

**ANDRÉ DANTAS DE MEDEIROS**

**APLICAÇÕES AVANÇADAS DE APRENDIZADO DE MÁQUINA E  
FERRAMENTAS DE ANÁLISE DE IMAGENS PARA CLASSIFICAÇÃO E  
FENOTIPAGEM DE SEMENTES**

Tese apresentada à Universidade Federal de Viçosa, como parte das exigências do Programa de Pós-Graduação em Fitotecnia, para obtenção do título de *Doctor Scientiae*.

Orientador: Laércio Junio da Silva

**VIÇOSA - MINAS GERAIS  
2023**

**Ficha catalográfica elaborada pela Biblioteca Central da Universidade  
Federal de Viçosa - Campus Viçosa**

T

M488a  
2023  
Medeiros, André Dantas de, 1994-  
Aplicações avançadas de aprendizado de máquina e  
ferramentas de análise de imagem para classificação e  
fenotipagem de sementes / André Dantas de Medeiros. – Viçosa,  
MG, 2023.

1 tese eletrônica (77 f.): il. (algumas color.).

Inclui anexo.

Orientador: Laércio Junio da Silva.

Tese (doutorado) - Universidade Federal de Viçosa,  
Departamento de Agronomia, 2023.

Inclui bibliografia.

DOI: <https://doi.org/10.47328/ufvbbt.2024.038>

Modo de acesso: World Wide Web.

1. Soja - Semente - Qualidade. 2. Sementes - Qualidade -  
Análise. 3. Sementes - Fisiologia. 4. Sementes - Radiografia.  
5. Imagens multiespectrais. 6. Raios X. 7. ImageJ (Programa de  
computador). 8. Aprendizado do computador. I. Silva, Laércio  
Junio da, 1984-. II. Universidade Federal de Viçosa.  
Departamento de Agronomia. Programa de Pós-Graduação em  
Fitotecnia. III. Título.

CDD 22. ed. 633.3421

**ANDRÉ DANTAS DE MEDEIROS**

**APLICAÇÕES AVANÇADAS DE APRENDIZADO DE MÁQUINA E  
FERRAMENTAS DE ANÁLISE DE IMAGENS PARA CLASSIFICAÇÃO E  
FENOTIPAGEM DE SEMENTES**

Tese apresentada à Universidade Federal de Viçosa, como parte das exigências do Programa de Pós-Graduação em Fitotecnia, para obtenção do título de *Doctor Scientiae*.

Orientador: Laércio Junio da Silva

APROVADA: 17 de outubro de 2023


Assentimento:

Documento assinado digitalmente  
 **ANDRE DANTAS DE MEDEIROS**  
Data: 07/02/2024 13:01:58-0300  
Verifique em <https://validar.iti.gov.br>

---

André Dantas de Medeiros

Autor

Documento assinado digitalmente  
 **LAERCIO JUNIO DA SILVA**  
Data: 05/02/2024 13:40:06-0300  
Verifique em <https://validar.iti.gov.br>

---

Laércio Junio da Silva  
Orientador

A minha mãe Izaltina Dantas de Medeiros (*in  
memoriam*). Seu amor, ensinamentos e valores  
são a base de quem me tornei.

*Dedico.*

## AGRADECIMENTOS

Agradeço ao meu Deus, por toda proteção e força concedida em todos os momentos, principalmente nas horas de dificuldade. Te agradeço meu Deus por se fazer presente em meu coração por onde quer que eu vá.

À minha família, em especial a Izaltina Dantas de Medeiros, Francisco Pereira de Medeiros e Francielio Dantas de Medeiros, pelo incentivo, carinho e apoio.

À minha noiva Letícia Tiezi Rombola, pelo incentivo e amor.

À Universidade Federal de Viçosa e ao Departamento de Fitotecnia, pela oportunidade de realização do curso. Ao Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq), pela concessão da bolsa.

Ao professor Laércio Junio de Silva, pela oportunidade, orientação, ensinamentos, apoio e amizade.

À professora Denise Cunha Fernandes dos Santos Dias, pelos ensinamentos, oportunidades concedidas e pelas significativas contribuições feitas a este trabalho.

Aos professores do Departamento de Fitotecnia. A todos os integrantes do GSem (Grupo de Estudos em Tecnologia de Sementes), pela ajuda na condução dos experimentos, companheirismo e amizade.

Aos meus colegas de curso, pelo companheirismo e por me acompanharem nessa jornada.

A todos que acreditam que a educação e a ciência podem mudar vidas.

**Muito obrigado!**

## **BIOGRAFIA**

André Dantas de Medeiros, filho de Francisco Pereira de Medeiros e Izaltina Dantas de Medeiros, nasceu no dia 05 de fevereiro de 1994, na cidade de Acari, Rio Grande do Norte, Brasil.

No ano de 2012 concluiu o ensino médio integrado com o curso de Técnico em Agricultura pela Escola Agrícola de Jundiá, atual Unidade Acadêmica Especializada em Ciências Agrárias, Macaíba, RN. Iniciou o curso de graduação em Engenharia Agrônoma em fevereiro de 2013, na Universidade Federal do Rio Grande do Norte, RN, concluindo em julho de 2017. Em agosto do mesmo ano, iniciou na Universidade Federal de Viçosa, MG, o Mestrado no Programa de Pós-graduação em Fitotecnia, na área de Tecnologia e Produção de Sementes, concluindo em fevereiro de 2019. Em março do mesmo ano, iniciou na Universidade Federal de Viçosa, MG, o Doutorado no Programa de Pós-graduação em Fitotecnia, na área de Tecnologia e Produção de Sementes.

## RESUMO

MEDEIROS, André Dantas de, D.Sc., Universidade Federal de Viçosa, outubro de 2023. **Aplicações avançadas de aprendizado de máquina e ferramentas de análise de imagem para classificação e fenotipagem de sementes.** Orientador: Laércio Junio da Silva.

As tecnologias ópticas modernas, complementadas por algoritmos de inteligência artificial, estão revolucionando a análise de amostras biológicas na agricultura, particularmente na avaliação da qualidade de sementes. Por meio da integração de técnicas sofisticadas de análise de imagens, incluindo imagens RGB, raios X e imageamento multiespectral, esta pesquisa introduziu novas abordagens e ferramentas inovadoras. Dentre as ferramentas estão o Ilastik, que proporciona a utilização do aprendizado de máquina interativo para classificação de sementes e plântulas de soja, e a IJCropSeed, macro projetada para permitir a análise de imagens de raios X de sementes de uma variedade de culturas agrícolas. Além disso, foram desenvolvidos e testados modelos de aprendizado de máquina, abrangendo tanto métodos interativos quanto tradicionais. Estes modelos demonstraram uma eficácia notável, alcançando níveis de precisão superiores a 90%, o que representa um marco significativo no campo. No que se refere ao uso do imageamento multiespectral, com ênfase em comprimentos de onda específicos, foi observada alta consistência nos modelos desenvolvidos para avaliar o potencial fisiológico dos lotes de semente, considerando diferentes genótipos, lotes e safras. Além disso, a pesquisa destacou correlações entre os aspectos físicos das sementes, seus componentes espectrais e o desempenho fisiológico subsequente. Estes avanços tecnológicos apresentam uma nova oportunidade de ganho eficiência e qualidade operacional na classificação do potencial fisiológico das sementes e a predição do vigor das plântulas. O uso dessas abordagens otimiza a análise, tornando-a rápida, objetiva e altamente eficaz, e destaca a correlação entre aspectos físicos, componentes espectrais e o desempenho fisiológico das sementes.

Palavras-chave: Imagens RGB. Imagens multiespectrais. Imagens de raios X. ImageJ. Aprendizado de máquina.

## ABSTRACT

MEDEIROS, André Dantas de, D.Sc., Universidade Federal de Viçosa, October, 2023. **Advanced machine learning applications and image analysis tools for seed classification and phenotyping.** Adviser: Laércio Junio da Silva.

Modern optical technologies, complemented by artificial intelligence algorithms, are revolutionizing the analysis of biological samples in agriculture, particularly in seed evaluation. Through the integration of sophisticated image analysis techniques, including RGB imaging, X-ray, and multispectral imaging, this research introduced new approaches and innovative tools. Among these tools are Ilastik, which enables the use of interactive machine learning for the classification of soybean seeds and seedlings, and IJCropSeed, a macro designed to facilitate the analysis of X-ray images of seeds from a variety of agricultural crops. Furthermore, machine learning models were extensively developed and tested, encompassing both interactive and traditional methods. These models demonstrated remarkable efficacy, achieving accuracy levels exceeding 90%, which represents a significant milestone in the field. In terms of the use of multispectral imaging, with an emphasis on specific wavelengths, high consistency was observed in the models developed to assess the physiological potential of seed lots, considering different genotypes, batches, and crops. Additionally, the research highlighted fundamental correlations between the physical aspects of the seeds, their spectral components, and subsequent physiological performance. These technological advances facilitate the classification of the physiological potential of seeds and the prediction of seedling vigor. The use of these approaches optimizes analysis, making it rapid, objective, and highly effective, and underscores the correlation between physical aspects, spectral components, and the physiological performance of seeds.

Keywords: RGB Images. Multispectral Images. X-ray Images. ImageJ. Machine Learning.

## SUMÁRIO

INTRODUÇÃO GERAL .....	10
REFERÊNCIAS .....	12
CAPÍTULO I - APRENDIZADO DE MÁQUINA INTERATIVO PARA CLASSIFICAÇÃO DA QUALIDADE DE SEMENTES E PLÂNTULAS DE SOJA .....	14
RESUMO .....	14
ABSTRACT .....	15
INTRODUÇÃO .....	16
MATERIAL E MÉTODOS .....	18
Material vegetal e aquisição de imagens .....	18
Pré-processamento e segmentação de imagens .....	18
Classificação do aspecto físico externo e qualidade fisiológica da semente .....	18
Classes do aspecto físico externo da semente .....	19
Classes de qualidade fisiológica .....	19
Classificação interativa do aspecto físico externo e vigor da semente .....	19
Algoritmos externos para classificação do aspecto físico externo das sementes e qualidade fisiológica .....	20
RESULTADOS .....	22
Modelos de aprendizado de máquina para classificação de sementes .....	24
Modelos de aprendizado de máquina para classificação de qualidade fisiológica .....	25
Resumo da Classificação .....	25
Relação entre aspecto físico externo da semente e crescimento da plântula .....	26
DISCUSSÃO .....	28
CONCLUSÕES .....	32
REFERÊNCIAS .....	33
CAPÍTULO II - IJCROPSEED: UMA FERRAMENTA DE ACESSO ABERTO PARA ANÁLISE DE ALTO RENDIMENTO DE RADIOGRAFIAS DE SEMENTES DE CULTURAS .....	36
RESUMO .....	36
INTRODUÇÃO .....	38
MATERIAL E MÉTODOS .....	40
Preparação da amostra e aquisição de imagens radiográficas .....	41
Visão geral do IJCropSeed .....	41
Validação de performance .....	45
Experimento com espécie modelo ( <i>C. abyssinica</i> ) .....	45
Aquisição e processamento das imagens .....	45
Análises fisiológicas .....	45
RESULTADOS .....	47

Processamento das imagens de raios X.....	47
Desempenho e validação da IJCropSeed .....	48
Experimento com sementes de <i>Crambe abyssinica</i> .....	49
Modelos de aprendizado de máquina para classificação de sementes.....	49
DISCUSSÃO .....	52
REFERÊNCIAS.....	56
<b>CAPÍTULO III - MARCADORES FENOTÍPICOS MULTIESPECTRAIS COMBINADOS COM ALGORITMOS DE APRENDIZADO DE MÁQUINA PARA CLASSIFICAÇÃO DE LOTES DE SEMENTES DE SOJA.....</b>	
RESUMO .....	59
INTRODUÇÃO .....	61
MATERIAL E MÉTODOS.....	62
Amostra .....	62
Imageamento multiespectral .....	62
Análises fisiológicas .....	63
Classes de sementes.....	63
Análise de dados e modelos de aprendizado de máquina .....	64
RESULTADOS.....	66
Análise fisiológica .....	66
Análise multivariada de componentes principais.....	68
Modelos de aprendizado de máquina.....	69
DISCUSSÃO .....	71
REFERÊNCIAS.....	74

## INTRODUÇÃO GERAL

A qualidade das sementes é um fator determinante para a produtividade e sucesso das culturas. As sementes carregam consigo o potencial genético da cultivar, e a sua qualidade reflete diretamente na segurança alimentar e na sustentabilidade agrícola (Ainsworth et al., 2012; Finch-Savage e Bassel, 2016).

A avaliação de sementes tem uma história de grandes descobertas, onde tradicionalmente, depende de técnicas manuais, frequentemente caracterizadas por sua subjetividade e susceptibilidade a variabilidades (ElMasry et al., 2020). No entanto, a tecnologia de sementes encontra-se em um ponto de inflexão, onde avanços tecnológicos promissores, especialmente na área de imagens, estão redefinindo os paradigmas de avaliação. A inspeção com o uso de raios X, especificamente a abordagem 2D, emergiu como uma técnica crucial, permitindo análises internas das sementes com grande precisão (Medeiros et al., 2020a). Esta técnica, ao revelar a estrutura interna da semente, cria um panorama mais claro da sua integridade física e potencial fisiológico (Dell'Aquila, 2009; Mahajan et al., 2018). Paralelamente, o imageamento multiespectral apresentou-se como uma poderosa ferramenta para análise de sementes. Ao analisar a interação das sementes com diversos comprimentos de onda de luz, é possível identificar variações sutis que podem ser correlacionadas com atributos de qualidade, indo desde a integridade física até a sanidade (Barboza da Silva et al., 2021; Batista et al., 2022; Olesen et al., 2015; Wilkes et al., 2016). Complementar ao espectro multiespectral, as imagens RGB – uma técnica amplamente acessível e versátil – também têm seu valor na avaliação de sementes. Embora menos sofisticada que o imageamento multiespectral, sua aplicação, quando integrada com técnicas de análise avançadas, pode fornecer insights significativos sobre a qualidade da semente (Baek et al., 2020; de Medeiros et al., 2020; Zhang et al., 2018).

Entretanto, o verdadeiro potencial destas técnicas de imagem é potencializado e aproveitado com eficiência quando combinado com aplicação de algoritmos avançados de aprendizado de máquina. Algoritmos como Redes Neurais, que têm mostrado notável eficiência no processamento e análise de imagens, estão sendo adaptados para decifrar características intrincadas e nuances nas imagens de sementes (Li et al., 2019). Além disso, métodos como Support Vector Machine (SVM) e Random Forest, reconhecidos por sua robustez em classificações e regressões, estão sendo empregados para correlacionar características visuais das sementes com seu potencial fisiológico e genético (Baek et al., 2019; Barboza da Silva et al., 2021). Estes algoritmos, ao identificarem padrões muitas vezes imperceptíveis à análise

humana, possibilitam um grau de precisão e detalhamento altamente acurado. Assim, a combinação da visão computacional com o aprendizado de máquina está revolucionando a tecnologia de sementes, permitindo avaliações mais acuradas e rápidas.

Além disso, novas ferramentas de processamento de imagens e visão computacional de livre acesso estão suportando aplicações feitas para classificação de sementes e plântulas de forma automática. Softwares como Ilastik e ImageJ são exemplos promissores desse tipo de abordagem que têm o poder de transformar imagens brutas em análises detalhadas e precisas, muitas vezes superando as limitações humanas de interpretação (Berg et al., 2019; Schneider et al., 2012). Neste contexto, o desenvolvimento de ferramentas analíticas robustas, simples de usar e práticas integradas a essas aplicações é fundamental e criam perspectivas para revolucionar a forma como percebemos e avaliamos a qualidade das sementes. Porém, apesar dos avanços significativos, a necessidade de abordagens mais abrangentes, que considerem diferentes genótipos e condições de cultivo, é fundamental na validação de novas técnicas e tecnologias.

Esta pesquisa, portanto, visa contribuir no aprofundamento do uso de técnicas de raios X, imagens multiespectrais e RGB, e na elaboração de ferramentas analíticas robustas orientada a qualidade de sementes. O objetivo é investigar e propor novas abordagens utilizando técnicas de imagens e modelos de inteligência artificial para a avaliação da qualidade das sementes.

## REFERÊNCIAS

- Ainsworth, E.A., Yendrek, C.R., Skoneczka, J.A., Long, S.P., 2012. Accelerating yield potential in soybean: potential targets for biotechnological improvement. *Plant. Cell Environ.* 35, 38–52. <https://doi.org/10.1111/j.1365-3040.2011.02378.x>
- Baek, I., Kusumaningrum, D., Kandpal, L.M., Lohumi, S., Mo, C., Kim, M.S., Cho, B.K., 2019. Rapid measurement of soybean seed viability using Kernel-based multispectral image analysis. *Sensors (Switzerland)* 19, 1–16. <https://doi.org/10.3390/s19020271>
- Baek, J., Lee, E., Kim, N., Kim, S.L., Choi, I., Ji, H., Chung, Y.S., Choi, M.-S.S., Moon, J.-K.K., Kim, K.-H.H., 2020. High throughput phenotyping for various traits on soybean seeds using image analysis. *Sensors (Switzerland)* 20, 1–9. <https://doi.org/10.3390/s20010248>
- Barboza da Silva, C., Oliveira, N.M., de Carvalho, M.E.A., de Medeiros, A.D., de Lima Nogueira, M., dos Reis, A.R., 2021. Autofluorescence-spectral imaging as an innovative method for rapid, non-destructive and reliable assessing of soybean seed quality. *Sci. Rep.* 11, 17834. <https://doi.org/10.1038/s41598-021-97223-5>
- Batista, T.B., Mastrangelo, C.B., de Medeiros, A.D., Petronilio, A.C.P., Fonseca de Oliveira, G.R., dos Santos, I.L., Crusciol, C.A.C., Amaral da Silva, E.A., 2022. A Reliable Method to Recognize Soybean Seed Maturation Stages Based on Autofluorescence-Spectral Imaging Combined With Machine Learning Algorithms. *Front. Plant Sci.* 13. <https://doi.org/10.3389/fpls.2022.914287>
- Berg, S., Kutra, D., Kroeger, T., Straehle, C.N., Kausler, B.X., Haubold, C., Schiegg, M., Ales, J., Beier, T., Rudy, M., Eren, K., Cervantes, J.I., Xu, B., Beuttenmueller, F., Wolny, A., Zhang, C., Koethe, U., Hamprecht, F.A., Kreshuk, A., 2019. Ilastik: Interactive Machine Learning for (Bio)Image Analysis. *Nat. Methods* 16, 1226–1232. <https://doi.org/10.1038/s41592-019-0582-9>
- de Medeiros, A.D., Capobiango, N.P., da Silva, J.M., da Silva, L.J., da Silva, C.B., dos Santos Dias, D.C.F., 2020. Interactive machine learning for soybean seed and seedling quality classification. *Sci. Rep.* 10, 11267. <https://doi.org/10.1038/s41598-020-68273-y>
- Dell'Aquila, A., 2009. Development of novel techniques in conditioning, testing and sorting seed physiological quality. *Seed Sci. Technol.* 37, 608–624.

<https://doi.org/10.15258/sst.2009.37.3.10>

- ElMasry, G., ElGamal, R., Mandour, N., Gou, P., Al-Rejaie, S., Belin, E., Rousseau, D., 2020. Emerging thermal imaging techniques for seed quality evaluation: Principles and applications. *Food Res. Int.* 131, 109025. <https://doi.org/10.1016/j.foodres.2020.109025>
- Finch-Savage, W.E.E., Bassel, G.W.W., 2016. Seed vigour and crop establishment: extending performance beyond adaptation. *J. Exp. Bot.* 67, 567–591. <https://doi.org/10.1093/jxb/erv490>
- Li, Y., Jia, J., Zhang, L., Khattak, A.M., Sun, S., Gao, W., Wang, M., 2019. Soybean Seed Counting Based on Pod Image Using Two-Column Convolution Neural Network. *IEEE Access* 7, 64177–64185. <https://doi.org/10.1109/ACCESS.2019.2916931>
- Mahajan, S., Mittal, S.K., Das, A., 2018. Machine vision based alternative testing approach for physical purity, viability and vigour testing of soybean seeds (*Glycine max*). *J. Food Sci. Technol.* <https://doi.org/10.1007/s13197-018-3320-x>
- Olesen, M.H., Nikneshan, P., Shrestha, S., Tadayyon, A., Deleuran, L.C., Boelt, B., Gislum, R., 2015. Viability prediction of *ricinus cummunis* L. Seeds using multispectral imaging. *Sensors (Switzerland)* 15, 4592–4604. <https://doi.org/10.3390/s150204592>
- Schneider, C.A., Rasband, W.S., Eliceiri, K.W., 2012. NIH Image to ImageJ: 25 years of image analysis. *Nat. Methods* 9, 671–675. <https://doi.org/10.1038/nmeth.2089>
- Wilkes, T., Nixon, G., Bushell, C., Waltho, A., Alroichdi, A., Burns, M., 2016. Feasibility Study for Applying Spectral Imaging for Wheat Grain Authenticity Testing in Pasta. *Food Nutr. Sci.* 07, 355–361. <https://doi.org/10.4236/fns.2016.75037>
- Zhang, C., Si, Y., Lamkey, J., Boydston, R.A., Garland-Campbell, K.A., Sankaran, S., 2018. High-throughput phenotyping of seed/seedling evaluation using digital image analysis. *Agronomy* 8, 1–14. <https://doi.org/10.3390/agronomy8050063>

# **APRENDIZADO DE MÁQUINA INTERATIVO PARA CLASSIFICAÇÃO DA QUALIDADE DE SEMENTES E PLÂNTULAS DE SOJA**

## **RESUMO**

Novas soluções de visão computacional combinadas com algoritmos de inteligência artificial podem ajudar a reconhecer padrões em imagens, reduzindo a subjetividade e otimizando o processo de análise. O objetivo deste estudo foi propor uma abordagem baseada em métodos de aprendizado de máquina interativos e tradicionais para classificar sementes e plântulas de soja de acordo com seu aspecto físico externo e potencial fisiológico. Além disso, correlacionamos o aspecto físico externo das sementes ao seu desempenho fisiológico. Imagens de sementes e plântulas de soja foram usadas para desenvolver modelos usando abordagens de baixo custo e software de acesso gratuito. Os modelos desenvolvidos mostraram alto desempenho, com acurácia geral atingindo 0,94 para classificação de sementes e plântulas. A alta precisão dos modelos que foram desenvolvidos com base em aprendizado de máquina interativo e tradicional demonstrou que o método pode ser facilmente utilizado para classificar sementes de soja de acordo com seu aspecto físico externo, bem como para classificar o vigor das plântulas de soja de forma rápida e não subjetiva. O aspecto físico externo das sementes de soja está fortemente correlacionado com seu desempenho fisiológico.

Palavras-chave: Software Ilastik. Fenotipagem de alto rendimento. Aspecto físico externo da semente. Potencial fisiológico de sementes.

## **ABSTRACT**

New computer vision solutions combined with artificial intelligence algorithms can help recognize patterns in images, reducing subjectivity and optimizing the analysis process. The aim of this study was to propose an approach based on interactive and traditional machine learning methods to classify soybean seeds and seedlings according to their appearance and physiological potential. In addition, we correlated the appearance of seeds to their physiological performance. Images of soybean seeds and seedlings were used to develop models using low-cost approaches and free-access software. The models developed showed high performance, with overall accuracy reaching 0.94 for seeds and seedling classification. The high precision of the models that were developed based on interactive and traditional machine learning demonstrated that the method can easily be used to classify soybean seeds according to their appearance, as well as to classify soybean seedling vigor quickly and non-subjectively. The appearance of soybean seeds is strongly correlated with their physiological performance.

**Keywords:** Ilastik software. High-throughput phenotyping. Seed appearance. Seed physiology.

## INTRODUÇÃO

A soja [*Glycine max* (L.) Merr.] é uma das principais commodities do agronegócio mundial. As sementes de soja são ricas em aminoácidos, gorduras, vitaminas, minerais e são uma excelente fonte de proteínas, sendo chave para a segurança alimentar global. Alcançar uma alta produtividade da soja depende do estabelecimento bem-sucedido das plantas, que se baseia no uso de sementes de alta qualidade (Finch-Savage and Bassel, 2016).

A qualidade da semente é muito suscetível a condições ambientais e procedimentos pós-colheita, como beneficiamento mecânico e secagem artificial (Esteve Agelet et al., 2012). Essas condições podem resultar em vários danos às sementes e alterações em seu aspecto físico externo, como amassados, ataques de patógenos, ruptura do tegumento, danos por umidade e sementes esverdeadas. Muitos esforços têm sido feitos para evitar esses danos e melhorar a qualidade das sementes. Na maioria dos casos, esses esforços baseiam-se na inspeção visual de lotes de sementes e métodos químicos que são destrutivos, subjetivos e demorados (Liu et al., 2015).

Métodos de aprendizado de máquina têm apoiado a recente revolução em visão computacional (Berg et al., 2019). Novas abordagens que combinam o uso de algoritmos de aprendizado de máquina com visão computacional trouxeram perspectivas novas e promissoras para a análise da qualidade de produtos agrícolas, especialmente sementes (Lin et al., 2019; Mahajan et al., 2018; Momin et al., 2017). Com essas tecnologias, muitas das limitações enfrentadas hoje pelos métodos tradicionais de inspeção visual de sementes poderiam ser resolvidas.

Softwares de código aberto colaborativo têm oferecido soluções poderosas em vários campos de pesquisa com bioimagens. Devido à sua flexibilidade e transparência no trato com novas tecnologias, essas ferramentas têm se destacado (Dietz et al., 2020), abrindo novas possibilidades de aplicação em áreas ainda pouco exploradas. Entre eles, o Ilastik é um software de código aberto que permite o desenvolvimento de modelos baseados em aprendizado de máquina interativo de imagens, fácil de usar, ideal para usuários sem conhecimento computacional substancial (Berg et al., 2019). Este software tem sido utilizado em pesquisas recentes para medir a confluência da cultura de células Hep G2 em micrografias de contraste de fases (Yordanov, 2020), triagem de alto rendimento para quantificar danos de tripes (Visschers et al., 2018), quantificação e análise espacial de características em imagens histológicas do cérebro de roedores (Yates et al., 2019), entre outros. No entanto, ainda não há informações sobre a aplicação de aprendizado de máquina interativo via Ilastik na ciência das plantas, particularmente para estudos com sementes e plântulas.

Considerando que o uso da análise de imagem na tecnologia de sementes proporcionou grandes avanços na avaliação da qualidade das sementes, o software Ilastik poderia ser integrado a essa abordagem, tornando essa técnica ainda mais eficiente. Assim, este estudo teve como objetivo propor uma abordagem baseada em métodos de aprendizado de máquina interativos e tradicionais para classificar sementes e plântulas de soja de acordo com seu aspecto físico externo e potencial fisiológico. Além disso, relacionamos o aspecto físico externo da semente ao seu desempenho fisiológico.

## **MATERIAL E MÉTODOS**

### **Material vegetal e aquisição de imagens**

Lotes comerciais de sementes de soja produzidos na safra 2019/2020 foram adquiridos, dos quais 700 sementes foram selecionadas. As sementes foram divididas em sete grupos de acordo com aspectos físicos, sanitários e fisiológicos. Os grupos foram compostos por sementes danificadas mecanicamente, sementes infestadas por fungos e sementes verdes (apresentando baixa degradação de clorofila). Inicialmente, imagens das sementes foram adquiridas por um scanner Epson Perfection V800. Em seguida, as sementes foram testadas quanto à capacidade de germinação. Estas sementes foram colocadas em rolos de papel de germinação umedecidos com água equivalente a 2,5 vezes a massa do papel seco e mantidas em germinador (Mangelsdorf) a 25 °C por 3 dias. As plântulas produzidas foram avaliadas individualmente através de análise de imagem para determinar o vigor da semente. As imagens das plântulas foram adquiridas através de um scanner HP modelo Scanjet 200 fixado em uma posição invertida dentro de uma caixa de alumínio. Ambas as imagens, de sementes e plântulas, tinham resolução de 300 dpi.

### **Pré-processamento e segmentação de imagens**

O pré-processamento e a segmentação das imagens foram realizados no software Ilastik (Berg et al., 2019) utilizando a ferramenta de classificação de pixels. Foi produzida uma segmentação semântica das imagens. Duas classes de segmentação foram definidas: 'semente ou plântula' e 'fundo'. Os recursos das imagens basearam-se em cor e intensidade de pixel, descritores de borda e textura, calculados como filtros pré-suavizados com um sigma variando de 0,3 a 1. Em seguida, pixels pertencentes às regiões de cada classe foram selecionados para o treinamento do modelo. O classificador Random Forest foi usado para a classificação de pixels (padrão Ilastik). A probabilidade de o pixel pertencer às classes de segmentação semântica ('semente/plântula' ou 'fundo') foi estimada para cada pixel da imagem. O classificador treinado foi aplicado a todas as imagens, sementes e plântulas individuais, em modo de lote, com mapas de probabilidade exportados para cada imagem.

### **Classificação do aspecto físico externo e qualidade fisiológica da semente**

Neste estudo, desenvolvemos classificações independentes para o aspecto físico externo da semente de soja e qualidade fisiológica, considerando o crescimento da plântula e os dados da semente não germinada. Os métodos utilizados para ambas as classificações foram semelhantes, portanto, as mesmas descrições foram apresentadas uma única vez.

### **Classes de aspecto físico da semente**

As sementes foram classificadas em sete classes diferentes: 1) Semente de alta qualidade (HQS) - semente quase redonda, firme, com tegumento liso e de cor única; 2) Semente amassada (KNS) - sementes com superfícies irregulares e depressões devido a danos mecânicos; 3) Sementes com mancha púrpura (PSS) - sementes mostrando descoloração de rosa a roxo claro ou escuro, que variava em tamanho, desde uma pequena mancha até cobrir toda o tegumento da semente; 4) Sementes quebradas (BRS) - sementes com integridade estrutural comprometida, apenas um dos cotilédones presentes ou porções significativas da sementes ausente; 5) Rasgo no tegumento da semente (SCT) - sementes mostrando rupturas no tegumento; 6) Semente danificada por umidade (MDS) - sementes mostrando rugas na região oposta ao hilo; 7) Sementes esverdeadas (GRS) - sementes com tegumento ou cotilédones esverdeados.

### **Classes de qualidade fisiológica**

As plântulas foram classificadas em duas classes: 1) Plântula vigorosa (VSD) - plântula morfológicamente saudável mostrando crescimento vigoroso do hipocótilo e raízes; e 2) Plântula fraca (WSD) - plântula mostrando ausência, subdesenvolvimento ou deformação de alguma estrutura essencial (cotilédones, hipocótilo, folhas primordiais e raízes); além disso, foi estabelecida a classe 3) Semente não germinada (NGS) - semente incapaz de germinar após três dias em condições adequadas.

### **Classificação interativa de aspecto físico e vigor da semente**

A classificação interativa foi realizada no software Ilastik, usando a ferramenta integrada de classificação de objetos. Primeiro, as imagens de treinamento foram inseridas (10% do total de imagens foi utilizado) e seus respectivos mapas de probabilidade foram obtidos na etapa de pré-processamento. Em seguida, 103 variáveis foram calculadas para cada semente e plântula, incluindo descritores de forma baseados em casco convexo e esqueleto, e estatísticas de propriedades de intensidade de cinza. Finalmente, treinamos o classificador selecionando classes individuais conhecidas.

O classificador treinado foi aplicado a toda a série de imagens em modo de lote, com descritores e mapas de previsão individuais exportados em CSV e PNG, respectivamente. Essa abordagem não exigiu hardware com um alto nível de capacidade de processamento. Realizamos toda a classificação em um Intel Core®™ CPU i5-42000 @ 1.60 GHz e 4 GB de RAM."

## Algoritmos externos para classificação do aspecto físico externo das sementes e qualidade fisiológica

Os descritores gerados pelo Ilastik para cada semente e plântula foram usados para desenvolver modelos de classificação usando técnicas de aprendizado de máquina baseadas em três métodos: Análise Discriminante Linear (LDA), Random Forest (RF) e Máquina de Vetor de Suporte (SVM). Os dados foram divididos em conjuntos de treinamento e validação, numa proporção de 70% e 30%, respectivamente. Os três classificadores foram desenvolvidos no software R 3.6.3 (R Core Team, 2019). Para LDA, os pacotes `Mass` (<https://cran.r-project.org/web/packages/MASS/index.html>) e `flipMultivariates` (<https://github.com/Displayr/flipMultivariates/>) foram utilizados; Para os modelos de random forest, a biblioteca `randomForest` (<https://cran.r-project.org/web/packages/randomForest/>) foi usada, com 500 árvores de decisão e hiperparâmetros padrão; Para SVM, a biblioteca `e1071` (<https://cran.r-project.org/web/packages/e1071/index.html>) foi usada, com um kernel radial e parâmetros padrão.

### Validação dos Modelos

Os modelos desenvolvidos com o método iterativo usando o software Ilastik foram validados utilizando 90% do conjunto de dados. Esses dados não haviam sido usados para treinamento. Os modelos externos desenvolvidos foram validados através de validação cruzada (10 dobras) e através de um conjunto de validação independente que não havia sido usado anteriormente. Eles foram avaliados com base em dados de Verdadeiro Positivo (TP), Falso Positivo (FP), Verdadeiro Negativo (TN) e Falso Negativo (FN). As métricas de Acurácia, Kappa, Precisão, Sensibilidade e Especificidade foram calculadas:

$$\text{Acurácia} = \frac{TP+TN}{TP+TN+FP+FN} \quad (1)$$

$$\text{Kappa} = 2 * \frac{TP*TN-FP*FN}{TP*FN+TP*FP+2*TP*TN+FN^2+FN*TN+FP^2+FP*TN} \quad (2)$$

$$\text{Precisão} = \frac{TP}{TP+FP} \quad (3)$$

$$\text{Sensibilidade} = \frac{TP}{TP+FN} \quad (4)$$

$$\text{Especificidade} = \frac{TN}{TN+FP} \quad (5)$$

### **Associação entre o aspecto físico externo da semente e qualidade fisiológica**

Para entender a relação entre o aspecto físico externo da semente e sua qualidade fisiológica, fizemos uma associação entre as classes de sementes e suas respectivas classes de plântulas. O comprimento da plântula foi medido e os dados foram usados para obter os índices de crescimento, uniformidade e vigor, e o comprimento da raiz usando o software Vigor-S® (Castan et al., 2018). A análise de componentes principais multivariada foi aplicada a esses dados, adotando as classes de sementes como indivíduos e os índices de vigor da plântula como vetores.

## RESULTADOS

### Modelos de aprendizado de máquina para classificação de sementes

Foram desenvolvidos e comparados quatro modelos diferentes para classificar sementes de soja com base em aspectos visíveis relevantes usando uma ferramenta de alto desempenho. O modelo baseado em aprendizado de máquina interativo mostrou uma precisão geral de 0,83 (Tabela 1). As classes BRS e SCT tiveram as maiores taxas de acerto, com sensibilidade superior a 0,97. Em contraste, as classes HQS e MDS tiveram menor precisão individual, com uma taxa maior de falsos positivos e falsos negativos. Pelo menos 21% das sementes da classe HQS foram confundidas com as sementes da classe MDS, enquanto a confusão aumentou para 30% na direção oposta.

**Tabela 1.** Matrizes de confusão e métricas da classificação interativa de sementes de soja de acordo com o aspecto físico externo. O classificador Random Forest foi aplicado e 10% do total de imagens foram usados para treinamento.

Classe <sup>1</sup>	Alta qualidade	Amassada	Mancha púrpura	Quebrada	Rasgo de tegumento	Dano por umidade	Esverdeada
<b>n = 630</b>							
Alta qualidade	<b>70</b>	1	3	0	2	27	0
Amassada	0	<b>76</b>	17	0	0	0	6
Mancha púrpura	1	2	<b>64</b>	0	0	0	2
Quebrada	0	0	0	<b>87</b>	0	1	0
Rasgo de tegumento	0	3	3	3	<b>88</b>	1	2
Dano por umidade	19	0	1	0	0	<b>61</b>	3
Esverdeada	0	8	2	0	0	0	<b>77</b>
Acurácia	0.92	0.95	0.94	0.99	0.98	0.92	0.96
Kappa	0.68	0.78	0.77	0.97	0.91	0.65	0.85
Precisão	0.68	0.77	0.93	0.99	0.88	0.73	0.89
Sensibilidade	0.78	0.84	0.71	0.97	0.98	0.68	0.86
Especificidade	0.94	0.96	0.99	1	0.98	0.96	0.98

<sup>1</sup> Nas colunas estão as verdadeiras classes de sementes, e nas linhas estão as classes estimadas

Os descritores de semente gerados pelo software Ilastik foram testados aplicando três métodos diferentes de aprendizado de máquina (Tabela 2). Em geral, foi observado alto desempenho nos três modelos desenvolvidos. O método LDA se destacou, alcançando uma acurácia de 0,93 no conjunto de validação cruzada e 0,94 no conjunto de validação

independente, e um coeficiente Kappa acima de 0,91. Na validação independente, as classes SCT e BRS tiveram as maiores taxas de acerto nos modelos LDA, RF e SVM.

**Tabela 2.** Matrizes de confusão e métricas obtidas com modelos externos de classificação de sementes usando descritores de sementes gerados pelo Ilastik.

Método	Classe	Treinamento (n = 490)	Validação cruzada Acertos (Total)	Validação (n =210)
LDA	Alta qualidade	68 (70)	-	29 (30)
	Amassada	64 (70)	-	29 (30)
	Mancha púrpura	69 (70)	-	28 (30)
	Quebrada	68 (70)	-	30 (30)
	Rasgo de tegumento	70 (70)	-	30 (30)
	Dano por umidade	66 (70)	-	25 (30)
	Esverdeada	69 (70)	-	27 (30)
	Acurácia	0.98	0.93±0.03	0.94
	Kappa	0.96	0.92±0.04	0.93
	Precisão	0.97	0.93±0.04	0.94
	Sensibilidade	0.97	0.93±0.03	0.94
	Especificidade	0.99	0.99±0.01	0.99
	RF	Alta qualidade	64 (70)	-
Amassada		65 (70)	-	27 (30)
Mancha púrpura		70 (70)	-	28 (30)
Quebrada		69 (70)	-	30 (30)
Rasgo de tegumento		70 (70)	-	29 (30)
Dano por umidade		64 (70)	-	24 (30)
Esverdeada		66 (70)	-	26 (30)
Acurácia		0.96	0.89±0.06	0.90
Kappa		0.95	0.87±0.07	0.88
Precisão		0.96	0.90±0.06	0.90
Sensibilidade		0.96	0.89±0.07	0.90
Especificidade		0.99	0.98±0.07	0.98
SVM		High-quality seed	67 (70)	-
	Alta qualidade	63 (70)	-	26 (30)
	Amassada	68 (70)	-	28 (30)
	Mancha púrpura	68 (70)	-	30 (30)
	Quebrada	70 (70)	-	30 (30)
	Rasgo de tegumento	64 (70)	-	24 (30)
	Dano por umidade	67 (70)	-	26 (30)
	Acurácia	0.95	0.92±0.03	0.91
	Kappa	0.95	0.91±0.04	0.89
	Precisão	0.95	0.92±0.03	0.91
	Sensibilidade	0.95	0.92±0.03	0.90
	Especificidade	0.99	0.99±0.01	0.98

### Modelos de aprendizado de máquina para classificação de qualidade fisiológica

Os dados fisiológicos foram usados para classificar as sementes em três classes de acordo com a capacidade de germinação da semente e o crescimento da plântula: plântulas vigorosas (VSD), plântulas fracas (WSD) e sementes não germinadas (NGS). Quatro modelos foram desenvolvidos para esse fim - o primeiro baseado em aprendizado de máquina iterativo

e os outros usando a abordagem tradicional de aprendizado de máquina para analisar os dados gerados pelo software Ilastik. Note-se que, embora o conjunto de dados deste estudo tivesse 700 sementes, apenas 600 foram usadas para o desenvolvimento e avaliação desses modelos. A classe BRS foi excluída, pois geralmente não é usada para avaliar o vigor da semente.

Um modelo de classificação de alta precisão dentro do método de aprendizado de máquina iterativo Ilastik foi usado (Tabela 3). O modelo alcançou uma acurácia média de 0,97 e valores acima de 0,92 para as métricas kappa, precisão, sensibilidade e especificidade para a classe VSD. A maior taxa de falsos negativos foi encontrada na classe WSD e a menor taxa na classe VSD.

**Tabela 3.** Matrizes de confusão e métricas da classificação de aprendizado de máquina iterativo Ilastik de sementes de soja de acordo com sua qualidade fisiológica.

Classe <sup>1</sup>	Plântula vigorosa	Plântula fraca n = 600	Semente não-germinada
Plântula vigorosa	<b>230</b>	18	0
Plântula fraca	1	<b>157</b>	9
Semente não-germinada	0	10	<b>174</b>
Acurácia	0.97	0.94	0.97
Kappa	0.93	0.85	0.93
Precisão	0.93	0.94	0.95
Sensibilidade	0.99	0.85	0.95
Especificidade	0.95	0.98	0.98

<sup>1</sup> Nas colunas estão as verdadeiras classes de sementes, e nas linhas estão as classes estimadas

Modelos baseados em classificação externa também foram satisfatórios (Tabela 4). Para esses modelos, foram utilizados os dados gerados pelo software Ilastik para plântulas e sementes não-germinadas. O conjunto de treinamento compreendeu 70% dos dados. O método Random Forest teve um desempenho ligeiramente melhor do que os outros nos conjuntos de validação independente, mostrando precisão de 0,93 e 0,94, respectivamente. A taxa de acerto para cada classe variou de acordo com o classificador, mas, em geral, houve um erro de classificação mais significativo para a classe WSD.

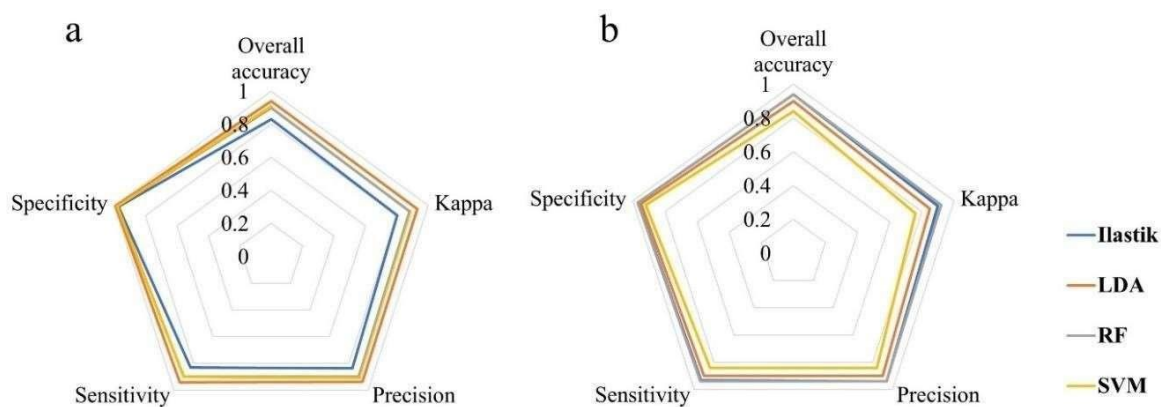
**Tabela 4.** Matrizes de confusão e métricas obtidas com classificação externa usando descritores do Ilastik de plântulas de soja e sementes de acordo com sua qualidade fisiológica.

Método	Classes	Treinamento (n =422)	Validação cruzada	Validação independente (n =178)
LDA	Plântula vigorosa	159 (162)	-	64 (69)
	Plântula fraca	122 (131)	-	44 (55)
	Semente não-germinada	129 (129)	-	52 (54)
	Acurácia	0.97	0.92±0.04	0.90
	Kappa	0.96	0.89±0.06	0.85
	Precisão	0.97	0.92±0.04	0.90
	Sensibilidade	0.97	0.93±0.04	0.90
	Especificidade	0.99	0.96±0.02	0.95
RF	Plântula vigorosa	160 (162)	-	67 (69)
	Plântula fraca	120 (131)	-	49 (55)
	Semente não-germinada	128 (129)	-	52 (54)
	Acurácia	0.97	0.93±0.03	0.94
	Kappa	0.95	0.89±0.04	0.92
	Precisão	0.97	0.93±0.03	0.94
	Sensibilidade	0.97	0.93±0.03	0.94
	Especificidade	0.98	0.96±0.02	0.97
SVM	Plântula vigorosa	158 (162)	-	60 (69)
	Plântula fraca	110 (131)	-	42 (55)
	Semente não-germinada	127 (129)	-	48 (54)
	Acurácia	0.94	0.89±0.03	0.84
	Kappa	0.90	0.83±0.04	0.76
	Precisão	0.94	0.89±0.02	0.84
	Sensibilidade	0.93	0.89±0.03	0.84
	Especificidade	0.97	0.94±0.02	0.92

### Resumo da Classificação

Em geral, as classificações baseadas em aprendizado de máquina iterativo e na abordagem tradicional foram precisas. No entanto, especialmente para classificações baseadas no aspecto físico externo da semente, os modelos de classificação externa tiveram um desempenho melhor (Figura 1a). Para esses modelos, é importante enfatizar que 70% dos dados foram usados para treinamento. Para a classificação baseada na qualidade fisiológica da semente, os modelos mostraram desempenho semelhante (Figura 1b). Note que o método Ilastik foi ligeiramente melhor (valores mais altos) na maioria das métricas do que os algoritmos LDA

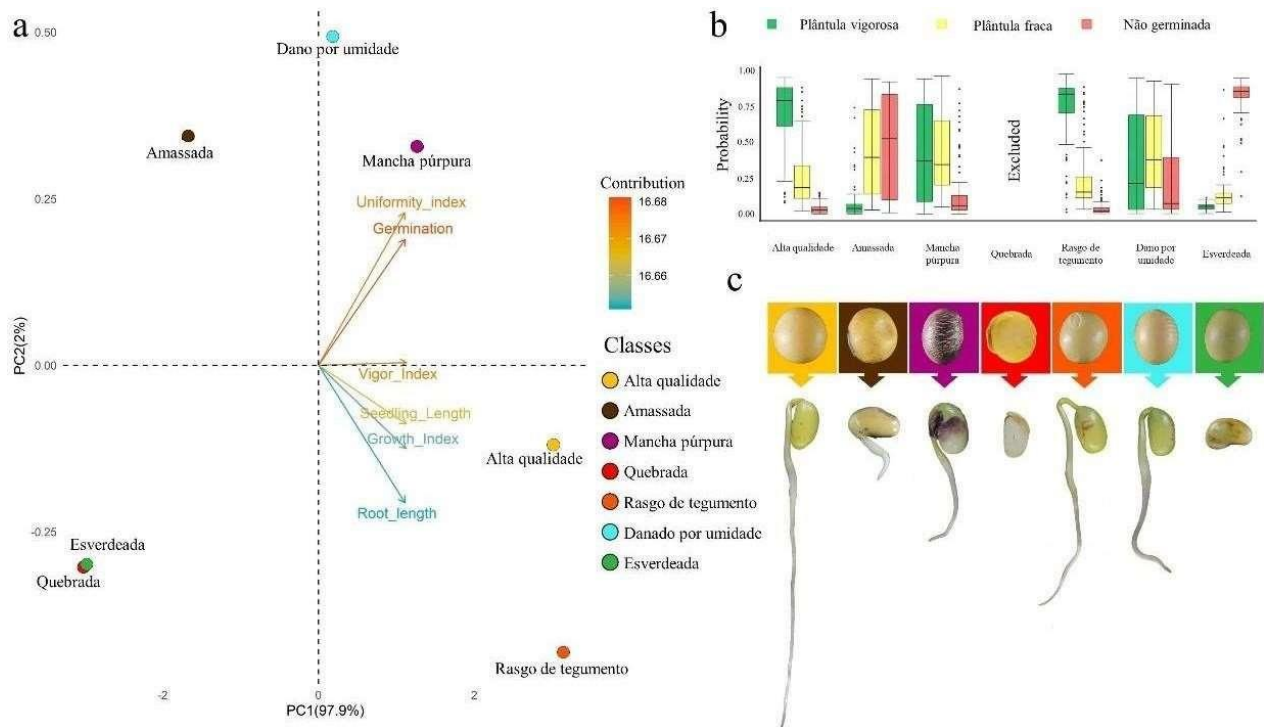
e SVM, mesmo usando apenas 10% dos dados para treinamento, contra 70% no método de classificação externa.



**Figura 1.** Gráfico para as métricas dos classificadores de aprendizado de máquina testados. Classificações no aspecto físico externo da semente (a) e qualidade fisiológica (b).

### **Relação entre o aspecto físico externo da semente e crescimento da plântula.**

Correlação entre aspecto físico externo da semente e crescimento da plântula. As sementes exibiram diferentes desempenhos fisiológicos de acordo com seu aspecto físico externo (Figura 2). Na análise de componentes principais multivariada, os vetores de qualidade fisiológica foram posicionados à direita do diagrama de ordenação, indicando que os indivíduos localizados em escores negativos do PC1 tinham valores significativamente menores para essas variáveis (Figura 2a). Em detalhes, as classes HQS e SCT mostraram valores mais altos de comprimento da plântula e índices mais elevados de vigor, uniformidade e crescimento (Figura 2a). Para essas classes, todas as sementes germinaram e geraram uma alta proporção de plântulas vigorosas (Figura 2b). Além disso, nas classes PSS e MDS, as sementes geraram principalmente plântulas vigorosas e plântulas fracas. A MDS também teve uma alta proporção de sementes não germinadas. Para a classe KNS, as sementes geraram principalmente plântulas fracas ou não germinaram. Finalmente, as sementes da classe GRS, em sua maioria, não germinaram. O aspecto físico externo da plântula que predominou em cada classe de semente é mostrado na Figura 2c.



**Figura 1.** Correlação entre o aspecto físico externo da semente e crescimento da plântula. Biplot da análise de componentes principais mostrando a importância dos parâmetros de qualidade da semente para a dispersão de classes (a), probabilidade de gerar plântulas vigorosas e plântulas fracas ou ausência de germinação das sementes individuais de acordo com sua classe (b), e aspecto físico externo da semente e o aspecto predominante das plântulas em cada classe de semente (c).

## DISCUSSÃO

A qualidade da semente de soja tem um impacto direto no seu preço de mercado e afeta o estabelecimento da plântula no campo (Lin et al., 2019). Métodos de inspeção visual são atualmente utilizados por laboratórios para avaliar precisamente a qualidade da semente de soja, incluindo avaliações de plântulas. No entanto, esses métodos são subjetivos, demorados e geralmente destrutivos (Liu et al., 2015). Neste estudo, apresentamos uma abordagem baseada em métodos de aprendizado de máquina interativos e tradicionais para classificar automaticamente sementes e plântulas de soja de forma eficiente e sem recursos dispendiosos.

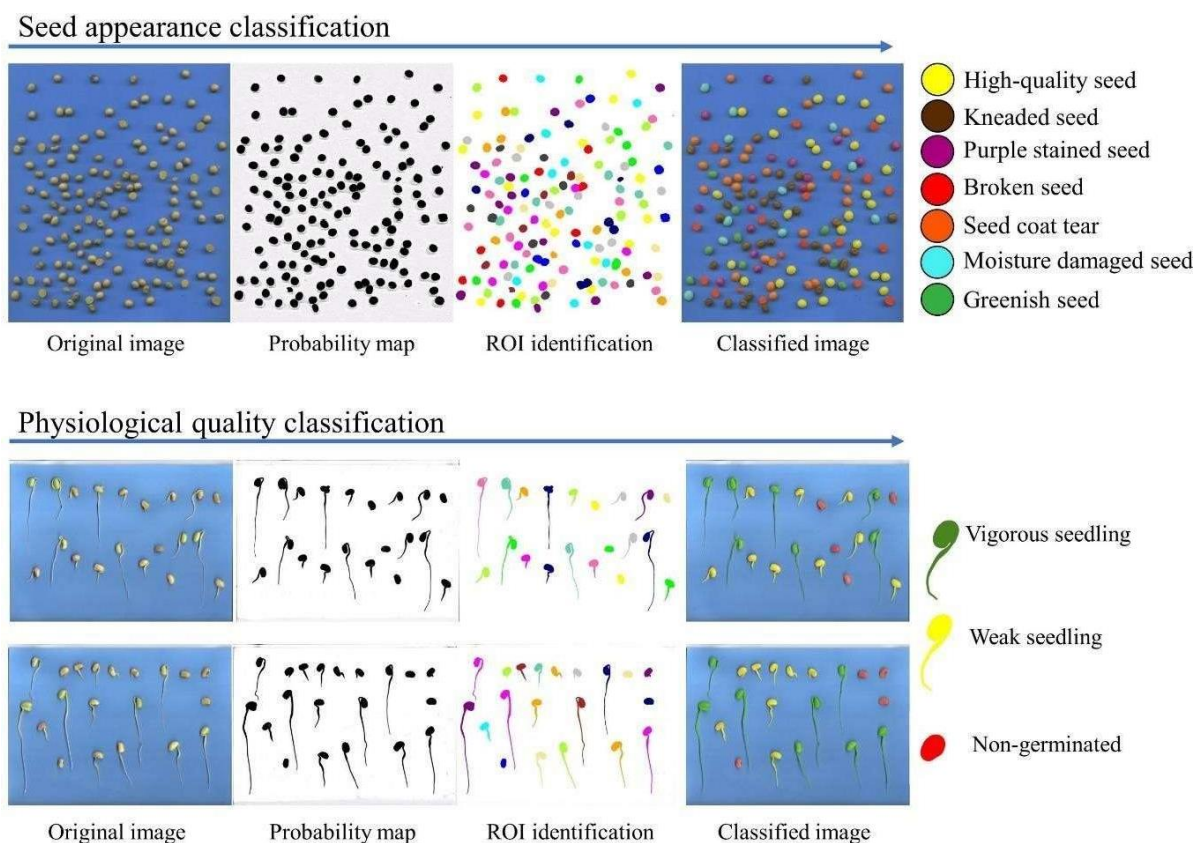
Os métodos propostos foram altamente satisfatórios. As precisões encontradas na discriminação das sementes em suas diferentes classes do aspecto físico externo foram altas, variando de 0,92 a 0,99 para o modelo de aprendizado de máquina interativo e superior a 0,90 de acurácia geral no conjunto de validação independente para classificação externa (Tabela 1 e Tabela 2). Esses resultados indicam alto potencial para aplicação prática desses métodos. Neste estudo, usamos vários lotes de sementes, o que é um bom primeiro passo em direção à generalização, e eleva a esperança quanto à possível generalização para fins industriais. Os modelos desenvolvidos podem ser inseridos nos programas de controle de qualidade de empresas de sementes ou laboratórios de análise de sementes, ou até mesmo em estudos em centros de pesquisa. Estamos cientes de que essa abordagem não dispensa totalmente o esforço humano, pois alguns testes manuais são necessários para atender aos padrões ISTA (Boelt et al., 2018), mas ela pode ser usada no processo de triagem de lotes em relação à qualidade das sementes e na identificação das razões para perda de qualidade de forma rápida e precisa. Vale ressaltar que este é um estudo pioneiro usando métodos de aprendizado de máquina interativos para classificar sementes de soja de acordo com seu aspecto físico externo.

Em estudos anteriores, pesquisadores tentaram identificar danos em sementes de soja usando diferentes técnicas. Lin et al. (2019) descreveram a representação conjunta do bag-of-feature multi-modal (JMBoF) para inspecionar a qualidade de aspecto físico externo de sementes de soja pós-colheita secas usando imagens obtidas no espectro visível. O algoritmo proposto alcançou uma precisão de 82% no conjunto de testes. Mahajan et al. (2018) usaram imagens RGB e de raios-X para avaliar a pureza física, viabilidade e vigor das sementes de soja. O classificador desenvolvido, baseado em uma rede neural, alcançou uma precisão de 91%. Momin et al. (2017) desenvolveram um sistema de visão por computador para detectar impurezas em amostras de soja colhidas e alcançaram uma precisão variando de 75% a 98%. Liu et al. (2015) desenvolveram um sistema para identificar e eliminar sementes de soja

danificadas usando visão por computador, tecnologias de processamento de imagem, redes neurais e controle mecânico automatizado. O sistema apresentou uma precisão média de reconhecimento de 97%. Várias outras técnicas usadas para avaliar a qualidade das sementes de soja incluem espectroscopia Raman (Lee et al., 2013; Schulmerich et al., 2012), espectroscopia no infravermelho próximo, e ressonância magnética nuclear (Krishnan et al., 2004; Pietrzak et al., 2002).

Os métodos de aprendizado de máquina aplicados neste estudo foram eficientes para a classificação da qualidade fisiológica das sementes. Encontramos uma acurácia geral de 0,94 no método interativo e uma acurácia de até 0,94 na classificação externa usando o conjunto de validação independente para o RF (Tabela 3 e Tabela 4). Essa técnica abre novas perspectivas para análises rápidas do vigor de sementes de soja e permite a fenotipagem de plântulas de soja para estudos genéticos. Outros softwares foram desenvolvidos para analisar a qualidade das sementes de soja, principalmente considerando o comprimento das plântulas, como SVIS (Sako et al., 2001), VIGOR-S (Castan et al., 2018), SAPL (Medeiros and Pereira, 2018) e GroundEye (<https://www.tbit.com.br/>). No entanto, alguns softwares são comerciais, o que restringe seu uso. Aqui, apresentamos uma alternativa gratuita e eficiente para a análise de sementes/plântulas que pode ser usada por muitos pesquisadores, laboratórios e instituições interessadas na aplicação de um método altamente eficiente para a classificação de vigor de sementes e plântulas de soja.

Um resumo das etapas do método interativo aplicado à classificação de sementes e plântulas de soja é mostrado na Figura 3. A aquisição de imagem é feita em um fundo azul para facilitar a segmentação e melhorar o mapa de probabilidade da ROI (Região de Interesse). A identificação de sementes e plântulas individuais pode ser vista imediatamente após o mapa de probabilidade. A previsão da classificação de cada semente ou plântula é mostrada pela cor dos grupos respectivos.



**Figura 3.** Representação das etapas de aprendizado de máquina interativo e classificação de qualidade fisiológica em sementes de soja.

Os modelos desenvolvidos podem ser aprimorados para lidar com mais classes, e o software Ilastik fornece todas as ferramentas necessárias para isso. Todo o fluxo de trabalho do Ilastik (recursos de segmentação genérica, classificadores não lineares, modelos gráficos probabilísticos) está envolto em uma interface intuitiva para treinamento interativo de classificadores e pós-processamento de algoritmos de segmentação e classificação (Berg et al., 2019).

Os algoritmos de aprendizado de máquina aplicados à classificação externa neste estudo têm sido usados para conjuntos de dados complexos de fenotipagem de plantas (Rahaman et al., 2015). Métodos como LDA, RF e SVM têm sido usados com sucesso em ciências das plantas. O LDA é um algoritmo de aprendizado de máquina popular que aumenta a distância entre as classes e reduz a separabilidade dentro da classe, combinando linearmente recursos e criando estimativas de limite. A germinação e o vigor de sementes de *Jatropha curcas* foram previstos com precisão usando o classificador LDA combinado com recursos morfométricos e de integridade tecidual obtidos a partir de imagens de raios X de sementes (Medeiros et al., 2020). O RF, por sua vez, é um classificador não linear formado por muitas árvores de decisão. Ele gera a classificação final com base em um sistema de votação para cada árvore, e no final, o algoritmo

seleciona a classe mais votada (Geurts et al., 2009; Rahaman et al., 2019). Em áreas de produção de milho na China, foi desenvolvido um classificador baseado em RF usando imagens de sensoriamento remoto de alta resolução. O classificador conseguiu diferenciar campos de produção de sementes e grãos e identificar variedades de milho (Zhang et al., 2020). No Ilastik, o método de aprendizado de máquina interativo usa um classificador RF com 100 árvores para a classificação de pixels (Berg et al., 2019). Por fim, o SVM é um algoritmo supervisionado que projeta dados em um espaço de recursos de dimensão superior e identifica um hiperplano para separar classes com a margem mais significativa possível. A vantagem de um hiperplano é sua robustez a valores extremos, reduzindo consideravelmente as classificações erradas (Lu et al., 2020). Além disso, em um estudo recente com plantas individuais de *Arabidopsis thaliana* sob diferentes níveis de diversos estresses abióticos, um algoritmo SVM usando concentrações de microRNA como recursos de entrada foi capaz de prever o estresse da planta com alta precisão ( $R^2 = 0,96$ ) (Vakilian, 2020). Esses estudos mostram o grande potencial das técnicas de aprendizado de máquina.

Embora nossos resultados sejam muito promissores, algumas limitações foram observadas. A presença combinada de diferentes classes na mesma semente individual é comum na soja. Por exemplo, uma semente pode ser esverdeada e mecanicamente amassada. Isso pode ter implicações negativas para a classificação se o objetivo for quantificar com precisão os danos individuais. Outra limitação está relacionada ao método de aquisição. Embora o uso de scanners e câmeras 2D seja a abordagem mais acessível e direta para obter imagens, a imagem 2D não cobre todas as faces da semente. Em nosso estudo, essa limitação causou confusão entre as classes HQS e MDS (Tabela 1), uma vez que o dano por umidade em alguns casos não era muito evidente. Em estudos futuros, imagens 3D podem ser aplicadas. Fatores genéticos e ambientais regulam o aspecto físico externo da semente, que está fortemente correlacionada com o desempenho fisiológico da semente, como mostrado no presente estudo (Figura 2). Curiosamente, sementes com tegumento rompido mostraram alto vigor, semelhante às sementes de alta qualidade. Acredita-se que essas rupturas sejam causadas por fatores genéticos associados às condições ambientais durante a maturação da semente. Embora seja muito comum, há poucos relatos sobre o efeito desse traço na qualidade fisiológica da semente (Machado et al., 2019).

Além disso, as sementes mecanicamente amassadas resultaram em um grande número de sementes não germinadas e plântulas fracas (Figura 2b). Esse dano é causado por impactos

mecânicos durante a colheita ou processamento das sementes. Dependendo da localização do dano, as sementes podem gerar uma plântula anormal ou não germinar. Sementes com teor de água extremamente baixo são mais suscetíveis a quebrar durante as operações mecânicas na colheita e no processamento. Essas sementes estavam na classe BRS, que praticamente não tinha sementes germinadas.

Sementes com mancha púrpura geralmente resultaram em plântulas fracas (Figura 2b). Essa característica é principalmente causada pelo fungo *Cercospora kikuchii*. Relatos anteriores mostraram uma correlação fraca entre essa característica e a qualidade fisiológica da semente, embora comprometa a sua comercialização.

Condições ambientais durante a produção de sementes e principalmente após a maturação das sementes podem levar a danos por umidade. Essas sementes geralmente têm baixa qualidade (Ebone et al., 2019), o que resulta em plântulas fracas e um número significativo de sementes não germinadas, como observado no presente estudo (Figura 2b). Além disso, altas temperaturas e déficit de água durante a maturação das sementes podem levar à formação de sementes esverdeadas. Essas sementes têm baixa taxa de germinação (Figura 2b).

A avaliação rápida da qualidade das sementes é essencial para a indústria de sementes. A tomada de decisões rápidas quanto à eliminação ou destino de lotes de sementes economiza tempo e recursos. Portanto, ferramentas que possam avaliar com precisão os lotes de sementes e identificar sementes de baixa qualidade fisiológica são de grande importância. Este estudo mostrou que é possível classificar sementes de acordo com seu aspecto físico externo, e essas características estão fortemente correlacionadas com seu potencial fisiológico. Portanto, a abordagem proposta tem potencial para aplicação na classificação de lotes de sementes de soja para triagem rápida, não destrutiva, não subjetiva e eficiente das sementes. Além disso, este método foi capaz de classificar eficaz e precisamente o vigor das plântulas de soja.

## CONCLUSÕES

O método de aprendizado de máquina interativo para a classificação de sementes de soja com base no seu aspecto físico externo é preciso. Essa abordagem identifica efetivamente sementes com danos e classifica as plântulas em níveis de vigor. É recomendado o uso dos algoritmos LDA, RF e SVM para classificar sementes e plântulas de soja com base em dados gerados com o software Ilastik. Sementes de soja com alterações na degradação da clorofila, manchas fúngicas e danos mecânicos têm baixa qualidade fisiológica.

## REFERÊNCIAS

- Berg, S., Kutra, D., Kroeger, T., Straehle, C.N., Kausler, B.X., Haubold, C., Schiegg, M., Ales, J., Beier, T., Rudy, M., Eren, K., Cervantes, J.I., Xu, B., Beuttenmueller, F., Wolny, A., Zhang, C., Koethe, U., Hamprecht, F.A., Kreshuk, A., 2019. Ilastik: Interactive Machine Learning for (Bio)Image Analysis. *Nat. Methods* 16, 1226–1232. <https://doi.org/10.1038/s41592-019-0582-9>
- Boelt, B., Shrestha, S., Salimi, Z., Jørgensen, J.R., Nicolaisen, M., Carstensen, J.M., 2018. Multispectral imaging – a new tool in seed quality assessment? *Seed Sci. Res.* 28, 222–228. <https://doi.org/10.1017/S0960258518000235>
- Castan, D.O.C., Gomes-Junior, F.G., Marcos-Filho, J., 2018. Vigor-S, a new system for evaluating the physiological potential of maize seeds. *Sci. Agric.* 75, 167–172. <https://doi.org/10.1590/1678-992x-2016-0401>
- Dietz, C., Rueden, C.T., Helfrich, S., Arena, E.T., Horn, M., Eglinger, J., Evans III, E.L., McLean, D.T., Novitskaya, T., Ricke, W.A., Sherer, N.M., Zijlstra, A., Berthold, M.R., Eliceiri, K.W., 2020. Integration of the ImageJ Ecosystem in the KNIME Analytics Platform. *Front. Comput. Sci.* 2, 8. <https://doi.org/10.3389/FCOMP.2020.00008>
- Ebone, L.A., Caverzan, A., Chavarria, G., 2019. Physiologic alterations in orthodox seeds due to deterioration processes. *Plant Physiol. Biochem.* <https://doi.org/10.1016/j.plaphy.2019.10.028>
- Esteve Agelet, L., Gowen, A.A., Hurburgh, C.R., O'Donnell, C.P., 2012. Feasibility of conventional and Roundup Ready® soybeans discrimination by different near infrared reflectance technologies. *Food Chem.* 134, 1165–1172. <https://doi.org/10.1016/j.foodchem.2012.02.144>
- Finch-Savage, W.E., Bassel, G.W., 2016. Seed vigour and crop establishment: extending performance beyond adaptation. *J. Exp. Bot.* 67, 567–591. <https://doi.org/10.1093/jxb/erv490>
- Geurts, P., Irrthum, A., Wehenkel, L., 2009. Supervised learning with decision tree-based methods in computational and systems biology. *Mol. Biosyst.* 5, 1593. <https://doi.org/10.1039/b907946g>
- Krishnan, P., Joshi, D.K., Maheswari, M., Nagarajan, S., Moharir, A. V., 2004. Characterisation of soybean and wheat seeds by nuclear magnetic resonance spectroscopy. *Biol. Plant.* 48, 117–120. <https://doi.org/10.1023/B:BIOP.0000024286.23683.05>

- Lee, H., Cho, B.K., Kim, M.S., Lee, W.H., Tewari, J., Bae, H., Sohn, S.I., Chi, H.Y., 2013. Prediction of crude protein and oil content of soybeans using Raman spectroscopy. *Sensors Actuators, B Chem.* 185, 694–700. <https://doi.org/10.1016/j.snb.2013.04.103>
- Lin, P., Xiaoli, L., Li, D., Jiang, S., Zou, Z., Lu, Q., Chen, Y., 2019. Rapidly and exactly determining postharvest dry soybean seed quality based on machine vision technology. *Sci. Rep.* 9, 1–11. <https://doi.org/10.1038/s41598-019-53796-w>
- Liu, D., Ning, X., Li, Z., Yang, D., Li, H., Gao, L., 2015. Discriminating and elimination of damaged soybean seeds based on image characteristics. *J. Stored Prod. Res.* 60, 67–74. <https://doi.org/10.1016/j.jspr.2014.10.001>
- Lu, Q., Forbes, M.G., Loewen, P.D., Backström, J.U., Dumont, G.A., Gopaluni, R.B., 2020. Support vector machine approach for model-plant mismatch detection. *Comput. Chem. Eng.* 133, 106660. <https://doi.org/10.1016/j.compchemeng.2019.106660>
- Machado, B.R., Pereira, C.C. de O., Carvalho, É.P.B., Silva, J.M.J.O. da, Mendes, L.M. de O., Lima, M.L. da P., Silva, J.M.J.O. da, Menezes, I.P.P. de, 2019. Effect of tear/crack on soybean (*Glycine max*) seed coat, physiological quality and pathology of the seed. *Aust. J. Crop Sci.* 988–994. <https://doi.org/10.21475/ajcs.19.13.06.p1780>
- Mahajan, S., Mittal, S.K., Das, A., 2018. Machine vision based alternative testing approach for physical purity, viability and vigour testing of soybean seeds (*Glycine max*). *J. Food Sci. Technol.* <https://doi.org/10.1007/s13197-018-3320-x>
- Medeiros, A.D. de, Pereira, M.D., 2018. SAPL ® : a free software for determining the physiological potential in soybean seeds. *Pesqui. Agropecuária Trop.* 48, 222–228. <https://doi.org/10.1590/1983-40632018v4852340>
- Medeiros, A.D. de, Pinheiro, D.T., Xavier, W.A., Silva, L.J. da, Dias, D.C.F. dos S., 2020. Quality classification of *Jatropha curcas* seeds using radiographic images and machine learning. *Ind. Crops Prod.* 146, 112162. <https://doi.org/10.1016/j.indcrop.2020.112162>
- Momin, M.A., Yamamoto, K., Miyamoto, M., Kondo, N., Grift, T., 2017. Machine vision based soybean quality evaluation. *Comput. Electron. Agric.* 140, 452–460. <https://doi.org/10.1016/j.compag.2017.06.023>
- Pietrzak, L.N., Frégeau-Reid, J., Chatson, B., Blackwell, B., 2002. Observations on water distribution in soybean seed during hydration processes using nuclear magnetic resonance imaging. *Can. J. Plant Sci.* 82, 513–519. <https://doi.org/10.4141/P01-150>
- Rahaman, M.M., Ahsan, M.A., Chen, M., 2019. Data-mining Techniques for Image-based Plant Phenotypic Traits Identification and Classification. *Sci. Rep.* 9, 19526.

<https://doi.org/10.1038/s41598-019-55609-6>

- Rahaman, M.M., Chen, D., Gillani, Z., Klukas, C., Chen, M., 2015. Advanced phenotyping and phenotype data analysis for the study of plant growth and development. *Front. Plant Sci.* <https://doi.org/10.3389/fpls.2015.00619>
- Sako, Y., Mcdonald, M.B., Fujimura, K., Evans, A.F., Bennett, M.A., 2001. A system for automated seed vigour assessment. *Seed Sci. Technol.* 29, 625–636.
- Schulmerich, M. V., Walsh, M.J., Gelber, M.K., Kong, R., Kole, M.R., Harrison, S.K., McKinney, J., Thompson, D., Kull, L.S., Bhargava, R., 2012. Protein and oil composition predictions of single soybeans by transmission Raman spectroscopy. *J. Agric. Food Chem.* 60, 8097–8102. <https://doi.org/10.1021/jf301247w>
- Vakilian, K.A., 2020. Machine learning improves our knowledge about miRNA functions towards plant abiotic stresses. *Sci. Rep.* 10, 3041. <https://doi.org/10.1038/s41598-020-59981-6>
- Vischers, I.G.S., van Dam, N.M., Peters, J.L., 2018. An objective high-throughput screening method for thrips damage quantitation using Ilastik and ImageJ. *Entomol. Exp. Appl.* 166, 508–515. <https://doi.org/10.1111/eea.12682>
- Yates, S.C., Groeneboom, N.E., Coello, C., Lichtenthaler, S.F., Kuhn, P.-H.H., Demuth, H.-U.U., Hartlage-Rübsamen, M., Roßner, S., Leergaard, T., Kreshuk, A., Puchades, M.A., Bjaalie, J.G., 2019. QUINT: Workflow for Quantification and Spatial Analysis of Features in Histological Images From Rodent Brain. *Front. Neuroinform.* 13, 1–14. <https://doi.org/10.3389/fninf.2019.00075>
- Yordanov, Y.I., 2020. Hep G2 cell culture confluence measurement in phase-contrast micrographs—a user-friendly, open-source software-based approach. *Toxicol. Mech. Methods* 30, 146–152. <https://doi.org/10.1080/15376516.2019.1695303>
- Zhang, L., Liu, Z., Ren, T., Liu, D., Ma, Z., Tong, L., Zhang, C., Zhou, T., Zhang, X., Li, S., 2020. Identification of Seed Maize Fields With High Spatial Resolution and Multiple Spectral Remote Sensing Using Random Forest Classifier. *Remote Sens.* 12, 362. <https://doi.org/10.3390/rs12030362>

## **IJCROPSEED: UMA FERRAMENTA DE ACESSO ABERTO PARA ANÁLISE DE ALTO RENDIMENTO DE RADIOGRAFIAS DE SEMENTES DE CULTURAS**

### **RESUMO**

Tecnologias ópticas, capazes de analisar propriedades físicas de amostras biológicas estão ganhando crescente interesse na agricultura moderna. O uso de raios X para análise de propriedades internas de produtos agrícolas, como as sementes, pode fornecer informações sobre sua qualidade de maneira não-destrutiva. No entanto, as avaliações das imagens radiográficas de forma visual são morosas, subjetivas e altamente propensas à erros. Portanto, se faz necessário o desenvolvimento de métodos que possibilitem a realização dessas análises de forma eficiente e assertiva. Para esse fim, foi desenvolvida uma ferramenta de livre acesso, de código aberto e de fácil utilização para análise de alto rendimento de imagens radiográficas de sementes de várias culturas agrícolas, chamada IJCropSeed. Além disso, foi realizado um experimento, no qual foram desenvolvidos modelos de aprendizagem de máquina a partir das informações obtidas com a ferramenta para prever a capacidade de germinação das sementes e o vigor das plântulas de *Crambe abyssinica*. Os resultados revelaram que a IJCropSeed apresentou alto desempenho para análise de radiografias digitais das 24 culturas agrícolas avaliadas, com rapidez e elevada precisão de segmentação das imagens. O uso de parâmetros obtidos com a ferramenta, combinados com os modelos de aprendizado de máquina, mostrou-se altamente eficiente para classificar a qualidade de sementes de *C. abyssinica*, sendo um método não-destrutivo e altamente eficaz.

Palavras-chave: ImageJ; Fenotipagem de sementes. Aprendizado de máquina. Qualidade de sementes. Raios X.

## ABSTRACT

Optical technologies that can analyze physical properties of biological samples are increasingly drawing interest in modern agriculture. The use of X-rays for analysis of internal properties of agricultural products, such as seeds, has proven its worth in providing information regarding their quality in a non-destructive manner. However, visual evaluations of radiographic images are time-consuming, subjective, and highly prone to error. Therefore, it is necessary to develop methods that allow these analyses to be performed in an efficient and assertive manner. To that end, a free-access, open-source, and easy-to-use tool called IJCropSeed has been developed for high-throughput analysis of radiographic images of seeds from several agricultural crops. In addition, an experiment was conducted in which machine learning models were developed from the information obtained from the tool to predict the seed germination capacity and seedling vigor of *Crambe abyssinica*. The results showed that IJCropSeed had a high performance for the analysis of digital radiographic images of the 24 agricultural crops evaluated, with high speed and high precision of segmentation of the images. The use of parameters obtained with the tool, in combination with the machine learning models, proved to be highly efficient in classifying the quality of *C. abyssinica* seeds. It is a non-destructive and highly effective method.

Keywords: ImageJ. High throughput phenotyping. Machine Learning. Seed quality. X-ray.

## INTRODUÇÃO

As sementes das culturas agrícolas são estruturas detentoras de toda a tecnologia alcançada pelo melhoramento de plantas. Elas armazenam todo potencial genético produtivo das culturas, especialmente aquelas nas quais estão inseridos avanços da biotecnologia e as sementes de híbridos. No cenário de constante incerteza devido às mudanças climáticas, a qualidade das sementes é um fator chave para manutenção da produção agrícola, refletindo diretamente sobre a segurança alimentar (Finch-Savage and Bassel, 2016).

A qualidade das sementes inclui uma série de componentes complexos, relacionados a aspectos fisiológicos, físicos, bioquímicos e sanitários (ISTA, 2019). A avaliação precisa dos componentes de qualidade apresenta dificuldades, principalmente para aqueles relacionados a aspectos como viabilidade e vigor de sementes, devido ao tempo, a subjetividade e às restrições tecnológicas que envolvem a análise (ElMasry et al., 2019). A aplicação de tecnologias ópticas como a análise de raios X 2D, microtomografia computadorizada, ressonância magnética e imageamento multiespectral na inspeção da qualidade de sementes, tem reduzido algumas dessas limitações, oferecendo análises precisas e não destrutivas.

Nos últimos anos, a inspeção em sementes com uso de raios X 2D tem crescido, trazendo vantagens adicionais relacionadas a sua simplicidade de utilização e menor custo em comparação com outros métodos mais avançados (Medeiros et al., 2020a). Várias pesquisas com uso de raios X 2D demonstraram uma forte relação entre aspectos físicos internos das sementes com o seu potencial fisiológico (Dell'Aquila, 2009; Gagliardi and Marcos-Filho, 2011; Leão-Araújo et al., 2019; Medeiros et al., 2020a). A análise radiográfica quando feita de forma visual por analistas experientes ou com suporte computacional via métodos semiautomáticos oferecerem uma análise sensível, mas pode exigir mais tempo e esforços (Kotwaliwale et al., 2014). Neste sentido, o desenvolvimento de métodos de alto rendimento para análise de imagens radiográficas de sementes é necessário.

Softwares colaborativos de livre acesso e inclusivos para análise de imagens podem ser um possível caminho para implementação desse tipo de aplicação. O ImageJ é um clássico exemplo de sucesso de arquitetura de software aberta projetada para desenvolvimento de aplicações em diversas áreas do meio científico, oferecendo soluções para a prototipagem rápida de algoritmos de processamento de imagem (Schneider et al., 2012). A sua facilidade de uso, linguagem de macro gravável e arquitetura extensível de plugins tem possibilitado o desenvolvimento de ferramentas para análise de imagens neuro-histológicas (Timothy and Forlano, 2019), fenotipagem simultânea da mucilagem solúvel e parâmetros morfométricos de

sementes (Miart et al., 2018), quantificação histológica de seções de caule de milho (Legland et al., 2017), entre outras.

Avanços no campo da inteligência artificial tem dado suporte para a combinação de análise de imagens via softwares de livre acesso com métodos de aprendizagem de máquina para diversas finalidades. Recentemente, foi demonstrado a possibilidade de classificação de sementes *Jatropha curcas* L. quanto à sua qualidade com uso de dados de características fenotípicas obtidas por análise por imagem de raios X com abordagens de aprendizado de máquina (Medeiros et al., 2020a).

Diante disso, esta pesquisa teve como objetivo desenvolver uma ferramenta de livre acesso e de fácil utilização para análise de alto rendimento de imagens radiográficas de sementes de culturas agrícolas. Além disso, foi realizado um experimento, no qual foram desenvolvidos modelos de aprendizagem de máquina a partir das informações obtidas com a ferramenta para prever a capacidade de germinação das sementes e o vigor das plântulas individuais de *Crambe abyssinica* H.

## MATERIAL E MÉTODOS

Uma ferramenta de livre acesso para análise de alto rendimento de imagens radiográficas de sementes foi desenvolvida, chamada IJCropSeed. Atualmente, a ferramenta foi ajustada para analisar radiografias de sementes de 24 culturas agrícolas (Tabela 1). A IJCropSeed foi desenvolvida em linguagem macro do software ImageJ® (Schneider et al., 2012) para ser facilmente utilizada e passível de modificações em seu código, visando atender a demandas atuais e futuras. O código-fonte da IJCropSeed, documentação e os tutoriais em vídeo estão disponíveis gratuitamente em <https://sites.google.com/ufv.br/ijcropseed>. A IJSeedCrop é licenciada sob a GNU General Public License v3.0 (GPLv3), uma licença livre e de código aberto, com objetivo de compartilhar o código com a comunidade e incentivar a contribuição para o desenvolvimento.

**Tabela 1.** Identificação das culturas agrícolas inseridas na IJCropSeed.

ID	Nome comum	Scientific name <sup>1</sup>
1	Cabbage- Repolho	<i>Brassica oleracea</i> var. <i>capitata</i> L.
2	Cauliflower - Cauliflower	<i>Brassica oleracea</i> var. <i>botrytis</i> L.
3	Chicory - Chicória	<i>Cichorium intybus</i> L.
4	Coffee - Coffee	<i>Coffea</i> L.
5	Cowpea beans – Feijão	<i>Vigna unguiculata</i> (L.) Walp.
6	Crambe – Crambe	<i>Crambe abyssinica</i> Hochst. ex R.E.Fr.
7	Crotalaria – Crotalária	<i>Crotalaria juncea</i> L.
8	Lentil- Lentilha	<i>Lens culinaris</i> Medik.
9	Maize- Milho	<i>Zea mays</i> L.
10	Melon – Melão	<i>Cucumis melo</i> L.
11	Mung bean – Feijão mungo	<i>Vigna radiata</i> (L.) R.Wilczek
12	Okra – Quiabo	<i>Abelmoschus esculentus</i> Moench
13	Onion – Cebola	<i>Allium cepa</i> L.
14	Panicum grass – Grama Panicum	<i>Panicum maximum</i> Jacq.
15	Papaya – Mamão	<i>Carica papaya</i> L.
16	Pea – Ervilha	<i>Pisum sativum</i> L.
17	Pepper – Pimenta	<i>Capsicum chinense</i> Jacq.
18	Sorghum – Sorgo	<i>Sorghum bicolor</i> (L.) Moench
19	Soybean – Soja	<i>Glycine max</i> (L.) Merr.
20	Sunflower – Girassol	<i>Helianthus annuus</i> L.
21	Sweet corn – Milho doce	<i>Zea mays</i> L.
22	Tomato – Tomate	<i>Solanum lycopersicum</i> L.
23	Urochloa grass – Grama Urochloa	<i>Urochloa brizantha</i> (A.Rich.) R.D.Webster
24	Wheat - Trigo	<i>Triticum</i> L.

<sup>1</sup>Nomes científicos consultados no International Plant Names Index (IPNI- <https://www.ipni.org/>).

Neste artigo, a eficiência da IJCropSeed para realizar a análise de imagens radiográficas de 24 culturas (Tabela 1) foi avaliada individualmente. Ainda, a espécie *Crambe abyssinica* foi utilizada como modelo para pesquisas utilizando-se a análise das radiografias e os dados fisiológicos das sementes individuais.

### **Preparação da amostra e aquisição de imagens radiográficas**

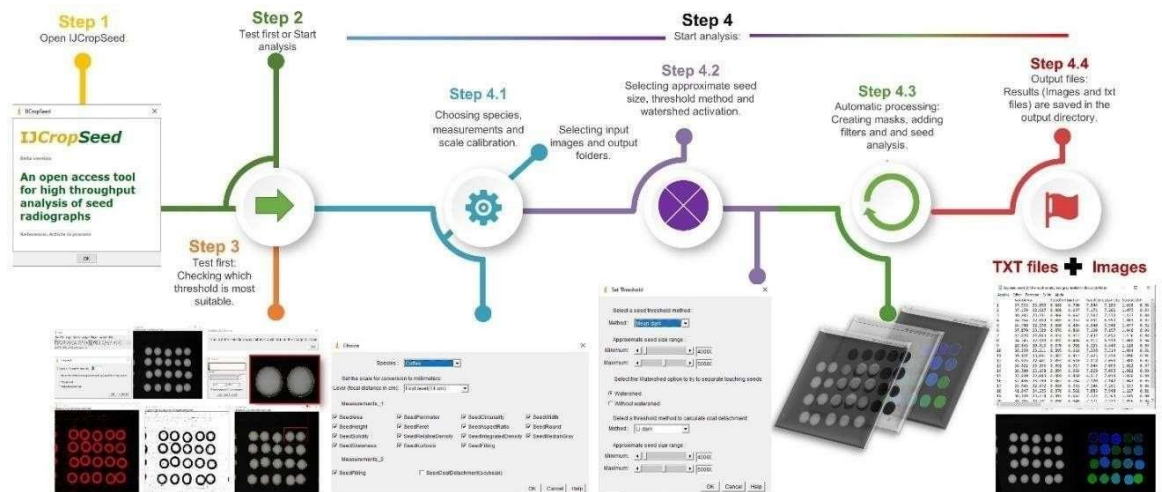
Para realizar a aquisição das imagens, as sementes de cada espécie foram fixadas previamente em papel adesivo, em grupos variando de 9 a 50 sementes, dependendo da espécie. As imagens radiográficas foram geradas por um dispositivo Faxitron, modelo MX-20 (Faxitron X-ray Corp. Wheeling, IL, EUA). O equipamento foi ajustado para uma tensão de 23kV e as sementes foram expostas à radiação por 10 segundos, a uma distância focal variando entre 30 e 41,6 cm. O contraste da imagem foi calibrado de acordo com cada espécie, de forma a aumentar o contraste da semente em relação ao fundo. As imagens foram salvas em arquivos no formato TIFF (*Tagged Image File Format*) e em seguida foram analisadas.

### **Visão geral do IJCropSeed**

A IJCropSeed é uma ferramenta escrita em linguagem macro do ImageJ® e foi desenvolvida para analisar em alto rendimento radiografias digitais de sementes de diversas espécies agrícolas. Nesta seção, apresentamos o pipeline da metodologia utilizada.

Para executar a macro, o script deve ser aberto no ImageJ® (*Plugins > Macros > Executar*). Em seguida, o arquivo “IJCropSeed.ijm” deve ser selecionado. Alternativamente, o arquivo pode ser arrastado e solto na janela do software ImageJ®, seguido do comando “Executar”.

A IJCropSeed usa apenas funções internas do ImageJ®, não sendo necessária a instalação de outros plugins para seu pleno funcionamento. O pipeline resumido do funcionamento da ferramenta proposta pode ser visualizado na Figura 1.



**Figura 1.** Esquema do fluxo de trabalho realizado pela IJCropSeed na análise de radiografias digitais de sementes.

Após a abertura da macro (Step 1), o usuário pode escolher entre duas opções (Step 2): Testar primeiro ou iniciar análise. A opção testar primeiro (Step 3) tem a finalidade de permitir que o usuário verifique de forma preliminar qual método de limiarização é mais adequado para a seleção da semente. Nesta etapa, o usuário é direcionado para a escolha do diretório e qual imagem será testada. Em seguida, é aplicado de forma automática um pré-processamento na imagem (suavização e reconhecimento de bordas) e uma caixa de diálogo é aberta. Na caixa de diálogo é requerido que o usuário selecione, com base no histograma, o método de limiarização mais adequado (17 métodos estão disponíveis; o método padrão é o “Mean dark”) para seleção da semente. Por fim, são aplicadas de forma automática operações matemáticas de morfologia na imagem (erosion e opening), resultando na criação de uma seleção envolvendo cada semente. Para checar se a seleção foi adequada, basta clicar na imagem e a ferramenta de aproximação será ativada.

Ao selecionar a opção iniciar análise, o Step 4 é ativado. Essa etapa permite a análise em alto rendimento de todas as imagens contidas na pasta de entrada. A etapa 4 pode ser entendida resumidamente em quatro sub-etapas. De início uma janela é aberta com opções de escolha da cultura a ser analisada (24 culturas estão atualmente cadastradas), definições de escala e descritores desejados (Step 4.1). Em definições de escala, oito opções estão ativas, a primeira opção mantém em unidade de pixels. As seis opções seguintes são baseadas em níveis da prateleira do dispositivo Faxitron utilizado nesta pesquisa, convertendo pixels em milímetros. Na última opção, o usuário define o valor para conversão de pixels em outra unidade. Na macro IJCropSeed, a tabela de resultados foi personalizada. Nela estão contidos 17 descritores relacionadas a morfometria e integridade tecidual das sementes (Tabela 2). Para

sementes de *C. abyssinica* é possível extrair até 29 descritores (15 descritores da semente e 14 descritores do embrião).

**Tabela 2.** Lista dos descritores gerados a partir da análise de imagens radiográficas das sementes.

<b>Trait (Característica)</b>	<b>Descrição</b>
<b>Medições 1</b>	
Área da Semente	Área (unidade escolhida): Área dentro do polígono definido pelo perímetro
Perímetro da Semente	Perímetro (unidade escolhida): Comprimento do limite externo da seleção
Circularidade da Semente	Circularidade: $Circularidade = 4 \cdot \pi \cdot Area / Perimeter^2$
Largura da Semente	Largura (unidade escolhida): Largura do menor retângulo que engloba a seleção.
Altura da Semente	Altura (unidade escolhida): Altura do menor retângulo que engloba a seleção.
Feret da Semente	Feret (unidade escolhida): A maior distância entre dois pontos ao longo do limite da seleção
Razão de Aspecto da Semente	Razão de aspecto = eixo maior/eixo menor
Redonda da Semente	Redonda = $4 \cdot Área / (\pi \cdot eixo maior^2)$
Solidez da Semente	Solidez = $Área / Área convexa$
Densidade Relativa da Semente	Cinza médio (cinza·unidade escolhida <sup>-1</sup> ): Valor médio de cinza dentro da seleção
Densidade Integrada da Semente	Densidade integrada (cinza·mm·pixel <sup>-1</sup> ): A soma dos valores de escala de cinza na partícula
Cinza Mediano da Semente	Cinza mediano (cinza·unidade escolhida <sup>-1</sup> ): Mediana de escala de cinza
Assimetria da Semente	Assimetria: O terceiro momento em relação à média
Curtose da Semente	Curtose: O quarto momento em relação à média
Preenchimento da Semente	Preenchimento (%): Determinado pela relação entre os espaços vazios e preenchidos da seleção após a imagem ser binarizada pelo limiar inicialmente definido.
<b>Medições 2</b>	
Preenchimento da Semente	Preenchimento (%): Duas máscaras são criadas, uma selecionando um método de limiar que seleciona pixels com valores mais altos na seleção e outra capaz de selecionar toda a semente, incluindo buracos. No final, as máscaras binarizadas

	são subtraídas e o resultado é a porcentagem de tecidos com a maior densidade (cinza definido pelo primeiro método) dentro da semente.
Desprendimento do Tegumento da Semente (soja)	Desprendimento do Tegumento (%): Porcentagem do espaço vazio dentro das sementes. Para as sementes de soja, é chamado de desprendimento de tegumento.

---

Após as definições iniciais, a pasta de entrada, em que estão contidas as imagens radiográficas, e a pasta de saída, local no qual os resultados serão gerados, devem ser selecionadas. Na sequência, uma caixa de diálogo é aberta (Step 4.2) para que seja selecionado o método de limiarização mais adequado para as sementes, a faixa de tamanho aproximado da semente (baseada na escala selecionada, Step 4.1), e uma opção de *Watershed* está disponível para separação de sementes tocantes. Após esse procedimento, inicia-se o processamento automático das imagens (Step 4.3).

Nesta última etapa, a IJCropSeed lida com todo o processo de análise. As imagens são processadas em lote, o que torna o método mais fácil, rápido e intuitivo para o usuário. O processo resume-se em um pré-processamento das imagens utilizando a aplicação de filtros de suavização, reconhecimento e maximização de bordas, binarização e aplicação de operações morfológicas (erosão e abertura). Após o pré-processamento, as regiões de interesse (neste caso, as sementes) são identificadas na imagem binária e analisadas com redirecionamento para a imagem original.

Por fim, para cada radiografia analisada, são gerados como resultado um arquivo texto (no formato TXT) e uma imagem (no formato TIF) na pasta previamente escolhida como saída. No arquivo texto estão contidos todos os descritores selecionados para cada semente individualmente. No arquivo de imagem, está contida a radiografia original no plano esquerdo e a imagem processada no plano direito com as sementes individualmente identificadas.

A IJCropSeed executa os processos de segmentação de uma maneira intuitiva. Na imagem processada é possível visualizar que ela atribui um número específico e uma cor à cada semente analisada, permitindo ao usuário pesquisar facilmente os parâmetros medidos nas tabelas de resultados para cada semente individualmente.

A maioria das etapas principais da IJCropSeed estão interligadas, ou seja, ao final de cada processo, o usuário poderá ser redirecionado para a opção anterior e escolher outro caminho, sem o abortamento da macro.

## **Validação de performance**

Para validar a segmentação usando a IJCropSeed, as imagens radiográficas de sementes das 24 espécies para as quais foi desenvolvida foram avaliadas manualmente na plataforma Fiji, com uso da ferramenta *Wand*. Posteriormente, os dados obtidos por meio da análise manual e com a utilização da IJCropSeed foram analisados por meio regressões lineares e gráficos de pontos de dispersão para obtenção da correlação entre as duas metodologias. O tempo despendido nas análises das radiografias de cada espécie pelo método manual e pela IJCropSeed foi registrado. O software R 3.6.1 (R Core Team, 2019) foi utilizado nas análises estatísticas.

## **Experimento com espécie modelo (*C. abyssinica*)**

Visando demonstrar a aplicação da ferramenta IJCropSeed na análise de qualidade de sementes, foi conduzido um experimento com sementes de *C. abyssinica*. Foram utilizadas sementes com diferentes níveis de qualidade fisiológica. As sementes foram produzidas em campo experimental localizado no município de Viçosa, Estado de Minas Gerais, Brasil (20°46'01.7"S 42°52'05.6"W) entre os anos de 2012 e 2016. Neste experimento, foram utilizados os descritores obtidos com a análise da radiografia para desenvolvimento de modelos de aprendizagem de máquina aplicados na predição da capacidade de germinação das sementes e o vigor das plântulas.

## **Aquisição e processamento das imagens**

As sementes foram fixadas em papel adesivo, em grupos de 50 sementes. Este procedimento foi realizado para permitir a identificação individual de cada semente nas análises posteriores, sendo que cada semente foi analisada individualmente.

As imagens radiográficas foram geradas por um dispositivo Faxitron, modelo MX-20 (Faxitron X-ray Corp. Wheeling, IL, EUA). O equipamento foi ajustado para uma tensão de 23kV e as sementes foram expostas à radiação por 10 segundos, a uma distância focal de 41,6 cm. O contraste da imagem foi calibrado para 16383 (width) x 3124 (center). As imagens foram salvas em arquivos no formato TIFF (*Tagged Image File Format*) e em seguida foram analisadas no software ImageJ® com uso da IJCropSeed.

## **Análises fisiológicas**

As mesmas sementes utilizadas na análise de raios X foram avaliadas quanto à sua qualidade fisiológica. Para tanto, foi conduzido o teste de germinação, utilizando-se caixas

plásticas transparentes (11 x 11 x 3 cm), com duas folhas de papel para germinação no fundo, umedecidos com quantidade de água equivalente a 2,5 vezes a massa do papel seco. O conjunto foi mantido em germinador a 25°C, por sete dias. O número de sementes germinadas foi contabilizado diariamente. Foram consideradas germinadas as sementes que apresentaram raiz primária maior que 2 mm. Após a última contagem, as sementes foram classificadas em sementes germinadas e sementes não germinadas (as que não apresentaram protrusão radicular até 7 dias de avaliação e/ou não viáveis). As plântulas geradas foram classificadas em normais (as que apresentaram todas as estruturas bem desenvolvidas) e anormais (plântulas mal desenvolvidas ou com defeitos graves).

### **Classificação da qualidade das sementes**

Os descritores obtidos com a análise das radiografias das sementes foram utilizados para desenvolver modelos de aprendizagem de máquina para classificação de sementes em diferentes classes, com base em seus dados de qualidade fisiológica. As seguintes classes foram estabelecidas: i) sementes germinadas e ii) sementes não germinadas. Em seguida, as plântulas originadas de sementes germinadas foram classificadas em i) plântulas normais, apresentando todas as estruturas bem desenvolvidas; e ii) plântulas anormais, apresentando algum defeito ou má-formação.

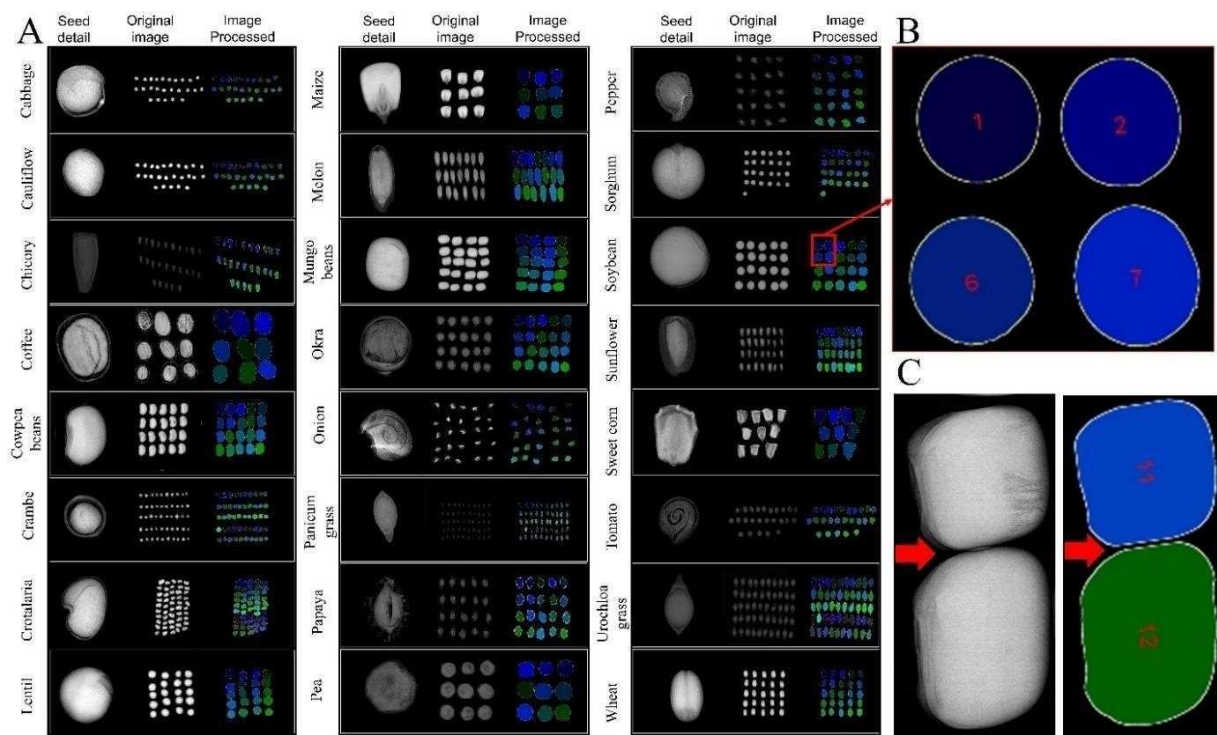
As sementes foram divididas em conjuntos de treinamento e validação. Neste estudo, foi utilizado dois métodos de aprendizagem de máquina, sendo eles a Análise Discriminante Linear (LDA) e o Random Forest. Os modelos foram desenvolvidos utilizando o conjunto de treinamento e foram validados por meio de validação cruzada e utilizando um conjunto de validação independente, que não foi utilizado durante a construção dos modelos. Os dois classificadores foram implementados em linguagem R 3.6.1 (R Core Team, 2019). Na LDA foi utilizado o pacote Mass, e para o Random Forest foi utilizada a biblioteca randomForest, com 500 árvores de decisão e hiperparâmetros padrão.

Para o desenvolvimento dos modelos, os dados obtidos a partir das imagens radiográficas foram arranjados em uma matriz X (os preditores) e os dados resultantes do teste de germinação no vetor Y (variável resposta). Neste estudo, os modelos foram avaliados por meio das métricas de acurácia, Kappa.

## RESULTADOS

### Processamento das imagens de raios X

As configurações utilizadas no aparelho de raios X foram adequadas para gerar as imagens com bom contraste entre as sementes e o fundo, além de gerar um bom detalhamento das estruturas internas das sementes da maioria das culturas testadas. Conforme ilustrado na Figura 2, a ferramenta IJCropSeed processou as radiografias digitais de mais de 20 culturas agrícolas (Tabela 1). Na imagem é possível observar as variações de formas, tamanhos e estruturas internas das sementes de cada cultura em detalhes.



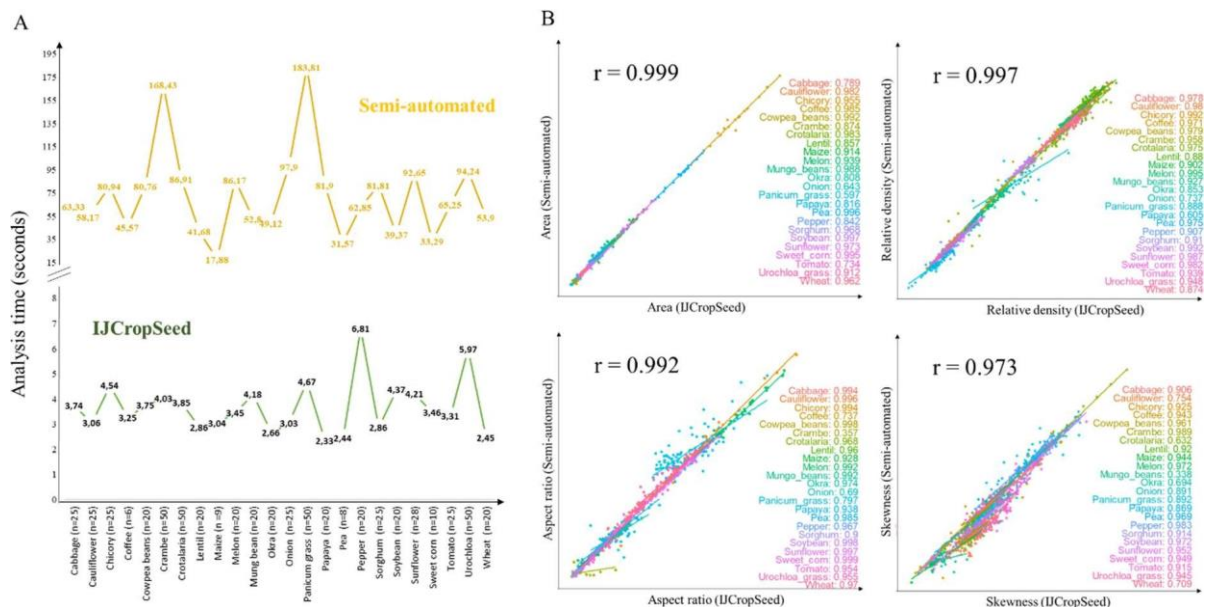
**Figura 2.** Radiografias digitais originais e processadas de sementes de 24 culturas agrícolas analisadas pela IJCropSeed (A). Detalhes da imagem de sementes de soja processadas (B) e detalhe da separação de sementes de feijão mungo tocantes.

Observa-se que na imagem processada as sementes foram coloridas e a elas foram atribuídas um número (Figura 2B). Nas imagens em que as sementes foram alinhadas horizontalmente e verticalmente, a numeração se deu de forma crescente da esquerda para direita e de cima para baixo. No caso de sementes tocantes, a opção *watershed* funcionou bem em sementes com formato convexo suave e que não se sobrepuseram muito, conforme ilustra a Figura 2C em sementes de *Vigna radiata*.

## Desempenho e validação da IJCropSeed

Para avaliar o desempenho da IJCropSeed, seus resultados foram comparados com o método semiautomático utilizando a ferramenta *wand* da plataforma Fiji. Pelo método semiautomático, as sementes foram selecionadas individualmente, o que permitiu um ajuste mais refinado da seleção.

A IJCropSeed apresentou desempenho superior quanto à velocidade de análise. O tempo médio para a análise de uma radiografia pela IJCropSeed foi 3.68s (processador Intel Core i5-4200U CPU 1.60GHz RAM 4.00GB), contra 73.0s gasto pelo método semiautomático (Figura 3A). O tempo de análise pelo método semiautomático variou principalmente em função do número de sementes na radiografia. Já a variação observada pela IJCropSeed se deu tanto pelo número de indivíduos, quanto pelo número de grupos de parâmetros avaliados, uma vez que cada análise é feita de forma independente.



**Figura 3.** Validação estatística da eficiência da macro IJCropSeed para análise de imagens radiográficas de sementes. Ensaio de tempo de análise de processamento (A), análise de correlação das medições com a IJCropSeed e pelo método semiautomático para *area*, *aspect ratio*, *densidade relativa* e *assimetria* (B). Os coeficientes de correlação de Pearson foram determinados por meio de análises de regressão linear (n =1202, correspondente a todas as sementes).

Para validar a eficiência da IJCropSeed para a segmentação das imagens radiográficas das sementes, todos os parâmetros foram medidos em todas as sementes utilizando a macro e correlacionados com medições manuais utilizando a plataforma Fiji (Figura 3B). Na figura 3B quatro descritores foram utilizados para representar o grau de associação entre os métodos, sendo dois relacionadas a parâmetros morfométricos (*area* e *aspect ratio*) e dois relacionadas à

integridade tecidual (*densidade relativa* e *assimetria*). Correlações significativas foram obtidas para todos os parâmetros ( $P < 0.001$ ), com fortes correlações entre os dados obtidos de forma semiautomática e por meio da macro para *area* ( $r = 0.99$ ), *aspect ratio* ( $r = 0.99$ ), *densidade relativa* ( $r = 0.99$ ) e *assimetria* ( $r = 0.97$ ).

### **Experimento com sementes de *Crambe abyssinica***

Para a classificação baseada na germinação de sementes e vigor de plântulas de *C. abyssinica*, foram observadas diferenças nas médias entre as classes para a maioria dos descritores. Observou-se que sementes das classes germinada e vigorosa obtiveram médias mais elevadas para parâmetros relacionados ao nível de cinza na radiografia, constatada pelos descritores densidade relativa, densidade integrada, cinza mediano e preenchimento. Além disso, a mesma tendência foi observada para descritores morfométricos do embrião (e.g. área, largura, altura, *feret*, *round*).

### **Modelos de aprendizado de máquina para classificação de sementes**

Foram desenvolvidos quatro diferentes modelos a partir da Análise Discriminante Linear (LDA) e Random Forest visando à classificação das sementes de *C. abyssinica* com base na capacidade de germinar e no vigor das plântulas.

### **Classificação entre sementes germinadas e não germinadas**

Os modelos de classificação desenvolvidos para discriminar as sementes quanto a sua capacidade de germinar foram estabelecidos para identificar se descritores morfométricos e de integridade tecidual das sementes, obtidos por meio da análise de radiografias digitais, podem ser utilizados para prever a qualidade fisiológica das sementes. Os resultados da classificação pelos modelos desenvolvidos com a LDA e Random forest são apresentados na Tabela 3 para os dados de treinamento, validação cruzada e validação independente.

**Tabela 3.** Matrizes de confusão dos modelos LDA e Random Forest para participação em classes de sementes de *C. abyssinica* 'Germinadas' e 'Não Germinadas', em conjuntos de treinamento, validação cruzada e validação independente.

Método	Característica	Treinamento	Validação cruzada	Validação independente
		(n =344)		(n =156)
		Acertos (Total)		
LDA	Germinada	259 (261)	-	23(33)
	Não-Germinada	66(83)	-	119(123)
	Acurácia	0.94	0.92±0.04	0.91
	Kappa	0.83	0.77±0.16	0.71
Random Forest	Germinada	261(261)	-	30 (33)
	Não-Germinada	83(83)	-	117(123)
	Acurácia	1	0.93±0.03	0.94
	Kappa	1	0.80±0.10	0.83

Os modelos de LDA e Random forest obtidos apresentaram uma acurácia superior a 90% para todos os conjuntos de dados, sendo considerado altamente preciso para classificar sementes germinadas e não germinadas de *C. abyssinica*.

### Classificação entre plântulas normais e anormais

Os modelos construídos para classificar sementes de *C. abyssinica* capazes de produzir plântulas normais e anormais revelou uma precisão moderada. Os resultados demonstram uma precisão de classificação em torno de 80% para validação cruzada em ambos os métodos de aprendizagem de máquina. Na validação independente, o método *Random forest* foi ligeiramente superior a LDA na precisão de classificação (0.76) (Tabela 4).

**Tabela 4.** Matrizes de confusão dos modelos LDA e *Random Forest* para partição em classes de sementes de *C. abyssinica* que originaram plântulas normais e plântulas anormais.

Método	Característica	Treinamento	Validação cruzada	Validação independente
		(n =344)		(n =156)
		Acertos (Total)		
LDA	Plântulas normais	182(194)	-	74(83)
	Plântulas anormais	113(150)	-	41(73)
	Acurácia	0.85	0.80±0.06	0.73
	Kappa	0.7	0.59±0.12	0.46
Random Forest	Plântulas normais	194 (194)	-	69(83)
	Plântulas anormais	150 (150)	-	51(73)
	Acurácia	1	0.80±0.04	0.76
	Kappa	1	0.60±0.09	0.53

## DISCUSSÃO

A segurança alimentar e a qualidade de produtos agrícolas são uma preocupação constante em todo o mundo (El-Mesery et al., 2019). As sementes, como o principal insumo agrícola, têm um efeito profundo na uniformidade de desenvolvimento e rendimento da produção agrícola. Portanto, é essencial que as sementes tenham alta qualidade (Xia et al., 2019). A qualidade das sementes, pode estar intimamente relacionada à sua estrutura interna (Du et al., 2019). A análise de raios X é um método não destrutivo amplamente usado para avaliar as estruturas internas das sementes, sendo indicado pela *International Seed Testing Association* (ISTA, 2019). No entanto, o processo de análise visual da radiografia pode ser considerado moroso, sujeito à erros e impraticável em larga escala (Medeiros et al., 2020b). Neste trabalho, nós apresentamos uma ferramenta de livre acesso, fácil utilização e de código aberto para automatização das análises de imagens radiográficas de sementes de culturas agrícolas, chamada IJCropSeed.

Os resultados revelaram que a macro IJCropSeed é altamente satisfatória para análise de radiografias digitais de várias culturas agrícolas. A ferramenta possui um tempo de análise expressivamente menor e uma precisão de segmentação elevada, quando comparado ao método semiautomático (Figura 3). Esses resultados são importantes, pois indicam alto potencial de aplicação da ferramenta para estudos de fenotipagem de características morfométricas e de integridade tecidual de sementes. O uso de ferramentas de fenotipagem de alto rendimento, baseadas em análise de imagem de sementes abre uma gama de possibilidades para análises mais profundas, como as anatômicas, metabolômicas e genômicas (Jahnke et al., 2016).

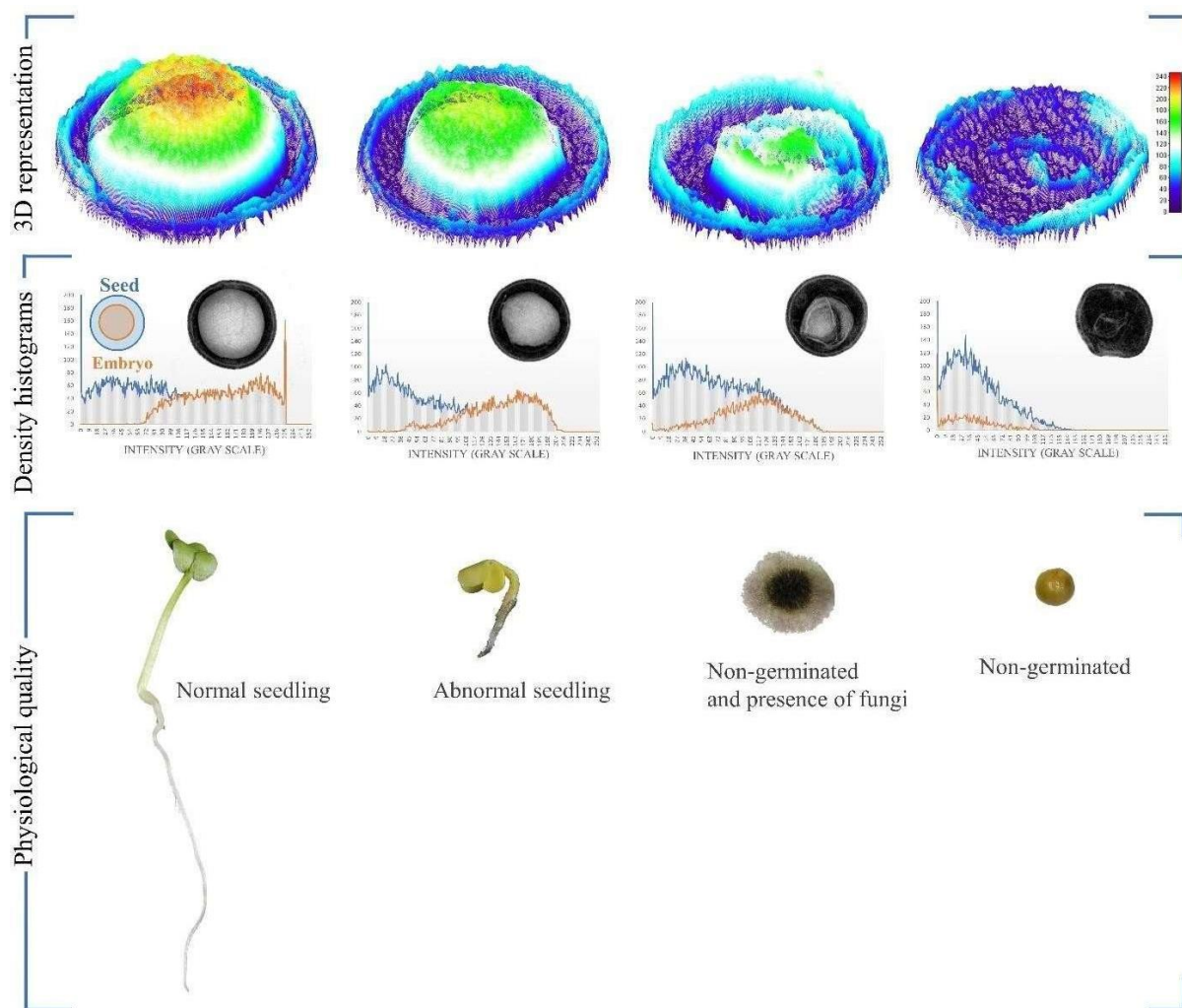
A linguagem macro do ImageJ<sup>®</sup>, na qual a IJCropSeed foi escrita, é uma linguagem de programação processual interpretada cuja estrutura e sintaxe lógica se assemelham ao C, C++ e Java. Ela foi escolhida por oferecer uma série de vantagens a comunidade interessada. Alguns dos benefícios ao utilizá-la são: O usuário não precisa ter experiência com métodos de programação orientada a objetos; a linguagem é versátil e poderosa, e apresenta sintaxe acessível a programadores iniciantes; a linguagem contém uma biblioteca de centenas de funções internas (Timothy and Forlano, 2019).

Uma das características mais interessantes da IJCropSeed é que o usuário pode incorporar mais funções internas do ImageJ<sup>®</sup>, e estruturar ainda mais seu código com instruções de controle e suas próprias funções, oferecendo infinitas possibilidades de aplicação. O código atualmente contempla métodos de processamento para radiografias digitais de sementes de 24 culturas agrícolas (Figura 2). O algoritmo desenvolvido foi pensado como a combinação dos

princípios globais de segmentação e baseado em limites, projetado para que um usuário iniciante possa implementar a segmentação e processamento de outras culturas agrícolas seguindo a mesma lógica de programação das culturas já inseridas, enriquecendo ainda mais a ferramenta, ou oferecendo base para criação de outras.

Com objetivo de demonstrar a aplicação prática da utilização dos parâmetros providos da IJCropSeed, foram desenvolvidos modelos de aprendizagem de máquina baseados em LDA e Random Forest para prever a capacidade de germinação e o vigor de plântulas individuais de *C. abyssinica*. Com base nos resultados obtidos (Tabelas 3 e 4) foi possível constatar elevadas acurácias, indicando a possibilidade de utilização dos parâmetros físicos obtidos com a IJCropSeed para classificar sementes de *C. abyssinica* quanto a sua capacidade de germinação e o vigor de suas plântulas. As variações dos parâmetros físicos obtidos com a análise de raios X entre sementes individuais podem ser atribuídas a alterações nas propriedades morfológicas e anatômicas das sementes, que implicam na menor resistência à passagem de raios X, em sementes de baixa qualidade, gerando imagens com níveis mais baixos de cinza, reflexo da sua menor integridade tecidual (Medeiros et al., 2020a).

Os parâmetros que mais contribuíram para o desenvolvimento dos modelos, em geral, foram aqueles relacionados à integridade tecidual das sementes e embriões (e.g. densidade relativa da semente e embrião, densidade integrada, assimetria, preenchimento), embora parâmetros morfométricos do embrião (e.g. área, circularidade, solidez) também tenham tido importância considerável (material suplementar 2). A Figura 4 mostra a representação da densidade tecidual avaliada nas radiografias das sementes de *C. abyssinica*, e sua associação com aspectos fisiológicos relacionados à germinação e vigor de sementes. Nela, observa-se que a semente com maior densidade tecidual da semente e embrião (concentração de pixels em níveis mais altos na escala de cinza) apresentou plântula bem desenvolvida. Por outro lado, foi observado que quanto menor a densidade dos tecidos, intensificou-se a redução do potencial germinativo das sementes e o vigor das plântulas.



**Figura 4.** Associação entre densidade de tecidos com a qualidade de sementes de *C. abyssinica*. Representação em cores 3D dos níveis de cinza de cada pixel das radiografias, histogramas de distribuição de cinza nas sementes e nos embriões, suas respectivas plântulas, e sementes não germinadas.

É importante ressaltar que a relação entre parâmetros físicos e fisiológicos nem sempre é observada, visto que as sementes como organismos biológicos complexos podem sofrer influência de outros fatores (Medeiros et al., 2020a). No entanto, diante dos promissores resultados observados no presente trabalho, novas perspectivas são abertas para estudos adicionais da associação de parâmetros obtidos com técnicas de alto rendimento baseadas em análise de imagens com processos mais complexos ligados a fisiologia, bioquímica e genética em sementes.

Na tecnologia de sementes, esses achados são muito valiosos, uma vez que a identificação rápida de sementes de boa e má qualidade contribuem para otimização do processamento e controle de qualidade da produção de sementes (Huang et al., 2015; Rahman and Cho, 2016). A implicação prática disso é que sementes com alta qualidade, capazes de gerar

plântulas vigorosas, promovem um estabelecimento bem sucedido do estande da plantas no campo, pois estão menos sujeitas a estresses bióticos e abióticos, o que garante maiores ganhos ao produtor (Finch-Savage and Bassel, 2016).

Os métodos de aprendizagem de máquina utilizados nesse trabalho também foram utilizados com êxito em pesquisas recentes. A LDA foi aplicada em investigações taxonômicas de ameixas (Sarigu et al., 2017), caracterização e diversidade de germoplasma de maçã (Sau et al., 2018, 2019) e em classificação da qualidade de sementes de *Jatropha curcas* L. (Medeiros et al., 2020a). Resumidamente, a abordagem da LDA baseia-se na classificação que visa maximizar a razão entre a variação entre classes e a variação dentro da classe, com a finalidade de alcançar a máxima separabilidade (Maione and Barbosa, 2018). O *Random Forest*, por sua vez, foi empregado em estudos de predição de amadurecimento de mamão (Santos Pereira et al., 2018), na avaliação dos níveis de suficiência de clorofila em dendezeiros maduros ( *Elaeis guineensis* ) (Amirruddin et al., 2020), e em estudos com sistema de imagem multiespectral combinado com quimiometria para avaliar a qualidade de sementes de melancia (Liu et al., 2019). O *Random Forest* opera construindo uma infinidade árvores de decisão na etapa do treinamento e produzindo a classe das árvores individuais. No momento da previsão, em esquema de votação, é obtida uma classe de cada árvore de decisão, e ao final o algoritmo seleciona a classificação mais votada sobre todas as árvores (Santos Pereira et al., 2018). A combinação entre métodos de aprendizagem de máquina com análise de imagens radiográficas de alto rendimento para análise de sementes traz novas e promissoras perspectivas para o setor e representa um novo marco do avanço tecnológico para a indústria sementeira.

A ferramenta IJCropSeed promoveu uma análise rápida e precisa das radiografias digitais de sementes de várias culturas agrícolas. O uso de parâmetros obtidos com a ferramenta combinados com os modelos de aprendizado de máquina mostrou-se altamente eficiente para classificar em diferentes níveis de qualidade sementes de *C. abyssinica*. Por fim, o método avaliado é não-destrutivo e altamente eficaz.

## CONCLUSÕES

A IJCropSeed é uma ferramenta de código aberto eficiente para análises de alto rendimento de radiografias digitais de sementes de várias culturas agrícolas. Parâmetros gerados pela ferramenta oferecem suporte para análises morfométricas e de integridade tecidual das sementes. Os modelos de aprendizagem de máquina desenvolvidos a partir de dados obtidos com a IJCropSeed apresentaram elevadas acurácias para classificar sementes de *C. abyssinica* quanto ao seu potencial fisiológico.

## REFERÊNCIAS

- Amirruddin, A.D., Muharam, F.M., Ismail, M.H., Ismail, M.F., Tan, N.P., Karam, D.S., 2020. Hyperspectral remote sensing for assessment of chlorophyll sufficiency levels in mature oil palm (*Elaeis guineensis*) based on frond numbers: Analysis of decision tree and random forest. *Comput. Electron. Agric.* 169, 105221.  
<https://doi.org/10.1016/j.compag.2020.105221>
- Dell'Aquila, A., 2009. Development of novel techniques in conditioning, testing and sorting seed physiological quality. *Seed Sci. Technol.* 37, 608–624.  
<https://doi.org/10.15258/sst.2009.37.3.10>
- Du, Z., Hu, Y., Ali Buttar, N., Mahmood, A., 2019. X-ray computed tomography for quality inspection of agricultural products: A review. *Food Sci. Nutr.* 7, 3146–3160.  
<https://doi.org/10.1002/fsn3.1179>
- El-Mesery, H., Mao, H., Abomohra, A., 2019. Applications of Non-destructive Technologies for Agricultural and Food Products Quality Inspection. *Sensors* 19, 846.  
<https://doi.org/10.3390/s19040846>
- ElMasry, G., Mandour, N., Wagner, M.-H., Demilly, D., Verdier, J., Belin, E., Rousseau, D., 2019. Utilization of computer vision and multispectral imaging techniques for classification of cowpea (*Vigna unguiculata*) seeds. *Plant Methods* 15, 24.  
<https://doi.org/10.1186/s13007-019-0411-2>
- Finch-Savage, W.E., Bassel, G.W., 2016. Seed vigour and crop establishment: extending performance beyond adaptation. *J. Exp. Bot.* 67, 567–591.  
<https://doi.org/10.1093/jxb/erv490>
- Gagliardi, B., Marcos-Filho, J., 2011. Relationship between germination and bell pepper seed structure assessed by the X-ray test. *Sci. Agric.* 68, 411–416.

<https://doi.org/10.1590/S0103-90162011000400004>

- Huang, M., Wang, Q.G.G., Zhu, Q.B.B., Qin, J.W.W., Huang, G., 2015. Review of seed quality and safety tests using optical sensing technologies. *Seed Sci. Technol.* 43, 337–366. <https://doi.org/10.15258/sst.2015.43.3.16>
- ISTA, 2019. International rules for seed testing. International Seed Testing Association, Bassersdorf.
- Jahnke, S., Roussel, J., Hombach, T., Kochs, J., Fischbach, A., Huber, G., Scharr, H., 2016. *pheno Seeder* - A Robot System for Automated Handling and Phenotyping of Individual Seeds. *Plant Physiol.* 172, 1358–1370. <https://doi.org/10.1104/pp.16.01122>
- Kotwaliwale, N., Singh, K., Kalne, A., Jha, S.N., Seth, N., Kar, A., 2014. X-ray imaging methods for internal quality evaluation of agricultural produce. *J. Food Sci. Technol.* 51, 1–15. <https://doi.org/10.1007/s13197-011-0485-y>
- Leão-Araújo, É.F., Gomes-Junior, F.G., da Silva, A.R., Peixoto, N., de Souza, E.R.B., 2019. Evaluation of the desiccation of *campomanesia adamantium* seed using radiographic analysis and the relation with physiological potential. *Agron. J.* 111, 592–600. <https://doi.org/10.2134/agronj2018.05.0302>
- Legland, D., El-Hage, F., Méchin, V., Reymond, M., 2017. Histological quantification of maize stem sections from FASGA-stained images. *Plant Methods* 13, 1–11. <https://doi.org/10.1186/s13007-017-0225-z>
- Liu, W., Xu, X., Liu, C., Zheng, L., 2019. Rapid discrimination of high-quality watermelon seeds by multispectral imaging combined with chemometric methods. *J. Appl. Spectrosc.* 85, 1044–1049. <https://doi.org/10.1007/s10812-019-00757-w>
- Maione, C., Barbosa, R.M., 2018. Recent applications of multivariate data analysis methods in the authentication of rice and the most analyzed parameters: A review. *Crit. Rev. Food Sci. Nutr.* 0, 1–12. <https://doi.org/10.1080/10408398.2018.1431763>
- Medeiros, A.D. de, Pinheiro, D.T., Xavier, W.A., Silva, L.J. da, Dias, D.C.F. dos S., 2020a. Quality classification of *Jatropha curcas* seeds using radiographic images and machine learning. *Ind. Crops Prod.* 146, 112162. <https://doi.org/10.1016/j.indcrop.2020.112162>
- Medeiros, A.D. de, Zavala-León, M.J., Silva, L.J. da, Oliveira, A.M.S., Dias, D.C.F. dos S., 2020b. Relationship between internal morphology and physiological quality of pepper seeds during fruit maturation and storage. *Agron. J.* <https://doi.org/10.1002/agj2.20071>
- Miart, F., Fontaine, J.-X., Pineau, C., Demailly, H., Thomasset, B., Van Wuytswinkel, O., Pageau, K., Mesnard, F., 2018. MuSeeQ, a novel supervised image analysis tool for the

- simultaneous phenotyping of the soluble mucilage and seed morphometric parameters. *Plant Methods* 14, 112. <https://doi.org/10.1186/s13007-018-0377-5>
- R Core Team, 2019. R Development Core Team. *R A Lang. Environ. Stat. Comput.* <https://doi.org/http://www.R-project.org>
- Rahman, A., Cho, B.-K.K., 2016. Assessment of seed quality using non-destructive measurement techniques: A review. *Seed Sci. Res.* 26, 285–305. <https://doi.org/10.1017/S0960258516000234>
- Santos Pereira, L.F., Barbon, S., Valous, N.A., Barbin, D.F., 2018. Predicting the ripening of papaya fruit with digital imaging and random forests. *Comput. Electron. Agric.* 145, 76–82. <https://doi.org/10.1016/j.compag.2017.12.029>
- Sarigu, M., Grillo, O., Lo Bianco, M., Uccesu, M., d’Hallewin, G., Loi, M.C., Venora, G., Bacchetta, G., 2017. Phenotypic identification of plum varieties (*Prunus domestica* L.) by endocarps morpho-colorimetric and textural descriptors. *Comput. Electron. Agric.* 136, 25–30. <https://doi.org/10.1016/j.compag.2017.02.009>
- Sau, S., Uccesu, M., D’hallewin, G., Bacchetta, G., 2019. Potential use of seed morpho-colourimetric analysis for Sardinian apple cultivar characterisation. *Comput. Electron. Agric.* 162, 373–379. <https://doi.org/10.1016/j.compag.2019.04.027>
- Sau, S., Uccesu, M., Dondini, L., De Franceschi, P., D’hallewin, G., Bacchetta, G., 2018. Seed morphometry is suitable for apple-germplasm diversity-analyses. *Comput. Electron. Agric.* 151, 118–125. <https://doi.org/10.1016/j.compag.2018.06.002>
- Schneider, C.A., Rasband, W.S., Eliceiri, K.W., 2012. NIH Image to ImageJ: 25 years of image analysis. *Nat. Methods* 9, 671–675. <https://doi.org/10.1038/nmeth.2089>
- Timothy, M., Forlano, P.M., 2019. A versatile macro-based neurohistological image analysis suite for ImageJ focused on automated and standardized user interaction and reproducible data output. *J. Neurosci. Methods* 324, 108286. <https://doi.org/10.1016/j.jneumeth.2019.04.009>
- Xia, Y., Xu, Y., Li, J., Zhang, C., Fan, S., 2019. Recent advances in emerging techniques for non-destructive detection of seed viability: A review. *Artif. Intell. Agric.* 1, 35–47. <https://doi.org/10.1016/j.aiia.2019.05.001>

# **MARCADORES FENOTÍPICOS MULTIESPECTRAIS COMBINADOS COM ALGORITMOS DE APRENDIZADO DE MÁQUINA PARA CLASSIFICAÇÃO DE LOTES DE SEMENTES DE SOJA**

## **RESUMO**

O reconhecimento de padrões espectrais em imagens biológicas apoiado por tecnologia óptica avançada e algoritmos de inteligência artificial abre novas oportunidades para assegurar a qualidade das sementes comercializadas na cadeia agrícola global, em especial para cultura da soja. Nesse estudo, foi investigado o uso de marcadores fenotípicos obtidos a partir de imagens multiespectrais para avaliação do potencial fisiológico de sementes de soja. Além disso, foram desenvolvidos e avaliados diferentes modelos de inteligência artificial para classificação automática das sementes em níveis de potencial fisiológico. Foram utilizadas 19 bandas espectrais aplicadas em mais de 16 mil sementes de 18 cultivares de soja, produzidas em duas safras no Brasil. Os modelos desenvolvidos alcançaram alta performance, com acurácias superiores a 90%. Os resultados obtidos nos oferecem uma visão entusiástica do uso do imageamento multiespectral combinado com algoritmos de inteligência artificial como ferramenta primária para identificação acurada do potencial fisiológico de sementes de soja.

Palavras-chave: Método não-destrutivo. Análise de imagens. Espectroscopia. Qualidade de sementes. Redes Neurais.

## **ABSTRACT**

The recognition of spectral patterns in biological images, supported by advanced optical technology and artificial intelligence algorithms, offers up new opportunities to ensure the quality of seeds traded in the global agricultural supply chain, especially for soybean cultivation. In this study, the use of phenotypic markers derived from multispectral images for assessing the physiological potential of soybean seeds was investigated. Furthermore, various artificial intelligence models for the automatic classification of seeds based on their physiological potential were developed and evaluated. 19 spectral bands were used on more than 16,000 seeds from 18 soybean cultivars, produced in two harvests in Brazil. The developed models achieved high performance, with accuracies exceeding 90%. The results offer us an enthusiastic view of the use of multispectral imaging combined with artificial intelligence algorithms as a primary tool for accurate identification of the physiological potential of soybean seeds.

**Keywords:** Non-destructive method. Image analysis. Spectroscopy. Seed quality. Neural Networks.

## INTRODUÇÃO

A soja é uma das principais culturas agrícolas globais e esforços contínuos são feitos para potencializar sua produtividade (Ainsworth et al., 2012). A produtividade é fortemente influenciada pela qualidade das sementes, crucial para o estabelecimento de plântulas (Finch-Savage e Bassel, 2016). Apesar dos testes tradicionais avaliarem aspectos como pureza, germinação e vigor, eles enfrentam desafios como lentidão, falta de padronização, interpretação subjetiva e mão de obra para realizá-los (ElMasry et al., 2019a).

O uso de novas abordagens utilizando tecnologias modernas de imagem combinadas com técnicas de inteligência artificial tem sido cada vez mais utilizado para ajudar na avaliação da qualidade de sementes e superar as limitações enfrentadas com o uso dos testes tradicionais (ElMasry et al., 2019b; Xia et al., 2019). Dentre as técnicas emergentes em grande potencial está o imageamento multiespectral. A aplicação de imagens multiespectrais em sementes de soja pode ser feita analisando as diferentes faixas espectrais de luz que os tecidos das sementes interagem. Uma vez que cada semente apresenta características particulares de estrutura celular e composição química, a luz é absorvida e refletida de maneira única em cada indivíduo. Ao capturar imagens em diferentes comprimentos de onda, é possível identificar essas diferenças e usá-las para classificar as sementes de acordo com vários atributos de qualidade, incluído pureza (Shrestha et al., 2015; Wilkes et al., 2016), sanidade (Boelt et al., 2018; França-Silva et al., 2020; Olesen et al., 2011; Rego et al., 2020a), integridade física e qualidade fisiológica (Barboza da Silva et al., 2021; Batista et al., 2022; Sendin et al., 2018).

Modelos avançados de aprendizado de máquina, como Redes Neurais, Support Vector Machine (SVM) e Random Forest, estão sendo utilizados e validados para a modelagem e previsão do potencial fisiológico das sementes (Baek et al., 2019a; Barboza da Silva et al., 2021; de Medeiros et al., 2020). Grande parte desses algoritmos têm a capacidade de identificar relações tanto lineares quanto não lineares, atingindo notável precisão em classificações. Estes avanços tecnológicos e analíticos estão contribuindo para o desenvolvimento de abordagens mais precisas e confiáveis na agricultura, em especial para a área de tecnologia de sementes (ElMasry et al., 2020).

Pesquisas recentes têm apresentado maiores evidências do potencial da técnica para seleção de sementes de soja de alto potencial fisiológico. Resultados experimentais demonstraram que sistema hiperespectral foi altamente preciso (> 95%) para determinar a viabilidade de sementes de soja (Baek et al., 2019b). Em outro estudo, com a utilização de imagens de autofluorescência multiespectral foi possível classificar sementes de soja em

diferentes níveis de envelhecimento acelerado artificial, com destaque para intensidade da autofluorescência no comprimento de onda de excitação de 365 e 660 nm, atribuída principalmente às clorofilas (Barboza da Silva et al., 2021). Correlações significativas entre variáveis de reflectância e aspectos fisiológicos de germinação e vigor também foram observadas, principalmente na faixa do infravermelho, obtidas com sistema multiespectral de baixo custo (Soares et al., 2023). Apesar das evidências científicas existentes, ainda há muitas lacunas a serem preenchidas no que diz respeito à aplicação da técnica. É preciso investigar sua aplicação de uma forma mais ampla, englobando diferentes genótipos e lotes, incluindo diferentes safras, a fim de obter e validar modelos de classificação mais abrangentes.

Neste sentido, o objetivo deste trabalho foi investigar o uso de imagens multiespectrais combinado com algoritmo de aprendizado de máquina para classificação do potencial fisiológico de sementes de soja advindas de diferentes genótipos, lotes e safras. Nele também buscamos explorar como as características espectrais das sementes de soja podem ser relacionadas com suas características físicas e fisiológicas, e como essas informações podem ser usadas para desenvolver métodos de avaliação de qualidade de sementes de soja mais precisos e confiáveis.

## **MATERIAL E MÉTODOS**

### **Amostra**

Neste estudo foram utilizados 84 lotes de 17 cultivares comerciais de soja (HO APOREÍ IPRO, HO JURUENA IPRO, HO CORUMBA IPRO, MONSOY 8349 IPRO, MONSOY 8644 IPRO, SYN 1687 IPRO, TMG2776IPRO, BMX EXTREMA IPRO, BMX OLIMPO IPRO, BMX TORMENTA CE, BMX BÔNUS IPRO, BMX DOMÍNIO IPRO, BMX DESAFIO RR, BMX FORTALEZA IPRO, BMX VORAZ IPRO, BMX FOCO IPRO, BMX ÚNICA IPRO), totalizando 16.474 sementes individuais avaliadas. As amostras foram obtidas de lotes produzidos nas safras 2020/21 e 2021/22 na região do Cerrado Brasileiro.

### **Imageamento multiespectral**

O imageamento multiespectral foi aplicado usando um equipamento VideometerLab4™. As imagens foram capturadas de cada semente em 19 comprimentos de onda. Este sistema captura e combina imagens multiespectrais de alta resolução (2192 × 2192 pixels). As imagens foram segmentadas e 19 variáveis foram mensuradas para cada semente -

reflectância em 365 (UV), 405 (violeta), 430 (índigo), 450 (azul), 470 (azul), 490 (ciano), 515 (verde), 540 (verde), 570 (amarelo), 590 (âmbar), 630 (vermelho), 645 (vermelho), 660 (vermelho), 690 (vermelho escuro), 780 (vermelho escuro), 850, 880, 940 e 970 nm (infravermelho próximo).

### **Análises fisiológicas**

O potencial fisiológico dos lotes foi avaliado por meio dos testes de tetrazólio (TZ viabilidade, TZ vigor), germinação (G) e envelhecimento acelerado (EA), conforme descritos abaixo:

O teste de tetrazólio foi feito com duas repetições de 50 sementes, pré-condicionadas em papel toalha por 16 h a 25 °C em germinador de sementes. Em seguida, as sementes foram colocadas em copos plásticos e submersas em 50 mL de solução de tetrazólio (2-3-5, cloreto de trifênil tetrazólio) a 0,075 % e mantidas a 40 °C por 3,5 h em BOD no escuro. Após o processo de coloração, as sementes foram lavadas com água e classificadas individualmente quanto à viabilidade e vigor, de acordo com os critérios descritos por França-Neto e Krzyzanowski (2020).

Para avaliar a germinação das sementes de soja, foi utilizado o método de rolo de papel toalha. O teste foi conduzido com 8 repetições de 50 sementes cada, colocadas em rolos de papel toalha umedecidos e mantidos a 25 °C em germinador com 8 horas de luz por dia. A porcentagem média de germinação foi registrada no 8º dia após a semeadura, seguindo as recomendações da ISTA (International Seed Testing Association).

O teste de envelhecimento acelerado foi realizado com quatro repetições de 50 sementes. As sementes foram colocadas em uma tela dentro de uma caixa plástica tipo "gerbox" contendo 40 ml de água. A caixa foi mantida em incubadora BOD a 41 °C por 48 horas. Após esse período, as sementes foram colocadas para germinar em condições padrão, a 25 °C, conforme descrito no teste de germinação. A avaliação da germinação foi feita no 5º dia após a semeadura (Marcos-Filho, 2020).

### **Classes de sementes**

Baseando-se no potencial fisiológico do lote, as sementes foram rotuladas em três classes – Alta qualidade, média qualidade e descarte. As classes foram definidas com base nos dados obtidos das análises fisiológicas por técnicos treinados, realizados na empresa de sementes que forneceu os lotes.

## **Análise de dados e modelos de aprendizado de máquina**

As imagens multiespectrais coletadas em diferentes comprimentos de onda foram convertidas inicialmente em matrizes de dados com aproximadamente ½ milhão de observações para as análises. Neste estudo, a análise multivariada de componentes principais (PCA) foi aplicada como uma análise exploratória, a fim de identificar possíveis padrões nos dados extraídos das imagens, passíveis de associação com aspectos de qualidade das sementes. A PCA foi executada usando o software R e os dois principais componentes foram visualizados em gráficos de dispersão bidimensionais. Além disso, a proporção de variação explicada por cada componente principal foi calculada para determinar a importância relativa de cada componente na representação dos dados.

Na sequência, foram desenvolvidos os modelos de aprendizado de máquina baseados em Rede Neural (Rede de baixa complexidade – Modelo padrão com 2 camadas ocultas e 50 neurônios sem otimização exaustiva de hiperparâmetros, Rede de alta complexidade – Modelo com otimização exaustiva de hiperparâmetros), Árvore de Decisão e Random Forest para classificar as sementes quanto ao potencial fisiológico. Para o desenvolvimento dos modelos para classificação das sementes foi adotado o seguinte *workflow* de processamento:

- 1) Separação dos dados em conjuntos de treinamento (90%) e teste (10%). Ao usar apenas 10% dos dados para teste, foi possível avaliar a capacidade de generalização do modelo para novos dados sem comprometer a quantidade de dados disponíveis para treinamento.
- 2) Aplicação da validação cruzada para avaliar o modelo com diferentes subconjuntos dos dados de treinamento.
- 3) Otimização dos hiperparâmetros do modelo, experimentando diferentes combinações de hiperparâmetros e escolhendo a combinação que resultou em melhor performance.
- 4) Cálculo da performance média do modelo com base nas avaliações realizadas na validação cruzada.
- 5) Escolha do modelo com a melhor performance média e treinar novamente com todos os dados de treinamento.
- 6) Avaliação do modelo final com os dados de teste, para avaliar sua capacidade de generalização para novos dados em nível de semente e lote.

A avaliação dos modelos de *machine learning* foi realizada com base na matriz de confusão e no cálculo de medidas de desempenho, como a Acurácia, Coeficiente Kappa, Sensibilidade e Especificidade, conforme equações abaixo:

$$Acurácia = \frac{\sum_{i=1}^n 1(y_i = \hat{y}_i)}{n}$$

Onde  $n$  é o número total de previsões,  $y_i$  é a classe real da amostra  $i$ , e  $\hat{y}_i$  é a classe prevista pelo modelo. A função  $1$  é uma função indicadora que retorna 1 se a condição dentro dos parênteses é verdadeira e 0 caso contrário.

$$Kappa = \frac{p_o - p_e}{1 - p_e}$$

Onde  $p_o$  é a proporção de previsões corretas realizadas pelo modelo e  $p_e$  é a proporção de acertos esperados baseado na distribuição das classes no conjunto de dados.

$$Sensibilidade = \frac{TP}{TP + FN}$$

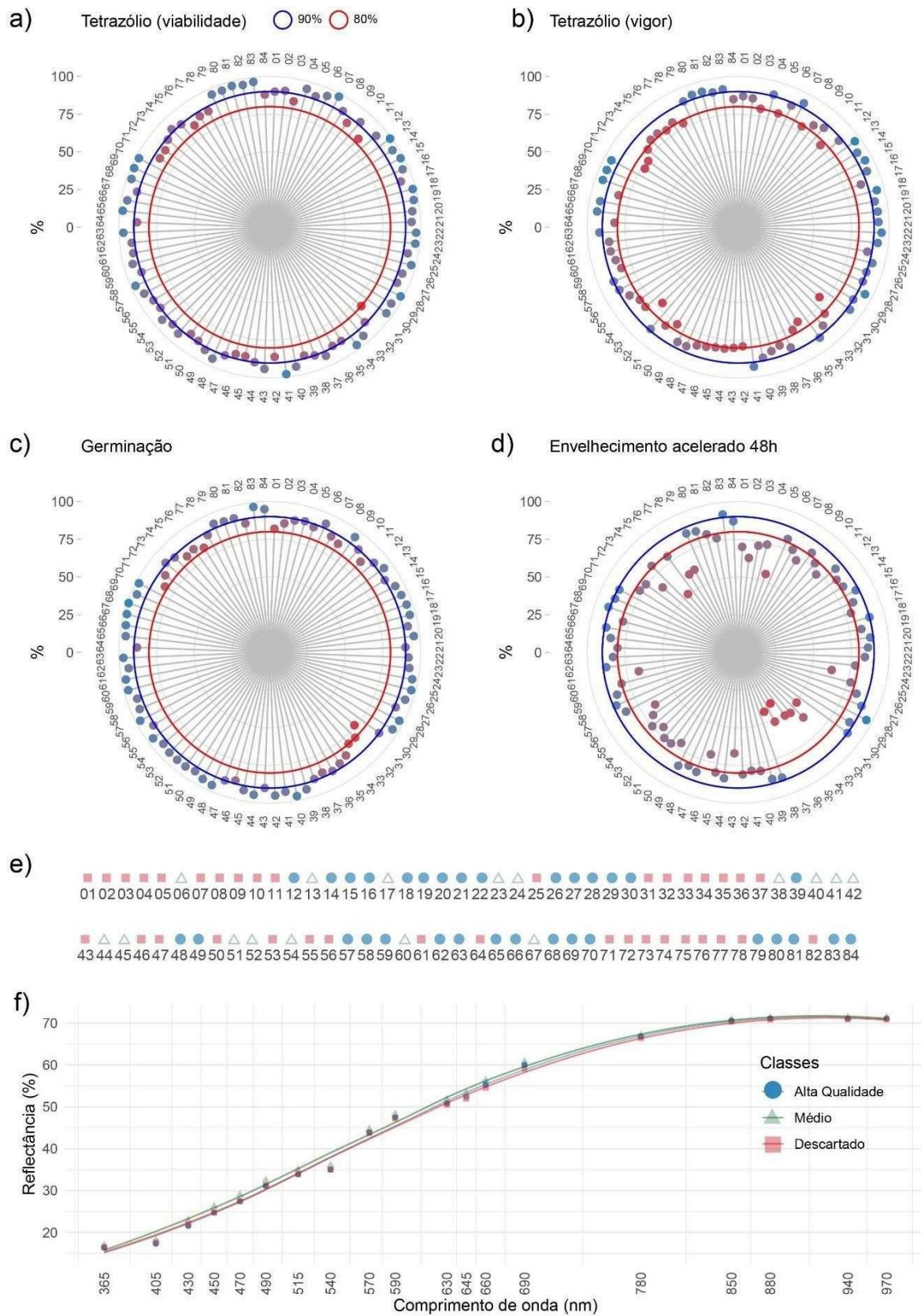
$$Especificidade = \frac{TN}{TN + FP}$$

Onde TP representa verdadeiros positivos, TN representa verdadeiros negativos, FP representa falsos positivos e FN representa falsos negativos.

## RESULTADOS

### **Análise fisiológica**

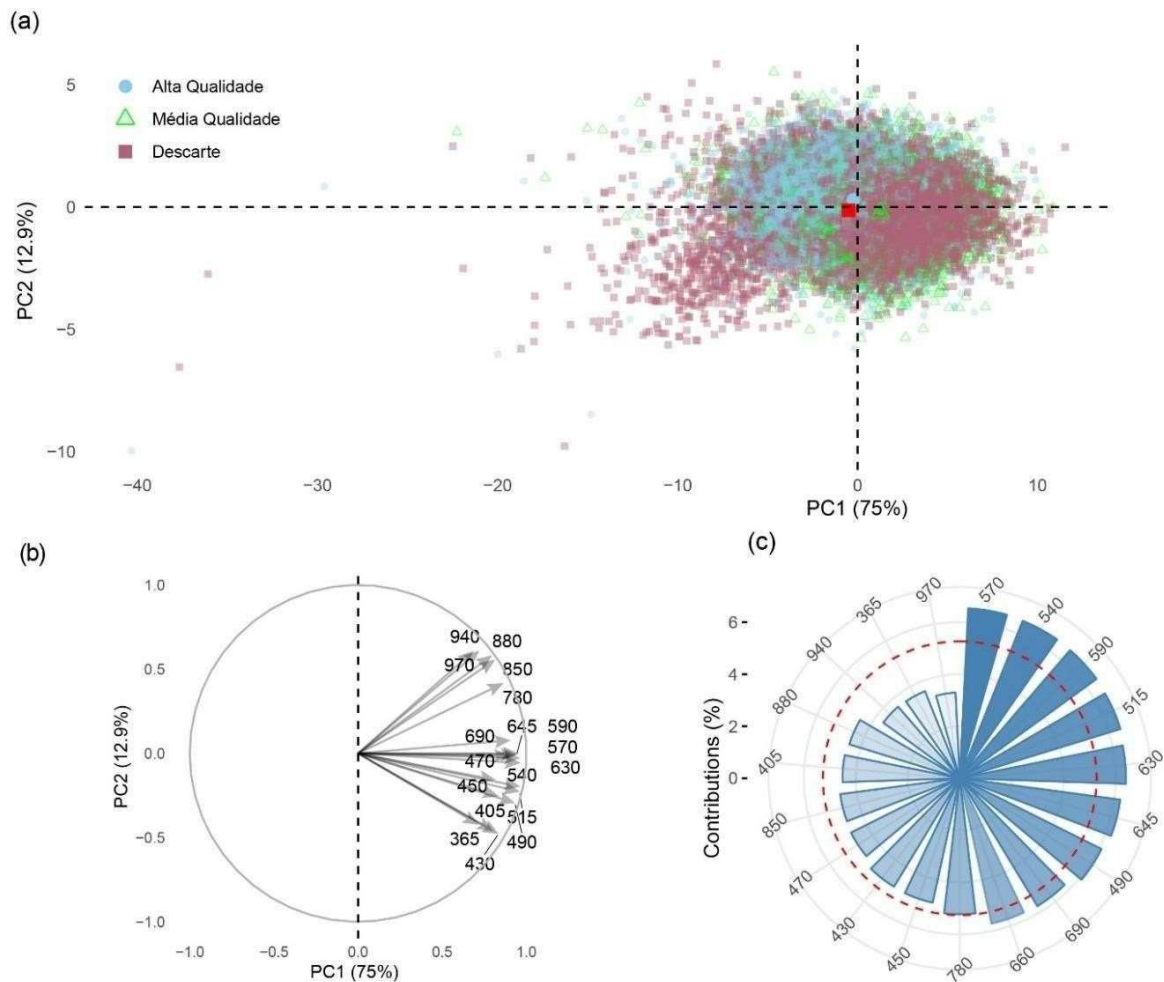
Os lotes apresentaram diferentes níveis de potencial fisiológico baseado nos testes de germinação e vigor aplicados (Figura 1). Na Figura 1a e b observa-se que a maioria dos lotes atingiram valores de viabilidade superiores a 90% e níveis de vigor superiores a 80% baseado no teste de tetrazólio. Para germinação também foi observado elevado desempenho dos lotes com apenas um único lote com padrão inferior a 80% (Figura 1 c). Diferentes dos demais testes, os resultados do envelhecimento acelerado amplificaram as diferenças entre os lotes, com valores que variaram de 63 a 98% de germinação (Figura 1 d). Na Figura 1e observa-se a classificação atribuída para cada lote de acordo com seu desempenho fisiológico. Com relação aos valores de reflectância nas diferentes faixas do espectro, de uma forma geral, observa-se que houve uma leve diferença entre as médias espectrais de cada classe de potencial fisiológico, com ênfase principalmente na região entre 500 e 600 nm (Figura 1.f).



**Figura 1.** Variáveis fisiológicas obtidas a partir dos testes de tetrazólio (a,b), germinação (c) e Envelhecimento acelerado (d). Classes de qualidade associadas aos lotes de sementes (e) e seu respectivo comportamento espectral em 19 bandas multiespectrais.

## Análise multivariada de componentes principais

Por meio da aplicação da análise de componentes principais, notou-se significativa sobreposição no diagrama de ordenação central entre as sementes que constituíam os lotes. Apesar de aproximadamente 88% da variabilidade dos dados serem explicados pelos primeiros dois componentes, não se conseguiu discernir um padrão de categorização evidente entre as sementes das respectivas classes. O círculo de correlação (Figura 2b) permitiu constatar que as sementes com valores mais elevados de reflectância estavam majoritariamente localizadas no lado direito do diagrama de ordenação central. Adicionalmente, os comprimentos de onda que mais influenciaram os componentes principais foram identificados como sendo 570, 540, 590, 515, 630, 645, 490, 690, 660 nm (Figura 2).

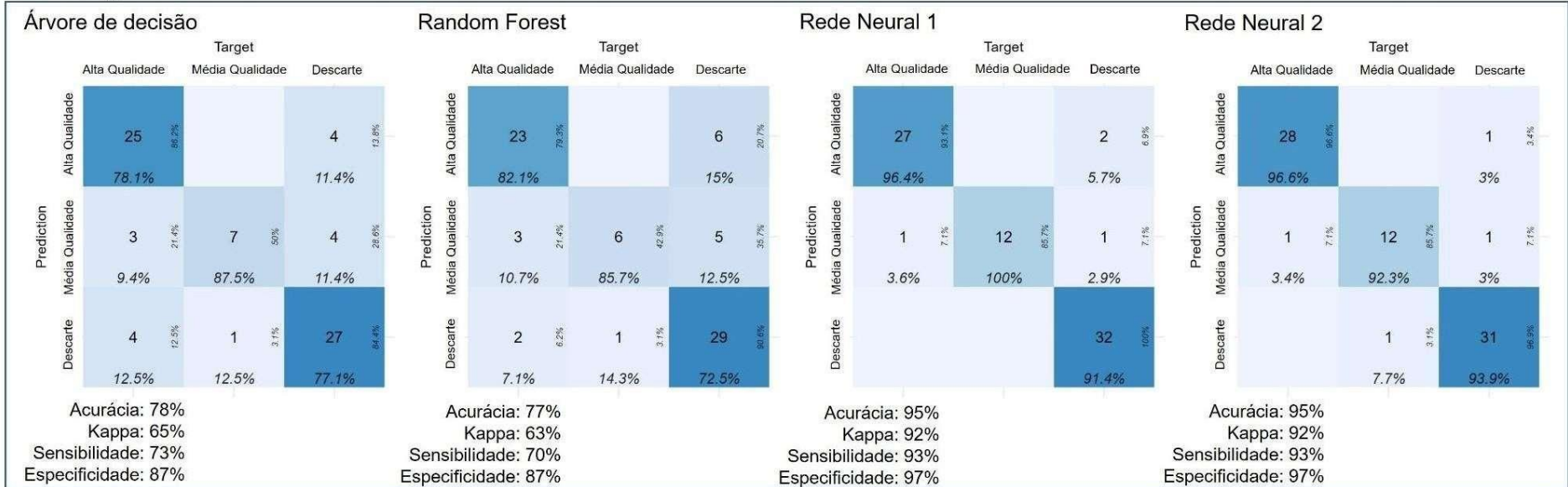


**Figura 2.** Diagrama de ordenação central (a), círculo de correlações (b) e importância das variáveis para o componente 1(c) da análise de componentes principais.

### **Modelos de aprendizado de máquina**

Os descritores obtidos a partir da análise multiespectral das sementes foram utilizados para desenvolver quatro modelos de aprendizado de máquina (Figura 3). Na validação cruzada foram observadas acurácias elevadas ( $\geq 95\%$ ) para os modelos de rede neural de baixa complexidade (Rede Neural 1) e de alta complexidade (Rede Neural 2). O coeficiente Kappa também se mostrou elevado, bem como as demais métricas de sensibilidade e especificidade ( $\geq 92\%$ ). Os modelos baseados em árvores (Árvore de decisão e *Random Forest*) apresentaram menor desempenho.

# Training



# Test

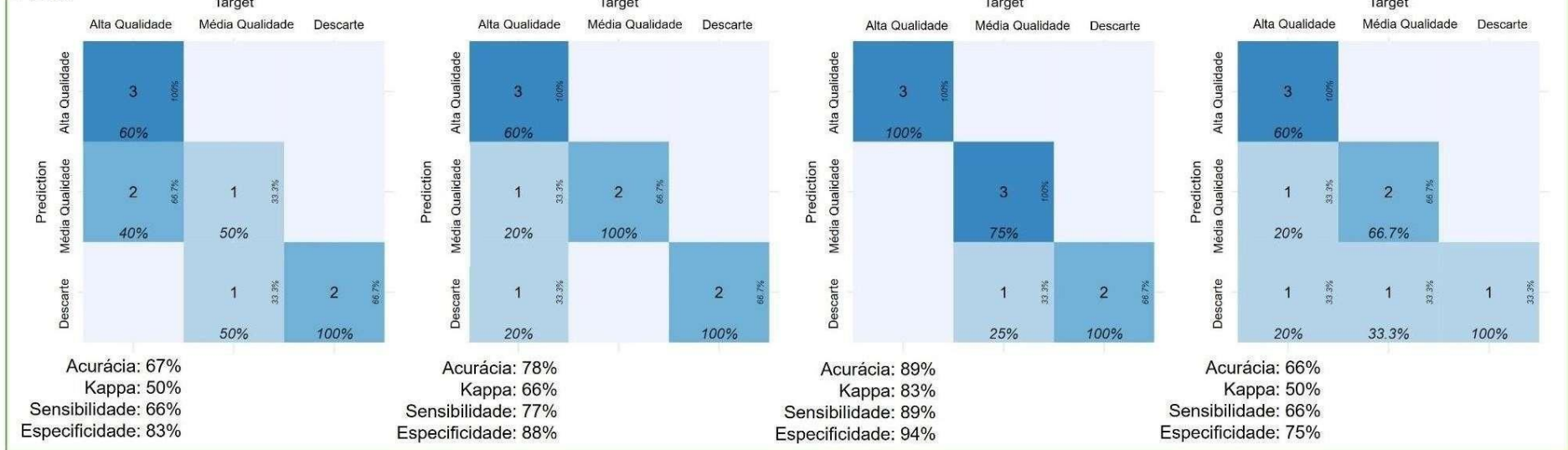


Figura 3. Performance dos modelos de aprendizado de máquina na classificação de lotes de sementes de soja.

Os resultados observados nos dados de teste mostraram superioridade das redes neurais de baixa complexidade, mantendo um alto nível de acurácia (89%), com altas métricas de Kappa (83%), sensibilidade (89%) e especificidade (94%). Os modelos baseados em árvore de decisão mantiveram a performance intermediária observada no treinamento e, por outro lado, as redes neurais de alta complexidade apresentaram baixo desempenho.

## DISCUSSÃO

A análise multiespectral é uma ferramenta poderosa na avaliação da composição química de diferentes materiais, incluindo sementes (ElMasry et al., 2019a; Tsakanikas et al., 2015). Neste estudo, a aplicação da análise multiespectral foi utilizada para avaliar a qualidade de sementes de soja advindas de diferentes lotes, genótipos e anos de produção, com o objetivo de determinar o seu potencial fisiológico aplicado à tomada de decisão sobre o destino dos lotes de sementes.

Por meio da análise multivariada de componentes principais (PCA), observou-se que vários comprimentos de onda na faixa do visível, mais especificamente em 490, 515, 540, 570, 590, 630, 645, 660 e 690 nm, revelaram-se essenciais para a interpretação da variabilidade dos dados. Estes comprimentos de onda abrangem região espectral que vai do azul ao vermelho no espectro. A luz nestas regiões é substancialmente modulada pela presença de clorofila e outros pigmentos que têm picos de absorção na região do azul (cerca de 400-500 nm) e vermelho (cerca de 600-700 nm) (Barboza da Silva et al., 2021; Buschmann et al., 2000; Galletti et al., 2020). Em sementes, estes comprimentos de ondas podem ser úteis para avaliar a maturidade e a presença de pigmentos que podem estar associados à qualidade da semente. Além disso, muitos carotenoides (pigmentos vegetais amarelo-laranja) absorvem fortemente nessa faixa (Batista et al., 2022). Em soja, o principal carotenoide é a luteína, no entanto há também quantidades menores de  $\beta$ -caroteno e zeaxantina, que desempenham um papel crucial na proteção celular (Gebregziabher et al., 2023; Monma et al., 1994). Esses carotenoides atuam como antioxidantes, protegendo as sementes de soja contra danos oxidativos e garantindo a preservação da integridade e viabilidade da semente durante o armazenamento e a germinação (Lee et al., 2013; Monma et al., 1994). Além dos carotenoides e da clorofila, é possível considerar a importância das antocianinas no contexto da análise espectral de sementes. Os comprimentos de onda específicos mencionados, especialmente aqueles na faixa do azul (cerca de 400-500 nm) e vermelho (cerca de 600-700 nm), são relevantes para a absorção de luz pelas antocianinas (Barboza da Silva et al., 2021). Essa absorção pode ser um indicador da presença e concentração de antocianinas nas sementes, oferecendo insights sobre sua maturidade,

qualidade e viabilidade (Batista et al., 2022). Além disso, as propriedades antioxidantes das antocianinas podem contribuir significativamente para a integridade e longevidade das sementes.

Os resultados mostraram alta eficiência e potencial da técnica de imageamento multiespectral combinado aos modelos de *machine learning* aplicados. Foram obtidas elevadas acurácias (> 90%) com um forte nível de concordância do coeficiente Kappa (90%) para classificar as sementes em suas respectivas classes de qualidade. Isso permite que os produtores e multiplicadores de sementes possam selecionar as sementes de melhor qualidade para comercialização e estabeleçam preços justos sobre os lotes comercializados, o que aumenta a eficiência do sistema produtivo. Além disso, os modelos criados podem ser aplicados em laboratórios de análise de rotina de sementes, ou mesmo em pesquisas.

Dentre os modelos criados, as redes neurais artificiais foram as que alcançaram maior desempenho, em especial a rede de baixa complexidade que obteve maior capacidade de generalização. Esse modelo, baseado em sistemas neurológicos, consiste em várias camadas de nós, cada uma responsável por extrair características complexas a partir dos dados de entrada (LeCun et al., 2015). O algoritmo de treinamento, conhecido como *backpropagation*, ajusta os pesos das conexões entre os nós para minimizar a diferença entre a saída prevista e a saída real (Goodfellow et al., 2016). Em outras palavras, a rede aprende a associar as medidas de refletância espectral a uma classificação de qualidade da semente ajustando os pesos das conexões entre os seus neurônios, de modo a minimizar o erro entre as saídas previstas e as classificações conhecidas dos dados de treinamento. Em outras abordagens de classificação utilizando redes neurais, pesquisadores demonstraram o seu potencial para diferenciar cultivares (Shrestha et al., 2015; Wilkes et al., 2016), identificar níveis de deterioração em sementes (Barboza da Silva et al., 2021; ElMasry et al., 2019b; Soares et al., 2023), identificar padrões de maturação fisiológica (Batista et al., 2022), identificar precocemente contaminação patogênica (França-Silva et al., 2020; Rego et al., 2020b), entre outros.

Neste estudo, avaliamos vários lotes de sementes com um bom nível de balanceamento entre as classes, o que é uma boa base para generalização. Embora alguns testes tradicionais sejam necessários para atender aos padrões oficiais (Brasil, 2009; ISTA, 2023), a abordagem utilizada neste trabalho pode ajudar a gerar laudos de qualidade mais rápidos, precisos, menos subjetivos, não sendo necessário destruir a amostra para obter o resultado. No entanto, embora tenhamos obtido resultados promissores, encontramos algumas limitações. Ao atribuir o rótulo de lote (classe de vigor) a todas as sementes presentes, resultou em uma classificação

inconsistente a nível individual, pois sementes não germinadas foram rotuladas como de alta qualidade em lotes de alta qualidade, e sementes germinadas foram rotuladas como descarte quanto o lote era de baixa qualidade. Isso prejudica a precisão do treinamento. Para melhorar a precisão, em trabalhos futuros será realizada a classificação individual de cada semente, atribuindo um rótulo único a cada uma delas. Além disso, será avaliada a aplicação de técnicas de *deep learning* em uma próxima abordagem.

É importante destacar que este é um estudo inovador, que usa tecnologia óptica multiespectral avançada e aprendizado de máquina para classificar o potencial fisiológico de sementes de soja, sendo potencialmente disruptivo para o setor sementeiro e importante para o fortalecimento da agricultura 5.0.

## CONCLUSÕES

Neste estudo, a combinação de imagens multiespectrais com algoritmos de aprendizado de máquina demonstrou ser uma abordagem promissora na classificação do potencial fisiológico de sementes de soja, considerando diferentes genótipos, lotes e safras. O algoritmo de rede neural artificial, especificamente, apresentou uma alta performance, com uma acurácia de 89% na classificação. Os comprimentos de onda 490, 515, 540, 570, 590, 630, 645, 660 e 690 nm foram fundamentais na análise e interpretação da variabilidade dos dados. Esses comprimentos de onda, localizados principalmente na região do espectro que vai do azul ao vermelho são fortemente influenciados por clorofilas e carotenoides que se relacionam diretamente com aspectos da qualidade de sementes, como maturidade fisiológica e proteção celular.

## REFERÊNCIAS

- Ainsworth, E.A., Yendrek, C.R., Skoneczka, J.A., Long, S.P., 2012. Accelerating yield potential in soybean: potential targets for biotechnological improvement. *Plant. Cell Environ.* 35, 38–52. <https://doi.org/10.1111/j.1365-3040.2011.02378.x>
- Baek, I., Kim, M., Cho, B.-K., Mo, C., Barnaby, J., McClung, A., Oh, M., 2019a. Selection of Optimal Hyperspectral Wavebands for Detection of Discolored, Diseased Rice Seeds. *Appl. Sci.* 9, 1027. <https://doi.org/10.3390/app9051027>
- Baek, I., Kusumaningrum, D., Kandpal, L.M., Lohumi, S., Mo, C., Kim, M.S., Cho, B.K., 2019b. Rapid measurement of soybean seed viability using Kernel-based multispectral image analysis. *Sensors (Switzerland)* 19, 1–16. <https://doi.org/10.3390/s19020271>
- Barboza da Silva, C., Oliveira, N.M., de Carvalho, M.E.A., de Medeiros, A.D., de Lima Nogueira, M., dos Reis, A.R., 2021. Autofluorescence-spectral imaging as an innovative method for rapid, non-destructive and reliable assessing of soybean seed quality. *Sci. Rep.* 11, 17834. <https://doi.org/10.1038/s41598-021-97223-5>
- Batista, T.B., Mastrangelo, C.B., de Medeiros, A.D., Petronilio, A.C.P., Fonseca de Oliveira, G.R., dos Santos, I.L., Crusciol, C.A.C., Amaral da Silva, E.A., 2022. A Reliable Method to Recognize Soybean Seed Maturation Stages Based on Autofluorescence- Spectral Imaging Combined With Machine Learning Algorithms. *Front. Plant Sci.* 13. <https://doi.org/10.3389/fpls.2022.914287>
- Boelt, B., Shrestha, S., Salimi, Z., Jørgensen, J.R., Nicolaisen, M., Carstensen, J.M., 2018.

- Multispectral imaging – a new tool in seed quality assessment? *Seed Sci. Res.* 28, 222–228. <https://doi.org/10.1017/S0960258518000235>
- Brasil, 2009. Regras para análise de sementes, Ministério da agricultura Pecuária e Abastecimento. MAPA/ACS, Brasília, DF. <https://doi.org/10.2307/2261447>
- Buschmann, C., Langsdorf, G., Lichtenthaler, H.K., 2000. Imaging of the Blue, Green, and Red Fluorescence Emission of Plants: An Overview. *Photosynthetica* 38, 483–491. <https://doi.org/10.1023/A:1012440903014>
- de Medeiros, A.D., Capobiango, N.P., da Silva, J.M., da Silva, L.J., da Silva, C.B., dos Santos Dias, D.C.F., 2020. Interactive machine learning for soybean seed and seedling quality classification. *Sci. Rep.* 10, 11267. <https://doi.org/10.1038/s41598-020-68273-y>
- ElMasry, G., ElGamal, R., Mandour, N., Gou, P., Al-Rejaie, S., Belin, E., Rousseau, D., 2020. Emerging thermal imaging techniques for seed quality evaluation: Principles and applications. *Food Res. Int.* 131, 109025. <https://doi.org/10.1016/j.foodres.2020.109025>
- ElMasry, G., Mandour, N., Al-Rejaie, S., Belin, E., Rousseau, D., 2019a. Recent Applications of Multispectral Imaging in Seed Phenotyping and Quality Monitoring—An Overview. *Sensors* 19, 1090. <https://doi.org/10.3390/s19051090>
- ElMasry, G., Mandour, N., Wagner, M.-H., Demilly, D., Verdier, J., Belin, E., Rousseau, D., 2019b. Utilization of computer vision and multispectral imaging techniques for classification of cowpea (*Vigna unguiculata*) seeds. *Plant Methods* 15, 24. <https://doi.org/10.1186/s13007-019-0411-2>
- Finch-Savage, W.E.E., Bassel, G.W.W., 2016. Seed vigour and crop establishment: extending performance beyond adaptation. *J. Exp. Bot.* 67, 567–591. <https://doi.org/10.1093/jxb/erv490>
- França-Neto, J. de B., Krzyzanowski, francisco carlos, 2020. Teste de Tetrazólio em Sementes de Soja, in: *Vigor de Sementes: Conceitos e Testes*. p. 601.
- França-Silva, F., Rego, C.H.Q., Gomes-Junior, F.G., Moraes, M.H.D. de, Medeiros, A.D. de, Silva, C.B. da, 2020. Detection of *Drechslera avenae* (Eidam) Sharif [*Helminthosporium avenae* (Eidam)] in Black Oat Seeds (*Avena strigosa* Schreb) Using Multispectral Imaging. *Sensors* 20, 3343. <https://doi.org/10.3390/s20123343>
- Galletti, P.A., Carvalho, M.E.A., Hirai, W.Y., Brancaglioni, V.A., Arthur, V., Barboza da Silva, C., 2020. Integrating Optical Imaging Tools for Rapid and Non-invasive Characterization of Seed Quality: Tomato (*Solanum lycopersicum* L.) and Carrot (*Daucus carota* L.) as Study Cases. *Front. Plant Sci.* 11.

- <https://doi.org/10.3389/fpls.2020.577851>
- Gebregziabher, B. s., Zhang, S., Azam, M., Qi, J., Agyenim-boateng, K. g., Feng, Y., Liu, Y., Li, J., Li, B., Sun, J., 2023. Natural variations and geographical distributions of seed carotenoids and chlorophylls in 1 167 Chinese soybean accessions. *J. Integr. Agric.* 22, 2632–2647. <https://doi.org/10.1016/j.jia.2022.10.011>
- Goodfellow, I., Bengio, Y., Aaron Courville, 2016. *Deep learning*. MIT Press, Cambridge, MA.
- ISTA, 2023. *International rules for seed testing*. International Seed Testing Association, Wallisellen, Switzerland. <https://doi.org/https://doi.org/10.15258/istarules.2023.I>
- LeCun, Y., Bengio, Y., Hinton, G., 2015. Deep learning. *Nature* 521, 436–444. <https://doi.org/10.1038/nature14539>
- Lee, J., Hwang, Y.-S., Lee, J.-D., Chang, W.-S., Choung, M.-G., 2013. Metabolic alterations of lutein,  $\beta$ -carotene and chlorophyll a during germination of two soybean sprout varieties. *Food Chem.* 141, 3177–3182. <https://doi.org/10.1016/j.foodchem.2013.06.006>
- Marcos-Filho, J., 2020. Teste de envelhecimento acelerado, in: *Vigor de Sementes: Conceitos e Testes*. p. 601.
- Monma, M., Ito, M., Saito, M., Chikuni, K., 1994. Carotenoid Components in Soybean Seeds Varying with Seed Color and Maturation Stage. *Biosci. Biotechnol. Biochem.* 58, 926–930. <https://doi.org/10.1271/bbb.58.926>
- Olesen, M.H., Carstensen, J.M., Boelt, B., 2011. Multispectral imaging as a potential tool for seed health testing of spinach (*Spinacia oleracea* L.). *Seed Sci. Technol.* 39, 140–150. <https://doi.org/10.15258/sst.2011.39.1.12>
- Rego, C.H.Q., França-Silva, F., Gomes-Junior, F.G., Moraes, M.H.D. de, Medeiros, A.D. de, Silva, C.B. da, 2020a. Using multispectral imaging for detecting seed-borne fungi in cowpea. *Agriculture* in press.
- Rego, C.H.Q., França-Silva, F., Gomes-Junior, F.G., Moraes, M.H.D. de, Medeiros, A.D. de, Silva, C.B. da, 2020b. Using Multispectral Imaging for Detecting Seed-Borne Fungi in Cowpea. *Agriculture* 10, 361. <https://doi.org/10.3390/agriculture10080361>
- Sendin, K., Manley, M., Williams, P.J., 2018. Classification of white maize defects with multispectral imaging. *Food Chem.* 243, 311–318. <https://doi.org/10.1016/j.foodchem.2017.09.133>
- Shrestha, S., Deleuran, L.C., Olesen, M.H., Gislum, R., 2015. Use of multispectral imaging in varietal identification of tomato. *Sensors (Switzerland)* 15, 4496–4512.

<https://doi.org/10.3390/s150204496>

- Soares, J.M., de Medeiros, A.D., Pinheiro, D.T., Rosas, J.T.F., da Silva, L.J., Machado, D.L.M., Dias, D.C.F.D.S., 2023. Low-cost system for multispectral image acquisition and its applicability to analysis of the physiological potential of soybean seeds. *Acta Sci. - Agron.* 45. <https://doi.org/10.4025/ACTASCIAGRON.V45I1.57060>
- Tsakanikas, P., Pavlidis, D., Nychas, G.J., 2015. High throughput multispectral image processing with applications in food science. *PLoS One* 10, 1–15. <https://doi.org/10.1371/journal.pone.0140122>
- Wilkes, T., Nixon, G., Bushell, C., Waltho, A., Alroichdi, A., Burns, M., 2016. Feasibility Study for Applying Spectral Imaging for Wheat Grain Authenticity Testing in Pasta. *Food Nutr. Sci.* 07, 355–361. <https://doi.org/10.4236/fns.2016.75037>
- Xia, Y., Xu, Y., Li, J., Zhang, C., Fan, S., 2019. Recent advances in emerging techniques for non-destructive detection of seed viability: A review. *Artif. Intell. Agric.* 1, 35–47. <https://doi.org/10.1016/j.aiia.2019.05.001>