



CLASSIFICAÇÃO DE IMAGENS DE SEMENTES DE SOJA COLORIDAS PELO TETRAZÓLIO BASEADA EM APRENDIZADO PROFUNDO

SOUZA JUNIOR, M.¹; FRANÇA-NETO, J.B.²; SAITO, P.T.M¹; BUGATTI, P.H.¹

¹Universidade Tecnológica Federal do Paraná - UTFPR, Campus Cornélio Procópio, Cornélio Procópio, PR, marcelojunior.2015@alunos.utfpr.edu.br; ²Embrapa Soja.

O Brasil é um dos maiores exportadores de soja do mundo, com uma produção de 114.095,8 mil toneladas, em uma área plantada de 33.909,4 mil hectares, correspondendo cerca de 55,68% da área cultivada no país, ocupando a segunda posição como maior produtor de soja do mundo (CONAB, 2017).

Para uma maior produtividade da cultura da soja, a utilização de sementes de alta qualidade é um fator extremamente importante, pois com o uso das mesmas, possibilitam o estabelecimento de uma população adequada de plantas, que são mais produtivas. A qualidade da semente pode variar em função de alguns fatores, como danos mecânicos, deterioração por umidade, ataques de percevejos, entre outros (Moreano et al., 2013). Mas, para que o sucesso seja alcançado, o controle de qualidade deve ser rápido e preciso, fornecendo resultados corretos e de forma eficiente (França-Neto et al., 1998).

Com a relativa lentidão na obtenção dos resultados do teste de germinação, o processo de tomada de decisões na indústria de sementes se torna limitado. Além disso, na sua forma tradicional de avaliação, este teste não fornece informações sobre o vigor da semente, e não permite a identificação dos fatores que afetam a sua qualidade.

Entre os testes para a avaliação da qualidade fisiológica de sementes, o teste de tetrazólio destaca-se pelo grande número de informações que fornece sobre os lotes de sementes, visando diagnosticar as causas responsáveis pela redução da qualidade das mesmas (França-Neto et al., 1998).

Apesar do teste ser relativamente simples e não utilizar equipamentos e reagentes caros, para a execução do mesmo é necessário que o analista de sementes seja bem treinado nas técnicas do teste, possuindo conhecimento das estruturas anatômicas da semente, entre outras. Dessa maneira, o conhecimento para a realização do mesmo não é trivial.

A partir de tais premissas, fica clara a necessidade da proposta e desenvolvimento de metodologias automáticas para realização da determinação da viabilidade e do vigor de sementes de soja pelo teste de tetrazólio, por meio de análise de imagens. O presente trabalho tem como objetivo principal apresentar uma metodologia para avaliação automática desses parâmetros em sementes de soja, aplicando técnicas de visão computacional agregadas ao aprendizado profundo.

O termo aprendizado profundo é utilizado para salientar o problema de treinar redes neurais artificiais que desempenham o aprendizado de características de forma hierárquica, onde as características dos níveis mais altos da hierarquia são formadas pela combinação das características de baixo nível (Schmidhuber, 2015). O aprendizado profundo nada mais é que uma técnica de aprendizado de máquina, onde muitos estágios de processamento de informações são explorados para a classificação de padrões e para a aprendizagem de recurso ou representação (Wan et al., 2014).

A aprendizagem profunda utiliza um grafo ponderado em camadas para modelar o problema de reconhecimento. Em cada camada, um conjunto de características abstratas é conectado às características nas camadas superiores e/ou inferiores adjacentes, onde o peso é associado com cada aresta que conecta os recursos (Chen et al., 2015).



Segundo Wan et al. (2014) o aprendizado profundo é encontrado em diversas áreas de pesquisa como: redes neurais, otimização e reconhecimento de padrões, processamento de sinais, entre outras.

Uma das abordagens relacionadas à aprendizagem profunda são as redes neurais por convolução (*convolutional neural networks - CNNs*). Essas redes possuem como base a ideia de identificar campos perceptivos locais (*local receptive fields*), realizar o compartilhamento de pesos (*weight sharing*), bem como aplicar operações de convolução e subamostragem (*pooling*) (Bezerra, 2016).

Com a finalidade de contribuir com o desenvolvimento de metodologias automáticas para a realização da definição do vigor de sementes de soja, neste trabalho foram realizados experimentos utilizando redes neurais por convolução, para a classificação automática de uma base de imagens de sementes de soja, submetidas ao teste de tetrazólio. A base de imagens foi gerada na Embrapa Soja com 188 imagens de 94 amostras de sementes de soja separadas em suas determinadas classes de vigor (Tabela 1), de acordo com seus respectivos tipos de danos. Na base original, cada amostra possui duas imagens (porção externa e interna) para cada semente.

Para os experimentos iniciais deste trabalho, a partir da base original, foi criada uma nova base de imagens (Tabela 2), em que as imagens de cada classe foram separadas de acordo com os lados (direito e esquerdo) e as porções (externa e interna) da semente. A classe M (dano mecânico) foi excluída, por conter somente uma amostra, tornando inviável a classificação e validação da mesma. A nova base de imagens contém um total de 372 imagens divididas em seis classes.

Após a geração desta nova base, a mesma foi dividida em 80% para etapa de treinamento e 20% para a etapa de validação. Posteriormente, as amostras foram submetidas a uma rede neural por convolução (CNN), onde foi realizada a classificação supervisionada das amostras.

O resultado preliminar da classificação desta base de imagens apresentou uma acurácia de 41%. No entanto, para melhorar tais resultados, faz-se necessária a aquisição de novas imagens de amostras de sementes de soja submetidas ao teste de tetrazólio, com os mesmos tipos de danos e também outros danos e suas determinadas classificações para cada classe. Portanto, como trabalho futuro, será realizada a criação de uma nova base de imagens, em conjunto com a Embrapa Soja, cujos técnicos atuarão no processo de classificação das novas amostras sementes de soja por imagem. Além disso, serão investigadas estratégias de Data Augmentation (Okafor et al., 2017; Shijie et al., 2017), com o intuito de produzir novas amostras (sintéticas), em diferentes classes, de maneira que contribua com o treinamento da rede neural por convolução, que necessita de uma grande quantidade de dados.

Referências

- BEZERRA, E. Introdução à aprendizagem profunda. In: Simpósio Brasileiro de Banco de Dados, 31., Salvador, 2016. **Minicursos...** Salvador: Sociedade Brasileira de Computação, 2016. p.1-30.
- CHEN, C. R.; LEE, G. G.; XIA, Y.; LIN, W. S.; SUZUMURA, T.; LIN, C. Efficient multi-training framework of image deep learning on gpu cluster. In: 2015 IEEE INTERNATIONAL SYMPOSIUM ON MULTIMEDIA (ISM), [s.l.], Institute of Electrical and Electronics Engineers (IEEE), 2015. p. 489-494. Disponível em: <<http://dx.doi.org/10.1109/ism.2015.119>>.
- CONAB. **Acompanhamento da safra brasileira: grãos**, v.4, safra 2016/17 - décimo segundo levantamento. Brasília, DF: CONAB, 2017. 158 p. Disponível em: <<http://www.conab.gov.br>>.
- FRANÇA-NETO, J. B.; KRZYZANOWSKI, F. C.; COSTA, N. P. **O teste de tetrazólio em sementes de soja**. Londrina: EMBRAPA-CNPSo, 1998. 72p. (Documentos, 116).



MOREANO, T. B.; BRACCINI, A. L.; SCAPIM, C. A.; FRANÇA-NETO, J. B.; KRZYZANOWSKI, F. C.; MARQUES, O. J. Physical and physiological qualities of soybean seed as affected by processing and handling. **Journal of Seed Science**, v.35, n.4, p.466-477, 2013.

OKAFOR, E.; SMIT, R.; SCHOMAKER, L.; WIERING, M. Operational data augmentation in classifying single aerial images of animals. In: CONFERENCE ON INNOVATIONS IN INTELLIGENT SYSTEMS AND APPLICATIONS (INISTA), 2017. p. 354–360.

SCHMIDHUBER, J. Deep learning in neural networks: An overview. **Neural Networks**, v.61, p.85–117, 2015.

SHIJIE, J.; PING, W.; PEIYI, J.; SIPING, H. Research on data augmentation for image classification based on convolution neural networks. In: 2017 CHINESE AUTOMATION CONGRESS (CAC), 2017. p.4165–4170.

WAN, J.; WANG, D; HOI, S. C. H.; WU, P.; ZHU, J.; ZHANG, Y.; LI, J. Deep learning for content-based image retrieval: a comprehensive study. **Proceedings of the ACM International Conference on Multimedia - MM '14**, p. 157–166, 2014.

Tabela 1. Distribuição de amostras de sementes de soja por classe de vigor, classificadas pelo teste de tetrazólio, de acordo com o tipo de dano anotado.

Classe	Tipo de Dano	Total de Amostras	Total Imagens
M	Mecânico	1	2
MPU	Mecânico, Percevejo, Umidade	9	18
MU	Mecânico, Umidade	35	70
P	Percevejo	3	6
PU	Percevejo, Umidade	27	54
SD	Sem Danos	3	6
U	Umidade	16	32
Total		94	188

Tabela 2. Distribuição de amostras de sementes de soja por classe de vigor, classificadas pelo teste de tetrazólio, de acordo com o tipo de dano: base de imagens nova.

Classe	Tipo de Dano	Total de Amostras	Total Imagens
MPU	Mecânico, Percevejo, Umidade	9	36
MU	Mecânico, Umidade	35	140
P	Percevejo	3	12
PU	Percevejo, Umidade	27	108
SD	Sem Danos	3	12
U	Umidade	16	64
Total		93	372