Identification of bean varieties according to color features using artificial neural network

A. Nasirahmadi[1]* and N. Behroozi-Khazaei[2]

[1] Department of Agricultural Machinery. Ferdowsi University of Mashhad. P.O. Box 91775. 1163 Mashhad, Iran
[2] Department of Agricultural Machinery. University of Kurdistan. Sanandaj, Iran

Abstract

A machine vision and a multilayer perceptron artificial neural network (MLP-ANN) were applied to identify bean varieties, based on color features. Ten varieties of beans, which were grown in Iran (Khomein1, KS21108, Khomein2, Sarab1, Khomein3, KS21409, Akhtar2, Sarab2, KS21205, and G11870) were collected. Six color features of the bean and six color features of the spots were extracted and used as input for MLP-ANN classifier. In this study, 1000 data sets were used, 70% for training, 15% for validating and 15% for testing. The results showed that the applied machine vision and neural network were able to classify bean varieties with 100% sensibility and specificity, except with Sarab1 with sensibilities of 100%, 73.3%, 60% for the training, validation and testing processes, respectively and KS21108 with specificities of 100%, 79% and 71%, respectively for the aforementioned processes. Considering total sensibilities of 100%, 97.33%, 96% and also specificities of 100%, 97.9% and 97.1% for training, validation and testing of beans, respectively, the ANN could be used as a effective tool for classification of bean varieties.

Additional key words: classification; *Phaseolus vulgaris;* image processing; machine vision.

Introduction

Beans (*Phaseolus vulgaris* L.) are legumes which are mostly produced and consumed as dry and green seeds in the world. Legumes are major sources of proteins and nutrients such as oil, fiber, starch, vitamins and mineral elements (Laurent *et al*., 2010). The analysis and classification of seeds are essential processes for the final step of crop production (Granitto *et al*., 2002). In today's competitive market, since consumers tend to use healthy and homogenous products, producers have to present products which are sorted according to physical characteristics like appearance, size, color, internal health and variety. Furthermore, identification of grain varieties helps farmers to use suitable grains for planting and marketing, because in this way the grains they sell would have the essential standards for marketers (Chen *et al*., 2010). It is very difficult to sort and separate products manually. Manual inspection involves labor intensive work and the decision made thereof can be very subjec-tive depending on the mood and condition of the person involved. Furthermore, this manual procedure can be very time consuming and inefficient especially when dealing with high production volumes. Considering these reasons, new methods like machine vision systems for grading and sorting have been introduced. Color is a property of products which is measured with machine vision and has been a great help in identifying objects for many years. The process of color classification in-volves extraction of useful information concerning the spectral properties of the object surfaces and discove-ring the best match from a set of known descriptions or class models to implement the recognition task (Sahin, 1997). Also color properties have been widely used for apple (*Malus × domestica*) quality evaluation, and mostly for defect detection (Leemans *et al*., 1998; Leemans & Destain, 2004) and tomato (*Solanum lycopersicum*) quality (Sarkar & Wolfe, 1985). Pydipati *et al*. (2006) used a machine vision technology for eva-luating the quality of grains and fruits. They distin-guished citrus diseases using the structural characte-

Abbreviations used: ANN (artificial neural network); HIS (hue, saturation, intensity); LM (learning algorithm); MLP (multi-layer perceptron); MSE (mean square error).

ristics of leaf color. Thorp & Dierig (2011) developed an image processing algorithm to extract information of flower features with hue, saturation and intensity (HSI) color space. They stated that their methods are simple, practical and quite generic, and could be easily used for many digital image processing applications which are based on color features. Their results showed that this system has the advantage of high accuracy and quick response in comparison with other measurement methods.

Classification of grain varieties by image analysis method has been investigated by several researchers. Shahin & Symons (2001) and Venora *et al*. (2007) used seed color and size with a flatbed imaging system to determine color grading, in Sicilian and Canadian lentil landraces and cultivars, finding that color features were good predictors. Also Venora *et al*. (2009) used color, shape and size of Italian bean landraces for identification of varieties. Similar experiments were performed by Chen *et al*. (2010) who identified five corn varieties. Color features have been extensively applied for grain quality evaluation. In a similar trend, Kılıc *et al*. (2007) developed a system based on size and color for classification of beans. Results showed that this method was able to correctly classify beans. Also Laurent *et al*. (2010) developed a computer vision system to study the hard-to-cook characteristic of bean with color and histogram features. The results showed that the color changes of beans are related to the hard-to-cook aspect. Guevara-Hernández & Gómez-Gil (2011) developed a machine vision system to classify wheat and barley grain kernels. Results of the system indicated that accuracies higher than 99% can be achieved when morphologic, color, and texture features are extracted from the grain kernels.

Color features and artificial neural network (ANN) classifier have been used by many researchers. Arribas *et al*. (2011) and Al Ohali (2011) used RGB color space and ANN classifier for sunflower (*Helianthus annuus*) leaves classification and grading of date fruit into three quality categories, respectively. Anami *et al*. (2011) classified different agriculture and horticulture products by neural network, finding that color and texture features were able to significantly detect normal and affected products. Abdullah *et al*. (2001) investigated the quality of oil palm (*Elgaeis guineensis*) with color features (HSI color space), finding that the vision system was able to correctly classify samples at a greater than 90% success rate. Huang (2007) also investigated the process for detecting *Phalaenopsis* seedling diseases based on texture and color features by neural network; results showed that diseases can be segmented and classified significantly using this method.

Considering the importance of identification of beans varieties for presenting to the market, this study aimed to classify ten bean cultivars using a method based on machine vision which extracted color features followed by ANN.

Material and methods

Varieties used and image acquisition

In this study, ten major commercial Iranian bean varieties were considered (Fig.1). Bean samples were acquired from the National Station of Bean Research in Khomein (Markazi province, Iran). The machine vision system was composed of three parts imaging, processing and information display unit. The imaging unit was responsible for recording the images and included a camera and a lighting system. To take pictures, a 576×720 pixels resolution Samsung camera (SCC-101 PA) was used. Low-level vision processing tasks need good lighting in the work environment. Hence, good and uniform illumination from external light source is essential for machine vision applications. In this research a fluorescent lamp (36 W) was used for lighting and a fiberglass sheet in front of it was utilized for producing homogenous lighting. The processing unit included three actions: transferring the camera signals to the computer, developing and extracting the image characteristics. Transferring the information from the camera to the computer was done by capture card (ACEDVio, Canopus). Image development was carried out to remove noise and non-homogenous lighting effects from images. The main purpose of the image processing was to extract features, which is done by the image processing science (González & Woods, 2007).

Image processing algorithm

Image analysis consisted of smoothing, image segmentation, features extraction, and data analysis. All pictures were captured in the RGB color space (Fig. 2a). Image process algorithm used Gaussian filter to remove the noise of picture and also image segmentation was designed to separate the region of interest from background. The blue color was selected for the
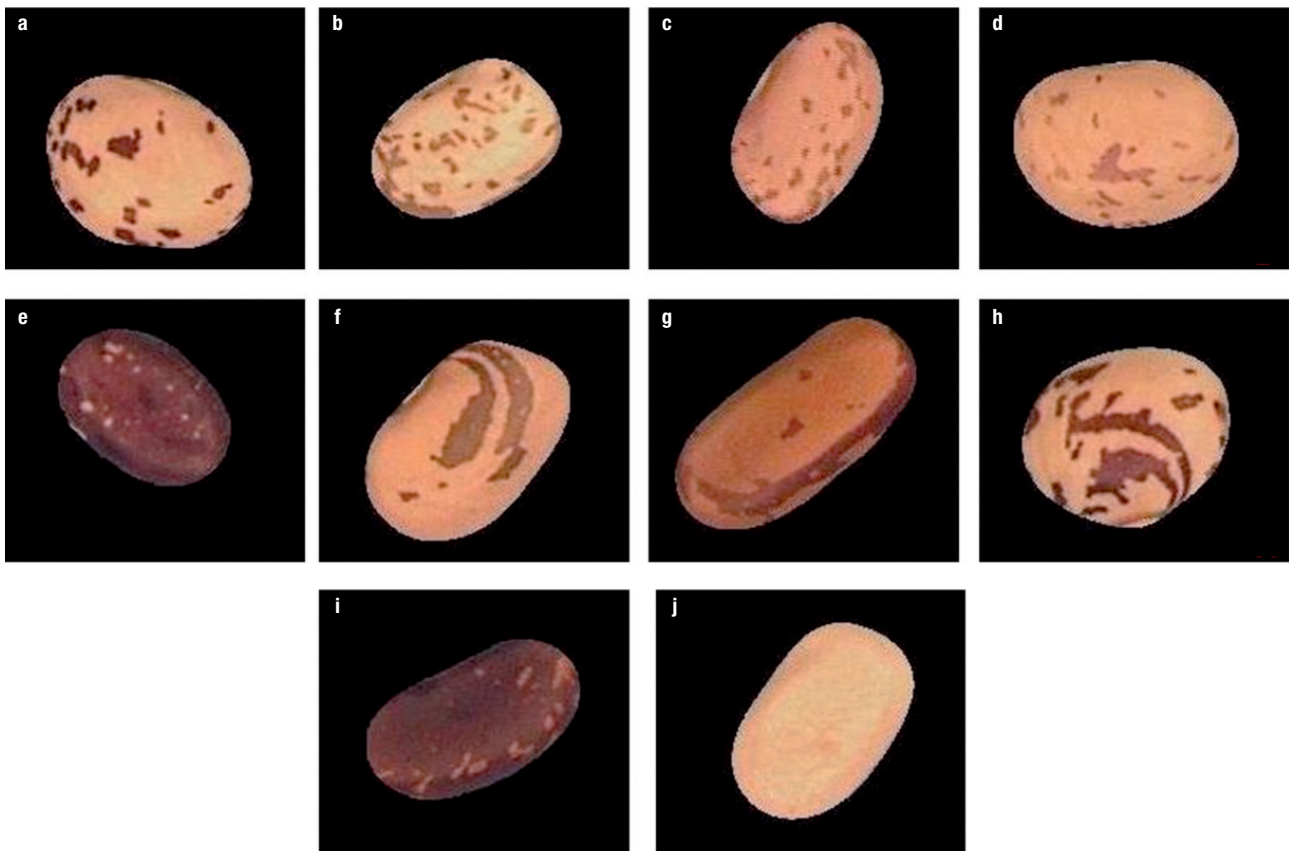
Figure 1. Color features of ten bean varieties: (a) Khomein1, (b) KS21108, (c) Khomein2, (d) Sarab1, (e) Khomein3, (f) KS21409, (g) Akhtar2 , (h) Sarab2, (i) KS21205, (j) G11870.

background, because of its clear color contrast with the beans. For this reason, the image of B channel was subtracted from R channel. In the R-B image, the object pixels (pixels of beans) are higher than zero because the red values of object pixels are higher than blue values (Fig. 2b). Then R-B gray image was converted into binary [0, 1] image with threshold and 1 assigned to the object and 0 assigned to the background (Fig. 2c). Then erosion and dilation orders with disk structure were used for smoothing of edge (Gonzalez & Woods, 2007). To extract the color features, the bina-

ry image was used for masking on RGB image (Fig. 2d). This image was used for two purposes; first, for extracting R_{mean}, G_{mean} and B_{mean}; second, for converting to HSI (Ruiz-Ruiza *et al.*, 2009), and extracting H_{mean}, S_{mean} and I_{mean}. Spot color features can increase the classification accuracy for the beans have high color similarity. The S channel in HSI was separated and gray image of S converted to binary with thresholding method and 1 assigned to the spots and 0 assigned to the plane of beans. Like beans, a similar method was used for masking on RGB and HSI to extract color features of spots (Fig. 3).
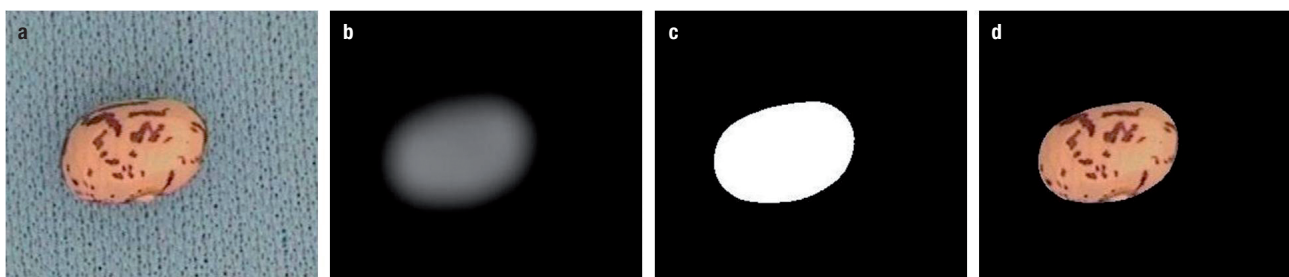

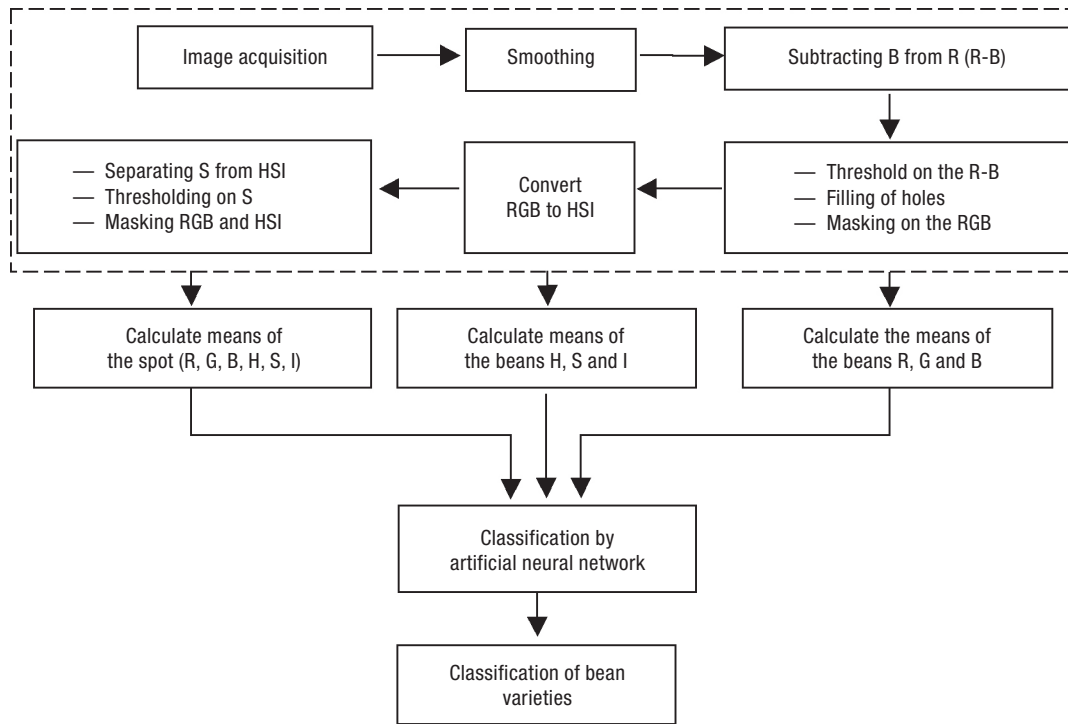**Figure 2.** RGB (a), R-B (b), binary of R-B (c), final image (d).

Figure 3. Algorithm of the coded program.

Artificial neural network

ANN is a non-linear modeling technique which can provide the classification abilities and lets the extension of computer vision technology into the area of color, at human brain levels of performance (Du & Sun, 2006). The learning procedure for developing a neural network can be either supervised or unsupervised. To obtain a unique solution for the network system, the sample size (number of input patterns) should be larger than or equal to the number of independent variables (sum of the number of weights and other possible variables) (Peres *et al.*, 2011). In this study, a multilayer perceptron (MLP) was employed as the modeling network for classification and a supervised learning algorithm (LM) was used for training the ANN. ANN model contained an input layer, an output layer and one or more hidden layers. In this study, the MLP network applied had four layers: an input layer, two hidden layers and an output layer. The input layer had a maximum of 12 neurons, one for each mean color parameter in the RGB and HSI color space for bean and spot. The two hidden layers had a variable number of hidden units, the optimal number of which was determined using a trial and error strategy. The output layer was equal to the number of classes, in this case 10 neurons

corresponding to the ten varieties. The MLP-ANN used TANSIG function in the hidden layers and linear function in the output layer. In this study, 1000 data sets were used, 700 (70%) for training, 150 (15%), for validating and 150 (15%) for testing. Due to the different ranges of each input, we normalized the inputs between [–1, 1]. Therefore, the ANNs were trained on the first subset (training set), and its performance was monitored using the second subset (validation set). Finally, the latter subset (test set) was used to check the predictive performance of the network since the data included in this subset was not used in the network development. In ANN, if the output layer equals [1 0 0 0 0 0 0 0 0 0] matrix, it means the Khomein1 variety, and if it equals [0 1 0 0 0 0 0 0 0 0] matrix it means the KS21108 variety and so on for other classes. In this method, each bean belongs to a class whose neurons have the highest value in the output layer. In this study the general sensibility and specificity of MLP-ANN classifier were calculated based on Peres *et al.* (2011):

$$\text{Sensibility (\%)} = \frac{N_c}{N_{tc}} \times 100 \qquad [1]$$

$$\text{Specificity (\%)} = \frac{N_b}{N_{tb}} \times 100 \qquad [2]$$

Table 1. Descriptive statistics (mean ± standard deviation) of color features measured in this research

Variety	Whole seed						Spot seed					
	R	G	B	H	S	I	R	G	B	H	S	I
Khomein1	190.51±5.60	127.01±6.29	88.03±4.39	0.07±0.08	0.56±0.01	0.76±0.02	142.56±8.62	62.96±7.68	28.52±5.18	0.05±0.04	0.81±0.02	0.49±0.03
KS21108	195.64±3.56	130.93±3.47	89.06±4.28	0.06±0.01	0.55±0.02	0.77±0.01	167.94±8.05	99.94±5.72	56.29±4.20	0.06±0.04	0.66±0.02	0.65±0.03
Khomein2	188.25±5.71	117.34±4.69	81.95±4.77	0.05±0.06	0.55±0.01	0.75±0.02	148.11±3.81	76.45±3.97	40.75±3.86	0.05±0.06	0.72±0.02	0.57±0.02
Sarab1	199.32±2.96	130.56±3.74	90.39±3.13	0.06±0.02	0.54±0.01	0.80±0.01	179.55±6.35	104.47±7.28	64.31±8.17	0.06±0.04	0.64±0.04	0.69±0.02
Khomein3	114.0±10.77	55.63±4.20	54.76±2.90	0.47±0.18	0.52±0.02	0.44±0.04	174.55±9.22	110.72±6.47	102.14±7.40	0.12±7.40	0.42±0.03	0.68±0.04
KS21409	179.57±9.46	109.47±7.27	77.35±3.64	0.05±0.02	0.57±0.02	0.71±0.04	136.87±10.6	79.96±7.55	52.51±4.23	0.05±0.04	0.62±0.02	0.53±0.04
Akhtar2	149.01±10.9	73.77±6.06	46.99±2.90	0.05±0.01	0.67±0.02	0.56±0.04	103.87±7.34	49.55±6.31	35.46±4.81	0.07±0.02	0.66±0.03	0.40±0.03
Sarab2	177.61±6.77	114.72±9.58	86.43±4.34	0.10±0.11	0.53±0.01	0.73±0.04	118.87±6.20	57.20±4.11	42.12±7.54	0.14±0.07	0.66±0.06	0.44±0.03
KS21205	89.00±6.03	46.11±2.89	45.16±1.57	0.29±0.09	0.51±0.02	0.35±0.02	140.96±6.47	83.14±6.72	71.33±6.02	0.06±0.01	0.51±0.03	0.55±0.02
G11870	207.97±5.32	156.79±10.4	106.17±5.06	0.08±0.01	0.52±0.04	0.84±0.01						

where N_c is the number of samples of a specific group correctly classified, N_{tc} is the total number of samples belonging to that specific group, N_b is the number of samples of a specific group classified as belonging to that group and N_{tb} is the total number of samples of any group classified as belonging to that specific group.

Results and discussion

Twelve color features were extracted from the bean images. Table 1 lists descriptive statistics of the bean color features computed from images. They are seed color components (whole seed and spot seed), in the RGB model (which stands for Red, Green and Blue channels) and HSI model (that stands for Hue, Saturation and Intensity channels, measured as mean grey levels). Twelve of the color feature values were plotted against bean varieties (Fig. 4). As can be seen in this

figure, it is clear that less overlapping features are more efficient in separating the bean varieties. There were complex relationships between bean varieties and color features. It was difficult to develop a simple model, like a linear model to predict bean varieties on these color features. Therefore all the 12 variables of seed color and spot information were inserted in the model of the MLP-ANN to identify the 10 varieties. The neural networks, with four layers, were trained and selected using two different data subsets (training and validation sets, respectively) and were then tested, with the experimental data not used in the two previous steps. Several neural networks were trained and the network which gave the minimum mean square error (MSE) of the validation subset was chosen (Table 2). The two hidden layers of the selected network had different numbers of neurons, being 20 and 10, respectively.

Finally the selected MLP-ANN (12:20:10:10) was used to evaluate the ability of this multivariate tech-

Table 2. Artificial neural network (ANN) results

Number of neurons		Mean square error (MSE)		
Firsh hidden layer	Second hidden layer	Train	Validation	Test
10	10	0.0075	0.0200	0.0192
10	15	0.0084	0.0176	0.0129
10	20	0.0003	0.0148	0.0122
10	25	0.0011	0.0182	0.0176
10	30	0.0091	0.0193	0.0177
20	**10**	**0.0025**	**0.0133**	**0.0119**
20	15	0.0074	0.0173	0.0147
20	20	0.0020	0.0178	0.0171
20	25	0.0012	0.0197	0.0204
20	30	0.0019	0.0218	0.0187

Bold values belong to the selected network with minimum MSE of the validation subset.
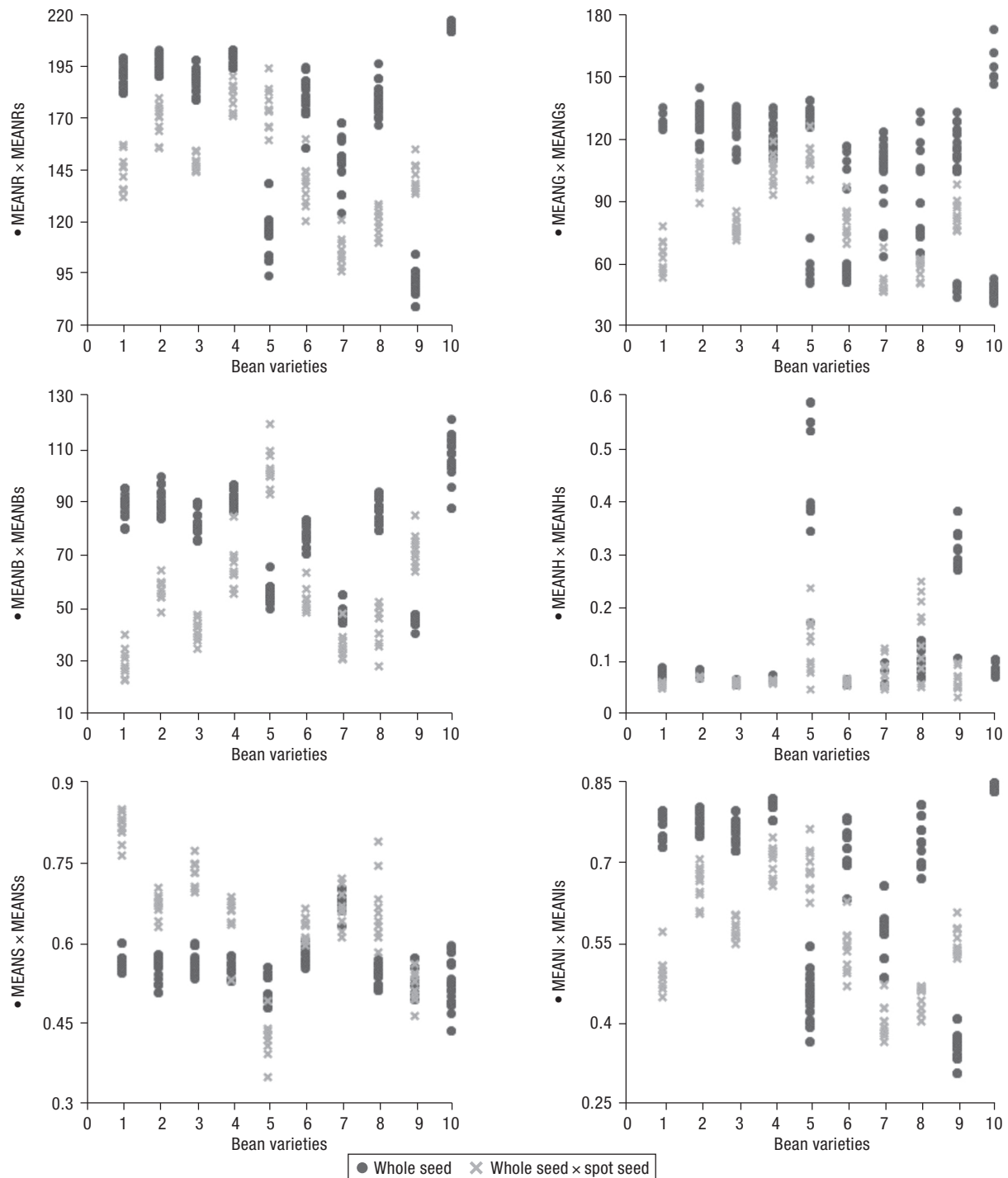
Figure 4. Relationship between color features (R: Red of whole seed, Rs: Red of spot seed, G: Green of whole seed, Gs: Green of spot seed, B: Blue of whole seed, Bs: Blue of spot seed, H: Hue of whole seed, Hs: Hue of spot seed, S: Saturation of whole seed, Ss: Saturation of spot seed, I: Intensity of whole seed, Is: Intensity of spot seed) and bean varieties: (1) Khomein1, (2) KS21108, (3) Khomein2, (4) Sarab1, (5) Khomein3, (6) KS21409, (7) Akhtar2, (8) Sarab2, (9) KS21205, (10) G11870.

nique for beans classification. The results obtained for the mentioned specific network are presented in Table 3. The results of this table showed that the selected neural network was able to correctly classify beans according to the varieties with a satisfactory sensibility and specificity of 100% (for the data included in the training,

Table 3. The artificial neural network (ANN) analysis: results for training, validation and test groups

Variety[1]	Data	Predicted beans membership[1]										Total	Sensibility (%)
		A	B	C	D	E	F	G	H	I	J		
A	Training	70	0	0	0	0	0	0	0	0	0	70	100
	Validation	15	0	0	0	0	0	0	0	0	0	15	100
	Test	15	0	0	0	0	0	0	0	0	0	15	100
B	Training	0	70	0	0	0	0	0	0	0	0	70	100
	Validation	0	15	0	0	0	0	0	0	0	0	15	100
	Test	0	15	0	0	0	0	0	0	0	0	15	100
C	Training	0	0	70	0	0	0	0	0	0	0	70	100
	Validation	0	0	15	0	0	0	0	0	0	0	15	100
	Test	0	0	15	0	0	0	0	0	0	0	15	100
D	Training	0	0	0	70	0	0	0	0	0	0	70	100
	Validation	0	4	0	11	0	0	0	0	0	0	15	73.3
	Test	0	6	0	9	0	0	0	0	0	0	15	60
E	Training	0	0	0	0	70	0	0	0	0	0	70	100
	Validation	0	0	0	0	15	0	0	0	0	0	15	100
	Test	0	0	0	0	15	0	0	0	0	0	15	100
F	Training	0	0	0	0	0	70	0	0	0	0	70	100
	Validation	0	0	0	0	0	15	0	0	0	0	15	100
	Test	0	0	0	0	0	15	0	0	0	0	15	100
G	Training	0	0	0	0	0	0	70	0	0	0	70	100
	Validation	0	0	0	0	0	0	15	0	0	0	15	100
	Test	0	0	0	0	0	0	15	0	0	0	15	100
H	Training	0	0	0	0	0	0	0	70	0	0	70	100
	Validation	0	0	0	0	0	0	0	15	0	0	15	100
	Test	0	0	0	0	0	0	0	15	0	0	15	100
I	Training	0	0	0	0	0	0	0	0	70	0	70	100
	Validation	0	0	0	0	0	0	0	0	15	0	15	100
	Test	0	0	0	0	0	0	0	0	15	0	15	100
J	Training	0	0	0	0	0	0	0	0	0	70	70	100
	Validation	0	0	0	0	0	0	0	0	0	15	15	100
	Test	0	0	0	0	0	0	0	0	0	15	15	100
Total	Training	70	70	70	70	70	70	70	70	70	70	700	100
	Validation	15	19	15	11	15	15	15	15	15	15	150	97.33
	Test	15	21	15	9	15	15	15	15	15	15	150	96
Specificity (%)	Training	100	100	100	100	100	100	100	100	100	100	100	
	Validation	100	79	100	100	100	100	100	100	100	100	97.9	
	Test	100	71	100	100	100	100	100	100	100	100	97.1	

[1] A: Khomein1, B: KS21108, C: Khomein2, D: Sarab1, E: Khomein3, F: KS21409, G: Akhtar2, H: Sarab2, I: KS21205, J: G11870.

validation and testing data sets). It should be mentioned that Sarab1 variety was an exception with sensibilities of 100%, 73.3%, 60% and the KS21108 variety with specificities of 100%, 79% and 71% for the training, validation and testing process, respectively. So, the MLP-ANN model proposed in the present study, together with the 12 color features used, is a reliable practical tool for classification of the beans variety.

Moreover, the analysis of the results presented in Table 3 shows that Sarab1 was misclassified as KS21108, probably due to the high overlapping in color features of these two varieties diagrams (Fig. 4). On the other hand, ANOVA results showed that there were no significant differences ($p \leq 0.05$) among six color features of Sarab1 variety and those of KS21108. A similar work was done by Shahin & Symons (2001) on identification

of seeds based on image analysis features, observing that neural classifiers achieved an overall accuracy of more than 90% for grading of lentil. Kılıc *et al*. (2007) investigated the quality of beans based on length, width, average, variance, skewness and kurtosis values by ANN classifier. Generally 90.6% of the beans were correctly classified; 99.3% of white beans, 93.3% of yellow-green damaged beans, 69.1% of black damaged beans, 74.5% of low damaged beans and 93.8% of highly damaged beans were correctly classified too.

In this research we used only 12 color features to identify 10 bean varieties by ANN classifier with a general sensibility of 100%, 97.33%, 96% and specificity of 100%, 97.9% and 97.1% for training, validation and testing, respectively. Applied ANN helped us use fewer variables than linear discriminant analysis (LDA) method of Venora *et al*. (2009) for classification. Therefore the image processing algorithm and ANN classification process was quite convenient and rapid.

As final conclusions, the MLP-ANN method confirmed to be a practical and effective model for bean varieties classification. MLP-ANN model contained an input layer, an output layer and two hidden layers. This model shows high overall sensitivity and specificity (higher than 96% and 97%, respectively). To obtain these results, 12 color features (R_{mean}, G_{mean}, B_{mean}, H_{mean}, S_{mean} and I_{mean}) of bean and their spots were used. This recommended method, compared with previous methods, is a good way for a rapid and cheap classification of bean varieties.

Acknowledgements

References

Abdullah MZ, Guan LC, Mohd Azemi BMN, 2001. Stepwise discriminant analysis for colour grading of oil palm using machine vision system. Food Bioprod Process 79: 223-231.

Al Ohali Y, 2011. Computer vision based date fruit grading system: design and implementation. KSU_CIS 23: 29-36.

Anami BS, Pujari JD, Yakkundimath R, 2011. Identification and classification of normal and affected agriculture/horticulture produce based on combined color and texture feature extraction. IJCAES I (III): 356-360.

Arribas JI, Sánchez-Ferrero GV, Ruiz-Ruiz G, Gómez-Gil J, 2011. Leaf classification in sunflower crops by computer vision and neural networks. Comput Electron Agr 78: 9-18.

Chen X, Xun Y, Li W, Zhang J, 2010.Combining discriminant analysis and neural networks for corn variety identification. Comput Electron Agr 71: S48-S53.

Du CJ, Sun DW, 2006. Learning techniques used in computer vision for food quality evaluation: a review. J Food Eng 72: 39-55.

Gonzalez R, Woods R, 2007. Digital image processing, 3rd ed. Prentice-Hall.

Granitto PM, Navone HD, Verdes PF, Ceccatto HA, 2002. Weed seeds identification by machine vision. Comput Electron Agr 33: 91-103.

Guevara-Hernández F, Gómez-Gil J, 2011. Machine vision system for classification of wheat and barley grain kernels. Span J Agric Res 9: 672-680.

Huang KY, 2007. Application of artificial neural network for detecting *Phalaenopsis* seedling diseases using color and texture features. Comput Electron Agr 57: 3-11.

Kılıç K, Boyacı IH, Köksel H, Küsmenoğlu I, 2007. A classification system for beans using computer vision system and artificial neural networks. J Food Eng 78: 897-904.

Laurent B, Ousman B, Dzudie1 T, Carl MF Mb, Emmanuel T, 2010. Digital camera images processing of hard-to-cook beans. J Eng Technol Res 2: 177-188.

Leemans V, Destain MF, 2004. A real-time grading method of apple based on features extracted from defects. J Food Eng 61: 83-89.

Leemans V, Magein H, Destain MF, 1998. Defects segmentation on 'Golden Delicious' apples by using colour machine vision. Comput Electron Agr 20: 117-130.

Peres AM, Baptista P, Malheiro R, Dias LG, Bento A, Pereira JA, 2011. Chemometric classification of several olive cultivars from Trás-os-Montes region (northeast of Portugal) using artificial neural networks. Chemom Intell Lab Syst 105: 65-73.

Pydipati R, Burks TF, Lee WS, 2006. Identification of citrus disease using color texture features and discriminant analysis. Comput Electron Agr 52: 49-59.

Ruiz-Ruiza G, Gómez-Gilb J, Navas-Gracia LM, 2009. Testing different color spaces based on hue for the environmentally adaptive segmentation algorithm (EASA). Comput Electron Agr 68: 88-96.

Sahin F,1997. A radial basis function approach to a color image classification problem in a real time industrial application. Master's thesis, Virginia Polytechnic Institute, Blacksburg.

Sarkar N, Wolfe RR, 1985. Feature extraction techniques for sorting tomatoes by computer vision. TASAE 28: 970-979.

Shahin MA, Symons SJ, 2001. A machine vision system for grading lentils. Can Biosyst Eng 43: 7-14.

Thorp KR, Dierig DA, 2011. Color image segmentation approach to monitor flowering in lesquerella. Ind Crops Prod 34: 1150-1159.

Venora G, Grillo O, Shahin MA, Symons SJ, 2007. Identification of Sicilian landraces and Canadian cultivars of lentil using an image analysis system. Food Res Int 40: 161-166.

Venora G, Grillo O, Ravalli C, Cremonini R, 2009. Identification of Italian landraces of bean (*Phaseolus vulgaris* L.) using an image analysis system. Sci Hortic 121: 410-418.