-based features for weed mapping using unmanned aerial vehicle imagery, Int. J. Appl. Earth Obs. Geoinf. 67 (2018) 43–53, https://doi.org/10.1016/j.jag.2017.12.012.