(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:

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

Where $S_{\lambda_{norm}}^{N_j}(\lambda_i)$ 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*;

$S_{\lambda_{max}}^{N_j}$ 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:

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

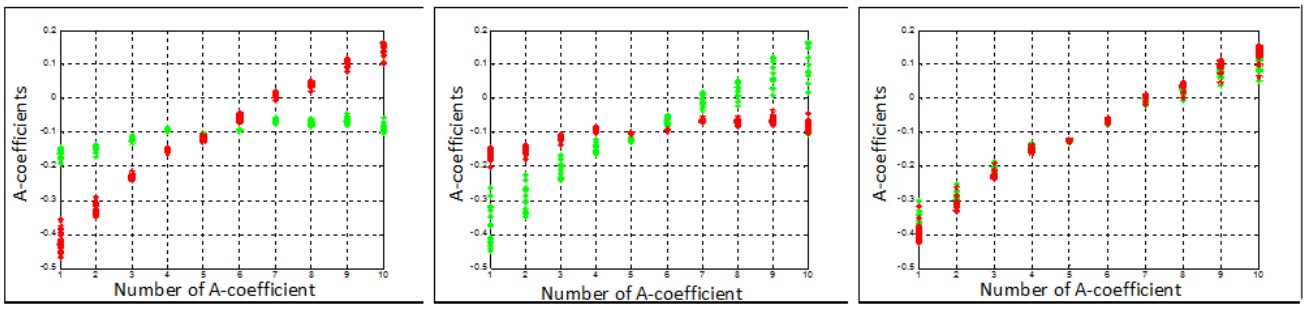
Where *k* is *k*-th value of wavelength λ, nm;

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 ($A_1 \div A_{10}$.) 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 ($A_1 \div A_{10}$).

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 $A_1 \div A_{10}$, 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 ($A_1 \div A_{10}$.) 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

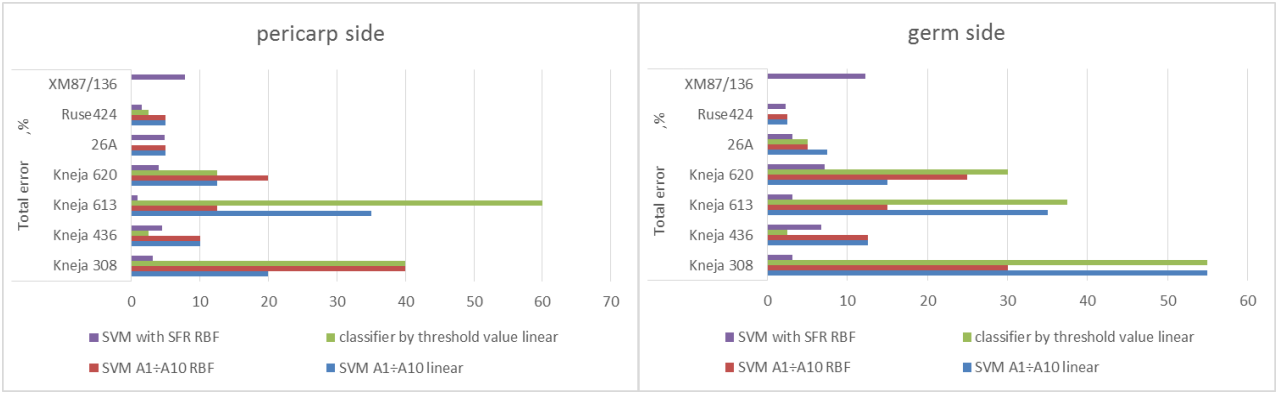


Fig. 5. Comparison between total error values using different classification methods

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.