Comparative Study on the Application of Visible and Near Infrared Hyperspectral Imaging for Fusarium Disease Assessment of Corn Seeds

PLAMEN DASKALOV, ELEONORA KIRILOVA, VIOLETA MANCHEVA AND TSVETELINA GEORGIEVA
Department of Automatics and Mechatronics
Ruse University "Angel Kanchev"
Ruse, 7017, Studentska Str. 8
BULGARIA
daskalov@uni-ruse.bg, ekirilova@uni-ruse.bg

Abstract: - A comparison between the classification of healthy and Fusarium Moniliforme diseased corn seeds through color image analysis and NIR Imaging techniques is presented in the paper. Results from SVM classification with selected color feature sets (from GDA, SFFS, Fisher's discriminant ratio together with Scatter matrices) and spectra data analysis including classification of AR models with SVM and threshold value classifiers are shown for corresponding seeds from 7 Bulgarian varieties. It can be concluded that total error rate varies widely within for classification of spectra data for all varieties (0÷65%), whereas for classification of color images it retains close values (0.7÷13.3%) both for pericarp side and for the germ side.

Key-Words: - Corn kernels, Fusarium Moniliforme, Image processing, NIR spectral data, AR models, Classification

1 Introduction

In recent years there has been a continuous increase in the food products quality criteria worldwide. This requires the search for new methods of objectively qualifying, which are as much as possible suitable for use in laboratory and industrial environments. The general trend is mainly to increase the effectiveness of these methods, which is to improve the accuracy of the assessments, to reduce the time they are being carried out, and, above all, to minimize subjectivity in the process of grading. Grains and cereal foods are an essential part of our food. Of the grain crops in Bulgaria, corn is in the second place of cultivation and importance after wheat. From 10 to 30% of the annual production of corn grains in the world is harmed due to diseases caused by development of microorganisms and molds. In doing so, the quality of grains that cause allergies and toxic reactions in humans and animals is impaired.

The Fusarium Moniliforme disease causes not only external but also internal changes in corn kernels. The external changes are alterations in the shape (underdeveloped, wizened) and colors of the corns. Usually diseased corns have grayish-white or normal color with spots and coatings with rose-red nuance. The internal changes in the Fusarium diseased corn kernels are connected with the amount and form of moisture in the kernels.

In addition to external product quality, interest is also intrinsic. Since internal defects can not be assessed subjectively because they are not visible to the naked eye, they require new technological solutions and techniques that provide information about the internal condition of the products. Effective solutions in the assessment of internal indicators were obtained by analyzing the spectral characteristics of objects formed by a suitable spectrophotometer. Monographs [2, 8] and numerous review papers [1, 4, 10, 12, 13, 16] were published with detailed analysis and results as well as extensive information on spectral analysis in the visible (VIS) and near infrared in the area of agricultural and other food products quality assessment. NIR spectroscopy has a good accuracy (of over 95%) in classification of different qualities of cereal crops [3, 11, 14, 15, 17]. The main advantages of presented methods are the non-destructive analysis of the objects and the ability to evaluate their internal state.

The purpose of this work is to be done comparative study on the application of Visible and Near Infrared Hyperspectral Imaging to differentiate between healthy and diseased corn seeds. For this purpose there is need to choose features and formulate the criteria for identification of Fusarium Moniliforme infected corns seeds through the implementation of visible imaging methods on the one hand and introduction of the spectral characteristics of diffuse reflection using methods based on linear discrete models, and the selection of a suitable classification approach on the other hand. The influence of the kernel variety on the identification accuracy also is taken into account.

2 Materials and Methods

The corn samples used in this study were provided and certified by ‘Bulgarian Institute of Corn’ from the city of Kneja. Corn samples from individual cobs and 7 varieties, including sound kernels and diseased with Fusarium Moniliforme were inspected in a random order by a human inspector. Seed varieties are as follows: Kneja 620, Kneja 613, Kneja 446, Kneja 436, Kneja 308(Fig.1.), Ruse 424, XM 87/136 and 26A.

The spectral characteristics of diffuse reflection of corn kernels form variety Kneja 308 are shown in Fig. 1 also. The characteristics of the other five varieties look similar.

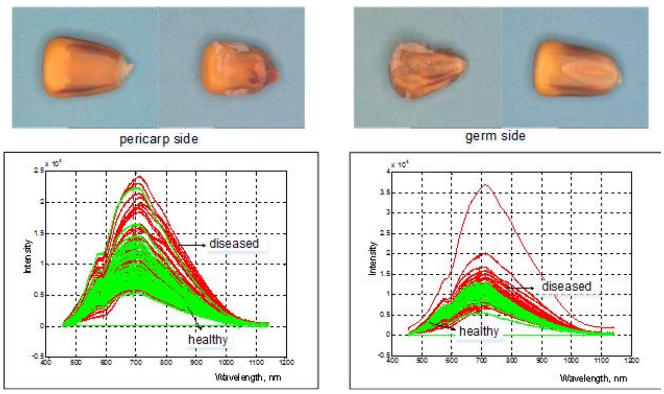


Fig. 1. Images and spectral characteristics of seeds from variety Kneja 308

2.1 Color Image Acquisition and Preprocessing

To obtain color images of seeds was used image acquisition system including CCD color camera and a fluorescent ring body – Philips TL-E 22W/54-765 (with diameter of 30 cm) perpendicular above the working scene around the camera in order to achieve uniform illumination of the object without the appearance of shadows. Fluorescent lighting was chosen due to several reasons, including less generation of infrared wavelengths that tend to bias video camera sensors, mentioned in investigation by [11]. The corn kernels were placed on a blue horizontal support, to achieve easy to remove background, which is generally different from the color of both kernels and infection with Fusarium Moniliforme. The camera Sony VIDO CC422S with a zoom lens of 5-50 mm focal length (F:1.6, Tokina, Japan) was placed at a working distance of 26 cm and a zoom adjusted to obtain an image with the kernel in full screen. For the purposes of preliminary analysis, corn kernels are captured by two distinct, opposing sides coded on the survey as “pericarp side” and a “germ side”. Images are sized (352 x 289) pixels with 8 bit encoding for the red (R), green (G) and blue (B) components of RGB color model and stored in bmp format.

After capturing of the source image of current seed follows procedure that removes background, which represents a threshold segmentation of image by H (Hue) component of HSV color model. After this, current image is called “residual”. For each residual image are obtained 17 continuous so-called color features according to the following procedure:

- Calculate R, G and B as the average values respectively of R, G and B components for all pixels of the residual image of grain;
- convert image into 5 other color models: Lab, XYZ, HSV, YCbCr, xyY. For each model are calculated averages of all three components;
- From each of selected regions of interest (ROIs), the following features are generated: R, G, B, H, S, V, L, a, b, X, Y, Z, Ycbcr, cb, cr, xm, ym from 6 color models respectively: RGB, HSV, Lab, Ycbcb, XYZ, xyY. All the color features are first-order statistics and are derived from the respective values of color components for all the pixels of seed image ROI;
- Forming 17-dimensional vector for current observation considering the fact that component Y is the same in models XYZ and xyY, but not in the model YCbCr, Y features of the latest model are called Ycbcr in vectors of observation:

$$\vec{x} = \left(R,G,B,L,a,b,X,Y,Z,H,S,V,Ycbcr,Cb,Cr,x,y\right)^T .$$

2.1.1 NIR Image Acquisition and Preprocessing

To obtain spectral characteristics, an Ocean Optics spectrophotometer is used for measurement in the visible and near infrared area. The system for obtaining spectral characteristics of diffuse reflection of corn seeds includes a portable computer, Spectrophotometer USB4000-VIS-NIR, range 200 ÷ 1140 nm; light source LS-1, with a range of 360 ÷ 2500 nm and a QR200-7-VIS-NIR probe for measuring the diffuse reflection from subject surface with a range of 200 ÷ 2500 nm.

The Ocean Optics – Spectra Suite software is used to record data from corn kernel measurements, it provides the capability to measure the spectral characteristics of Intensity, Absorbance, Transmission, and Reflection. The resulting characteristics are expressed, respectively, by the coefficient of intensity characterizing the magnitude of the charge of each pixel, *i*. the number of photons per pixel (count / pixel), and the absorption, slip, and diffusion coefficients measured in %(max. 100%). On the basis of the preliminary measurements [7] it was found that for obtaining the spectral characteristic of each corn seed, it should be scanned 10 times (Scans to average = 10), then only one being averaged. The Integration Time = 300ms, and the number of smoothings is Boxcar width = 15.

700 spectral characteristics of diffuse reflection were taken – 350 of healthy and 350 of diseased corn kernels in the range from 456 to 1140.5 nm.

The spectral characteristic of each kernel was obtained for its pericarp side and the germ side, pursuant to the methodology described in [7]. These characteristics show the dependency of the wavelength λ, nm on the intensity of the reflected radiation from the seeds S_λ, in absolute values.

The taken spectral characteristics, shown on Fig. 4 and Fig. 5, are normalized through following equation:

$$S_{\lambda n} = \frac{S_\lambda(\lambda_i)}{S_{\lambda max}} , \qquad (1)$$

where $S_\lambda(\lambda_i)$ is value of the intensity of the reflected radiation with wavelength λ_i,
$S_{\lambda max}$ is the maximum value of the intensity of the reflected radiation for the n-th spectral characteristic.
There are numerous methods for formation of training and test sets when there are big databases.

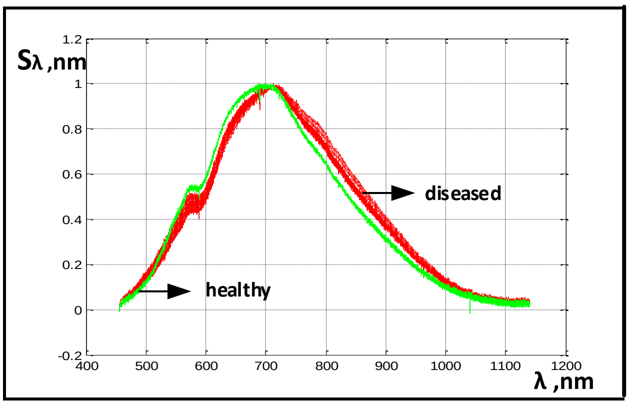


Fig. 2. Regulated spectral characteristics of variety Kneja 308 of corn kernels for pericarp side

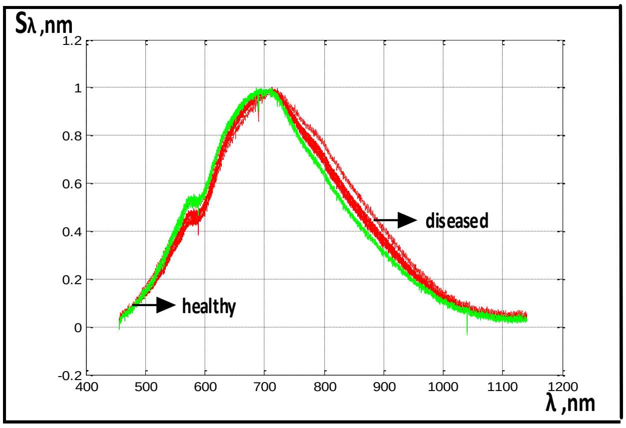


Fig. 3. Regulated spectral characteristics of variety Kneja 308 of corn kernels for germ side

In one of the approaches for selection of representative objects, the objects are selected so that they are uniformly distributed over the output database. A well-known representative approach is the Kennard and Stone Algorithm. When using it, the desired number of objects to be selected from the output ones is given [6]. The obtained spectral characteristics are distributed into two sets. The training set includes 30 spectra of healthy and 30 spectra of diseased corn kernels of one variety or totally 210 spectra of healthy and 210 spectra of diseased kernels for all the seven varieties. The test set includes 20 spectra of healthy and 20 spectra of diseased corn kernels of one variety or totally 140 spectra of healthy and 140 spectra of diseased kernels for the seven varieties.

3 Classification results and discussion

To evaluate the performance of classification procedures Total error rate e_0 was used as criterion:

$$e_0 = \frac{FP + FN}{TP + TN + FP + FN}.100, \ \% \qquad (2)$$

where TP, TN, FP and FN are number of true – positives, true-negatives, false-positives and false-negatives, respectively.

3.1 Classification by image analysis

The first goal, namely, to select those color features that are independent of one another within each class and rich in discriminatory information with respect to the classification problem at hand is achieved in our previous researches [5]. Feature subset selection by General Discriminant Analysis (GDA) (Sequential Forward floating selection

(SFFS), Fisher's discriminant ratio together with Scatter matrices - J3 criterion were tried out. The implementation of procedures was carried out in MATLAB environment. Procedures for determining the relevant sets of features for all varieties were met when setting the two values for the maximum number of features, respectively 7 in method Scalar Feature Ranking (SFR) and 3 in the method Best feature combination (BFC). Table 1 show results for all of the used feature selection methods including all data sets and varieties.

Support Vector Machine classification- SVM method was used. It makes use of the so-called kernel. The most commonly used kernels are linear kernels, polynomial kernels, and radial basis functions. First, the average test data set classification accuracy obtained from a single SVM was estimated. The implementation of the method was made for radial basis function (RBF) with a width σ and regularization constant C.

The optimal values of the regularization constant C and the kernel width have been selected experimentally. To select the values, a "qualified guess" was made from several experiments, first. Then, several loops were run to refine the values by keeping one parameter fixed and adjusting the other one, interchangeably. The selected feature sets were applied to single SVMs.

The classification procedure has to assign the samples in two classes (healthy and infected seeds); software package STATISTICA 8 (StatSoft) is used for implementing the classifiers. The training sample comprises 75% of the overall sample and V-fold cross validation (V = 10) is applied. The parameters of the classifiers gamma and C are experimentally chosen in fixed ranges during training phase, respectively for C = 1÷10 (minimum - 1, maximum - 10, with the increment unit), and for gamma - in the range of 0.091 to 0.333.

Different sets of features received by three different feature subset selection methods has been tested with SVM classifier. Table 1 presents the total error rate values for all the varieties and sets "pericarp side", and "germ side". Total error rate varies in the range of 0.7÷13.3% both for pericarp side and for the germ side. The SVM classifier using features derived through the Scalar Feature Ranking method is resulting in the best classification accuracy with range of total error rate of 0.9-12.2 %.

Table 1. Classifier performance – Total error rate results using SVM classifier and color features

Seed Variety	**Side of capturing**	**Total error rate e_0, %**						
		17color features	**Features from GDA**		**Features from Scalar Feature Ranking**		**Features from Best feature combination**	
Kneja 308	pericarp	3.11	G,H,S,V,L,a,b,X,Ycbcr,cb,cr,xm,ym	3.11	B,S,b,Z,cb,xm,ym	3.11	B,S,Z	3.56
	germ	4	G,H,S,V,L,a,b,X,Ycbcr,cb,cr,xm,ym	4	B,S,b, Z,cb,xm,ym	3.11	B,S,Z	7.11
Kneja 436	pericarp	4.44	B,H,S,V,L,a,b,Y, Ycbcr,cb,cr,xm,ym	4.44	B,S,b,cb,cr,xmym	4.44	B,b,cb	4.89
	germ	3.56	B,H,S,V,L,a,b,Y, Ycbcr,cb,cr,xm,ym	3.56	B,S,b,Z,cb,xm,ym	6.67	B,b,Z	7.56
Kneja 613	pericarp	0.89	B,H,S,V,L,a,b,Z, Ycbcr, xm,ym	0.89	B,S,b,cb,cr,xm ym	0.89	S,xm,ym	0.89
	germ	3.56	H,S,V,L,a,b,Y, Ycbcr , cr,xm,ym	3.56	H,S,a,b,cb,cr, xm, ym	3.11	H,S,b	3.56
Kneja 620	pericarp	4.44	B,H,S,V,L,a,b, Z,cb,xm,ym	4.44	B,H,S,b,cb,xm,ym	4	H,S,ym	4.44
	germ	7.11	B,H,S,V,L,a,b, Z,cb,xm,ym	7.11	H,S,a,b,cb,xm,ym	7.11	H,S,b	6.67
XM 87/136	pericarp	7.78	B,H,S,V,L,a,b, Z,cb,xm,ym	7.78	R,V,L,b,Y,cb, ym	7.78	b,Y,ym	7.78
	germ	12.22	B,H,S,V,L,a,b, Z,cb,xm,ym	13.33	G,L,a,b,Y,cb, ym	12.22	G,L, ym	12.22
Ruse 424	pericarp	0.74	G,H,S,V,L,a,b, Y, Ycbcr, xm,ym	0.74	B,S,b,X,cb,xm,ym	1.48	B,S, ym	0.74
	germ	2.22	G,B,H,S,V,L,a,b, Z,xm,ym	1.48	G,S,L,b,cb,xm,ym	2.22	G,S,L	5.93
26A	pericarp	4.44	B,H,S,V,L,a,b,Z, cr , xm,ym	4.44	B,H,S,b,Z,xm, ym	4.89	B, b, ym	4.44
	germ	4	H,S,V,L,a,b,Z, Ycbcr, cb, xm,ym	3.56	B,H,V,b,Z,xm, ym	3.11	H,V, ym	2.22

3.2. Classification by analysis of spectral characteristics of diffuse reflection intensity and linear discrete models

The discrete parameter models reflect the discreet behavior of the object only in moments that are multiples of the so- sampling rate (tact, measurement) T0.

From obtained spectral characteristics of diffuse reflection intensity, it can be summarized that the amplitude of the coefficient of intensity S(λ) for the healthy and infected grains varies greatly. In order to eliminate the effect of this amplitude on further processing, each spectral characteristic is normalized to its maximum value by the dependence:

$$S^{N_j}_{\lambda_{norm}}(\lambda_i) = \frac{S_\lambda(\lambda_i)}{S^{N_j}_{\lambda_{max}}} \quad , \qquad (3)$$

Where $S^{N_j}_{\lambda_{norm}}(\lambda_i)$ is the normalized spectral characteristic of a corn grain with number N_j, *j=1÷50;*

$S_\lambda(\lambda_i)$ is value of the coefficient of intensity at wavelength λ_i , *i=1÷3648*;

$S^{N_j}_{\lambda_{max}}$ is maximum value of the coefficient of intensity for a spectral characteristic of corn seed with number N_j.

The received normalized spectral characteristic are not linear and could not be described with typical curves. Both nonparametric regression and parametric linear discrete models can be used to obtain their exact description. The main problem with non-parametric ones is the ability to omit existing non-linear relationships between variables, which is undesirable when demanding accurate quality assessment.

Therefore, parametric mathematical models are preferred. Of these, the most common application is the ARMA and its private cases of autoregressive (AR) and creep model (MA), because they are not sophisticated and accurate enough to reproduce the spectral characteristics.

The spectral characteristic of each maize grain is approximated by an equation of type of autoregression (AR) of the type:

$$S_\lambda(k) + A_1 S_\lambda(k-1) + \cdots + A_n S_\lambda(k-n) = e(k) \qquad (4)$$

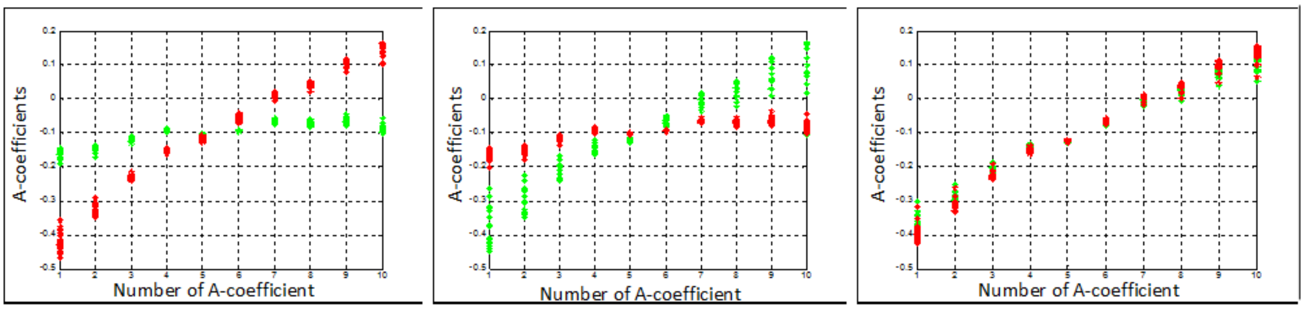
Where k is k-th value of wavelength λ, nm;

A_i, i= (1÷10) – coefficients of autoregressive model;

$S_\lambda(k)$– coefficient of intensity for the k-th value of the wavelength λ; *n*– order of the autoregression model; *e(k)* – difference between the model and the real spectral characteristics for the k -th wavelength value λ.

The lines of the autoregressive pattern represent derivatives of the corresponding order *n*. Most of the researchers in the state-of-the-art present spectral characteristics with derivatives - most commonly through first and second derivatives. Therefore, in this study for the representation and approximation of the spectral characteristics of corn seeds, discrete AR models are investigated.

The n-th series of the linear discrete model, describing the spectral characteristics of the healthy class and diseased class of kernels, is analyzed and series of discrete model of Autoregression (AR) type are obtained. The results show that model series number is not a determining indicator for identification of the kernels. The 10th series of the model is dominating for seven varieties of corn kernels. That is why only the 10th series will be used in order to receive coefficients of the model from training set. Therefore ten coefficients should be calculated for each variety. Three cases for A-coefficients of the AR model are obtained and they are shown on Fig. 4. In the first two cases, the a) and b) group of healthy kernels is clearly distinguishable from the group of diseased kernels. In the third case – c) – there is an overlapping of class healthy and diseased.



a) b) c)

Fig. 4. Three cases for the A-coefficients of the AR model

In order to formulate the identification criterion, the limit value of ALV should be determined between the two classes of kernels – healthy and diseased. To this end, the maximum distance ΔA between class healthy and class diseased should be calculated for each of ten calculated coefficients. The next step is to select the biggest distance from the calculated ΔA. Table 2 shows the conditions for decision in the three cases, for A-coefficients of the AR model for different varieties, and the total error rate evaluated using classifier by threshold value. A_{zdr_min} and A_{zdr_max} are the minimum and maximum value, respectively, of the coefficient A_i , i= (1÷10) for healthy kernels; A_{zar_min} and A_{zar_max} are the minimum and maximum value of the coefficient A_i, i= (1÷10) for the diseased kernels.

As there is no clear distinction between healthy and diseased class in variant c), the mean values A_{zdr_mean} and A_{zar_mean} of the coefficient A_i , i= (1÷10) for the both classes of corn kernels should be determined.

The biggest distance from calculated distances for the ten coefficients is chosen and it corresponds to the A1 coefficient. The limit value of first coefficient A1LV between class healthy and the class diseased is determined (shown on Table 3), thus on the grounds of the received coefficient variants, shown on Fig. 4, the conditions for identification of healthy and diseased corn kernels from the test set are formulated and also presented in Table 2.

Table 2. Conditions for decision for the A-coefficients of the AR model for different varieties, and the total error rate evaluated using classifier by threshold value

Seed Variety	value in terms of coefficient A	Condition for decision	Total error e_0 , %
Kneja 308	c	If $A_{izar\_min} \equiv A_{izdr\_max}$, $\Delta A = \frac{A_{izdr\_mean} + A_{izar\_mean}}{2}$	47.5
Kneja 436	b	If $A_{izar\_min} > A_{izdr\_max}$, $\Delta A = A_{izar\_min} - A_{izdr\_max}$	2.5
Kneja 613	c	If $A_{izar\_min} \equiv A_{izdr\_max}$, $\Delta A = \frac{A_{izdr\_mean} + A_{izar\_mean}}{2}$	48.7
Kneja 620	c	If $A_{izar\_min} \equiv A_{izdr\_max}$, $\Delta A = \frac{A_{izdr\_mean} + A_{izar\_mean}}{2}$	21.2
XM 87/136	a	If $A_{izdr\_min} > A_{izar\_max}$, $\Delta A = A_{izdr\_min} - A_{izar\_max}$	2.5
Ruse 424	b	If $A_{izar\_min} > A_{izdr\_max}$, $\Delta A = A_{izar\_min} - A_{izdr\_max}$	0
26A	a	If $A_{izdr\_min} > A_{izar\_max}$, $\Delta A = A_{izdr\_min} - A_{izar\_max}$	1.2

Table 3. Limit values between healthy and diseased

variety	Limit value A1LV for pericarp side	Limit value A1LV for germ side
Kneja 308	- 0.2696	- 0.2788
Kneja 436	- 0.2168	- 0.2296
Kneja 613	- 0.2786	- 0.2987
Kneja 620	- 0.3792	- 0.3915
26 A	- 0.2895	- 0.3317
XM 87/136	- 0.2925	- 0.2743
Ruse 424	- 0.2812	- 0.3038

The three variants of obtaining the coefficients A_i , i= (1÷10) from the AR models of Fig.4 allow the SVM method to be used both for linearly separable objects (Fig.4, a) and b)) and for non-linearly separable ones (Fig.4, c). Therefore, the realization of the method is performed for both types of kernel functions - linear and radial-base function (RBF). Only input coefficients A_1 are used as input vectors, and all the ten coefficients (A_1÷A_{10}.) are the second time. Thus, the task of classifying corn seeds is limited to classification except for one-dimensional (A_1) and multi-dimensional descriptions (A_1÷A_{10}).

The results from classification of test samples using the SVM method in both versions (RBF and linear) are presented in Table 4. For comparison, the Table 5 shows the accuracy obtained with SVM classifier with coefficients of AR models A_1÷A_{10}, the classifier by threshold value and AR models and SVM classification with color features by Method Scalar Feature Ranking. It is based on the summarized results for healthy and infected (pericarp side) and for healthy and infected (germ side).

The results from Table 4 show that the use of ten coefficients (A_1÷A_{10}.) for classification give higher results in both cases of SVM classifier as linear and nonlinear than only using the first coefficient (A_1).

For Kneja varieties 436, 26A, Rousse 424 and XM87 / 136 percent of the classification remains the same as using one coefficient A_1 and using all $A_1 \div A_{10}$.

The highest percentage of recognition (100%) is achieved for variety XM87 / 136 using both methods -SVM and the threshold value classification for AR models.

For Kneja 436, 26A and Rousse 424 varieties there is a slight decrease in classification accuracy when using the SVM classifier compared to classification by threshold value. But overall, the rate of recognition remains high - within the range of 90 ÷ 97.5%.

For varieties with overlapping coefficient distributions (Kneja 308, Kneja 613 and Kneja 620), the application of SVM and AR models gives significantly better results. Classification accuracy increased by 20% for pericarp side of capturing and 25% for germ side for Kneja 308 seeds; with 47.5% for the pericarp side and 22.5% for the germ side for Kneja 613; and 15% for germ side for variety Kneja 620. For pericarp side of Kneja variety 620 - it remains the same (87.5%).

The classification results are summarized in Fig.5 corresponding to classification of independent test samples of images of seeds- pericarp side and germ side respectively.

Table 4. Results from approximating the spectral characteristics of the diffuse reflection intensity accuracy for test samples by SVM method e_0, %

Classifier	Type of classifier	Coefficients of AR models	Total error e_0, %													
			Kneja 308		Kneja 436		Kneja 613		Kneja 620		26A		Ruse424		XM87/136	
			pericarp	germ	pericarp	germ	pericarp	germ	pericarp	germ	pericarp	germ	pericarp	germ	pericarp	germ
SVM	linear	A_1	57.5	45	10	12.5	37.5	42.5	27.5	55	5	2.5	5	2.5	0	0
	RBF		55	65	10	10	62.5	60	25	22.5	5	2.5	5	2.5	0	0
	linear	$A_1 \div A_{10}$	20	55	10	12.5	35	35	12.5	15	5	7.5	5	2.5	0	0
	RBF		40	30	10	12.5	12.5	15	20	25	5	5	5	2.5	0	0

Table 5. Comparison between total error values using different classification methods

Classifier	Type of classifier	Total error e_0, %													
		Kneja 308		Kneja 436		Kneja 613		Kneja 620		26A		Ruse424		XM87/136	
		pericarp	germ	pericarp	germ	pericarp	germ	pericarp	germ	pericarp	germ	pericarp	germ	pericarp	germ
SVM A1÷A10	linear	20	55	10	12.5	35	35	12.5	15	5	7.5	5	2.5	0	0
	RBF	40	30	10	12.5	12.5	15	20	25	5	5	5	2.5	0	0
classifier by threshold value	linear	40	55	2.5	2.5	60	37.5	12.5	30	0	5	2.5	0	0	0
SVM with SFR	RBF	3.11	3.11	4.44	6.67	0.89	3.11	4	7.11	4.89	3.11	1.48	2.22	7.78	12.22

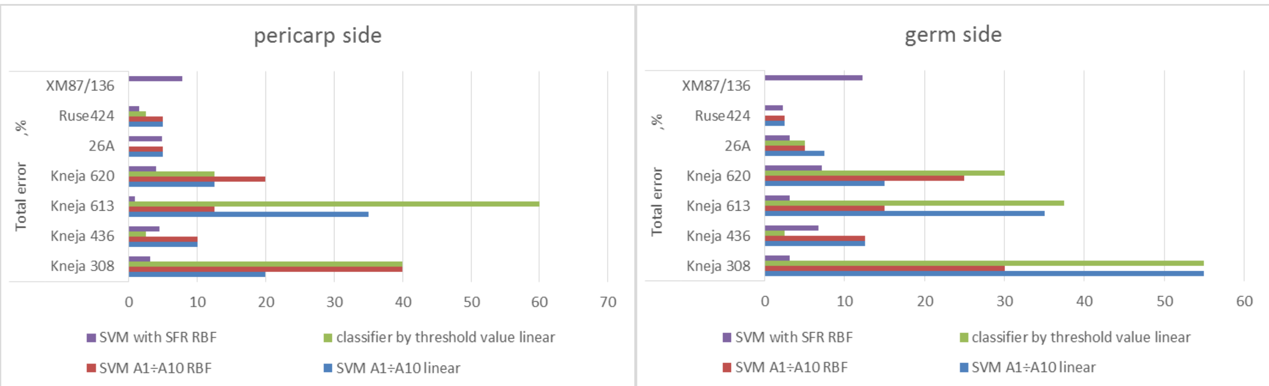


Fig. 5. Comparison between total error values using different classification methods

There are compared results from classification by color analysis and SVM algorithms, and spectral analysis classification including AR models with SVM and threshold value classifiers. It can be seen that total error rate varies widely within for classification of spectra data for all varieties, whereas for classification of color images it retains close values both for pericarp side and for the germ side. There are two varieties which differ as XM87 / 136 reaches highest percentage of recognition (100%) or 0% total error rate in classification based on spectra data, so variety Kneja 613 reaches very high total error of 12-60% using spectra data, while significantly lower values 0.89% using color analysis.

4 Conclusion

Comparing obtained results for color image analysis it has been found that in terms of the minimum total average error for all varieties best results gives the Support Vector Machine classifier using features derived through the Scalar Feature Ranking method. It is resulting in the range of 0.9-12.2 % according to the side of image capturing and variety.

In regard to spectra data technique where the coefficients of linear parametric models of discrete type Autoregresion (AR) were analyzed, and identification criterion is based on the limit value of ALV between the class healthy and class infected seeds, so the maximum distance between the two classes - ΔA for the 10th order of AR-model is used to determine the limit value. It can be concluded that combining AR models and SVM classifiers, and using all ten coefficients of AR models gives a higher recognition rate for all varieties. Regardless of the improved SVM classifiers, accuracy of classification using spectral data for some varieties is still not satisfactory.

The presented results show that total error rate varies widely within for classification of spectra data for all varieties (0÷65%), whereas for classification of color images it retains close values (0.7÷13.3%) both for pericarp side and for the germ side.

Acknowledgement

The study was supported by contract of University of Ruse “Angel Kanchev”, № BG05M2OP001-2.009-0011-C01, "Support for the development of human resources for research and innovation” at the University of Ruse “Angel Kanchev”. The project is funded with support from the Operational Program " Science and Education for Smart Growth 2014 - 2020" financed by the European Social Fund of the European Union..

References:

[1] Berardo N., Pisacane V., Battilani P., Scandolara A., Pietri A., Marocco A., Rapid detection of kernel rots and mycotoxins in maize by near-infrared reflectance spectroscopy, *Journal of Agricultural and Food Chemistry*, Vol.53, 2005, pp.8128-8134.

[2] Bishop Christopher M., *Pattern recognition and machine learning*, Springer- Verlag., 2006.

[3] Chen Ho-Hsien, C. Thing, The development of a machine-vision system for shiitake garding, *Journal of Food quality* 27, 2004, pp.352-365.

[4] Chen X., Y. Xun et al.,Combining discriminant analysis and neural networks for corn variety identification, *Computers and Electronics in Agriculture* 71S, 2010, pp. 48–S53.

[5] Daskalov Plamen, Kirilova E., Georgieva Tz., Performance of an automatic inspection system for classification of Fusarium Moniliforme damaged corn seeds by image analysis, *MATEC Web of Conferences* 210, 02014, 2018. https://doi.org/10.1051/matecconf/201821002014

[6] Facchin S., J. Trierwieler, V. Conz, Soft sensor design: A new approach for variable selection., *2nd Mercosur Congress on Chemical Engineering.*

[7] Mancheva, V., Pl.Daskalov, R. Tsonev Ts. Draganova. Creation of spectral characteristics database for Fusarium diseased corn seeds recognition, *Proceedings of the University of Rousse "Angel Kanchev"*, Vol. 48, 3.1, 2009, pp. 150-157.

[8] Olsson J., *Modern methods in cereal grain mycology.*, 2000, PhD Thesis., Sweden.

[9] Pearson Tom, Machine vision system for automated detection of stained pistachio nuts, Lebensm.-Wiss. *u.-Technol.*, 29, 1996, pp.203–209.

[10] Pettersson H., Aberg L., Near infrared spectroscopy for determination of mycotoxins in cereals. *Food Control*, 14, 2003, pp. 229-232.

[11] Steenhoek L. W., M. K. Misra, C. R. Hurburgh Jr., C.J. Bern, Implementing a computer vision system for corn kernel damage evaluation, *Applied Engineering in Agriculture* Vol. 17(2), 2001, pp.235 – 240.

[12] Singh Chandra B., D. Jayas, J. Paliwal, N. White, Detection of midge-damaged wheat kernels using short-wave near-infrared hyperspectral and digital colour imaging, *Biosystems engineering* 105, 2010, pp.380-387

[13] Shahin M., S. Symons, Detection of Fusarium damaged kernels in Canada Western Red Spring wheat using visible/near-infrared hyperspectral imaging and principal component analysis, *Computers and electronics in agriculture,* 75, 2011, pp. 107- 112.

[14] Taghizadeh Masoud, A. Gowen, C. ODonnell, Comparison of hyperspectral imaging with conventional RGB imaging for quality evaluation of Agaricus bisporus mushrooms, *Biosystems engineering* 108, 2011, pp.191-194.

[15] Wang J., K. Nakano, S. Ohashi, Y. Kubota, K. Takizawa, Y. Sasaki, Detection of external insect infestations in jujube fruit using hyperspectral reflectance imaging, *Biosystems engineering* 106, 2010, pp.345-351.

[16] Xing Juan, S. Symons, M. Shahin, D. Hatcher, Detection of sprout damage in Canada Western Red Spring wheat with multiple wavebands using visible/near-infrared hyperspectral imaging, *Biosystems engineering* 106 (2010), pp.188-194.

[17] Zhang, N., Chaisattapagon, C., Effective criteria for weed identification in wheat fields using machine vision. *Transactions of the ASAE* 38 (3), 1995, pp.965-974.