



ECOLOGIA E O FUTURO DA BIOSFERA

III CLAE

IX CEB

10 a 17 de setembro de 2009



[www.seb-ecologia.org.br](http://www.seb-ecologia.org.br)



## UTILIZAÇÃO DE REDES NEURAIS ARTIFICIAIS PARA CLASSIFICAÇÃO POR ORIGEM DE CASTANHAS DO BRASIL

R. G. de Andrade<sup>1</sup>

V. O. Roda.<sup>2</sup>; J. D. C. Pessoa<sup>3</sup>

1 - Universidade de São Paulo, Escola de Engenharia, Departamento Engenharia Elétrica, Avenida Trabalhador São - carlense nº 400, 13566 - 590, São Carlos-SP, Brasil. Telefone: 55 16 21072809 - raphael@cnpdia.embrapa.br.

2 - Embrapa Instrumentação Agropecuária, Laboratório de Instrumentação e Pós Colheita, Rua XV de Novembro nº 1452, 13560 - 970, São Carlos-SP, Brasil.

### INTRODUÇÃO

A castanha do Brasil (*Bertholletia excelsa* H.B.K.) é a semente da castanheira, uma árvore da família *Lecythidaceae*, nativa da Floresta Amazônica. Ela é uma das maiores e mais tradicionais fontes de renda da região norte do país. Sua produção está distribuída entre os estados do Pará, Amazonas, Acre, Mato Grosso, Rondônia e Amapá, sendo os três primeiros responsáveis pela maior parte de toda produção [4] [5]. Essa produção faz do Brasil o maior exportador de castanhas não processadas. Nos últimos anos foram exportadas em média de 10 mil toneladas de castanhas por ano (10 milhões de dólares em média) [1].

A pós - colheita das castanhas do Brasil tem duas fases: a exploratória e a de beneficiamento. Nas duas fases não há uma classificação prévia para controle das sementes como, por exemplo, sua origem. A falta de controle sanitário e rastreabilidade estão fazendo com que os países europeus e os Estados Unidos imponham cada vez mais barreiras na importação das castanhas [6].

A rastreabilidade pode ser definida como uma forma de manter as informações referentes à origem e ao destino dos produtos. Assim a qualidade do produto tem maior credibilidade e confiabilidade junto aos consumidores. Com objetivo de classificar as castanhas conforme sua origem, este trabalho utilizou conceitos de redes neurais artificiais (RNA). Conforme [2], as RNAs são modelos matemáticos baseados no funcionamento dos neurônios biológicos e são muito utilizadas para classificação e reconhecimento de padrões em diversas áreas de conhecimento.

O trabalho de Simões, 2003 [7] utilizou a arquitetura de RNA conhecida por *Multi Layer Perceptron* (MLP) para classificação das laranjas. Os dados foram adquiridos a partir de imagens digitalizadas no espaço de cor RGB e foram aplicadas na camada de entrada (um neurônio para R, G e B). Também foi utilizado uma camada intermediária de 20 neurônios e para a camada de saída foram definidos 5 padrões visuais que variavam do verde escuro até o laranja

escuro (também um neurônio para cada padrão

Baseado nessas informações, este trabalho utilizou na RNA três espaços de cores para a distinção das castanhas: XYZ, RGB e  $L^*a^*b^*$  (definidos pela *Commission Internationale de l'Éclairage*).

### OBJETIVOS

Identificar as castanhas do Brasil de diferentes regiões utilizando dados de cores e redes neurais artificiais.

### MATERIAL E MÉTODOS

Foram conseguidas amostras de três regiões diferentes: Acre, Amapá e Rondônia. As castanhas da região do Acre foram colhidas no Seringal Filipinas (Reserva Extrativista Chico Mendes na cidade de Epitaciolândia), as do Amapá na Reserva Extrativista do rio Cajari (município de Laranjal do Jari) e as de Rondônia foram compradas na região de Machadinho do Oeste.

Com um colorímetro de marca Konica Minolta, modelo CR - 400, foram adquiridos dados relativos ao espaço de cor XYZ com iluminante C dessas amostras. Os dados adquiridos foram convertidos para o espaço RGB e o  $L^*a^*b^*$  de acordo com as matrizes de conversão encontradas em trabalho de Pascale, 2006 [3]. Os valores adquiridos e os convertidos foram utilizados separadamente na camada de entrada da RNA que foi modelada utilizando o software livre Weka 3.6.0.

A RNA utilizada foi a MLP e sua arquitetura é composta de 3 neurônios na camada de entrada, 6 neurônios na camada intermediária e mais 3 neurônios na camada de saída. Na camada de entrada foram inseridos os valores dos espaços de cores, um para cada neurônio. Já na camada de saída, os neurônios eram uma para cada estado de origem.

Essa MLP foi definida após vários testes variando o espaço de cor na entrada da RNA utilizando *cross - validation* com

10 *folds*. Então, a RNA foi configurada com uma taxa de aprendizagem de 0.7, fator de momentum de 0.2 e 650 épocas.

Após a configuração, os dados foram divididos em 2 arquivos para entrada no Weka. Um dos arquivos foi utilizado para treinamento (70% dos dados) e o outro para teste (30% dos dados) da RNA. Esses dois arquivos foram criados utilizando um total de 150 amostras e posteriormente um total de 300 amostras.

Todos os testes foram feitos utilizando um computador com processador Intel Core 2 Duo 2GHz e 2Gbytes de memória ram DDR2.

## RESULTADOS

Os primeiros testes foram feitos utilizando 150 amostras no espaço de cor XYZ e apresentaram valores muito baixos (65% de acerto com os dados de teste). Isso ocorreu devido à proximidade dos valores no espaço de cores das amostras. Por causa disso, foram testados mais dois espaços de cor (RGB e  $L^*a^*b^*$ ) com o dobro da quantidade de castanhas. Assim, nos teste de *cross - validation* os resultados foram bem parecidos com 75% de acerto para os espaços de cores XYZ e RGB e 78% de acerto para o espaço de cor  $L^*a^*b^*$ . Para os dados de treinamento esses resultados também foram semelhantes, com acerto de 77% para XYZ e 80% para RGB e  $L^*a^*b^*$ . Já para aplicação da RNA com os dados de teste, a variação de acerto foi maior com 74% para XYZ, 78% para RGB e 81% para  $L^*a^*b^*$ .

Os resultados foram superiores aos testes feitos com 150 amostras principalmente nos testes de *cross - validation* e com os dados de teste. Isso ocorreu justamente por conta do aumento do espaço amostral, mesmo que o aumento não tenha acontecido com os dados de treinamento. Outro ponto importante é que o erro quadrático médio (EQM) se manteve em 0.3 para todos os testes.

Esses dados nos mostram a diferença de resultado ao mudar o espaço de cor para entrada da RNA.

Assim, os dados convertidos a partir do espaço XYZ tiveram melhores resultados. Esse resultado varia de acordo com as definições da sua rede, porém neste caso com espaço de cor  $L^*a^*b^*$  foram obtidos os maiores resultados.

Outro ponto importante é que as castanhas que deveriam ser reconhecidas como do Acre ou do Amapá, em todos os testes foram facilmente confundidas com as castanhas de Rondônia pela RNA. No teste de *cross - validation* utilizando espaço de cor RGB com total de 300 amostras (100 para cada estado de origem), a quantidade de castanhas identificadas corretamente foi mais baixo para as amostras de Rondônia com 68 acertos.

Pode - se observar também que 5 amostras do Acre foram identificadas como amostras do Amapá e que com 4 amostras aconteceu o contrário. Isso representa 3% do total de amostras que é quase 4 vezes menos do que o total de amostras do Acre e Amapá que foram confundidas com

sendo as de Rondônia. Porém os maiores erros aconteceram entre amostras do Amapá e Rondônia (21 e 17). Isso aconteceu por terem cores bem próximas o que influenciou bastante nos resultados da rede.

## CONCLUSÃO

O objetivo deste trabalho foi distinguir as castanhas do Brasil conforme seu estado de origem. Isso foi feito utilizando dados de três espaços de cores na entrada de uma arquitetura MLP, que mostrou melhor desempenho com o espaço de cor  $L^*a^*b^*$ , classificando 81% das amostras corretamente. Também se pode observar a diferença de resultados entre os espaços de cores.

Outro ponto importante foi à dificuldade da RNA em distinguir castanhas de Rondônia das outras. Apesar disso, pode - se concluir que é possível iniciar um processo de rastreamento pelo menos na questão da origem das castanhas. Isso porque esses resultados podem ser melhorados testando outros tipos de RNA ou até mesmo utilizando outras técnicas de classificação.

(Os autores agradecem aos colegas do projeto Kamukaia, em especial a Dra. Lúcia Helena de Oliveira Wadt (Embrapa Acre), pela disponibilização das amostras e discussão dos resultados. Agradecem à Embrapa pelo apoio financeiro através do projeto Kamukaia (sobre manejo de produtos florestais não madeireiros na Amazônia).

Agradecem também ao Ms. Ednaldo José Ferreira (Embrapa Instrumentação Agropecuária) por importante ajuda intelectual.)

## REFERÊNCIAS

- [1] BRASIL. Ministério do Desenvolvimento Indústria e Comércio Exterior Secretaria de Comércio Exterior. Disponível em: <<http://aliceweb.desenvolvimento.gov.br/alice.asp>>. Acessado em: 27 Abr. 2009
- [2] Kovács, L. Z. Redes Neurais Artificiais: Fundamentos e Aplicações. 2ª Edição. Collegium Cognition, 1996.
- [3] Pascale, D. RGB Coordinates of the Macbeth ColorChecker. The BabelColor Company, 2006.
- [4] Pennacchio, H. L. Castanha do Brasil, período até 28/04/2006. Disponível em: <[http://www.conab.gov.br/conabweb/\\_a\\_28042006\\_castanha\\_do\\_brasil.pdf](http://www.conab.gov.br/conabweb/_a_28042006_castanha_do_brasil.pdf)>. Acesso em: 29 Mar. 2009.
- [5] Pennacchio, H. L. Castanha do Brasil, Proposta de Preço Mínimo. Disponível em: <[http://www.conab.gov.br/conabweb/download/precos\\_minimos/proposta\\_de\\_precos\\_minimos\\_safra\\_2006\\_07\\_castanha\\_do\\_brasil.pdf](http://www.conab.gov.br/conabweb/download/precos_minimos/proposta_de_precos_minimos_safra_2006_07_castanha_do_brasil.pdf)>. Acesso em: 29 Mar. 2009.
- [6] Silva, F. A. Aplicação de Microondas no Processo de Beneficiamento da Castanha do Brasil (*Bertholletia excelsa*). Dissertação. Unicamp: Campinas - SP, 2002.
- [7] Simões, A. S.; Costa, A. H. R. Classificação de Laranjas Baseadas em Padrões Visuais. São Paulo - SP, Abril, 2003