



## Hyperspectral imaging coupled with multivariate analysis and artificial intelligence to the classification of maize kernels

Fariba Alimohammadi<sup>1</sup>, Mansour Rasekh<sup>1\*</sup>, Amir Hosein Afkari Sayyah<sup>1</sup>, Yousef Abbaspour-Gilandeh<sup>1</sup>, Hamed Karami<sup>1</sup>†, Vali Rasooli Sharabiani<sup>1</sup>, Ambra Fioravanti<sup>2</sup>, Marek Gancarz<sup>3,4\*</sup>, Pavol Findura<sup>5,6</sup>, and Dariusz Kwaśniewski<sup>3</sup>

<sup>1</sup>Department of Biosystems Engineering, University of Mohaghegh Ardabili, Ardabil 56199-11367, Iran

<sup>2</sup>Institute of Sciences and Technologies for Sustainable Energy and Mobility, National Research Council (STEMS-CNR), Via Canal Bianco 28, 44124 Ferrara, Italy

<sup>3</sup>Faculty of Production and Power Engineering, University of Agriculture in Kraków, Balicka 116B, 30-149 Kraków, Poland

<sup>4</sup>Institute of Agrophysics, Polish Academy of Sciences, Doświadczalna 4, 20-290 Lublin, Poland

<sup>5</sup>Faculty of Agriculture, University of South Bohemia in České Budějovice, Studentská 1668, 370 05 České Budějovice, Czech Republic

<sup>6</sup>Faculty of Engineering, Slovak University of Agriculture in Nitra, Hlinku 2, 949 76 Nitra, Slovakia

Received November 3, 2021; accepted March 8, 2022

**Abstract.** Maize (*Zea mays*) is one of the key crops in the world, taking third place after wheat and rice in terms of cultivated area. This study aimed to demonstrate the potential of non-destructive hyperspectral imaging in the wavelength range of 400-1000 nm to discriminate between and classify maize kernels in three cultivars by using non-destructive hyperspectral imaging in the wavelength range of 400-1000 nm. Three cultivars of maize kernels were exposed to hyperspectral imaging with 20 replications. Predictor variables included 28 intensities of reflection wave for spectral imaging and 4 variables in terms of the weight, length, width, and thickness of a single kernel. The classification was successfully performed through Linear Discriminant Analysis and Artificial Neural Network methods, taking into account 32, 15, and 5 predictor variables. According to the results, Linear Discriminant Analysis with 32 predictor variables is characterized by a high degree of accuracy (95%). The most important predictor variables included the reflection wave intensity of the third peak, the wavelength intensity of 490 nm, the wavelength intensity of 580 nm, and the weight and thickness of a single kernel.

**Keywords:** maize, classification, hyperspectral imaging, artificial intelligence

### INTRODUCTION

Maize (*Zea mays*) is one of the key crops cultivated across the world, taking third place after wheat and rice in terms of area under cultivation (Harris *et al.*, 2007; Bajus *et al.*, 2019;

Khorsand *et al.*, 2020; Kapela *et al.*, 2020). The importance of this crop and its considerable area under cultivation is associated with its compatibility with various climatic conditions, making it one of the main food sources in temperate, tropical, subtropical, and humid regions. In Iran, this plant is the most important crop with the highest area under cultivation after wheat, rice, and barley (Harris *et al.*, 2007).

There are different cultivars of maize kernels, which makes their classification imperative to ensure high quality. Moreover, ensuring quality is fundamental to the development of sustainable agriculture; hence, techniques such as drying, refrigeration, and the application of an edible coating are necessary to maintain the quality of agricultural products. On the other hand, effective and efficient methods, usually used in seed and seedling preparation and improvement centres, silos, and mechanized warehouses, should be developed to evaluate and classify the quality of these crops (Benthien *et al.*, 2020; Pekel *et al.*, 2020). Traditional methods used to classify different cultivars of maize kernels include the use of fluorescence imaging, protein electrophoresis, and molecular markers of deoxyribonucleic acid (DNA), which are usually time-consuming

\*Corresponding authors e-mail: rasekh@uma.ac.ir, m.gancarz@ipan.lublin.pl

† Senior author: hamedkarami@uma.ac.ir

and complicated (Cheng *et al.*, 2014; Cheng and Sun, 2015; Kamruzzaman *et al.*, 2011; Gowen *et al.*, 2007; Dziki *et al.*, 2020; Tan *et al.*, 2020; Alsalem *et al.*, 2021).

In recent times, spectroscopic techniques have been introduced and used extensively to assess food quality and safety. These analysis methods are generally non-destructive and can be used to analyse physical-chemical information in both raw and processed foods. In particular, spectroscopy in the range of visible and near-infrared (NIR) radiation and attenuated total reflection (ATR) have been used to evaluate the types of a particular food and to confirm the geographical origin of many products (Wang *et al.*, 2015; Anjos *et al.*, 2015).

Furthermore, hyperspectral imaging (HSI) is an emerging technology that combines imaging or machine vision techniques with spectroscopy in a single system, thereby obtaining a sample image from visible to near-infrared wavelengths (Jackman *et al.*, 2008; Valous *et al.*, 2009). It is possible to obtain spatial and spectral information simultaneously using HSI (El-Masry *et al.*, 2011 and 2012; Kamruzzaman *et al.*, 2012; Wu *et al.*, 2012). This method is widely used in agriculture, including: i) the detection of cereal fungal infection (Bauriegel *et al.*, 2011), ii) the analysis of the hardness of maize kernels (Delwiche *et al.*, 2013), and iii) seed variety identification and classification (Zhang *et al.*, 2012; Yang *et al.*, 2015; Liu *et al.*, 2021).

Williams *et al.* (2016) evaluated maize kernels in three classes of hardness (hard, medium, and soft) using NIR hyperspectral imaging. Pixel-wise and object-wise approaches were used to group the maize kernels according to hardness. Pixel-wise classification assigned a class to each pixel of the single kernel and did not lead to acceptable results due to a high level of classification errors. However, more accurate results were obtained (with a sensitivity and specificity of 0.75 and 0.97) using a predefined threshold and classifying all of the kernels based on the number of correctly predicted pixels. Object-wise classification was performed using two featured extraction methods of score histograms and mean spectra. Score histograms performed better concerning the classification of hard kernels (sensitivity and specificity of 0.93 and 0.97), while mean spectra performed better concerning the classification of medium kernels (sensitivity and specificity of 0.93 and 0.93). Both of the featured extraction methods may be recommended in the classification of maize kernels at a production scale (Williams and Kucheryavskiy, 2016).

The selection of wavelength and the conversion of features are the two major methods used to reduce the dimensionality of hyperspectral data. Wavelength selection is the process of finding the most representative wavelength of the original wavelengths and involves uninformative variable elimination (UVE) and sequential projection algorithms (SPA) (de Araújo Gomes *et al.*, 2015; Hu *et al.*, 2016). Wavelength selection can be used to minimize spectral redundancy to improve classification accuracy. The integration of feature conversion and wavelength selection

(known as feature conversion/reduction) has been widely used in the non-destructive examination of the quality of agricultural products. Sun *et al.* (2016) reported the correct discrimination rate of 98.33% in discriminating between three types of black beans through the fusion of spectral and image information extracted from 13 optimal wavelengths under the first principal component for the shoot. Zhang *et al.* (2012) used a system based on visible and near-infrared (VIS/NIR) hyperspectral images and the near-infrared spectrum to discriminate between maize kernel cultivars by combining textural and spectral features. They used back-propagation neural network classification models and least-squares support vector machines, leading to the best correct discrimination rate of 98.89% for the predicted models. According to this study, the HIS model with data fusion was very suitable for the analysis of the non-destructive classification of different maize kernel cultivars.

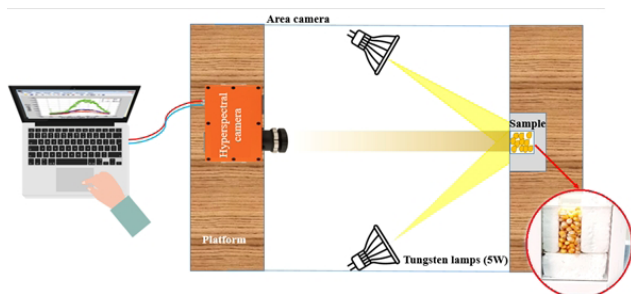
In recent times, some researches have also used deep ANN-based relevant latent representation learning for hyperspectral imaging (HI) classification (Sellami and Tabbone, 2022), active deep learning for HI classification with uncertainty learning (Lei *et al.*, 2021), and proxy-based deep learning framework for spectral-spatial HI classification: efficient and robust (Yuan *et al.*, 2021).

The objectives of this study included the examination of the ability to identify kernels of three maize cultivars based on a hyperspectral imaging technique and the ability to discriminate between these cultivars through the step-wise method and to the elimination of some variables using LDA and ANN methods. Also, in this research, spectral data were obtained based on the level of grain mass, which may be used to facilitate imaging operations, and can be developed in practical applications.

## MATERIALS AND METHODS

The maize kernels of three single-cross cultivars: SC 703, SC 704, and SC 705 were obtained from the Ardebil Agricultural and Natural Resources Research Centre in Parsabad-Moghan, Iran. The kernels were then transferred to the Biophysical Properties Laboratory of the Biosystems Engineering Department of the University of Mohaghegh Ardabili, Iran. Three samples of each kernel with a weight of 20 g were placed at a temperature of 105°C for 24 hours in a laboratory oven to determine the initial water content of the maize kernels. After recording the dry weight of the samples, the initial water content of the maize kernels was calculated to be 10.5%. A total number of 60 samples were used with 20 replications for each sample to discriminate between 3 maize cultivars. Also, a digital scale with an accuracy of 0.001 g was used to measure the weight of the kernels.

The system for collecting hyperspectral images is shown in Fig. 1. For the most part it consists of a specially designed light-insulated box in which an image sensor, a camera equipped with a charge coupled device (CCD),



**Fig. 1.** Schematic of hyperspectral system and sample chamber.

a light source and the sample which are located inside the box and a computer outside the box supported by access to an information and control system (Spectrometer-OPTC-1000, IRAN). The finished wooden and rectangular box has two supports, one for the camera that can move during imaging and the other one fixed for the sample chamber. The distance from the camera lens to the sample is 1 m and imaging is obtained in a horizontal direction. The samples (kernel mass) were placed in a small wooden chamber with thin glass on the side where the light is located. As a light source, two tungsten lamps (5 W) were used to illuminate the samples and placed with an angle of 45° to avoid undesired reflections. The camera used is a line scanning camera that can capture an image with maximum dimensions in pixels of 720 x 440 mm and a minimum width of 0.2 mm in a period of time ranging from 10 to 110 ms. The CCD of the camera may be used to reveal radiation over a range of 190-1150 nm with a resolution of 1 nm. Nevertheless, in the case of the ranges of 190 to 400 and 1000 to 1150 nm a high degree of noise is present, hence only radiation wavelengths of between 400 and 1000 nm will be considered to create a suitable and stable image.

For each mass sample, 300 scans were taken over a period of 300 ms, and 1279 wavelengths for each hyperspectral reflectance image. In this way, a special block was prepared from an image with dimensions of 1279 x 300 x 720 mm. After each imaging, 20 kernels from the sample mass were randomly selected and weighed using a digital scale. After that, their dimensions (length, width, and thickness) were measured using a precision calliper. The variables were selected based on an initial statistical evaluation. In Table 1 the weights and dimensions of the seed samples for the 3 maize cultivars are summarized along with their means and standard deviations.

A spectral range of 740-420 nm was considered for the analysis, from which wavelengths assumed to facilitate the discrimination between maize cultivars were selected. The wavelengths of the spectral samples were selected on the basis that the set of spectral data could cover an average of the total range of the extractable spectra ranging from 400 to 1000 nm. For this purpose, 28 spectral wavelengths were selected in this range. Based on an analysis of the obtained spectral data, wavelengths of 420, 430, 435 (first peak of the curve), 450, 460, 480, 490, 500, 510, 520, 540, 546 (second peak of the curve), 550, 570, 578 (third peak of the curve), 580, 600, 610, 630, 640, 660, 670, 690, 700, 710, 720, 740 and 750 (fourth peak of the curve), were effective as significant wavelengths in the prediction model. Therefore, these 28 wavelengths were used in the next stages of processing.

The collected data were analysed using linear discriminant analysis (LDA) and artificial neural network (ANN) analysis. LDA is commonly used as a dimension reduction tool to find a linear combination of new variables from the original data. LDA aims to maximize the between-class variance and minimize the within-class variance. ANN is one of the most common methods of artificial intelligence. One of the most important applications of neural networks is pattern recognition. Pattern recognition may be implemented using a feed-forward neural network trained in the same way. A network consisting of three input, output, and hidden layers was used to classify the maize cultivars. Performance was calculated using cross-entropy (CE%). The following equations may be used to determine the number of nodes in the hidden layer (Rasekh and Karami, 2021a):

$$\leq 2N_i + 1, \tag{1}$$

$$\left( \frac{N_i + N_o}{2} \right), \tag{2}$$

$$\frac{2N_i}{3}, \tag{3}$$

$$\sqrt{N_i N_o}, \tag{4}$$

$$2N_i. \tag{5}$$

Based on these equations, ( $N_i$ ) and ( $N_o$ ) indicate the number of nodes in the input and output layers, respectively. Overall, 60% of the data were used for training and 40% for validation and testing. The confusion matrix was used to select the best model; the confusion matrix calculates the prediction accuracy of models. This matrix compares the predicted and actual values. The columns and rows in the confusion matrix correspond to the predicted and actual classes, respectively. The diagonal cells in the confusion matrix correspond exactly to the classified observations

**Table 1.** Weight data and dimensions of samples of measured corn cultivars

Repeat	Weight (g)			Length (mm)			Width (mm)			Thickness (mm)		
	SC 703	SC 704	SC 705	SC 703	SC 704	SC 705	SC 703	SC 704	SC 705	SC 703	SC 704	SC 705
Min	0.224	0.277	0.389	8.900	11.04	10.77	7.705	7.335	7.885	4.765	3.890	5.985
Max	0.241	0.291	0.288	10.20	10.23	10.31	6.500	7.560	7.165	4.250	4.135	4.980
Mean	0.243	0.266	0.283	10.43	10.88	10.35	6.755	6.969	7.137	4.456	4.205	4.834
SD	0.026	0.035	0.053	0.674	0.590	0.580	0.637	0.541	0.512	0.455	0.477	0.671

(Rasekh and Karami, 2021a; Rasekh *et al.*, 2021b), while the parameters of sensitivity, specificity, accuracy, and precision were used to analyse the performance of the system (Rasekh and Karami, 2021a; Tatli *et al.*, 2022):

$$\text{Sensitivity} = \frac{TP}{TP + FN}, \quad (6)$$

$$\text{Specificity} = \frac{TN}{TN + FP}, \quad (7)$$

$$\text{Precision} = \frac{TP}{TP + FP}, \quad (8)$$

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FN + FP}, \quad (9)$$

$$\text{AUC} = \frac{\text{Sensitivity} + \text{Precision}}{2}. \quad (10)$$

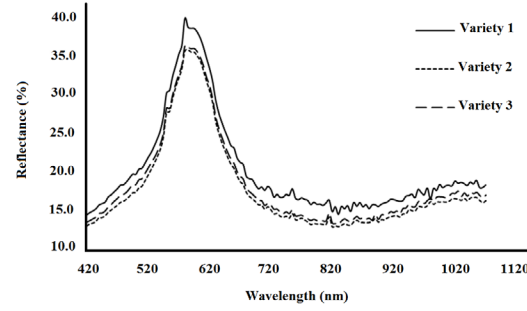
In the above mentioned equations true positive ( $TP$ ), true negative ( $TN$ ), false positive ( $FP$ ) and false negative ( $FN$ ) are used to calculate the specified parameters and all values are dimensionless. The parameter of specificity reflects the proportion of the samples that received a negative result (true negative rate). The parameter of precision, which is also known as a positive predictive value, provided an indication of the repeatability of the data as indicated by the closeness of data clustering in data plots. The parameter of sensitivity is defined as the ratio of  $TP$  samples to total  $TP$  and  $FN$  samples. The Precision ( $P$ ) is defined as the ratio of  $TP$  samples to the total  $TP$  and  $FP$  samples. The area under the curve ( $AUC$ ) is a measure of the ability of a classifier to distinguish between classes and is used as a summary of the ROC curve.

LDA and ANN analyses using the SPSS statistical program (SPSS, Chicago, IL) and Matlab® (ver. 2014a) software (Mathworks, Inc., Natick, MA, USA) were used to classify the cultivars. The conventional method considered all of the variables in the analysis, while the stepwise method eliminated some variables and included only those with the highest impact.

## RESULTS AND DISCUSSION

The reflectance of the raw spectra is shown in Fig. 2. According to the figure, all of the significant wavelengths were between 420 to 1070 nm. Also, the differences between the three varieties of maize may easily be observed. The highest reflectance was observed in the region around the red band (up to 35 to 40%), while in the near-infrared band the reflection is lower. However, there were some fluctuations in all of the trends of the curve obtained, this may be due to some of the noise which occurred during data acquisition by the hyperspectral camera and environmental illuminations.

This analysis was performed based on 32 predictor variables, the 28 intensities of the reflectance spectrum and the 4 variables of weight, length, width, and thickness of a single kernel. Also the analysis was performed based on 5 predictor variables, it eliminated variables that did not have a significant effect on the potential of the process to



**Fig. 2.** Percentage of reflectance in three different maize varieties.

discriminate between cultivars and considered only five variables that could significantly discriminate between the three maize cultivars. The variables with the greatest importance as predictor variables included the reflection wave intensity of the third peak, the wavelength intensity at 490 nm, the wavelength intensity at 580 nm, and the weight and thickness of a single kernel. Finally, an analysis was conducted with 15 predictor variables, including 11 intensities of reflectance spectra and 4 variables of the weight, length, width, and thickness of a single kernel.

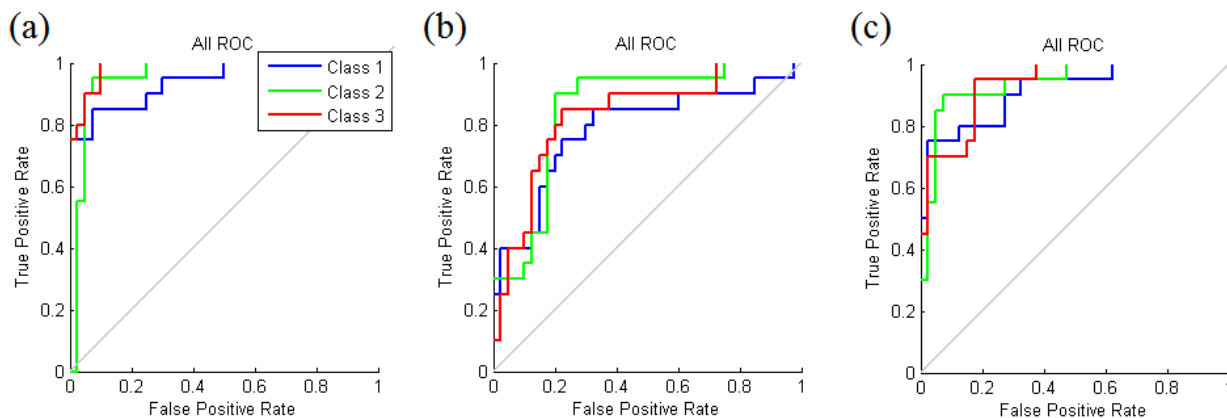
According to the data of 32 predictor variables, 32 neurons were considered for the input layer, and 3 neuron layers were considered for the output layer given the classification of 3 maize cultivars. The number of hidden layer nodes was determined using Eqs 1 to 5. Therefore, the neural network with a structure of 32-17-3 had the highest accuracy in the classification of 3 maize cultivars.

In the case of 5 predictor variables, 5 and 3 neurons were considered for the input and output layers, respectively, and the hidden layer was considered to be equal to 8 layers. Therefore, the neural network with a structure of 5-8-3 had the highest degree of accuracy in the classification of 3 maize cultivars. Finally, a network with 15 predictor variables was

**Table 2.** Statistical results of the artificial neural network classification models using 32, 5, and 15 predictor variables

Topology	Stage	Samples	Accuracy	Error*	CE**
32-17-3	Training	36	100	0	1.27
	Validation	12	91.7	8.3	3.23
	Testing	12	50	50	3.41
	Overall	60	88.3	11.7	0.26
5-8-3	Training	36	75	25	0.86
	Validation	12	75	25	1.28
	Testing	12	58	41	1.28
	Overall	60	71.7	28.3	0.48
15-10-3	Training	36	88.9	11.1	0.87
	Validation	12	66.7	33.3	1.53
	Testing	12	75	25	1.54
	Overall	60	81.7	18.3	0.34

\*Percentage error indicates the fraction of samples that are misclassified. A value of 0 means no misclassifications, 100 indicates maximum misclassifications; \*\*Minimizing cross-entropy results in a favourable classification. Lower values are better. Zero means no error.



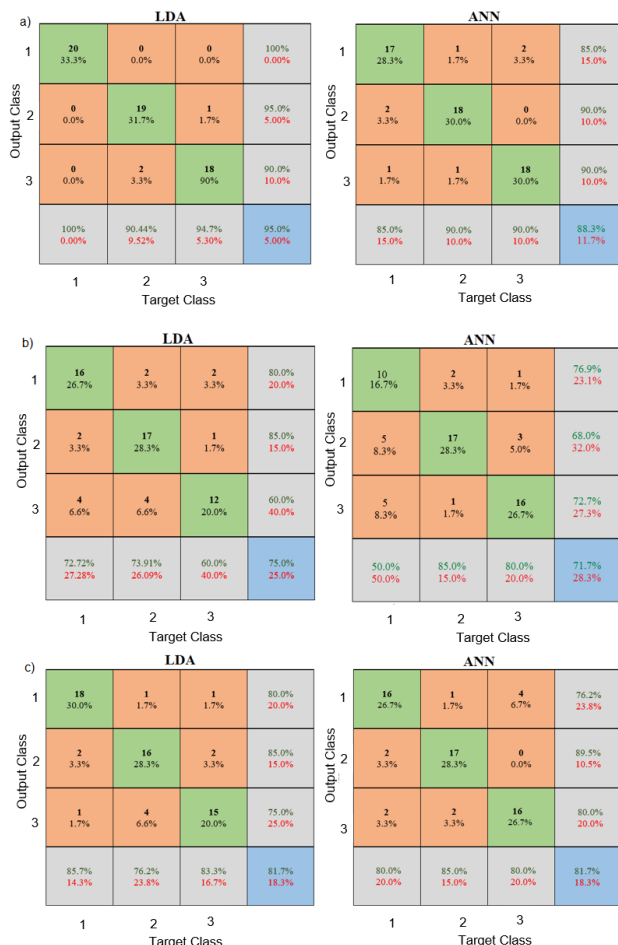
**Fig. 3.** Receiver operating characteristics (ROC) curve showing the true positive (sensitivity; y-axis) and false-positive (specificity; x-axis) rates of the classification Model ANN using, a) 32, b) 5, and c) 15 predictor variables.

chosen, the neurons considered for the input, hidden and output layers were 15, 10 and 3, respectively. A neural network with a structure of 15-10-3 had the highest degree of accuracy concerning the classification of 3 maize cultivars, according to which the CE was equal to 0.34% for overall data, its results are shown in Table 2. According to the results, the highest degree of accuracy was obtained with 32 predictor variables and the amount of error and CE were 11.7 and 0.26, respectively. Furthermore, the value of the receiver operating characteristic (ROC), was very sensitive for the classification of the 3 maize cultivars (true positive rate; 0.883). Table 2 shows that the inputs data had a high degree of accuracy (88.3%) overall for the classification of 3 maize cultivars. This model wasn't under or over fitted because the training CE values were lower than the testing stage, which indicates that they produced a high performance.

In a ROC curve the true positive rate (Sensitivity) is plotted as a function of the false positive rate (100-Specificity) for different cut-off points. Each point on the ROC curve represents a sensitivity/specificity pair corresponding to a particular decision threshold. A test with perfect discrimination (no overlap in the two distributions) has an ROC curve that passes through the upper left corner (100% sensitivity, 100% specificity). Therefore, the closer the ROC curve is to the upper left corner, the higher the overall accuracy of the test. ROC analyses showing that maize kernel classification had large areas under the curve (with high F values) provided favourable indications of the effective classification of maize kernels.

Furthermore, in their receiver operating characteristic (ROC) curve (Fig. 3a), it may be observed that the 3 classifications with 32 variable predictors had a very high sensitivity (true positive rate; 0.96). Figure 3b shows the performance curve of the neural network model based on 5 predictor variables. The real positive rate and the false positive rate for classifying the 3 maize cultivars were 0.717 and 0.811, respectively. By comparing the accuracy of the results those obtained with this model with respect

to the one based on 32 predictor variables, the results were lower. The reason for this may be related to the exclusion of predictor variables which leads to a reduction in accuracy. Finally, Fig. 3c provides an indication of the receiver operating characteristics curve, based on 15 predictor variables. The real positive rate and the false positive rate for classifying 3 maize cultivars were 0.819 and 0.878, respectively.



**Fig. 4.** Confusion matrix for the classification of three maize cultivars using; a) 32, b) 5, and c) 15 predictor variables.

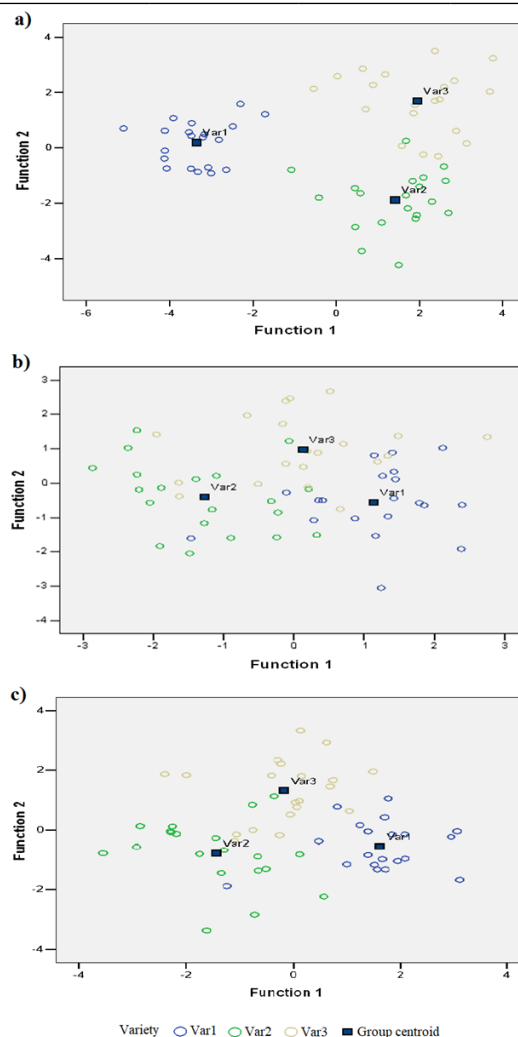
**Table 3.** Performance parameters for classification of three maize cultivars using 32, 5 and 15 predictor variables

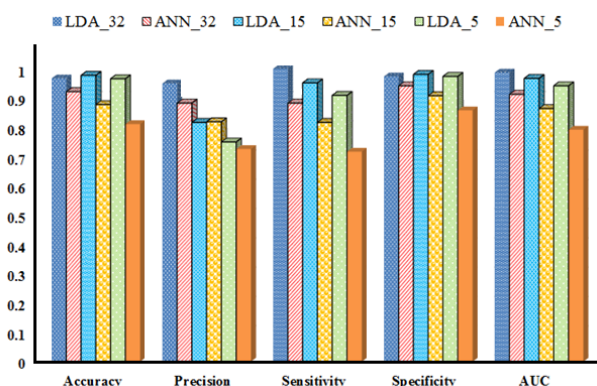
Models	Predictor variable	Variety	Accuracy	Precision	Sensitivity	Specificity	AUC	
LDA	32	Var1	1.000	1.000	1.000	1.000	1.000	
		Var2	0.950	0.950	1.000	0.974	0.987	
		Var3	0.950	0.900	1.000	0.949	0.974	
	5	Average per class		0.967	0.950	1.000	0.974	0.987
			Var1	0.952	0.800	0.727	0.979	0.853
			Var2	0.986	0.850	1.000	0.984	0.992
		Average per class		0.962	0.600	1.000	0.959	0.980
			Var1	0.966	0.750	0.909	0.974	0.942
			Var2	0.976	0.900	0.857	0.989	0.923
	15	Average per class		0.976	0.750	1.000	0.974	0.987
			Var1	0.978	0.817	0.952	0.981	0.967
			Var2	0.900	0.850	0.850	0.925	0.888
Average per class			0.933	0.900	0.900	0.950	0.925	
		Var1	0.922	0.883	0.883	0.942	0.913	
		Var2	0.783	0.769	0.500	0.925	0.847	
ANN	32	Var1	0.817	0.680	0.850	0.800	0.740	
		Var2	0.817	0.680	0.850	0.800	0.740	
		Var3	0.833	0.727	0.800	0.850	0.789	
	5	Average per class		0.811	0.726	0.717	0.858	0.792
			Var1	0.850	0.762	0.800	0.875	0.818
			Var2	0.917	0.895	0.850	0.950	0.922
		Average per class		0.867	0.800	0.800	0.900	0.850
			Var1	0.878	0.819	0.817	0.908	0.864
			Var2	0.878	0.819	0.817	0.908	0.864

The classification, based on 15 predictor variables, has shown a lower accuracy than that of 32 predictor variables, but higher than that of 5 predictor variables.

Figure 4 shows the confusion matrix obtained from the classification of 3 maize cultivars using LDA and ANN methods. As shown in Fig. 4a, the accuracy of LDA and ANN methods with 32 variable predictors were 95% and 88.3%, respectively. According to Fig. 4, there are two numerical digits in the upper and lower sides within the diagonal cells, indicating the number and the correct discrimination rate, respectively. For example, in the first cell corresponding to cultivar 1, 33.3% of the total observed datasets were classified correctly, and since there were no instances of incorrect classification, the accuracy of classification in the LDA method was 100% for cultivar 1. Also, in Fig. 4b, with 5 variable predictors, the accuracy of the LDA and ANN methods was 75 and 71.7%, respectively. In addition, these values with 15 predictor variables, were 81.7% and 81.9%, respectively.

Considering Equations 6 to 10, the performance parameters of the LDA and ANN methods in classifying the three maize cultivars are summarized in Table 3. The confusion matrix calculates the performance parameters of the discriminant models (Karami *et al.*, 2020a, 2020b; Rasekh and Karami, 2021b; Combrzyński *et al.*, 2021). According to the results of Table 4, it may be seen that the accuracy of the LDA method with different predictor variables is more than 96%, while this value from the ANN method ranged from 81 to 92%. According to the results of Table 3, the accuracy of data classification was 95 and 85% using the LDA and ANN methods, respectively.

**Fig. 5.** Canonical discriminant functions of LDA with; a) 32, b) 5, and c) 15 predictor variables.



**Fig. 6.** Average performance parameters of different models in the classification of 3 maize cultivars based on the number of predictor variables.

It is evident that the LDA method is more accurate than the ANN one in the classification of three maize cultivars. Similar results have been reported by other researchers using LDA and ANN methods (Karami *et al.*, 2020c; Khorramifar *et al.*, 2021; Karami *et al.*, 2021; Rasekh *et al.* 2021a).

According to Fig. 5a, the three maize cultivars were correctly identified according to the canonical discriminant function analysis. As shown, cultivars 2 and 3 had some overlap, while cultivar 1 did not have any overlap with the other two cultivars. In addition, according to Fig. 5b, in canonical discriminant functions, the three maize cultivars were classified using the LDA method with a 75% accuracy. As may be seen in Fig. 5b, the centres of the groups have equal distances from each other, and the samples of all three groups overlap. Therefore, this method provides less accuracy in the identification of maize cultivars than other methods. Finally, according to Fig. 5c, in canonical discriminant functions, three maize cultivars were classified with 81.7% accuracy using the LDA method, and there was some overlap among all three samples.

Figure 6 shows the test results for the accuracy of the classification of 3 maize cultivars based on the number of predictor variables. The statistical results were calculated using equations (6-10) and their mean was reported. Among the models tested, the LDA based on the 32 predictor variables provided the best performance in classifying 3 maize cultivars, in particular, the highest sensitivity, precision, and AUC. Based on the obtained results, the overall accuracy of all of the LDA models was higher than that of the ANN methods and they can be used to classify maize cultivars with great accuracy.

Williams *et al.* (2016) evaluated corn kernels at three hardness levels (hard, medium and soft) with the help of hyperspectral imaging in the NIR range for classification. They achieved a level of accuracy of 0.75 and 0.97, respectively, with a Pixel-wise approach in sensitivity and specificity. Object-wise classification was performed using two methods for feature extraction – score histograms and mean spectra. The model based on score histograms

performed better for hard kernel classification (sensitivity and specificity of 0.93 and 0.97), while that of mean spectra gave better results for medium kernels (sensitivity and specificity of 0.95 and 0.93). In this study, imaging was performed on a single grain.

Zhang *et al.* (2012) used a system based on an observable HSI model and near-infrared spectrum to detect corn seed cultivars by combining textural and spectral characteristics. To do this, they used back-propagation neural network classification models and a partial square support vector machine (LS-SVM) and also the best detection accuracy of 98.89% was obtained for the predicted models. This study showed that the HSI model combined with the experimental data was very suitable for the non-destructive classification analysis of different maize grain cultivars.

In the present study, the classification of 3 maize seed cultivars was performed using hyperspectral imaging and LDA and ANN methods using three analytical methods, i.e. considering the total of 32, 5 and 15 predictor variables. In this study, a direct linear diagnostic analysis method, which considered the highest number of predictor variables, produced the highest level of accuracy in distinguishing the maize cultivars from each other (95%) and in reducing the number of predictor variables, the ANN method was found to be more accurate than the LDA method (85 and 88.3%).

## CONCLUSIONS

1. This study has successfully shown that a hyperspectral imaging analysis may be used to rapidly classify maize cultivars if coupled with LDA and ANN methods.

2. The LDA and ANN methods, with the use of 32 predictor variables, offered an accuracy of 95 and 88.3%, respectively. Furthermore, in the stepwise method and by eliminating some variables, the LDA and ANN methods had an accuracy of 75 and 71.7%, respectively, in the case of the 5 predictor variables. Finally, based on 15 predictor variables, the accuracy of both the LDA and ANN models was 81.7%. The LDA method is also more accurate than the ANN method. Therefore, it is reasonable to surmise that the discrimination accuracy decreases with a lower number of predictor variables. In this study, the LDA method had the highest level of accuracy in discriminating between the maize cultivars considering the 32 predictor variables. The most important predictors for discriminating between cultivars included the reflection wave intensity of the third peak, the wavelength intensity at 490 nm, the wavelength intensity at 580 nm, the weight of a single kernel, and the thickness of a single kernel. The length and width of a single kernel did not affect the classification of the cultivars significantly.

3. Hyperspectral imaging technology can be used successfully along with the measurement of seed weight, the dimensions of the seed, seedling preparation, improvement centres, silos, mechanized warehouses, and in places where it is necessary to sort and separate maize kernel cultivars.

4. This study has shown that a hyperspectral imaging system can rapidly classify maize cultivars by using LDA methods. Due to the great number of advantages (non-destructive, rapid, real-time) hyperspectral imaging systems can be widely applied to classify different cultivars of seeds.

**Data Availability Statement:** The datasets used and/or analysed during the current study are available from the corresponding author upon reasonable request.

**Conflicts of interest:** The authors declare that there is no conflict of interest regarding the publication of this paper.

#### REFERENCES

- Alsalem M., Salehi A., Zhao J., Rewald B., and Bodner G., 2021.** Combining image analyses tools for comprehensive characterization of root systems from soil-filled rhizobox phenotyping platforms. *Int. Agrophys.*, 35(3), 257-268, <https://doi.org/10.31545/intagr/143121>
- Anjos O., Campos M.G., Ruiz P.C., and Antunes P., 2015.** Application of FTIR-ATR spectroscopy to the quantification of sugar in honey. *Food Chem.*, 169, 218-223, <https://doi.org/10.1016/j.foodchem.2014.07.138>
- Bajus P., Mraz M., Rigo I., Findura P., Fürstenzeller A., Kielbasa P., and Malaga-Tobola U., 2019.** The influence of drying temperature and moisture of corn seeds planted on their damage. *Agric. Engin.*, 23(1), 5-12, <https://doi.org/10.1515/agriceng-2019-0001>
- Bauriegel E., Giebel A., and Herppich W.B., 2011.** Hyperspectral and chlorophyll fluorescence imaging to analyse the impact of *Fusarium culmorum* on the photosynthetic integrity of infected wheat ears. *Sensors*, 11(4), 3765-3779, <https://doi.org/10.3390/s110403765>
- Cheng J.-H. and Sun D.-W., 2015.** Rapid quantification analysis and visualization of *Escherichia coli* loads in grass carp fish flesh by hyperspectral imaging method. *Food Bioproc. Technol.*, 8(5), 951-959, <https://doi.org/10.1007/s11947-014-1457-9>
- Cheng J.-H., Sun D.-W., Pu H., and Zeng X.-A., 2014.** Comparison of visible and long-wave near-infrared hyperspectral imaging for colour measurement of grass carp (*Ctenopharyngodon idella*). *Food Bioproc. Technol.*, 7(11), 3109-3120, <https://doi.org/10.1007/s11947-014-1325-7>
- Combrzyński M., Oniszcuk T., Kupryaniuk K., Wójtowicz A., Mitrus M., Milanowski M., Soja J., Budziak-Wieczorek I., Karcz D., Kamiński D., Kulesza S., Wojtunik-Kulesza K., Kasprzak-Drozd K., Gancarz M., Kowalska I., Ślusarczyk L., and Matwijczuk A., 2021.** Physical properties, spectroscopic, microscopic, x-ray, and chemometric analysis of starch films enriched with selected functional additives. *Materials*, 14, 2673, <https://doi.org/10.3390/ma14102673>
- Delwiche S.R., Souza E.J., and Kim M.S., 2013.** Limitations of single kernel near-infrared hyperspectral imaging of soft wheat for milling quality. *Bios. Engin.*, 115(3), 260-273, <https://doi.org/10.1016/j.biosystemseng.2013.03.015>
- de Araújo Gomes A., Schenone A.V., Goicoechea H.C., and de Araújo M.C.U., 2015.** Unfolded partial least squares/residual bilinearization combined with the Successive Projections Algorithm for interval selection: enhanced excitation-emission fluorescence data modeling in the presence of the inner filter effect. *Anal. Bioanal. Chem.*, 407(19), 5649-5659, <https://doi.org/10.1007/s00216-015-8745-8>
- Dziki D., Tarasiuk W., Lysiak G., and Jochymek P., 2020.** The study of particle size distribution of micronized oat bran layer. *Agric. Engin.*, 24(2), 45-54, <https://doi.org/10.1515/agriceng-2020-0016>
- ElMasry G., Iqbal A., Sun D.-W., Allen P., and Ward P., 2011.** Quality classification of cooked, sliced turkey hams using NIR hyperspectral imaging system. *J. Food Engin.*, 103(3), 333-344, <https://doi.org/10.1016/j.jfoodeng.2010.10.031>
- ElMasry G., Sun D.-W. and Allen P., 2012.** Near-infrared hyperspectral imaging for predicting colour, pH and tenderness of fresh beef. *J. Food Engin.*, 110(1), 127-140, <https://doi.org/10.1016/j.jfoodeng.2011.11.028>
- Gowen A.A., O'Donnell C.P., Cullen P.J., Downey G. and Frias J.M., 2007.** Hyperspectral imaging - an emerging process analytical tool for food quality and safety control. *Trends Food Sci. Technol.*, 18(12), 590-598, <https://doi.org/10.1016/j.tifs.2007.06.001>
- Harris D., Rashid A., Miraj G., Arif M., and Shah H., 2007.** 'On-farm' seed priming with zinc sulphate solution - A cost-effective way to increase the maize yields of resource-poor farmers. *Field Crops Res.*, 102(2), 119-127, <https://doi.org/10.1016/j.fcr.2007.03.005>
- Huang M., Zhao W., Wang Q., Min Z., and Zhu Q., 2015.** Prediction of moisture content uniformity using hyperspectral imaging technology during the drying of maize kernel. *Int. Agrophys.*, 29(1), 39-46, <https://doi.org/10.1515/intag-2015-0012>
- Hu M.-H., Dong Q.-L., Liu B.-L., and Opara U.L., 2016.** Prediction of mechanical properties of blueberry using hyperspectral interactance imaging. *Postharvest Biol. Technol.*, 115, 122-131, <https://doi.org/10.1016/j.postharvbio.2015.11.021>
- Jackman P., Sun D.-W., Du C.-J., Allen P., and Downey G., 2008.** Prediction of beef eating quality from colour, marbling and wavelet texture features. *Meat Sci.*, 80, 1273-81, <https://doi.org/10.1016/j.meatsci.2008.06.001>
- Kamruzzaman M., ElMasry G., Sun D.-W., and Allen P., 2011.** Application of NIR hyperspectral imaging for discrimination of lamb muscles. *J. Food Engin.*, 104(3), 332-340, <https://doi.org/10.1016/j.jfoodeng.2010.12.024>
- Kamruzzaman M., ElMasry G., Sun D.-W., and Allen P., 2012.** Prediction of some quality attributes of lamb meat using near-infrared hyperspectral imaging and multivariate analysis. *Anal. Chim. Acta*, 714, 57-67, <https://doi.org/10.1016/j.aca.2011.11.037>
- Kapela K., Sikorska A., Niewęglowski M., Krasnodębska E., Zarzecka K., and Gugala M., 2020.** The impact of nitrogen fertilization and the use of biostimulants on the yield of two maize varieties (*Zea mays* L.) cultivated for grain. *Agronomy*, 10(9), 1408, <https://doi.org/10.3390/agronomy10091408>
- Karami H., Rasekh M., and Mirzaee-Ghaleh E., 2020a.** Comparison of chemometrics and AOCS official methods for predicting the shelf life of edible oil. *Chemom. Intell. Lab. Syst.*, 206, 104165, <https://doi.org/10.1016/j.chemolab.2020.104165>



- Karami H., Rasekh M., and Mirzaee-Ghaleh E., 2020b.** Qualitative analysis of edible oil oxidation using an olfactory machine. *J. Food Meas. Charact.*, 14(5), 2600-2610, <https://doi.org/10.1007/s11694-020-00506-0>
- Karami H., Rasekh M., and Mirzaee-Ghaleh E., 2020c.** Application of the E-nose machine system to detect adulterations in mixed edible oils using chemometrics methods. *J. Food Meas. Charact.*, 44(9), e14696, <https://doi.org/10.1111/jfpp.14696>
- Karami H., Rasekh M., and Mirzaee-Ghaleh E., 2021.** Identification of olfactory characteristics of edible oil during storage period using metal oxide semiconductor sensor signals and ANN methods. *J. Food Proc. Preserv.*, 45(10), e15749, <https://doi.org/10.1111/jfpp.15749>
- Khorrarnifar A., Rasekh M., Karami H., Malaga-Tobola U., and Gancarz M., 2021.** A machine learning method for classification and identification of potato cultivars based on the reaction of MOS type sensor-array. *Sensors*, 21(17), 5836, <https://doi.org/10.3390/s21175836>
- Khorsand A., Rezaverdinejad V., Asgarzadeh H., Majnooni-Heris A., Rahimi A., and Besharat S., 2020.** Response of maize and black gram yield and water productivity to variation in canopy temperature and crop water stress index. *Int. Agrophys.*, 34(3), 381-390, <https://doi.org/10.31545/intagr/126439>
- Lei Z., Zeng Y., Liu P., and Su X., 2021.** Active deep learning for hyperspectral image classification with uncertainty learning. *IEEE Geoscience and Remote Sensing Letters*, 1-5, <https://doi.org/10.1109/LGRS.2020.3045437>
- Benthien J., Heldner S., Seppke B., and Hörbelt J., 2020.** Comparison of two techniques of pattern recognition in the image analysis-based wheat stalk length characterization. *Agric. Engin.*, 24(2), 1-8, <https://doi.org/10.1515/agriceng-2020-0011>
- Liu W., Zeng S., Wu G., Li H., and Chen F., 2021.** Rice seed purity identification technology using hyperspectral image with LASSO logistic regression model. *Sensors*, 21(13), 4384, <https://doi.org/10.3390/s21134384>
- Pekel A.Y., Çalık A., Kuter E., Alataş M.S., Öklen S.B., Kızıl A., Bulat M., and Cengiz Ö., 2020.** Impact of chemical and physical properties on flowability characteristics of corn distillers dried grains with solubles. *Int. Agrophys.*, 34(2), 195-202, <https://doi.org/10.31545/intagr/117502>
- Rasekh M. and Karami H., 2021a.** E-nose coupled with an artificial neural network to detection of fraud in pure and industrial fruit juices. *Int. J. Food Propert.*, 24(1), 592-602, <https://doi.org/10.1080/10942912.2021.1908354>
- Rasekh M. and Karami H., 2021b.** Application of electronic nose with chemometrics methods to the detection of juices fraud. *J. Food Proc. Preserv.*, 45(2), e15432, <https://doi.org/10.1111/jfpp.15432>
- Rasekh M., Karami H., Wilson A.D., and Gancarz M., 2021a.** Classification and identification of essential oils from herbs and fruits based on a mos electronic-nose technology. *Chemosensors*, 9(6), 142, <https://doi.org/10.3390/chemosensors9060142>
- Rasekh M., Karami H., Wilson A.D., and Gancarz M., 2021b.** Performance analysis of mau-9 electronic-nose mos sensor array components and ann classification methods for discrimination of herb and fruit essential oils. *Chemosensors*, 9(9), 243, <https://doi.org/10.3390/chemosensors9090243>
- Sellami A. and Tabbone S., 2022.** Deep neural networks-based relevant latent representation learning for hyperspectral image classification. *Pattern Recognit.*, 121, 108224, <https://doi.org/10.1016/j.patcog.2021.108224>
- Sun J., Jiang S., Mao H., Wu X., and Li Q., 2016.** Classification of black beans using visible and near infrared hyperspectral imaging. *Int. J. Food Propert.*, 19(8), 1687-1695, <https://doi.org/10.1080/10942912.2015.1055760>
- Tan D., Guo L., Liu J., Fan Y., and Li Q., 2020.** Response of dry matter translocation and grain yield of summer maize to biodegradable film in the North China plain. *Int. Agrophys.*, 34(1), 87-94, <https://doi.org/10.31545/intagr/112269>
- Tatli S., Mirzaee-Ghaleh E., Rabbani H., Karami H., and Wilson A.D., 2020.** Rapid detection of urea fertilizer effects on voc emissions from cucumber fruits using a MOS E-Nose Sensor Array. *Agronomy*, 12, 35, <https://doi.org/10.3390/agronomy12010035>
- Valous N.A., Mendoza F., Sun D.-W., and Allen P., 2009.** Colour calibration of a laboratory computer vision system for quality evaluation of pre-sliced hams. *Meat Sci.*, 81(1), 132-141, <https://doi.org/10.1016/j.meatsci.2008.07.009>
- Wang Q.G., Zhu Q.B., Qin J.W., and Huang G., 2015.** Review of seed quality and safety tests using optical sensing technologies. *Seed Sci. Technol.*, 43, 3, 337-366, <https://doi.org/10.15258/sst.2015.43.3.16>
- Williams P.J. and Kucheryavskiy S., 2016.** Classification of maize kernels using NIR hyperspectral imaging. *Food Chem.*, 209, 131-138, <https://doi.org/10.1016/j.foodchem.2016.04.044>
- Wu D., Sun D.-W., and He Y., 2012.** Application of long-wave near infrared hyperspectral imaging for measurement of color distribution in salmon fillet. *Innov. Food Sci. Emerg. Technol.*, 16, 361-372, <https://doi.org/10.1016/j.ifset.2012.08.003>
- Yang X., Hong H., You Z., and Cheng F., 2015.** Spectral and image integrated analysis of hyperspectral data for waxy corn seed variety classification. *Sensors*, 15(7), 15578-15594, <https://doi.org/10.3390/s150715578>
- Yuan Y., Wang C., and Jiang Z., 2021.** Proxy-based deep learning framework for spectral-spatial hyperspectral image classification: efficient and robust. *IEEE Trans. Geosci. Remote Sens.*, 1-15, <https://doi.org/10.1109/TGRS.2021.3054008>
- Zhang X., Liu F., He Y., and Li X., 2012.** Application of hyperspectral imaging and chemometric calibrations for variety discrimination of maize seeds. *Sensors*, 12(12), 17234-17246, <https://doi.org/10.3390/s121217234>