



7º Congresso Brasileiro de Agroinformática
Agroinformática e a sustentabilidade do agronegócio e dos recursos naturais



21 a 25 de setembro de 2009
Universidade Federal de Viçosa • Viçosa/MG



**ANÁLISE DE COMPONENTES PRINCIPAIS E REDES NEURAIS ARTIFICIAIS NO
PROCESSAMENTO DIGITAL DE IMAGENS PARA IDENTIFICAÇÃO DA
SIGATOKA NEGRA EM BANANAIS**

SILVIA HELENA MODENÈSE GORLA DA SILVA¹

HUGO DO NASCIMENTO BENDINI²

WÍLSON DA SILVA MORAES³

LÚCIO ANDRÉ DE CASTRO JORGE⁴

VILMAR ANTONIO RODRIGUES⁵

CARLOS ROBERTO PADOVANI⁶

RESUMO: O presente trabalho investigou a aplicação, mais especificamente o processamento digital de imagens, com análise de componentes principais e redes neurais artificiais como ferramentas de apoio para uma melhor identificação dos estádios iniciais do desenvolvimento da Sigatoka Negra, em nível de campo, para que medidas de controle sejam adotadas mais rapidamente e, assim, reduzir danos e prejuízos causados pela doença na bananicultura. Foram coletadas imagens digitais de folhas de bananeiras infectadas com a Sigatoka Negra nos estádios 1, 2 e 3, sadia e com fitotoxidez por óleo. A seguir, extraíram-se histogramas dos componentes de imagens no sistema RGB para 256 tons de cinza, das amostras. Com isso, fez-se necessário a aplicação de uma técnica de seleção de atributos, a análise de componentes principais. Assim, conseguiu-se reduzir as 768 variáveis de entrada de cada exemplo, que nada mais são que combinações lineares das variáveis de entradas originais. Por último, realizou-se o treinamento das redes neurais artificiais para reconhecimento de cada uma das classes citadas.

PALAVRAS-CHAVE: Sigatoka Negra, processamento digital de imagens, redes neurais artificiais, análise de componentes principais.

**PRINCIPAL COMPONENTS ANALYSIS AND ARTIFICIAL NEURAL NETWORKS
IN THE DIGITAL IMAGE PROCESSING FOR IDENTIFICATION OF BLACK
SIGATOKA IN BANANA CROP**

ABSTRACT: This study investigated the application, specifically the digital processing of images, with principal component analysis and artificial neural networks as tools to support for better identification of the early stadiuns 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. Images were collected from digital leaves of banana crop infected by the Black Sigatoka in stages 1, 2 and 3, 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 samples. This made it necessary to apply a technique of attributes selection, the principal component analysis. Thus, it was able to reduce the 768 input variables of each example, 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.

KEY-WORDS: Black Sigatoka, digital image processing, artificial neural networks, principal components analysis.

¹ Tecnóloga em Processamento de Dados, Campus Experimental de Registro/UNESP: silvia@registro.unesp.br.

² Aluno de graduação do curso de Agronomia, Campus Experimental de Registro/UNESP: h_bendini@registro.unesp.br

³ Engenheiro Agrônomo, Campus Experimental de Registro/UNESP: wilson@registro.unesp.br

⁴ Engenheiro Elétrico, Centro Nacional de Pesquisas em Instrumentação Agropecuária – CNPDIA/EMBRAPA: lucio@cnpdia.embrapa.br

⁵ Engenheiro Agrimensur, Campus Experimental de Registro/UNESP: vilmar@registro.unesp.br

⁶ Matemático, Departamento de Bioestatística – IBB/UNESP – Botucatu/SP – Brasil.

1. INTRODUÇÃO

A Sigatoka Negra, causada pelo fungo *Mycosphaerella fijiensis* Morelet, é a mais severa e destrutiva doença da bananeira em todas as áreas produtoras do mundo, sendo responsável por perdas de produção superiores a 50%. A doença ocorre nas folhas provocando estrias marrons e manchas negras necróticas que reduzem os tecidos fotossintetizantes e, conseqüentemente, os rendimentos brutos. O primeiro registro da doença, em território brasileiro, deu-se no estado do Amazonas no ano de 1998 e, em 2004, a doença foi constatada em bananais do Vale do Ribeira. Atualmente, encontra-se disseminada em todas as regiões produtoras do estado de São Paulo (FERRARI, 2005).

O controle da doença é feito por meio de pulverizações alternadas de fungicidas sistêmicos e protetores. Entretanto, o monitoramento semanal é uma tarefa trabalhosa que requer uma estrutura mínima e envolve despesas contínuas com deslocamento e treinamento de técnicos capazes de identificar os estádios de desenvolvimento da doença. Os registros semanais dos estádios da doença e da emissão de folhas são transferidos para planilhas que calculam o estado de evolução atual da severidade e indicam a necessidade ou não da aplicação de fungicidas, cuja decisão é auxiliada pelos dados climáticos do local. Esse método pode induzir a erros de observação, uma vez que é preciso distinguir as lesões mortas (sobre efeitos de fungicidas) das lesões vivas, as lesões de Sigatoka Negra de Sigatoka Amarela, viroses e de fitotoxidez do óleo mineral, aplicado junto da calda fungicida, pois tarefas que necessitam de experiências e subjetividade do ser humano são altamente suscetíveis a erros. Outras imprecisões de controle podem ser atribuídas às características geomorfológicas, cobertura vegetal, estradas de difícil acesso a bananais e às propriedades de pequeno porte, que tornam praticamente impossível realizar um diagnóstico preciso usando os métodos tradicionais.

No mundo globalizado, a agricultura intensiva depende, de forma crescente e irreversível, da utilização de insumos e tecnologias modernas, a fim de que os resultados sejam beneficiados cada vez mais com os avanços tecnológicos.

Uma alternativa interessante e atraente para lidar com este tipo de problema envolvendo a agricultura de precisão, trata-se do processamento digital de imagens associado com a análise multivariada de dados. Neste sentido, pretendeu-se na presente pesquisa, utilizar a aplicação da agricultura de precisão; mais especificamente o processamento digital de imagens com análise de componentes principais (ACP), uma técnica da estatística multivariada de dados (JOHNSON; WICHERN, 2002), e redes neurais artificiais (RNA) como ferramentas de apoio para uma melhor identificação dos estádios iniciais do desenvolvimento da Sigatoka Negra, em nível de campo, para que medidas de controle sejam adotadas mais rapidamente e, assim, reduzir danos e prejuízos causados pela doença na bananicultura. A ACP tem por objetivo substituir um conjunto de variáveis correlacionadas por um conjunto de novas variáveis não correlacionadas entre si, as quais são combinações lineares das variáveis iniciais e apresentadas em ordem decrescente da magnitude de suas variâncias (HOTELLING, 1936; JOHNSON; WICHERN, 2002). As RNA se baseiam em uma forma de computação não algorítmica inspirada no cérebro humano (BRAGA et al., 2000).

2. MATERIAL E MÉTODOS

Os Em Moraes et al. (2005) desenvolveu-se um aplicativo computacional que realiza estimativa da severidade da Sigatoka Negra, baseado no método de Foure (1988). Esse método descreve a evolução da doença em seis fases ou estádios. Com esse aplicativo, realiza-se o monitoramento da doença em bananais do Vale do Ribeira. No entanto, o monitoramento é realizado com base na classificação dos estádios segundo o avaliador. Sabe-se que, tarefas que envolvam a subjetividade do ser humano são altamente suscetíveis a erros. Assim, este trabalho teve por objetivo utilizar o processamento digital de imagens na

identificação dos sintomas da doença. Para tanto, o método consistiu-se de quatro fases distintas: coleta e digitalização das amostras, segmentação e extração dos histogramas das imagens, seleção de atributos e classificação.

Primeiramente, foram coletadas folhas de bananas, das quais extraíram-se 10 amostras de aproximadamente 2cm x 2cm para cada uma das classes: Sigatoka Negra nos estádios 1, 2 e 3, Sigatoka Amarela, fitotoxidez por óleo e folhas sadias, totalizando 60 amostras. Essas amostras foram levadas ao laboratório do Campus Experimental de Registro e, posteriormente, escaneadas em escaner de mesa da marca HP a uma resolução de 640 pixels. No entanto, nesse primeiro experimento, como será visto no próximo capítulo, os resultados não foram tão satisfatórios. Assim, realizou-se um segundo experimento onde foram coletadas 20 amostras de folhas para cada uma das referidas classes, com exceção da classe Sigatoka Amarela que, de acordo com especialista da área, não necessita-se desse conhecimento para o monitoramento. Dessa maneira, no segundo experimento, fez-se um total de 100 amostras. Na segunda fase, procurou-se realizar a segmentação e um estudo de frequência das intensidades de cores dos pixels. Para tal, construiu-se uma ferramenta computacional, com o uso do software HALCON (MVTec, 2000), que decompõe cada imagem nos três componentes de cores do sistema RGB. De cada componente de imagens, organizou-se um histograma da intensidade dos níveis de cinza (256 níveis). Assim, cada imagem passou a ter 769 ($3 \cdot 256 + 1$) variáveis para representá-la, ou seja, a frequência de cada um dos níveis de cinza para cada um dos componentes de cores RGB, representado as 768 variáveis de entrada e, também, uma variável de saída representando a classe de cada imagem ou exemplo.

Com o número excessivo de variáveis de entrada (768), fez-se necessário, na etapa três do trabalho, a aplicação de uma técnica de seleção de atributos para que se conseguisse diminuir o número de variáveis de entrada e, assim, facilitar computacionalmente o treinamento das RNA na fase de classificação das imagens. Para isto, aplicou-se a técnica multivariada de dados expressa na análise de componentes principais que permite a redução do espaço paramétrico, praticamente, sem prejuízo da informação biológica. A ferramenta selecionada para indução dos algoritmos a serem utilizados foi a Weka (WITTEN; FRANK, 1999). A Weka é uma ferramenta para Aprendizado de Máquina (MITCHELL, 1997) desenvolvida na Universidade de Waikato, na Nova Zelândia, escrita em linguagem JAVA, com código aberto, que permite alterações nas funções a serem utilizadas, bem como a inserção de implementações não suportadas. A operacionalização da Weka é de domínio público disponível no endereço eletrônico: <http://www.cs.waikato.ac.nz/ml/weka>.

Por último, finalizando o modelo proposto, tem-se a etapa de interpretação da imagem. A etapa de interpretação é a mais “inteligente” do processamento digital da imagem, pois permite obter a compreensão e a descrição final do fenômeno. É na etapa de interpretação de imagem que utiliza-se a técnica de RNA para classificação dos referidos padrões coletados. A rede neural empregada foi uma rede neural multicamadas, treinada com algoritmo *backpropagation*. Nesta fase de classificação também utilizou-se a ferramenta Weka.

Toda a execução da experimentação computacional foi realizada em um microcomputador Intel Pentium 4 3,06 GHz, com 2 Gbytes de memória RAM e 200 Gbytes de memória auxiliar.

Continuando o procedimento operacional, para verificar a validação (consistência) do modelo proposto, foram utilizados alguns procedimentos e técnicas convencionais de confiabilidade. No caso de tarefas de classificação, uma medida comumente utilizada é a frequência (taxa) de erro do classificador h , também conhecida como probabilidade (taxa) de classificação incorreta ou errônea. A probabilidade clássica do erro é obtida utilizando a expressão

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

, a qual acumula as discordâncias entre a classe original de cada amostra com o rótulo atribuído pelo classificador induzido ponderada pelo total de amostras.

Utilizando a função indicadora, tem-se que se a indicação de classificação for concordante, há uma classificação correta, ou seja, expressão lógica verdadeira e, neste caso, o contador indica o valor 1, caso contrário, expressão lógica falsa, e o valor indicado consiste de 0. O procedimento é obtido repetido para todas as amostras e o operador matemático vai acumulando o resultado até o número n de unidades amostrais. O complemento da probabilidade de erro é tomado como a precisão do classificador (BARANAUSKAS, 2000). Para obter uma estimativa não viciada de erro do classificador, utilizou-se de técnicas de amostragem para manipulação dos conjuntos de treinamento e teste. Neste trabalho, considerou-se a metodologia *r-fold cross-validation* (EFRON; GONG, 1983; EFRON; TIBSHIRANI, 1993), que divide fraciona o conjunto de dados total, de tamanho n , em r partes mutuamente exclusivas (*folds*) de tamanhos iguais a n/r amostras. As amostras nos $(r - 1)$ *folds* são usadas para treinamento e a hipótese induzida é testada no *fold* remanescente. Este processo é repetido r vezes, cada vez considerando um *fold* diferente para teste. O erro no *cross-validation* é a média dos erros calculados em cada um dos r *folds* de teste (BARANAUSKAS, 2000).

3. RESULTADOS E DISCUSSÕES

Após as amostras serem digitalizadas, realizou-se a segmentação e o estudo de frequências das imagens decompostas no sistema RGB. Nesta fase, gerou-se um arquivo texto contendo as 768 variáveis, representando as frequências de intensidade dos 256 níveis de cinza para cada uma das três imagens decompostas, dos 60 primeiros padrões e, posteriormente, dos 100 padrões do segundo experimento. Nesse momento, fez-se um primeiro teste de classificação com as RNA, na ferramenta Weka, e verificou-se que não se conseguia ter um resultado factível, devido à grande quantidade de variáveis de entrada. Dessa maneira, teve-se a necessidade de se ter a seleção de atributos aplicando a análise de componentes principais. Com isso, gerou-se um arquivo contendo somente 22 variáveis de entrada, no primeiro experimento e 11 no segundo, que são combinações lineares das 768 variáveis originais. Assim, os arquivos de entrada para a fase de classificação foram compostos por 22 variáveis de entrada, no primeiro experimento e 11 no segundo, e uma variável de saída representando a classe de um determinado padrão. Nessa quarta fase, portanto, foi realizado o treinamento das RNA Perceptron multicamadas, treinadas com o algoritmo *backpropagation*, com termo *momentum*. A taxa de aprendizado e o termo *momentum* adotados, que são parâmetros do algoritmo de aprendizado escolhidos de maneira empírica, foram respectivamente iguais a 0,2 e 0,3.

Como resultado de classificação do primeiro experimento, a RNA conseguiu uma taxa de erro médio nos conjuntos de teste de 48,33%. Já no segundo experimento, onde procurou-se coletar amostras mais representativas conseguiu-se um erro médio significativamente mais baixo, nos conjuntos de teste, igual a 18%. A alta taxa de erro do experimento um, talvez, deve ter ocorrido pelo motivo das amostras não representarem bem as classes designadas. Dessa maneira, teve-se a necessidade da realização de outros testes com um maior número de amostras para cada classe, o que originou o segundo experimento.

4. CONSIDERAÇÕES FINAIS

Este trabalho investigou a utilização de técnicas multivariadas e o uso de Redes Neurais Artificiais no processamento digital de imagens para identificação de estádios da Sigatoka Negra. Para realização desta tarefa, foram aplicadas as técnicas de análise de componentes principais e Redes Neurais Artificiais. A primeira técnica foi utilizada na seleção de atributos e, a segunda, utilizada na fase de classificação. Os resultados obtidos no experimento um indicam que as taxas de acertos não se mostraram suficientes para um bom reconhecimento

dos padrões. Isto mostrou a necessidade de, para melhora de classificação, de se repetir o experimento, coletando-se um maior e mais representativo número de amostras para cada classe estudada. Com isso, conseguiu-se melhorar significativamente o desempenho do classificador. Também, poderão ser feitos testes para reconhecimento da Sigatoka Negra com o uso de imagens aéreas, servindo de ferramenta, principalmente, para um estudo da distribuição geográfica da doença. Com o uso do processamento digital de imagens e de técnicas de interpretação de imagens estudados puderam ser estabelecidos os primeiros estudos de padrões para identificação dos estádios da Sigatoka Negra, bem como da Sigatoka Amarela e fitotoxidez por óleo. Assim, esse estudo serviu como primeiro trabalho e possibilitou a identificação de outros trabalhos futuros que servirão de ferramentas para o uso racional de energia, otimizando a aplicação de fungicidas. Ao mesmo tempo, será possível localizar geograficamente a presença exclusiva de Sigatoka Negra e, ou Sigatoka Amarela e mapear regiões ou microbacias, importantes para a implantação de um Sistema de Previsão bio-climática da doença.

5. REFERÊNCIAS BIBLIOGRÁFICAS

- 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.; GONG, G. A leisurely Look at the bootstrap, the jackknife, and crossvalidation. **The American Statistician**, Washington, DC, v.37, n.1, p.36-48. 1983.
- 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.
- FOURÉ, E. Stratégies de lutte contre la Cercospora noire des bananiers et des plantains provoqué par *Mycosphaerella fijiensis* Morelet. L'avertissement biologique au Cameroun. Evaluation des possibilités de amélioration. **Fruits**, Paris, v. 43, n. 5: p. 269 – 274. 1988.
- GONZALEZ, R. C. 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.
- MORAES, W. da S.; FUKUDA, E.; MODONESE-GORLA DASILVA, S. H.; MENDONÇA, J. C.; LIMA, J. D. MENDES, C. DA S. Aplicativo para estimativa biológica da Sigatoka-negra (*Mycosphaerella fijiensis* Morelet). **Fitopatologia Brasileira**, Brasília, DF, v. 30 (Supl): p.193. 2005.
- MVTec Software GmbH. MVTec Halcon. Germany, January, 2000.(<http://mvtec.com>).
- 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.