

## Potencialidades das redes neurais artificiais na avaliação de recursos genéticos em bancos de germoplasma



**Maria da Cruz Chaves Lima Moura**

*Eng. Agr. pela U.E. do Maranhão (1988), Me. (Horticultura) pela UNESP (1994); Dr. pela UFV (2003); Pós Doc. em melhoramento genético vegetal pela U.F. Norte Fluminense Darcy Ribeiro (2006). Atualmente: U.F. do Maranhão, Curso Agronomia, Chapadinha/MA-Brasil*



**Alcinei Místico Azevedo**

*Eng. Agr. pela U.F. Vales do Jequitinhonha e Mucuri (2010), Me. UFLA (2012), Dr. UFV (2015). Atualmente na Universidade Federal de Minas Gerais, Instituto de Ciências Agrárias, Montes Claros - MG, Brasil, com Estatística Experimental.*



**Derly José Henriques da Silva**

*Eng. Agr. pela UFV (1985), Me. pela UFV (1992), Dr. pela USP (1999), Pós Doc. Universidade da Flórida, Gainesville, USA. Atualmente na Universidade Federal de Viçosa, MG – Brasil, com Genética e Melhoramento de Plantas.*



**Cosme Damião Cruz**

*Eng. Agr. pela UFV (1980), Me. pela UFV (1984), e Dr. pela USP (1990), atualmente pela Universidade Federal de Viçosa, MG, Brasil, com Genética Quantitativa.*

### Resumo

As reservas genéticas estão ameaçadas devido alterações da paisagem, no Clima e ausência de medidas de conservação. Em função disso, vários Bancos de Germoplasma foram criados no Mundo. É necessário conhecer dados dos genótipos armazenados nestes Bancos de germoplasma, processa-los e gerar informações úteis para tomada decisões e reduzir os gastos com os recursos financeiros, técnicos e humanos na condução das pesquisas. Recentemente, surgiram técnicas computacionais para acelerar a caracterização dos Recursos Genéticos, e conseqüentemente, seu uso têm aumentado. Logo, a adoção das técnicas analíticas, com abordagem estocástica ou inteligência computacional, torna-se indispensável. Neste contexto, o uso de redes neurais artificiais é um instrumento de grande potencial, pois ajudam a vencer esse grande desafio, principalmente acelerando a adoção rotineira dessas metodologias no meio acadêmico. Difusão do conhecimento sobre princípios da inteligência

computacional, publicações de artigos sobre aplicações nos diferentes campos da pesquisa em recursos genéticos e o desenvolvimento de aplicativos computacional de fácil utilização são valiosas contribuições para a ciência. E atualmente, é campo aberto para atuação de grupos de pesquisa e para a formação de núcleos de competência.

Palavras-chave: Biometria, inteligência computacional, biodiversidade.

### **Abstract**

The gene pools are threatened as a result of landscape change and Climate and the absence of conservation measures. As a result, several Genebank networks were created in the world. It needs to know this data, process them and generate useful information for making important decisions in the search for optimal solutions in view of the reduced amount of financial, technical and human resources in the conduct of research. Recently, there was a computational technique to accelerate the characterization of genetic resources and consequently its use. However, the adoption of analytical techniques with stochastic approach or computational intelligence, it is essential and in this context, the use of artificial neural networks is a potentially great tool to be able to help overcome this great challenge, especially accelerating the routine adoption of these methodologies in academia. Dissemination of knowledge about principles of computational intelligence, published papers on applications in different fields of research in genetic resources and the development of computational easy to use applications are valuable contributions to science and is currently open field for action research groups and the formation of competence centers.

Keywords: Biometrics, computational intelligence, biodiversity.

### **Bancos de Germoplasma**

As reservas genéticas estão ameaçadas em resultado de alterações da paisagem e do Clima e à ausência de medidas de conservação. Foi com essa preocupação e pensamento visionário do cientista russo Nikolai Ivanovik Vavilov, que criou-se o primeiro Banco de Germoplasma (BAG), na década 20. Seu sonho era acabar com a fome (garantir a segurança alimentar) no seu país e no mundo. O seu maior legado para a humanidade foi mapear sete Centros de diversidades das espécies cultivadas no mundo, em duas décadas, totalizando 5 continentes e 52 países, formando uma das maiores coleções de plantas do mundo, com 300 mil exemplares, armazenados em São Petersburgo.

Dando continuidade ao pensamento de Vavilov, no Brasil, foram criados vários BAG's, em diferentes épocas, sendo o primeiro em 1930, o BAG do Instituto Agrônomo de Campinas (IAC-SP) e o segundo, na década de 60, o Banco Germoplasma de Hortaliças (BGH) na Universidade Federal de Viçosa-MG, atualmente com 7.800 acessos armazenados, no entanto, só foram caracterizados em torno de 20% da sua totalidade da coleção.

Na década de 70, foi criada a Embrapa-CENARGEN que atualmente gerencia o quarto maior BAG do mundo com capacidade de armazenar 250 mil amostras de sementes, porém já constam 110 mil amostras armazenadas.

Muitas metodologias têm evoluído ao longo dos anos, como técnicas de genotipagem, técnicas de análise química, técnicas estatísticas, etc. Porém, o processo de fenotipagem, principalmente relacionados à caracteres agromorfológicos dos recursos genéticos não têm evoluído, e ainda são os mesmos realizados desde o início da experimentação agrícola. Dentre os principais empecilhos dos métodos de fenotipagem utilizados atualmente está a necessidade de intensa mão de obra e tempo para as avaliações. O que torna limitado o número possível de germoplasma a serem estudadas. Conseqüentemente, avanços devem ser feitos para tornar possível a fenotipagem com alta eficiência.

A adoção das técnicas analíticas, com abordagem estocástica ou inteligência computacional, torna-se indispensável e, neste contexto, o uso de redes neurais artificiais é um instrumento de grande potencial por ser capaz de ajudar a vencer esse grande desafio, principalmente acelerando a adoção rotineira dessas metodologias no meio acadêmico.

### **Redes Neurais Artificiais em Bancos de Germoplasma**

As Redes Neurais Artificiais (RNAs) surgiram como uma área da informática cujo objetivo básico era o de criar modelos artificiais do cérebro humano, de forma a permitir que computadores "pensem". Desta forma, uma RNA é, basicamente, um modelo não paramétrico do cérebro humano que reconhece padrões e regularidades nos dados que lhe são apresentados. Elas podem aprender por experiência e fazer generalizações com base em seu conhecimento passado.

A associação das redes neurais artificiais (RNAs) com métodos tradicionais é uma estratégia promissora, com as vantagens de ser não-paramétrico, requererem pequenas amostras para treinamento e tolerar a perda de dados. Em função de suas características naturais e do paralelismo interno inerente à sua arquitetura, as RNAs são capazes de solucionar problemas de grande complexidade. Conseqüentemente, apresentam possibilidade de melhor desempenho quando comparadas com os modelos estatísticos convencionais e conseqüentemente vem sendo usada nas mais variadas áreas do conhecimento.

Existem dezenas de diferentes modelos de RNA descritos na literatura, tais como o perceptron multicamadas (Multi Layer Perceptron - MLP), Redes de função de Base Radial (RBF), Redes de Função Sample (SFNN), Redes de Fourier, Redes Wavelet, redes auto-organizáveis de Kohonen, dentre outras. As propriedades e capacidade que tornam RNAs potencialmente úteis são: Não-linearidade: um neurônio artificial pode utilizar funções lineares ou não-lineares; Adaptabilidade: capacidade da RNA adaptar suas estruturas a partir de modificações do meio ambiente; Mapeamento de Entrada-

Saída: com base em exemplos de entrada e saída a RNA é capaz de se adaptar para minimizar o erro de mapeamento; Resposta a Evidências: capacidade de retornar não somente a resposta a uma entrada, mas também a crença ou confiança nesta decisão tomada; E, informação Textual: o conhecimento que a RNA possui é representado pela própria estrutura e estado da rede.

### **Aplicações das redes neurais artificiais na fenotipagem de germoplasma**

A capacidade das RNAs em modelar sistemas de natureza complexa, faz com que esta técnica tenha grande potencial de aplicação na fenotipagem de germoplasma. Com vantagens de automatizar processos, que feitos manualmente necessitariam de muito tempo. Neste sentido, a utilização de fotografias e RNAs tornam-se de imensa funcionalidade.

Obtendo-se um banco de imagens por fotografias retiradas com distância e iluminação uniforme, é possível por meio de redes neurais encontrarem associações existentes entre imagens digitalizadas e características mensuradas manualmente, como altura de plantas, diâmetro do caule, número de frutos, uniformidade na maturação, coloração de frutos, formato de frutos, coloração de folhas, arquitetura de planta, etc. Estas associações entre imagens e características agromorfológicas são encontradas durante o treinamento das RNAs, e armazenadas nos chamados pesos sinápticos. A partir destes pesos sinápticos, é possível fenotipar com alta eficiência várias características simultaneamente por meio de imagens. Em outras palavras, após o treinamento de RNAs, todos os vários descritores necessários para caracterizar um determinado acesso, podem ser obtidos de forma simultânea e rápida por meio de fotografias digitalizadas e redes neurais artificiais. A utilização desta estratégia possibilita uma redução muito grande de mão de obra, tempo e custo, viabilizando a avaliação de um número muito maior de acessos.

Outra grande dificuldade no processo de fenotipagem realizada de forma manual é a avaliação da resistência das plantas à pragas e doenças. Nestas situações, geralmente, a resistência é estimada baseando-se no grau de severidade, avaliada por escalas de notas. A avaliação efetuada por escala de notas requer avaliadores altamente treinados, os quais precisam avaliar individualmente cada planta. O grande tempo necessário para essa avaliação, aliado ao desgaste físico do avaliador, faz com que as avaliações tenham baixa acurácia, o que contribui para baixa herdabilidade. E, além disso, limita o número de plantas a serem avaliadas. Considerando o menor tempo para a fenotipagem por redes neurais e imparcialidade na obtenção dos resultados, torna-se mais uma vez viável esta estratégia. E além de possibilitar a avaliação de um número muito maior de plantas, utilizando menos mão de obra e tempo, têm-se ainda medições mais acuradas, o que proporciona maior herdabilidade.

No mesmo sentido, as RNAs tem grande potencial também na fenotipagem de distúrbios fisiológicos decorrentes de estresses abióticos, como déficit hídrico, déficit nutricional, alta temperatura, acidez no solo, etc. Possibilitando também a avaliação de um maior número de plantas, a necessidade de menos mão de obra, menos custo e

a obtenção de resultados imparciais, e conseqüentemente de maior acurácia.

### **Aplicações das redes neurais artificiais na obtenção de informações importantes para o manejo de germoplasma**

Na organização em bancos de germoplasma, uma grande dificuldade que se tem é a classificação das plantas quanto à sua espécie, o que geralmente requer conhecimentos aprofundados de taxonomia. Muitas pesquisas têm sido conduzidas mostrando que é possível diferenciar espécies bastante próximas por meio de imagens e redes neurais artificiais, obtendo-se alta eficiência. Isso é uma estratégia muito importante, possibilitando a organização de bancos de germoplasma, os quais geralmente armazenam um grande número de espécies, dispensando conhecimentos aprofundados em taxonomia.

Outra informação de grande importância em bancos de germoplasma é o estudo de dissimilaridade, o que geralmente é feito por análise multivariada utilizando dados agromorfológicos, moleculares, ou ambos. Porém, outra estratégia possível é o estudo da dissimilaridade por RNAs do tipo mapas auto-organizáveis de Kohonen (Self Organizing Maps - SOM). Neste tipo de RNA a própria rede se auto-organiza em relação às particularidades existentes entre as características avaliadas nos acessos, identificando subconjuntos (clusters) que contenham similaridade. Outra técnica de análise multivariada utilizada frequentemente em bancos de germoplasma é a análise discriminante, o que também pode ser feito por RNAs, principalmente do tipo MLP ou RBF.

Outra aplicação das RNAs que vêm sendo estudadas é para a predição de valores genéticos. Tais estratégias tem como principal objetivo possibilitar a seleção de indivíduos realmente superiores, os quais podem ser direcionados à programas de melhoramento.

A aplicação das RNAs na seleção indireta é também uma estratégia viável. A seleção indireta é recomendada quanto a característica de principal interesse é de difícil mensuração ou de baixa herdabilidade. Neste caso, tradicionalmente é utilizada uma característica correlacionada com a de interesse, e que seja de fácil mensuração. Porém, esta estratégia também pode ser substituída pelo uso de RNAs. A utilização de RNAs para este fim apresenta vantagens, pois neste caso, a variável de interesse pode ser predita por várias outras características simultaneamente, e por permitir ajuste não linear, proporciona maior precisão na seleção indireta. Neste sentido, muitos trabalhos com RNAs têm sido desenvolvidos, mostrando eficiência na predição do teor de carotenoides por dados colorimétricos, área foliar pelas dimensões das folhas, predição características importantes por dados obtidos pelo NIR (espectrômetros de infravermelho próximo), etc.

Na seleção genômica ampla, o uso de RNAs tem mostrado também desempenho superior aos encontrados utilizando estatísticas baseadas em análise bayesiana. Neste caso, as RNAs além de possibilitar maior acurácia na seleção genômica apresenta a

vantagem de não exigir conhecimentos estatísticos tão aprofundados, como é necessário ao se utiliza técnicas de estatística bayesiana. A possibilidade de se trabalhar com a seleção genômica ampla pode ser de grande aplicação em bancos de germoplasma, pois a partir de dados moleculares obtidos em plântulas é possível prever o desempenho agrônomo da planta. Isto permite a seleção precoce, e conseqüentemente, economia de tempo e mão de obra. Esta estratégia, embora possa ser aplicada em plantas de ciclo curto, é mais vantajosa em plantas perenes.

### **Perspectivas para o uso de redes neurais artificiais em bancos de germoplasma**

Atualmente, um dos principais softwares utilizados no estudo das redes neurais é o MATLAB da empresa Mathworks. Outra boa opção é o software R (R Development Core Team), por meio dos pacotes “AMORE”, “brnn”, “monmlp”, “neuralnet”, “RSNNS”, “nnet”, etc. Estes dois softwares permitem a utilização de vários tipos de RNAs para os mais variados objetivos. No entanto, a necessidade do conhecimento de algoritmos de programação na utilização destes softwares vem sendo uma das principais justificativas para seu pouco uso. Porém, recentemente, os usos de RNAs vêm sendo facilitado por softwares como o GENES, o qual dispensa o conhecimento de algoritmos de programação. Além disso, o assunto é cada vez mais abordado em cursos de curta duração e disciplinas de pós-graduação. Conseqüentemente, em curto prazo, as técnicas de RNAs serão bastante utilizadas, o que trará grandes vantagens para os bancos de germoplasma.

### **Conclusões**

Os princípios teóricos desenvolvidos por inúmeros pesquisadores têm permitido o planejamento e a execução mais eficiente de programas que tenham como objetivo a conservação e utilização racional de recursos genéticos em bancos de germoplasma. Para adoção de ações mais apropriadas é necessário a realização de experimentos fidedignos, nos quais são obtidos grande volume de informações. É a partir do processamento adequado destes dados, que os parâmetros populacionais são estimados e os fenômenos biológicos são modelados e interpretados. Nesta etapa de análise e interpretação de resultados, o uso da abordagem baseada em inteligência computacional se mostra promissora. Difusão do conhecimento sobre princípios da inteligência computacional, publicações de artigos sobre aplicações nos diferentes campos da pesquisa em recursos genéticos e o desenvolvimento de aplicativos computacional de fácil utilização são valiosas contribuições para a ciência. E atualmente, é campo aberto para atuação de grupos de pesquisa e para a formação de núcleos de competência.