The papers published in this CD are of full responsability of their authors

ISSN 1982-3797

International Conference of Agricultural Engineering
XXXVII Brazilian Congress of Agricultural Engineering
International Livestock Environment Symposium - ILES VIII

August 31st to September 4th, 2008
Iguassu Falls City - Brazil

INTERNATIONAL CONFERENCE OF AGRICULTURAL ENGINEERING
CIGR 2008 / BRAZIL

CENTRAL THEME
Technology for All:
Sharing the Knowledge for Development

Hosts

SBEA ASABE TECHNALLER

Sponsors

CNPq CONFEA CREA A SERVIÇO DO BRASIL MASSEY FERGUSON BANCO DO BRASIL ÁGUA BOA ITAIPU BINACIONAL

JOHN DEERE CAPES PG unioeste FUNDAÇÃO ARAUCÁRIA British Embassy Brasilia SCIENCE AND INNOVATION

Support

Funep UNIVERSIDADE ESTADUAL DE MARINGÁ unesp Jaboticabal

Proceedings

DIGITAL IMAGE PROCESSING FOR IDENTIFICATION OF BLACK SIGATOKA

SILVIA HELENA MODENESE GORLA DA SILVA[1]; WÍLSON DA SILVA MORAES[2]; HUGO DO NASCIMENTO BENDINI[3]; CARLOS ROBERTO PADOVANI[4]; LUCIO ANDRÉ DE CASTRO JORGE[5]; KELLY BOTIGELI SEVEGNANI[6]

[1] Researcher, Msc, Campus Experimental de Registro/UNESP/Registro – Brazil. e-mail: silvia@registro.unesp.br
[2] Researcher, PhD, Campus Experimental de Registro/UNESP/Registro and Apta/Pariquera-Açu – Brazil
[3] Undergraduate student, Campus Experimental de Registro/UNESP/Registro – Brazil
[4] Researcher, PhD, Instituto de Biociências/Bioestatística/UNESP/Botucatu – Brazil
[5] Researcher, Msc, CNPDIA/EMBRAPA/São Carlos - Brazil
[6] Researcher, PhD, Campus Experimental de Registro/UNESP/Registro - Brazil

Presented at
CIGR INTERNATIONAL CONFERENCE OF AGRICULTURAL ENGINEERING
XXXVII CONGRESSO BRASILEIRO DE ENGENHARIA AGRÍCOLA – CONBEA 2008
Brazil, August 31 to September 4, 2008

ABSTRACT: This study investigated the application, specifically the digital processing of images, with main component analysis and artificial neural networks as tools to support for better identification of the early stages of the development of the Black Sigatoka, so that control measures are adopted more quickly and thus reduce injuries and damage caused by the disease in banana plantation. 10 images were collected from digital leaves infected by the Black Sigatoka in stages 1, 2 and 3, Yellow Sigatoka, healthy and with fitotoxicity by oil. Then, it was extracted histograms of the components of the system RGB images to 256 shades of gray, from the 60 examples. This made it necessary to apply a technique for selection of attributes, the principal component analysis. Thus, it was able to reduce the 768 input varieties of each example to 22, which are linear combinations of the varieties of unique entries. Finally, there was the training of artificial neural networks for recognition of each of the mentioned classes.
KEYWORDS: Black Sigatoka, digital image processing, artificial neural networks, principal components analysis.

INTRODUCTION: The Black Sigatoka, caused by the fungus *Mycosphaerella fijiensis* Morelet, is the most severe and destructive disease of banana producers in worldwide areas, being responsible for production losses exceeding 50%. The disease occurs on the leaves causing brown streaks and spots that reduce black necrotic photosynthesis tissue and, consequently, the gross income. The first record of the disease in Brazilian territory was in the state of Amazon in 1998 and in 2004, the disease was detected in Vale do Ribeira banana plantation. Currently, it is disseminated in all producing regions of the state of São Paulo (FERRARI, 2005). The Figure 1 illustrates the ocorrency of the Black Sigatoka.
The control of the disease is made by spraying alternate of systemic fungicides and protectors. Meanwhile, the weekly monitoring is a laborious task that requires a minimum structure and involves progressive spending to shift and train technicians able to identify the stages of development of the disease. The records of weekly stages of the disease and the issue of leaves are transferred to spreadsheets that calculate the state of current developments what indicates the severity and the need or not the application of fungicides, whose decision is assisted by the local weather data. This method can lead to errors of observation, because we have to distinguish between lesions healed (on effects of fungicides) of the lesions alive, injuries from Black and Yellow Sigatoka , viruses and fitotoxicity of mineral oil, applied to the syrup fungicide because tasks that require experience and subjectivity of human beings are highly susceptible to errors.
Other inaccuracies of control can be attributed to geomorphologic features, vegetation, roads, difficult access to banana plantations and small properties, which makes it virtually impossible to achieve a precise diagnosis using traditional methods.

In the globalized world, intensive agriculture depends increasingly and irreversibly on the use of inputs and modern technologies so that the results are increasingly enhanced with the technological advances.

An interesting and attractive alternative for dealing with this type of problem involving agriculture is the processing of digital images associated with the analysis of multivariate data. In this sense, it was intended in the current study to use the application of precision farming; more specifically the digital processing of images with principal components analysis (PCA), a technique of multivariate statistical data (JOHNSON & WICHERN, 2002), and Artificial neural networks (ANN) as tools to support for better identification of the early stages of the development of the Black Sigatoka, so that the control measures could be adopted more quickly and thus reduce injuries and damage caused by the disease in banana plantation . The PCA aims to replace a number of varieties correlated by a set of new varieties not correlated with each other, which are linear combinations of initial varieties and presented in descending order of magnitude of their variances (HOTELLING, 1936; JOHNSON & WICHERN, 2002). As the RNA bases on a form of algorithmic computation not inspired in human brain (BRAGA et al., 2000).

FIGURE 1. Ocorrency of the Black Sigatoka.

METHODOLOGY: to carry out the study, the methodology consisted up of four distinct phases: collection and scanning of samples, targeting and extraction of histograms of images, selection of attributes and classification.

First, 10 samples were collected from 2x2 cm sheets for each of the classes: Black Sigatoka in stages 1, 2 and 3, Yellow Sigatoka and fitotoxicity by oil. Those samples were brought to the laboratory of the Campus Center in Registro, and then scanned on the scanner table Brand HP to a resolution of 140 x 140 pixels.

Second, it was studied the frequency and intensity of the color of the pixels. For that, it was built up a tool computer, using software HALCON, which decomposes each image in the three color components of the system RGB (GONALEZ & WOODS, 1992). Each of these components of images, estimated to be a histogram of the intensity levels of gray (256 levels). Thus, each picture had a 769 varieties to represent it, or the frequency of each one of the levels of gray for each of the RGB components of color, represented the 768 input varieties and also a variable output representing the class of each image.

With the excessive number of input varieties (768), it was necessary, in the third step of the study, the application of a technique for the selection of attributes to reduce the number of input varieties and thus facilitate computingly the training of Artificial Neural Networks in the process of classification of images. For that, it was applied the technique of main component analysis, which is a technique of multivariate analysis of data for which the tool selected for induction of algorithms was the Weka (WITTEN & Frank, 1999). The Weka is a tool for Machine Learning (MITCHELL, 1997) developed at the University of Waikato, New Zealand, written in Java language, with open source, which allows changes in functions to be used, and the insertion of implementations not supported. The Weka is a tool available on the public domain address: http://www.cs.waikato.ac.nz./ml/weka.

Finally, it came to pass the interpretation of the image. The stage of interpretation is the most "intelligent" digital processing of the image, as it allows to get the understanding and final description of the phenomenon. It is in step interpretation of the image that was used the technique of ARN for classification of these patterns collected. The neural network used was a neural network sandwich, trained with back propagation algorithm (RUMELHART, 1986).

The results verified in the experiments conducted in this study were to be analyzed according measures and statistical tests widely disseminated. In the case of classification of tasks, a measure commonly used is the error rate of the classifier h, also known as rate incorrect classification The error rate is obtained using Equation 1, which compares the real class obtained from each example with the label assigned by the classifier induced. If the expression is true, yields are 1 and otherwise 0, where n is the number of examples. The increase in the rate of error is the accuracy of the classifier (BARANAUSKAS, 2000).

$$e(h) = \frac{1}{n}\sum_{i=1}^{n}(y_i \neq f(x_i) \qquad (1)$$

To better estimate of the error of the classifier it was used samples techniques for manipulation of sets of training and testing. This work is the methodology used r-fold cross-validation (EFRON & TIBSHIRANI, 1993). The methodology *r-fold cross validation* partitionates data sets r total in equal parts, using r-1 folds to train and to test the classifier. They are made, therefore, r different training and testing. The errors are evaluated in each partition and it is estimated, for the purpose of comparison of algorithms, the average and standard deviation of the errors of all partitions.

RESULTS AND DISCUSSION: According to the methodology developed, first, ten samples of leaves were collected up for each of the patterns studied: stages 1, 2 and 3 of the Black Sigatoka, sound, and Yellow Sigatoka fitotoxicity by mineral oil. Then, the samples were scanned and then in phase two was held on targeting and testing of frequencies from RGB images broken in the system. At this stage, generated by a text file containing the 768 varieties, representing the frequencies of intensity of 256 levels of gray for each of the three images broken, of the 60 standards. At that time, it was made a first test of classification with the ARN, the tool Weka, and it was found that could not have a result, due to the large amount of input varieties. Thus, it was necessary to take the selection of attributes to apply to the main component analysis. With that, it was generated a file containing only 22 variables, which are linear combinations of the original 768 variables. Thus, the file of entry to the tage of classification was composed by 22 varieties for entry and exit of a variety representing the class of a particular standard. In this fourth step, therefore, the training of ARN Multilayer Perceptron was made, trained with the backpropagation algorithm, ending momentum. The rate of learning and adopted term momentum, which are parameters of the algorithm of learning adopted chosen from empirically, were respectively equal to 0.2 and 0.3. The Figure 2 represents the architecture of ARN used in the classification.

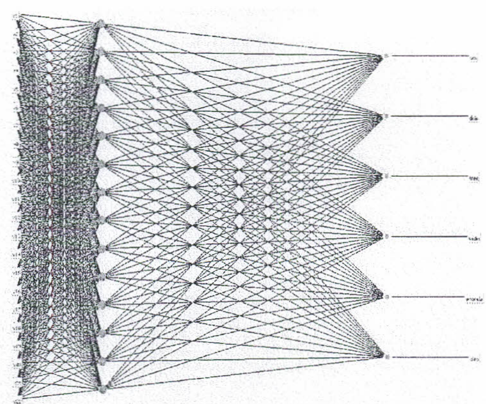

FIGURE 2. Artificial Neural Network Multilayer used in the training of patterns.

As a result of the classification ARN has a hit rate of the average sets of tests of 51.67%. The Figure 3 illustrates the confusion matrix of test folds as a result of classification of ARN.

a	b	c	d	e	f	classified as
5	1	0	0	3	1	I a = um
3	3	0	1	0	3	I b = dois
1	0	7	0	1	1	I c = tres
1	0	0	8	0	1	I d = sadia
4	0	2	0	3	1	I e = amarela
2	1	0	0	2	5	I f = oleo

FIGURA 2. Confusion matrix of test folds **.**

Analyzing FIGURE 3, it appears that the more classes were incorrectly classified as class two (b), (e) Yellow Sigatoka and with fitotoxicity by mineral oil (f). This problem may have occurred, perhaps, because of the samples does not represent well the designated classes. Thus, it's necessary to carry out other tests with a larger number of samples for each class. Also, they should be investigated, as future work, other classifiers, and other methods of digital processing of images.

CONCLUSION: With the use of digital image processing and interpretation techniques for imaging studies could be established patterns of the first studies to identify the stages of Black and Yellow Sigatoka and the fitotoxicity by oil. This study served as the first work in the identification of other possible future work that serve as tools for the rational use of energy, optimizing the application of fungicides. At the same time, you can find geographically the exclusive presence of Black or Yellow Sigatoka and map regions or micro basins, important for the deployment of a system Forecast bio-climatic of the disease.

REFERENCES:

BARANAUSKAS, J.A. **Extração automática de conhecimento por múltiplos indutores**. Tese de Doutorado**.** Instituto de Ciências Matemáticas e de Computação, USP, 2000.

BRAGA, A. P.; CARVALHO, A. C. P. F.; LUDERMIR, E T. B. **Redes Neurais Artificiais: Teoria e Aplicações;** Rio de Janeiro: Livro Técnico e Científico, 2000.

EFRON, B.; TIBSHIRANI, R. J. **An Introduction to the Bootstrap**. New York: Chapman & Hall. 436p, 1993.

FERRARI, J. T.; NOGUEIRA, E. M. C; GASPAROTTO, L.; HANADA, R. E.; LOUZEIRO, L. M. **Ocorrência da Sigatoka-negra em bananais no Estado de São Paulo.** Arquivos do Instituto Biológico**,** v. 72: 133-134, 2005.

GONZALEZ, R. C. e WOODS, R. E. **Digital Image Processing.** Reading, Addison Wesley, 716p.,1992.

HOTELLING, H. Simplified calculation of principal components. **Psychometrika**, v.1: 27-35, 1936.

JOHNSON, R.A.; WICHERN, D.W. **Applied Multivariate Statistical Analysis,** 5ª ed, New Jersey: Prentice-Hall, 767p., 2002

Mitchell, T. M. **Machine learning**. Boston: McGraw Hill Companies Inc. 414p., 1997.

RUMELHART, D. E.; HINTON, G. E., WILLIAMS, R. J. Learning representations by back-propagation errors. **Nature**, v. 323:533-536, 1986.

WITTEN, I.H.; FRANK, E. Data Mining: Practical Machine Learning Tools and Techniques with Java Implementations. Morgan Kaufmann, 1999.