

## **MACHINE LEARNING PARA CLASSIFICAÇÃO DE LOTES DE SEMENTES DE SOJA**

**ASCOLI, C. A.<sup>1</sup>, GADOTTI, G. I.<sup>2</sup>, PINHEIRO, R. M.<sup>3</sup>, BERNARDY, R. <sup>4</sup>, MONTEIRO, R. C<sup>5</sup>**

<sup>1</sup> Engenheira Agrícola e Ambiental, mestre em Agronomia, aluna de especialização, Faculdade de Agronomia Eliseu Maciel Universidade Federal de Pelotas, Capão do Leão, RS, [ascolicarla@gmail.com](mailto:ascolicarla@gmail.com)

<sup>2</sup> Engenheira Agrícola, doutora em Ciência e Tecnologia de Sementes- Centro de Engenharias/Faculdade de Agronomia Eliseu Maciel, Universidade Federal de Pelotas, Capão do Leão, RS. [gizele.gadotti@ufpel.edu.br](mailto:gizele.gadotti@ufpel.edu.br)

<sup>3</sup> Engenheiro Agrônomo, doutorando em Ciência e Tecnologia de Sementes – Faculdade de Agronomia Eliseu Maciel, Universidade Federal de Pelotas, Capão do Leão, RS. [romario.ufacpz@hotmail.com](mailto:romario.ufacpz@hotmail.com)

<sup>4</sup> Engenheiro Agrícola, mestrando em Ciências Ambientais, Centro de Engenharias, Universidade Federal de Pelotas, Pelotas, RS. [ruanbernardy@yahoo.com.br](mailto:ruanbernardy@yahoo.com.br)

<sup>5</sup> Engenheira Agrícola, mestranda em Ciência e Tecnologia de Sementes- Centro de Engenharias/Faculdade de Agronomia Eliseu Maciel, Universidade Federal de Pelotas, Capão do Leão, RS. [ritamonteiro@gmail.com](mailto:ritamonteiro@gmail.com)

Apresentado no

L Congresso Brasileiro de Engenharia Agrícola - CONBEA 2021

08 a 10 de novembro de 2021 - Congresso On-line

**RESUMO:** A análise de testes que determinam a qualidade de sementes gera um grande número de informações que se torna quase impossível, à capacidade intelectual humana, realizar uma análise rápida e eficaz dentro de um laboratório de controle de qualidade, a curto prazo. O objetivo foi avaliar as diferenças na qualidade fisiológica de sementes de soja em diferentes cultivares através de técnica de aprendizado de máquina de modo a ranquear os lotes conforme a qualidade. Utilizou os dados de germinação, envelhecimento acelerado, tetrazólio, emergência de plântulas e peso de mil sementes de 93 lotes, com três cultivares. A avaliação dos lotes foi realizada em duas fases, uma após a colheita e outra com seis meses de armazenamento. Foram analisados os classificadores *Random Forest*, *Multi Layer Perceptron* (MLP), J48 e *Classification Via Regression* (CVR), auxiliados pela técnica *Feature Resampler*. Verificou-se que o *Random Forest* e o CVR obtiveram maior acurácia nos dados, sendo o primeiro capaz de classificar melhor os lotes rejeitados, com menos falsos positivos. Concluímos que os algoritmos, são ferramentas importantes na inteligência artificial, com maiores valores de acurácia e precisão possibilitam a classificação dos lotes de sementes de soja por técnica de aprendizado de máquina.

**PALAVRAS-CHAVE:** inteligência artificial, agricultura, controle de qualidade

## **MACHINE LEARNING FOR SOYBEAN SEED LOT CLASSIFICATION**

**ABSTRACT:** The analysis of tests that determine the quality of seeds generates a large amount of information that makes it almost impossible for the human intellectual capacity to carry out a quick and effective analysis within a quality control laboratory in the short term. The objective was to evaluate the differences in the physiological quality of soybean seeds in different cultivars through machine learning techniques to rank the lots according to quality. It was analyzed germination, accelerated aging, tetrazolium, seedling emergence, and weight of a thousand seeds test from 93 lots with three cultivars. The evaluation of the lots was carried out in two phases, one after harvest and another with six months of storage. Random Forest, Multi Layer Perceptron (MLP), J48, and Classification Via Regression (CVR) classifiers were analyzed, aided by the Feature Resampler technique. It was found that Random Forest and

CVR obtained greater accuracy in the data, the first being able to better classify the rejected batches, with fewer false positives. We concluded that it is possible to classify soybean seed lots with great accuracy and precision through artificial intelligence and machine learning techniques.

**KEYWORDS:** artificial intelligence, agriculture, quality control

**INTRODUÇÃO:** Analisar todos os testes que determinam a qualidade de sementes gera um tanto de informações que se torna quase impossível, a capacidade intelectual humana proceder uma análise rápida e eficaz dentro de um laboratório de controle de qualidade, a curto prazo (PINHEIRO et al., 2021). Portanto, resultados errôneos podem implicar em prejuízos econômicos para as empresas sementeiras. Para tornar esta fonte útil, é necessário avaliar os dados gerados nos testes de controle de qualidade dos lotes das sementes. Assim, se faz adequado o uso da metodologia de aprendizado de máquina para a otimização do tempo e os recursos gastos em testes laboratoriais repetitivos. Desta forma, o objetivo deste trabalho foi investigar o desempenho de algoritmos de classificação por meio de técnica de aprendizado de máquinas baseado em diferenças de qualidade fisiológica das sementes de soja em várias cultivares, de modo a ranquear os lotes conforme a sua qualidade fisiológica.

**MATERIAL E MÉTODOS:** O trabalho foi conduzido no Laboratório Interno de Análises de Sementes (LAIS) de uma empresa localizada na cidade de Sinop – MT e foram utilizados 36 lotes de sementes de soja da cultivar A e 22 lotes de sementes da cultivar B e 7 lotes de sementes da cultivar C, cedidas pela empresa, se tratando de genética própria e com indicação para o cultivo no Estado do Mato Grosso, sendo produzidas na safra 18/19. As sementes foram avaliadas em duas etapas. Na primeira fase, logo após a colheita e a segunda com seis meses de armazenamento. A condição de armazenamento foi em ambiente refrigerado mantido de forma constante a 13 °C e 60% de umidade relativa do ar (condições de armazenamento praticadas pela empresa). Os testes realizados foram: germinação, envelhecimento acelerado (EA), tetrazólio, emergência de plântulas e peso de mil sementes. Os dados gerados para base de treinamento por aprendizado de máquina supervisionado foram 93 linhas com 42 atributos, onde 54 lotes foram aceitos para comercialização, 27 rejeitados e 12 chamados intermediários. Os lotes intermediários não apresentaram características de alto vigor e nem de baixo vigor, considerando que podem ser utilizados em determinado momento. Para as técnicas de mineração de dados foram utilizados quatro classificadores: J48, Random Forest, CVR e MLP. Procedimento inicial foi validação cruzada, onde se dividiu o conjunto de dados, treinamento e teste com 10 subconjuntos. Para verificar quais algoritmos seriam mais precisos utilizou-se acurácia e matriz de confusão. Com os resultados obtidos determinou-se a melhor técnica de aprendizagem. Os dados dos testes também foram submetidos a procedimentos estatísticos de comparação de média através de análise de variância - ANOVA e constatando-se diferença estatística, as médias foram comparadas pelo Teste de Tukey, a 5% de significância.

**RESULTADOS E DISCUSSÃO:** O conjunto de treinamento para cada modelo compreendeu 81,7% de acurácia dos dados para *Random Forest* e *Classification Via Regression* e em seguida 79,6% para J48 e *Multi Layer Perceptron* com 74,2%, sendo estes últimos os que apresentaram método com desempenho inferior. Após o treinamento, o algoritmo *Random Forest* gerou 92,6% de recall e 90,9% de precisão média em relação ao conjunto de dados de teste real para a classe de aceitos. E para a classe de rejeitados mostrou 92,6% de recall e 73,59% de precisão. Na classe de intermediários mostrou 8,3% de recall e

25% de precisão. A metodologia proposta foi adequada, pois as precisões encontradas na discriminação de sementes em suas diferentes classes de lotes aceitos foram altas, variando de 0,85 a 0,90 de precisão entre os modelos de aprendizado de máquina (Tabela 1 e 2).

Tabela 1. Matriz de confusão do classificador *Random Forest* de lotes de sementes de soja.

	a	b	c	
<b>50</b>	3	1		a=aceito
0	<b>25</b>	2		b=rejeitado
5	6	<b>1</b>		c=intermediário

Tabela 2. Matriz de confusão do classificador CVR lotes de sementes de soja.

	a	b	c	
<b>51</b>	2	1		a=aceito
3	<b>23</b>	1		b=rejeitado
6	4	<b>2</b>		c=intermediário

Esses resultados indicam alto potencial de aplicação prática desses métodos e podem ser facilmente interpretados para análise de lotes de sementes em outras culturas. No entanto, ainda é indispensável o esforço humano, pois alguns testes manuais são necessários para atender as normativas (BOELT et al., 2018). Em estudo com soja realizado a partir de imagens, de Medeiros et al. (2020) observaram valores para o conjunto de dados com precisão de 93% destacando seu bom desempenho. Em estudos com conjuntos de dados sintéticos para fenotipagem de sementes, Toda et al. (2020) mostrou 96% de recall e 95% de precisão média, com rede neural, em relação ao conjunto de dados de teste do dado real. Neste mesmo trabalho de Moraes (2020) observou que a árvore decisória gerada naquele estudo pelo J48 foi baseada no atributo de vigor. Naquele trabalho havia somente um teste de vigor utilizado, desta forma sugeriu-se investigar outros testes que determinam o vigor de sementes em trabalhos futuros. Nos dados deste estudo foi considerado vários testes de vigor, realizados em dois períodos (inicial e seis meses de armazenamento). De acordo com Tillmann et al. (2019), é de fundamental importância testes de vigor para uma classificação mais eficiente de lotes. Em comparação com o trabalho de Moraes (2020), este estudo apresentou mais resultados de acurácia para as classes “rejeitados” e “intermediários”, o que se deve ao fato de obtermos mais atributos baseados em vigor.

**CONCLUSÕES:** Os algoritmos com melhor desempenho foram *Random Forest* e *Classification Via Regression*, apresentando os maiores valores de acurácia e precisão e assim eles podem ser inseridos como ferramenta de inteligência artificial para classificação dos lotes

de sementes, até mesmo em outras culturas. Esses resultados indicam alto potencial de aplicação prática desses métodos no setor sementeiro.

**AGRADECIMENTOS:** O presente trabalho foi realizado com apoio da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Código de Financiamento 001. CNPq e FAPERGS pela concessão de bolsas.

## **REFERÊNCIAS:**

BOELT, B., SHRESTHA, S., SALIMI, Z., JØRGENSEN, J., NICOLAISEN, M., CARSTENSEN, J. Multispectral imaging - A new tool in seed quality assessment? **Seed Science Research**, v. 28, n. 3, p. 222-228, 2018. doi: 10.1017/S0960258518000235.

DE MEDEIROS, AD, CAPOBIANGO, NP, DA SILVA, JM., DA SILVA, LJ, SILVA, CB, DIAS, DCFS. *Interactive machine learning for soybean seed and seedling quality classification*. **Scientific Reports**, v. 10, 11267, 2020. <https://doi.org/10.1038/s41598-020-68273-y>.

MORAES, N.A.B. **Predição de ranqueamento de lotes de sementes de milho por inteligência artificial**. 29 f. 2020. Dissertação (Mestrado). Programa de Pós-graduação em Ciência e Tecnologia de Sementes, Faculdade de Agronomia Eliseu Maciel, Universidade Federal de Pelotas. 2020.

PINHEIRO, R. M.; GADOTTI, G. I.; MONTEIRO, R. C. M.; BERNARDY, R. Inteligência artificial na agricultura com aplicabilidade no setor sementeiro. **Diversitas Journal**, 2021.

TILMANN, M.A.A; TUNES, L.M. de; ALMEIDA, A. da S. Análise de Sementes. In: PESKE, S.T.; VILLELA, F. A.; MENEGHELLO, G.E. **Sementes: Fundamentos Científicos e Tecnológicos**. 4a. Edição 2019. P.147-257.

TODA, Y., OKURA, F., ITO, J., OKADA, S. KINISHITA, T. TSUJI, H, SAISHO, D. Training instance segmentation neural network with synthetic datasets for crop seed phenotyping. **Communications Biology**, v.3, n.173, 2020. Doi: <https://doi.org/10.1038/s42003-020-0905-5>.