An Approach for Identifying of *Fusarium* Infected Maize Grains by Spectral Analysis in the Visible and Near Infrared Region, SIMCA Models, Parametric and Neural Classifiers

Tsvetelina Draganova*, Plamen Daskalov, Rusin Tsonev

Department of Automatics, Information and Control Engineering, University of Rousse
8 Studentska Str., 7017 Rousse, Bulgaria
E-mail: cgeorgieva@ru.acad.bg

* *Corresponding author*

Received: March 23, 2010 **Accepted: June 24, 2010**

Published: July 30, 2010

Abstract: *An approach for identifying of Fusarium infected single maize grains based on diffuse reflectance in visible and near infrared region is proposed in the paper. Spectral characteristics were collected in the range 400-2500 nm in steps of 2 nm. Soft independent modeling of class analogy (SIMCA) is used for data processing. Maize grains classification is based on SIMCA classifier and Probabilistic neural network (PNN).* Recognition accuracy which is achieved for both classes of grains is respectively 99.89% for healthy, and 93.7% for infected.

Keywords: *NIR spectroscopy, Maize grains, SIMCA method, Probabilistic neural network, K-means classifier.*

Introduction

Maize is one of the main products used in agriculture. In food industry is used to produce corn flour, starch, glucose, corn oil, corn meal and others. Seeds are subject to inspection for purity, germination, moisture and disease contaminatin. A number of requirements are defined in Bulgarian standards and are used when evaluating the quality of various agricultural products – fruits, vegetables, cereals and more.

Most important disease is *Fusarium spp. Fusarium verticillioides* is toxic for both people and animals. Methods for diagnosing the *Fusarium* disease are divided into two main groups – non-destructive and destructive. Destructive methods are based on chemical analysis. Methods based on gas-liquid chromatography and chromatography [2, 11, 16, 18] and overall spectrometry are the most often used and accurate, but they are uncomfortable when analyzing a large amount of sample. Inspection of seed processing requires a large amount of samples, limited funds invested and short processing time of the data that these methods do not allow.

Another similar method is fluorimetric [9]. This method is able to detect very low levels of *Fusarium*. Time for individual analysis of one seed is less than 15 min. A disadvantage of this method is the use of specific apparatus and its preliminary study within 20 min before analyzing each sample.

Non-destructive methods for diagnosis of *Fusarium* are possible to reduce the time for sample preparation and determination of the disease. There are three technologies for the detection of *Fusarium*: electronic nose, near infrared spectroscopy and image analysis of whole grain. The electronic nose resembles the human nose and accurately identified 91% taking the amount of toxins in the sample, but after 48 hours. Wiwart [22] used digital image processing of *Fusarium culmorum* and *Fusarium avenaceum* infected corn seed. HSI color model was used for representing of all samples of infected and healthy seeds. Wiwart defined the limits of the three components for healthy and infected seeds. The distribution of seeds according to the values of H and S makes it possible to distinguish samples are carried by almost 90% success rate. This method gives good results provided that the external manifestation of the disease – the color of the coating is clearly distinguishable from the color of the seed.

Diagnostics of *Fusarium* and levels of toxins in the seeds depends on many factors. The requirements for production of quality seeds are aimed at easy to use procedures, relatively fast data processing and low demands on the technical realization of systems for seeds diagnostics. The degree of precision and accuracy of methods which are used are of great importance.

Both methods – near infrared spectroscopy and analysis of the images have great potential in the process of diagnosing the seeds. Both methods are non-destructive and easy to implement and can explore and evaluate the *Fusarium* disease. Main steps in both methods are presented on Fig. 1.

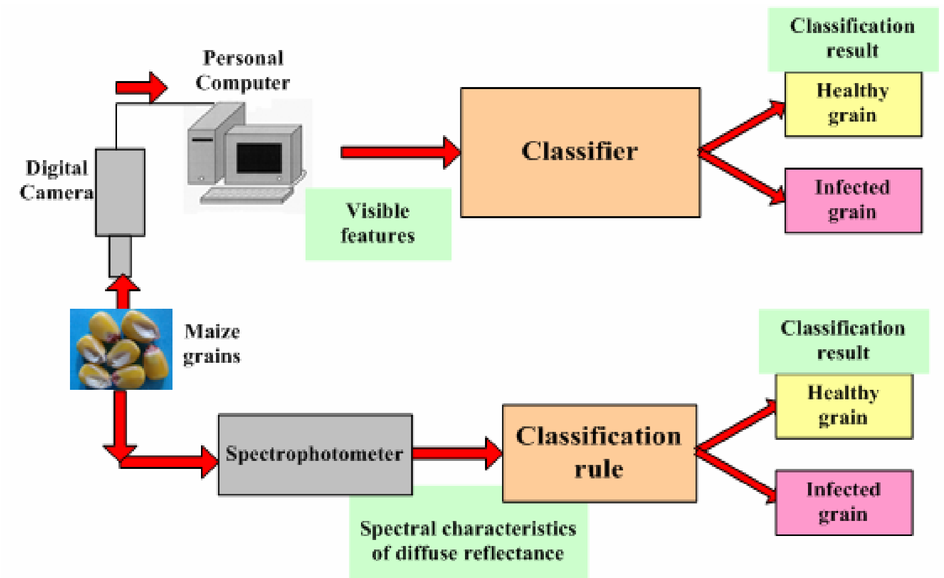

Fig. 1 Main steps in both methods – near infrared spectroscopy and analysis of the images

The first approach analyze the external features of the disease and includes the following steps: receiving digital image in RGB color model, its preliminary processing and conversion into other color models with a personal computer [4], select the image for signs of disease [7], determining the status of seed (healthy or infected) with an appropriate classifier [8]. Using a classifier, based on Probabilistic neural network (PNN), classification accuracies were 98% (healthy) and 94% (disease corn seeds). The classification accuracies achieved by using a classifier, based on Fuzzy logic were respectively 100% (healthy) and 80% (disease corn seeds). Disadvantage of the method based on image processing is applicable only in cases when grains have visual infection features.

The second approach which analyze the internal quality and structure of grains is based on spectral analysis. In recent years, spectroscopic techniques in the visible (VIS) and near infrared (NIR) are considered as an alternative to traditional laboratory methods of analysis for a number of advantages – speed and non-destructives [1, 5, 6, 10, 20, 21]. Proposed method includes: obtaining of spectral characteristics of diffuse reflection of maize grains by spectrophotometer; principal component analysis, definition of rule based on these components to determine the status of maize grain.

Method based on grain spectral characteristic not depends on available visible features.

Therefore an approach based on visible and near infrared spectroscopy, Soft independent modeling of class analogy (SIMCA) modeling of spectral data and neural network is proposed in the paper.

Spectral data processing

Maize sample

The maize sample was prepared by Cereal Research Centre and the grains were selected from the most popular variety of corn in Bulgaria. The sample was divided into two sets-training and test. Training set was used for training the classifiers and the test set was used for assessment the recognition accuracy. Training set includes 600 spectra from healthy and diseased maize grains. Test set includes 300 spectra from healthy and diseased maize grains.

Spectral characteristic acquisition

Spectra of diffuse reflectance were obtained with an automated system for spectral measurements which includes NIRSystem 6500 spectrophotometer (FOSS NIRSystem, Silver Spring, MD, USA) and personal computer (shown in Fig. 2). The spectrum was measured in 1050 points in the range from 400 to 2498 nm in 2 nm.

Fig. 2 Automated system for spectral data acquisition

Methodology, which has collected spectral characteristics of the grains (Fig. 3) includes the following steps:

- two sets maize grains (healthy and diseased);
- each set consists of 150 seeds;
- each spectral characteristics is acquired from both grain sides (of the germ and the other side);
- thus forming 300 spectral characteristics of each grain group and after three repeated measurements for a group of healthy and infected grains group were collected in 900 spectral characteristics.

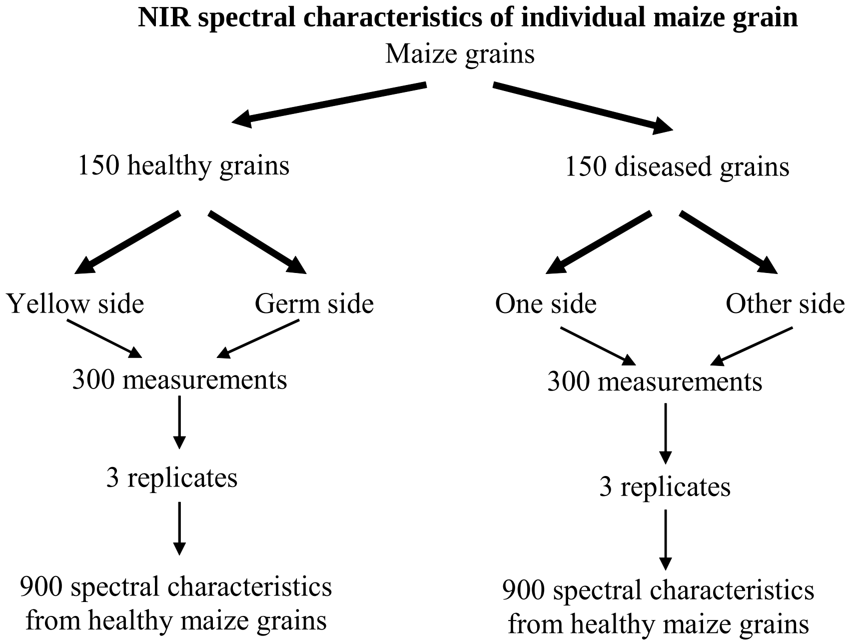

Fig. 3 Methodology, which has collected spectral characteristics of the grains

Acquired spectral data are shown in Fig. 4.

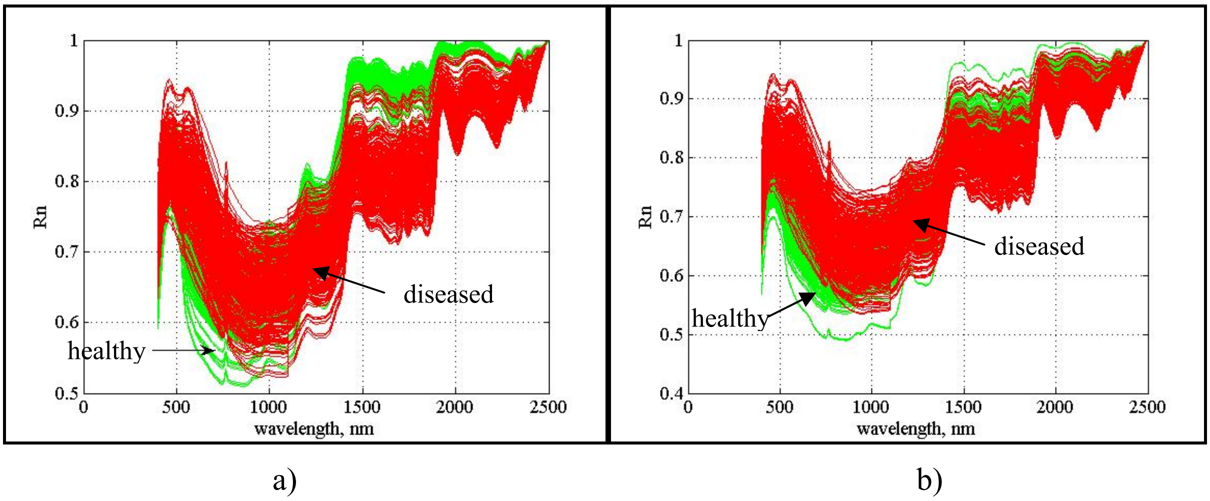

Fig. 4 Maize grain spectral data: a) obtained from the germ side, b) obtained from the other side of grain (green – healthy, red – *Fusarium* infected grains)

Obtained data show that they have similarly shape. The result is that the raw spectra couldn't be used as a classification feature.

Principal component analysis

Processing of spectral data using the method of principal components. It was used first for reducing the data (spectral characteristics have 1050 points) and second for obtained an informative components which will be usefully for classification of grains.

The analysis of main components (Principal component analysis – PCA) [14] is a powerful method for visualization and graphical representation of the relationship between samples and between variables. It also allows to reduce the dimensionality of the data. PCA finds linear combinations of original independent variables that describe the maximum proportion of variation. In the regression of the main components are first assessed the factors that describe the optimal spectral information for all samples. If each of the used wavelengths comparable to one coordinate axis and strike the resulting spectral values of each sample, it will match point in m-dimensional space (m-number of wavelengths used). The purpose of the receipt of principal components is to find another system of axes. The first axis (first principal component) corresponds to this direction, in which variations are greatest. The second axis is perpendicular to the first and in the direction in which the largest described by the first axis variations, etc. This process continues until all describe variations in the spectra through the main components. The percentage of scheduled part of the total variance decreases with increasing sequence number of the component. The spectrum of each individual sample is described by its coordinates with respect to this new set of axes.

A software program Pirouette Version 3.11 was used for spectral data analysis, The data were analyzed using SIMCA. This method used principal components as features for object classification.

Class projection in new set of axes – PC1 (principal component 1), PC2 (principal component 2) and PC3 (principal component 3) is presented in Fig. 5. The results show that after PCA it can be obtain absolutely definite classes of healthy and diseased grains which are described only by three features – first three components. Loadings for first three components are illustrated in Fig. 6.

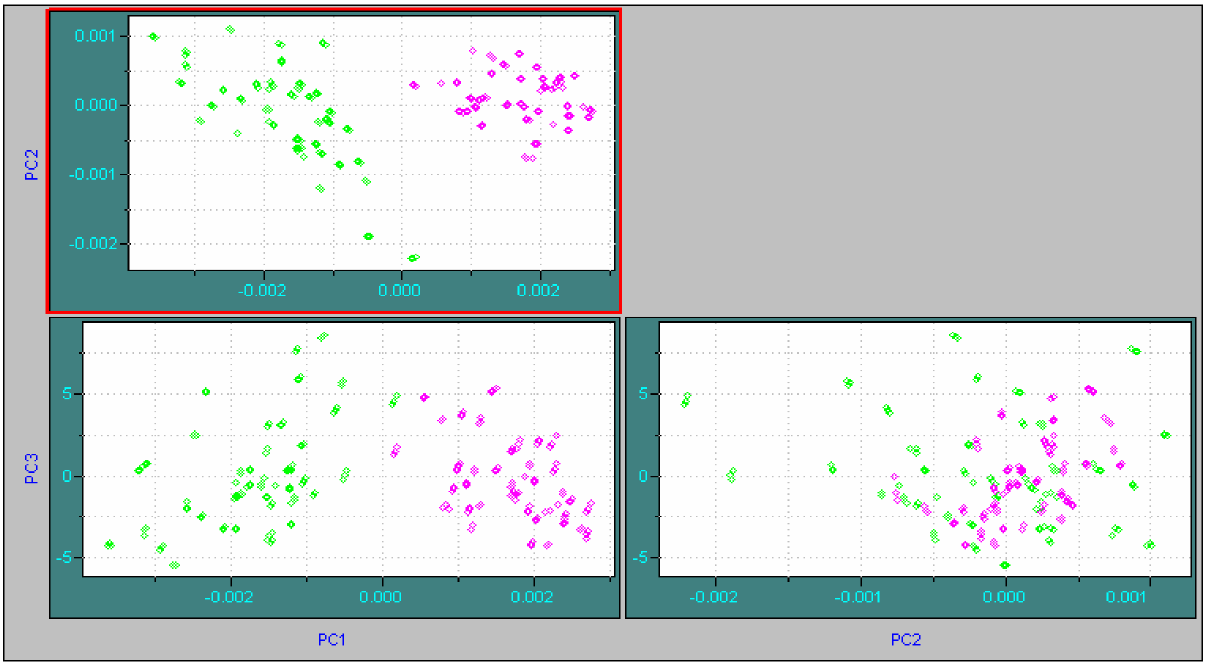

Fig. 5 Class projection in set of axes – PC1, PC2 and PC3

Soft independent modelling of class analogy (SIMCA)
SIMCA was implemented to create qualitative models of the respective classes based on NIR spectra of grains. Since that time functionality of the method has been demonstrated and enhancements offered by a number of researchers especially in food quality and safety [3, 15]. SIMCA is supervised methods [12] which is based on principals of similarity, i.e. it necessary some preliminary knowledge of samples. It implements on two steps. First, SIMCA develops models for each class based on PCA. PCA transforms the original data to a new coordinate

system such that the greatest variance by any projection of the data comes to lie on the first coordinate (called the first principal component), the second greatest variance on the second coordinate, and so on. Once each class has its own model, new samples could be classified to one or another classes according to their spectra. Samples from training set were used to develop SIMCA models for respective classes. The obtained models were tested using samples from independent test data sets.

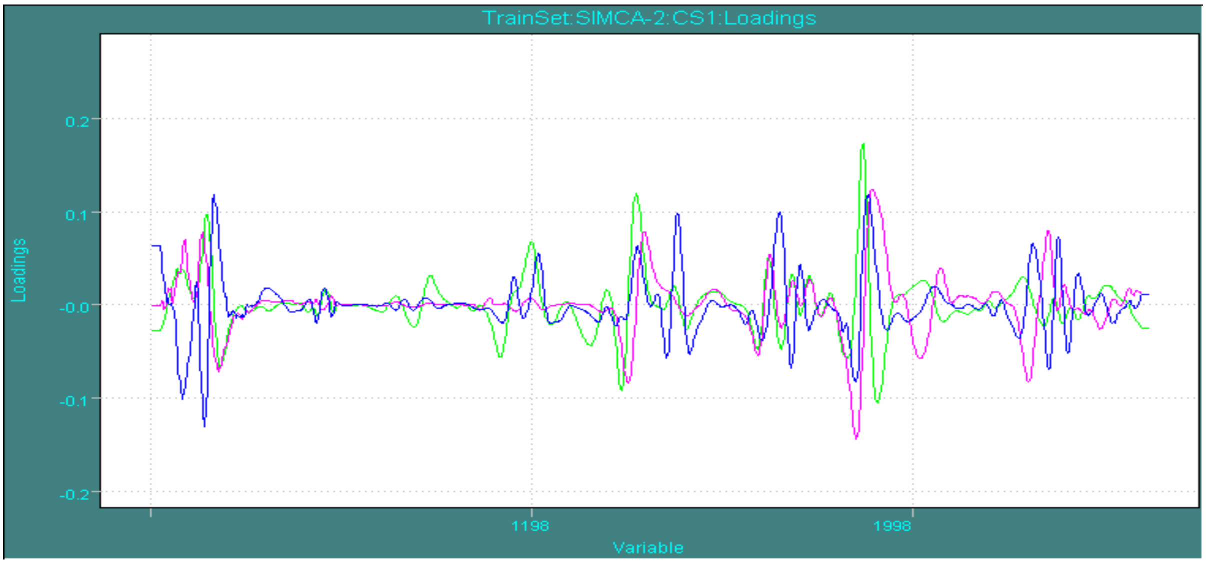

Fig. 6 Loadings for first three components

SIMCA provides information on wavelength, which is based on the distinction between classes discrimininating power and wavelength, affecting model modeling power and to meet what they (the water absorption, chemical compounds, proteins, fats, sugars composition of chitin and polysaccharides λ-glucan, present in cell walls of fungi such *Fusarium* etc.).

Parametric classifier
Classification is based on parametric approach of the type K-means. K-means [13, 19] is an unsupervised learning algorithms. The procedure classify a spectra data set through a certain number of classes (assume *k* classes) which are fixed priori – class healthy and class diseased grains.

The main steps of the K-means algorithm are:

1. Determine the coordinates of the center for each class (whichever is the center point).
2. Determine the distance from each object to the center.
3. The site is allocated to the class based on the minimum distance to the center.

Inputs parameters of the classifier are first three principal components of each spectra. MATLAB program was used for obtaining of principal components and for classifier realization.

Probabilistic neural network
PNN is used as a classifier. The inputs of the network are first three principal components. The network output is the grain kind – healthy or diseased. PNN classifier is realized in MATLAB. The PNN classifier has a parameter which is called smoothing factor. It is determined experimentally (Fig. 7) [17].

This smoothing factor σ determines the Gaussian window width. The optimal smoothing factor is defined as the values of σ from 0.01 to 10 with step 0.01 in percentage of correct recognition for seeds of representative samples for both classes healthy and infected. For class healthy seeds, with increasing σ, recognition accuracy remained constant – 100%, while for infected – grew up to a constant value – 80% for the σ increased from 0 to 0.3. For optimal value of the smoothing factor σ was determined 0.3, where the accuracy recognition for both classes healthy and diseased seeds is maximum.

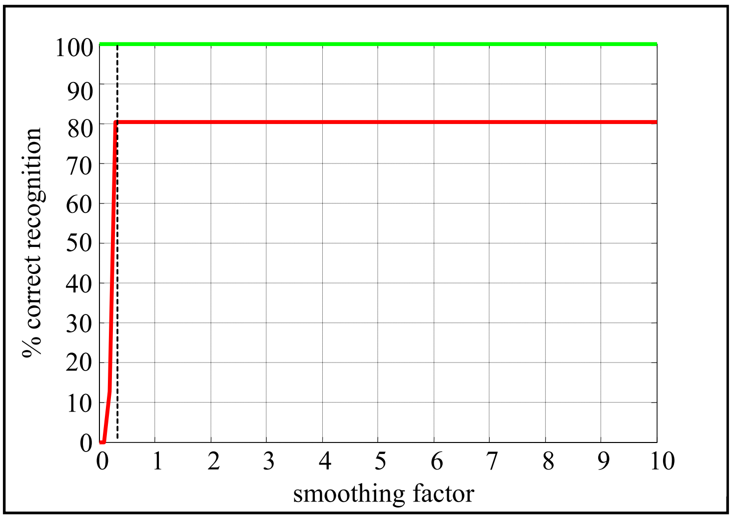

Fig. 7 Experimentally determination of smoothing factor

Results

Summary of classification results are presented in Table 1 and Table 2. Results of test set classification (spectral data were obtained from the germ grain side) are given in Table 1. Results of test set classification (spectral data were obtained from the other grain side) are given in Table 2.

Table 1. Results of test set classification (germ grain side)

Grain class	Class healthy			Class *Fusarium* diseased		
	N_c	N_{inc}	TP, %	N_c	N_{inc}	TP, %
SIMCA	148	2	98.7	144	6	96.0
PNN	149	1	99.3	148	2	98.7
K-means	143	7	95.3	136	14	90.6

Table 2. Results of test set classification (grain – other side)

Grain class	Class healthy			Class *Fusarium* diseased		
	N_c	N_{inc}	TP, %	N_c	N_{inc}	TP, %
SIMCA	149	1	99.3	140	10	93.3
PNN	149	1	99.3	149	1	99.3
K-means	142	8	94.6	138	12	92.0

The N_c is the number of corn kernels which are correct recognized, N_{inc} – the number of corn kernels which are incorrect recognized and *True positive* (TP) is calculated based on the following equations:

$$TruePositive = \frac{N_c}{Total\_class\_samples} \times 100,\ \%. \quad (1)$$

Obtained results show that SIMCA and PNN classifiers give similar and better classification accuracy than the parametric classifier K-means. There is a difference between the results of the germ and the other grain side. This means that the grain will be oriented for obtaining the spectral data. The best classification accuracy is achieved with PNN based classifier for class healthy and disease grains – 99.3% and 98.7% respectively. Similar result was achieved using SIMCA classifier – class healthy and disease grains – 98.7% and 96% respectively. K-means classifier result doesn't reach the accuracy which is defined in the Bulgarian standards.

Conclusion

An approach to identify single corn kernels infected with *Fusarium*, based on the absorption of seeds obtained by measuring their diffuse reflection in the visible and near infrared region is considered. Using SIMCA statistical models allow a relatively good accuracy to classify maize grains in two classes healthy and infected. It was found that using the SIMCA statistical models can be obtained and the characteristic wavelengths that are relevant for the separation of grains. The obtained results show that the position of grains (germ and other side) is a significant factor for disease detection, it is necessary prior orientation of the grains.The best accuracy of identification for both classes of maize grains – healthy and infected is achieved – 99.3% when the spectral data were obtained from the other side of the grain.

The next step of this work will be to develop new sophisticate methods for informative spectral data extracting which will be used as a feature for detection of *Fusarium* infected maize grains.

Acknowledgement

The study was supported by contract № BG051PO001-3.3.04/28, "Support for the scientific staff development in the field of engineering research and innovation". The project is funded with support from the Operational Program "Human Resources Development" 2007-2013, financed by the European Social Fund of the European Union.

References

1. Balasuriya J., T. Katsutomo, K. Shinichiro, M. Vassileva, S. Yoshida, R. Tsenkova (2010). Near Infrared Spectroscopy and Aquaphotomics: Novel Approach for Rapid in Vivo Diagnosis of Virus Infected Soybean. Biochemical and Biophysical Research Communications, 397(4), 685-90.
2. Berthiller F., C. Dall'Asta, R. Schuhmacher, M. Lemmens, G. Adam, R. Krska (2005). Masked Mycotoxins: Determination of a Deoxynivalenol Glucoside in Artificially and Naturally Contaminated Wheat by Liquid Chromatography − Tandem Mass Spectrometry, Journal of Agricultural Food Chemistry, 53(9), 3421-3425.
3. Cassells J., R. Reuss, B. Osborne, I. Weslwy (2007). Near Infrared Spectroscopic Studies of Changes in Stored Grain, Journal of Near Infrared Spectroscopy, 15, 161-167.
4. Daskalov P., R. Tzonev, T. Draganova, K. Arvanitis, N. Sigrimis (2004). Possibilities for Fusarium Disease Recognition of Corn Kernels by Color Image Analysis and Spectral Reflectance Processing, International Scientific Conference AgEng – Engineering the Future, Leuven, Belgium, September 12-16, 2004.

5. Damyanov C. (2006). Non-destructive Quality Recognition in Systems for Automatic Sorting of Food Products, UFT – Plovdiv.
6. Damyanov C. I., A. S. Georgiev, L. F. Kostadinova (2003). Non-destructive Technique for On-line Determination of Fruit and Vegetable Quality, Proceedings of the 3rd International Conference on Quality in Chains "Integrated View on Fruit and Vegetable", Wageningen, The Netherlands, Acta Horticulturae, 604(2), 577-584.
7. Draganova Ts. (2005). Analysis of Informative Color Features, used for Fusarium Recognition of Corn Seeds, Proceedings of University of Rousse, Bulgaria, 106-112.
8. Draganova Ts., R. Tsonev, P. Daskalov (2006). Corn Kernels Classification using Neural Networks and Fuzzy Logic, Proceedings of the Union of Scientists Rousse, Energy Efficiency and Agricultural Engineering, 681-688.
9. Hafner M., Z. Kubus, M. Freudenschuss, E. M. Binder, R. Krska (2008). Rapid Fluorometric Test for the Quantitative Determination of Deoxynivalenol in Raw Cereals, Journal of Mycotoxin Research, 23(1), 3-6.
10. Huang H., H. Yu, H. Xu, Y. Ying (2008) Near Infrared Spectroscopy for On/In-line Monitoring of Quality in Foods and Beverages: A review, Journal of Food Engineering, 87(3), 303-313.
11. Klötzel M., B. Gutsche, U. Lauber, H-U. Humpf (2005). Determination of 12 Type A and B Trichothecenes in Cereals by Liquid Chromatography − Electrospray Ionization Tandem Mass Spectrometry, Journal of Agricultural Food Chemistry, 53(23), 8904-8910.
12. Otto M. (2007). Chemometrics-Statistics and Computer Application in Analytical Chemistry, Wiley-VCH.
13. Patrick T. (2007). Visualization with Data Clustering, University of Fribourg, DIVA Seminar.
14. Ringnér M. (2008). What is Principal Component Analysis? Nature Biotechnology, 26(3), 303-304.
15. Roggo Y., L. Duponchel, C. Ruckebusch, J.-P. Huvenne (2003). Statistical Tests for Comparison of Quantitative and Qualitative Models Developed with Near Infrared Spectral Data, Journal of Molecular Structure, 654(1-3), 253-262.
16. Sokolović M., B. Šimpraga (2006). Survey of Trichothecene Mycotoxins in Grains and Animal Feed in Croatia by Thin Layer Chromatography, Food Control, 17(9), 733-740.
17. Steenhoek L. W., M. K. Misra, W. D. Batchelor, J. L. Davidson (2001). Probabilistic Neural Networks for Segmentation of Features in Corn Kernel Images, Engineering in Agriculture, 17(2), 225-234.
18. Sulyok M., F. Berthiller, R. Krska (2006). Rainer Schuhmacher Development and Validation of a Liquid Chromatography/Tandem Mass Spectrometric Method for the Determination of 39 Mycotoxins in Wheat and Maize, Rapid Communications in Mass Spectrometry, 20(18), 2649-2659.
19. Trygg J., E. Holmes, T. Lundstedt (2007). Chemometrics in Metabonomics, Journal of Proteome Research, 6(2), 469-479.
20. Tsenkova R., H. Meilina, S. Kuroki, D. Burns (2009). Near Infrared Spectroscopy using Short Wavelengths and Leave-one-cow-out Cross-validation for Quantification of Somatic Cells in Milk, Journal of Near Infrared Spectroscopy, 17(6), 345-351.
21. Veleva-Doneva P., Ts. Draganova, S. Atanasova, G. Beev (2008). Detection of Infected with *Staphylococcus* and *Streptococcus* Milk Samples by Spectral Analysis, Proceedings of University of Rousse, Bulgaria, 141-145.
22. Wiwart M., I. Koczowska, A. Borusiewicz (2001). Estimation of Fusarium Head Blight of Triticall using Digital Image Analysis of Grain, CAIP, LNCS, 2124, 563-569.

Assist. Prof. Tsvetelina D. Draganova, Ph.D.
E-mail: cgeorgieva@uni-ruse.bg

Dr Tsvetelina D. Draganova is an Assistant Professor of the Department of Automatics, Information and Control Systems, University of Ruse, Bulgaria. She has taught various subjects in Computer control systems and Systems for automatic control of environment parameters to different levels of students, ranging from Associated degree up to Master levels.

She holds the degrees of M.Sc. in Automatics and Computer Technology and Ph.D. Hers research interests are in Machine vision, Pattern recognition, Quality control of agricultural products using image and spectral data processing, Fuzzy logic and Neural networks. She takes part in project № BG051PO001-3.3.04/28 titled "Support for the scientific staff development in the field of engineering research and innovation" funded with support from the Operational Program "Human Resources Development" 2007-2013 and research project № DO 02-143/16.12.2008 titled "Development of Intelligent Technologies for Assessment of Quality and Safety of Food and Agricultural Products", financed by the National Science Fund of the Bulgarian Ministry of Education and Science from 2008 to 2011.

Assoc. Prof. Plamen I. Daskalov, Ph.D.
E-mail: daskalov@uni-ruse.bg

Dr Plamen I. Daskalov is an Associate Professor of the Department of Automatics, Information and Control Systems, University of Ruse, Bulgaria. He has taught various subjects in Computer systems for control and automation, Industrial communication networks, MATLAB Programming and Computer systems for measurement to different levels of students, ranging from Associated degree up to Master levels.

He holds the degrees of M.Sc. and Ph.D. in Automatics and Control Engineering. His research interests are in Quality control of agricultural products using visual images and spectral data processing; Application of information technologies in agriculture; Automatic control of the microclimate processes in livestock buildings; Application of computer systems in livestock breeding and agriculture; Identification and modelling of processes and equipment.

Assoc. Prof. Rusin S. Tsonev, Ph.D.
E-mail: rtzonev@uni-ruse.bg

Dr Rusin S. Tsonev is an Associate Professor of the Department of Automatics, Information and Control Systems, University of Ruse, Bulgaria. He has taught various subjects in System identification, Quality control, Specialized systems for management and control.

He holds the degrees of M.Sc. and Ph.D in Automatics and Computer Technology. His research interests are in System identification, Pattern recognition and Quality control of agricultural products.