

UNIVERSIDADE FEDERAL DE VIÇOSA

**Espectroscopia no infravermelho próximo e imagens multiespectrais para a
classificação do potencial fisiológico de sementes de algodão**

Bruno Gomes de Noronha
Doctor Scientiae

**VIÇOSA - MINAS GERAIS
2024**

BRUNO GOMES DE NORONHA

**Espectroscopia no infravermelho próximo e imagens multiespectrais para a
classificação do potencial fisiológico de sementes de algodão**

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*.

Orientadora: Denise C. F. dos S. Dias

**VIÇOSA - MINAS GERAIS
2024**

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

T

Noronha, Bruno Gomes de, 1987-
N852e
2024
Espectroscopia no infravermelho próximo e imagens multiespectrais para a classificação do potencial fisiológico de sementes de algodão / Bruno Gomes de Noronha. – Viçosa, MG, 2024.

1 tese eletrônica (60 f.): il. (algumas color.).

Orientador: Denise Cunha Fernandes dos Santos Dias.
Tese (doutorado) - Universidade Federal de Viçosa,
Departamento de Agronomia, 2024.

Inclui bibliografia.

DOI: <https://doi.org/10.47328/ufvbbt.2025.154>

Modo de acesso: World Wide Web.

1. Algodão - Semente - Qualidade. 2. Germinação.
3. Sementes - Fisiologia. 4. Imagens multiespectrais.
5. Espectroscopia de infravermelho. 6. Aprendizado do computador. I. Dias, Denise Cunha Fernandes dos Santos, 1958-. II. Universidade Federal de Viçosa. Departamento de Agronomia. Programa de Pós-Graduação em Fitotecnia. III. Título.

CDD 22. ed. 633.5121

BRUNO GOMES DE NORONHA

**Espectroscopia no infravermelho próximo e imagens multiespectrais para a
classificação do potencial fisiológico de sementes de algodão**

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*.

APROVADA: 18 de dezembro de 2024.

Assentimento:

Bruno Gomes de Noronha
Autor

Denise Cunha Fernandes dos Santos Dias
Orientadora

Essa tese foi assinada digitalmente pelo autor em 01/04/2025 às 12:03:48 e pela orientadora em 01/04/2025 às 14:50:36. As assinaturas têm validade legal, conforme o disposto na Medida Provisória 2.200-2/2001 e na Resolução nº 37/2012 do CONARQ. Para conferir a autenticidade, acesse <https://siadoc.ufv.br/validar-documento>. No campo 'Código de registro', informe o código **UBEW.KDYL.Y7GB** e clique no botão 'Validar documento'.

Para Bruno Rafael Morais de Macedo, um ser humano incrível, um amigo e um ser
de luz inenarrável
Dedico.

AGRADECIMENTOS

A Deus, por me levantar todas as vezes que precisei.

Ao Universo por me inspirar a ter Fé, que sonhos podem virar realidade.

A Universidade Federal de Viçosa, Departamento de Agronomia e a Pós-graduação em Fitotecnia pela oportunidade.

A professora Denise Dias, por ter aceitado essa missão, por compartilhar conhecimento com tanto amor, e criar uma nação baseada na educação e respeito, obrigado por todos os ensinamentos e palavras de carinho.

Ao professor Laércio Junio, por todo carinho e dedicação para com o progresso da ciência, por ser um homem honesto e principalmente pela amizade e ensinamentos.

Ao Grupo Fantástico GSEM! obrigado por ser minha casa, em muitos momentos vocês foram o meu lar, o meu abrigo, meu local de aconchego, onde pude fazer amigos, trocar conhecimento e me divertir como toda família faz;

A Kamylla de Faria, Julia Soares, Augusto Limão, Maura Brochado, Ítalo de Jesus e Daniel Pinheiros, minha vida não teria sido a mesma se não tivesse conhecido vocês, obrigado por tudo e por tanto, amo vocês!

A André Medeiros, meu irmão, meu amigo, e parte disso é culpa sua, te agradeço infinitamente por ter me ajudado a subir mais esse degrau na vida, te amo!

Ao meu noivo Rodolpho Ernesto Russi e toda família Russi, por todo amor, cuidado e companheirismo, te amo!

Ao Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq), pela concessão da bolsa.

O presente trabalho foi realizado com apoio da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior – Brasil (CAPES) – Código de Financiamento 001.

“É bom ter um objetivo para a jornada à frente, mas, no fim, o que importa é a
jornada”
(Ursula K. Le Guin)

RESUMO

NORONHA, Bruno Gomes de, D.Sc., Universidade Federal de Viçosa, dezembro de 2024. **Espectroscopia no infravermelho próximo e imagens multiespectrais para a classificação do potencial fisiológico de sementes de algodão.** Orientadora: Denise Cunha Fernandes dos Santos Dias.

A cultura do algodão tem grande importância mundial, e a qualidade das sementes é um fator crítico para o sucesso das lavouras. Métodos rápidos e não destrutivos para avaliação do potencial fisiológico das sementes são essenciais para o controle de qualidade nas empresas. Assim, este trabalho teve como objetivo avaliar a eficiência da espectroscopia no infravermelho próximo (NIR) e de imagens multiespectrais, associadas a modelos de aprendizado supervisionado, para a predição e classificação da qualidade fisiológica de sementes de algodão. Foram utilizadas sementes de um lote comercial, que foram envelhecidas artificialmente a 41 °C e 100% de umidade relativa (UR) por 0, 48 e 96 horas. As sementes de cada tratamento foram submetidas aos testes de germinação, emergência, comprimento de plântulas e condutividade elétrica, sendo então categorizadas em alto vigor (controle), médio vigor (48 horas) e baixo vigor (96 horas), com base nos resultados desses testes. No primeiro ensaio, o objetivo foi avaliar a eficiência da espectroscopia no infravermelho próximo para a classificação do potencial fisiológico de sementes de algodão. Foram obtidos espectros NIR a partir das sementes individuais, totalizando 100 leituras para cada amostra. Os espectros foram processados utilizando diferentes métodos de pré-processamento, incluindo normalização, suavização e remoção de ruído. Em seguida, foram desenvolvidos modelos utilizando a Análise Discriminante por Mínimos Quadrados Parciais (PLS-DA). O modelo PLS-DA apresentou resultados promissores, com acurácia e kappa de 100% na fase de treinamento e de 95% e 93% na fase de teste, respectivamente. Os resultados confirmam que a espectroscopia NIR é uma ferramenta eficaz para classificar o potencial fisiológico das sementes de algodão, podendo ser incorporada como uma etapa rápida e complementar ao processo de controle de qualidade. No segundo ensaio, o objetivo foi investigar a viabilidade do uso de imagens multiespectrais na classificação da qualidade de sementes de algodão, utilizando algoritmos de inteligência artificial como ferramenta para otimizar o controle de qualidade. Para isso, foi desenvolvido um protótipo de baixo custo (Chroma Seeds) para aquisição de imagens multiespectrais. As imagens, capturadas em comprimentos de onda específicos (395–940 nm), foram processadas no software ImageJ® para obter dados de reflectância. Modelos preditivos

foram desenvolvidos utilizando o software Orange Data Mining e algoritmos de classificação supervisionada, como Tree, Random Forest, Gradient Boosting, Support Vector Machine (SVM) e Neural Network. Os modelos de rede neural e SVM foram eficientes, com acurácia de 100% e kappa de 99% na fase de teste, confirmando a eficácia da abordagem baseada em imagens multiespectrais na classificação do potencial fisiológico das sementes. Os resultados obtidos nos dois estudos demonstram que tanto a espectroscopia NIR quanto a análise de imagens multiespectrais são ferramentas eficazes para classificar lotes de sementes de algodão quanto ao potencial fisiológico, podendo ser integradas em programas de controle de qualidade para otimizar o processo de seleção de sementes.

Palavras-chave: Espectroscopia NIR; Vigor de sementes; Deterioração; Imagens Multiespectrais; Aprendizado de Máquina

ABSTRACT

NORONHA, Bruno Gomes de, D.Sc., Universidade Federal de Viçosa, December, 2024. **Near-infrared spectroscopy and multispectral imaging for classification of physiological potential of cotton seeds.** Adviser: Denise Cunha Fernandes dos Santos Dias.

Cotton cultivation holds great global importance, and seed quality is a critical factor for the success of cotton crops. Fast and non-destructive methods for assessing the physiological potential of seeds are essential for quality control in seed companies. Thus, this study aimed to evaluate the efficiency of near-infrared spectroscopy (NIR) and multispectral imaging, combined with supervised learning models, for predicting and classifying the physiological quality of cotton seeds. Seeds from a commercial lot were artificially aged at 41 °C and 100% relative humidity (RH) for 0, 48, and 96 hours. Seeds from each treatment were subjected to germination, seedling emergence, seedling length, and electrical conductivity tests, and were subsequently categorized into high vigor (control), medium vigor (48 hours), and low vigor (96 hours) based on the results of these tests. In the first trial, the objective was to evaluate the efficiency of near-infrared spectroscopy for classifying the physiological potential of cotton seeds. NIR spectra were obtained from individual seeds, totaling 100 readings for each sample. The spectra were processed using different preprocessing methods, including normalization, smoothing, and noise removal. Subsequently, models were developed using Partial Least Squares Discriminant Analysis (PLS-DA). The PLS-DA model showed promising results, with accuracy and kappa values of 100% in the training phase and 95% and 93% in the test phase, respectively. The results confirm that NIR spectroscopy is an effective tool for classifying the physiological potential of cotton seeds and can be incorporated as a fast and complementary step in the quality control process. In the second trial, the objective was to investigate the feasibility of using multispectral imaging to classify cotton seed quality, employing artificial intelligence algorithms as a tool to optimize the quality control process. For this purpose, a low-cost prototype (Chroma Seeds) was developed for multispectral image acquisition. Images, captured at specific wavelengths (395–940 nm), were processed using the ImageJ® software to obtain reflectance data. Predictive models were developed using the Orange Data Mining software and supervised classification algorithms such as Tree, Random Forest, Gradient Boosting, Support Vector Machine (SVM), and Neural Network. The Neural Network and SVM models were highly effective, with accuracy of 100% and kappa of 99% in the test phase, confirming the efficiency of the multispectral

imaging-based approach for classifying the physiological potential of seeds. The results obtained in both studies demonstrate that both NIR spectroscopy and multispectral imaging analysis are effective tools for classifying cotton seed lots in terms of physiological potential, and they can be integrated into quality control programs to optimize the seed selection process.

Keywords: NIR Spectroscopy; Seed Vigor; Deterioration; Multispectral Imaging; Machine Learning

SUMÁRIO

CAPITULO I.....	11
PREDIÇÃO DA QUALIDADE DE SEMENTES DE ALGODÃO POR ESPECTROSCOPIA NO INFRAVERMELHO PRÓXIMO.....	11
RESUMO	12
ABSTRACT	13
INTRODUÇÃO.....	14
MATERIAL E MÉTODOS.....	16
Análises Fisiológicas.....	17
Primeira contagem da germinação (PC):	17
Emergência de plântulas	17
Índice de velocidade de emergência (IVE)	17
Condutividade elétrica (CE).....	17
Análises de Imagem.....	17
Teste de Raios X	17
Análise computadorizada de plântulas	18
Análises por Espectroscopia de Infravermelho NIR	18
Delineamento experimental e análise estatística	20
RESULTADOS E DISCUSSÃO.....	20
CONCLUSÃO.....	31
REFERÊNCIAS	31
CAPITULO II.....	38
CLASSIFICAÇÃO DO POTENCIAL FISIOLÓGICO DE SEMENTES DE ALGODÃO POR IMAGENS MULTIESPECTRAIS E MODELOS DE APRENDIZADO SUPERVISIONADO	38
RESUMO:	39
ABSTRACT	40
INTRODUÇÃO.....	41
MATERIAL E MÉTODOS.....	43
Determinação do grau de umidade.....	43
Teste de germinação.....	43
Primeira contagem de germinação	43
Emergência de plântulas	43
Índice de velocidade de emergência (IVE)	44
Condutividade elétrica	44
Comprimento de plântulas	44
Delineamento experimental e análise estatística	44
Imagens Multiespectrais	44
Modelos de Predição.....	48
RESULTADOS E DISCUSSÃO.....	50
CONCLUSÃO.....	57
REFERÊNCIAS	57

Capitulo I

**PREDIÇÃO DA QUALIDADE DE SEMENTES DE ALGODÃO POR ESPECTROSCOPIA
NO INFRAVERMELHO PRÓXIMO**

RESUMO

O algodão é uma *commodity* agrícola de importância mundial e o Brasil é o quarto maior produtor, sendo que, um dos principais desafios na produção é obter sementes de alta qualidade. Na atualidade, há uma grande demanda por ferramentas rápidas, não invasivas e menos subjetivas para avaliação da qualidade das sementes a fim de agilizar as tomadas de decisões sobre os lotes produzidos a cada safra. A espectroscopia no infravermelho próximo (NIR) é uma técnica promissora nesse contexto. Assim, o objetivo desta pesquisa foi avaliar a metodologia baseada em FT-NIR para predição dos níveis de qualidade em sementes de algodão. Foram utilizadas sementes de um lote comercial que foram envelhecidas artificialmente a 41 °C, 100% UR por 0 (controle), 48 e 96 h. As sementes de cada tratamento foram submetidas aos testes de germinação, emergência, comprimento de plântulas e condutividade elétrica. Foram ainda obtidos os espectros NIR a partir das sementes individuais, totalizando 100 leituras para cada amostra. As amostras foram categorizadas em alto vigor (controle), médio vigor (48 horas) e baixo vigor (96 horas), com base nos resultados dos testes de germinação e vigor, esses espectros foram processados por meio de diferentes métodos para comparação, em seguida, foram desenvolvidos modelos utilizando a Análise Discriminante por Mínimos Quadrados Parciais (PLS-DA). O modelo obtido a partir dos dados espectrais NIR e da PLS-DA, utilizando a centralização como pré-processamento, demonstrou acurácia e coeficiente *kappa* de 100% na etapa de treinamento e na etapa de teste esses valores foram de 95 e 93 %, respectivamente. Isso demonstra que a espectroscopia no infravermelho próximo (NIR) foi eficiente para classificar sementes de algodão com diferentes níveis de deterioração.

Palavras chaves: *Gossypium hirsutum* L.; vigor; FT-NIR; potencial fisiológico

ABSTRACT

Cotton is an agricultural commodity of global importance and Brazil is the fourth largest producer. One of the main challenges in production is to obtain high-quality seeds. Currently, there is a great demand for fast, non-invasive and less subjective tools to assess seed quality in order to speed up decision-making about the lots produced each harvest. Near-infrared spectroscopy (NIR) is a promising technique in this context. Thus, the objective of this research was to evaluate the FT-NIR-based methodology for predicting quality levels in cotton seeds. Seeds from a commercial lot that were artificially aged at 41 °C, 100% RH for 0 (control), 48 and 96 h were used. The seeds of each treatment were subjected to germination, emergence, seedling length and electrical conductivity tests. NIR spectra were also obtained from individual seeds, totaling 100 readings for each sample. The samples were categorized as high vigor (control), medium vigor (48 hours) and low vigor (96 hours), based on the results of the germination and vigor tests. These spectra were processed using different methods for comparison, and then models were developed using Partial Least Squares Discriminant Analysis (PLS-DA). The model obtained from the NIR spectral data and PLS-DA, using centralization as preprocessing, demonstrated accuracy and *kappa* coefficient of 100% in the training stage and in the testing stage these values were 95 and 93%, respectively. This demonstrates that near-infrared (NIR) spectroscopy was efficient in classifying cotton seeds with different levels of deterioration.

Keywords: *Gossypium hirsutum* L.; vigor; FT-NIR; seed quality; physiological potential

INTRODUÇÃO

O algodão (*Gossypium hirsutum* L.) é uma das culturas mais amplamente cultivadas no mundo, especialmente no Brasil. É a principal fonte de fibra natural, utilizado para produção de óleo vegetal e farinhas de alta qualidade proteica (Zhu *et al.*, 2019). Além disso, possui grande importância socioeconômica, gerando milhares de empregos diretos e indiretos (Saeed *et al.*, 2020). Atualmente, área plantada de algodão ultrapassa os 2 milhões de hectares a produção estimada de algodão em pluma gira em torno de 3,82 milhões de toneladas para a safra 2024/25, com área total plantada no Brasil de 2 milhões de hectares, a maior das últimas seis safras, sendo os estados do Mato Grosso e Bahia os maiores produtores (CONAB, 2024).

Para se obter estande adequado em campo, é importante que as sementes tenham alta qualidade física, fisiológica, genética e sanitária, proporcionando, por exemplo, maior velocidade e uniformidade de emergência de plântulas vigorosas que irão refletir na uniformidade e potencial produtivo da lavoura (Mayrinck *et al.*, 2020; Mattioni *et al.*, 2012). A utilização de sementes de baixa qualidade resulta em plantas que não podem ser equiparadas em desenvolvimento e produtividade às obtidas via sementes de alta qualidade, podendo acarretar perdas econômicas significativas (Krzyzanowski *et al.*, 2020; Nagel; Börner, 2010; Oliveira *et al.*, 2009).

A desuniformidade de maturação das sementes de algodão na planta, a propensão das sementes à deterioração em campo e no armazenamento e a susceptibilidade a danos por pragas e doenças, são fatores que contribuem para reduzir o potencial fisiológico das sementes. É importante ressaltar que o beneficiamento das sementes pode contribuir significativamente para a redução da sua qualidade, devido a possíveis danos ocasionados pelo descaroçamento mecânico para retirada das fibras e pelo processo de deslntamento, o qual utiliza ácidos para a retirada do línter. Esses aspectos refletem diretamente na qualidade das sementes ao final do processo produtivo, de modo que os padrões mínimos de germinação exigido para comercialização das sementes é de 70% para sementes básicas e 75% para sementes certificadas (Brasil, 2013), inferiores ao exigido para outras espécies de grandes culturas.

O alto teor de óleo presente nas sementes de algodão é um dos fatores que contribuem para deterioração. Sabe-se que este é um processo progressivo, irreversível e provocado principalmente pela peroxidação de lipídios (Finch-Savage *et al.*, 2013). No entanto, o ambiente no qual as sementes são expostas exerce muita influência sobre a intensidade e velocidade desse processo (Goel *et al.*, 2003).

Técnicas de envelhecimento artificial têm sido utilizadas em estudos sobre os mecanismos ligados à deterioração das sementes (Ribeiro *et al.*, 2021). O envelhecimento artificial promove estresse, provocado pela exposição de altas temperaturas e umidade relativa às sementes, o que acelera o processo de deterioração. Assim, o teste de envelhecimento acelerado pode ser utilizado como ferramenta para provocar a deterioração das sementes (Limão *et al.*, 2025), de modo que o desempenho das sementes após o teste permite indicar quais lotes apresentam maior e menor tolerância à deterioração (Silva *et al.*, 2024; Silva *et al.*, 2023; Goel *et al.*, 2003).

A identificação e discriminação de lotes de sementes quanto ao potencial fisiológico é relevante para os programas de controle de qualidade das sementes. Na atualidade, o monitoramento para avaliar o potencial fisiológico das sementes, demanda de métodos rápidos, não subjetivos, confiáveis e, se possível, não destrutivos, interessantes, sobretudo, para agilizar a tomada de decisão.

A avaliação da qualidade das sementes de algodão usualmente é realizada por meio do teste de germinação (Brasil, 2009) e complementada pelos testes de vigor. Os testes de vigor mais recomendados para a avaliação do potencial fisiológico de sementes de algodão são o envelhecimento acelerado (Krzyzanowski *et al.*, 2020), germinação a baixa temperatura (Dias e Marcos-Filho, 2020) e o teste de tetrazólio (Krzyzanowski *et al.*, 2020). Dentre estes, o teste de tetrazólio é o único que fornece resultados em período relativamente curto, ou seja, 24 horas. Apesar de eficientes, essas análises são destrutivas, ou seja, para cada teste realizado é necessária uma quantidade de sementes para avaliação que não podem ser utilizadas para outras avaliações além disso esses procedimentos são onerosos e tem um custo elevado dado a quantidade de amostras que uma empresa pode realizar ao longo da safra, podendo levar alguns dias para a obtenção e interpretação dos resultados. Além disso, envolvem interpretações por parte do avaliador, o que pode interferir na precisão dos resultados. Todas essas questões citadas podem ter impactos nos custos e nas tomadas de decisões quanto ao manejo e destino dos lotes.

Desta forma, é de extrema importância a integração de novas tecnologias que permitam a obtenção de resultados com menor subjetividade, em curto período e de maneira não destrutiva, permitindo a utilização das sementes para outras avaliações. Dentre essas tecnologias, o teste de Raios X, a análise automatizada de imagens de sementes e plântulas (Alvarenga; Marcos-Filho; Gomes Junior, 2012) e a espectroscopia no infravermelho próximo – NIR (Cordeiro, 2023) vêm se destacando pelo grande potencial para utilização nos programas de controle interno de qualidade de sementes.

Neste contexto, a espectroscopia no infravermelho próximo pode representar uma alternativa promissora — não como método substitutivo para a análise de qualidade de sementes, mas como um método mais rápido para agilizar a tomada de decisões. A vantagem dessa técnica é a capacidade de realizar análises sucessivas em um curto período de tempo, gerando uma grande quantidade de informações, com menor necessidade de mão de obra, maior rapidez e menor custo, sendo não poluente, sem utilização de produtos químicos ou reagentes, além de reduzir os erros na condução e interpretação dos testes convencionais. (Bianchini *et al.*, 2021; Egidio *et al.*, 2010; Elmasry *et al.*, 2019; Gaitán-Jurado *et al.*, 2008; Zhu *et al.*, 2019).

A técnica de espectroscopia no infravermelho próximo (780-2500 nm) permite analisar as vibrações moleculares (C-H, N-H e O-H) presentes em amostras de sementes, que estão diretamente ligadas à composição de carboidratos, lipídeos e proteínas, por possuírem ligações que se alteram em virtude do comprimento de onda, modificando o seu estado vibracional. Pesquisas recentes relacionadas à avaliação da qualidade de sementes de diversas espécies têm gerado resultados promissores como em soja (Soares *et al.* 2024; Bazoni *et al.*, 2017), tomate (Shrestha *et al.*, 2016), milho (Ambrose *et al.*, 2016) e mais recentemente em lentilha (Limão *et al.*, 2025), já para algodão as análises foram realizadas para determinação de parâmetros de qualidade de caroços destinados a indústria (Huang *et al.*, 2013; Gaitán-Jurado *et al.*, 2008), havendo assim a necessidade de estabelecer uma metodologia para qualidade de sementes.

Diante do exposto, este trabalho teve como objetivo avaliar a eficiência da espectroscopia no infravermelho próximo para a classificação do potencial fisiológico de sementes de algodão.

MATERIAL E MÉTODOS

A pesquisa foi conduzida no Laboratório de Pesquisa de Sementes, do Departamento de Agronomia da Universidade Federal de Viçosa (DAA-UFV). Foram utilizadas sementes de algodão (*Gossypium hirsutum* L.), provenientes da Cooperativa de Produtores Rurais de Catuti-MG. As sementes foram divididas em amostras e submetidas à deterioração utilizando o princípio do teste de envelhecimento acelerado. Para tanto, foram utilizadas caixas do tipo gerbox com tela metálica, com o fundo do recipiente contendo 40 mL de água destilada. As sementes foram dispostas uniformemente sobre a tela sem contato com a água, as caixas foram fechadas e mantidas em câmara do tipo B.O.D a 41 °C, por períodos de 0, 48 e 96 horas, tendo o tratamento controle sementes não envelhecidas (Marcos-Filho, 2020).

Após cada tratamento de envelhecimento, foram retiradas amostras de quatro repetições de 20 sementes cada, para determinação do grau de umidade em estufa, a 105 °C por 24 h.

Os resultados foram expressos em porcentagem (Brasil, 2009). Em seguida, as sementes de cada tratamento foram submetidas aos testes descritos a seguir.

Análises Fisiológicas

Germinação:

Foram empregadas quatro repetições de 50 sementes, utilizando rolo de papel toalha umedecido com água na quantidade 2,5 vezes o peso do papel. O teste foi conduzido em germinador regulado à temperatura de 25°C e as avaliações efetuadas aos 4 e 12 dias após a sementeira, segundo os critérios estabelecidos pelas Regras para Análise de Sementes (Brasil, 2009).

Primeira contagem da germinação (PC):

Conduzido juntamente com o teste de germinação, foi determinada a porcentagem de plântulas normais no quarto dia após a sementeira.

Emergência de plântulas

Foram utilizadas quatro repetições de 50 sementes. A sementeira ocorreu no espaçamento de 5,0 x 5,0 cm e profundidade de 1,0 cm em bandejas de 10 kg contendo solo e areia (1:1). A contagem final foi realizada 14 dias após a sementeira, computando-se as plântulas emergidas (Krzyzanowski *et al.*, 2020).

Índice de velocidade de emergência (IVE)

Foi estabelecido juntamente com o teste de emergência, realizando contagens diárias até 14 dias após a sementeira e, empregando a fórmula de Maguire (1962).

Condutividade elétrica (CE)

Quatro repetições de 50 sementes foram pesadas (precisão de 0,001 g) e colocadas em copos plásticos contendo o volume de 70 ml de água destilada. Os copos foram mantidos em incubadora BOD a 25 °C por 24 h. Após esse período, foi realizada a leitura da condutividade elétrica por meio de condutímetro (Digimed CD 21) e os resultados foram expressos em $\mu\text{S}\cdot\text{cm}^{-1}\cdot\text{g}^{-1}$ (Vieira e Marcos-Filho, 2020).

Análises de Imagem

Teste de Raios X

Para a aquisição das imagens, as sementes, em grupos de 10, foram fixadas previamente em papel adesivo, totalizando 100 sementes por tratamento. As imagens radiográficas foram

geradas por um dispositivo Faxitron, modelo MX-20 (Faxitron X-ray Corp. Wheeling, IL, EUA) disponível no Laboratório de Entomologia da Universidade Federal de Viçosa. O equipamento foi ajustado para uma tensão de 23 kV e as sementes expostas à radiação por 10 segundos, a uma distância focal de 30 cm. As imagens foram salvas em arquivos no formato TIFF (*Tagged Image File Format*), e em seguida analisadas obtendo-se os valores de área, perímetro, circularidade, proporção do comprimento e largura, solidez, média dos valores de cinza e densidade integrada por meio do programa ImageJ® (Schneider; Rasband e Eliceiri, 2012), utilizando a metodologia conforme Noronha *et al.*, 2018.

Análise computadorizada de plântulas

Foram utilizadas as plântulas normais provenientes do teste de germinação que as foram dispostas em fundo azul, fotografadas e, posteriormente, analisadas pelo programa ImageJ obtendo-se os comprimentos da parte aérea, radicular e total de cada plântula (Medeiros; Pereira, 2018, Noronha *et al.* 2018).

Análises por Espectroscopia de Infravermelho NIR

As sementes foram classificadas como alto vigor (controle), médio vigor (48 horas) e baixo vigor (96 horas) de acordo com os resultados dos testes de qualidade fisiológica. Foram realizadas leituras de 100 sementes por tratamento, de forma individual para cada semente, totalizando 300 espectros obtidos em espectrômetro de transformada de Fourier - FT-NIR (*Thermo Scientific Antaris II*), disponível no Laboratório de Agroenergia da Universidade Federal de Viçosa, com faixa espectral de 1000 a 2500 nm. As sementes isoladas foram dispostas em suporte de tamanho adequado às características das sementes. Cada espectro foi representado por uma média de 32 varreduras medidas com resolução de 8 nm, resultando em 3112 variáveis (bandas espectrais). Este instrumento opera em uma esfera integradora com um módulo de refletância difusa e os espectros foram expressos em Absorbância ($\log 1/R$).

Pré-processamento

Os espectros NIR foram transformados usando diferentes pré-processamentos matemáticos: *Standard Normal Variate* (SNV): A normalização de variável padrão é uma técnica usada para transformar os dados de forma que cada variável tenha média zero e desvio padrão igual a um. *Multiplicative Scatter Correction* (MSC): A Correção de Dispersão Multiplicativa é uma técnica que visa eliminar variações sistemáticas nos espectros causadas por fatores como o instrumental ou condições experimentais. Essa correção é feita dividindo os dados espectrais pela média e multiplicando pela média dos valores espectrais, restaurando assim a dispersão dos dados para um valor padrão, melhorando a precisão nas análises

subsequentes. Derivadas de Savitzky-Golay: As derivadas de Savitzky-Golay são aplicadas para suavizar e obter as derivadas de sinais ou espectros. Esse método usa um filtro polinomial para ajustar os dados e calcular as derivadas de uma maneira que minimize o ruído. É muito utilizado em espectroscopia, pois pode realçar as características do espectro, como picos ou mudanças de forma, sem distorcer as informações subjacentes. Esses métodos são usados a fim de remover ou minimizar qualquer contribuição espectral indesejada e normalizar a razão sinal/ruído.

Para o processamento foi utilizando o algoritmo de classificação supervisionada *Partial Least Squares – Discriminant Analysis* (PLS-DA). O banco de dados foi dividido em 70% para o treinamento e 30% para a validação e avaliando o modelo utilizando as métricas de Acurácia e coeficiente de concordância de *kappa* e matrizes de confusão. Por meio da matriz de confusão, foram extraídos os valores de verdadeiros positivos (TP), falsos positivos (FP), verdadeiros negativos (TN) e falsos negativos (FN), com a finalidade de calcular as métricas *sensitivity*, *accuracy*, *balanced accuracy*, *F1-score*, *Neg. Pred Value*, *Posi. Pred. Value*, *precision* e *specificity*, empregando as equações 1 a 6, como descritas por Brodersen *et al.*, (2010).

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

$$\text{Kappa} = \frac{P_o - P_e}{1 - P_e} \quad (2)$$

$$\text{Sensibilidade} = \frac{TP}{TP + FN} \quad (3)$$

$$\text{Acurácia balanceada} = \frac{\text{Sensibilidade} + \text{Especificidade}}{2} \quad (4)$$

$$\text{F1} = \frac{2 \cdot (\text{Precisão} \cdot \text{Sensibilidade})}{\text{Precisão} + \text{Sensibilidade}} \quad (5)$$

$$\text{Especificidade} = \frac{TN}{TN + FP} \quad (6)$$

Onde: TP = Verdadeiros Positivos; TN = Verdadeiros Negativos; FP = Falsos Positivos; FN = Falsos Negativos; P_o = Proporção de concordância observada; P_e = Proporção de concordância esperada pelo acaso.

A análise de componentes principais (PCA) foi aplicada aos dados como uma ferramenta exploratória para observar as relações e variações dos atributos de qualidade em função dos tratamentos quanto ao nível de vigor.

Delineamento experimental e análise estatística

Os testes foram conduzidos em delineamento inteiramente casualizado (DIC) com quatro repetições para cada tratamento: Controle (0), 48 horas e 96 horas de envelhecimento artificial. Os dados foram submetidos à análise de variância, e para comprimento de parte aérea, radicular e total os dados foram transformados para $x' = \sqrt{x}$, para garantir o atendimento aos pressupostos da análise de variância. As médias obtidas para cada tratamento foram comparadas pelo teste de Tukey a 5% de probabilidade utilizando o software estatístico R 12.1 (R CORE TEAM, 2024). Para o modelo de classificação supervisionada (PLS-DA) dos dados espectrais, foi utilizado o pacote Nira (Medeiros *et al.*, 2020).

RESULTADOS E DISCUSSÃO

Foram observadas diferenças significativas entre os tratamentos de acordo com os testes de germinação e vigor, no qual é possível observar que as sementes submetidas à alta temperatura e umidade relativa apresentaram redução no potencial fisiológico quando comparadas ao controle (Tabela 1). Pelos resultados de germinação e primeira contagem de germinação, sementes do tratamento controle (não deterioradas) apresentaram desempenho superior às demais, enquanto sementes envelhecidas por 96 h apresentaram o pior desempenho. Dessa forma, foi possível estratificar os tratamentos em três níveis de qualidade fisiológica, ou seja, sementes de alto potencial fisiológico (controle), sementes com desempenho intermediário (envelhecidas por 48 h) e sementes de baixo potencial fisiológico (envelhecidas por 96 h).

Para a emergência de plântulas, resultados superiores foram obtidos para as sementes dos tratamentos controle e 48 h, que não diferiram entre si, com menores valores para sementes submetidas a 96 h. Já nos demais testes de qualidade fisiológica foi possível observar melhor desempenho para as sementes do controle em relação às dos tratamentos de envelhecimento por 48 e 96 h, que não diferiram estatisticamente entre si (Tabela 1).

Nesse contexto, pode-se considerar que altas temperaturas e umidade relativa aceleram o metabolismo celular e o consumo de reservas, o que favorece à incidência de patógenos, formação de espécies reativas de oxigênio prejudiciais tais como: radicais superóxidos e oxigênio singlete e a degradação por peroxidação lipídica, que fazem parte de rotas metabólicas importantes, e em algumas situações são essenciais, porém o desequilíbrio destas moléculas afeta negativamente causando deterioração, fato este já reportado em sementes de algodão por apresentarem alto teor de lipídeos em sua composição (Goel; Sheoran, 2003b; Sano, 2016).

Tabela 1. Grau de umidade (U), primeira contagem de germinação (PC), germinação (G), emergência (E), índice de velocidade de emergência (IVE), condutividade elétrica (CE), comprimento da parte aérea (P.A), comprimento da parte radicular (P.R), comprimento total (C.Total), área, perímetro da semente (Perim), Circularidade (Círc.), Proporção (Propor), solidez, round, densidade integrada (Dens. Int.) e densidade relativa (Dens. Rel) de sementes de algodão submetidas aos períodos de deterioração por 0 (controle), 48 e 96 horas.

Tratamentos	U -----%	PC	G	E	IVE índice	CE $\mu\text{S cm}^{-1} \text{ g}^{-1}$	Comp.PA -----cm-----	Comp.RA	Comp.Total
Controle	11 a	95 a	96 a	92 a	2,18 a	70,7 c	3.12 (9,88) a	4.05 (17.08) a	5.12 (26.96) a
48 horas	11 a	76 b	77 b	83 a	1,44 b	91,5 b	2.15 (4,83) b	3.10 (9.67) b	3.78 (14.49) b
96 horas	10.5 a	65 c	65 c	50 b	1,17 c	109 a	1.74 (3,05) b	2.23 (5,01) b	2.83 (8,06) b
C.V (%)	6.88	3.17	2.22	7.89	9.23	11.61	14.41	17.64	15.16

Tratamentos	Área cm^2	Perim cm	Circ -----índice-----	Propor	Solidez	Round	Dens Int. pixel	Dens. Int. Int. pixel cm^2	Dens. Rel Sum. Pixel
Controle	0.582 a	4.98 b	0.31 a	1.52 a	0.66 a	0.66 a	115.11 a	66.82 a	2794.39 a
48 horas	0.517 b	5.54 a	0.22 b	1.54 a	0.65 b	0.65 a	97.88 b	50.82 b	2125.56 b
96 horas	0.5034 b	4.83 b	0.27 a	1.51 a	0.67 ab	0.67 a	100.3 b	54.6 b	2283.45 b
C.V (%)	2.04	3.08	3.67	8.88	2.29	1.06	8.93	4.51	4.51

Médias originais entre parênteses para comprimento de plântulas. Médias seguidas pela mesma letra, em cada coluna, não diferem entre si pelo teste Tukey, a 5% de probabilidade.

O monitoramento da qualidade das sementes de algodão realizado por testes fisiológicos é de extrema importância para o estabelecimento de um plantio de sucesso. No entanto, tendo em vista o tempo de análise e o uso de metodologias baseadas na avaliação visual, buscam-se ferramentas e tecnologias neste monitoramento que garantam celeridade na seleção de lotes e também que possam agregar informações novas em uma análise mais profunda.

A partir do teste de Raios X, foi possível calcular medidas relacionadas a área, perímetro, circularidade, proporção, solidez e redondeza que trazem informações sobre a biometria das sementes, além disso foi possível estimar a densidade dos tecidos embrionários, uma vez que há relação entre a integridade destes com a densidade, como reportado em trabalhos com soja (Pinheiro *et al.*, 2021) e pinhão-manso (Bianchini *et al.*, 2021).

Pelo teste de Raios X, o controle apresentou uma área de $0,58 \text{ cm}^2$ e densidade integrada de $66,82 \text{ (pixels}^{-1}\text{)}$, que foram maiores quando comparadas aos tratamentos 48 e 96 horas, já para o perímetro e circularidade o tratamento 48 horas apresentou os maiores valores em relação ao controle e sementes envelhecidas por 96 horas, as características de proporção (razão entre comprimento e largura das sementes), solidez (relacionado a textura) e round (relacionado a forma da semente elipse ou circular) não apresentaram diferenças estatísticas, já para os

indicadores de densidade, foi possível observar que o tratamento controle apresentou os maiores valores em relação aos tratamentos 48 e 96 horas, indicando uma redução na densidade dos tecidos embrionários nas sementes que foram submetidas ao envelhecimento (Tabela 1).

Essa diminuição da densidade e do preenchimento pode estar relacionada à formação de plântulas anormais e sementes mortas (Corrêa, 2017; Oliveira; Carvalho; Davide, 2003) uma vez que tecidos íntegros apresentam maior radiopacidade do que tecidos deteriorados, as sementes que foram submetidas ao envelhecimento acelerado exibiram diminuição da densidade dos tecidos (Figura 1).

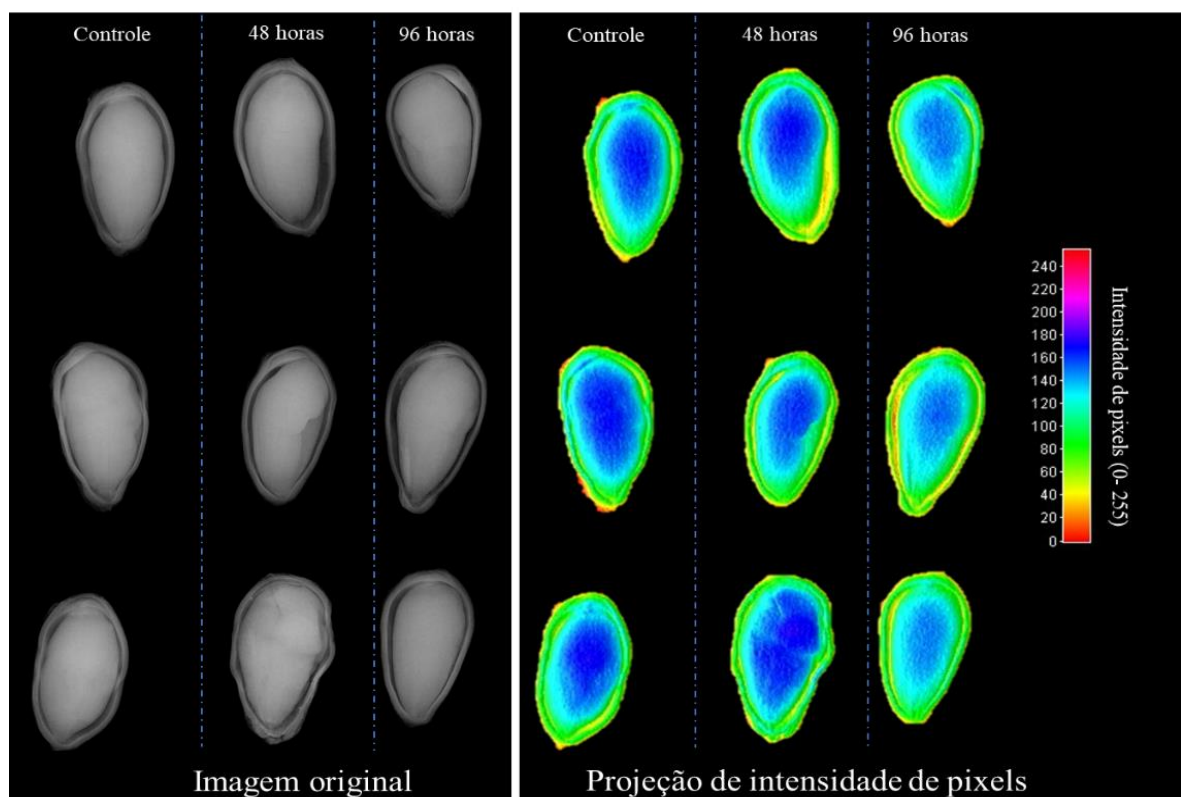


Figura 1. Raios X de sementes de algodão submetidas ao envelhecimento acelerado.

Pela análise de componentes principais (PCA), foi possível explicar 82,7% da variação nos dois primeiros componentes. O diagrama de ordenação central apresentou a disposição dos três tratamentos em relação ao nível de vigor. Paralelamente, foi possível observar os vetores dos atributos físicos e fisiológicos, que apresentaram correlação negativa com a variável bioquímica de condutividade elétrica (CE). A condutividade elétrica está relacionada à integridade das membranas, à capacidade de reorganização celular durante a embebição e à quantidade de lixiviados liberados. Nesse caso, quanto mais altos forem os valores de condutividade elétrica, menor será a qualidade das sementes (Vieira; Marcos-Filho, 2020; Finch-Savage *et al.*, 2013), o que confirma os resultados observados nas análises, já que sementes de baixa qualidade apresentam altos valores de condutividade elétrica. Além disso, os vetores das variáveis relacionadas à qualidade fisiológica mostraram forte correlação com sementes não envelhecidas (Figura 2).

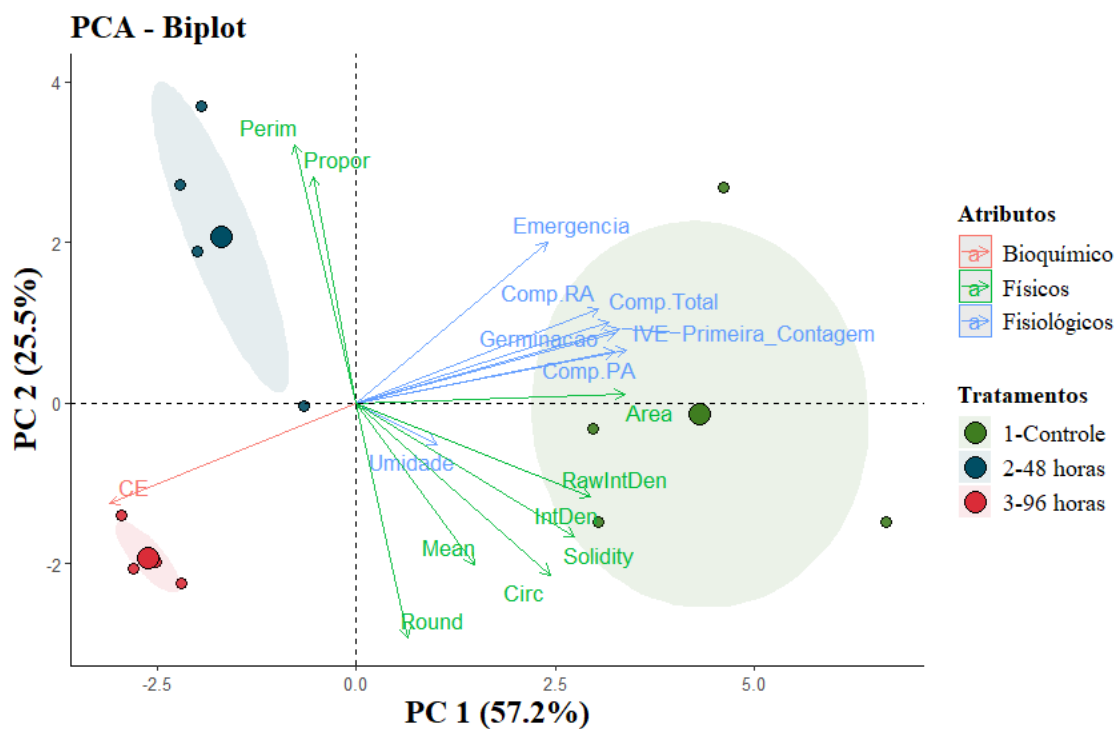


Figura 2. Biplot da análise de componentes principais (PCA) pela combinação das variáveis bioquímica físicas e fisiológicas e dos três tratamentos de sementes de algodão classificados quanto ao vigor.

Os tratamentos foram então definidos como: Alto vigor (Controle), médio vigor (sementes envelhecidas por 48 horas) e baixo vigor (sementes que foram envelhecidas por 96 horas) conforme os resultados dos testes de avaliação do potencial fisiológico. Foram então coletados 100 espectros por tratamento totalizando 300 espectros de infravermelho dos três

níveis de qualidade de sementes de algodão registrados na faixa entre 1.000 e 2.500 nm, totalizando 3112 bandas espectrais. Observa-se que nos espectros totais houve uma variação da intensidade da assinatura espectral em função do tratamento (Figura 3A). A média dos dados originais permitiu identificar de forma mais clara essas diferenças, sendo a classe de sementes considerada de alto vigor a que possui maior intensidade de absorbância, seguido pela classe de qualidade intermediária e baixa, respectivamente (Figura 3B).

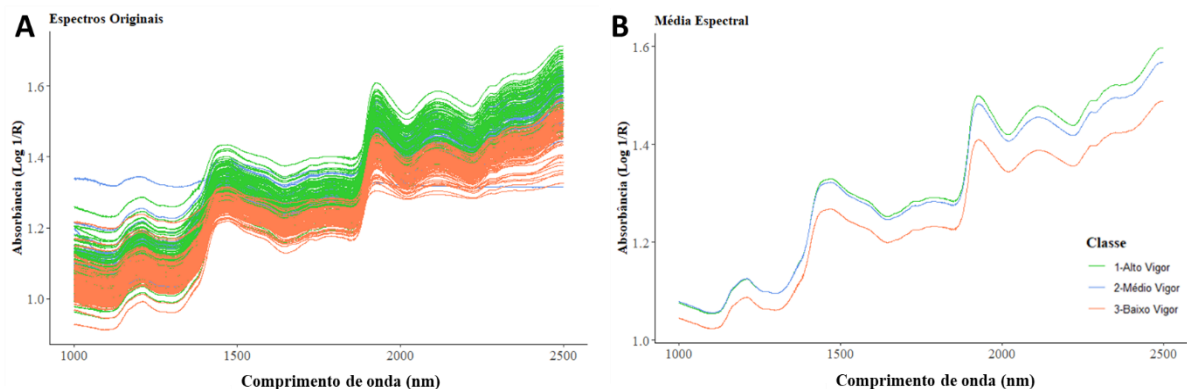


Figura 3. Espectros originais (A) e média espectral (B) de sementes de algodão submetidas ao envelhecimento acelerado por 0, 48 e 96 h.

Os pré-processamentos de dados espectrais têm sido úteis para reduzir o impacto dos ruídos instrumentais, magnitudes diferentes e alterações da linha de base que não contribuem para o modelo. O pré-processamento dos dados antes da classificação é capaz de aprimorar a característica buscada nos espectros e obter um melhor ajuste no modelo de classificação (Mortensen *et al.*, 2021).

Verificou-se, portanto, a necessidade de testar diferentes métodos de processamentos dos dados. Foram testados métodos de correção tais como: Centralização e Autoescalamento que são baseados numa média central, *Multiplicative Scatter Correction* (MSC) e *Standard Normal Variate* (SNV) e métodos derivativos (Savitzky-Golay) que reduzem a sobreposição de pico enquanto suaviza o sinal.

O pré-processamento recomendado para obter o modelo de calibração foi definido pela capacidade de predição da classe de qualidade fisiológica, avaliado pela acurácia e coeficiente *kappa*. A acurácia é frequentemente utilizada para avaliação de problemas de classificação de aprendizado de máquina. Essa métrica avalia a proporção da classificação correta sobre os resultados obtidos. Porém, é bastante influenciada pela proporção dos dados. Em dados desproporcionais ela pode ser maximizada por considerar informações menores como irrelevantes.

Como forma de confirmar os resultados, também foi calculado o coeficiente *kappa*, um método estatístico que descreve a intensidade da concordância entre dois conjuntos de dados. Um *kappa* de 1 indica concordância perfeita, enquanto um *kappa* de 0 indica concordância equivalente ao acaso (Vieira; Garrett, 2005). Na Tabela 2 encontram-se os valores de acurácia e *kappa* obtidos para os modelos de calibração multivariada, construído por meio da PLS-DA, a partir dos espectros NIR em diferentes estratégias de pré-processamento.

Tabela 2. Resultados de acurácia e coeficiente *kappa* para treinamento e teste dos diferentes métodos de pré-processamento obtidos pelos modelos de classificação via PLS-DA para os tratamentos de sementes de algodão.

Pré- Processamento	Treinamento (N=210)		Teste (N=90)	
	Acurácia	<i>Kappa</i>	Acurácia	<i>Kappa</i>
Dados Originais	0.95	0.93	0.94	0.92
Centralização	1.00	1.00	0.95	0.93
Autoescalonamento	0.94	0.91	0.96	0.93
MSC	0.84	0.76	0.86	0.78
SNV	0.82	0.73	0.86	0.78
1ª Derivada Savitzky-Golay (1; 2; 7)	1.00	1.00	0.93	0.90
2ª Derivada Savitzky-Golay (2; 2; 7)	1.00	1.00	0.79	0.68

n = número de espectros usados para treinamento e teste do modelo de classificação. MSC = Multiplicative Scatter Correction; SNV = Standard Normal Variate. Derivadas de Savitzky-Golay (ordem da derivada, polinômio, janela amostral) respectivamente.

Em geral, os métodos de processamento dos espectros melhoram de forma significativa a capacidade de predição do modelo. O melhor índice de classificação foi obtido pela centralização, sendo esta escolhida para o modelo final por apresentar índices de acurácia e *kappa* balanceados tanto para o treinamento como para o teste, sem que fossem observados *underfitting* ou *overfitting*.

Enquanto os espectros originais permitiram a classificação com acurácia de 95% e *kappa* de 93% para a etapa de treinamento e 94% e 92% para o teste respectivamente, a centralização atingiu acurácia de 100% e *kappa* de 100% para etapa de treinamento e 95% e 93% para o teste (Tabela 2). A interpretação de Landis e Koch (1977) para este valor de *kappa* considera a concordância quase perfeita (0,80 – 0,99), e demonstra que o pré-processamento teve boa capacidade de discriminar classes de sementes de algodão em função da deterioração.

Na figura 4A observa-se os 300 espectros das sementes de algodão e suas respectivas classes, bem como na figura 4B as médias de cada classe de vigor após o processamento utilizando a primeira derivada de centralização.

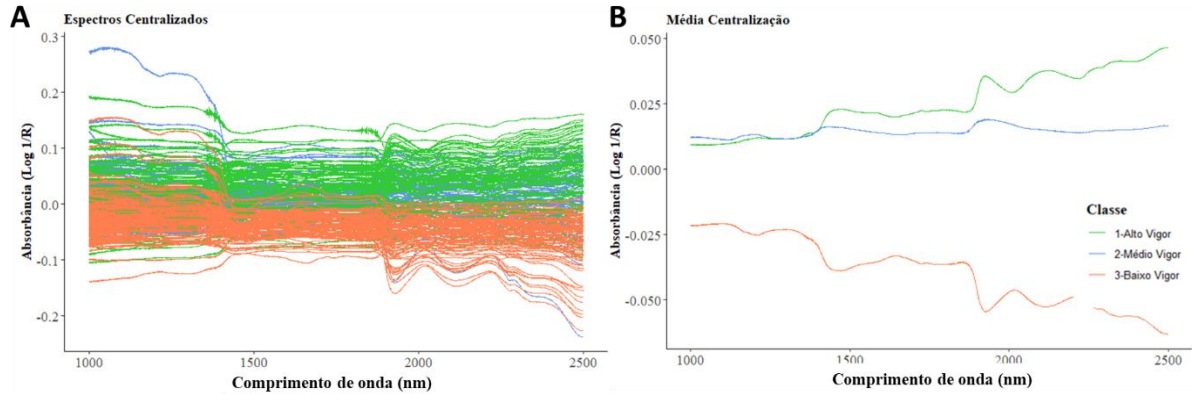


Figura 4. Espectros de sementes de algodão submetidas ao envelhecimento acelerado por 0, 48 e 96 h após processamento utilizando o método de centralização

Utilizando imagens hiperespectrais NIR-HSI Soares *et al.* (2016), obtiveram resultados semelhantes para sementes de algodão com uma acurácia de 98% utilizando o modelo PLS-DA na classificação de variedades e qualidade fisiológica de sementes.

Fica evidente a separação da classe de alto vigor (controle) com as classes de médio e baixo vigor (48 e 96 horas, respectivamente) a partir de 1500 nm. É importante ressaltar que entre 1500-1700 nm ocorre o primeiro *overtone* de estiramento ligados a molécula C-H (hidrocarbonetos) e O-H (água e álcool), 1700-1900 nm há uma relação de estiramento com ligações do tipo N-H (aminas e proteínas) e 1900-2500, ligações do tipo C-H, C=O (lipídeos e hidrocarbonetos) (Santos; Bonfim, 2023).

As principais reservas de armazenamento em sementes de algodão são constituídas em média por 39% de proteínas, 33% de lipídeos e 15% de carboidratos, Marcos-Filho (2015), relata que a deterioração geralmente é associada a uma série de alterações metabólicas, quem em sua maioria são deletérias, alta umidade e temperatura influenciam diretamente na conformação das moléculas de reservas, sendo proteínas e lipídeos mais sensíveis a essas alterações. Neste caso, foi observado que as sementes que foram submetidas ao envelhecimento acelerado apresentaram diferenças de assinatura espectral quando comparada com o tratamento controle que foi considerado de alto vigor, fato este que pode estar associado à peroxidação lipídica e desnaturação de proteínas, porém vale ressaltar que em sementes de gramíneas por exemplo, o acúmulo de açúcares redutores, apresenta efeitos mais pronunciados (Biachini *et al.*, 2001).

Essas alterações metabólicas ocasionadas pelo envelhecimento acelerado refletem diretamente na assinatura espectral das sementes, permitindo que modelos de aprendizado de máquina, como o PLS-DA, sejam empregados para identificar padrões e classificar o vigor das sementes com alta precisão. No presente estudo, o modelo PLS-DA foi aplicado para classificar os espectros obtidos, e os resultados demonstraram acurácias consistentes na distinção das sementes de algodão, reforçando a capacidade desse método em detectar variações associadas à qualidade fisiológica, principalmente, neste caso, relacionados a deterioração causadas por alta temperatura e umidade.

Desta forma, é possível observar que o modelo teve bom desempenho na classificação na etapa de treinamento, as medidas de desempenho para todas as classes: Alto, médio e baixo vigor, relacionados aos tratamentos controle, 48 e 96 horas de envelhecimento acelerado respectivamente, foram >90% (Figura 5A), analisando a matriz de confusão, 62 sementes foram corretamente classificadas como de alto vigor, enquanto 4 foram incorretamente atribuídas à classe de médio vigor e outras 4 à classe de baixo vigor. Para a classe de médio vigor, 3 sementes foram equivocadamente classificadas como de alto vigor (Figura 5B). Para as métricas do teste, os valores foram >80%, mantendo assim, uma boa relação entre treinamento e teste na previsão de novas amostra (Figura 5C), e na matriz de confusão, 25 sementes da classe alto vigor foram corretamente classificadas e 5 delas classificadas de forma incorreta atribuídas a classe baixo vigor, para a classe baixo vigor, apenas 2 sementes foram alocadas na classe de médio vigor (Figura 5D).

Esses resultados reforçam a importância do desenvolvimento de metodologias de análise mais rápidas e precisas para a tomada de decisão, complementando os testes de qualidade tradicionais. A separação eficiente de lotes de sementes de alto, médio e baixo vigor é essencial para o mercado de sementes, permitindo a oferta de lotes com qualidade diferenciada e maior previsibilidade no desempenho das sementes em campo.

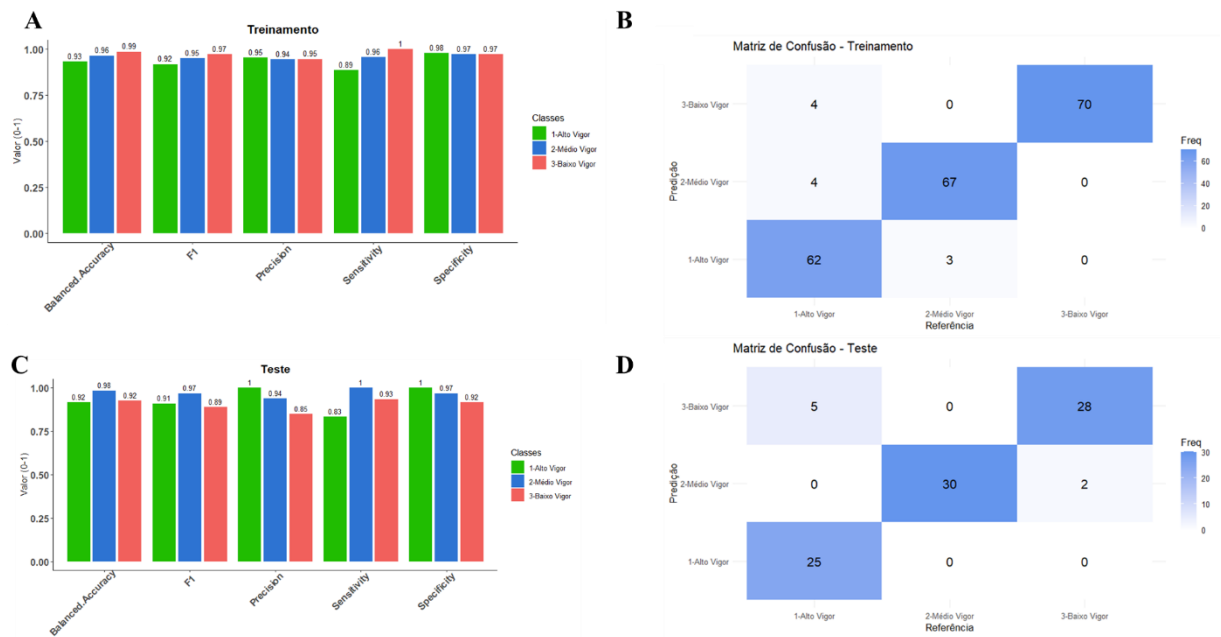


Figura 5. Métricas de desempenho e matriz de confusão da classificação dos tratamentos de sementes de algodão em níveis de vigor por meio do modelo PLS-DA utilizando os espectros após o processamento de centralização.

Os comprimentos de onda que mais contribuíram para a classificação das sementes em diferentes níveis de qualidade, com base na importância das variáveis para a separação das classes de vigor, estão ilustrados na Figura 6. A linha em vermelho representa a contribuição de pelo menos 50% para a classificação do modelo, destacando as faixas espectrais mais influentes na análise. Analisando as contribuições espectrais, observa-se que as regiões de comprimento de onda entre 1000, 1400, 1900, 2050 e 2500 nm emergem como as mais importantes para a classificação das sementes em diferentes níveis de qualidade. Essas faixas espectrais estão associadas a picos de reflectância de compostos químicos essenciais, como água, carboidratos, proteínas e lipídeos, que possuem características espectrais específicas.

Por exemplo, a faixa entre 1400 e 1900 nm é amplamente reconhecida pela absorção de água (estiramento H-O). A água é um componente fundamental das sementes, desempenhando papéis cruciais nas reações bioquímicas e no processo de metabolismo. Durante a deterioração das sementes, a atividade metabólica diminui, e a água livre disponível nas sementes é alterada, o que influencia diretamente suas propriedades espectrais (Carvalho; Nakagawa, 2000). O monitoramento da reflectância nesses comprimentos de onda permite detectar mudanças na quantidade e no estado da água nas sementes, fornecendo um indicador precoce do processo de envelhecimento. Estudos como o de Dufour (2009) indicam que a deterioração das sementes

está frequentemente associada à perda de umidade controlada, e essas mudanças espectrais podem ser utilizadas para avaliar a qualidade das sementes.

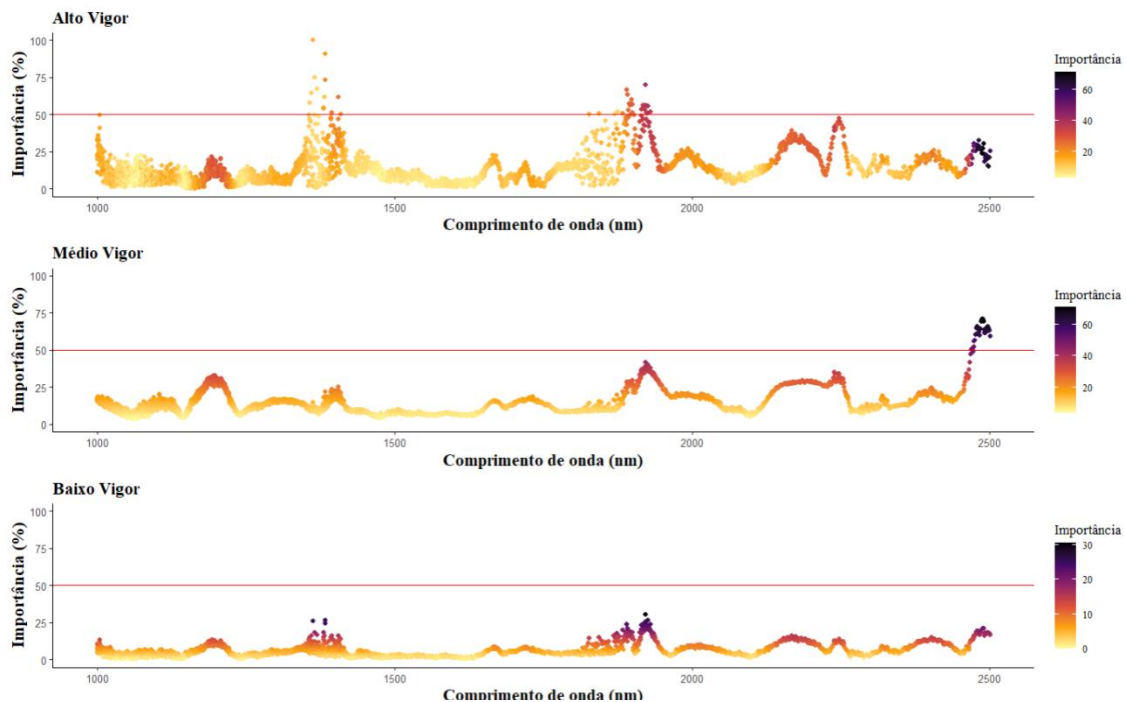


Figura 6. Importância das variáveis de comprimento de ondas para classificação via PLS-DA dos níveis de vigor de sementes de algodão. Linha vermelha corresponde a 50% de contribuição.

Além disso, a faixa entre 1000 e 1400 nm está relacionada a estiramentos assimétricos de ligações C-H, particularmente associados a carboidratos, compostos que são importantes fontes de energia para as sementes. Durante o processo de deterioração, esses compostos, especialmente os açúcares solúveis, podem ser degradados devido à ação de enzimas, afetando as propriedades fisiológicas da semente. A reflectância nessa região pode, portanto, refletir a integridade estrutural e funcional dos carboidratos, sendo um marcador da saúde das sementes.

Já as regiões espectrais em torno de 2050 e 2500 nm estão associadas a estiramentos de ligações N-H (proteínas) e C-H (lipídeos). As proteínas e lipídeos são componentes essenciais das células das sementes, envolvidos em processos metabólicos e de estrutura celular. Durante a deterioração das sementes, esses compostos sofrem modificações, como a oxidação dos lipídeos e a desnaturalização das proteínas, o que compromete a funcionalidade e a viabilidade das sementes. A espectroscopia nas faixas de 2050 e 2500 nm pode, assim, ajudar a monitorar esses processos de degradação, fornecendo informações detalhadas sobre o estado de conservação das sementes.

Os trabalhos de Dufour (2009) e Mehrotra (2000), corroboram essa relação entre a deterioração de compostos químicos nas sementes e os padrões espectrais observados. A

degradação da água, carboidratos, proteínas e lipídeos resulta em alterações nos parâmetros espectrais que podem ser detectadas por espectroscopia, oferecendo uma maneira eficaz e não destrutiva de avaliar a qualidade das sementes. Além disso, os estudos indicam que a deterioração das sementes é um processo complexo e multifatorial, influenciado por fatores ambientais e fisiológicos (Krzyzanowski; Marcos-Filho, 2020)

Portanto, ao analisar as contribuições espectrais dessas faixas de comprimento de onda, é possível não apenas classificar as sementes em diferentes níveis de qualidade, mas também entender melhor os processos bioquímicos e fisiológicos envolvidos na deterioração das sementes. Essa abordagem proporciona uma ferramenta poderosa para a avaliação rápida e precisa da qualidade das sementes, com implicações diretas para o armazenamento e uso de sementes em programas de melhoramento genético e agricultura de precisão.

Apesar de ser desafiador quantificar compostos específicos em sementes usando NIR devido à sobreposição de bandas espectrais relacionadas a mais de um composto (Dufour, 2009), o modelo PLS-DA, permitiu classificar as sementes de algodão com base em seu potencial fisiológico. Como isso, foi possível ao analisar a resposta vibracional das ligações químicas à radiação NIR e identificar as regiões mais importantes do espectro eletromagnético para distinguir os diferentes níveis de vigor das sementes.

Esse estudo demonstra que é possível calibrar equipamentos NIR para prever o potencial fisiológico das sementes de algodão. Mesmo que essa técnica não substitua os testes convencionais, a espectroscopia no infravermelho próximo pode ser utilizada para uma triagem rápida e como um teste complementar para o descarte de lotes com baixa qualidade nas diferentes etapas do processo de produção de sementes. Hoje já existem equipamentos acessíveis tanto de bancada como também portáteis para aquisição dos espectros no setor sementeiro. Além disso, o custo para a obtenção dos dados é relativamente baixo. A incorporação dessa etapa ao processo produtivo pode dar celeridade ao setor, principalmente nas etapas de recepção nas unidades de beneficiamento, auxiliando na seleção rápida de lotes, uma vez que os testes de germinação e vigor são mais demorados. A agilização nas tomadas de decisão interna quanto ao manejo de lotes é fundamental, evitando gastos desnecessários com processamento de lotes de baixo potencial fisiológico, resultando em economia de tempo, espaço e recursos.

CONCLUSÃO

A espectroscopia no infravermelho próximo (NIR) foi eficiente para classificar sementes de algodão em diferentes níveis de deterioração.

O modelo PLS-DA obtido por meio os dados espectrais NIR, aplicando a centralização como pré-processamento, resultou em 95% acurácia na predição de sementes com diferentes classes de vigor.

REFERÊNCIAS

- ALVARENGA, R. O., Marcos-Filho, J., & Gomes Junior, F. G. (2012). Avaliação do vigor de sementes de milho superdoce por meio da análise computadorizada de imagens de plântulas. *Revista Brasileira de Sementes*, 34, 488-494.
- AMBROSE, A. *et al.* Comparative nondestructive measurement of corn seed viability using Fourier transform near-infrared (FT-NIR) and Raman spectroscopy. *Sensors and Actuators, B: Chemical*, 2016. v. 224, p. 500–506.
- ANDERSON, M.D.; PRASAD, T.K., STEWART, C.R. Changes in isozyme profiles of catalase, peroxidase, and glutathione reductase during acclimation to chilling in mesocotylus of maize seedlings. *Plant Physiology*, 109:1247-1257, 1995.
- ASSOCIATION OF OFFICIAL SEED ANALYSIS (AOSA), ed. *Seed Vigor testing handbook*. Ithaca, N. Y., 2009 (Contribution nº 32 to the Handbook on Seed Testing).
- BAZONI, C. H.; IDA, E. I.; BARBIN, D. F.; KUROZAWA, L. E. Near-infrared spectroscopy as a rapid method for evaluation physicochemical changes of stored soybeans. *Journal of Stored Products Research*, v.73, p. 1-6, 2017. <https://www.sciencedirect.com/science/article/abs/pii/S0022474X17301315>
- BEAUCHAMP, C., FRIDOVICH, I. Superoxide dismutase: improved assays and an assay applicable to acrylamide gels. *Analytical Biochemistry*, 44:276-287, 1971.
- BIANCHINI, A.L.; BRACCINI, M.C.L.; SCAPIM, C.A.; Mecanismos de deterioração de sementes. *Informativo ABRATES*, v.11, n.1, p.10-15, 2001.

- BIANCHINI, V. De J. M. *et al.* Multispectral and X-ray images for characterization of *Jatropha curcas* L. seed quality. *Plant Methods*, 2021. v. 17, n. 1, p. 1–14. Disponível em: <<https://doi.org/10.1186/s13007-021-00709-6>>.
- BRASIL. Ministério da Agricultura, Pecuária e Abastecimento. Instrução Normativa nº 45, de 17 de setembro de 2013. Estabelece critérios para o controle da produção e da comercialização de sementes e mudas no Brasil. *Diário Oficial da União: Brasília, DF*, 20 set. 2013. Disponível em: https://www.gov.br/agricultura/pt-br/assuntos/insumos-agropecuarios/insumos-agricolas/sementes-e-mudas/publicacoes-sementes-emudas/copy_of_INN45de17desetembrode2013.pdf Acesso em: 9 dez. 2024.
- BRASIL. Ministério da Agricultura, Pecuária e Abastecimento. Regras para análise de sementes. Ministério da Agricultura, Pecuária e Abastecimento. Secretaria de Defesa Agropecuária. Brasília, DF: MAPA/ ACS, 2009. 395p.
- CAKMAK, C.J.; HORST, W.J. Effect of aluminum on lipid peroxidation, superoxide dismutase, catalase, and peroxidase activities in root tips of soybean (*Glycine max*). *Physiologia Plantarum*, v.83, p.463–468, 1991
- CARREIRO SOARES, S. F. *et al.* Classification of individual cotton seeds with respect to variety using near-infrared hyperspectral imaging. *Analytical Methods*, 2016. v. 8, n. 48, p. 8498–8505.
- CARVALHO, N. M.; NAKAGAWA, J. **Sementes: ciência, tecnologia e produção**. Jaboticabal: FUNEP, 2000. 429p.
- CHANCE, B., MAEHLEY, A.C. Assay of catalases and peroxidases. *Methods in Enzymology*, 2:764-775, 1955.
- CONAB. Companhia Nacional de Abastecimento. Disponível em: <https://www.conab.gov.br/info-agro/safras/serie-historica-das-safras/itemlist/category/898-algodao>. Acesso em 30/12/2024.
- CORDEIRO, Maria Barbosa da Silva. Espectroscopia NIR e algoritmos de seleção de variáveis para determinação de teor de água e classificação de sementes de algodão. 2023.
- CORRÊA, Plínio Duarte. Avaliação da morfologia interna e do vigor de sementes de algodão utilizando técnicas de análise de imagens. 2017. Tese de Doutorado. Universidade de São Paulo.
- CUI, H. F. *et al.* Automatic and rapid discrimination of cotton genotypes by near infrared spectroscopy and chemometrics. *Journal of Analytical Methods in Chemistry*, 2012. v. 1, n. 1.
- DEL LONGO, O.T., GONZÁLEZ, A., PASTORI, G.M., TRIPPI, V.S. Antioxidant defenses under hyperoxygenic and hyperosmotic conditions in leaves of two lines of maize with differential sensitivity to drought. *Plant Cell Physiology*. 34:1023-1028, 1993.
- DIAS, D.C.F.S.; MARCOS-FILHO, J. Teste de germinação a baixa temperatura. In: KRZYŻANOWSKI, F.C.; VIEIRA, R. D.; FRANÇA-NETO, J.B.; MARCOS-FILHO, J. **Vigor de sementes: conceitos e testes**. Londrina: ABRATES, c. 7, p.318-332, 2020.

- DUFOUR, É. Principles of infrared spectroscopy. *In*: SUN, D. (Org.). Infrared Spectroscopy for Food Quality Analysis and Control. 1. ed. National University of Ireland, Dublin: Elsevier Inc., 2009. p. 3-26.
- EGIDIO, V. DI *et al.* NIR and MIR spectroscopy as rapid methods to monitor red wine fermentation. *European Food Research and Technology*, 2010. v. 230, n. 6, p. 947–955.
- EKRONY, D.M.; EGLI, D.B. Relationship of seed vigor to crop yield: A review. *Crop Science*, v.31, p.816-822, 1991
- VIEIRA, R. D.; KRZYZANOWSKI, F. C. Teste de condutividade elétrica. *In*: Krzyzanowski, F. C.; Vieira, R. D.; França-Neto, J. (Eds.). Vigor de sementes: conceitos e testes. Londrina: ABRATES, Cap.4, 1999.
- ELMASRY, G. *et al.* Recent applications of multispectral imaging in seed phenotyping and quality monitoring—An overview. *Sensors (Switzerland)*, 2019. v. 19, n. 5, p. 1–32.
- FINCH-SAVAGE, Bill. Seeds: Physiology of development, germination and dormancy (3rd edition) - J.D. Bewley, K.J. Bradford, H.W.M. Hilhorst H. Nonogaki. 392 pp. Springer, New York – Heidelberg – Dordrecht – London 2013 978-1-4614-4692-7. *Seed Science Research*, [s. l.], v. 23, n. 4, p. 289, 2013. Disponível em: <https://doi.org/10.1017/s0960258513000287>.
- GAITÁN-JURADO, A. J. *et al.* Near infrared applications in the quality control of seed cotton. *Journal of Near Infrared Spectroscopy*, 2008. v. 16, n. 4, p. 421–429.
- GIANNOPOLITIS, C.N., RIES, S.K. Superoxide dismutases: Occurrence in higher plants. *Plant Physiology*, 59:309–314, 1977.
- GOEL, Anuradha; GOEL, Ajay Kumar; SHEORAN, Inder Singh. Changes in oxidative stress enzymes during artificial ageing in cotton (*Gossypium hirsutum* L.) seeds. *Journal of plant physiology*, v. 160, n. 9, p. 1093-1100, 2003a.
- GOULART, Augusto César Pereira. Fungos em Sementes de Soja Detecção, Importância e Controle. 2. ed. rev. e aum. Brasília/DF: Embrapa, 2018. 71 p. Disponível em: <https://www.embrapa.br/busca-de-publicacoes/-/publicacao/1097768/fungos-em-sementes-desoja-deteccao-importancia-e-controle>. Acesso em: 15 dez. 2023.
- GRUNVALD, A. K. *et al.* Predição do teor de óleo em sementes de genótipos de girassol por espectroscopia no infravermelho próximo (NIR). *Acta Scientiarum - Agronomy*, 2014. v. 36, n. 2, p. 233–237.
- HAVIR, E.A., McHALE, N.A. Biochemical and developmental characterization of multiple forms of catalase in tobacco leaves. *Plant Physiology*, 84:450-455, 1987.
- HUANG, Z.; SHA, S.; RONG, Z.; CHEN, J.; HE, Q.; KHAN, D. M.; ZHU, S. Feasibility study of near infrared spectroscopy with variable selection for non-destructive determination of quality parameters in shell-intact cottonseed. *Industrial Crops and Products*, v. 43, p. 654-660, 2013. <https://www.sciencedirect.com/science/article/abs/pii/S0926669012004852>
- ISHIKAWA, H. *et al.* Development of calibration model to predict nitrogen content in single seeds of cowpea (*Vigna unguiculata*) using near infrared spectroscopy. *Journal of Near Infrared Spectroscopy*, 2017. v. 25, n. 3, p. 211–214.

- JIA, S. *et al.* Variety identification method of coated maize seeds based on near-infrared spectroscopy and chemometrics. *Journal of Cereal Science*, 2015. v. 63, p. 21–26. Disponível em: <<http://dx.doi.org/10.1016/j.jcs.2014.07.003>>.
- KAR, M., MISHRA, D. Catalase, peroxidase, and polyphenoloxidase activities during rice leaf senescence. *Plant Physiology*, 57:315-319, 1976.
- KOSHIBA, T. Cytosolic ascorbate peroxidase in seedling and leaves of maize (*Zea mays*), *Plant Cell Physiology*, 34:713-481, 1993.
- KRZYZANOWSKI, F.C *et al.* Testes de vigor baseado em desempenho de plântulas. In: KRZYZANOWSKI, F.C.; VIEIRA, R. D.;FRANÇA-NETO, J.B.; MARCOS-FILHO, J. **Vigor de sementes: conceitos e testes**. Londrina: ABRATES, p.79-127, 2020.
- KRZYZANOWSKI, F.C.; MARCOS-FILHO, J. Deterioração controlada. In: KRZYZANOWSKI, F.C.; VIEIRA, R. D.;FRANÇA-NETO, J.B.; MARCOS-FILHO, J. **Vigor de sementes: conceitos e testes**. Londrina: ABRATES, c. 7, p.248-275, 2020.
- LIMÃO, M. A. R.; DIAS, D. C. F. S.; NASCIMENTO, W. M.; NORONHA, B. G.; SOARES, J. M.; SILVA, L. J. da. Near-infrared spectroscopy and accelerated ageing in evaluating the vigour of lentil seeds. *Revista Ciência Agronômica*, v. 56, e202493547, 2025. <https://doi.org/10.5935/1806-6690.20250034>
- MAGUIRE, J. D. Speed of germination – aid in selection and evaluation for seedling emergence and vigor. *Crop Science*, v. 2, n. 2, p. 176-177, 1962.
- MARCOS FILHO, J. *Fisiologia de Sementes de Plantas Cultivadas*. 2ª edição. ABRATES, Londrina, 2015.
- MARCOS FILHO, J. Teste de envelhecimento acelerado. In: Krzyzanowski, F.C.; Vieira, R.D.; França-Neto, J.B. (Eds.). *Vigor de sementes: conceitos e testes*. Londrina: ABRATES. Cap.3, p1-24, 1999.
- MARCOS-FILHO, J. Teste de envelhecimento acelerado. In: KRZYZANOWSKI, F.C.; VIEIRA, R.D.; FRANÇA-NETO, J.B. (Ed.). *Vigor de sementes: conceitos e testes*. Londrina: ABRATES, 2020. Cap. 3, p.3.1-3.24.
- MATTIONI, F. *et al.* Vigor de sementes e desempenho agrônômico de plantas de algodão. *Revista Brasileira de Sementes*, 2012. v. 34, n. 1, p. 108–116.
- MAYRINCK, Livia Giro *et al.* Use of near infrared spectroscopy in cotton seeds physiological quality evaluation. *Journal of Seed Science*, v. 42, 2020.
- MEDEIROS, A.D.; PEREIRA, M.D. SAPL®: a free software for determining the physiological potential in soybean seeds. *Pesquisa Agropecuária Tropical*, Goiânia, v. 48, n. 3, p. 222-228, 2018.
- MEDEIROS, A.D.d.; SILVA, L.J.d.; RIBEIRO, J.P.O.; FERREIRA, K.C.; ROSAS, J.T.F.; SANTOS, A.A.; SILVA, C.B.d. Machine Learning for Seed Quality Classification: An Advanced Approach Using Merger Data from FT-NIR Spectroscopy and X-ray Imaging. *Sensors* 2020, 20, 4319. Doi: <https://doi.org/10.3390/s20154319>

- MEDEIROS, M. L. S; CRUZ-TIRADO, J. P; LIMA, A. F; NETTO, J. M. RIBEIRO, A. P. B; BASSEGIO, D; GODOY, H. T; BARBIM, D. F. Assessment oil composition and species discrimination of Brassicas seeds based on hyperspectral imaging and portable near infrared (NIR) spectroscopy tools and chemometrics. *Journal of Food Composition and Analysis*, v. 107, p. 104403, 1 abr. 2022. <https://doi.org/10.1016/j.jfca.2022.104403>
- MEHROTRA, R. (2000). *Infrared Spectroscopy, Gas Chromatography/Infrared in Food Analysis*. *Encyclopedia of Analytical Chemistry*. Doi: 10.1002/9780470027318.a1013.
- MORTENSEN, A.K.; GISLUM, R.; JØRGENSEN, J.R.; BOELT, B. The Use of Multispectral Imaging and Single Seed and Bulk Near-Infrared Spectroscopy to Characterize Seed Covering Structures: Methods and Applications in Seed Testing and Research. *Agriculture* 2021, 11, 301. <https://doi.org/10.3390/agriculture11040301>
- NAGEL, M.; BÖRNER, A. The longevity of crop seeds stored under ambient conditions. *Seed Science Research*, 2010. v. 20, n. 1, p. 1–12.
- NAKAGAWA, J. Testes de vigor baseados no desempenho das plântulas. In: KRZYZANOWSKI, F.C.; VIEIRA, R.D.; FRANÇA-NETO, J.B. (Eds.). *Vigor de sementes: conceitos e testes*. Londrina: ABRATES, 1999. cap. 2, p.2.1-2.24.
- NAKANO, Y., ASADA, K. Hydrogen peroxide is scavenged by ascorbate-specific peroxidases in spinach chloroplasts. *Plant Cell Physiology*, 22:867–880, 1981.
- NORONHA, B. G. D., MEDEIROS, A. D. D., & PEREIRA, M. D. (2018). Avaliação da qualidade fisiológica de sementes de *Moringa oleifera* Lam. *Ciência Florestal*, 28, 393-402.
- OLIVEIRA, A.C.S.; MARTINS. G.N.; SILVA, R.F.; VIEIRA, H.D. Testes de vigor em sementes baseados no desempenho de plântulas. *Revista Científica Internacional*, v.2, p.1-21, n.4, 2009.
- OLIVEIRA, Luciana Magda de; CARVALHO, Maria Laene Moreira de; DAVIDE, Antonio Claudio. Utilização do teste de raios-X na avaliação da qualidade de sementes de canafístula (*Peltophorum dubium* (Sprengel) Taubert). *Revista Brasileira de Sementes*, v. 25, p. 116-120, 2003.
- PANDE, R.; MISHRA, H. N. Fourier Transform Near-Infrared Spectroscopy for rapid and simple determination of phytic acid content in green gram seeds (*Vigna radiata*). *Food Chemistry*, 2015. v.148, p. 880–884. Disponível em: <<http://dx.doi.org/10.1016/j.foodchem.2014.09.049>>.
- PEIXOTO, P.H.P., CAMBRAIA, J., SANTANNA, R., MOSQUIM, P.R., MOREIRA, M.A. Aluminum effects on lipid peroxidation and on the activities of enzymes of oxidative metabolism in sorghum. *Revista Brasileira de Fisiologia Vegetal*, 11:137-143, 1999.
- RAMALHEIRO, J. P. da S. C. Contribuição para caracterização bioquímica do estado de maturação de azeitonas de diferentes variedades. Dissertação de Mestrado em Engenharia Alimentar – Qualidade e Segurança Alimentar, universidade Técnica de Lisboa. 2009. 51p

- RIBEIRO, J. P. O; MEDEIROS, A. D; CALIARI, I. P; TRANCOSO, A.C; MIRANDA, R. M; FREITAS, F. C. L; SILVA, L. J; DIAS, D. C. F. S. FT-NIR and linear discriminant analysis to classify chickpea seeds produced with harvest aid chemicals. *Food Chemistry*, v. 342, p. 128324, 16 abr. 2021
- RINGSTED, T. *et al.* Near-Infrared Spectroscopy Using a Supercontinuum Laser: Application to Long Wavelength Transmission Spectra of Barley Endosperm and Oil. *Applied Spectroscopy*, 2016. v. 70, n. 7, p. 1176–1185.
- ROSSATO, R. *et al.* Predicting rapeseed oil content with near-infrared spectroscopy. *Pesquisa Agropecuária Brasileira*, 2013. v. 48, n. 12, p. 1601–1604.
- SAEED, M. F. *et al.* Storage conditions deteriorate cotton and wheat seeds quality: An assessment of Farmers' awareness in Pakistan. *Agronomy*, 2020. v. 10, n. 9.
- SANO, N. *et al.* Staying alive: Molecular aspects of seed longevity. *Plant and Cell Physiology*, 57(4), 660-674. 2016. Doi: <https://doi.org/10.1093/pcp/pcv186>
- SANTOS, M. B. H. *et al.* Non-destructive NIR spectrometric cultivar discrimination of castor seeds resulting from breeding programs. *Journal of the Brazilian Chemical Society*, 2014. v. 25, n. 5, p. 969–974.
- SANTOS, S. F. dos; BOMFIM, M. A. D. *Instrumentação de espectroscopia no infravermelho próximo para previsão bromatológica do farelo de trigo*. *Revista Agrária Acadêmica*, v. 6, n. 3, p. 70–81, 2023. Disponível em: <https://agrariacad.com>. Acesso em: 19 mar. 2025.
- SCHNEIDER, C., RASBAND, W. & ELICEIRI, K. NIH Image to ImageJ: 25 years of image analysis. *Nat Methods* **9**, 671–675 (2012). <https://doi.org/10.1038/nmeth.2089>
- SHRESTHA, S.; KNAPIČ, M.; ŽIBRAT, U.; DELEURAN, L. C.; GISLUM, R. Single seed near-infrared hyperspectral imaging in determining tomato (*Solanum lycopersicum* L.) seed quality in association with multivariate data analysis. *Sensors and Actuators B: Chemical*, v. 237, p. 1027-1034, 2016. <https://www.sciencedirect.com/science/article/abs/pii/S0925400516314034>
- SILVA, M. F. da *et al.* Artificial aging for predicting the storability of soybean seeds via GGE biplot. *Journal of Seed Science*, v. 46, p. e202446001, 2024.
- SILVA, M. F. da *et al.* Near infrared spectroscopy for the classification of vigor level of soybean seed. *Revista Ciência Agronômica*, v. 55, p. e20238703, 2023.
- SILVA, P. P., Freitas, R. A., Cícero, S. M., Marcos-Filho, J., & Nascimento, W. M. (2014). Análise de imagens no estudo morfológico e fisiológico de sementes de abóbora. *Horticultura Brasileira*, 32(2), 210-214.
- SOARES, Júlia Martins *et al.* Near-infrared spectral evaluation of physiological potential, biochemical composition and enzymatic activity of soybean seeds. **Journal of Seed Science**, [s. l.], v. 46, 2024. Disponível em: <https://www.scielo.br/j/jss/a/MGpcZp4zYGLsCKcQF4SVvXp/?lang=en>.

VIEIRA, R. D; MARCOS-FILHO, J. Teste de condutividade elétrica. In: KRZYZANOWSKI, F.C.; VIEIRA, R. D.; FRANÇA-NETO, J.B.; MARCOS-FILHO, J. **Vigor de sementes: conceitos e testes**. Londrina: ABRATES, p.182-244, 2020.

ZHU, S. *et al.* Near-infrared hyperspectral imaging combined with deep learning to identify cotton seed varieties. *Molecules*, 2019. v. 24, n. 18.

Capítulo II

CLASSIFICAÇÃO DO POTENCIAL FISIOLÓGICO DE SEMENTES DE ALGODÃO POR IMAGENS MULTIESPECTRAIS E MODELOS DE APRENDIZADO SUPERVISIONADO

RESUMO:

A obtenção de sementes de alta qualidade é um desafio crucial na produção de algodão. Testes tradicionais como de germinação e vigor são trabalhosos e em grande parte subjetivos, destacando a necessidade de ferramentas complementares aos testes que sejam rápidos e menos onerosos. Imagens multiespectrais surgem como uma alternativa promissora, fornecendo informações espaciais e espectrais sobre atributos morfológicos e composição química. Este estudo avaliou o uso de imagens multiespectrais para classificar a qualidade de sementes de algodão. Sementes de um lote da região de Catuti-MG foram submetidas a envelhecimento acelerado em câmara BOD a 41 °C por 24, 48, 72 e 96 horas, além do controle sem envelhecimento. Foi desenvolvido um protótipo para captura de imagens espectrais de baixo custo: Croma Seeds. As imagens, capturadas em comprimentos de onda específicos (395–940 nm), foram processadas no ImageJ para obter dados de reflectância. Modelos preditivos foram desenvolvidos utilizando o Orange Data Mining e algoritmos de classificação supervisionada tais como *Tree*, *Random Forest*, *Gradient Boosting*, *Support Vector Machine (SVM)* e *Neural Network*. Foi possível observar que nos testes tradicionais de germinação e vigor, houve diferença estatística entre o tratamento controle e sementes que foram submetidas ao envelhecimento acelerado, observando uma queda acentuada na qualidade fisiológica de sementes, sobretudo para o período de 96 horas e aumento na condutividade elétrica. A análise de componentes principais (PCA) explicou 88.62% da variância, diferenciando os tratamentos. O modelo de Neural Network e SVM demonstraram eficácia quando comparado aos outros modelos, com acurácia de 100% e *kappa* de 99%, no teste utilizando espectros das sementes de algodão. Imagens multiespectrais foram eficientes em fornecer dados para a classificação de sementes de algodão quanto ao potencial fisiológico e estágios de deterioração. Os modelos de predição utilizando os algoritmos *Neural Network*, *Support Vector Machine* e *Gradient Boosting* demonstraram ser os mais eficazes para a classificação dos espectros das sementes de algodão quanto ao potencial fisiológico.

Palavras chaves: *Gossypium* L.; espectroscopia aplicada; análise de imagens; inteligência artificial; modelagem preditiva

ABSTRACT

Obtaining high-quality seeds is a crucial challenge in cotton production. Traditional methods, such as germination and vigor tests, are laborious and largely subjective, highlighting the need for complementary tools that are fast and less costly. Multispectral imaging emerges as a promising alternative, providing spatial and spectral information about morphological attributes and chemical composition. This study evaluated the use of multispectral imaging to classify the quality of cotton seeds. Seeds from a batch from the Catuti-MG region were subjected to accelerated aging in a BOD chamber at 41 °C for 24, 48, 72, and 96 hours, in addition to a control group without aging. A low-cost spectral imaging prototype, Croma Seeds, was developed. The images, captured at specific wavelengths (395–940 nm), were processed using ImageJ to obtain reflectance data. Predictive models were developed using Orange Data Mining and supervised classification algorithms such as Tree, Random Forest, Gradient Boosting, Support Vector Machine (SVM), and Neural Network. It was observed that, in traditional germination and vigor tests, there was a statistical difference between the control treatment and seeds subjected to accelerated aging, with a sharp decline in seed physiological quality, particularly after 96 hours, and an increase in electrical conductivity. Principal Component Analysis (PCA) explained 88.62% of the variance, differentiating the treatments. The Neural Network and SVM models demonstrated effectiveness when compared to other models, with 100% accuracy and a 99% *kappa* in the test using spectra from cotton seeds. Multispectral imaging proved to be effective in providing data for classifying cotton seeds in terms of physiological potential and deterioration stages. The predictive models using Neural Network, Support Vector Machine, and Gradient Boosting algorithms were shown to be the most effective for classifying cotton seed spectra based on physiological potential.

Keywords: *Gossypium L.*; applied spectroscopy; image analysis; artificial intelligence; predictive modeling

INTRODUÇÃO

O algodão é amplamente utilizado na produção de fibras têxteis, óleos e subprodutos industriais, tornando sua cadeia produtiva essencial para muitos países (Zhu *et al.*, 2019). A obtenção de sementes de algodão de alta qualidade, porém, enfrenta desafios significativos. Além de resistência a estresses bióticos, como pragas e doenças, e abióticos, como seca e calor, as sementes devem apresentar uniformidade genética e altas taxas de germinação para atender às demandas dos produtores (Queiroga *et al.*, 2009). A complexidade desse processo é agravada pela sensibilidade das sementes ao armazenamento e pelos problemas relacionados à longevidade, que podem afetar sua viabilidade e vigor. Isso destaca a importância de métodos eficazes de avaliação e manejo de qualidade (Mattioni *et al.*, 2012).

Atualmente, a avaliação da qualidade das sementes é realizada predominantemente por meio de testes de germinação e vigor, que demandam trabalho intensivo e a expertise de analisadores de sementes qualificados (Ambrose *et al.*, 2016). Embora esses métodos tradicionais sejam amplamente utilizados, eles apresentam limitações em termos de eficiência e objetividade. Nesse contexto, a adoção de tecnologias inovadoras tem revolucionado o setor, não como substituto, mas como complemento aos testes tradicionais. A automação das análises laboratoriais tem facilitado a tomada de decisões estratégicas, como a alocação de lotes de sementes com base em sua qualidade (Bianchini *et al.*, 2021).

Entre essas tecnologias, a imagem multiespectral destaca-se como uma ferramenta promissora para a análise de qualidade de sementes. Essa técnica permite a avaliação de diversos atributos de qualidade, como morfocolorimetria e composição química (Carreiro Soares *et al.*, 2016). O método consiste em expor as sementes a diferentes comprimentos de onda de luz sequencialmente, combinando essa exposição com sistemas computacionais que identificam propriedades físico-químicas das amostras. No setor de grãos, essa tecnologia já é utilizada para detectar danos causados por insetos, infecções fúngicas, sementes quebradas e outras alterações que afetam a qualidade (Conceição *et al.*, 2024). Entretanto, no setor sementeiro, a análise de qualidade fisiológica ainda representa um desafio para essa tecnologia (Mesquita *et al.*, 2022).

A aplicação de imagens multiespectrais na análise de sementes tem avançado especialmente com o uso de algoritmos de *machine learning* para interpretação dos dados espectrais. Modelos como *Artificial Neural Network* (ANN), *Support Vector Machine* (SVM) e *Random Forest* têm demonstrado grande potencial para identificar padrões complexos,

associando características físico-químicas das sementes a indicadores de qualidade (Barboza da Silva *et al.*, 2021; Soares *et al.*, 2024). Essa integração permite o treinamento de modelos capazes de diferenciar automaticamente sementes saudáveis de danificadas, prever taxas de germinação e até identificar a presença de contaminantes ou doenças (Boelt *et al.*, 2018; França-Silva *et al.*, 2020). Dessa forma, a análise torna-se mais precisa, rápida e objetiva, reduzindo a subjetividade e a dependência de análises manuais, o que é essencial para a modernização do setor de sementes.

Apesar dos avanços, a implementação dessa tecnologia ainda enfrenta barreiras. Os equipamentos de análise multiespectral são caros e muitas universidades com poucos recursos têm dificuldade em acessá-los para fins de pesquisa, o que limita o avanço tecnológico na área.

Nesse contexto, o presente trabalho teve como objetivo i) desenvolver um protótipo de baixo custo para aquisição de imagens multiespectrais; ii) investigar a viabilidade do uso das imagens multiespectrais na classificação da qualidade de sementes de algodão, utilizando algoritmos de inteligência artificial como ferramenta para otimizar o controle de qualidade. Essa abordagem pode contribuir para o desenvolvimento de métodos mais eficientes e acessíveis para a avaliação de sementes, fortalecendo a cadeia produtiva do algodão.

MATERIAL E MÉTODOS

A pesquisa foi conduzida no Laboratório de Pesquisa de Sementes, do Departamento de Agronomia da Universidade Federal de Viçosa (DAA-UFV). Foram utilizadas sementes oriundas da região de Catuti-MG, já deslintadas. Inicialmente amostras foram separadas para formação dos tratamentos com diferentes estágios de deterioração. Para tanto, procedeu-se com o envelhecimento acelerado de amostras no qual, as sementes foram dispostas em caixas do tipo gerbox com tela, contendo 40 ml de água destilada, as caixas foram vedadas de modo a obter 100% UR em seu interior e mantidas em câmara do tipo BOD a 41° C por períodos de 24, 48, 72 e 96 horas tendo como o controle sementes não envelhecidas (Marcos-Filho, 2020).

Determinação do grau de umidade

Foi obtido pelo método de estufa a 105 °C, por um período de 24 horas, segundo as Regras para Análise de Sementes (Brasil, 2009). Utilizou-se quatro repetições de 25 sementes para cada tratamento e os resultados foram expressos em porcentagem média com base no peso úmido.

Teste de germinação

Realizado conforme as Regras para Análise de Sementes (Brasil, 2009), foram utilizadas quatro repetições de 50 sementes em rolo de papel, umedecido com a quantidade de água equivalente a 2,5 vezes o peso do papel seco. Após a semeadura, confeccionou-se rolos que foram mantidos em germinador à temperatura de 25°C. As avaliações foram realizadas aos 4 e 12 dias após a semeadura e os resultados expressos em porcentagem de plântulas normais obtidas no decimo segundo dia.

Primeira contagem de germinação

Foi realizada juntamente com o teste de germinação, calculando-se a porcentagem de plântulas normais obtidas no quarto dia após a semeadura (Brasil, 2009).

Emergência de plântulas

Foram utilizadas quatro repetições de 25 sementes e semeadas a 2,0 cm de profundidade em bandejas plásticas contendo mistura de solo e areia na proporção de 1:2, umedecido até atingir 60% da capacidade de retenção (Brasil, 2009). O teste foi conduzido em câmara de crescimento vegetal e foram consideradas emergidas, as plântulas que se apresentaram com 2

em acima da superfície do substrato. A porcentagem de emergência foi calculada computando-se o total de plântulas emergidas após a completa estabilização do estande.

Índice de velocidade de emergência (IVE)

Calculado a partir dos valores diários obtidos do número de plantas emergidas, do primeiro ao último dia de duração do teste de emergência, conforme Maguire (1962).

Condutividade elétrica

Quatro repetições de 50 sementes foram pesadas em balança de precisão 0,001 g, e em seguida, as sementes de cada repetição foram colocadas em copos plásticos contendo volume de 75 ml de água destilada. Os copos foram mantidos em incubadora BOD regulada a 25 °C por períodos de 16 horas. Após cada período, os copos foram retirados da incubadora, os exsudatos foram agitados com um bastão de vidro e a condutividade elétrica das soluções de embebição foi determinada em condutivímetro (Digimed modelo CD 21) devidamente calibrado. Os resultados obtidos foram divididos pelo peso das sementes das respectivas repetições e o resultado expresso em $\mu\text{S}/\text{cm}/\text{g}$ de sementes (Vieira; Marcos-Filho, 2020).

Comprimento de plântulas

Para avaliação do comprimento de plântulas as sementes foram submetidas aos procedimentos adaptados descritos por Nakagawa (1999), após cinco dias de instalação do teste, foi mensurado o comprimento de radícula e do hipocótilo nas plântulas normais, para tanto utilizou-se imagens das plântulas que foram capturadas e mensuradas através do programa ImageJ, com os valores expressos em centímetros.

Delineamento experimental e análise estatística

Os testes foram conduzidos em delineamento inteiramente casualizado (DIC) com cinco tratamentos (Controle, 24, 48, 72 e 96 horas) e quatro repetições. Os dados foram testados quanto à distribuição normal dos erros pelo teste de Shapiro-Wilk e à homogeneidade de variâncias pelo teste de Bartlett. Os dados referentes aos testes para a caracterização da qualidade foram submetidos à análise de variância e as médias obtidas foram comparadas pelo teste de Tukey a 5% de probabilidade. Para as análises estatísticas foi utilizado o software estatístico R 12.1 (R CORE TEAM, 2024). e o pacote Exp.Des.pt (Ferreira *et al.* 2014).

Imagens Multiespectrais

Para Análise multiespectral foi desenvolvido um protótipo de captura de imagens multiespectrais nomeado de Cromas Seed, que forneceu imagens UV-VIS-NIR. O protótipo funciona utilizando o princípio de iluminação de banda estreita, composto com 8 tipos de diodo emissor de luz (LED) nas faixas de 395, 460, 520, 585, 620, 740, 850 e 940 nm e uma câmera do tipo NOIR-Monocromática de resolução 3280 x 2464 pixels, controlados por um microcomputador (Raspiberry pi 4 - modelo B), sistema operacional Raspbian (PI OS – Bookworm) e linguagem Python (Figura 1). Para avaliação da intensidade e confirmação das bandas espectrais emitidas pelo conjunto de LEDs, foi realizada quatro leituras para cada LED espectral em espectroradiômetro L1-180 que utiliza sensor de imagem linear do tipo CMOS que captura intensidade luminosa em incrementos de 1 nm (largura de banda de 12 nm) entre 380 – 1000 nm (Figura 2), disponível no Laboratório de Caracterização Óptica da Universidade Federal de Viçosa.

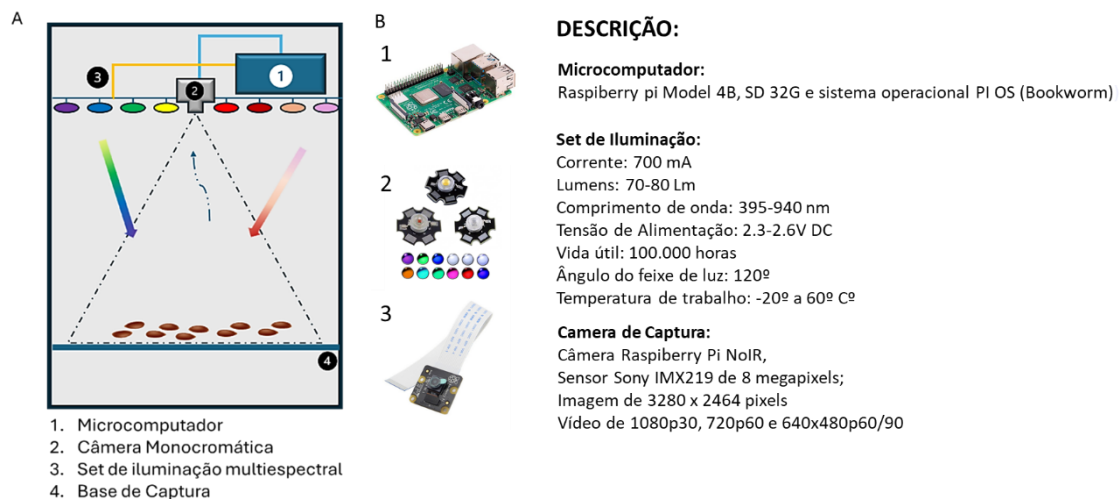


Figura 1. Disposição dos componentes no equipamento (A), imagem dos componentes utilizados (B) e descrição das características dos principais periféricos do protótipo Cromas Seeds.

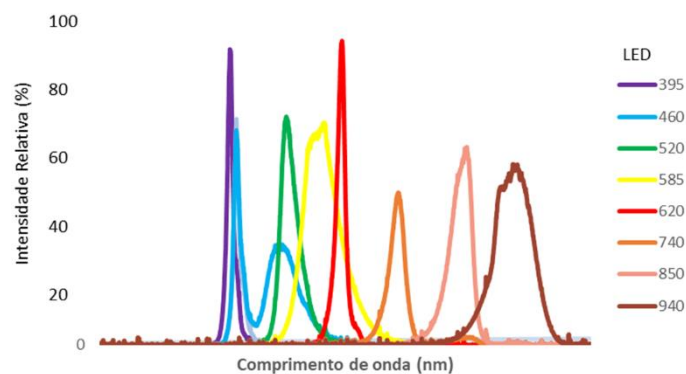


Figura 2. Porcentagem de intensidade relativa dos conjuntos de LEDs utilizados na iluminação do Cromas Seeds, em função das bandas espectrais, destacando os picos de maior intensidade nas faixas correspondentes.

As amostras de sementes previamente descritas foram dispostas dentro do equipamento em fundo azul confeccionado de etileno vinil acetato (EVA) a uma distância do conjunto de LEDs e câmera de 40 cm. Para o processamento das imagens foi utilizado o programa ImageJ (Schindelin *et al.* 2012). As imagens foram importadas em formato sequência salvas no formato *Tiff*, onde procedeu-se com a abertura das imagens, empilhamento das bandas, conversão em escala de cinza (8 bits), seleção da área de interesse pelo método *Threshold* e obtenção dos valores médios de cinza dos pixels de cada semente para cada banda espectral (Figura 3).

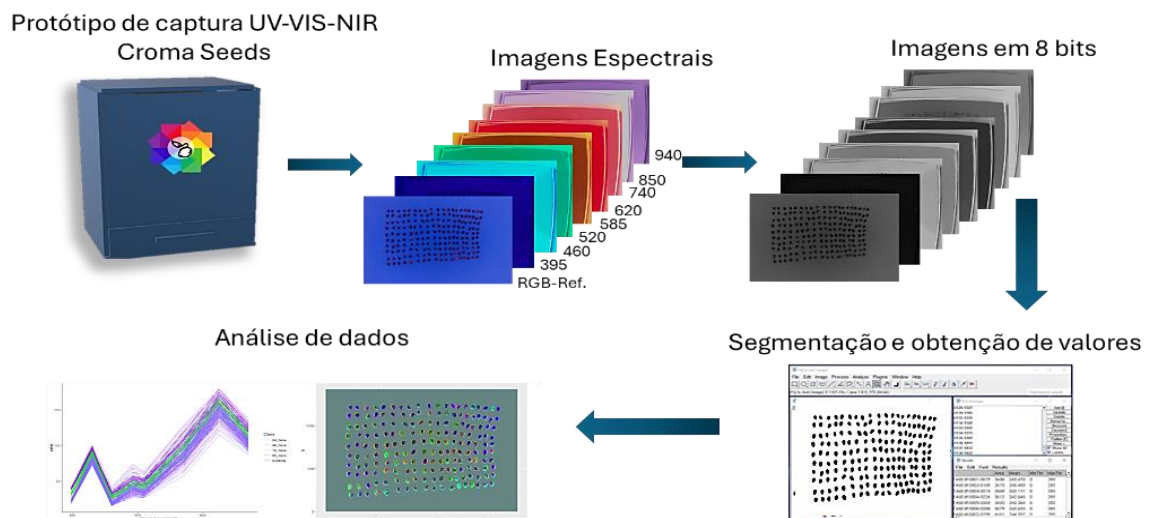


Figura 3. Etapas de Captura, empilhamento e processamento das imagens utilizando o ImageJ para segmentação da área de interesse e obtenção dos valores espectrais.

Para obtenção dos valores de reflectância foram utilizados como alvo de reflectância o Sulfato de Bário P.A (BaSO_4) e Carvão ativado em pó P.A ambos da marca Dinâmica® (Figura 4A). Estes materiais foram escolhidos dado a facilidade de acesso, durabilidade, não apresentar toxicidade e também pela capacidade de reflectância conhecida do sulfato de bário que é aproximadamente 98% na faixa de 350 a 1000 nm e carvão ativado apresenta de 3% de reflectância na mesma faixa o que compreende a região do ultravioleta, visível e parte do infravermelho próximo (Sanches *et al.*, 2009; Knighton; Bugbee, 2005). Os valores calculados para o branco e preto estão dispostos na Tabela 1.

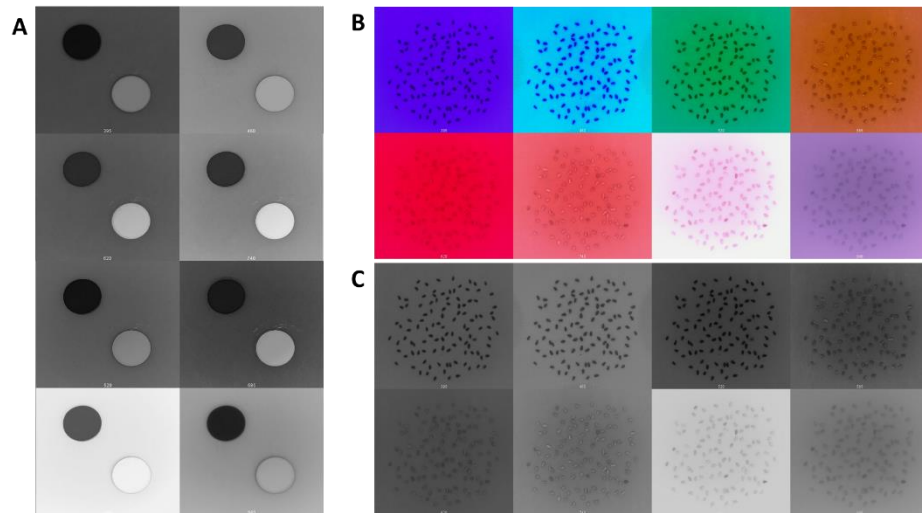


Figura 4. Imagens obtidas pelo Croma Seeds, Alvos de reflectância (A), imagens das sementes em cada banda espectral (B), imagens após conversão para 8 bits.

Tabela 1. Valores de pixels extraídos das amostras de referência para cálculo de reflectância utilizando sulfato de bário como branco e carvão ativado como preto para cada uma das bandas espectrais.

Banda Espectral	Branco	Preto
395 nm	117	14.5
460 nm	162	14.7
520 nm	128	18.1
585 nm	149	15.3
620 nm	178	19.1
740 nm	220	19.1
850 nm	240	28.9
940 nm	240	30.5
Média	179.3	20
Desv. Padrão	48.9	6.3
C.V %	27.3	31.3
Reflectância %	98	3

Após a obtenção destes valores, de cada semente individualmente, foi realizado o cálculo para transformação de números digitais 8 bits (0-255) que comumente são conhecidos como valores de pixels, em reflectância %. Essa conversão é possível, pois valores digitais registrados pelos sensores ópticos correspondem à intensidade da luz refletida pelos objetos em relação à luz incidente, assumindo que essa relação é linear, assim, ao aplicar uma fórmula de calibração, considerando os padrões de referência, é possível converter esses números digitais em valores percentuais de reflectância, representando a proporção de luz refletida por cada semente em relação à luz que a ilumina. Isso permite a análise precisa das propriedades ópticas e físicas das sementes e é calculado da usando a seguinte equação de acordo com Pereira *et al.*, (2015):

$$\text{Reflectância \%} = \frac{(\text{Valor do pixel} - \text{Valor de pixel preto})}{(\text{Valor de pixel branco} - \text{Valor de Pixel preto})} * \text{Reflectância do Branco}$$

Modelos de Predição

O conjunto de dados inicial, foi composto por 500 sementes por tratamento (Controle, 24, 48, 72 e 96 horas de envelhecimento acelerado) perfazendo um total de 2500 sementes, posteriormente esse baco de dados foi subdividido em 70% e 30% para treinamento e teste, respectivamente. Os dados após a coleta das curvas espectrais a partir das amostras de sementes, as mesmas foram plotadas de maneira individual para a verificação da qualidade com relação à presença de ruídos e estes eliminados, ao final o conjunto de dados foi composto por 2344. Para prevenir o desbalanceamento dos tratamentos no treinamento foi utilizada a técnica *SMOTE* que gera amostras sintéticas para as classes minoritárias que neste caso foram os tratamentos 72 e 96 horas. Para realizar a classificação de dados supervisionada, os algoritmos utilizados foram: *Tree*, *Random Forest*, *Support Vector Machine (SVM)*, *Gradient Boosting* e *Neural Network* (Figura 5). Todos disponíveis no programa Orange data mining (Demsar *et al.*, 2013).

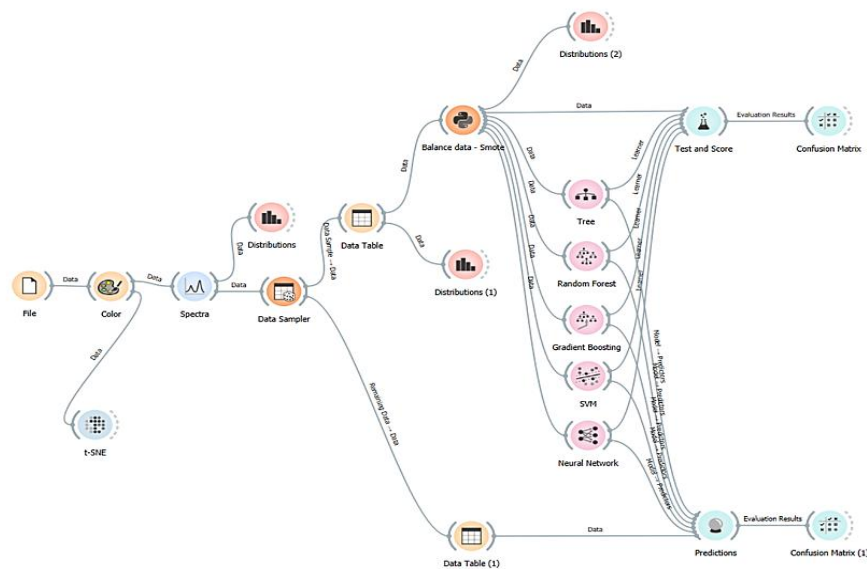


Figura 5. Fluxograma criado no Orange Data Mining para classificação supervisionada. Os dados foram divididos em 70% para treinamento e 30% para teste, seguidos pela aplicação de técnicas de amostragem e balanceamento e implementação de cinco algoritmos de aprendizado de máquina: Tree, Random Forest, Gradient Boosting, SVM e Neural Network para análise dos dados. A avaliação do desempenho foi realizada por meio de matrizes de confusão e métricas de assertividade, garantindo uma comparação detalhada entre os modelos utilizados.

Para verificar a precisão dos algoritmos foram utilizadas as seguintes métricas de avaliação: Área sob a curva ROC. Acurácia (*Correctly Classified Accuracy*), F1 Score: Média harmônica entre precisão e recall, Precisão: Proporção de positivos verdadeiros entre as previsões positivas, Recall: Proporção de positivos verdadeiros detectados, Coeficiente de Correlação de Matthews, Especificidade: Proporção de negativos verdadeiros, LogLoss: Perda logarítmica (avalia a confiança das previsões probabilísticas) e Coeficiente de *Kappa* de Cohen: concordância ajustada ao acaso, também foram avaliadas as matrizes de confusão com a finalidade de observar o percentual de assertividade de cada modelo, tanto no treinamento como no teste, de acordo com Lever *et al.*, (2016), para o treinamento e validação dos modelos de predição foi utilizado o programa Orange Data Mining (Demsar *et al.*, 2013).

RESULTADOS E DISCUSSÃO

A alta umidade e temperatura apresentaram impacto negativo significativo na qualidade fisiológica de sementes de algodão, conforme observado na Tabela 2. À medida que o período de envelhecimento acelerado aumentou, os valores de germinação (G) reduziram-se gradativamente de 92% (controle) para 65% (96 horas). De forma similar, os valores de emergência (E) decaíram de 88% (controle) para 46% (96 horas), indicando uma redução expressiva na viabilidade e no vigor das sementes (Tabela 2).

Tabela 2. Umidade (U), primeira contagem de germinação (PC), germinação (G), emergência (E), índice de velocidade de emergência (IVE), condutividade elétrica (CE), comprimento da parte aérea (P.A), comprimento da parte radicular (P.R), comprimento total (C.Total) de sementes de algodão submetidas aos períodos de envelhecimento acelerado por 0 (controle), 24, 48, 72 e 96 horas. Médias seguidas pela mesma letra na coluna, não diferem entre si pelo teste de Tukey a 5% de probabilidade.

Tratamentos	U	PC	G	E	IVE	CE	Comp.PA	Comp.RA	Comp.Total
	-----%	-----%	-----%	-----%	índice	$\mu\text{S cm}^{-1} \text{g}^{-1}$	-----cm-----	-----cm-----	-----cm-----
Controle	11 a	95 a	97 a	92 a	2.17 a	70.68 c	9.88 a	17.08 a	26.96 a
24 horas	10.5 a	87 b	93 a	89 a	2.12 a	80.5 bc	11.75 a	16.30 a	28.05 a
48 horas	11 a	76 c	77 b	83 b	1.44 a	91.45 ab	4.82 b	9.67 ab	14.49 b
72 horas	10.75 a	66 d	67 c	67 c	1.16 b	92.3 ab	3.26 b	5.12 b	8.39 b
96 horas	10.5 a	65 d	65 c	46 c	1.14 c	108.91 ab	3.04 b	5.01 b	8.06 b
C.V (%)	7.1	2.74	2.66	7.19	7.76	10.48	25.57	41.21	28.34

O índice de velocidade de emergência (IVE) seguiu a mesma tendência de declínio, passando de 2.17 (controle) para 1.14 (96 horas), reforçando a diminuição no desempenho germinativo. Em contraste, a variável condutividade elétrica (CE) apresentou um aumento progressivo, de 70.68 $\mu\text{S/cm/g}$ (controle) para 108.91 $\mu\text{S/cm/g}$ (96 horas), sinalizando maior lixiviação de íons causada pela deterioração das membranas celulares (Tabela 2).

Com relação ao crescimento das plântulas, o comprimento da parte aérea (P.A) reduziu de 9.88 cm (controle) para 3.04 cm (96 horas), enquanto o comprimento da parte radicular (P.R) diminuiu de 17.08 cm (controle) para 5.01 cm (96 horas). O comprimento total (C.Total), por consequência, variou de 26.96 cm (controle) para 8.06 cm (96 horas), indicando que os efeitos do envelhecimento acelerado impactam tanto o desenvolvimento inicial da plântula quanto o potencial de estabelecimento da planta (Tabela 2).

Haja vista que a composição das reservas da semente de algodão é basicamente lipoprotéica, contendo aproximadamente 39% de proteínas, 33% de lipídeos e 15% de carboidratos as sementes tornam-se susceptíveis a peroxidação lipídica e a desnaturação de

proteínas, processos estes que tem radicais livres como protagonistas de ações metabólicas, que apesar de fazerem parte como sinalizadores em algumas reações essenciais, sobretudo quando o metabolismo está mais intenso, o desbalanceamento destes radicais causam uma serie de danos irreversíveis aos tecidos embrionários, sobretudo as moléculas de lipídeos que estão sujeitas a constantes ataques por espécies reativas de oxigênio o que altera sua estrutura e função (Marcos-Filho, 2015). Outra característica está relacionada a altas temperaturas e a desnaturação de proteínas (perca da conformidade e função) e também desativação de enzimas essenciais nos processos metabólicos, fatos estes descritos por Bewley *et al.*, (2013) que considera que um dos primeiros eventos bioquímicos verificados quando as sementes são expostas a condições desfavoráveis é o dano à síntese de proteínas durante as primeiras horas de embebição, sementes não viáveis ou com baixo potencial fisiológico não conseguem sintetizar quantidade suficiente de proteínas.

Desta forma, alterações nos processos metabólicos irão causar uma serie de reações que poderão impedir a semente de expressar sua máxima qualidade. Esses resultados evidenciam que o envelhecimento acelerado, causado por condições de alta umidade e temperatura, compromete significativamente causando deterioração e queda da qualidade fisiológica das sementes.

Essa relação de queda de qualidade em função do período de envelhecimento acelerado fica mais evidente quando observa-se a figura 6, que por meio da aplicação de análise de componentes principais, notou-se significativa discriminação dos tratamentos em função do tempo de envelhecimento. Os dois primeiros componentes explicaram 88.62% da variação observada. Os vetores relacionados as variáveis de qualidade fisiológica, estão diretamente correlacionadas de forma positiva ao tratamento controle (maior viabilidade e vigor), enquanto há distribuição dos demais tratamentos ao logo da PC 1 para a esquerda, onde se encontra o vetor de condutividade elétrica que está correlacionado com os tratamentos de menor qualidade que foram expostos por mais tempo ao envelhecimento acelerado, evidenciando que houve estratificação dos tratamentos quando ao nível de viabilidade e vigor (Figura 6).

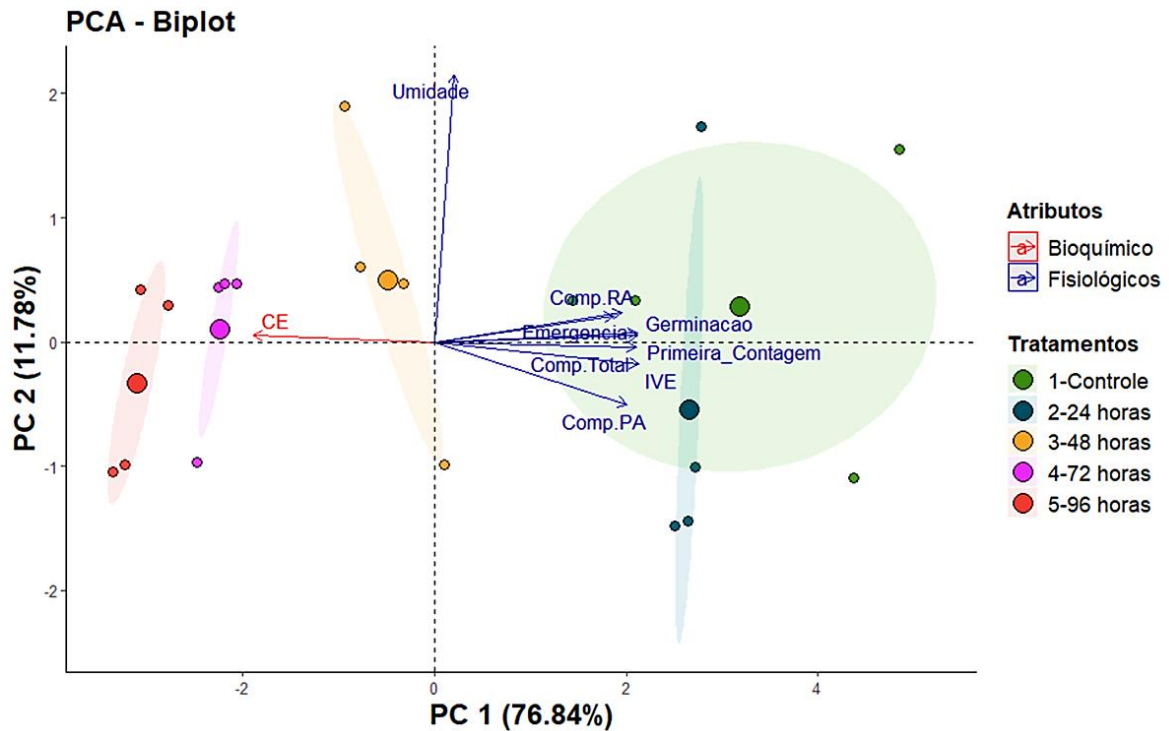


Figura 6. Biplot da Análise de Componentes Principais das variáveis de qualidade fisiológica de sementes de algodão submetidas a diferentes tempos de envelhecimento acelerado.

Nesta pesquisa optou-se por esse método de deterioração com o objetivo de formar sementes com diferentes níveis de qualidade, pois no setor sementeiro, é de extrema importância a identificação de lotes de maior qualidade, essa etapa de triagem acontece ainda nos primeiros momentos quando é realizada a colheita, muito embora a qualidade de sementes se faça no campo, com bom manejo de solo, água, nutrientes, pragas e doenças, ainda assim é preciso selecionar materiais que ofereçam maior qualidade ao mercado, etapa essa realizada em laboratório com a realização dos testes de germinação e vigor como os descritos anteriormente. (Mesquita *et al.*, 2022; Queiroga *et al.*, 2009).

A implementação de ferramentas e técnicas que ofereçam informações complementares de forma ágil, objetiva e economicamente acessível é uma necessidade crescente no mercado de sementes, embora ainda enfrente desafios significativos (Mesquita *et al.* 2022; Xia *et al.*, 2019). Nesta pesquisa, buscou-se introduzir um equipamento multispectral de baixo custo, integrado a tecnologias avançadas como algoritmos de aprendizado de máquina, com o objetivo de gerar dados inovadores que possam ser correlacionados aos testes tradicionais. Além disso, o estudo também visa avaliar a capacidade do equipamento em identificar sementes com diferentes níveis de qualidade fisiológica.

Outro aspecto relevante é a proposta de democratizar o acesso a esse tipo de tecnologia, considerando que equipamentos desse porte ainda exigem alto investimento no setor de sementes. O avanço de pesquisas, associado ao desenvolvimento de novos métodos analíticos, pode otimizar processos e melhorar a tomada de decisão, principalmente no que se refere à classificação e destinação de lotes com base na qualidade.

Com base nesse contexto, os dados espectrais obtidos desempenham um papel fundamental, permitindo explorar novas possibilidades analíticas e consolidando a eficiência do equipamento na distinção entre diferentes níveis de qualidade fisiológica, é possível observar que não houve alterações na linha de base, dado que as imagens apresentam menos ruídos do que sinais elétricos quando comparados (Figura 7).

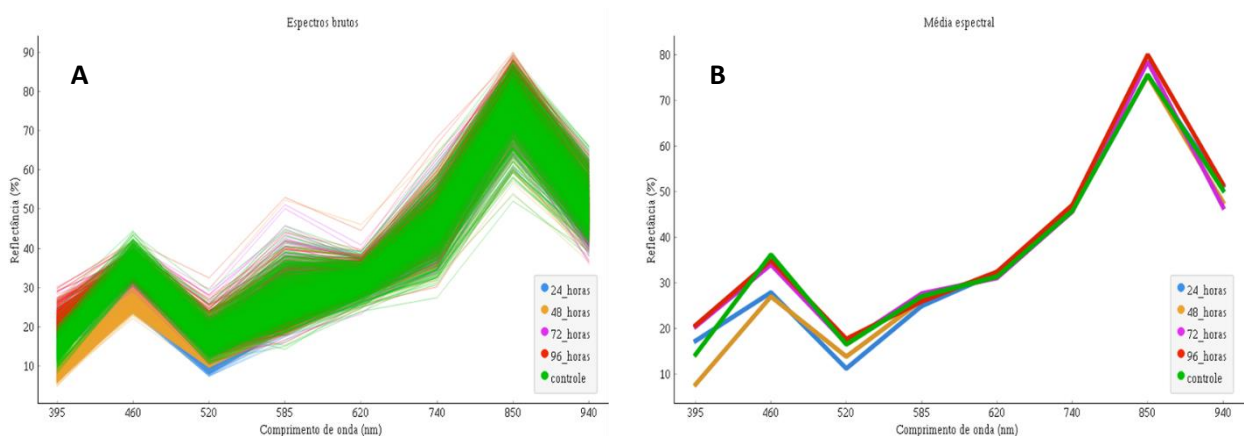


Figura 7. Espectros de sementes de algodão obtidos com o Cromas Seeds, totalizando 2500 observações em cinco períodos de envelhecimento acelerado (A), com a média espectral de cada tratamento (B).

Os dados espectrais correspondentes a oito bandas (395, 460, 520, 585, 620, 740, 850 e 940 nm) dos cinco tratamentos de envelhecimento foram utilizados para construção de cinco modelos de aprendizado de máquina, utilizando 70% dos dados para a etapa de treinamento (Tabela 3). Os modelos de *Neural Network* e *SVM*, respectivamente apresentaram índices de assertividade elevados em todas as métricas avaliadas (99%) com menores perdas logarítmicas, 0.048 e 0.032 respectivamente. O *Gradient Boosting* e *Random Forest* também apresentaram altos valores de assertividade em todas as métricas, porém menores que as anteriores, sendo o *Gradient Boosting* o que apresentou menor valor de perda logarítmica entre todos os modelos 0.008. A árvore de decisão foi a que teve o menor desempenho quando comparado com as demais, mas mesmo assim, ainda com valores elevados de assertividade >90%, porém apresentou o maior valor de perda logarítmica 1.606. O LogLoss mede o grau de incerteza nas previsões de um modelo, penalizando respostas que estão mais distantes das classes

verdadeiras. Diferentemente de apenas verificar se uma previsão está correta, o LogLoss considera a confiança da previsão (as probabilidades atribuídas). Isso significa que quanto mais próximo o modelo estiver da probabilidade correta menor será o LogLoss (Tabela 3).

Tabela 3. Métricas da etapa de treinamento dos modelos de aprendizado de máquina: *Tree*, *Random Forest*, *Gradient Boosting*, *Support Vector Machine* (SVM) e *Neural Network* na classificação de sementes de algodão submetidas a cinco períodos de envelhecimento acelerado. Curva ROC (AUC), Acurácia (CA), F1-Score, Precisão (Prec), Recall, Coeficiente de correlação de Mathews (MCC), Perda logarítmica (LogLoss) e Coeficiente de conformidade *Kappa*.

Modelo	Treinamento								
	n =1515								
	AUC	CA	F1	Prec	Recall	MCC	Spec	LogLoss	<i>Kappa</i>
Tree	0.959	0.929	0.929	0.929	0.929	0.912	0.982	1.606	0.939
Random Forest	0.996	0.962	0.962	0.962	0.962	0.953	0.991	0.242	0.974
Gradient Boosting	0.998	0.980	0.980	0.980	0.980	0.975	0.995	0.008	0.983
SVM	1.000	0.994	0.994	0.994	0.994	0.993	0.999	0.032	0.993
Neural Network	1.000	0.991	0.991	0.991	0.991	0.988	0.998	0.048	0.989

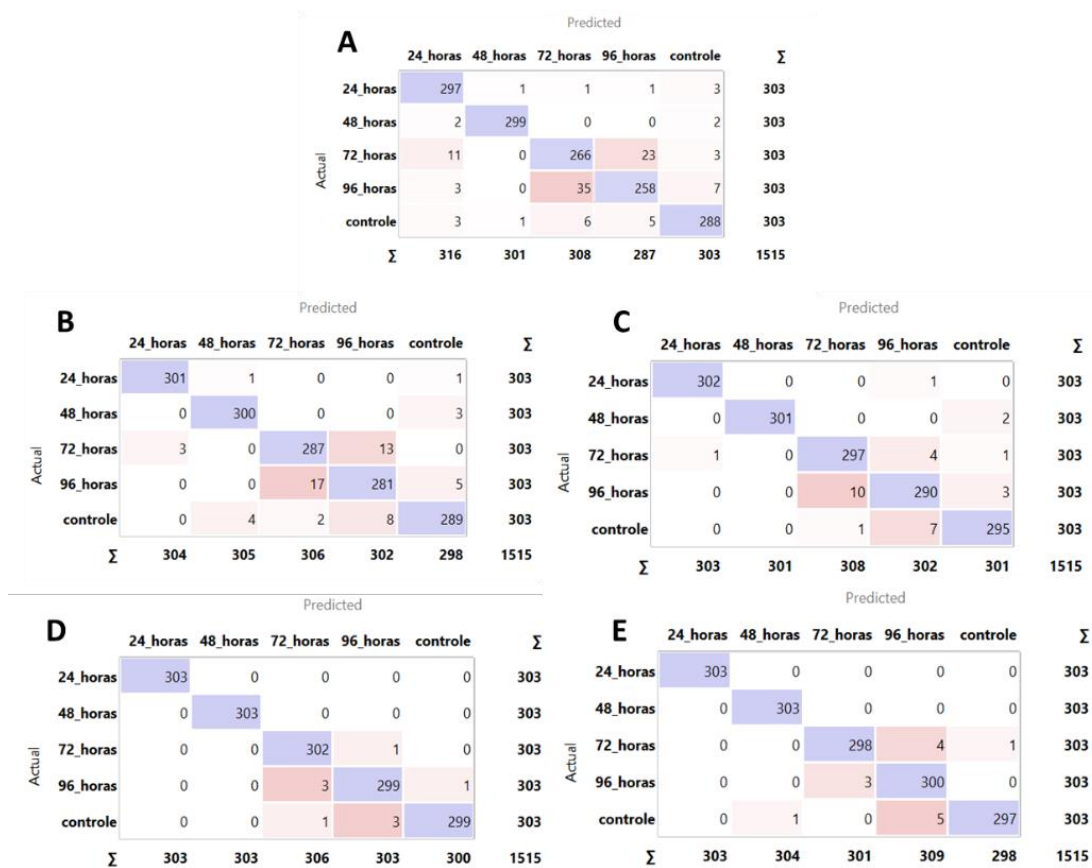


Figura 4. Resultados das Matrizes de confusão da etapa de treinamento dos modelos (A) Tree, (B) Random Forest, (C) Gradient Boosting, (D) SVM e (E) Neural Network para classificação de sementes de algodão submetidas a tempos de envelhecimento acelerado.

Observa-se que na figura 8, as matrizes de confusão corroboram com as métricas observadas, sendo bastante útil para avaliar o desempenho. O modelo Tree, foi o que apresentou mais erros de classificação principalmente entre os tratamentos de 72 e 96 horas. Vale ressaltar que como atribuiu-se o tempo de envelhecimento a todas as sementes pertencentes aquele grupo, evidentemente algumas sementes possuem qualidade e outras não, esse fato foi reportado por Barbosa da Silva *et al.*, (2021) e também por Medeiros (2023) quando descrevem sobre a forma de rotular a classe de sementes de forma geral e não individualmente. Mesmo diante desse fato, os outros modelos foram capazes de classificar com maior precisão do que o modelo de Tree.

Ao observar a consistência dos modelos na fase de teste com dados que foram externos ao modelo, foi possível verificar que as métricas permaneceram altas para Neural Network, SVM, Gradient Boosting >99% e perda logarítmica baixa <0.09. O modelo de Tree ainda conseguiu manter uma boa concordância entre treinamento e teste, porém as métricas foram as menores quando comparadas as demais (Tabela 4).

Tabela 4. Métricas da etapa de teste dos modelos de aprendizado de máquina: *Tree*, *Random Forest*, *Gradient Boosting*, *Support Vector Machine* (SVM) e *Neural Network* na classificação de sementes de algodão submetidas a cinco períodos de envelhecimento acelerado. Curva ROC (AUC), Acurácia (CA), F1-Score, Precisão (Prec), Recall, Coeficiente de correlação de Mathews (MCC), Perda logarítmica (LogLoss) e Coeficiente de conformidade *Kappa*.

Modelo	Teste								
	n =829								
	AUC	CA	F1	Prec	Recall	MCC	Spec	LogLoss	<i>Kappa</i>
Tree	0.958	0.936	0.936	0.936	0.936	0.919	0.984	1.711	0.919
Random Forest	0.996	0.955	0.956	0.956	0.955	0.944	0.989	0.217	0.936
Gradient Boosting	0.998	0.978	0.978	0.978	0.978	0.973	0.995	0.097	0.973
SVM	1.000	0.998	0.998	0.998	0.998	0.997	0.999	0.019	0.997
Neural Network	1.000	0.996	0.996	0.996	0.996	0.995	0.999	0.042	0.995

Os erros de classificação observados na matriz de confusão da etapa de teste, mostram que apesar de serem pequenos quando comparados ao tamanho do banco de dados, os modelos Tree, Random Forest e Gradient Boosting foram os que tiveram menor desempenho na assertividade. Já para SVM e Neural Network, os erros foram menores, apenas 2 e 3 sementes em um universo de 829 sementes, respectivamente (Figura 9).

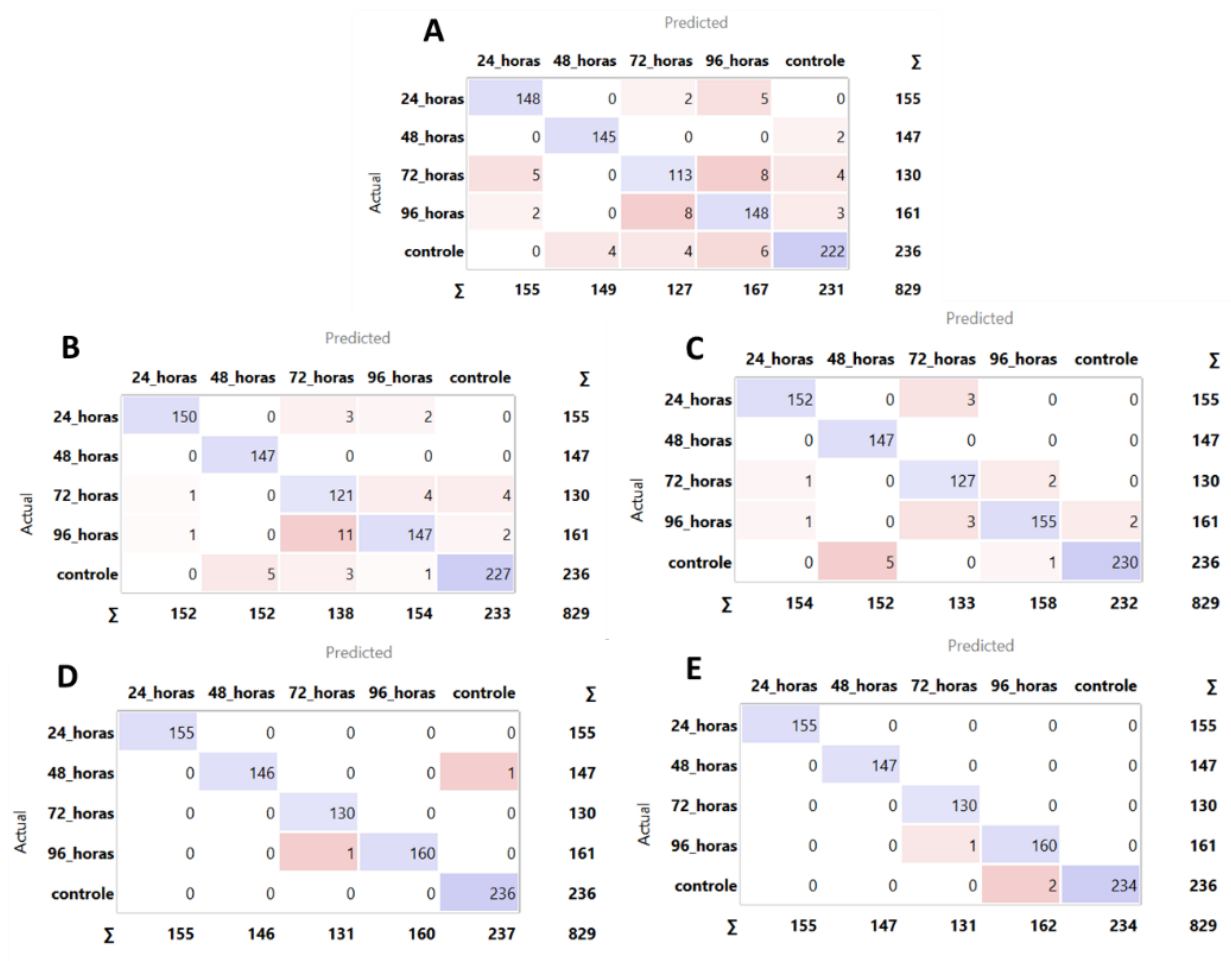


Figura 5. Resultados das Matrizes de confusão da etapa de teste dos modelos (A) Tree, (B) Randon Forest, (C) Gradient Boosting, (D) SVM e (E) Neural Network para classificação de sementes de algodão submetidas a tempos de envelhecimento acelerado.

Resultados semelhantes foram obtidos com outras espécies de sementes cultivadas como por exemplo: soja em um sistema de captura de baixo custo (SOARES *et al.*, 2022), pinhão-manso (BIANCHINI *et al.*, 2021), e algodão através de imagens hiperespectrais (SOARES *et al.*, 2016) e na discriminação quanto ao estágio de maturação em sementes de soja, por meio de autofluorescência e imagens multiespectrais (BARBOSA *et al.*, 2022). Esses estudos mostram que a combinação de algoritmos de classificação com técnicas de processamento de imagem proporciona resultados consistentes na identificação de sementes danificadas ou doentes e na classificação de sua qualidade, contribuindo para a adoção de práticas agrícolas mais sustentáveis e criando oportunidades para a automação do setor de sementes.

CONCLUSÃO

Imagens multiespectrais foram eficientes em fornecer dados para a classificação de sementes de algodão quanto ao potencial fisiológico e estágios de deterioração.

Os modelos de predição utilizando os algoritmos *Neural Network*, *Support Vector Machine* e *Gradient Boosting* foram eficazes para a classificação dos espectros das sementes de algodão quanto ao potencial fisiológico.

REFERÊNCIAS

- ALVARENGA, Renata Oliveira; MARCOS-FILHO, Júlio. Vigor evaluation of stored cotton seeds, including the Seed Vigor Imaging System (SVIS®). *Journal of Seed Science*, [s. l.], v. 36, n. 2, p. 222–230, 2014. Disponível em: <https://doi.org/10.1590/2317-1545v32n2944>
- ALVES, Lucas Matheus Carvalho Ribeiro *et al.* Influência do tempo e condições de conservação na qualidade das sementes de gergelim do Banco Ativo de Germoplasma. *Contribuciones a Las Ciencias Sociales*, [s. l.], v. 17, n. 12, p. e12489, 2024. Disponível em: <https://doi.org/10.55905/revconv.17n.12-092>.
- AMBROSE, Ashabahebwa *et al.* Comparative nondestructive measurement of corn seed viability using Fourier transform near-infrared (FT-NIR) and Raman spectroscopy. *Sensors and Actuators B Chemical*, [s. l.], v. 224, p. 500–506, 2015. Disponível em: <https://doi.org/10.1016/j.snb.2015.10.082>.
- BARBOSA 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. Disponível em: <https://doi.org/10.1038/s41598-021-97223-5>
- BATISTA TB, MASTRANGELO CB, DE MEDEIROS AD, PETRONILIO ACP, FONSECA DE OLIVEIRA GR, DOS SANTOS IL, CRUSCIOL CAC AND AMARAL DA SILVA EA (2022) A Reliable Method to Recognize Soybean Seed Maturation Stages Based on Autofluorescence-Spectral Imaging Combined with Machine Learning Algorithms. *Front. Plant Sci.* 13:914287. doi: 10.3389/fpls.2022.914287
- BIANCHINI, Vitor *et al.* Multispectral and X-ray images for characterization of *Jatropha curcas* L. seed quality. *Plant Methods*, [s. l.], v. 17, n. 1, 2021c. Disponível em: <https://doi.org/10.1186/s13007-021-00709-6>.
- BRASIL. Ministério da Agricultura, Pecuária e Abastecimento. Regras para análise de sementes. Ministério da Agricultura, Pecuária e Abastecimento. Secretaria de Defesa Agropecuária. Brasília, DF: MAPA/ ACS, 2009. 395p.

- CARREIRO SOARES, S. F. *et al.* Classification of individual cotton seeds with respect to variety using near-infrared hyperspectral imaging. *Analytical Methods*, 2016. v. 8, n. 48, p. 8498–8505.
- CONCEIÇÃO, R. R. P., QUEIROZ, V. A. V., MEDEIROS, E. P., ARAÚJO, J. B., ARAÚJO, D. D. S., MIGUEL, R. A., ... & Simeone, M. L. F. (2024). Determinação do teor de fumonisina em milho usando tecnologia de imagem hiperespectral no infravermelho próximo (NIR-HSI) e métodos quimiométricos. *Brazilian Journal of Biology*, 84, e277974.
- CUI, Hai-Feng *et al.* Automatic and Rapid Discrimination of Cotton Genotypes by Near Infrared Spectroscopy and Chemometrics. *Journal of Analytical Methods in Chemistry*, [s. l.], v. 2012, p. 1–7, 2012. Disponível em: <https://doi.org/10.1155/2012/793468>.
- DA SILVA, Fabiano França. **Relações entre a análise multiespectral e o potencial fisiológico de sementes de soja**. 2021. Tese de Doutorado. Escola Superior de Agricultura “Luiz de Queiroz”.
- DA SILVA, Martha Freire *et al.* Near infrared spectroscopy for the classification of vigor level of soybean seed. *Ciência Agrônômica/Revista Ciência Agrônômica*, [s. l.], v. 55, 2023. Disponível em: <https://doi.org/10.5935/1806-6690.20240005>.
- DE JESUS MARTINS BIANCHINI, Vitor *et al.* Multispectral and X-ray images for characterization of *Jatropha curcas* L. seed quality. *Plant Methods*, [s. l.], v. 17, n. 1, 2021a. Disponível em: <https://doi.org/10.1186/s13007-021-00709-6>.
- DEMSAR, Janez; CURK, Tomaž; ERJAVEC, Aleš; GORUP, Črt; HOČEVAR, Tomaž; MILUTINOVIČ, Mitar; MOŽINA, Martin; POLAJNAR, Matija; TOPLAK, Marko; STARIČ, Anže; ŠTAJDOHAR, Miha; UMEK, Lan; ŽAGAR, Lan; ŽBONTAR, Jure; ŽITNIK, Marinka; ZUPAN, Blaž. Orange: data mining toolbox in Python. *Journal of Machine Learning Research*, v. 14, p. 2349–2353, 2013. Disponível em: <http://jmlr.org/papers/v14/demsar13a.html>. Acesso em: 19/03/2025.
- FERREIRA, Eric B.; CAVALCANTI, Pórtya P.; NOGUEIRA, Denismar A. ExpDes: An R Package for ANOVA and Experimental Designs. *Applied Mathematics*, [s. l.], v. 05, n. 19, p. 2952–2958, 2014. Disponível em: <https://www.scirp.org/journal/paperinformation?paperid=51204>.
- KNIGHTON, Nick; BUGBEE, Bruce. A mixture of barium sulfate and white paint is a low-cost substitute reflectance standard for spectralon®. *Tech Instrum*, v. 11, p. 4-6, 2005.
- KRZYŻANOWSKI, F.C. *et al.* Testes de vigor baseado em desempenho de plântulas. In: KRZYŻANOWSKI, F.C.; VIEIRA, R. D.; FRANÇA-NETO, J.B.; MARCOS-FILHO, J. Vigor de sementes: conceitos e testes. Londrina: ABRATES, p.79-127, 2020.
- LEVER, Jake; KRZYWINSKI, Martin; ALTMAN, Naomi. Model selection and overfitting. *Nature Methods*, [s. l.], v. 13, n. 9, p. 703–704, 2016. Disponível em: <https://www.nature.com/articles/nmeth.3968>.
- MAGUIRE, J. D. Speed of germination – aid in selection and evaluation for seedling emergence and vigor. *Crop Science*, v. 2, n. 2, p. 176-177, 1962.
- MARCOS FILHO, J. Fisiologia de Sementes de Plantas Cultivadas. 2ª edição. ABRATES, Londrina, 2015.

- MARCOS FILHO, J. Teste de envelhecimento acelerado. In: Krzyzanowski, F.C.; Vieira, R.D.; França-Neto, J.B. (Eds.). Vigor de sementes: conceitos e testes. Londrina: ABRATES. Cap.3, p1-24, 2021.
- MASILAMANI, P. *et al.* Role of Near - Infrared Spectroscopy in Seed Quality Evaluation: A Review. *Agricultural Reviews*, [s. l.], n. OF, 2020. Disponível em: <https://doi.org/10.18805/ag.r-1960>.
- MATTIONI, F. *et al.* Vigor de sementes e desempenho agrônômico de plantas de algodão. *Revista Brasileira de Sementes*, 2012. v. 34, n. 1, p. 108–116.
- MESQUITA PINHEIRO, Romário *et al.* Processamento de imagens como ferramenta importante para inteligência artificial no setor de sementes. *Revista Agraria Academica*, [s. l.], v. 5, n. 1, p. 89–101, 2022. Disponível em: <https://doi.org/10.32406/v5n1/2022/89-101/agrariacad>.
- OLESEN, Merete *et al.* Viability Prediction of *Ricinus cummunis* L. Seeds Using Multispectral Imaging. *Sensors*, [s. l.], v. 15, n. 2, p. 4592–4604, 2015. Disponível em: <https://doi.org/10.3390/s150204592>.
- PÁDUA, G.P.; VIEIRA, R.D. Deterioração de sementes de algodão durante o armazenamento. *Revista Brasileira De Sementes*, [s. l.], v. 23, n. 2, p. 255–262, 2001. Disponível em: <https://doi.org/10.17801/0101-3122/rbs.v23n2p255-262>.
- PANDE, R.; MISHRA, H. N. Fourier Transform Near-Infrared Spectroscopy for rapid and simple determination of phytic acid content in green gram seeds (*Vigna radiata*). *Food Chemistry*, 2015. v. 172, p. 880–884. Disponível em: <http://dx.doi.org/10.1016/j.foodchem.2014.09.049>.
- PEREIRA, Cláudia Donald *et al.* Guia de medição e cálculo para refletância e absortância solar em superfícies opacas (v. 1). **Florianópolis: UFSC**, 2015.
- QUEIROGA, V. P. *et al.* Qualidade Fisiológica De Sementes De Algodão Armazenadas Em Função De Diferentes Tratamentos E Cultivares. *Revista Brasileira de Produtos Agroindustriais*, 2009. v. 11, n. 1, p. 43–54.
- R Core Team (2023). R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria. <https://www.R-project.org/>.
- RAMOS, Nilza Patrícia *et al.* Envelhecimento acelerado em sementes de rúcula (*Eruca sativa* L.). *Revista Brasileira De Sementes*, [s. l.], v. 26, n. 1, p. 98–103, 2004b. Disponível em: <https://doi.org/10.1590/s0101-31222004000100015>.
- SALIMI, Zahra; BOELT, Birte. Classification of Processing Damage in Sugar Beet (*Beta vulgaris*) Seeds by Multispectral Image Analysis. *Sensors*, [s. l.], v. 19, n. 10, p. 2360, 2019. Disponível em: <https://doi.org/10.3390/s19102360>.
- SANCHES, I. D., TUOHY, M. P., HEDLEY, M. J., & BRETHERTON, M. R. (2009). *Large, durable and low-cost reflectance standard for field remote sensing applications. International Journal of Remote Sensing*, 30(9), 2309–2319. doi:10.1080/01431160802549377
- SANTOS, M. B. H. *et al.* Non-destructive NIR spectrometric cultivar discrimination of castor seeds resulting from breeding programs. *Journal of the Brazilian Chemical Society*, 2014. v. 25, n. 5, p. 969–974.

- SCHINDELIN, J., ARGANDA-CARRERAS, I., FRISE, E. *et al.* Fiji: an open-source platform for biological-image analysis. *Nat Methods* 9, 676–682 (2012). <https://doi.org/10.1038/nmeth.2019>
- SILVA, Freire Da *et al.* Artificial aging for predicting the storability of soybean seeds via GGE biplot *Journal of Seed Science*. [S. l.: s. n.], 2024. Disponível em: <http://dx.doi.org/10.1590/2317-1545v46270874>. .
- SILVA, Josias Conceição *et al.* Desempenho de sementes de algodão após o processamento e armazenamento. *Revista Brasileira De Sementes*, [s. l.], v. 28, n. 1, p. 79–85, 2006. Disponível em: <https://doi.org/10.1590/s0101-31222006000100011>.
- SOARES, Júlia Martins *et al.* Low-cost system for multispectral image acquisition and its applicability to analysis of the physiological potential of soybean seeds. *Acta Scientiarum Agronomy*, [s. l.], v. 45, p. e57060, 2022. Disponível em: <https://doi.org/10.4025/actasciagron.v45i1.57060>.
- SOARES, Sófacles Figueredo Carreiro *et al.* Classification of individual cotton seeds with respect to variety using near-infrared hyperspectral imaging. *Analytical Methods*, [s. l.], v. 8, n. 48, p. 8498–8505, 2016b. Disponível em: <https://doi.org/10.1039/c6ay02896a>.
- VIEIRA, R. D; MARCOS-FILHO, J. Teste de condutividade elétrica. In: KRZYZANOWSKI, F.C.; VIEIRA, R. D.; FRANÇA-NETO, J.B.; MARCOS-FILHO, J. Vigor de sementes: conceitos e testes. Londrina: ABRATES, p.182-244, 2020.
- WANG, Xuemeng *et al.* Non-Destructive Identification of Naturally Aged Alfalfa Seeds via Multispectral Imaging Analysis. *Sensors*, [s. l.], v. 21, n. 17, p. 5804, 2021. Disponível em: <https://doi.org/10.3390/s21175804>.
- XIA, Yu *et al.* Recent advances in emerging techniques for non-destructive detection of seed viability: A review. *Artificial Intelligence in Agriculture*, [s. l.], v. 1, p. 35–47, 2019. Disponível em: <https://doi.org/10.1016/j.aiia.2019.05.001>.
- ZHANG X, LIU F, HE Y, LI X. Application of hyperspectral imaging and chemometric calibrations for variety discrimination of maize seeds. *Sensors*. 2012; 12:17234–6. <https://doi.org/10.3390/s121217234>.
- ZHU, S. *et al.* Near-infrared hyperspectral imaging combined with deep learning to identify cotton seed varieties. *Molecules*, 2019. v. 24, n. 18.