



VI Workshop de Visão Computacional  
04-07/Julho/2010 - Presidente Prudente, SP - Brasil  
ISSN 175-6120

UNESP  
UNIVERSIDADE ESTADUAL "JULIO DE MESQUITA FILHO"  
ENGFOTO

opto

SBMAC

FAPESP

CNPq

orga

capit

## Extração de Características para Geração de um Classificador para Detecção Precoce do HLB em Citros

Diego Carlos Pereira da Silva<sup>1</sup>; Patrícia Pedroso Estevam Ribeiro<sup>1</sup>; Lúcio André de Castro Jorge<sup>2</sup>; Maria Stela Veludo de Paiva<sup>3</sup>; Débora Marcondes Milori<sup>2</sup>; Danilo Scavacini Gonçalves<sup>4</sup>, Camila Miranda Carvalho<sup>4</sup>; André Leonardo Venânico<sup>4</sup>; Fabíola Manhas Verbi Pereira<sup>4</sup>

- 1- Mestrando, Escola de Engenharia de São Carlos - Universidade de São Paulo - USP, Depto. de Engenharia Elétrica, São Carlos-SP, [dicapesilva@gmail.com](mailto:dicapesilva@gmail.com).
- 2- Pesquisador, Embrapa Instrumentação Agropecuária, São Carlos - SP, [lucio@cnpdia.embrapa.br](mailto:lucio@cnpdia.embrapa.br).
- 3- Professora, Escola de Engenharia de São Carlos - Universidade de São Paulo - USP, Depto. de Engenharia Elétrica, São Carlos-SP, [mstela@sc.usp.br](mailto:mstela@sc.usp.br).
- 4- Bolsista, Embrapa Instrumentação Agropecuária, São Carlos - SP.

### Abstract

*Greening or Huanglongbing (HLB) is one of the most serious threats for citrus production worldwide. São Paulo, Brazil is the most important citrus producers and, is making efforts for citrus diseases control. The bacterial pathogen is, mainly, under suppression control, applied by eradication of symptomatic and no-symptomatic plants. In this way, the detection and diagnostic of the related symptoms and, consequently, the eradication of the citrus trees are essential for higher economical losses prevention. In this way, our goal is to develop a new optical and data mining technique, applied in field conditions, to detect citrus diseases using a fluorescence imaging system. It was used a system that perform image of the chlorophyll fluorescence in the whole leaf. It was collected images of 3 leaves from 120 different greenhouse rootstocks. It has obtained fluorescence images and spectral data of citrus healthy leaves and contaminated leaves with HLB. In this paper it was described a new method to extract attributes by data fusion, carrying out different analysis to try to discriminate between safe and diseased leafs. The results show that the method is very accurate for HLB ultimately detection.*

### 1 Introdução

O Greening, também chamado de Huanglongbing (HLB) (BOVÉ, 2006), é uma das mais graves doenças dos citros presentes nos pomares do Brasil desde 2004 (FUNDECITRUS, 2009). Causada pela bactéria *Candidatus Liberibacter SSP* é transmitida pelo inseto *psílideo Diaphorina citri*, que ao se alimentar de plantas doentes transmite a mesma às demais plantas. A única forma de prevenção

consiste na erradicação pela raiz das plantas contaminadas, associada ao controle dos insetos transmissores. Por este motivo, a obtenção de um diagnóstico precoce é de suma importância, para evitar a propagação.

Dentre os métodos existentes atualmente para diagnosticar a doença, pode-se citar: o método de inspeção visual e a análise do PCR (Polymerase Chain Reaction) (INNIS et al, 1990). A inspeção visual só permite atuar depois de muitos meses e, enquanto isto, a doença prolifera-se. Por outro lado, o PCR seria mais rápido, porém, o custo da análise torna-o proibitivo de ser aplicado em escala necessária para o controle.

Algumas técnicas rápidas de detecção com potencial de uso em larga escala começam a ser aplicadas no estudo de doenças de plantas. Dentre elas se destaca a espectroscopia de fluorescência, utilizada em estudos de plantas, onde o sintoma ainda não é visível, tais como: a utilização para o diagnóstico do Cancro Cítrico (LINS, 2009), Morte Súbita dos Citros (MSC) e Declínio (ZAGHI, 2008). Também em frutos, Kim et al. (2008) desenvolveram um sistema de imagens pela utilização da técnica de Fluorescência Induzida por Laser. Além do uso da espectroscopia, o uso de imagem tem sido cada vez maior no estudo de doenças de diferentes culturas. Dentre eles, destaca-se a utilização de técnicas de análise de textura aplicadas nas imagens no visível (PYDIPATI et

al., 2006). Lenk et al. (2007) demonstraram a utilização das técnicas de fluorescência e refletância de imagens por meio da análise das folhas das plantas. Kim et al. (2007) desenvolveram um sistema *on-line* para obtenção de imagens hiperespectrais no visível e no infravermelho próximo, aplicado na classificação de maçãs.

Neste artigo pretende-se apresentar os resultados da aplicação de técnicas de seleção de características para geração de classificadores automáticos para detecção do Greening, a partir de dados e imagens de fluorescência.

Para a extração destas características e geração do classificador foram aplicadas algumas técnicas de Mineração de Dados descritas por Witten e Frank (2005).

Foram utilizadas neste artigo as técnicas de Fluorescência Induzida por Laser (FIL) e Imagens de Fluorescência Induzida por Radiação Ultravioleta. Foram gerados três vetores de características, sendo o primeiro composto pelos espectros obtidos com a aplicação da técnica FIL, o segundo composto pelos histogramas dos modelos de cores RGB, HSL e CIE  $La^*b^*$  obtidos das imagens de fluorescência e o terceiro com estatísticas de cada um dos histogramas de cores, ou seja, a média, mediana e coeficiente de curtose.

## 2 Aquisição de Dados

O Grupo Fischer (Grupo Fischer, 2010) disponibilizou à EMBRAPA Instrumentação Agropecuária 120 mudas de Valência com o porta-enxerto *Swingle*. Estas mudas foram plantadas no Instituto Agronômico de Campinas (IAC), localizado na cidade de Cordeirópolis - SP.

Todas as mudas foram podadas para que o crescimento vegetativo fosse estimulado e para que a idade das folhas fosse uniformizada. Desse total de 120 mudas, 60 delas foram inoculadas com o vírus da tristeza (MEISSNER FILHO et al., 2002) e com a bactéria *Candidatus Liberibacter asiaticus*, por meio do

método de enxertia de borbulha, e as restantes foram inoculadas apenas com o vírus da tristeza. Para esse experimento a coleta das folhas foi padronizada.

Durante 28 semanas foram coletadas amostras e analisadas pelas diferentes técnicas descritas a seguir. Para o armazenamento das informações obtidas neste experimento foi desenvolvido pela EMBRAPA Instrumentação Agropecuária o software Greening, em linguagem C# (Visual C# Developer Center, 2010) e as informações foram armazenadas num Banco de Dados PostgreSQL (PostgreSQL, 2010).

### 2.1 Imagem de Fluorescência Induzida por Radiação Ultravioleta

As imagens de fluorescência induzida por luz ultravioleta foram obtidas por meio de uma câmera digital (AxioCamMRc5, Zeiss, Göttingen/Alemanha) de 5M pixels de resolução acoplada a um estereomicroscópio (Lumar.v12, Zeiss, Göttingen/Alemanha). Para excitar a fluorescência foi utilizada uma lâmpada a vapor de mercúrio, sendo selecionado apenas o comprimento de onda da radiação ultravioleta. As folhas ficaram expostas por um período de 545 ms. As imagens obtidas foram aumentadas em 30 vezes com o auxílio de uma lupa e possuem as seguintes dimensões: 1292 (largura) x 968 (altura) pixels com 24 bits de intensidade (PEREIRA e MILORI, 2009). Na Figura 1 pode ser observada a imagem típica de fluorescência para uma folha da estufa, sendo a imagem de fluorescência (a) de uma planta sadia e (b) de uma planta inoculada com a bactéria, após um mês.

### 2.2 Fluorescência Induzida por Laser (FIL)

A Fluorescência Induzida por Laser utiliza a radiação de um laser para obter o espectro de fluorescência das folhas de citros analisadas (ZAGHI, 2008).

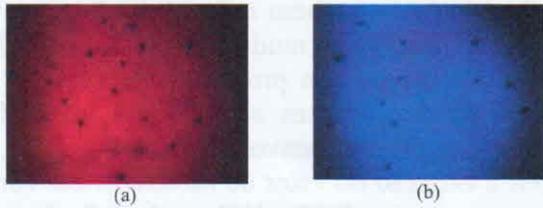


Figura 1 – (a) Imagem de fluorescência típica de uma planta saudável e (b) de uma planta inoculada após um mês.

Neste artigo, foram analisados os comprimentos de onda da faixa de 575 nm até 1.000 nm. Na Figura 2 apresentam-se os espectros típicos para (a) uma planta saudável e (b) uma planta inoculada após o primeiro mês.

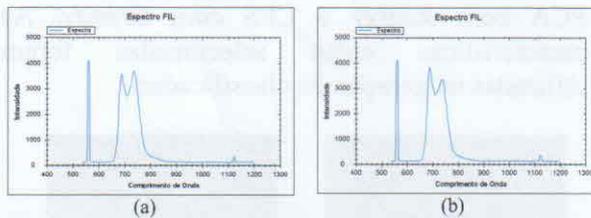


Figura 2 – (a) Espectro de fluorescência típico de uma planta saudável e (b) de uma planta inoculada após 1 mês.

### 3 Extração de características e Classificadores

#### 3.1 Vetor de Características

O primeiro vetor de características é formado por todos os componentes do espectro FIL, normalizado e com *off-set* zerado. Para tanto, todos os valores foram subtraídos do mínimo e divididos pelo máximo valor.

O segundo vetor de características foi formado pelas componentes de cor das imagens de Fluorescência Induzida por Laser, no modelo de cores RGB, HSL e CIE  $L^*a^*b^*$ . Da mesma forma foram normalizados pelos valores máximos de cada componente, visando a sua normalização entre 0 e 1.

O terceiro vetor de características foi composto dos valores da média, da mediana e do coeficiente de curtose, calculados dos

histogramas  $f(x)$  de cada componente de cor, conforme as equações (1), (2) e (3).

$$Média = \frac{1}{n} \cdot \sum_{i=0}^n f(x_i) \quad (1)$$

Na equação (1)  $n$  corresponde ao número de tonalidades do histograma  $f(x)$ . A mediana do histograma  $f(x)$  é a medida de tendência central e é calculada pela equação (2).

$$Mediana = f\left(\frac{n+1}{2}\right) \quad (2)$$

A curtose é a medida de dispersão que caracteriza o "achatamento" da curva da função de distribuição de cores e é calculada por (3):

$$Curtose = \frac{\mu_4}{\sigma^2} - 3 \quad (3)$$

Sendo  $\mu_4$  o quarto momento central, calculado da equação dos momentos estatísticos (4) e  $\sigma$  o desvio padrão, onde  $\bar{X}$  o valor médio de  $X$ .

$$\mu_k = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^k f(x_i) \quad (4)$$

#### 3.2 Seleção de Características e Classificadores

A API (*Application Programming Interface*) do *toolbox* WEKA (WEKA, 2010) foi utilizada neste trabalho para a extração das características e geração do classificador. Para tanto é necessária a utilização da combinação de um algoritmo de seleção de características, e de um algoritmo de busca, que é responsável por realizar a busca no subconjunto de características gerado pela execução dos algoritmos de seleção de características.

Neste artigo foram avaliados os seguintes algoritmos para serem utilizados na extração das características desses vetores: Seleção de características baseados na correlação (*Correlation-based Feature Selection - CFS*) de características e busca gulosa (*Greedy Stepwise*); Componentes Principais (PCA) e *Ranker*. O algoritmo CFS utilizado para seleção de características, avalia o valor de um subconjunto de características, considerando a habilidade preditiva de cada característica permitindo certo grau de redundância entre estas (HALL, 1998); o algoritmo *Greedy Stepwise* faz

uma busca gulosa nos subconjuntos de cada espaço de atributos (WITTEN e FRANK, 2005). O algoritmo PCA é utilizado para seleção de características, transformando os dados para um novo espaço de atributos (WITTEN e FRANK, 2005) segundo a maior variância dos dados; o algoritmo *Ranker* conjuntamente com a PCA para realizar a busca de componentes principais mais relevantes (WITTEN e FRANK, 2005).

Para a classificação desses vetores foram avaliados os algoritmos: o C4.5 e o *Multi Layer Perceptron* (MLP). O algoritmo C4.5 é baseado em uma árvore de decisão cuja a estrutura expressa os dados de maneira rápida e eficiente (QUINLAN, 1993). O algoritmo MLP é baseado em uma rede neural do tipo *Multi Layer Perceptron* com algoritmo *backpropagation* (WITTEN e FRANK, 2005).

O C4.5 foi escolhido pela simplicidade e o MLP pela facilidade de implementar em dispositivo embarcado para uso direto em campo.

#### 4. RESULTADOS E DISCUSSÃO

Foram realizados três experimentos, usando: a) só o vetor de atributos FIL; b) somente os componentes de cor a partir dos histogramas RGB, HSL e CIE La\*b\*; e c) componentes de cor fundidos com dados do espectro de fluorescência. Em cada um dos casos foram feitos estudos das características selecionadas e a melhor classificação encontrada pelos diferentes métodos testados.

##### 4.1 Resultados com as Imagens de Fluorescência a partir do Ultravioleta

Os padrões de fluorescência para as mudas inoculadas com o HLB se alteram ao longo do tempo, como pode ser observado na Figura 3, durante os oito meses do experimento.

No caso de mudas sadias observa-se que a variação é menor, permanecendo mais presente o padrão mais avermelhado. Um exemplo disto é apresentado na Figura 4, onde são evidenciados os histogramas HSL no início e final do experimento, para uma planta com HLB e sadia.

A fluorescência também sofre alteração com o desenvolvimento da muda, apesar da retirada das mudas seguir um protocolo que garante a realização das análises sempre em folhas de mesmo estágio de desenvolvimento.

Após a extração do vetor de características com as componentes RGB, HSL e La\*b\*, foram aplicados os métodos de seleção de características e de classificação descritos no item 3.2. Os resultados são apresentados na Figura 5.

O método J48 implementa o algoritmo C4.5. Este método juntamente com o MLP foi aplicado diretamente no vetor com todas as características de cor. Posteriormente, foram selecionadas as características pelos métodos de PCA com *Ranker* e CFS com *Greedy*. As características então selecionadas foram utilizadas na geração do classificador.

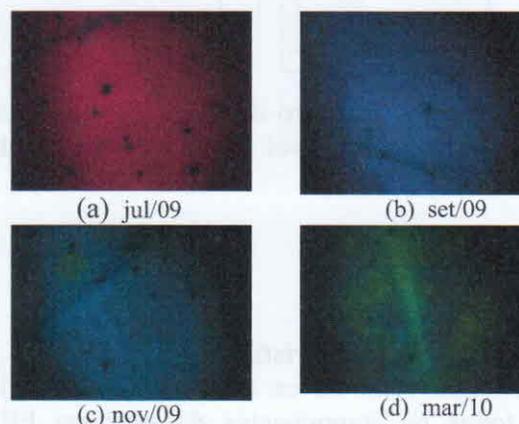


Figura 3: Variação da fluorescência percebida na imagem ao longo do tempo, para uma planta inoculada com o HLB.

Durante a classificação foram utilizados 10 *folds* com validação cruzada, aplicados a todas as imagens coletadas. Os resultados apresentados na Figura 5 mostram a porcentagem de acerto utilizando cada técnica.

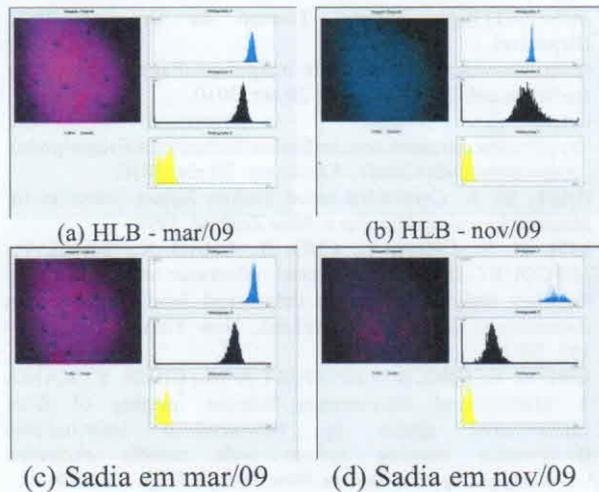


Figura 4: Imagens de fluorescência e histogramas HSL da muda com HLB (a) em mar/09 e (b) em nov/09 e da muda sadia (c) em mar/09 e (d) em nov/09.

Para a rede MLP foi utilizada a taxa de aprendizagem de 0.3, a taxa do momentum foi de 0.2, número máximo de épocas de treinamento foi limitado a 500 e a função de ativação é a sigmóide.

Em geral, os resultados com MLP melhoram a precisão. Também a seleção de características resulta em classificação superior e utilizam menos componentes tornando o processo mais rápido. A detecção só acontece com precisão a partir do quinto mês.

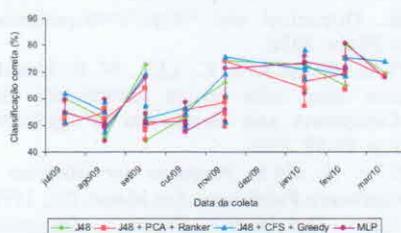


Figura 5: Classificações corretas a partir das imagens de fluorescência.

#### 4.2 Resultados com os Espectros FIL

Os padrões de fluorescência para as mudas inoculadas com o HLB se alteram ao longo do tempo, como pode ser observado na Figura 6. No exemplo apresentado são evidenciados os

espectros FIL no início e no final do experimento, para uma planta com HLB e outra sadia. Percebe-se que os padrões do sinal entre 650nm e 750nm são alterados de uma planta sadia para uma planta com HLB.

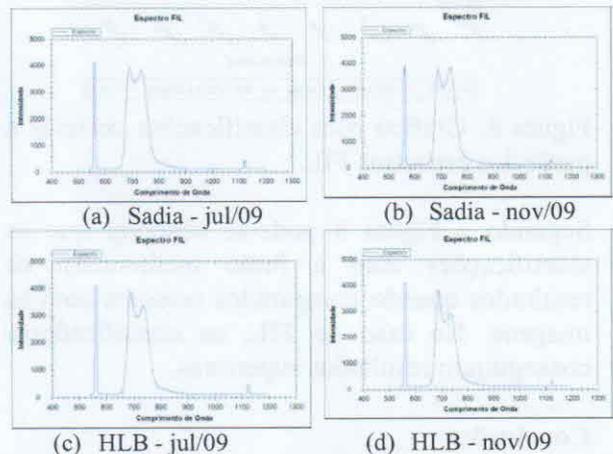


Figura 6: Espectro de fluorescência FIL da muda sadia (a) em mar/09 e (b) em nov/09, e com HLB (c) em mar/09 e (d) em nov/09.

Aplicando os algoritmos de classificação e seleção de características, foram obtidos os resultados de classificação da Figura 7. Nitidamente consegue-se uma detecção maior logo no primeiro mês.

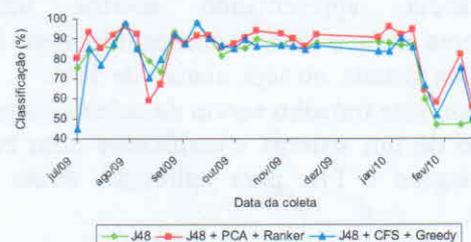


Figura 7: Gráfico com classificações corretas a partir dos espectros FIL.

#### 4.2 Resultados com a fusão dos Espectros FIL, histogramas imagens de Fluorescência e parâmetros estatísticos

Fazendo-se a seleção de características com os dados FIL, os histogramas RGB, HSI e LAB, e os parâmetros curtose, média e mediana dos

histogramas, obteve-se a classificação apresentada na Figura 8.

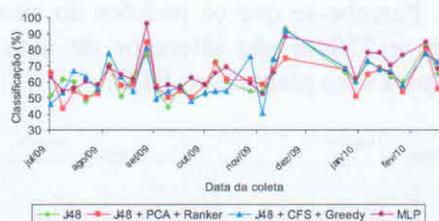


Figura 8: Gráfico com classificações corretas a partir dos espectros FIL.

Segundo a Figura 8 pode-se observar que as classificações com a fusão melhoraram os resultados quando comparados somente com as imagens. No caso do FIL, os classificadores conseguiram resultados superiores.

### Conclusões

Em geral, um número grande de atributos também leva a um erro maior na classificação. Porém, o ideal seria fazer uma busca exaustiva no espaço de estados para selecionar a melhor combinação de atributos para gerar o classificador, o que é inviável dependendo do número de combinações existentes.

Este trabalho apresenta um resultado inédito de classificador aplicado numa área de extrema importância, apresentando acertos muito superiores com o que é conseguido hoje nas inspeções visuais, ou seja, abaixo de 50%.

Também, este trabalho serviu de referência para geração de um sistema classificador com base em imagens e FIL para aplicação direta no campo.

### Agradecimentos

Agradecemos o apoio financeiro do CNPq (processo 578627/2008-6), fornecimento de mudas da empresa Fischer S.A., estufa e condução do experimento pelo Centro de Citricultura do IAC.

### Referências Bibliográficas

BOVÉ, J. M. Huanglongbing: a destructive, newly-emerging, century-old disease of citrus. *Journal of Plant Pathology*, Pisa, v. 88, p.7-37, 2006.

FUNDECITRUS. Manual Técnico do Greening 2009. Disponível em:

<<http://www.fundecitrus.com.br/ImageBank/PageFlip/pageflip.aspx?idPage=143>>. Acesso em: 29 abr. 2010.

Grupo Fischer. Disponível em:

<[http://www.citrosuco.com.br/fischer/fischer/sites/fischer/portal\\_grupo/home/index.html](http://www.citrosuco.com.br/fischer/fischer/sites/fischer/portal_grupo/home/index.html)>. Acesso em: 30 abr. 2010.

HALL, M. A. Correlation-based Feature Subset Selection for Machine Learning. *Hamilton*, New Zealand, 1998.

KIM, M. S.; CHEN, Y.; CHO, B.; CHAO, K.; YANG, C.; LEFCOURT, D. C. Hyperspectral reflectance and fluorescence line-scan imaging for online defect and fecal contamination inspection of apples. *SpringerLink*, New York, v. 1, p. 151-159, 2007.

KIM, M. S.; CHO, B.; LEFCOURT A. M.; CHEN, Y.; KANG, S. Multispectral fluorescence lifetime imaging of feces contaminated apples by time-resolved laser-induced fluorescence imaging system with tunable excitation wavelengths. *Applied Optics*, New York, v.47, p. 10, 2008.

INNIS, M. A.; GELFAND, D. H.; SNINSKY, J. J.; WHITE, T. J. (eds.), PCR Protocols: A guide to methods and applications, Academic Press: San Diego, 1990.

LENK, S.; CHAERLE, L.; PFÜNDEL, E. E.; LANGSDORF, G.; HAGENBEEK, D.; LICHTENTHALER, H. K.; STRAETEN, D. V. D.; BUSCHMANN, C. Multispectral fluorescence and reflectance imaging at the leaf level and its possible applications. *Journal of Experimental Botany*, Oxford, v. 58, p. 807-814, 2007.

LINS, E. C. C. C. Espectroscopia da Fluorescência na Citricultura. São Carlos: Instituto de física de São Carlos, Universidade de São Paulo, 2009.

MEISSNER FILHO, P. E.; SOARES FILHO, W. dos S.; VELAME, K. V. C.; DIAMANTINO, E. P.; DIAMANTINO, M. S. A. S. Reação de porta-enxertos híbridos ao Citrus tristeza virus. *Fitopatologia Brasileira*, Brasília, v. 27, p. 312-315, 2002.

PEREIRA, F. M. V.; MILORI, D. M. B. P. Identificação de variedades genéticas de laranja doce por meio de imagens de fluorescência. São Carlos, SP: Embrapa Instrumentação Agropecuária, 2009. 17 p. (Embrapa Instrumentação Agropecuária. Boletim de Pesquisa e Desenvolvimento, 27). PostgreSQL. Disponível em: <<http://www.postgresql.org/>>. Acesso em: 29 abr. 2010.

PYDIPATI, R.; BURKS, T. F.; LEE, W. S. Identification of citrus disease using color texture features and discriminant analysis. *Computers and electronics in agriculture*, New York, v. 52, p. 49-59, 2006.

QUINLAN, J. R. C4.5: Programs for Machine Learning. *Morgan Kaufmann Publishers*, San Mateo, CA, 1993.

Visual C# Developer Center. Disponível em: <<http://msdn.microsoft.com/pt-br/vcsharp/default.aspx>>. Acesso em: 29 abr 2010.

WEKA. Disponível em: <<http://www.cs.waikato.ac.nz/ml/weka/>>. Acesso em: 29 abr. 2010.

WITTEN, I. H.; FRANK, E. Data mining : practical machine learning tools and techniques. 2 ed. San Francisco, 2005.

ZAGHI, A. F. Diagnóstico de Doenças de Citros Utilizando espectroscopia de fluorescência. São Carlos: Instituto de química de São Carlos, Universidade de São Paulo, 2008.