

Identification of Foliar Diseases in Cotton Crop

A. A. Bernardes, J. G. Rogeri, N. Marranghello and A. S. Pereira.

Department of Computer Science and Statistics, IBILCE, Sao Paulo State University (UNESP), Sao Jose do Rio Preto – SP, Brazil.

A.F. Araujo and João Manuel R. S. Tavares

Department of Mechanical Engineering (DeMec), Faculty of Engineering, University of Porto (FEUP) Institute of Mechanical Engineering and Industrial Management (INEGI), Porto, Portugal.

ABSTRACT: The pathogens manifestation in plantations are the largest cause of damage in several cultivars, which may cause increase of prices and loss of crop quality. This paper presents a method for automatic classification of cotton diseases through feature extraction of leaf symptoms from digital images. Wavelet transform energy has been used for feature extraction while Support Vector Machine has been used for classification. Five situations have been diagnosed, namely: Healthy crop, Ramularia disease, Bacterial Blight, Ascochyta Blight, and unspecified disease.

1 Introduction

This work aims at identifying foliar diseases in cotton plantations. This cultivar is of great economic importance to Brazil, being one of the pillars of the textile industry, which consumes around one million tons of cotton fiber per year [1, 2].

The primary goal of the developed system has been to identify the existence of pathogens in cotton foliar. Once a disease is identified it has to be automatically classified through further processing of the corresponding image. The pathogens used in our tests are among those most frequently occurring in Brazil, which are very aggressive, rapidly spreading throughout the plantations, and that can be fought against only by the use of chemicals [1, 2, 3, 4]. Three diseases were used during the classification stage, as follows: Ramularia (RA), Bacterial Blight (MA), and Ascochyta Blight (AS). Alternatively we allowed the system to classify an image as not belonging to any of these classes (NONE).

In this work we used both RGB and HSV standards, as well as I3a and I3b channels, and grey level (GL) images. The color channels I3a and I3b are obtained from a transformation of the I1I2I3 color standard [5]. The energy of the wavelet transform was computed for each sub-band to build feature vectors that were utilized during the training phase of the support vector machine (SVM) used for image classification.

1.1 Generation of I3a and I3b Channels

Channels and I3a and I3b were created from changes carried out by Campbell [5] on the color channel of I1I2I3 model defined by Ohta *et al.* [6].

1.2 Wavelets and Image Processing

Currently, the application of wavelet transform has proven very efficient to perform complex analysis of non-stationary signals, with respect to time-frequency signal. There is a considerable increase in its application in a variety of biosignals, such as in neurophysiological signal analysis, medical signals and image diagnosis, pulmonary microvascular pressure estimation, mammograms, signal compression, brain signals analysis, and classification of voice disorders [7].

1.3 Support Vector Machine

Support Vector Machine (SVM) is a classification technique that is becoming increasingly used in science, when high performance computation is required for practical applications. The outcome of such a technique is often superior to results obtained by other machine learning algorithms such as Artificial Neural Networks [8, 9, 10]. SVMs have been used for pattern recognition, image processing, machine learning, bioinformatics, among others [11].

2 Material

A total of 420 images were used in this research. One set of images was provided by phytopathologist Dr. Nelson Dias Suasuna, researcher at Embrapa Cotton, in Campina Grande, Paraíba, Brazil [12]. This database was complemented by another set of images from the Forestry Images website [13].

Size, brightness, contrast and resolution of the images are quite different, resulting in a very heterogeneous database. To illustrate the database items some images of healthy regions of cotton leaves are shown in Figure 1. Opposite to this some images

of infected leaves with varying degrees of disease severity are shown in Figure 2. Figure 2(a) displays the leaf area at an early stage of MA disease; Figure 2(b) presents the infected leaf, at an intermediate stage; and Figure 2(c) shows the advanced stage.



Figure 1: Images of healthy leaf area of cotton [12, 13].

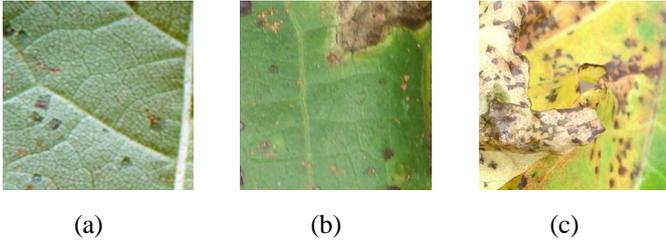


Figure 2: Severity degree of MA cotton disease: (a) initial; (b) intermediate; and (c), advanced stages of infection [12].

3 Proposed Method

The classification process was divided into two phases:

- **Phase 1:** Finding the best feature vector for each class;
- **Phase 2:** Create the final classification system from the best results obtained in the previous phase.

3.1 Phase 1: Finding the Best Feature Vector

This phase is aimed at finding the best feature vector to represent each of the classes to be considered during classification. To achieve this goal the following steps were accomplished:

- Decomposition of images into multiple channels (R, G, B, H, S, V, I3a, I3b, and GL);
- Application of the discrete wavelet transform (DWT) up to the third level;
- Computation of the energy for each sub-band and compose the feature vector;
- Creation of the SVM classification environment;
- Listing of the images used for training and testing;
- Evaluation of the best feature vectors.

3.1.1 Decomposition of the Image

The decomposition of the images is the first process the system executes. In this stage an image is decomposed into nine channels, namely: R, G, B, H, S, V, I3a, I3b and GL.

3.1.2 Application of the Discrete Wavelet Transform

Discrete Wavelet Transform (DWT) decomposition is applied up to the third level. When an image is decomposed as such it will have ten sub-bands, as illustrated in Figure 3. Note that each sub-band is identified by a number between 1 and 10. Region A1 and sub-bands 8, 9 and 10, are generated by the first level of decomposition of the DWT. Region A2 and sub-bands

5, 6 and 7 refer to the second level of decomposition, and the third level is formed by the sub-bands 1, 2, 3 and 4.

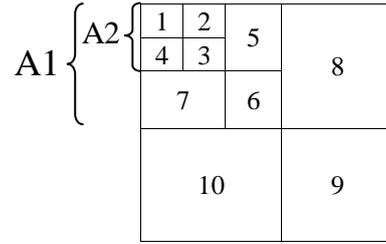


Figure 3: Schematic of the wavelet decomposition.

3.1.3 Computation of the Energy for Each Sub-band

After applying the DWT to the three levels, the energy for each wavelet sub-band is computed. Each value obtained is inserted into a feature vector as the one illustrated in Figure 4. The vector in this figure consists of ten elements, each of them is identified by a number corresponding to the number of the sub-band in Figure 3. The energy value computed for each sub-band is stored in the corresponding vector element.

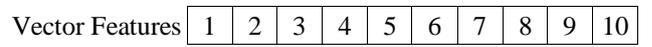


Figure 4: Example of the structure of a feature vector.

3.1.4 Creation of an SVM Classification Environment

The network architecture used is shown in Figure 5. Note that 10 input elements are used. To each input element is assigned the value of the element of the corresponding characteristic vector. In the hidden layer there are a number of neurons (N) equal to the number of training examples, making the net convergence easier [8, 9]. The Gaussian kernel has been used as the network mapping function (kernel).

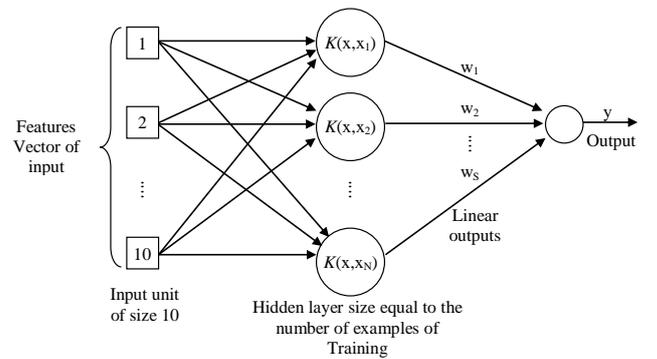


Figure 5: Architecture of SVM used

3.1.5 Listing of the Images Used for Training and Testing

The classification process was divided into two different steps:

- **Step 1:** Label the leaf images as healthy (SA) or injured (LE).

- **Step 2:** Find the best descriptor (feature vector) representing each class. It is performed only among the three types of pathogens (RA, MA and AS), which an image can be classified.

To find the best descriptor for each class, 12 wavelets were used. Table 1 displays the name of the wavelet coefficients, followed by the number of its support, and the abbreviations used in this work, which associates the name of a coefficient to a support number. The 12 wavelets were identified as: Bey18,

Coi12, Coi30, Dau4, Dau14, Dau34, Dau64, Dau76, Haar, Sym8 and Vai24.

Table 1: Coefficients and support numbers used

Coefficient Name	Number of Support	Abbreviation used
Beylkin	18	Bey18
Coiflets	12 e 30	Coi12 e Coi30
Daubechies	4, 14, 34, 64 e 74	Dau4, Dau14, Dau34, Dau64 e Dau74
Haar	1	Haar
Symmlets	8 e 16	Sym8 e Sym16
Vaidyanathan	24	Vai24

3.1.6 Evaluation of the Best Feature Vectors

First, 108 feature vectors are obtained. Then they are classified either as LE or as SA. After that the percentage of correct answers of the 108 feature vectors for the classes RA, MA and AS is computed. The best feature vectors for each class are listed in Table 2. Among classes SA and LE the best feature vector that separated the two classes was in the channel H, by using the wavelet Vai24, which for SA achieved 96.2% correct guesses, and for LE it reached 100% accuracy. For class MA the best result was in I3b channel coefficients, on Coi12, Sym16, which reached 97.1% accuracy. Class RA achieved 88.6% accuracy using the feature vector of channel H and wavelet Dau4. It can also be seen in Table 2 that the AS class presents the best result, achieving 88.6% accuracy in channel H, using coefficient wavelet Bey18.

Table 2: Best results achieved between the classes

Class Name	Channel	Coefficient	Percentage of Correct
SA	H	Vai24	96,2%
LE	H	Vai24	100%
MA	I3b	Coi12 e Sym16	97,1%
RA	H	Dau4	88,6%
AS	H	Bey18	88,6%

3.2 Creating the Final Classification System

In the previous section the best descriptors (feature vectors) for each class were identified. This section describes the methodology to compose the final classification system.

Figure 6 illustrates the general classification structure. Observe that four different classification systems were created. Each system was trained and tested with the best feature vectors, as discussed in the previous section.

4 Results and Discussion

Among 210 images used for the tests, 188 images were correctly classified (101 SA images, 34 MA images, 28 RA images, and 25 AS images) corresponding to slightly over 89.5% correct guesses.

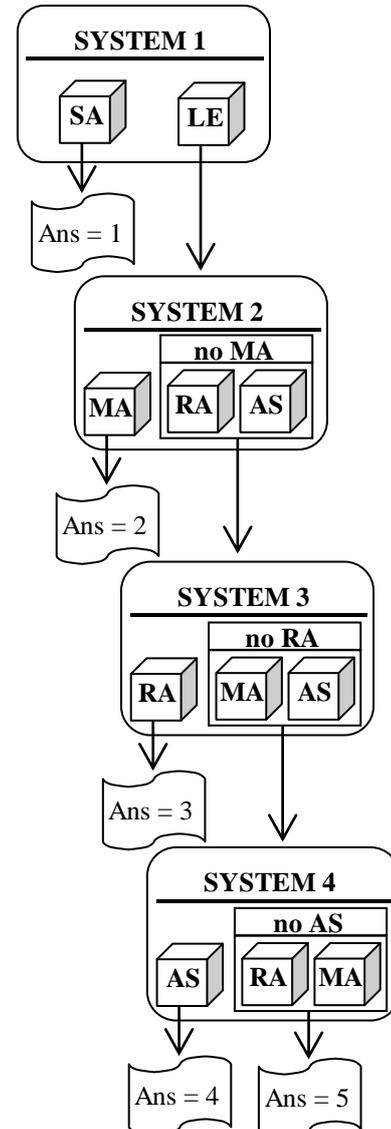


Figure 6: Final Classification System

For the healthy class (SA), which comprises 105 images, 101 were correctly classified, corresponding to about 96.2% of correct guesses. Table 3 displays the percentage of healthy image classification for each diagnosis. Note that 3.8% are false positives, half of which were classified as belonging to class Bacterial Blight, and another half as Ramularia Disease.

Table 3: Test results for 105 “Healthy” images

SA	False Positive			
	MA	RA	AS	NONE
96,2%	1,9%	1,9%	0,0%	0,0%

For Bacterial Blight class, 34 of the 35 images were correctly classified. Table 4 displays some details for this classification, where it can be seen that 97.1% of images were correctly classified, and only 2.9% were false positives.

Table 4: Test results for 35 “Bacterial Blight” images

MA	False Positive			
	SA	RA	AS	NONE
97,1%	0,0%	0,0%	0,0%	2,9%

For Ramularia class 28 of the 35 images were correctly classified. Table 5 shows details of “Ramularia” classification for each diagnosis. Note that 80% of the images were correctly classified, and 20% were false positives, being 8.6% classified as MA, 11.4% classified as AS, and no images were classified as healthy or as presenting no known disease.

Table 5: Test results for 35 “Ramularia” images

RA	False Positive			
	SA	MA	AS	NONE
80,0%	0,0%	8,6%	11,4%	0,0%

For the Ascochyta Blight class 25 of the 35 images were correctly classified. Table 6 displays the details of image classification for the “Ascochyta Blight” set. Note that 71.4% of images were correctly classified, and 29.6% were false positives, being 11.4% classified as MA, 14.3% classified as RA, and 2.9% classified as not belonging to any known class.

Table 6: Test results for 35 “Ascochyta Blight” images

AS	False Positive			
	SA	MA	RA	NONE
71,4%	0,0%	11,4%	14,3%	2,9%

Note from the healthy (SA) class of tables 3 through 6 that no images of injured leaves was classified as healthy. Thus, it follows that all 105 infected images, namely with pathogens Bacterial Blight, Ascochyta Blight and Ramularia, were classified as injured, achieving 100% accuracy of image separation, and consequently resulting in no false positive.

5 Comparison to Related Works

The present work has taken cotton culture as its subject. As far as we could determine no other research took such a cultivar as subject in the same sense we took.

Huang [14] presented a study on rubber trees with four different diagnosis (healthy, and three diseases) achieving a score of 97.2% correct guesses for injured leaves. A problem with Huang's work is the fact that only injured leaves were used.

Abduhlah *et al.* [15] classified three pathogens for rubber trees achieving 80% of correct guesses. However, Abdulah and co-workers' tests considered only one kind of pathogens.

Meunkaewjinda *et al.* [16] examined three diagnoses (healthy, and two diseases) for grape cultivars. In their study they achieved 86.3% correct guesses.

Phadicar and Sil [17] studied rice crops achieving 92% of correct diseases classification. However, only two classes were used for this study.

6 Conclusions

216 feature vectors were generated for our study, half of which used to identify the best feature vectors within SA and LE classes, and the other half used to find the best among MA, RA, and AS classes.

The best feature vectors were used during the final system classification. After sorting the healthy (SA) images out, the image set of supposedly infected leaves was classified within one of the four other sub-classes, namely: MA, RA, AS, and NONE. The final results achieved were: 96.2% accuracy for the SA class, 97.1% accuracy for the MA class, 80% accuracy for the RA class, and 71.4% accuracy for the AS class.

Moreover, the present work achieved better results with respect to other cultivars. Considering it used a small amount of samples to train the SVM, it produced descriptors adequately representing each class, even using a heterogeneous database, and it produced appropriate classification of a fairly large number of classes with almost 90% accuracy.

7 Acknowledgments

The authors are thankful to FUNDUNESP – Fundação para o Desenvolvimento da UNESP – Brazil and FAPESP – Fundação de Amparo à Pesquisa do Estado de São Paulo – Brazil for the financial support.

8 References

- [1] OLIVEIRA, M. C. Agricultural Protection Agency of Goiás (Agrodefesa). Program of Prevention and Control of Pests in Cotton. Available at: http://www.agrodefesa.go.gov.br/index.php?option=com_content&view=article&id=74. Accessed on January, 2010.
- [2] GALBIERI, R. Behavior of Cotton Genotypes in the Presence of Pathogens and Nematodes. Master Dissertation (Tropical and Subtropical Agriculture), Campinas, SP – Brazil, 2007. (in Portuguese)
- [3] KIMATI, H.; AMORIM, L.; BERGAMIN FILHO, A.; CAMARGO L. E. A.; and REZENDE, J. A. M. Handbook of Phytopathology. Vol. 2: Diseases of Cultivated Plants. Sao Paulo: Agronomic Ceres Ltda, 1997.
- [4] CURVELO, C. R. S.; RODRIGUES, F. A.; BERGER, P. G.; and REZENDE, D. C. Scanning electron microscopy of the infectious process of Ramularia aureola on cotton leaves. Trop. Plant Pathol. [online]. v.35, n.2, pp.108-113, 2010. ISSN: 1982-5676.
- [5] CAMARGO A.; SMITH J.S. An image-processing based algorithm to automatically identify plant disease visual symptoms. Biosystems Engineering, v.102, n.1, pp.9-21, 2009.
- [6] OHTA, Y.; KANADE, T.; and SAKAI, T. Color information for region segmentation. Computer Graphics and Image Processing. Department of Information Science, Kyoto, Japan, v.13, n.3, pp.222–241, July, 1980.
- [7] ADDISON, P. S.; WALKER, J.; and GUIDO, R. C. Time-frequency analysis of biosignals, Engineering in Medicine and Biology Magazine, IEEE, v. 28, n. 5, pp.14-29, September – October, 2009.
- [8] SOUZA, L. M. Intelligent Detection in Laryngeal Pathologies Based on Support Vector Machines and Wavelet Transform. Master Dissertation (Applied Computational

Physics) – University of Sao Paulo at Sao Carlos, SP, Brazil, 2011. (in Portuguese)

- [9] BISOGNIN, G. Using Support Machine to Prediction of Protein Tertiary Structure. Master Dissertation (Applied Computing) – University of Vale do Rio dos Sinos, Sao Leopoldo, RS - Brazil, 2007. (in Portuguese)
- [10] FONSECA, E.; GUIDO, R. C.; SCALASSARA, P. R.; MACIEL, C. D.; and PEREIRA, J. C. *Wavelet* Time-frequency Analysis and Least-Squares Support Vector Machine for the Identification of Voice Disorders. *Computers in Biology and Medicine*, Elsevier, v.37, n.4, pp.571-578, 2007.
- [11] YU, Z.; WONG, H.; and WEN, G. A Modified Support Vector Machine and its Application to Image Segmentation. *Image and Vision Computing*, v.29, pp.29-40, 2011.
- [12] SUASSUNA, N. D. Private Communication. Brazilian Company of Agricultural Research, Campina Grande, PB - Brazil.
- [13] FI. Forestry Images. A joint project of the Center for Invasive Species and Ecosystem Health, USDA Forest Service and International Society of Arboriculture. The University of Georgia - Warnell School of Forestry and Natural Resources and College of Agricultural and Environmental Sciences. Available at: <http://www.forestryimages.org>. Accessed on August, 2010.
- [14] HUANG, K. Application of artificial neural network for detecting Phalaenopsis seedling diseases using color and texture features. *Comput. Electron. Agric.* v.57, n.1, pp.3-11, May, 2007.
- [15] ABDULLAH, N. E.; RAHIM, A. A.; HASHIM, H.; and KAMAL, M. M. Classification of Rubber Tree Leaf Diseases Using Multilayer Perceptron Neural Network, Research and Development. SCORED 5th Student Conference. pp.1 – 6, 11-12 December, 2007.
- [16] MEUNKAJINDA, A.; KUMSAWAT, P.; ATTAKITMONGCOL, K.; and SRIKAEW, A. Grape leaf disease detection from color imagery using hybrid intelligent system. *Electrical Engineering/Electronics, Computer, Telecommunications and Information Technology*, 2008. ECTI-CON 2008. 5th International Conference on v.1, pp.513–516, 14-17 May, 2008.
- [17] PHADIKAR, S.; and SIL, J. Rice Disease Identification using Pattern Recognition Techniques; Proceedings of the 11th International Conference on Computer and Information Technology (ICCIT 2008), Khulna, Bangladesh. pp.420-423, 25-27 December, 2008.