

# Performance of an automatic inspection system for classification of *Fusarium Moniliforme* damaged corn seeds by image analysis

Plamen Daskalov<sup>1</sup>, Eleonora Kirilova<sup>1,\*</sup>, and Tzvetelina Georgieva<sup>1</sup>

<sup>1</sup>Department of Automatics and Mechatronics, University of Ruse, 8 Studentska Str., 7017 Ruse, Bulgaria

**Abstract.** This paper presents algorithms for pre-processing, feature selection and classifier design which are used for Parameter Refining in the task of development of an automatic system for recognition and grading of corn seeds with external signs of *Fusarium Moniliforme* disease. The abilities of several feature selection methods – FDR, Scatter matrices and Stepwise Discriminant Analysis and two classification methods - Support Vector Machine (SVM) and K-Nearest Neighbours (K-NN) are investigated. Design and implementation of the system also has been showed. The system could continually present one by one positioned corn kernels to CCD camera, perform a classification procedure of captured images and discharge seeds to assigned containers. The software was developed in LabVIEW environment including image analysis and classification procedures performed using MATLAB Script. Results for total error rate of 8.4% - 7.2% from preliminary classification related to 8480 seeds from 16 Bulgarian varieties and total error rate of 6.95% - 20.4% for experimental results obtained with the system during the control measurements of the seed sample are obtained.

## 1 Introduction

In the field of agricultural and food technology the task of qualify production is complex and complicated. Difficulties in solving it are linked to the diversity of products and their non-homogeneous nature. Each product or good which is subject to quality control has multiple specific features that need to be considered when drawing up the criteria for assessing the quality and for realization of their actual grading [6, 8, 20].

The main share in the production of agricultural products takes cultivation of cereals such as maize globally ranks third in sown area and first in production of grain in all cereals. Corn is a major crop in the economy of the Republic of Bulgaria. It ranks second after wheat in planted area and production level. Preparation of seed in Bulgaria even nowadays is carried out by drying and calibration of hybrid maize. In practice seeds grading is done manually in the so-called seed control inspection stations in accordance with the regulations of law. Corn diseases are one of the important factors that have a negative influence on the yield, and incidence rate is one of the main quality indicators. Most damage cause the following diseases: *Ustilago zea* Unger, *Sorosporium zeilianum*, *Fisarium* sp, *Helminthosporium turcicum*, *Puccinia sorghi*, *Nigrospora oryzae*. Causers are numerous different fungi from *Fusarium* species. Maize contamination by fungi not only renders grains unfit for human consumption by discoloration and reduction of nutritional value, but can also lead to mycotoxin production. Mycotoxins [13, 18, 19] are poisonous secondary metabolites produced by some fungi in staple foods and foodstuffs. The Bulgarian commission estimated the distribution levels of disease isolates by species in regard to corn seeds. Main and the most common are *F.moniliforme* and *F.graminearum*.

The state-of-the-art analyses in the subject area show that classical methods for microbiological analysis and quality control of the corn seeds are precise, reliable and comparable, but they require special laboratory conditions, chemical reagents, qualified experts and the like [19, 25, 26, 27]. The duration of analyses, their labour consumption and the destruction of the samples analysed make them inefficient for on-line control. Organoleptic methods are subjective and depend on the qualifications and the experience of the assessing expert. There is a need to develop a model of an automated system, which integrates a comprehensive non-invasive and objective approach to the recognition of the disease based on a study of Bulgarian maize varieties. Several aspects regarding the influence of varietal identity, year of harvest and side of capturing images also need to be studied. This will improve and simplify the classification process so that the feature development of an automated system to be facilitated.

A brief review on some of the main disease detection techniques on kernels [16, 23] proposed is presented next. Machine vision- based systems are well known [1] and preferred in the field of identification and automation of grain handling, but the variety of diseases and damages on kernels are leading the need of more researches. Several authors have applied digital image analysis and pattern recognition for objective identification of cereal grains [3, 5, 7, 10, 11, 12]. Various aspects of researches on the application of machine vision techniques for disease detection have been previously reported in the field of grain inspection. Detection of diseased kernels by image analysis is a relatively often discussed topic, recently, but most of researches are on wheat kernels. For example Singh et al. [14] used short-wave near-infrared hyperspectral and digital colour imaging for detection of healthy and midge-damaged wheat kernels. NIR hyperspectral image features combined with the top 10 colour image features

\* Corresponding author: [ekirilova@uni-ruse.bg](mailto:ekirilova@uni-ruse.bg)

gave the highest average classifier accuracy of 95.3–99.3%.

Yang et al. [22] reported several works on the effort to develop more efficient methods for separating Fusarium-damaged kernels from sound wheat kernels using high-powered pulsed LED system. Two parameters (slope and  $r^2$ ) from a regression analysis of the green response onto the red response were used as input parameters in a linear discriminant analysis (LDA). They have demonstrated that Fusarium-damaged and sound individual wheat kernels can be correctly categorized at up to 91% average accuracy.

Shahin and Symons [15] investigated the use of hyperspectral imaging in the visible-NIR spectral range (400–1000 nm) for the detection of varying degrees of fusarium damage in wheat. The study reported that hyperspectral imaging over 450–950nm can be used to detect varying degrees of fusarium damage in CWRS wheat. Using Linear Discriminant Analysis, sound and damaged kernels can be classified with an overall accuracy of 92% and the extent of fusarium damage can be predicted with an accuracy approaching 86%. A set of six wavebands can be used to achieve accuracies similar to those with the entire range of 450–950 nm.

However, these methods are good basis on which to stand but naturally they are not suitable directly to serve for rapid online inspection of Fusarium Moniliforme damaged corn seeds. Currently, the available researches [21] on identification of F. moniliforme on corn seeds by color features are not enough.

The objectives of this study were to establish valid inspection parameters from color characteristics of corn seeds to develop automatic inspection system for recognition and grading of sound and Fusarium Moniliforme damaged corn seeds and to study the corn seed classification accuracy and repeatability of the computer vision system as compared with results of human inspection.

## 2 Materials and methods

### 2.1. Data collection of corn samples

The corn samples used in this study were provided and certified by ‘Bulgarian Institute of Corn’ from the city of Kneja. They have been collected in two following crop years. A total of 8480 corn samples from individual cobs and 16 varieties, listed in Fig. 1 and Fig.2, including sound kernels and diseased with Fusarium Moniliforme were inspected in a random order by a human inspector.



Fig. 1. Seed varieties from crop year 2008.

Seed varieties from crop year 2008 shown in fig.1. are as follows: Kneja 620, Kneja 613, Kneja 446, Kneja 436, Kneja 308, Ruse 424, XM 87/136 and 26A.

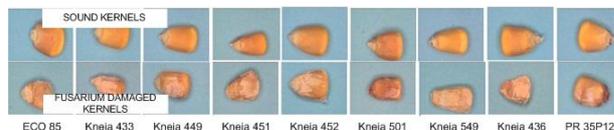


Fig. 2. Seed varieties from crop year 2009.

Seed varieties from crop year 2009, shown in fig.2 are as follows: ECO 85, Kneja 433, Kneja 436, Kneja 449, Kneja 451, Kneja 452, Kneja 501, Kneja 515, Kneja 549 and PR35 P12.

### 2.2. Image Acquisition and Preprocessing

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 Steenhoek et al., 2001. For the purposes of preliminary analysis, maize kernels are captured by two distinct, opposing sides coded on the survey as “pericarp” and a “germ side” (fig.3.). 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.

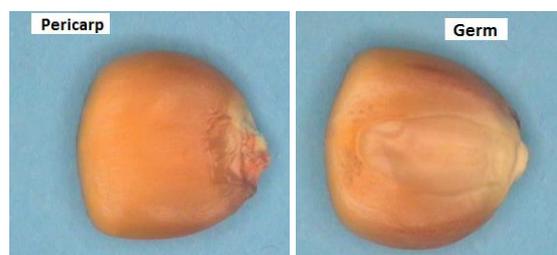


Fig.3. Images of seeds showing “pericarp” side and a “germ” side

After capturing the source image of the current seed Region of interest (ROI) has been selected by the Hue component. From each of the selected ROIs, the following features are generated: R, G, B, H, S, V, L, a, b, X, Y, Z, Ycbr, cb, cr, xm, ym from 6 color spaces respectively: RGB, HSV, Lab, Ycbcb, XYZ, xyY. All the color features [24] are first-order statistics and are derived from the respective values of the color components for all the pixels of the seed image ROI. 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 as follows

#### 2.2.1. Feature subset selection and classification

A major issue in classifier synthesis is the selection of informative features [9]. Feature reduction here is important with respect to the working time of the algorithm. The first goal - to select the most relevant

features, 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 as follows. The feature selection module in the presented work is based on the filter approach, which selects relevant features before the classification algorithm is applied.

**2.2.1.1. Stepwise Discriminant Analysis. (General Discriminant Analysis GDA) Sequential Forward floating selection (SFFS)**

Discriminant analysis [2] was used to identify those color features (variables) that contributed most to the between-group differences. The null hypothesis H0 was tested using the Wilks  $\lambda$ , defined as

$$\lambda = \frac{|W|}{|T|} = \frac{|W|}{|W + B|} \tag{1}$$

where  $|W|$ ,  $|B|$ ,  $|T|$  denote determinants of within-group, between-groups, and total variation matrices, bearing in mind that

$$W = \sum_{i=1}^k \sum_{j=1}^{n_i} (x_{ij} - \bar{x}_i)(x_{ij} - \bar{x}_i)' \tag{2}$$

$$B = \sum_{i=1}^k n_i (\bar{x}_i - \bar{x})(\bar{x}_i - \bar{x})' \tag{3}$$

$$T = \sum_{i=1}^k \sum_{j=1}^{n_i} (x_{ij} - \bar{x})(x_{ij} - \bar{x})' \tag{4}$$

$x_{ij}$  is a p-dimensional vector of the jth object in the ith population;  $\bar{x}_i$  is a p-dimensional vector of means in the ith population;  $\bar{x}$  is a vector of the overall mean.

Stepwise selection begins with no variables in the classification model. At each step of the process, the variables within and outside the model are evaluated. The variable within the model, at that particular step, which contributes least to the model as determined by the Wilks' Lambda  $\lambda$  method is removed from the model. Likewise, the variable outside the model that contributes most to the model and passes the test to be admitted is added. When no more steps can be taken, the number of variables in the model is reduced to its final form. Based on these analyses, the features were ranked by the level of contribution (determined by Correlation Coefficient (CC) and Average Squared Canonical Correlation (ASCC)) to the classifier. In this method, Wilks' Lambda is the likelihood ratio statistic for testing the hypothesis that the means of the classes on the selected variables are equal in the population. Lambda is close to zero if any two groups are well separated.

$$F_{wilks} = \left( \frac{n-g-p}{g-1} \right) \left( \frac{(1-\lambda_{p+1}/\lambda_p)}{\lambda_{p+1}/\lambda_p} \right) \tag{5}$$

where p – the overall number of features, g – number of classes,  $\lambda_p$  - Wilks  $\lambda$  for current feature sub-set.

The Wilks' Lambda (15) statistic for the overall discrimination is computed as the ratio of the determinant (det) of the within-groups variance/covariance matrix over the determinant of the total variance covariance matrix.

In accordance with the number of features from baseline feature space - 17, total number of 131,069 feature sets for each variety has been generated and tested by the GDA method. The computations were carried out with the Discriminant Analysis module of the Statistica software.

**2.2.1.2. Scalar Feature Ranking – SFR**

Feature subset selection technique proposed by Theodoridis and Koutrombas [17], which employs the FDR (Fisher's Discriminant Ratio) criterion (1) as well as a cross-correlation measure between pairs of features and scatter matrices in order to rank them, was used. Scatter matrices are among the most popular measures for quantifying the way feature vectors "scatter" in the feature space. Because of their rich physical meaning, a number of class- separability measures are built around them [17], so we choose one of the available measures, based on scatter matrices -  $J_3$  criterion.

The procedure included computing of class separability criterion for each feature, so first FDR criterion (16) is used to rank features in descending order.

$$FDR = \frac{(m_1 - m_2)^2}{(\sigma_1^2 + \sigma_2^2)} \tag{6}$$

where  $m_1$  and  $m_2$  are the respective mean values and  $\sigma_1^2$ ,  $\sigma_2^2$  - the respective variances associated with the values of a feature in two classes. As result, features are ranked in descending order according FDR criterion. The highest-ranked are identified and exhaustive search is employed to select the combination that maximizes the  $J_3$  criterion. The  $J_3$  criterion is defined as:

$$J_3 = trace\{S_w^{-1}S_b\} \tag{7}$$

Where  $S_w$  is the within-class scatter matrix, and  $S_b$  – between-class scatter matrix.

Let  $i_1$  be the index of the best one. The respective definitions are

$$S_w = \sum_{i=1}^c P_i S_i \tag{8}$$

Where  $P_i$  denotes the a priori probability of class  $i = 1, 2, \dots, c$  and  $S_i$  is the respective covariance matrix of class  $i$ .

**2.2.1.3. Best Feature Combination – BFC**

We assume that  $m$  features are chosen employing „Scalar Feature Ranking” technique, so the goal is to find the “best” combination of features. The Best Feature Combination technique [17] represents exhaustive search method, which follows the next steps: Read the training data and normalize the feature values; Rank the features using the normalized training data set employing the above described technique. To reduce the dimensionality of the feature space, work with seven highest-ranked features; Choose the best feature combination consisting of three features (out of the previously selected seven) using exhaustive search method; Classify the normalized feature vectors of the test data using k-NN classifier (k=3) and compute the classification error, employing the Leave-one-out (LOO) method.

The three-dimensional feature space, which results in the best performance of k-NN (k=3) classifier, in respect to the LOO Cross-validation error is chosen.

**2.2.2. Classifier performance evaluation**

During this stage, two types of classifiers are employed in the selected feature space. The selection of the most appropriate classifier in specific automated grading system is determined by a number of requirements, the most important of which are: enough classification accuracy, simplest possible classification algorithm, and computational efficiency, minimum time of execution (boost), easy implementation and possibility of additional settings, versatility and flexibility. Methods of pattern recognition, requiring long time for analysis and classification, or those that are based on large amounts of computational procedures are practically "not applicable" under consideration task. For the purpose of this work is investigated the possibility of image classification with K-Nearest Neighbours (k-nn) and Support Vector Machine (SVM). Random sampling is performed, as 75% of the observations are taking part in the train set, and 25 % in the test set.

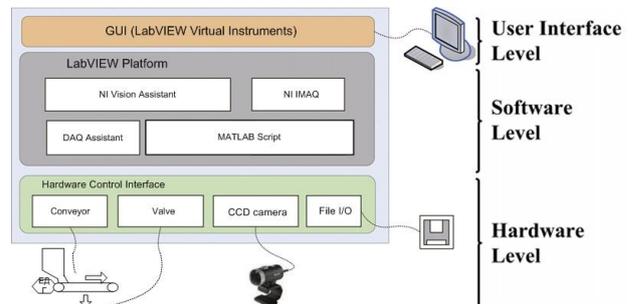
To evaluate the performance of the designed system Total error rate  $e_0$  was used as criterion for classifier performance evaluation.

$$e_0 = \frac{FP + FN}{TP + TN + FP + FN} \cdot 100, \% \tag{9}$$

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

**2.3. Design and Implementation of the System**

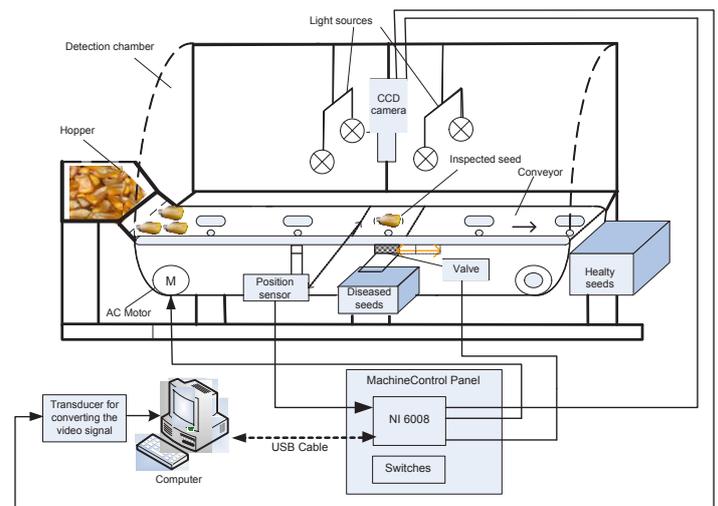
The global structure of the system is shown on fig.4. There are specified three general levels – Hardware level, Software level and User Interface level.



**Fig. 4.** Global structure of prototype system

**2.3.1. Hardware level**

The proposed automated system for inspecting of healthy and infected with Fusarium Moniliforme corn kernels (Fig. 5) is consisting of aluminium profiles measuring 38 cm. – Width, 63 cm. - Length and 72 cm. - Height. The movement of the kernels is ensured by conveyor belt, which is driven by a DC motor. Conveyor is opaque with a dark blue color and width 3 cm. along the belt off 3.5 cm apart are formed round holes with a diameter of 1.5 cm. The color of the belt is selected according to the fact that the blue color does not appear in either the colors inherent in corn kernels, or to typical color coating, which forms fungi from Fusarium Moniliforme. This allows a clear separation of the object from the background color, which is important when analysing only the pixels belonging to the object.



**Fig. 5.** Design schema of the system

2.3.2. Software level

In the software design level of the developed automated system, LabVIEW 8.6. (National Instruments, Inc.) is adopted as software development and integration platform. The proposed software is designed to obtain a reliable, simple and easy to use program for managing automated recognition system for classification of corn based on their color. NI IMAQ, NI Vision Assistant and NI DAQ packages were used to develop subroutines associated with image capturing, image processing and motion control, respectively.

Module NI Vision Acquisition software is the basic element in the development of this type of applications and includes the necessary drivers, namely NI-IMAQ and NI-IMAQdx through which easily controls the camera. IMAQdx driver allows use with different types of cameras – IEEE 1394 (FireWire), GigE Vision (Ethernet), and USB.

Fig 5 shows Diagram of synchronization of the system actions in case of healthy or diseased seed. The structure and working algorithm of the virtual instrument for Fusarium damage detection is shown in fig.6. The first step of the block diagram is image registration. Original image, which is captured by the camera, includes grain and background. To be performed correctly classification procedure is necessary in the area of inspection (corresponding to specified coordinates in the image to the location of the camera) is actually located object.

The template color searching function in LabVIEW Vision Development Module is used for searching the template color image of a seed in the ROI Region of interest to make decision whether there is properly positioned object. The identified ROI “A” and the cropped ROI area “B” for the current image can be seen on the Front panel of the VI (Fig.8.). The block "Cropped ROI area" is performed to remove the initial background image in order to reduce the quantity of pixels for further processing. The phase of classification is performed only if it is found correct positioning of the object. For the current image is formed three integer array with color data of each pixel corresponding to R, G and B components. These arrays are fed directly as input to MATLAB Script, since there is the algorithm for the classification of objects studied by analysis of the color components.

View of the camera, visualization of results and the light indicators are also part of the operation of the LabVIEW instrument. MATLAB, as part of the overall virtual instrument, is used for calculating the values of the 17 color features, and implementation of the algorithm for classification.

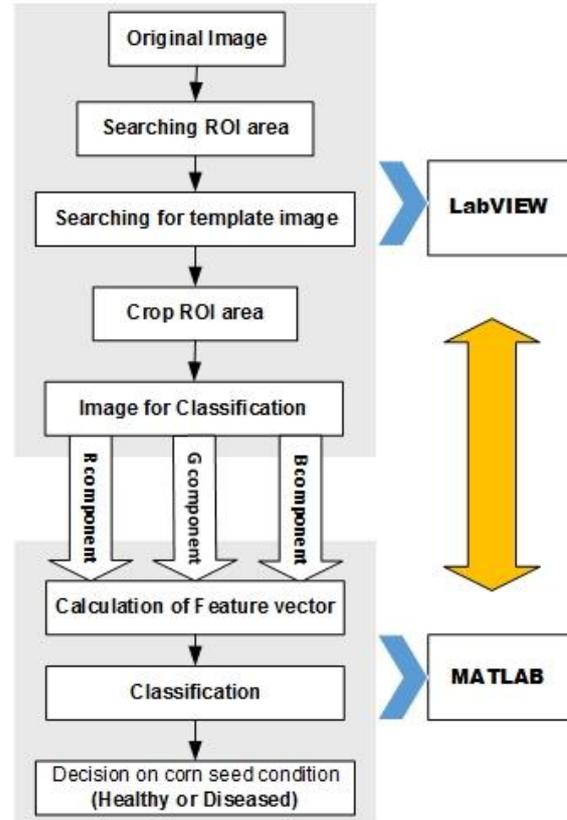


Fig. 6. Structure and working algorithm of the virtual instrument for Fusarium damage detection

2.3.3. Level User Interface

Figure 8 shows the front panel of the developed LabVIEW program for real – time image acquisition and system control.

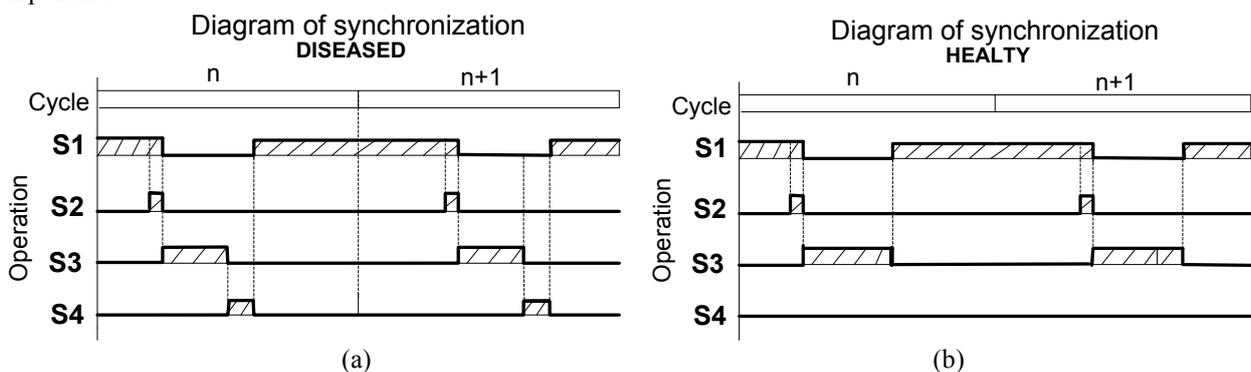
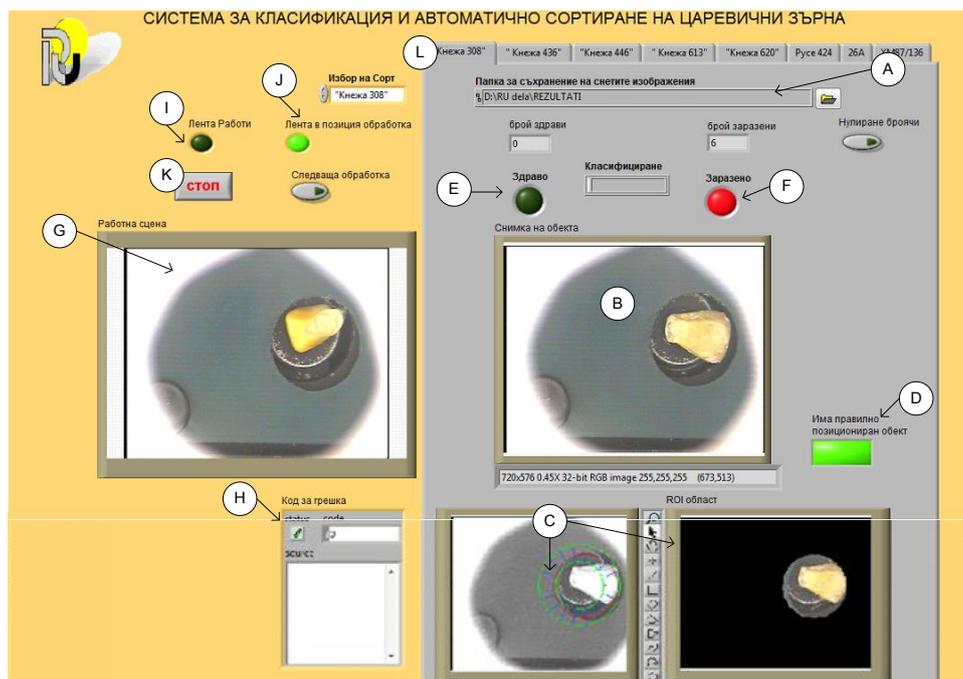


Fig.7. Diagram of synchronization of the system actions in case of healthy (a) or diseased (b) seed; Operation: S1 – motion of conveyor belt; S2 – position sensor trigger; S3 – image capture and decision trigger; S4 – valve trigger



**Fig.8.** Front panel of the developed virtual instrument

The left side of the VI shows the real-time field of view of the camera (Fig.8-G); the field ‘Error Code’(Fig.8-H), which indicates if there is an error obtained while executing program code; indications ‘Working Conveyor’ (Fig.8-I) and ‘Conveyor in position for processing’ (Fig.8-J); Stop button (Fig.8-K) .On the right there is the main menu for corn variety selection (Fig.8-L), by default this is variety Kneja 308. All of the available submenus allow performing the following actions: Selection of folder for storing the images from the available folders on your computer (Fig. 8- A); Visualization image of the last classified corn seed (Fig. 8- B); Visualization of the results of the procedure for searching correctly positioned object in front of the camera (ROI – Region of interest) (Fig. 8– C); LED indication for the presence of a properly positioned object in the ROI (Fig.8–D); LED indication of the classification procedure result - green indication for sound seed and red indication for damaged seed (Fig. 8– E); Fields for automatically counting classified sound and damaged seeds with option for zeroing the fields before start working (Fig.8-F).

### 3 Results and discussion

#### 3.1. Results from Feature Subset Selection

The computations were performed in MATLAB environment (Mathworks, USA) and STATISTICA (Six Sigma Version 8, StatSoft).

**Table 1** Feature selection results for different varieties

Variety	Feature selection method		
	GDA	SFR	BFC
<b>Kneja 308</b>	G,H,S,V,L,a,b, Y,Ycbcr,xm,ym	B,S,b, Z,cb,xm,ym	B,S,Z
<b>Kneja 436</b>	B,H,S,V,L,a,b,Z,cb,xm,ym	B,S,b,Z,cb,xm,ym	B,b,Z
<b>Kneja 446</b>	G,H,S,V,L,a,b, Y, Ycbcr, xm,ym	B,H,V,a,X,Z,ym	V,X,Z
<b>Kneja 613</b>	B,H,S,V,L,a,b,Z,cb,xm,ym	G,H,S,b,cb,xmy m	G,S,b
<b>Kneja 620</b>	B,H,S,V,L,a,b, Z,cb,xm,ym	H,S,b,X,Y,cb,xm,ym	H,X,cb
<b>XM87/136</b>	B,H,S,V,L,a,b, Z,cb,xm,ym	G,L,a,b,Y,cb,ym	b,Y,ym
<b>Ruse424</b>	H,S,V,L,a,b,X, Ycbcr,cr, xm,ym	S,L,b,Y,Ycbcr,cb,ym	L,Y,ym
<b>26A</b>	G,H,S,V,L,a,b, Y, Ycbcr, xm,ym	B,H,S,V,b,Z,ym	B, b, ym
<b>Kneja 433</b>	B,H,S,V,L,a,b, Z, cb,xm,ym	H,S,a,b,Ycbcr,cb,ym	H,S,a
<b>Kneja 449</b>	B,H,S,V,L,a,b, Z, Ycbcr,xm,ym	H,S,V,a,b,cb,ym	H,S,a
<b>Kneja 451</b>	B,H,S,V,L,a,b, Z, Ycbcr,xm,ym	G,H,S,b,cr,xm,ym	G,H,S
<b>Kneja 452</b>	B,H,S,V,L,a,b,X, cb,xm,ym	G,H,a,b,X,cb,ym	H,X,ym
<b>Kneja 501</b>	B,H,S,V,L,a,b,Y,cb,xm,ym	G,S,L,b,Y,cb,ym	L,Y,ym
<b>Kneja 515</b>	H,S,V,L,a,b, Z, Ycbcr, cb,xm,ym	R,G,V,b,X,cb,ym	V,b,X
<b>Kneja 549</b>	B,H,S,V,L,a,b,Z,Ycbcr, xm,ym	B,S,a,b,cb,xm,ym	S,a,b
<b>PR35 P12</b>	G, H,S,V,L,a,b, Y,Ycbcr,xm,ym	R, V,a,b,X,Y,cb,ym	a,b,ym

First color features were compared by applying one-way ANOVA (Analysis of variance) method followed by Tukey's post-hoc test using the software STATISTICA 8, with a 0.05 significance level. Applied ANOVA [24] showed that the factor variety has a statistically significant impact over all 17 color features, both healthy and infected seeds. This is also an indication that the coating on corn seed that is formed as a result of the disease Fusarium Moniliforme has different color characteristics for each variety.

By observing Table 1, it can be seen that there is no repeatability in the feature sets for different varieties, excepting Kneja 433 and Kneja 449 by method BFC and Kneja 501 and Ruse 424 by the same method. The features ym (xyY), S(HSV) and b (from Lab model) which contribute distinguishing between two classes are met in feature sets for 6 of varieties. The weakest features are: R(RGB), x(xyY) and Ycreb, cr(yCrCb), they are not included in more than one set.

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 and 3, the method Best feature combination. Results for all data sets and varieties when samples –germ side and pericarp side are combined into a single sample are given in Table 1.

### 3.2. Results from classification

SVM 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 without any involvement of the designing procedure proposed was estimated. The implementation of the method was made for radial basis function (RBF) with a width  $\sigma$  and is formulated as

$$K(x, y) = \exp(-\|x - y\|^2 / (2\sigma^2)) \quad (10)$$

The optimal values of the regularization constant  $C$  and the kernel width have been selected experimentally. 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  $C = 1 \div 10$  (minimum- 1, maximum - 10, with the increment unit), and for gamma - in the range of 0.091 to 0.333. The objective here is to combine a set of features and the most appropriate classifier for each one in order to improve the overall classification accuracy. Table 2 presents the average test data set classification accuracy obtained for all the varieties sets from a single SVM when using different combinations of features.

The evaluated results listed in Table 2 are the averaged values of total error rate of 16 varieties: 7.7%, dealing with feature set from GDA method; 7.8% - dealing with feature set from SFR method and 8.2%, - dealing with feature set from BFC method.

**Table 2.** Classifier performance results using Support vector machines

Seed Variety	Feature selection method		
	GDA	SFR	BFC
	Total error $e_0, \%$	Total error $e_0, \%$	Total error $e_0, \%$
Kneja 308	24	28.7	26.7
Kneja 436	4.3	5	6.4
Kneja 446	9.3	10.7	12
Kneja 613	4.7	3.3	3.3
Kneja 620	13.3	10.67	10
XM87/136	18.3	16.7	16.7
Ruse424	2.2	2.2	2.2
26A	4	6	5.93
Kneja 433	6.7	6.7	5.9
Kneja 449	5.7	5.7	7.6
Kneja 451	8.8	8.8	6.4
Kneja 452	3.2	3.2	3.2
Kneja 501	7.7	5.7	5.7
Kneja 515	3.85	3.8	3.8
Kneja 549	1.6	1.6	1.6
PR35 P12	5.2	5.22	13.04
Average	7.7%	7.8%	8.2%

**Table 3.** Classifier performance results using k-nearest neighbor method

Seed Variety	Feature selection method		
	GDA	SFR	BFC
	Total error $e_0, \%$	Total error $e_0, \%$	Total error $e_0, \%$
Kneja 308	18.7	20	30
Kneja 436	4.6	4.6	6.1
Kneja 446	8.7	8.7	15.3
Kneja 613	3.3	2	2
Kneja 620	10.7	10.7	9.3
XM87/136	23.3	20	18.3
Ruse424	2.2	2.2	2.2
26A	4.7	6	6.7
Kneja 433	7.4	2.9	3.7
Kneja 449	7.6	6.7	5.7
Kneja 451	9.6	8.8	7.2
Kneja 452	2.4	3.2	1.6
Kneja 501	8.6	10.5	11.4
Kneja 515	3.8	2.9	2.9
Kneja 549	3.2	3.2	3.2
PR35 P12	3.5	3.5	8.7
Average	7.6%	7.2%	8.4%

Regarding k-NN classifiers, results were obtained on the following terms: for each variety optimal number of nearest neighbors - k is selected experimentally; selected is k, which received the highest accuracy of cross-validation, as tested in five variants, respectively for k = 1,2,3,4 and 5; the measure of similarity, which is the second parameter setting k-NN classifier and depends on the characteristics of the input data in this case is chosen to be the Euclidean distance as (6); V-fold cross validation (V=10) was applied as well.

$$d(X, Y) = \sqrt{\sum_{i=1}^n (X_i - Y_i)^2} \tag{11}$$

The evaluated results listed in Table 3 regarding classification with k-nn includes averaged values of total error rate of 16 varieties respectively: 7.6%, dealing with feature set from GDA method; 7.2% - dealing with feature set from SFR method and 8.4%, - dealing with feature set from BFC method.

In practice in the field of recognition and sorting food stuffs, mostly the performance and applicability of classification algorithms are evaluated by various error of classifiers.

The evaluation of the classification error as the ratio of the wrongly classified samples to the total number of samples is often used with the status of a standard criterion for evaluating the quality of recognition systems. In the practice of automatic grading and sorting of food products, the classes are more than two, but the evaluations of the credibility of recognition in general include mainly two types of errors: highway and marginal [4].

In accordance with the above, the proposed classification approaches, which are integrated into the system are tested with a control sample of seed and evaluated by the computed values for total error  $e_0$  (9), highway errors  $e_i$  (12) and marginal errors  $g_i$  (13). Highway (major, basic) error is calculated according to the expression:

$$e_i = 100[(\sum_{k=1}^N m_{ik}) - m_{ii}] / m_i, [%] \tag{12}$$

where: i = 1, 2, ...,N - is the number of the respective class;  $m_i$ - the number of objects, assed on the basis of expert evaluation (EE) in class i;  $m_{ik}$ - the number of objects according EE in class i, which after the algorithmic classification (AC) have ended up in class k by mistake; $m_{ii}$ - the correctly identified objects in the class with the number i.

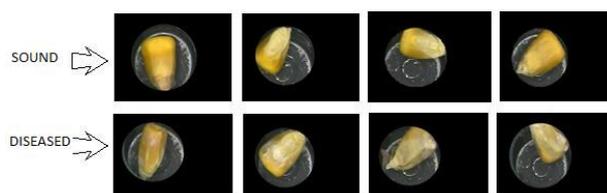
Marginal (alternative, field) error is determined by the expression:

$$g_i = 100[(\sum_{k=1}^N m_{ki}) - m_{ii}] / \sum_{k=1}^N m_{ki}, [%] \tag{13}$$

In a synthesized form summarizing the bialternative case of error, the highway error shows how

many objects have mistakenly gone to other classes, i.e. how many objects are not assigned to the class to which they actually belong, and the marginal – how many objects have mistakenly come from other classes and in this sense are equivalent to the type I and type II errors in mathematical statistics.

A series of tests have been conducted to validate the property of the automated grading system. A control sample of 230 corn seeds (120 healthy and 110 infected seeds) from a variety Kneja 451, were tested in automatic mode. When submitting in the inspection area, seeds are not steered in advance. Figure 9 shows images of sound and damaged corn seeds, captured during the experimental study with automated system.



**Fig.9.** Images of inspected seeds in grading system

Tables 4 and 5 represents Tables of errors, which summarize the accuracy results obtained by the system using previously proposed approaches:

- Classification with SVM classifier and three color features from method BFC
- Classification of seeds with k-nn classifier using three color features, which is ready to use instrument in Labview environment.

**Table 4.** Performance of the designed system with control set of seeds using SVM classifier and three color features.

Actual class	Predicted class		Highway (basic)error $e_i$ , %	Time for execution
	Sound	diseased		
Sound	114	6	5	9.8 s
diseased	10	100	9.09	9.8 s
<b>Marginal error, <math>g_i</math>, %</b>	<b>8.1%</b>	<b>5.7%</b>	<b>Total error</b>	<b>6.95 %</b>

**Table 5.** Performance of the designed system with control set of seeds using k-nn classifier

Actual class	Predicted class		Highway (basic)error $e_i$ , %	Time for execution
	Sound	diseased		
Sound	89	31	25.8	0.95 s
diseased	16	94	14.5	0.95 s
<b>Marginal error, <math>g_i</math>, %</b>	<b>15.2 %</b>	<b>24.8 %</b>	<b>Total error</b>	<b>20.4 %</b>

The time for analysis and classification of the current image is mainly responsible for the timing of overall

process for one seed. It determines the system's performance at all. In order to measure the speed of image processing, a counter block was placed in the Labview instrument. At the beginning of the image analysis, the procedure triggered the time counter. At the end of the image processing, a new signal stopped the clock, thus allowing us to measure the time that had elapsed.

The k-nearest neighbour classifier (k- nn) produced the best time, but also had highest value of total error (20,4%) and highest value of marginal error for class 'sound' – 15,2% (Table 5).

Analysis of results demonstrated that using SVM classifier the total error rate is lower (6.9%) compared to k-nn with the same number of features.

#### 4 Conclusion

The variety of corn seeds influences significantly on the color features of sound kernels as well as of the diseased area (*Fusarium Moniliforme*) of kernels. The attached two-factor analysis of variance showed that the factors variety and side of seed have a statistically significant impact on the studied 17 color features, both for healthy and infected seeds.

Different approaches are proposed for the classification task including feature subset selection by General Discriminant Analysis, Fisher's discriminant ratio together with Scatter matrices - J3 criterion and classification with k-nn and SVM. Tests analysis carried out with sixteen different varieties of corn seeds (Kneja 308, Kneja 436, Kneja 613, Kneja 620, 26A, Ruse 424, XM 87/136, ECO 85, Kneja 433, Kneja 449, Kneja 451, Kneja 452, Kneja 501, Kneja 515, Kneja 549 and PR35 P12) demonstrated that the varietal identity significantly influenced the selection of informative subsets of color features for classification of corn kernels. Tests analysis of impact of the side of capturing of seed on the accuracy of the classification procedure demonstrated that this influence could be compensated using the proper combination of feature set and classifier, corresponding to seed variety.

A comparative study to assess the impact of the methods used for feature selection and classification on recognition accuracy is made. It has been found that classification accuracy for k-nn and SVM without taking into account the side of capturing in the preliminary research reaches total error rate  $8.4 \div 7.2\%$  depending on seed variety. The classifier performance is estimated and results showed that the lowest total error rate of 2-20% according to the variety is achieved by classification with k-nn classifier and features selected by Scalar Feature ranking method.

A computer-based system, which has been reported in this paper allows real-time recognition and grading of corn seeds with external signs of *Fusarium Moniliforme* disease in laboratory practice. The designed automatic system has successfully incorporated visible-light imaging techniques to a modular mechatronic system

with quite precise functions of motion control, signal processing, and machine vision, based on LabVIEW development platform and MATLAB workbench. The proposed virtual instrument combines an easy realization of a graphical user interface, camera control unit and complex mathematical analysis. It allowed for the classification of the 16 varieties of corn seeds. In terms of processing time the k-nearest neighbour classifier is leading by the fact that it is used a k-nn built-in Labview environment module for classification. The classifier performance is estimated also experimentally with the system during control measurements of seed sample (variety Kneja 451) and results showed that total error rate is 6.95% for SVM classifier and 20.4% for k-nn classifier.

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. Bishop Christopher M., Pattern recognition and machine learning (Springer- Verlag., 2006)
2. Chen X., Y. Xun et al., Combining discriminant analysis and neural networks for corn variety identification, *Computers and Electronics in Agriculture* **71S**, S48–S53, (2010)
3. Choundary R., J. Paliwal, D.S. Jayas, Classification of cereal grains using wavelet, morphological, colour and textural features of non-touching kernel images, *Biosystem Engineering* **99**, 330-337, (2008)
4. Damyanov Ch., Analysis and evaluation of errors in training and testing of recognition systems, *Information, communication and control systems and technologies*, **1**, 13-19, (2012)
5. Delwiche, S.R., Pearson, T.C., Brabec, D.L., High-speed optical sorting of soft wheat for reduction of deoxynivalenol. *Plant Disease* **89**:1214-1219, (2005b)
6. Duda R., P.Hart, D.G. Stork, Pattern classification – 2nd Edition, (A Wiley- Interscience Publication, USA, 2001)
7. Granitto Pablo M., H.D. Navone, P.F. Verdes, H.A. Ceccatto, Weed seeds identification by machine vision, *Computers and Electronics in Agriculture* **33**, 91-103, (2002)
8. Gunasekaran Sundaram, Computer vision technology for food quality assurance, *Trends in Food science & Technology*, Vol.7, 245-256, (1996)
9. Guyon I., Elisseeff, An introduction to variable and feature selection, *Journal of Machine Learning Research*, Vol.3, 1157-1182, (2003)
10. Majumdar S., D.S. Jayas, Classification of bulk samples of cereal grains using machine vision, *J.Agric. Engng. Res.* **73**, 35-47, (1999)

11. Mollazade K., M. Omid, A. Arefi, Comparing data mining classifiers for grading raisins based on visual features, *Computers and Electronics in Agriculture* **84**, 124-131,( 2012)
12. Pearson T., Hardware based image processing for high-speed inspection of grains, *Computers and electronics in agriculture* **69**, 12-18, (2009)
13. Pittet A, Natural occurrence of mycotoxins in foods and feeds – an updated review. *Rev. Méd. Vét.* **149**: 479–492, (1998)
14. 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**, 380-387, (2010)
15. 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**,107-112, (2011)
16. 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)**, 235 – 240, (2001)
17. Theodoridis S., K. Koutrombas, *Pattern Recognition, Fourth Edition*, (Academic Press, 2009)
18. Visconti A., Problems associated with Fusarium mycotoxins in cereals. *Bulletin of the Institute for Comprehensive Agricultural Sciences Kinki University* 9, p. 39-55, (2001)
19. Visconti A., A. Girolamo, V. M. T. Lattanzio, V. Lippolis, M. Pascale, M. Solfrizzo, Novel analytical methods for Fusarium toxins in the cereal food chain *Colloque Fusariotoxines des Céréales – Arcachon – (11–13 septembre 2007)*
20. Webb Andrew, *Statistical Pattern Recognition, Second Ed*, Wiley, (2002)
21. Williams P., M. Manley, G. Fox, P. Geladi, Indirect detection of *Fusarium verticillioides* in maize (*Zea mays* L.) kernels by near infrared hyperspectral imaging, *J. Near Infrared Spectrosc.* **18**, 49-58, (2009)
22. Yang I –Chang, S. Delwiche, S. Chen, Y. Martin Lo, Enhancement of *Fusarium* head blight detection in free-falling wheat kernels using a bichromatic pulsed LED design, *Optical Engineering* Vol. **48(2)**, 023602-1 – 023602-11, (2009)
23. Yorulmaz Onur, T.C Pearson, A. E. Cetin, Detection of fungal damaged popcorn using image property covariance features, *Computers and Electronics in Agriculture* **84**, 47- 52, (2012)
24. Kirilova E., P. Daskalov, R. Tzonev, C. Draganova, Statistical characteristics of images of sound and *Fusarium* diseased sorn kernels, *Trakia Journal of Sciences*, Vol. **7**, Suppl. 2, pp 161-167, (2009)
25. Lattanzio V.M.T., Solfrizzo M., Powers S., Visconti A., Simultaneous determination of aflatoxins, ochratoxin A and *Fusarium* toxins by liquid chromatography/tandem mass spectrometry after multitoxin immunoaffinity clean up. *Rapid Communications in Mass Spectrometry*, 21, 3253-3261, (2007).
26. Maragos C.M., M. Busman, Rapid and advanced tools for mycotoxin analysis: A review, *Food additives and contaminants*, Vol 27(5), 688- 700 (2010)
27. Pittet A., Modern methods and trends in mycotoxin analysis, 117th annual conference of the society of food environmental chemistry, 8-9september, p. 424-444, (2005).