The Journal of Agricultural Science

cambridge.org/ags

Seed classification of three species of amaranth (*Amaranthus* spp.) using artificial neural network and canonical discriminant analysis

A. Bagheri[1], L. Eghbali[2] and R. Sadrabadi Haghighi[2]

[1]Department of Agronomy and Plant Breeding, Razi University, Kermanshah, Iran and [2]Department of Agronomy and Plant Breeding, Azad University, Mashhad, Iran

Crops and Soils Research Paper

Cite this article: Bagheri A, Eghbali L, Sadrabadi Haghighi R (2019). Seed classification of three species of amaranth (*Amaranthus* spp.) using artificial neural network and canonical discriminant analysis. *The Journal of Agricultural Science* **157**, 333–341. https://doi.org/10.1017/S0021859619000649

Received: 3 August 2018
Revised: 22 June 2019
Accepted: 6 September 2019

Key words:
Image processing; pattern recognition; seed identification; seed morphology; weed species

Author for correspondence:
A. Bagheri, E-mail: alireza884@gmail.com

Abstract

The current study was conducted in 2013 to identify the seeds of three species of *Amaranthus*, *Amaranthus viridis* L., *Amaranthus retroflexus* L. and *Amaranthus albus* L., by using the artificial neural network (ANN) and canonical discriminant analysis (CDA) methods. To begin with, photographs were taken of the seeds and 13 morphological characteristics of each seed extracted as predictor variables. Backward regression was used to find the most influential variables and seven variables were derived. Thus, predictor variables were divided into two sets of 13 and seven morphological characteristics. The results showed that the recognition accuracy of the ANN made using 13 and seven predictor variables was 81.1 and 80.3%, respectively. Meanwhile, recognition accuracy of the CDA using the seven and 13 predictor variables was 74.0 and 75.7%, respectively. Therefore, in comparison to CDA, ANN showed higher identification accuracy; however, the difference was not statistically significant. Identification accuracy for *A. retroflexus* was higher using the CDA method than ANN, while the ANN method had higher recognition accuracy for *A. viridis* than CDA. In addition, use of 13 predictor variables yielded a greater identification accuracy than seven. The results of the current study showed that using seed morphological characteristics extracted by computer vision could be effective for reliable identification of the similar seeds of *Amaranthus* species.

Introduction

Early identification of weed and other seeds, evaluation of changes in soil seedbank and purity of seed crops could lead to improvement of weed management (Granitto *et al*., 2005; Slaughter *et al*., 2008). Weed seed identification is usually more difficult than for cultivated varieties, as they tend to have large variation compared to seeds belonging to the same crop species (Chtioui *et al*., 1996). Seed identification is often specialized and done visually by experts, which is therefore subjective and requires a high level of skill (Chtioui *et al*., 1998). Visual identification of weed seeds is time consuming, tedious and, even with expert knowledge, inherently inconsistent (OuYang *et al*., 2010), and especially difficult when a high degree of similarity exists between species (Chtioui *et al*., 1998). In addition, due to the subjectivity of these methods, there is a risk of confusion between different inspectors under different circumstances (Majumdar and Jayas, 2000). Therefore, it is important to implement repeatable and rapid automated methods to identify and classify weed seeds (Venora *et al*., 2007).

New techniques such as machine vision have a bright future for the accurate automatic identification of weed seeds. Machine vision is divided into two parts: (a) measurement features and (b) pattern recognition based on the obtained features (Snyder and Qi, 2010). In this technique, characteristics of external variables such as size, shape, colour and surface texture of seeds can be extracted using imaging systems and classification methods (Chtioui *et al*., 1998). Thus, the seeds can be identified using extracted morphological features (Granitto *et al*., 2005). Image processing algorithms implemented by machine vision are more accurate and efficient in measuring seed size than highly experienced inspectors working by microscope (Venora *et al*., 2007). The benefits of such methods are considerable in seed classification. For example, Granitto *et al*. (2002) used machine vision techniques to identify 57 species of weeds and demonstrated promising results. Fawzi *et al*. (2010) studied 11 Silene species using light and electron microscopic morphology to determine the importance of seed coating as a taxonomic character. Seeds were kidney shaped or spherical–kidney shaped and their colour was green to brown. The length of the seeds was between 0.5 and 1.2 mm.

Some image processing algorithms are available for pattern recognition and to extract the seed morphological characteristics, of which canonical discriminant analysis (CDA) and artificial neural networks (ANN) are two main approaches (Granitto *et al*., 2002).

Canonical discriminant analysis is a supervised learning technique of statistical pattern recognition (Jain *et al.*, 2000). The application of statistical methods in pattern recognition was first formalized by Chow (1965). During the learning process, multivariate CDA defines optimal boundaries between the clusters of values in the parameter space. The performance of a classification depends on the separability of the classes. This suggests that the centres of clusters within the measurement space should be sufficiently separated. However, investigations of seed identification have shown that linear discriminators do not often yield satisfactory performances (Chtioui *et al.*, 1996). This method has been employed in agricultural science for various purposes, such as species genetic diversity, plant morphology, seed systematic and classification, and seed quality testing (Olesen *et al.*, 2011, 2015; Hoyo and Tsuyuzaki, 2013; Eizenga *et al.*, 2014; Padonou *et al.*, 2014; Pometti *et al.*, 2016; Roy *et al.*, 2016; Tungmunnithum *et al.*, 2016; Şeker and Şenel, 2017).

In recent years, ANN has become used widely for forecasting in various fields of research including finance, power generation, pharmaceutical, water and environmental resources (Li *et al.*, 2014, 2015; Qiu *et al.*, 2016; Velásco-Mejía *et al.*, 2016; Monteiro *et al.*, 2017). In agriculture, ANN is one of the main machine learning models which have been used widely (van Evert *et al.*, 2017). The main idea of ANN for processing data is based on the way the nervous system and brain function in order to learn and create knowledge. Biological neural networks are able to learn based on a system through which adaptive learning takes place, which means that the system is trained using different examples, so when new entries are entered, the system will produce 'the right answer', which is subjective (Kasabov, 1996). Artificial neural networks are based on training, and through this training, the mechanisms of the phenomenon are estimated (Kohonen, 2012). These networks demonstrate very high efficiency in estimation and approximation (Kasabov, 1996).

Considering the importance of weed seed identification, the present study attempted to identify three species of *Amaranthus* genus, commonly referred to as 'amaranth'. The species of this genus are among the most troublesome weeds in many crop production systems and cause substantial losses to many crops (Horak *et al.*, 1994; Sellers *et al.*, 2003; Horak and Loughin, 2009). The seeds of *Amaranthus* species are very similar, so that seed identification is very difficult and usually done based on the capsule characteristics (Horak *et al.*, 1994). In the current experiment, an attempt was made to classify three species of *Amaranthus*, including *Amaranthus albus* L. (white amaranth), *Amaranthus retroflexus* L. (red-root amaranth) and *Amaranthus viridis* L. (slender amaranth), based on shape characteristics of the seeds using machine vision technique and ANN and CDA as pattern recognition methods, to assess the accuracy of the methods.

Materials and methods

Seed collection and identification

The current experiment was conducted in 2013 to identify seeds of three species of the genus *Amaranthus*, using a machine vision approach and methods of ANN and CDA. In order to provide the seeds required, the target species were identified and seeds collected from farms around Mashhad (36°81′N, 59°82′E, 985 m a.s.l), in northeast Iran. According to the United Nations Environment Program, the climate of seed collection areas is considered as moderate semi-arid, with an average annual temperature of 14 °C and 253 mm of annual precipitation (Ashraf *et al.*, 2014). The farms belonged to Ferdowsi University of Mashhad and the commercial firm Astan Quds Razavi; almost 1200 ha were monitored. To this end, the three *Amaranthus* species were identified and collected by walking in farms, in two sampling stages including mid-October and mid-November. The distance between sampling points was >250 m, to ensure that no spatial correlation occurred among selected populations (Guisan and Theurillat, 2000). In order to consider the variation among populations, sampling was done from different locations and at two different times. To ensure correct identification of the collected species, the collected plants were transferred to the Institute of Plant Sciences, Ferdowsi University of Mashhad and identified by botany experts. Then, the seeds of each species were isolated and prepared by cleaning and drying before randomly selecting 200 seeds of each species for further analysis.

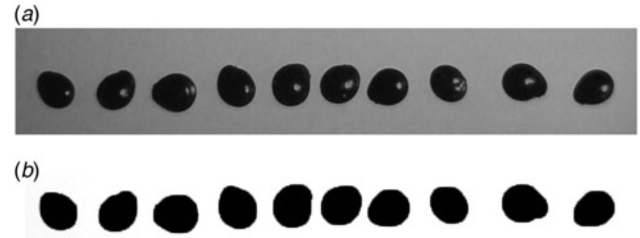

Fig. 1. Modifying images of amaranth species in order to increase the resolution of the seeds and background. (*a*) Original image and (*b*) modified image.

Extraction of the shape characteristics of the seeds

To extract the shape characteristics of the collected seeds, 200 images were taken of each species of *Amaranthus* using a digital camera (Sony, Cybershot DSC-W70, Japan) attached to a stereo microscope. Therefore, the images of 600 seeds of three species of *Amaranthus* were captured. Adobe Photoshop CS6 Extended software was used to modify the images, such as removing shadows, creating a good resolution between the seeds and the foreground and removal of other noises in the images (Fig. 1). The corrected images were segmented by thresholding (Liu *et al.*, 2005), using image processing software JMicroVision v 1.2.7 to process and extract the seed morphological features from segmented images. Hence, 13 morphological features, including area, perimeter, orientation, length, width, eccentricity, compactness, equivalent circular diameter, elongation, ellipticity, rectangularity, solidity and convexity were extracted. Each of the extracted 13 variables could have different effects on the identification of different species of amaranth seeds (Cervantes *et al.*, 2016). In order to find the most influential variables, the backward method was used by means of the software SPSS v. 17.00, and seven variables were derived.

Classification of the seeds

The ANN and CDA methods were used to identify the seeds. For this purpose, the two sets of normalized data of shape characteristics of each *Amaranthus* species were applied, which included a series of data (overall, 13 shape characteristics data were extracted and a set was obtained from the backward method). Data were normalized by Johansson Transformation or Box–Cox methods by means of the software Minitab v. 15.1.1.0. Normalized data can contribute to increasing the accuracy of seed identification. Dubey *et al.* (2006) used neural networks to identify three varieties of wheat and stated that if the data were not normal, the neural network was not able to classify the wheat seeds.

Canonical discriminant analysis was applied to highlight among-group variation and minimize within-group variation (Li *et al.*, 2013). In CDA, the group variables are transformed into an identity matrix and group means are calculated. A principal component analysis was conducted on the calculated means and eigenvalues obtained by dividing between-group variations by within-group variation. In order to obtain the canonical variables, the principal components are transformed into the space of the original variables, then the boundaries are obtained (Chen *et al.*, 2013). After data preparation, in order to classify the seeds by CDA, Wilks' Lambda method was applied to two sets of normalized shape data of *Amaranthus* species, using SPSS v. 17.00. The Wilks' Lambda evaluates the performance of the discriminant analyses and is the ratio of the within-group variation and the total variation. This statistic is used to test the significance of the discriminant function and provides an objective means of calculating the chance-corrected percentage of agreement between real and predicted groups (Khemiri *et al.*, 2018).

Artificial neural networks are information-processing systems consisting of networks of simple interconnected processing elements (neurons) which are able to construct a mathematical model to predict the complex behaviour of a phenomenon. An ANN consists of three layers: input, hidden and output layers. The input layer provides information on the studied phenomenon; the hidden layer performs computations in which the level of complexity is determined. In the hidden layer, transfer functions are specified to determine the learning process and relevant weights between corresponding neurons (Alvarez, 2009). The output layer transfers the determined data from the hidden layer to the outside of the network. Generally in ANN, data are compiled from the input layer, and after passing through the hidden layer are excluded from the output layer (Kasabov, 1996; Alvarez, 2009). There are different numbers of neurons in each layer. The number of neurons in the input and output layers is determined based on the purpose of the study and the research question. The neuron numbers in hidden layers may be adjusted by trial and error (Dubey *et al.*, 2006).

In order to identify the seed by ANN, the normalized morphological data were used as the input layer of neural networks. The number of input neurons was considered to be 13 (total morphological characteristics or predictive variables derived from the seeds) and seven (the predictor variables derived from backward regression). The number of neurons in the output layer was three, based on the number of *Amaranthus* species. Artificial neural network performance depends on the choice of the number of hidden layers (Ramchoun *et al.*, 2017): hence, during construction of the ANNs, one to ten hidden layers were used and tested. To classify seeds of the studied species of *Amaranthus*, various neural networks such as Multilayer Perceptrons Neural Networks, Generalized Feed Forward Neural Networks, Modular Neural Networks and Principal Component Analysis Neural Networks were tested and the best network was selected based on the highest classification accuracy. In the current study, the learning rules of Momentum, and Levenberg Marquardt and also the functions of TanhAxon, SigmoidAxon, Linear TanhAxon, Linear SigmoidAxon, SoftMaxAxon, LinearAxon and Axon were tested as transfer functions. After construction of the neural networks, the learning rule and transfer function with the highest classification accuracy were selected. The software NeuroSolution v. 5.00 was used to build the neural networks. In the current experiment, overall convergence of the training was obtained via a learning epoch (one cycle through a training process) of 1000, so further repeats of the learning epoch had no significant increase in network performance.

Precision validation

In the current experiment, to avoid increasing the error and overestimation, ANN training was terminated using the cross-validation stopping method, which stops the training network at the point of the smallest error in the validation data set (Amari *et al.*, 1997; Benedetti *et al.*, 2004). So, 15% of the data were allocated as the validation data set. Thus, with the determination of error between the desired output and the actual output, and its increase during network training, the training operation was stopped. In addition, 15% was allocated for network testing and evaluation of accuracy. The remaining 70% of the data were used for network training. In the method of least mean square error of the training data, 20% of the data were allocated for the network test and the remaining 80% were allocated for network training.

A *t* test was used for statistical comparison of seed classification accuracy between the two sets of input data in each of the ANN and CDA methods and also between the two methods.

Results

Seeds shape description

The mean and standard deviation of shape characteristics of *Amaranthus* species are shown in Table 1. The maximum value of the standard deviation of the three species of *Amaranthus* was observed in orientation, and the minimum in ellipticity, rectangularity, solidity and convexity (Table 1). The results showed that *A. retroflexus* had larger seeds than the other two species, and *A. viridis* had the smallest seeds. Other traits related to the seed size and shapes of the studied seeds are shown in Table 1.

Artificial neural network

Based on the results of the backward regression method, the predictor variables such as perimeter, length, width, eccentricity, compactness, elongation and rectangularity were the best predictors for *A. viridis*, *A. retroflexus* and *A. albus* (Table 2). The outputs of backward regression were used as a data set to build the neural networks. A Generalized Feed Forward Neural Network with five hidden layers for seven (extracted from backward regression) data series and a Principal Component Analysis Neural Network with three hidden layers for 13 (total data) data series were the best networks for identification and classification in *Amaranthus* species. In these networks, a stopping criterion of cross-validation performed better in comparison to a stopping criterion of minimum mean square error of the training set and yielded more proper networks with classification accuracy. Mean square errors of the training set and cross-validation of 13 normal input variables were 0.175 and 0.191, respectively. This network was stopped in epoch 204 (Fig. 2*a*). After reaching the minimum mean square error of cross-validation, the training process continued for some time to ensure proper network

Table 1. The mean and standard deviation (±) of morphological characteristics of 200 seeds of each of three species of *Amaranthus* spp.

Shape characteristic	Mean and standard deviation		
	A. albus	*A. retroflexus*	*A. viridis*
Area	0.7 ± 0.08	0.9 ± 0.11	0.7 ± 0.09
Perimeter	3.0 ± 0.22	3.3 ± 0.21	2.9 ± 0.19
Orientation[a]	88 ± 52.1	90 ± 46.0	88 ± 55.8
Length[b]	1.0 ± 0.06	1.1 ± 0.07	1.0 ± 0.07
Width[c]	0.9 ± 0.06	1.0 ± 0.07	0.9 ± 0.06
Eccentricity[d]	1.2 ± 0.08	1.4 ± 0.14	1.3 ± 0.09
Compactness[e]	1.0 ± 0.05	1.0 ± 0.02	1.0 ± 0.02
Equivalent circular diameter[f]	0.9 ± 0.05	1.0 ± 0.07	0.9 ± 0.06
Elongation[g]	0.9 ± 0.03	0.8 ± 0.04	0.9 ± 0.03
Ellipticity[h]	0.0 ± 0.02	0.0 ± 0.01	0.0 ± 0.02
Rectangularity[i]	0.3 ± 0.02	0.3 ± 0.02	0.3 ± 0.02
Solidity[j]	1.0 ± 0.01	1.0 ± 0.00	1.0 ± 0.00
Convexity[k]	1.0 ± 0.03	1.0 ± 0.01	1.0 ± 0.01

[a]Angle between the horizontal axis and the major axis of the ellipse equivalent to the seed (0–180°, anti-clockwise).
[b]Calliper length along the orientation axis of the seed.
[c]Calliper length along the orientation axis + 90° of the seed.
[d]Ratio between the major and the minor axis of the ellipse equivalent to the seed (first- and second-degree moments).
[e]Ratio of the area of the seed to the area of a circle with the same perimeter.
[f]Diameter of a circle with the same area as that of the seed.
[g]Ratio of the length to the width.
[h]Ratio of the area of an ellipse (formed with length and width as axes) to the area of the seed.
[i]Ratio of the area of a rectangle (formed with length and width as sides) to the area of the seed.
[j]Ratio of the area of the seed to the convex area.
[k]Ratio of the convex perimeter to the perimeter of the seed.

Table 2. Backward regression analysis of predictive variables with a significant effect on the classification of different species of *Amaranthus* spp

Step	Predictive variable	*t*	*P* value
1	Perimeter	−2.904	0.004
2	Length	2.860	0.004
3	Width[a]	4.363	0.000
4	Eccentricity[b]	2.546	0.011
5	Compactness[c]	2.620	0.009
6	Elongation[d]	4.736	0.000
7	Rectangularity[e]	1.953	0.051

[a]Calliper length along the orientation axis + 90° of the seed.
[b]Ratio between the major and the minor axis of the ellipse equivalent to the seed (first- and second-degree moments).
[c]Ratio of the area of the seed to the area of a circle with the same perimeter.
[d]Ratio of the length to the width.
[e]Ratio of the area of a rectangle (formed with length and width as sides) to the area of the seed.

training. It is recommended that network training should be continued for a period of time to eliminate the risk of lack of proper training data, after the first test error starts to increase (Masters, 1993). Also, mean square error of the network training and cross-validation that consisted of seven normal data input of predictor variables was 0.185 and 0.205, respectively, and the network was stopped in epoch 847 (Fig. 2*b*).

After testing neural networks, the network consisted of 13 normal input predictor variables with an overall classification accuracy of 81.1%, and species *A. retroflexus*, *A. viridis* and *A. albus* were classified by values of 90.3, 82.1 and 71.0%, respectively (Table 3). Furthermore, in the neural network consisting of seven normal input predictor variables, species *A. retroflexus*, *A. viridis* and *A. albus* were classified by values of 92.0, 82.9 and 66.1%, respectively (Table 3); however, this network had an overall classification of 80.3%. The results showed that the use of shape characteristics can be very helpful in identifying the seeds of three species of *Amaranthus*. Granitto *et al.* (2002) used six morphological, four colour and two textural characteristics for identification of seeds from 57 different weed species and illustrated that the maximum separation accuracy of seeds was in relation to their morphological characteristics.

The results of ANN for normalized data of *Amaranthus* species showed that the neural network built from 13 normal input predictor variables had higher accuracy compared with the neural network built from seven normal input predictor variables. Although the difference was not statistically significant, it can be concluded that increasing the predictor variables could increase the quality of neural network training.

Canonical discriminant analysis

The results of the CDA method on morphological characteristics of *Amaranthus* species showed that in both the 13 and seven normal predictor variables, discriminate functions significantly described the differences among the *Amaranthus* species in the model and fit the data well (Table 4). As a consequence, the three studied *Amaranthus* species were significantly classified

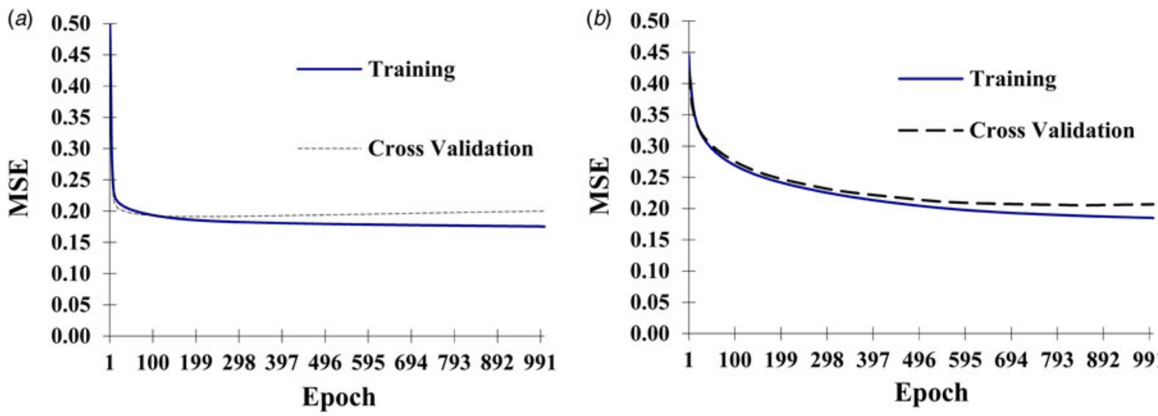

Fig. 2. The mean square error (MSE) of the training (continuous line) and cross-validation (broken line) procedure of networks including (*a*) seven and (*b*) 13 normal input variables.

Table 3. Studied species identification accuracy (%) of Artificial Neural Network on the normalized data of 13 and seven predictor variables (network stopping criterion, increase in the mean square error of validation process)

Actual/desired	*A. albus*	*A. retroflexus*	*A. viridis*
	13 normal input variables		
A. albus	66.1	4.0	14.6
A. retroflexus	6.8	92.0	2.4
A. viridis	27.1	4.0	82.9
	Seven normal input variables		
A. albus	71.0	9.7	14.3
A. retroflexus	0.0	90.3	3.6
A. viridis	29.0	0.0	82.1

Table 4. Summary of canonical discriminant functions was used in the analysis for the classification of *Amaranthus* species using 13 and seven normal input variables. All functions were significant ($P \leqslant 0.01$)

Function	Eigenvalue	% of Variance	Wilks' Lambda	χ^2
		13 normal input variables		
1	2.131	94.6	0.285	745.598
2	0.121	5.4	0.892	67.628
		Seven normal input variables		
1	2.451	94.9	0.256	806.620
2	0.132	5.1	0.884	73.302

from each other based on seed morphological characteristics. *A. retroflexus*, compared to *A. albus* and *A. viridis*, had more distinctive morphological features; however, *A. albus* and *A. viridis* showed more similarities to each other (Fig. 3).

Identification accuracy of *Amaranthus* species by the CDA method showed that in both the 13 and seven normal variables input, the highest identification accuracy was achieved for *A. retoflexus*, in which the accuracy of the seven normal variables (93.0%) was more than in the 13 variables input (92.5%). In addition, detection percentage in the species *A. albus* was more than *A. viridis* in the models with both input data set (Table 5). The overall identification percentage of applied models with the 13 and seven normal variables input were 75.7 and 74.0%, respectively, which did not show significant difference.

Discussion

Considering the characteristics that provide the potential to increase seed classification accuracy is very important in seed identification. Various characteristics such as shape, size, colour and texture of seeds have been considered for seed classification (Paliwal *et al*., 2001; Granitto *et al*., 2002, 2005; Liu *et al*., 2005; Dana and Ivo, 2008; Chen *et al*., 2010; OuYang *et al*., 2010). However, regarding the smooth surface and same colour of the surface of seeds examined in the current study, the identification of seeds of the three species was investigated based on the morphological characteristics. Seed morphology is considered as an effective factor for seed description and analysis of intra- and inter-specific differences between plant species and varieties, but for the species studied here, due to the small size of the seeds and high similarities between the seeds of the three species, visual identification of each is almost impossible by non-specialists. On the other hand, some situations, such as soil seedbank surveys, require identification of a large number of different weed species seeds (Gardarin *et al*., 2009) and using the visual method is very difficult in these cases.

The development of computer vision capabilities allows a reliable and fast identification and classification of seeds, even for non-specialists (Tellaeche *et al*., 2011). Using image processing to extract several quantified seed morphological features can be an efficient tool in comparative taxonomy (Cervantes *et al*., 2016). The developments of imaging systems are mainly based on the computation of geometrical characteristics of the seeds because they have forms (shape factor, aspect ratio, length ratio, etc.) which can be identified (Perez *et al*., 2000; Onyango and Marchant, 2003). Anouar *et al*. (2001) identified the seeds of four varieties of carrots based on size, using a machine vision system. In a study by OuYang *et al*. (2010), identification accuracy of five varieties of rice by ANN was 86.65%. Liu *et al*. (2005) evaluated a neural network to identify the seeds of six rice varieties and obtained an average identification accuracy of 84.83%.

The average identification accuracy of the ANN and CDA methods in the current study (based on seven and 13

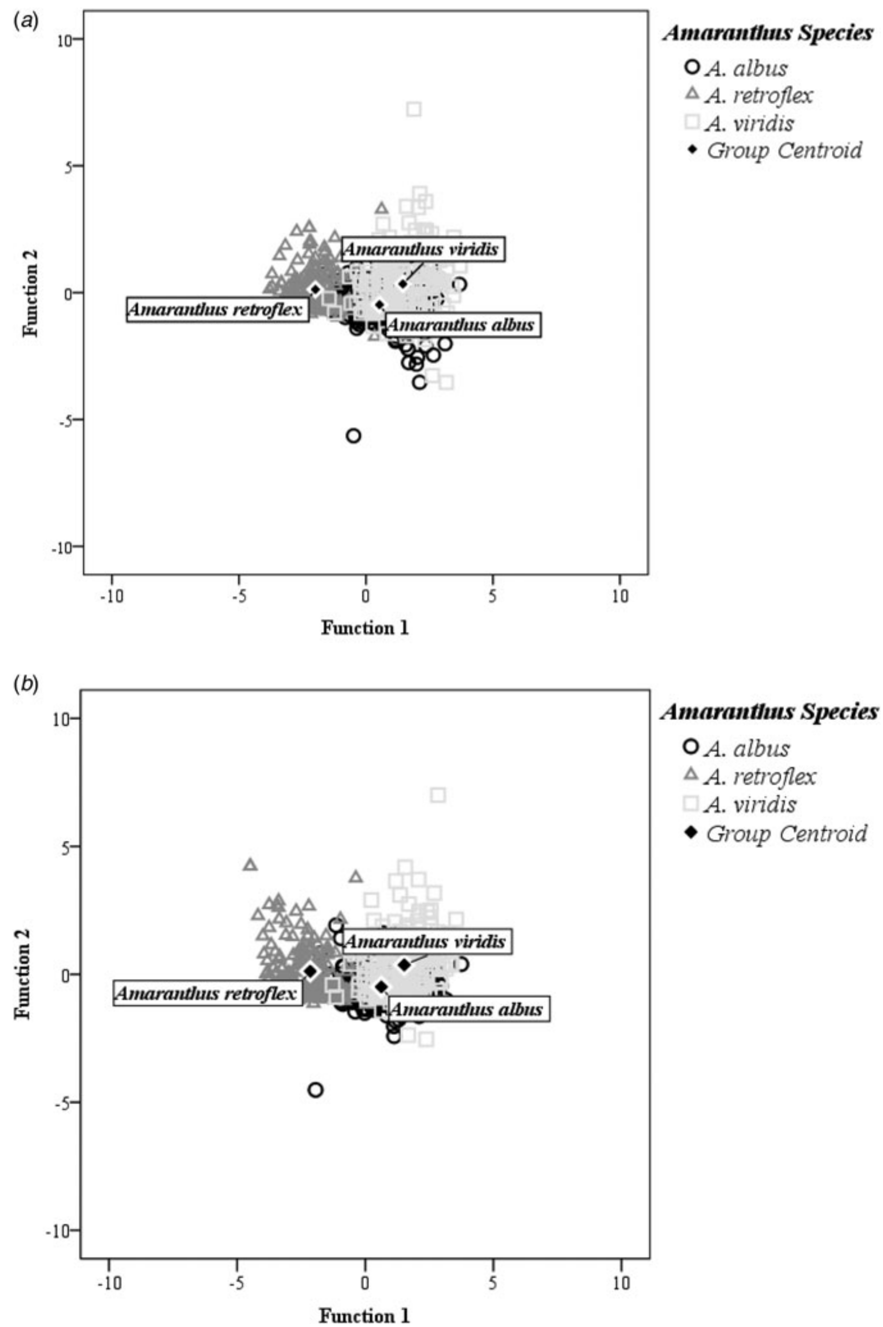

Fig. 3. Bi-plots of canonical discriminant functions for shape characteristics of three studied *Amaranthus* species. (*a*) Seven and (*b*) 13 normal input variables.

morphological features of the seeds) were >80 and 74%, respectively. Dubey *et al.* (2006) illustrated that the combination of ANN with image processing had the potential to identify different varieties of wheat and they were able to identify three varieties of wheat with an accuracy >80%. Liu *et al.* (2005) developed a neural network model to identify six varieties of rice seeds, with identification accuracies between 74 and 95%. Paliwal *et al.* (2001) used a neural network to classify the grains of two varieties of wheat, barley, oats and rye. They considered four morphological traits, namely Feret diameter, area, width and compactness, as input layer and reported identification accuracies for wheat and oats of about 97% and for barley and rye of about 88%. In the study of Shrestha *et al.* (2015), CDA was used for pairwise discrimination of 11 cultivars of tomato, with an accuracy between 85 and 100%. The results of the current study showed that *A. retroflexus* was identified as the highest accuracy in both ANN (90 and 92.3%) and CDA (92.5 and 93%) methods, while the other two species were identified with accuracy between 66.1 and 82.9% for ANN and 59 and 70% for CDA, respectively. All shape characteristics of *A. albus* and *A. viridis* were very similar; however, *A. retroflexus* differed strongly from the other two species in terms of area, perimeter, length and eccentricity. Owing to this, *A. albus* and *A. viridis* were misidentified as each other rather than *A. retroflexus*. *A. retroflexus* is an aggressive weed in semi-arid environments such as Mediterranean areas (Lovelli *et al.*, 2010). Accurate seed identification of the weed can be important in weed management programmes and the use of machine vision can be helpful in this regard.

In the current study, two sets of data were used ($n = 7$ and $n = 13$ normalized morphological data) in order to identify the

https://doi.org/10.1017/S0021859619000649 Published online by Cambridge University Press

Table 5. Studied species identification accuracy (%) of canonical discriminant analysis on the normalized data of 13 and seven predictor variables

Actual/desired	*A. albus*	*A. retroflexus*	*A. viridis*
A. albus	13 normal input variables		
	68.0	8.5	23.5
A. retroflexus	6.5	92.5	1.0
A. viridis	32.0	1.5	66.5
	Seven normal input variables		
A. albus	70.0	10.0	20.0
A. retroflexus	5.5	93.0	1.5
A. viridis	40.0	1.0	59.0

seeds of *Amaranthus* species using ANN and CDA. The results showed that all 13 seed morphological traits achieved higher classification accuracy than seven seed morphological traits in both ANN and CDA methods; however, the difference was not statistically significant. In some studies, the omission of some characteristics resulted in decreased seed identification. For example, Dana and Ivo (2008) used computer image analysis to describe seeds of 53 flax cultivars and stated that significant multivariate clustering was obtained by using a non-reduced data set composed of four morphological and three colour features of the seeds. In the current study, data reduction had no significant effect on seed identification, so a reduced data set (perimeter, length, width, eccentricity, compactness, elongation and rectangularity) could be suggested as the input data.

Comparison between ANN and CDA methods revealed that average accuracy of the studied species seed identification of ANN and CDA methods (for both the seven and 13 morphological features) was 80.7 and 74.8%, respectively, although this difference was not statistically significant. The CDA method showed higher accuracy in the identification of *A. retroflexus* species compared with the ANN method. Meanwhile, recognition accuracy in *A. viridis* was higher with ANN in comparison with CDA. Also, identification of *A. albus* species in ANN and CDA methods was almost the same. In total, the results indicated that recognition accuracy of the ANN method to identify the studied *Amaranthus* species was higher than CDA. Chtioui *et al.* (1996), in a study on comparison of discriminant analysis (DA) and ANN to identify weed seeds based on morphological and textural characteristics, reported that ANN had higher accuracy than DA. Ronge and Sardeshmukh (2014) developed the ANN and k-nearest neighbour (k-NN) methods for the classification of four Indian wheat seed varieties: 120 images (40 images of four classes, ten images of each class) were taken and converted into greyscale images. Texture features of wheat varieties were extracted. The feature group which gave highest percentage of accuracy in classification was determined. The ANN method showed average accuracies of 66.68–100%, while average accuracies of k-NN were 39–85%. Their results showed that ANN outperformed k-NN.

In most studies of automatic identification of plant seeds, different varieties of crops have been investigated (Dehghan-Shoar *et al.*, 1998; Majumdar and Jayas, 2000; Anouar *et al.*, 2001; Marini *et al.*, 2004; Liu *et al.*, 2005; Dubey *et al.*, 2006; Dana and Ivo, 2008; OuYang *et al.*, 2010) and less attention has been paid to weeds (Granitto *et al.*, 2002, 2005; Xinshao and Cheng, 2015); that is, in studies on weeds, the seeds of weed species from different families have hardly been considered. The seeds of different weed species are different in terms of size, shape and surface texture and can even be identified visually. Based on a review of scientific literature by the authors, seed identification of closely related species of a weedy genus has not been studied. Meanwhile, in the current study, three species of *Amaranthus* genus with very similar seeds were investigated. The classification accuracy, especially in the cases of *A. retroflexus* (in the both ANN and CDA methods) and *A. viridis* (in the ANN method), has excellent potential for identification of these species.

Conclusion

In the current study, the overall accuracy of the ANN and CDA methods was 80.7 and 74.8% in studied seed recognition. The identification of *A. retroflexus* was >90% in both ANN and CDA models. The identification accuracy of *A. viridis* in the neural network method was >80%; however, it was 66.5 and 59% in the CDA method for the total input data and the stepwise regression derived data, respectively. Although there is no significant difference between the overall accuracy of ANN and CDA methods, ANN had high accuracy in identifying two of the three studied species while CDA had an acceptable accuracy in identifying only the seeds of *A. retroflexus*. Weed species seed identification is a professional work carried out by specialists; however, using new methods of identification, it can be provided for non-specialists. Utilization of weed seed automatic identification techniques and application of the results could lead to the quick and easy identification required in agricultural research.

Financial Support. This research received no specific grant from any funding agency, commercial or not-for-profit sectors.

Conflict of interest. The authors have no conflict of interest to declare.

Ethical standards. Not applicable.

References

Alvarez R (2009) Predicting average regional yield and production of wheat in the Argentine Pampas by an artificial neural network approach. *European Journal of Agronomy* **30**, 70–77.

Amari S, Murata N, Muller KR, Finke M and Yang HH (1997) Asymptotic statistical theory of overtraining and cross-validation. *IEEE Transactions on Neural Networks* **8**, 985–996.

Anouar F, Mannino MR, Casals ML, Fougereux JA and Demilly D (2001) Carrot seeds grading using a vision system. *Seed Science and Technology* **29**, 215–225.

Ashraf B, Yazdani R, Mousavi-Baygi M and Bannayan M (2014) Investigation of temporal and spatial climate variability and aridity of Iran. *Theoretical and Applied Climatology* **118**, 35–46.

Benedetti S, Mannino S, Sabatini AG and Marcazzan GL (2004) Electronic nose and neural network use for the classification of honey. *Apidologie* **35**, 397–402.

Cervantes E, Martín JJ and Saadaoui E (2016) Updated methods for seed shape analysis. *Scientifica* **2016**, 1–10. http://dx.doi.org/10.1155/2016/5691825.

Chen X, Xun Y, Li W and Zhang J (2010) Combining discriminant analysis and neural networks for corn variety identification. *Computers and Electronics in Agriculture* **71**(suppl. 1), S48–S53.

Chen G, Bui TD, Krzyzak A and Krishnan S (2013) Small bowel image classification based on Fourier-Zernike moment features and canonical discriminant analysis. *Pattern Recognition and Image Analysis* **23**, 211–216.

Chow CK (1965) Statistical independence and threshold functions. *IEEE Transactions on Electronic Computers* **EC-14**, 66–68.

Chtioui Y, Bertrand D, Dattée Y and Devaux MF (1996) Identification of seeds by colour imaging: comparison of discriminant analysis and artificial neural network. *Journal of the Science of Food and Agriculture* **71**, 433–441.

Chtioui Y, Bertrand D and Barba D (1998) Feature selection by a genetic algorithm. Application to seed discrimination by artificial vision. *Journal of the Science of Food and Agriculture* **76**, 77–86.

Dana W and Ivo W (2008) Computer image analysis of seed shape and seed color for flax cultivar description. *Computers and Electronics in Agriculture* **61**, 126–135.

Dehghan-Shoar M, Hampton J and Haslett S (1998) Identification of, and discrimination among, lucerne (*Medicago sativa* L.) varieties using seed image analysis. *Plant Varieties & Seeds* **11**, 107–127.

Dubey BP, Bhagwat SG, Shouche SP and Sainis JK (2006) Potential of artificial neural networks in varietal identification using morphometry of wheat grains. *Biosystems Engineering* **95**, 61–67.

Eizenga GC, Ali ML, Bryant RJ, Yeater KM, McClung AM and McCouch SR (2014) Registration of the Rice Diversity Panel 1 for genome-wide association studies. *Journal of Plant Registrations* **8**, 109–116.

Fawzi NM, Fawzy AM and Mohamed AAA (2010) Seed morphological studies on some species of *Silene* l. (Caryophyllaceae). *International Journal of Botany* **6**, 287–292.

Gardarin A, Dürr C and Colbach N (2009) Which model species for weed seedbank and emergence studies? A review. *Weed Research* **49**, 117–130.

Granitto PM, Navone HD, Verdes PF and Ceccatto HA (2002) Weed seeds identification by machine vision. *Computers and Electronics in Agriculture* **33**, 91–103.

Granitto PM, Verdes PF and Ceccatto HA (2005) Large-scale investigation of weed seed identification by machine vision. *Computers and Electronics in Agriculture* **47**, 15–24.

Guisan A and Theurillat JP (2000) Assessing alpine plant vulnerability to climate change: a modeling perspective. *Integrated Assessment* **1**, 307–320.

Horak MJ and Loughin TM (2009) Growth analysis of four *Amaranthus* species. *Weed Science* **48**, 347–355.

Horak MJ, Peterson DE, Chessman DJ and Wax LM (1994) *Pigweed Identification: A Pictorial Guide to the Common Pigweeds of the Great Plains*. Manhattan, KS, USA: Kansas State University Agricultural Experiment Station and Cooperative Extension Service.

Hoyo Y and Tsuyuzaki S (2013) Characteristics of leaf shapes among two parental *Drosera* species and a hybrid examined by canonical discriminant analysis and a hierarchical Bayesian model. *American Journal of Botany* **100**, 817–823.

Jain AK, Duin RPW and Mao J (2000) Statistical pattern recognition: a review. *IEEE Transactions on Pattern Analysis and Machine Intelligence* **22**, 4–37.

Kasabov NK (1996) *Foundations of Neural Networks, Fuzzy Systems, and Knowledge Engineering*. Cambridge, MA, USA: MIT Press.

Khemiri S, Gaamour A, Ben Abdallah L and Fezzani S (2018) The use of otolith shape to determine stock structure of *Engraulis encrasicolus* along the Tunisian coast. *Hydrobiologia* **821**, 73–82.

Kohonen T (2012) *Self-Organization and Associative Memory*. Springer Series in Information Sciences no. 8. Germany, Berlin: Springer Berlin Heidelberg.

Li L, Boyd CE, Odom J and Dong S (2013) Identification of ictalurid catfish fillets to rearing location using elemental profiling. *Journal of the World Aquaculture Society* **44**, 405–414.

Li X, Zecchin AC and Maier HR (2014) Selection of smoothing parameter estimators for general regression neural networks – Applications to hydrological and water resources modelling. *Environmental Modelling & Software* **59**, 162–186.

Li X, Maier HR and Zecchin AC (2015) Improved PMI-based input variable selection approach for artificial neural network and other data driven environmental and water resource models. *Environmental Modelling & Software* **65**, 15–29.

Liu ZY, Cheng F, Ying YB and Rao XQ (2005) Identification of rice seed varieties using neural network. *Journal of Zhejiang University: Science B* **6**, 1095–1100.

Lovelli S, Perniola M, Ferrara A, Amato M and Di Tommaso T (2010) Photosynthetic response to water stress of pigweed (*Amaranthus retroflexus*) in a southern-Mediterranean area. *Weed Science* **58**, 126–131.

Majumdar S and Jayas DS (2000) Classification of cereal grains using machine vision: I. Morphology models. *Transactions of the ASAE* **43**, 1669–1675.

Marini F, Zupan J and Magrì AL (2004) On the use of counterpropagation artificial neural networks to characterize Italian rice varieties. *Analytica Chimica Acta* **510**, 231–240.

Masters T (1993) *Practical Neural Network Recipes in C++*. San Diego, USA: Academic Press.

Monteiro RVA, Guimarães GC, Moura FAM, Albertini MRMC and Albertini MK (2017) Estimating photovoltaic power generation: performance analysis of artificial neural networks, Support Vector Machine and Kalman filter. *Electric Power Systems Research* **143**, 643–656.

Olesen MH, Carstensen JM and Boelt B (2011) Multispectral imaging as a potential tool for seed health testing of spinach (*Spinacia oleracea* L.). *Seed Science and Technology* **39**, 140–150.

Olesen MH, Nikneshan P, Shrestha S, Tadayyon A, Deleuran LC, Boelt B and Gislum R (2015) Viability prediction of *Ricinus cummunis* L. seeds using multispectral imaging. *Sensors (Switzerland)* **15**, 4592–4604.

Onyango CM and Marchant JA (2003) Segmentation of row crop plants from weeds using colour and morphology. *Computers and Electronics in Agriculture* **39**, 141–155.

OuYang A, Gao R, Liu Y, Sun X, Pan Y and Dong X (2010) An automatic method for identifying different variety of rice seeds using machine vision technology. In Yue S, Wei HL, Wang L and Song Y (eds), *Sixth International Conference on Natural Computation (ICNC)*. Yantai, Shandong, China: Institute of Electrical and Electronics Engineers, pp. 84–88.

Padonou EA, Kassa B, Assogbadjo AE, Fandohan B, Chakeredza S, Glèlè Kakaï R and Sinsin B (2014) Natural variation in fruit characteristics and seed germination of *Jatropha curcas* in Benin, West Africa. *Journal of Horticultural Science and Biotechnology* **89**, 69–73.

Paliwal J, Visen NS and Jayas DS (2001) AE – automation and emerging technologies: evaluation of neural network architectures for cereal grain classification using morphological features. *Journal of Agricultural Engineering Research* **79**, 361–370.

Perez AJ, Lopez F, Benlloch JV and Christensen S (2000) Colour and shape analysis techniques for weed detection in cereal fields. *Computers and Electronics in Agriculture* **25**, 197–212.

Pometti CL, Bessega CF, Vilardi JC, Ewens M and Saidman BO (2016) Genetic variation in natural populations of *Acacia visco* (Fabaceae) belonging to two sub-regions of Argentina using AFLP. *Plant Systematics and Evolution* **302**, 901–910.

Qiu M, Song Y and Akagi F (2016) Application of artificial neural network for the prediction of stock market returns: the case of the Japanese stock market. *Chaos, Solitons & Fractals* **85**, 1–7.

Ramchoun H, Idrissi MAJ, Ghanou Y and Ettaouil M (2017) New modeling of multilayer perceptron architecture optimization with regularization: an application to pattern classification. *IAENG International Journal of Computer Science* **44**, 261–269.

Ronge RV and Sardeshmukh M (2014) Comparative analysis of Indian wheat seed classification. In *2014 International Conference on Advances in Computing, Communications and Informatics (ICACCI)*. Piscataway, NJ, USA: IEEE, pp. 937–942.

Roy S, Marndi BC, Mawkhlieng B, Banerjee A, Yadav RM, Misra AK and Bansal KC (2016) Genetic diversity and structure in hill rice (*Oryza sativa* L.) landraces from the North-Eastern Himalayas of India. *BMC Genetics* **17**, 1–15. doi: 10.1186/s12863-016-0414-1.

Şeker ŞS and Şenel G (2017) Comparative seed micromorphology and morphometry of some orchid species (Orchidaceae) belong to the related Anacamptis, Orchis and Neotinea genera. *Biologia (Poland)* **72**, 14–23.

Sellers BA, Smeda RJ, Johnson WG, Kendig JA and Ellersieck MR (2003) Comparative growth of six *Amaranthus* species in Missouri. *Weed Science* **51**, 329–333.

Shrestha S, Deleuran LC, Olesen MH and Gislum R (2015) Use of multispectral imaging in varietal identification of tomato. *Sensors (Switzerland)* **15**, 4496–4512.

Slaughter DC, Giles DK and Downey D (2008) Autonomous robotic weed control systems: a review. *Computers and Electronics in Agriculture* **61**, 63–78.

Snyder WE and Qi H (2010) *Machine Vision*. Cambridge, UK: Cambridge University Press.

Tellaeche A, Pajares G, Burgos-Artizzu XP and Ribeiro A (2011) A computer vision approach for weeds identification through Support Vector Machines. *Applied Soft Computing* **11**, 908–915.

Tungmunnithum D, Boonkerd T, Zungsontiporn S and Tanaka N (2016) Morphological variations among populations of *Monochoria vaginalis* s.l. (Pontederiaceae) in Thailand. *Phytotaxa* **268**, 57–68.

van Evert FK, Fountas S, Jakovetic D, Crnojevic V, Travlos I and Kempenaar C (2017) Big Data for weed control and crop protection. *Weed Research* **57**, 218–233.

Velásco-Mejía A, Vallejo-Becerra V, Chávez-Ramírez AU, Torres-González J, Reyes-Vidal Y and Castañeda-Zaldivar F (2016) Modeling and optimization of a pharmaceutical crystallization process by using neural networks and genetic algorithms. *Powder Technology* **292**, 122–128.

Venora G, Grillo O, Shahin MA and Symons SJ (2007) Identification of Sicilian landraces and Canadian cultivars of lentil using an image analysis system. *Food Research International* **40**, 161–166.

Xinshao W and Cheng C (2015) Weed seeds classification based on PCANet deep learning baseline. In *2015 Asia-Pacific Signal and Information Processing Association Annual Summit and Conference (APSIPA)*. Hong Kong: Asia-Pacific Signal and Information Processing Association, pp. 408–415.