Wavelet features selection approach for nondestructive Fusarium corn kernels recognition using spectral data processing

PLAMEN DASKALOV, TSVETELINA DRAGANOVA, VIOLETA MANCHEVA, RUSIN TSONEV Department of "Automatics, information and control engineering"

University of Rousse

8 Studentska Str., 7017 Rousse, BULGARIA

cgeorgieva@uni-ruse.bg http://www.uni-ruse.bg/faculties/ef/automation/site/index en.htm

Abstract: An approach for Fusarium diseased corn kernels recognition based on wavelet continuous transformation and probabilistic neural network (PNN) classification is presented in the paper. The near infrared diffuse reflectance characteristics are used as base for features extraction. The spectral data are fitted using continuous wavelet mexican hat transformations and their parameters are used for classification features. The range of the wavelet parameters "scale" and "time" are defined for seven most popular corn kernels variety in Bulgaria. A PNN classifier is created using wavelet coefficient $C_{a,b}$. The range of the PNN classifier smoothing parameter σ is determined experimentally. The classification accuracy is in the range of 75 to 100%.

Key-Words: Wavelet, Corn kernels, Fusarium disease, NIR spectral data, Recognition

1 Introduction

Fusarium disease, caused by several *Fusarium spp.*, is one of the major fungal diseases of cereals worldwide [9]. *Fusarium verticillioides* is the most significant, damaging and common pathogen in Bulgaria.

Three traditional diagnostics methods for identification of *Fusarium spp*. are used in Bulgaria. The first is based on visual assessment by expert. The second one is based on chemical analysis and the third is based on biology analysis of micro- and macro-morphological features of the fungal. Some other methods like fluorometric analysis, gas and liquid chromatography are reliable and accurately. The basic disadvantages of these destructive and non-destructive methods of detection are time consuming, requires highly skilled experts and each method varies in the requirements for technical investments. The comparative analysis indicate that the computer systems using image processing and spectral characteristics processing are the most common applied methods in precision farming and agriculture.

The Fusarium disease causes not only external but also internal changes in the corn kernels. The external changes are alterations in the shape (underdeveloped, wizened) and colors of the corns. Usually the 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.

The computer systems using image processing assess only the visible corn kernels changes. It is not enough to evaluate the internal quality of the kernels. Therefore near infrared spectroscopy (NIR) is used for assessment the corn kernels into two classes – healthy and diseased. The power of near infrared spectroscopy includes nondestructive qualitative and quantitative analysis [4,6,8]. In many cases, the most distinguished information is hidden in the frequency content of the signal. Fourier transform is a reversible transform, which goes back and forward between the raw and processed (transformed) signals. However, either of them is available at any given time. That is, no frequency information is available in the time-domain signal, and no time information is available in the Fourier transformed signal. The natural question that comes to mind is that is it necessary to have both the time and the frequency information at the same time. Fortunately, wavelet transform is capable of providing the time and frequency information simultaneously. Especially, the width of the window is changed as the transform is computed for every single spectral component, which is probably the most significant feature of the wavelet transform. Wavelet analysis generates an estimate of the local frequency content of a signal by representing the data using a family of wavelet functions that vary in scale and position. By using wavelet analysis, we can identify locally periodic trends in the signal. Therefore, this method is applied to analyze signals more and more [7,11]. In this paper we describe the use of wavelet transformation and NIR reflectance spectral data processing for recognition of healthy and Fusarium diseased corn kernels.

Quality of maize seeds is determined by a number of quality indicators, regulated in Bulgarian Standards [1].

According to Bulgarian Standards [2] allowed percentage of infected seeds in a sample is up to 5%.

The purpose of this article is to propose a method for effective classification of healthy and *Fusarium* infected corn kernels, which cover requirements of the Bulgarian quality standards and to determine the values of wavelet parameters "*scale - a*", "*time - b*", coefficient $C_{a,b}$ and probabilistic network smoothing factor for all corn variety.

2 Materials and methods

2.1 Corn samples

Images of healthy and *Fusarium verticillioides* diseased corn kernels are shown in Fig. 1.



Fig. 1. Images of corn kernels: a) healthy, b) *Fusarium verticillioides* diseased

Seven varieties of corn kernels were examined – Knezha 308, Knezha 436, Knezha 613, Knezha 620, 26A, XM87/136 and Ruse 424. They have been certified by the Maize Institute in the town of Knezha, Bulgaria since 2008. The corn kernels samples include Fusarium diseased and healthy corn kernels from the most popular variety corn kernels in Bulgaria. The expert assesses the variety of corn kernel – healthy or diseased and determines the variety of Fusarium species - *Fusarium verticillioides*. The set of samples was split into two subsets – train and test sets. The train set which includes 30 spectra from 30 healthy and 30 spectra from 30

diseased corn kernels of each corn variety was used to obtain the classification features. The test set (20 spectra from 20 healthy and 20 spectra from 20 diseased corn kernels of each corn variety) was used to test the accuracy of *Fusarium verticillioides* corn kernels recognition by NIR spectroscopy and wavelet transformation. All procedures are repeated for each side of the corn kernels – the side of the germ and the other side.

2.2. Instrumentation and NIR reflectance spectra

characteristics obtained Spectral were by spectrophotometer Ocean Optics in the visible and near infrared spectral area of 456 to 1140 nm. For each of the varieties have taken the spectral characteristics of intensity (Intensity) of 50 healthy and 50 infected grains for both sides - germ side and the other side. Total of 100 characteristics of healthy grains (50 - of the germ side and 50 of the other side) and 100 characteristics of contaminated grains (50 - of the germ side and 50 of the other side). The spectral characteristics of two corn kernels varieties are shown in Fig. 2. The characteristics of the other five varieties look similar.



a)corn variety Ruse 424



b) corn variety Knezha 613 Fig. 2 Spectral characteristics of intensity of 50 healthy and 50 Fusarium diseased corn kernels

2.3. Wavelets transformations

A wavelet is a mathematical function that can decompose a signal with a series of averaging and differencing calculations. They are able to differentiate and preserve details at various scales or resolutions.

Continuous wavelet transform (CWT) of a time-signal f(t) is defined by [12].

$$F_{\Psi}(a,b) = \frac{1}{\sqrt{a}} \int_{-\infty}^{\infty} f(t) \Psi\left(\frac{t-b}{a}\right) dt$$
(1)

The quantity $\Psi_{a,b}(t)$ given in the definition is referred to as the wavelet function. The *mexican hat family* was used as a wavelet function. This function:

$$\Psi(x) = \left(\frac{2}{\sqrt{3}}\pi^{-1/4}\right) \left(l - x^2\right) e^{-x^2/2}$$
(2)

is proportional to the second derivative function of the Gaussian probability density function.

The position variable b shifts the wavelet function along the time axis t of f(t) while the scale variable ^{*a*} expands or compresses the wavelet function $\Psi_{a,b}(t)$. Compared to Fourier transform, the scale variable a is equivalent to the inverse of the frequency. The definition also suggests that $F_{\Psi}(a,b)$ is the correlation coefficient between the wavelet function $\Psi_{a,b}(t)$ and the timesignal f(t) at the scale *a* and position *b* of $\Psi_{a,b}(t)$. The continuous wavelet transform was applied to the spectral reflectance data from healthy and Fusarium diseased corn kernels. A Matlab program was developed using Matlab R2007 (The MathWorksInc., Natick, MA, USA) for wavelet mathematically realization. CWT was applied to these signals to calculate the coefficients $F_{\Psi}(a,b)$, which were then displayed as a grey-scale map with the vertical and horizontal axes denoting respectively the scale a and position b, which is equivalent to time t, as shown in Fig. 3 and Fig. 4. Although there is general similarity between them, subtle differences are clearly visible at scale a in range of 14 to 80 and in time range from 40 to 250.

It is very important to define the values of a and b in which a maximum of the classification accuracy is achieved.

The values of b are obtained from *Discriminating power*, parameter in SIMCA [5]. For these values of b and for all range of scale a the percentage of correct recognition is achieved for both training samples – healthy and Fusarium diseased. This is shown on Fig. 5 and Fig. 6, where with green line are the values of correct percentage for healthy and with blue line – for diseased

kernels, in the left side. The red line shows the values of mean coefficients C_{ab} in the right side. Fig. 5 and Fig. 6 are for the corn variety Kneja 308, germ side and respectively for b = 81 and for b = 95.



Fig. 3. CWT coefficients - healthy corn kernels.



Fig. 4. CWT coefficients - Fusarium diseased corn kernels.



Fig. 5 The percentage of correct recognition for training samples – healthy and Fusarium diseased, and mean coefficients C_{ab} for the corn variety Kneja 308, germ side and b = 81



Fig. 6. The percentage of correct recognition for training samples – healthy and Fusarium diseased, and mean coefficients C_{ab} for the corn variety Kneja 308, germ side and b = 95

For all corn varieties the results are similar. The obtained results show that the maximum of the coefficients C_{ab} does not match to the maximum of the correct recognition in the range of scale a. Therefore the range of scale a where the percent correct recognition for the kernels from the training samples is 100 are used to define the value of a. This is shown in Table 1.

The empty cells in the columns for diseased kernels means that there the percent correct recognition is less than 100 and this range it is not used.

From these ranges we determine the value of "scale a" for each value of parameter "time b". The coefficients C_{ab} are calculated for each corn kernel (germ and the other side) for determined values of a and b for all corn varieties.



a) Germ side of corn kernel

Fig.7. C_{ab} coefficients for corn variety Kneja 308, b=67 and a=13 $\,$

| Table | 1 | Range | of scale a | for fixed v | alue of b |
|-------|-------------------------------|------------|-------------------|----------------|-----------|
| | 1 | Corn vari | ety Kneja 3 | 08 | |
| | | Range o | of scale a | | Defined |
| b | Other s | ide of the | Germ s | ide of the | value of |
| | ke | rnel | ke | rnel | a |
| (7 | healthy | diseased | healthy | diseased | 12 |
| 0/ | 9÷11 7×10 | - | 12-20 | 13-18 | 13 |
| 81 | /÷18 | - | 8-27 | 23-26 | 8 |
| 95 | - | - | 23÷34 | 33 | 33 |
| 301 | 24÷41 64÷175 | - | 6÷213 | - | 24 |
| 328 | 10÷141 | - | 18÷211 | - | 18 |
| 607 | 14÷47 | - | 21÷48 | - | 21 |
| | | Corn vari | iety Kneja 4 | 36 | |
| 38 | - | - | 3÷29 88÷102 | 2÷8 | 3 |
| 326 | 7÷27 | _ | 85÷92 | _ | 92 |
| 520 | $43 \div 84$ $02 \div 123$ | | 05*92 | | 2 |
| | 92-125 | Corn vari | l iety Kneja 6 | 513 | |
| 63 | 3÷25 | - | 3÷13 | 3÷21 | 3 |
| 196 | 1÷25 | 1÷24 | 133÷256 | - | 1 |
| 222 | - | 2÷17 | 117÷236 | 8 | 8 |
| 300 | 5÷106 | 7÷9 | 5÷186 | 5÷16 | 7 |
| 330 | 6÷140 | - | 5÷11 66÷174 | 3÷6 | 6 |
| 360 | 48÷136 | - | 76÷162 | 3÷20 | 76 |
| | | Corn vari | iety Kneja 6 | 520 | |
| 340 | 46÷95 | - | 75÷149 | - | 75 |
| | | Corn | variety 26A | | |
| 47 | 45÷64 | - | 22÷50 | 11÷26 30÷44 | 45 |
| 307 | 14÷15 29÷41 | - | 6÷8 | - | 29 |
| | 27.41 | Corn var | ioty Ruse 4 | 24 | |
| 66 | | | 5.22 | 17:20 | 17 |
| 00 | - | - | 3722 | 17÷20 | 1/ |
| 297 | 5÷25 39÷151 | - | 5÷189 | - | 5 |
| 332 | 77÷129 | - | 65÷174 | - | 77 |
| 597 | 17÷59 | - | 25÷54 | - | 25 |
| | | Corn varie | ety XM87/1 | 136 | |
| | 14÷20 | | | | |
| 84 | 37÷56 | - | 7÷19 | 4÷28 | 14 |
| | 78÷79 | | | - | |
| | 97÷125 | | | | |
| 182 | 148÷25 | _ | 49÷52 | 73÷76 | 148 |
| 2 | 2 | | 59÷256 | | 1.0 |
| | 6÷35 | | 31-60 | 3 | |
| 246 | 72÷88 | _ | 105÷239 | 107 ± 115 | 31 |
| 2 10 | 12.00 | | | 122 ± 170 | 51 |



b) Other side of corn kernel

Fig.7. C_{ab} coefficients for corn variety Kneja 308, b=67 and a=13

Red line is for diseased corn kernels and green – for healthy corn kernels. Two classes of corn kernels – healthy and diseased are not clearly separated. The results for the others corn kernel varieties are similarly.

2.4. Classification procedure

The coefficients $C_{a,b}$ have overlapped values therefore probabilistic classifiers are appropriate for this classification procedure. One of these classifiers is a probabilistic neural network (PNN) which is used for classification of the corn kernels.

The most important advantage of PNN is the simple structure, training manner and only one free parameter, the smoothing factor, which have to be adjusted by the user and this factor can be adjusted at run time without the requirement of network retraining.

The PNN has three layers. When an input is presented, the first layer computes distances from the input vector to the training input vectors, and produces a vector whose elements indicate how close the input is to a training input. The second layer sums these contributions for each class of inputs to produce a vector of probabilities as its net output. Finally, an compete transfer function on the output of the second layer picks the maximum of these probabilities, and produces a 1 for that class and 0 for the other classes.

The neural network has 1 hidden layer, 3 input nodes, and 2 outputs. The input node corresponds to the mean wavelet coefficient for both of classes – healthy and *Fusarium* diseased corn kernels. The outputs correspond to the two classes healthy and diseased corn kernels. A winner – take – all decision rule (largest output wins) gave the expected classification.

The activation function of a neuron, in the case of the PNN, is statistically derived from estimates of probability density functions (PDFs) based on training parameters [3]. Estimator for the PDF is:

$$p(x \mid s_i) = \frac{1}{(2\pi)^{m/2}} \sum_{j=1}^{n_i} exp\left[\frac{-(x - x_j^{(i)})!(x - x_j^{(i)})!}{2\sigma_i^2}\right]$$
(3)

where $p(x | s_i)$ is the probability of vector x occurring in $x_i^{(i)}$

set s_i, corresponding to the type of fault; $x_j^{v'}=j^{th}$ exemplar pattern or training pattern belonging to class s_i, type of fault; n_i is the cardinality of the set of patterns in class s_i; m is total number of training patterns; σ_i is smoothing parameter.

The summation layer is used to compute the sum from the previous layer:

$$p_{r}(s_{i} \mid x) = \frac{p(x \mid s_{i})p_{r}(s_{i})}{p(x)}$$
(4)

where $p_r(s_i | x)$, i=1,2,...,k is the priori PDF of the pattern in classes to be separated; $p_r(s_i)$ is priori probabilities of the classes; p(x) is assumed to be constant.

The decision rule is to select class S_i of the fault type, for which $p_r(S_i | x)$ is maximum.

The variable σ is a smoothing factor that, in effect, determines the Gaussian window width. For the twodimensional input case, a small value of σ gives distinct modes corresponding to the locations of training samples [10]. A smoothing factor value $\sigma=5$ was selected using the criterion of percent classification of the training dataset in range from 0.05 to 10.

3. Results

Summary of classification results are presented in Table 2. The N_c is the number of corn kernels which are correct recognized, *Total_class_samples* – the total number of corn kernels in the sample and True Positive (TP) is calculated based on the following equations:

$$TruePositive = \frac{N_c}{Total_class_samples}.100,\%$$
(5)

This parameter *TruePositive* is calculated for each of the corn kernels varieties, for each side of the kernel (germ and the other side), for each class – healthy and diseased, for determined values of wavelet parameters "scale a" and "time b" and for PNN smoothing factor value σ =5. The achieved percentage of correct recognition is in range from 75 to 100%. The grey rows in Table 2 show the values of the parameters *a* and *b* which allowed to achieve the percentage of correct recognition covered the requirements of the Bulgarian standards.

| Table 2 | Та | ble | 2 |
|---------|----|-----|---|
|---------|----|-----|---|

Summary of classification results

| iet | Value | <i>TruePositive</i> for the test samples, % | | | | |
|--------------|---------------|---|----------|------------------|----------|--|
| var | | Other side of the | | Germ side of the | | |
| N U. | 0I a and h | kernel | | kernel | | |
| c01 | a and b | healthy | diseased | healthy | diseased | |
| | b=67 | 100 | 85 | 95 | 100 | |
| | a=13 | | | | | |
| | b=81 | 75 | 80 | 100 | 90 | |
| | a=8 | | | | | |
| 308 | b=95 | 90 | 80 | 100 | 100 | |
| ćneja 3 | a=33 | 100 | | 100 | | |
| | b=301 | 100 | 85 | 100 | 95 | |
| ¥ | h = 229 | 100 | 80 | 05 | 80 | |
| | 0-320 3=18 | 100 | 80 | 95 | 80 | |
| | h = 607 | 100 | 75 | 100 | 75 | |
| | a=21 | 100 | 15 | 100 | 15 | |
| 36 | b=38 | 85 | 95 | 100 | 100 | |
| a 45 | a=3 | - | - | | | |
| neji | b=326 | 95 | 95 | 100 | 85 | |
| Kı | a=92 | | | | | |
| | b=63 | 100 | 90 | 100 | 95 | |
| | a=3 | | | | | |
| | b=196 | 100 | 100 | 60 | 45 | |
| ~ | a=1 | 0.0 | 100 | 0.5 | 100 | |
| 613 | b=222 | 90 | 100 | 95 | 100 | |
| ija | h = 200 | 100 | 100 | 100 | 100 | |
| ζne | 0-300 a=7 | 100 | 100 | 100 | 100 | |
| <u> </u> | h=330 | 95 | 95 | 100 | 100 | |
| | a=6 | ,,, | ,,, | 100 | 100 | |
| | b=360 | 100 | 100 | 100 | 85 | |
| | a=76 | | | | | |
| Kneja 620 | b=340 | 95 | 95 | 100 | 90 | |
| | a=75 | | | | | |
| 26A 1 | b=47 | 100 | 55 | 100 | 100 | |
| | a=45 | | | | | |
| | 1 00- | 100 | 0.0 | 100 | 0- | |
| | b=307 | 100 | 80 | 100 | 85 | |
| | a=29 | 100 | 05 | 100 | 100 | |
| Ruse 424 | D = 66 | 100 | 85 | 100 | 100 | |
| | a-1/ | 05 | 00 | 100 | 05 | |
| | 0-297 a=5 | 73 | 90 | 100 | 93 | |
| | h=332 | 100 | 100 | 100 | 95 | |
| | a=77 | 100 | 100 | 100 |)5 | |
| | b=597 | 100 | 80 | 100 | 85 | |
| | a=25 | 100 | | 100 | | |
| 36 | b=84 | 95 | 75 | 100 | 100 | |
| | a=14 | - | - | | | |
| 2//1 | b=182 | 100 | 80 | 100 | 100 | |
| 18, | a=148 | | | | | |
| ХN | b=246 | 100 | 90 | 95 | 100 | |
| | a=31 | | | | | |

The results show that the linear classifier do not recognized the corn kernels from class Fusarium diseased with not so high classification accuracy 93,3%. The PNN classifier recognized all corn kernels (healthy and Fusarium diseased) better than the linear.

4 Conclusion

The results show that the proposed method is effective for classification of healthy and *Fusarium* infected corn kernels. The proposed approach based on wavelet continuous transformation and probabilistic neural network (PNN) classification classifies the corn kernels with classification accuracy in the range from 75 to 100.

The wavelet parameters "*scale - a*", "*time - b*" and coefficient $C_{a,b}$ influence of the corn variety. The probabilistic network smoothing factor σ =5 is the optimal value for all corn variety.

Determined wavelet parameters "*scale - a*" and "*time - b*" and probabilistic network smoothing factor σ cover requirements of the Bulgarian quality standards.

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.] BDS 529-84. Cereal grains, 1985.
- [2.] BDS 14852-79. Methods for determining the damage from pests and diseases. 1980.
- [3.] Bose, N.K., Liang P. (1998). Neural network fundamentals with graphs, algorithms and applications. Tata McGraw Hill.
- [4.] Cassells J.A., Reuss R., Osborne B.G., Wesley I.J., (2007). Near infrared spectroscopic studies of changes in stored grain. J. Near Infrared Spectrosc., 15, 161 – 167.
- [5.] Daskalov P., Ts. Draganova, V. Mancheva, R. Tsonev, Assessment of corn variety influence to Fusarium disease diagnostic using spectral analysis and statistical approach. *Proceedings of University of Rousse*, Vol 49, 3.1, 2010, pp.119-124.
- [6.] Delwiche S. R., Gaines C. S., (2005). Wavelength selection for monochromatic and bichromatic sorting of fusarium-damaged wheat. *Applied Engineering in Agriculture*, Vol. 21(4), 681–688
- [7.] Koger, C., Bruce, L., Shaw, D., and Reddy, K. (2003). Wavelet analysis of hyperspectral reflectance data for detecting pitted morning glory (Ipomoealacunosa) in soybean (Glycinemax). *Remote Sensing of Environment*, 86,108–119.
- [8.] Mahesh, S., Manickavasagan A., Jayas D.S., Paliwal J., White N.D.G., (2008). Feasibility of near-infrared hypespectral imaging to differentiate Canadian wheat classes. *Biosystem Engineering*, 101, 50-57.
- [9.] Munkvold, G.P., Desjardins A.E.. (1997). Fumonisins in maize. Can we reduce their occurrence? *Plant disease*, 81, 556 – 564.
- [10.] Specht, D. F., (1996). Probabilistic neural networks and general regression neural networks. Fuzzy Logic and Neural network Handbook, ed. C.H. Chen, ch. 3. New York: McGraw-Hill, Inc.
- [11.] Vetterli M., (2001). Wavelets, Approximation, and Compression, IEEE Signal Processing Magazine, 59-73
- [12.] Xu,G., Hon-fu Y., and Wa-zhen L., (2000). Development of modern near infrared spectroscopic techniques and its applications. *Spectroscopy and Spectral Analysis*, 20, 134–142.