

Utilização de Redes Neurais Artificiais na classificação de danos em sementes de girassol

Antonio M. Magalhães Junior^{1†}, Paula R. Santos², Thelma Sáfadi³

¹Mestrando em Estatística e Experimentação Agropecuária, UFLA.

²Mestranda em Estatística e Experimentação Agropecuária, UFLA.

E-mail: pauillasant_s@hotmail.com.

³Departamento de Estatística, DES/UFLA. E-mail: safadi@ufla.br.

Resumo: *Redes Neurais Artificiais (RNA's) são reconhecidamente eficazes para a utilização em problemas de reconhecimento de padrões e classificação de dados. Este trabalho teve como objetivo a classificação de imagens radiografadas de sementes de girassol quanto ao seu nível de dano. Dessa maneira, as sementes de girassol foram radiografadas e as imagens geradas foram categorizadas em sementes cheias, com danos leves ou deformadas. Essas imagens foram redimensionadas buscando uniformizar suas dimensões e diminuir a quantidade de entradas para a RNA. Após isso, a equalização das imagens foi realizada, objetivando-se a melhora do contraste e assim acentuar as imperfeições presentes nas sementes. Uma RNA de topologia feed-forward e uma camada escondida foi utilizada e o conjunto de imagens foi dividido de forma aleatória, reservando parte das imagens para validação e teste da RNA treinada. Após mil treinos com cada configuração, o índice médio de acerto da RNA foi de aproximadamente 74,5% utilizando as três classes, de 77,4% para sementes cheias contra sementes com danos leves, 96,0% para sementes cheias contra sementes deformadas e 86,5% para sementes com danos leves contra sementes deformadas. Dessa forma, a abordagem utilizada pode ser aplicada para a classificação de sementes de forma automatizada, uma vez que obteve bons índices de acerto.*

Palavras-chave: análise de raio-X; análise de sementes; reconhecimento de padrões.

Abstract: *Artificial Neural Networks (ANN's) are recognized to be effective for use in pattern recognition and data classification problems. This work was aimed to classify X-ray images of sunflower seeds as to their level of damage. Therefore, the sunflower seeds were radiographed and the generated images were categorized into full seeds, seeds with slight damages or deformed seeds. These images were resized in order to standardize their dimensions and decrease the number of entries to the ANN. After this, the equalization of the images was performed, aiming at improving the contrast and thus accentuating the imperfections present in the seeds. A feed-forward topology ANN and a hidden layer was used and the set of images was randomly divided, reserving part of the images for validation and testing of the trained ANN. After a thousand training sessions with each configuration, the average accuracy of the neural network was 74.5% using the three classes, from 77.4% for full seeds versus seeds with slight damage, 96,0% for full seeds versus deformed seeds and 86.5% for seeds with slight damage versus deformed seeds. Therefore, the approach used can be applied to the classification of seeds in an automated way, since it obtained good indexes of correctness.*

Keywords: X-ray analysis; seed analysis; pattern recognition.

[†]Autor correspondente: jrjpng@hotmail.com.

Introdução

O mercado brasileiro de sementes é o terceiro maior do mundo, atrás apenas de Estados Unidos e China. A economia deste setor movimentou bilhões de reais, como na safra 2015/16 em que movimentou aproximadamente R\$ 10 bilhões (ABRASEM, 2015). Esta movimentação financeira poderá ser ainda maior nos próximos anos, uma vez que se estima um aumento na produção nacional da safra 2018/19, se comparada à anterior, devido ao crescimento da área plantada (CONAB, 2019). A produção brasileira de sementes se concentra na soja e milho, que juntos correspondem a 74% deste mercado. Entretanto, vem se destacando a produção de outros tipos de sementes como os de forrageiras e de olerícolas, que já correspondem a 11% e 6%, respectivamente, do mercado total de sementes (ABRASEM, 2015). Uma cultura olerícola que tem ganhado espaço no agronegócio brasileiro é a de girassol (*Helianthus annuus* L.), a qual apresenta boas propriedades como alto teor de óleo nos grãos e ótima qualidade de subprodutos. Além disso, possui facilidade de adaptação a diferentes condições climáticas, com boa semeadura e colheita (GRUNVALD et al., 2009).

A análise de sementes é de grande importância para se determinar a qualidade das mesmas. Testes para avaliar o potencial físico e fisiológico são realizados em laboratórios, a fim de determinar o potencial de germinação e outras características das sementes. Uma das técnicas mais utilizadas na análise de sementes é o raio-X, que avalia a qualidade das sementes por meio de imagens radiografadas. Este é um método que propicia avaliação rápida e não destrutiva das sementes, com a análise sendo realizada de forma visual e permitindo a diferenciação das sementes com boa formação e as que sofreram algum dano (BRASIL, 2009). Técnicas computacionais de processamento de imagens podem ser aplicadas para auxiliar na análise das sementes, visando a automatização do processamento das imagens geradas e redução da subjetividade deste tipo de análise.

Uma das técnicas computacionais que podem ser utilizadas é a aplicação Redes Neurais Artificiais (RNA's), que são reconhecidamente eficazes para os problemas de reconhecimento e classificação de dados (FAUSETT, 1994; RIPLEY, 1994; MEHTANI, 2011). Dessa forma, esse estudo buscou realizar a classificação das imagens radiografadas das sementes de girassol, com base em seu nível de dano, por meio da utilização de Redes Neurais Artificiais.

Metodologia

Obtenção das imagens

As imagens radiografadas das sementes de girassol, cultivar Hélio-250, foram obtidas pelo Laboratório de Análise de Sementes do Departamento de Agricultura da Universidade Federal de Lavras. As sementes passaram pelo processo de radiografia com intensidade de 22kV e exposição média de 11 segundos em um aparelho de raios-X Faxitron MX20, resultando em imagens digitalizadas que foram arquivadas em formato jpeg. As imagens passaram por análise visual por parte de um profissional da área e em seguida as sementes foram classificadas como sementes cheias, sementes levemente danificadas ou sementes deformadas. Como sementes cheias foram classificadas aquelas com embrião opaco na radiografia, levemente danificadas as com danos, porém com preservação do eixo embrionário e por fim as deformadas, com danos que afetam o eixo embrionário. Ao término

da classificação, o grupo de sementes cheias contou com 175 imagens, o de sementes com dano leve com 130 imagens e o de sementes deformadas com 140 imagens.

Processamento das imagens

A fim de uniformizar suas dimensões, as imagens foram redimensionadas para 250 x 130 pixels. Visando aumentar o contraste das imagens e facilitar a identificação dos danos, as imagens passaram por uma técnica de equalização chamada Equalização de Histograma Adaptativo de Contraste Limitado (CLAHE, em inglês). Esta técnica dividiu a imagem em pequenas seções chamadas *tiles* e equalizou os histogramas dessas seções para que o histograma da região de saída correspondesse aproximadamente a um histograma especificado, de modelo uniforme. Ao término da equalização das regiões, as seções vizinhas foram combinadas com a utilização de interpolação bilinear, buscando eliminar limites induzidos de forma artificial. A Figura 1 mostra uma imagem de cada classe e seus respectivos histogramas antes e depois do processo de equalização.

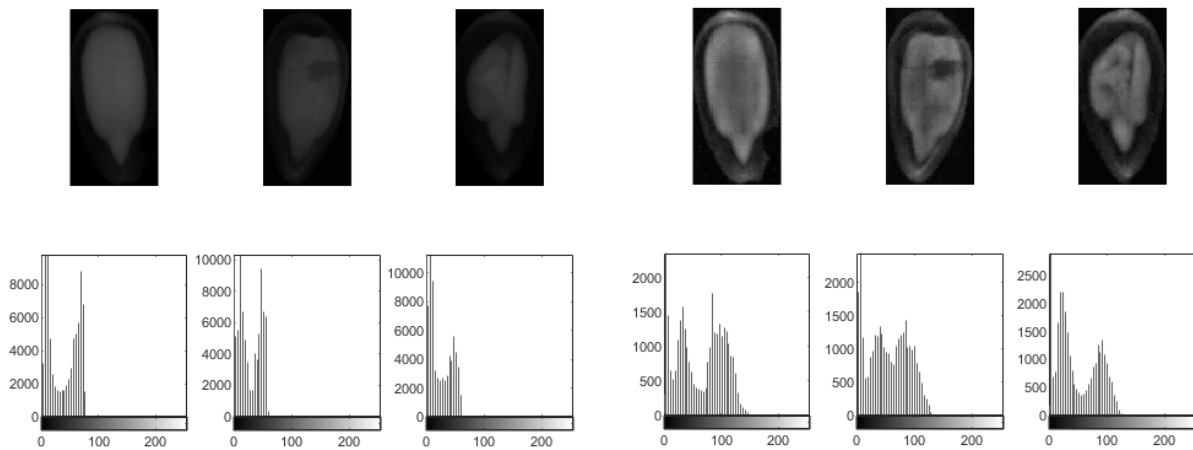


Figura 1: Exemplos de cada classe (cheia, com dano leve e deformada) e seus respectivos histogramas: a) antes da equalização e b) após a equalização.

Após o processo de equalização, as imagens foram novamente redimensionadas, dessa vez para 40% de suas dimensões, visando diminuir a quantidade de entradas para rede neural, uma vez que cada pixel configura um dado de entrada na rede. A matriz com os valores dos pixels foi então transformada em um vetor, onde cada linha da mesma foi concatenada com a linha anterior. Repetiu-se o processo para cada uma das imagens, obtendo enfim uma única matriz de 5200 x 445, onde 5200 é o número de parâmetros (pixels) ou entradas para a rede e 445 é a quantidade total de imagens.

Arquitetura da rede neural

A rede neural utilizada possui topologia do tipo *feed-forward*, com uma camada escondida utilizando função de ativação do tipo sigmoide e com função de ativação *softmax* na camada de saída. Foi definido empiricamente um número de 10 neurônios na camada

escondida. A rede neural recebeu treinamento supervisionado com função de retropropagação de gradiente conjugado em escala para a atualização dos pesos. Dessa forma, a configuração final da rede conta com a camada de entrada dos dados com 5200 neurônios, um neurônio para cada pixel, com a camada escondida contendo 10 neurônios e com a camada de saída com 3 neurônios para classificação utilizando as três classes em conjunto e com 2 neurônios para classificações utilizando as combinações de duas classes (Figura 2). Foram comparadas, quanto aos níveis médios de acerto, as classificações utilizando todas as classes em conjunto e também de duas em duas classes.

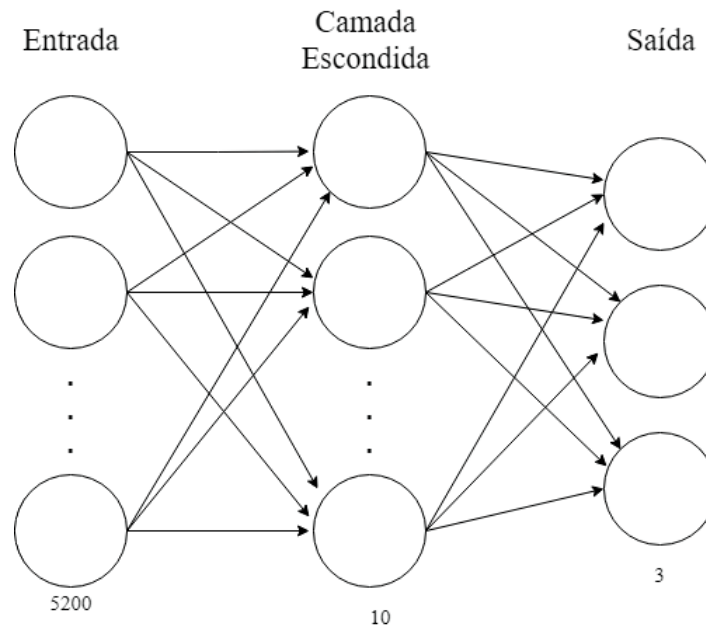


Figura 2: Arquitetura da rede neural artificial empregada na classificação das sementes utilizando as três classes.

Treinamento da rede neural

O treinamento da RNA foi realizado de forma supervisionada, onde foi fornecida para a rede a classificação esperada de parte do conjunto de imagens. No momento do treinamento, o conjunto de imagens foi dividido entre dados de treinamento (70%), validação (15%) e teste (15%). Os dados de treinamento foram apresentados à rede no treinamento para esta se ajuste de acordo com seu erro. Os de validação foram utilizados para medir o nível de generalização da rede e finalizar o treinamento quando este nível parasse de aumentar. Por fim, os dados de teste foram utilizados para medir o desempenho da rede, de forma independente ao treinamento.

Resultados e discussões

Os resultados indicam que a maior dificuldade reside na correta classificação entre as sementes cheias e as sementes com danos leves. Isso provavelmente ocorre em razão da pouca diferença visual entre essas duas classes. A rede neural foi treinada mil vezes para

cada um dos casos e o índice médio de acerto, quando consideradas as três classes em conjunto, foi de 74,4876% (Tabela 1). Já para as classificações utilizando pares de classes, os índices de acerto cresceram significativamente, com a rede neural atingindo nível médio de acerto de 77,4265% para sementes cheias contra sementes com danos leves (Tabela 2), de 95,9968% para sementes cheias contra sementes deformadas (Tabela 3) e de 86,4633% para sementes com danos leves contra sementes deformadas (Tabela 4).

Tabela 1: Matriz de confusão média com mil execuções da rede neural para a classificação de sementes cheias, com danos leves e deformadas.

		Classificação real			Taxa de acerto
		Cheias	Danos leves	Deformadas	
Classificado como	Cheias	153,0720	54,2250	7,5760	71,2386%
	Danos leves	18,6650	61,1290	15,1550	64,3809%
	Deformadas	3,2630	14,6460	117,2690	86,7515%
Taxa de acerto		87,4697%	47,0223%	83,7636%	74,4876%

Fonte: elaborada pelo autor

Tabela 2: Matriz de confusão média com mil execuções da rede neural para a classificação de sementes cheias e com danos leves.

		Classificação real		Taxa de acerto
		Cheias	Danos leves	
Classificado como	Cheias	153,1790	47,0280	76,5103%
	Danos leves	21,8210	82,9720	79,1770%
Taxa de acerto		87,5308%	63,8246%	77,4265%

Fonte: elaborada pelo autor

Tabela 3: Matriz de confusão média com mil execuções da rede neural para a classificação de sementes cheias e deformadas.

		Classificação real		Taxa de acerto
		Cheias	Deformadas	
Classificado como	Cheias	171,2150	8,8250	95,0983%
	Deformadas	3,7850	131,1750	97,1954%
Taxa de acerto		97,8371%	93,6964%	95,9968%

Fonte: elaborada pelo autor

Tabela 4: Matriz de confusão média com mil execuções da rede neural para a classificação de sementes com danos leves e deformadas.

		Classificação real		Taxa de acerto
		Danos leves	Deformadas	
Classificado como	Danos leves	114,8550	21,4060	84,2904%
	Deformadas	15,1450	118,5960	88,6758%
Taxa de acerto		88,3500%	84,7114%	86,4633%

Fonte: elaborada pelo autor

Os resultados são comparáveis a outros trabalhos onde os autores também realizaram a classificação de sementes de girassol, porém com a utilização de outras abordagens. Em

Sigmae, Alfenas, v.8, n,2, p. 569-575, 2019.

64^a Reunião da Região Brasileira da Sociedade Internacional de Biometria (RBRAS).

18^o Simpósio de Estatística Aplicada à Experimentação Agronômica (SEAGRO).

Sáfadi et al. (2016), foi utilizada a abordagem de espectro de ondaletas e Support Vector Machine (SVM) para a classificação de sementes e obtiveram um índice de acerto médio de 57% utilizando todas as classes e de 82% utilizando apenas as classes de sementes cheias e deformadas. Leite, Sáfadi e Carvalho (2013) optaram pela abordagem de análise de componentes independentes e análise discriminante e obtiveram índice de acerto de 82%. Enquanto Cassiano (2018) fez uso da abordagem de Expoente Direcional de Hurst e obteve índice de acerto de 74,5% quando consideradas todas as classes em conjunto e de 99,76% quando consideradas apenas as classes de sementes cheias e deformadas.

Conclusões

A utilização de Redes Neurais Artificiais para classificação de imagens radiografadas de sementes se mostrou como uma alternativa eficiente e de relativa facilidade de aplicação, uma vez que apresentou índices de acerto satisfatórios na classificação das sementes, demandando pouco pré-processamento das imagens de raio-X. Dessa forma, essa técnica pode ser aplicada para a classificação das sementes, automatizando o processo de classificação e reduzindo a subjetividade inerente ao se analisar sementes sem o auxílio de ferramentas computacionais.

Agradecimentos

Os autores agradecem o apoio financeiro da FAPEMIG e CAPES.

Referências bibliográficas

ASSOCIAÇÃO BRASILEIRA DE SEMENTES E MUDAS - ABRASEM. 2015. URL www.abrasem.com.br/wp-content/uploads/2013/09/Anuario_ABRASEM_2015.pdf.

BRASIL. Ministério da Agricultura, Pecuária e Abastecimento. Regras para análise de sementes. Brasília, 2009. 399p.

CASSIANO, F.R. *Expoente direcional de Hurst na análise de similaridade de imagens de sementes*. Tese (Doutorado em Estatística e Experimentação Agropecuária) - Universidade Federal de Lavras, Lavras, 2018. 50p.

COMPANHIA NACIONAL DE ABASTECIMENTO- CONAB. 2019. URL www.conab.gov.br/info-agro/safras/graos/boletim-da-safra-de-graos.

FAUSETT L. *Fundamentals of Neural Networks Architectures, Algorithms and Applications*. New Jersey: Prentice Hall. 1994. 476p.

GRUNVALD, A.K.; CARVALHO; C.G.P. DE; OLIVEIRA, A.C.B.; ANDRADE, C.A.B. Adaptabilidade e estabilidade de genótipos de girassol nos estados do Rio Grande do Sul e Paraná. *Ciência e Agrotecnologia (online)*, v.33, n.5, p.1195-1204, 2009.

Sigmae, Alfenas, v.8, n.2, p. 569-575, 2019.

64^a Reunião da Região Brasileira da Sociedade Internacional de Biometria (RBRAS).
18^o Simpósio de Estatística Aplicada à Experimentação Agronômica (SEAGRO).

LEITE, I.C.C.; SÁFADI, T.; CARVALHO, M.L. M. Evaluation of seed radiographic images by independent component analysis and discriminant analysis, *Seed Science and Technology*, v.41, 235-244, 2013.

MEHTANI, P.P. *Classification using Artificial Neural Networks*. Teses B. Tech, apresentado no departamento de ciência da computação e engenharia de Rourkela, Rourkela, p.1-40, 2011.

RIPLEY, B.D. Neural Networks and Related Methods for Classification. *Journal of the Royal Statistical Society. Series B (Methodological)*, v.56, n.3, p.409-456, 1994.

SÁFADI, T.; VIDA KOVIC, B.; KANG, M.; LEITE, I.C.C. Wavelet-based spectral for descriptors of detection of damage in sunflower seeds. *International Journal of Wavelets, Multiresolution and Information Processing*, London, v.14, n.4, p.235-244, 2016.

Sigmae, Alfenas, v.8, n,2, p. 569-575, 2019.

64^a Reunião da Região Brasileira da Sociedade Internacional de Biometria (RBRAS).
18^o Simpósio de Estatística Aplicada à Experimentação Agronômica (SEAGRO).