

UTILIZAÇÃO DA MODELAGEM MATEMÁTICA (REDES NEURAIS ARTIFICIAIS) NA CLASSIFICAÇÃO DE AUTOTETRAPLOIDES DE BANANEIRA (*Musa acuminata* COLLA)

USE OF MATHEMATICAL MODELING (ARTIFICIAL NEURAL NETWORKS) IN CLASSIFICATION OF BANANA AUTOTETRAPLOID (*Musa acuminata* COLLA)

Ana Catarina Lima de OLIVEIRA¹; Moacir PASQUAL²; Leila Aparecida Salles PIO³; Wilian Soares LACERDA⁴; Sebastião de OLIVEIRA E SILVA⁵

1. Doutoranda em Agronomia/Fitotecnia, Universidade Federal de Lavras - UFLA, Lavras, MG, Brasil e Professora do EBTT – IFS, N. Sra. Da Glória, SE, Brasil. kata_lima@yahoo.com.br. 2. Doutor em Genética e Melhoramento de Plantas, Professor Titular - UFLA, Lavras, MG, Brasil. 3. Doutora em Agronomia/Fitotecnia – UFLA, Lavras, MG, Brasil; 4. Doutor em Engenharia Elétrica, Professor Adjunto - UFLA, Lavras, MG, Brasil; 5. Doutor em Genética e Melhoramento de Plantas, Professor Colaborador UFRB, Cruz das Almas, BA, Brasil.

RESUMO: Objetivou-se desenvolver uma metodologia para possibilitar a classificação de plantas de bananeira submetidas à indução de duplicação cromossômica utilizando Redes Neurais Artificiais (RNA). Os dados utilizados neste trabalho foram retirados de uma tese já apresentada, cujo autor estudou a correlação entre a massa fresca de discos foliares e o conteúdo de DNA. A RNA foi implementada com a função de classificação. A taxa de aprendizado e o termo *momentum* adotados foram respectivamente iguais a 0,01 e 0,2, o número de épocas de treinamento foi 1000. Esses valores foram determinados por meio de tentativa e erro. Para o treinamento, 90% das plantas foram utilizadas e, para validação, 10% do total de 114 autotetraploides produzidos artificialmente por meio de exposição ao antimetabólito colchicina. A RNA classificou corretamente 10 das 11 amostras utilizadas para validação. A estatística Kappa foi de 63,33%, o que indica que a RNA pode ainda ser melhorada. A rede neural artificial do tipo *Multi Layer Perceptron* implementada é eficaz na seleção de poliploides desejáveis de bananeira Tong Dok Mak.

PALAVRAS-CHAVE: Modelagem Matemática. Aprendizagem. Neurônios.

INTRODUÇÃO

Embora exista um número expressivo de variedades de bananeira no Brasil, quando se considera preferência dos consumidores, produtividade, tolerância a pragas, porte adequado, resistência à seca e ao frio, restam poucas variedades com potencial agrônomo para utilização comercial (SILVA et al., 2002). A maioria das variedades comerciais é pouco produtiva, possui porte alto e ainda está associada à falta de resistência às principais doenças e pragas.

O melhoramento genético da bananeira, em especial dos genótipos diplóides (AA), visa ao desenvolvimento de variedades comerciais resistentes às principais doenças, tais como Mal-do-panamá, Sigatoka Amarela e Negra, Moko, dentre outras características indesejáveis. Porém, esse processo tem sido dificultado pela ausência de sementes ou pela inviabilidade frequente do pólen (SILVA et al., 1999). Esse fato dificulta o melhoramento convencional da bananeira, em especial nas bananeiras triplóides (AAA) pertencentes ao subgrupo Cavendish, devido à presença da partenocarpia e esterilidade.

A fim de contornar esse entrave, os programas de melhoramento genético têm dado atenção especial à obtenção de autotetraploides induzidos em bananeira. Esses autotetraploides são obtidos por meio da duplicação do número básico de cromossomos de um diplóide, o que resultará em um indivíduo com quatro genomas idênticos. Esse processo pode ser realizado com a utilização de substâncias antimetabólicas, sendo a colchicina a mais amplamente empregada. Esses tetraploides artificiais serão utilizados no cruzamento com outros diplóides, gerando triplóides com características desejáveis incorporadas. Um dos problemas encontrados no processo de indução artificiais de ploidia é a regeneração de plantas mixoploides, isto é, plantas que apresentam células com variação no número cromossômico em um mesmo tecido ou entre órgãos de uma mesma planta.

Após a indução da poliploidia, o procedimento seguinte é a confirmação da duplicação cromossômica, cujos principais métodos de análise são: contagem cromossômica, avaliação de caracterização citogenéticas e estimativa do conteúdo de DNA pela citometria de fluxo, e a contagem do número de cromossomos de células

mitóticas. Contudo, esse procedimento é moroso e trabalhoso, não recomendado, portanto, quando se trabalha com um grande número de plantas. Outro método é a caracterização citoanatômica, que se dá pela observação do “efeito gigas” no qual órgãos e células de uma planta aumentam proporcionalmente a quantidade de DNA, no entanto, isso pode levar a resultados errôneos, pois essas características podem perfeitamente ser influenciadas pelo ambiente de cultivo (SARI; ABAK; PITRAT, 1999; SOUZA; QUEIROZ, 2004; VILLA, 1995).

Assim para avaliar maior número de indivíduos de citometria de fluxo é útil apesar da facilidade a citometria de fluxo é um método caro e que requer mão-de-obra qualificada.

Nesse contexto, a modelagem matemática surge como uma ferramenta auxiliar no melhoramento genético dessa cultura, pois se pode realizar a classificação dessas plantas expostas a antimitóticos, em diplóides, mixoploides e tetraploides. Nos últimos anos, o desenvolvimento da inteligência artificial, em especial das redes neurais artificiais (RNA) aplicadas à agricultura, facilitou diversos trabalhos de previsão, reconhecimento e classificação e resolveu problemas nas diversas áreas do conhecimento. Esse tipo de rede visa à representação do comportamento humano através de modelos computacionais, de maneira geral projetados para modelar a maneira como o cérebro realiza uma tarefa particular ou uma função de interesse (RUSSEL; NORVIG, 1995).

Louro, Mendonça e Gonzaga (2006), na tentativa de minimizar a perda de fruto, aplicou um RNA no sistema de produção de tomate. Essa rede classificava o tomate quanto à preferência do consumidor. Esse autor relatou que houve uma diminuição da perda de frutos, visto que o fator humano não mais afetava a seleção, pois os critérios eram matemáticos e pré-estabelecidos. Entretanto, deve-se tomar cuidado na escolha das variáveis, porque grande quantidade de materiais não garante bom desenvolvimento da rede. Os tomates foram classificados como verdes ou maduros, pequenos ou grandes.

Por meio da classificação feita pela RNA e por métodos estatísticos, Shahin, Symons e Poysa (2006) classificaram as sementes de soja quanto à dimensão por meio de processamento de imagens.

Objetivou-se, com este trabalho, desenvolver uma metodologia que possibilite a classificação de plantas de bananeira submetidas à duplicação cromossômica utilizando redes neurais artificiais.

MATERIAL E MÉTODOS

Obtenção dos dados

Os dados utilizados neste trabalho foram obtidos junto à Universidade Federal de Lavras (UFLA) e se refere a experimentos relacionados à tese intitulada *Respostas Morfogênicas de Bananeira submetida à Poliploidização* localizadas no município de Cruz Das Almas, Bahia (COSTA, 2010).

Material vegetal

A variedade utilizada foi a Tong Dok Mak (TDM). Os ápices caulinares dessa variedade foram submetidos à duplicação cromossômica por meio de tratamento com colchicina (0; 1,25; 2,5; 3,75 e 5 mM, por 24 e 48 horas). Para aclimatização, as plantas alongadas e enraizadas em sala de crescimento de luz natural foram retiradas dos frascos de cultivo, submetidas à lavagem e poda das suas raízes e, imediatamente, transferidas para sacos de polietileno de três litros de capacidade preenchidos com substrato composto do substrato casca de Pinus grânulos médios.

Em seguida, as plantas foram mantidas em telado por 90 dias, coberto com filme de polietileno transparente (150 microns) e sombrite (50% de interceptação luminosa). A irrigação foi feita de forma manual e, a cada 15 dias, realizou-se a adubação foliar com macro e micronutrientes.

Discos foliares de plantas de bananeira variedade Tong Dok Mak tratadas com colchicina

Para a obtenção dos discos foliares, as folhas foram coletadas de plantas da mesma idade, nas primeiras horas de cada manhã, especificamente entre seis e sete horas. Imediatamente após a excisão, as folhas foram mantidas em sacos plásticos fechados para manter a umidade e evitar a perda excessiva de água por transpiração. Ao final da coleta, foram conduzidas ao laboratório para a determinação da massa fresca dos discos. Os discos foliares foram retirados na região mediana do limbo foliar da segunda folha expandida (direção ápice base). Para tanto, utilizou-se um vazador de 1,608 cm de diâmetro (2,0297 cm² de área interna).

Citometria de fluxo de plantas de bananeira variedade Tong Dok Mak tratadas com colchicina

As mesmas plantas, das quais foram retirados os discos foliares, foram analisadas utilizando a técnica de citometria de fluxo. O preparo das amostras e as análises foram conduzidos

nos Laboratórios de Genética e Imunologia da Universidade Federal de Juiz de Fora, Juiz de Fora, Minas Gerais.

Foram utilizados aproximadamente 20-30mg de tecido foliar fresco obtido a partir de uma amostra da folha mais jovem de cada planta. Como padrão externo de referência, utilizou-se a mesma quantidade de tecido foliar de *Glycine Max*. Para preparação das amostras e obtenção de suspensão nuclear, o tecido foliar foi triturado em placa de Petri contendo 1mL de tampão LB01.

A suspensão de núcleos foi aspirada através de duas camadas de gaze, com auxílio de uma pipeta plástica, e filtrada através de uma malha de 50 µm. Os núcleos foram corados pela adição de 25 µL de uma solução de 1mg/1mL de Iodeto de Propídeo. As amostras foram armazenadas em refrigerador e analisadas após 1-2 horas. Para cada amostra, pelo menos 5 mil núcleos foram analisados utilizando-se uma escala logarítmica. A análise foi realizada no citômetro FacsCalibur (BD, Biosciences, San Jose,

CA, USA), e os histogramas foram obtidos com o software Cell Quest (Becton Dickinson e Companhia, San Jose, CA, USA) e analisados estatisticamente no software WinMDI 2.8 (SCRIPPS RESEARCH INSTITUTE, 2011).

O conteúdo de DNA nuclear em (pg) das plantas foi estimado utilizando-se a fórmula a

$$\text{DNA amostra} = \left(\frac{\text{G1 Amostra}}{\text{G1 padrão}} \right) \times \text{DNA padrão}$$

seguir:

onde o DNA da amostra refere-se à quantidade de DNA nuclear da amostra, o G1 da amostra corresponde à posição do pico G1 da amostra. O G1 padrão à posição do pico G1 da espécie padrão de referência usado e o DNA padrão ao conteúdo de DNA da espécie padrão de referência utilizada.

Redes neurais

Cada planta foi classificada de acordo com seu conteúdo de DNA, como visto na Tabela 1.

Tabela 1. Quantidade de plantas classificadas de acordo com a ploidia. UFLA, Lavras, MG, 2011.

Ploidia	Número de Plantas
Diplóide	13
Mixoploide	36
Tetraploide	65

Esses dados foram correlacionados com a massa fresca de discos foliares das respectivas plantas, sendo este o atributo utilizado como variável de entrada para a implementação da RNA. Não foi necessária a padronização desses dados, pois todos já se encontravam entre 0 e 1. A mineração dos dados, que se encontravam em um arquivo de formato “.xls”, foi realizada no software livre WEKA®.

O algoritmo de treinamento de dados utilizado foi o de classificação e a variável de saída foi a ploidia (diplóide, mixoploide e tetraploide).

Os resultados foram obtidos utilizando *percentage split*. Esse método de amostragem particiona o conjunto de dados em partes que o usuário escolhe. Nesse caso, 90% dos dados foram escolhidos para treinamento (103 plantas) e 10% para a validação (11 plantas).

Foram utilizadas RNAs do tipo MLP - *MultiLayer Perceptron* (Perceptron de multicamadas) que possuem uma ou mais camadas escondidas. O algoritmo de treinamento utilizado foi o *backpropagation* (RUMELHART; HINTON; WILLIAMS, 1986), com termo *momentum* adicionado. A taxa de aprendizado e o termo

momentum adotados foram respectivamente iguais a 0,01 e 0,2. O número de camadas ocultas da rede foi igual a dois. Sendo assim, a arquitetura da rede foi 2, 10, 10, 3, ou seja, dois neurônios na camada de entrada, dez em cada camada oculta e três na camada de saída, o número de épocas de treinamento foi estipulado em 1000. Esses valores foram determinados por meio de tentativa e erro.

RESULTADOS E DISCUSSÃO

O tipo MLP de Rede Neural foi utilizado, pois os dados de saída não são linearmente separáveis. No conjunto de validação, apenas uma planta foi classificada incorretamente (Tabela 2).

Na Tabela 3, é observado o peso médio para os principais parâmetros de treinamento.

A Figura 1 apresenta o valor da estatística Kappa de 63,33%, que indica quão próximos estão os dados obtidos do observado. Esse valor de Kappa é satisfatório, porém pode ser melhorado. Costa (2010) encontrou correlação entre poliploides identificados pela técnica de discos foliares e aquelas analisadas por citometria de fluxo. Utilizando a estatística convencional, por meio da

análise de variância e de comparação das médias pelo Teste de Tukey a 1% de probabilidade, esse mesmo autor relatou que, por meio da massa fresca de discos foliares, foi possível diferenciar plantas diplóides não tratadas com antimitóticos de poliploides. Entretanto, esse tipo de análise não

possibilita a classificação de outras plantas, além das analisadas no experimento. A RNA, diferente disso, é uma ferramenta pela qual é possível a classificação de outras plantas, não apenas as analisadas.

Tabela 2. Matriz de confusão com as taxas de acerto médio obtidas para cada classe, de acordo com os conjuntos de teste. UFLA, Lavras, MG, 2011.

Ploidia	Mixoplóide	Tetraplóide	Diplóide
Mixoploide	9	0	0
Tetraploide	0	9	0
Diplóide	0	1	0

Tabela 3. Taxa de Verdadeiro Positivo, Falso Positivo, Precisão e Recall e Classes. UFLA, Lavras, MG, 2011.

	Taxa de Verdadeiro Positivo	Taxa de Falso Positivo	Precisão	Recall	Classes
Média Ponderada	0,909	0,409	0,827	0,909	Mixoploide Diplóide Tetraploide

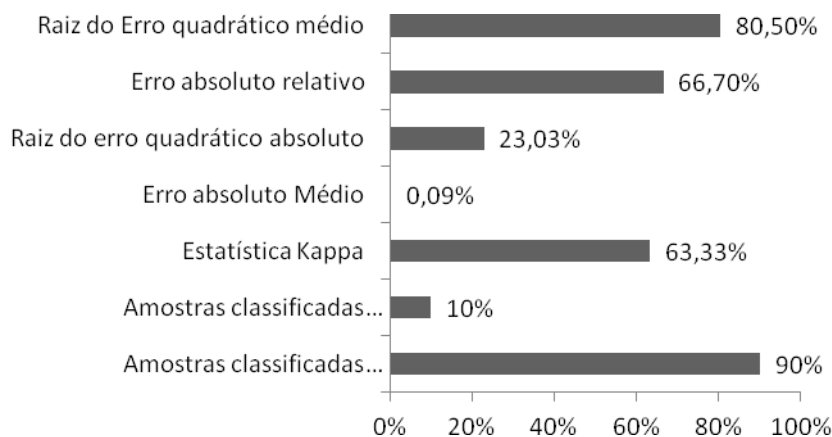


Figura 1. Valores de parâmetros após o treinamento. UFLA, Lavras, MG, 2011.

A rede neural apresentou resultados satisfatórios, porém, para a classificação, pode ser melhorada. Dessa forma, como pode diminuir o total de amostras analisadas no citômetro de fluxo, há apenas a necessidade de realizar as amostras nas que apresentarem características agrônômicas que levem à incerteza da classificação realizada com a utilização da RNA.

Chagas et al. (2009), trabalhando com a classificação dos níveis de degradação de pastagens na zona da mata mineira por meio de RNA, constatou que esse é um classificador viável para este fim, principalmente em grandes áreas. A eficiência da utilização de redes neurais se baseou na sua arquitetura. Nesse trabalho, os autores usaram uma rede do tipo MLP de uma camada interna com seis neurônios e os pesos foram

inicializados com pequenos pesos aleatórios (variando entre -0,5 e 0,5), com a taxa de aprendizagem de 0,2 e o número de épocas igual a 10000. Esses valores foram obtidos também pela tentativa e erro.

A rede neural artificial do tipo MLP também foi eficaz na classificação de três espécies de Iris (*Iris setosa*, *Iris virginica* e *Iris versicolor*) com base nas características da flor de cada uma. A rede também possuía apenas uma camada interna e a topologia adotada foi 4x3x2 (quatro neurônios na camada de entrada, três neurônios na camada oculta e dois neurônios na camada de saída). O algoritmo de treinamento utilizado foi o *backpropagation*. O

programa utilizado na execução dos treinamentos foi o Matlab®. O menor erro encontrado foi de 0,00001, quando a rede neural utilizou 80% dos dados para treinamento e os outros 20% para validação, resultado este considerado como satisfatório (PLOTZE, 2004).

CONCLUSÃO

A rede neural artificial do tipo MLP implementada é eficaz na pré-seleção de poliploides desejáveis de bananeira Tong Dok Mak.

ABSTRACT: The objective was to develop a methodology to enable the classification of banana crop subjected to induction of chromosome doubling using Neural Networks (NN). The data used in this study were taken from a thesis already presented, whose authors studied the correlation between fresh weight of leaf discs and DNA content. The NN was implemented by the ranking function. The learning rate and momentum term used were respectively equal to 0.01 and 0.2, the number of training epochs was 1000. These values were determined by trial and error. For training, 90% of the plants were employed, and for validation, 10% of the total of 114 autotetraploids artificially produced by exposure to antimetabolic agent colchicine. The NN correctly classified 10 of the 11 samples used for validation. Kappa statistics was 63.33%, which indicates that the NN can be further improved. The artificial neural network-type Multi Layer Perceptron is effectively implemented in the pre-selection of polyploid desirable banana Tong Dok Mak.

KEYWORDS: Mathematical Modeling. Learning. Neurons.

REFERÊNCIAS

- CHAGAS, C. S.; VIEIRA, C. A. O.; FILHO FERNANDES, E. I.; JÚNIOR, W. de C. Utilização de redes neurais artificiais na classificação de níveis de degradação em pastagens. **Revista Brasileira de Engenharia Agrícola e Ambiental**, Campina Grande, v. 13, n. 3, p. 319-327, 2009.
- COSTA, F. H. da S. **Respostas morfo genéticas de bananeira submetida à poliploidização**. 2010. 134 p. Tese (Doutorado em Fitotecnia) - Universidade Federal de Lavras, Lavras, 2010.
- LOURO, A. H. F.; MENDONÇA, M. M.; GONZAGA, A. Classificação de tomates utilizando redes neurais artificiais. In: WORKSHOP DE VISÃO COMPUTACIONAL, 2., 2006, São Carlos. **Anais...** São Carlos: UFSCar, 2006. p. 239-244.
- PLOTZE, R. O. **Identificação de espécies vegetais através da análise da forma interna de órgãos foliares**. 2004. 152 p. Dissertação (Mestrado em Ciência da Computação e Matemática Computacional) - Universidade de São Paulo, São Carlos, 2004.
- RUMELHART, D. E.; HINTON, G. E.; WILLIAMS, R. J. Learning internal representations by error propagation. In: RUMELHART, D. E.; MCCLELLAND, J. L. (Ed.). **Parallel distributed processing: explorations in the microstructure of cognition**. Cambridge: MIT, 1986. p. 45-76.
- RUSSEL, S. J.; NORVIG, P. **Artificial intelligence: a modern approach**. Upper Saddle River: Prentice-Hall, 1995. 932 p.

SARI, N.; ABAK, K.; PITRAT, M. Comparison of ploidy level screening methods in watermelon: *Citrullus lanatus* (Thunb.). **Scientia Horticulturae**, Amsterdam, v. 82, n. 3/4, p. 265-277, 1999.

SCRIPPS RESEARCH INSTITUTE. **WinMDI version 2.8**. Disponível em:
<<http://facs.scripps.edu/software.html>>. Acesso em: 10 jan. 2011.

SHAHIN, M. A.; SYMONS, S. J.; POYSA, V. W. Determining soybean seed size uniformity with image analysis. **Biosystems Engineering**, Saint Albans, v. 94, n. 2, p. 191-198, 2006.

SILVA, S. de O. et al. Bananeira. In: BRUCKNER, C. H. (Ed.). **Melhoramento de fruteiras tropicais**. Viçosa, MG: UFV, 2002. p. 101-157.

SILVA, S. DE O.; ALVES, E. J.; CORDEIRO, Z. J. M.; MATOS, A. P.; JESUS, S, C. Variabilidade genética e melhoramento da bananeira. In: QUEIRÓZ, M. A. de; GOEDERT, C. O.; RAMOS, S. R. R. (Ed.). **Recursos genéticos e melhoramento de plantas para o nordeste brasileiro**. Petrolina: EMBRAPA Semi-Árido; Brasília: EMBRAPA Recursos Genéticos e Biotecnologia, 1999. Disponível em:
<<http://www.cpatas.embrapa.br>>. Acesso em: 28 jun. 2011.

SOUZA, F. F.; QUEIRÓZ, M. A. Avaliação de caracteres morfológicos úteis na identificação de plantas poliploides de melancia. **Horticultura Brasileira**, Brasília, v. 22, n. 3, p. 516-520, maio/jun. 2004.

VILLA, V. B. **Análise citomorfoanatômica e eletroférica de variedades de *Solanum tuberosum* L. X. (*Solanum Tuberosum* L. X *Solanum chacoense* Bitt)**. 1995. 76 f. Dissertação (Mestrado em Genética e Melhoramento de Plantas) - Universidade Federal de Lavras, Lavras, 1995.