

JÚLIA MARTINS SOARES

**ESPECTROSCOPIA NO INFRAVERMELHO PRÓXIMO E MÉTODOS
QUIMIOMÉTRICOS PARA CLASSIFICAÇÃO DE SEMENTES DE SOJA QUANTO
AO POTENCIAL FISIOLÓGICO**

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

Orientador: Laércio Junio da Silva

**VIÇOSA - MINAS GERAIS
2023**

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

T

S676e
2023 Soares, Júlia Martins, 1998-
Espectroscopia no infravermelho próximo e métodos
quimiométricos para classificação de sementes de soja quanto ao
potencial fisiológico / Júlia Martins Soares. – Viçosa, MG, 2023.
1 dissertação eletrônica (55 f.): il. (algumas color.).

Orientador: Laércio Junio da Silva.
Dissertação (mestrado) - Universidade Federal de Viçosa,
Departamento de Agronomia, 2023.
Referências bibliográficas: f. 45-55.
DOI: <https://doi.org/10.47328/ufvbbt.2023.336>
Modo de acesso: World Wide Web.

1. *Glycine max*. 2. Soja - Semente - Qualidade.
3. Espectroscopia de infravermelho próximo. 4. Quimiometria.
I. Silva, Laércio Junio da, 1984-. II. Universidade Federal de
Viçosa. Departamento de Agronomia. Programa de
Pós-Graduação em Fitotecnia. III. Título.

CDD 22. ed. 633.3421

JÚLIA MARTINS SOARES

**ESPECTROSCOPIA NO INFRAVERMELHO PRÓXIMO E MÉTODOS
QUIMIOMÉTRICOS PARA CLASSIFICAÇÃO DE SEMENTES DE SOJA QUANTO
AO POTENCIAL FISIOLÓGICO**

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

APROVADA: 01 de março de 2023.

Assentimento:



Documento assinado digitalmente

JULIA MARTINS SOARES

Data: 30/05/2023 18:16:11-0300

Verifique em <https://validar.iti.gov.br>

Júlia Martins Soares
Autora



Documento assinado digitalmente

LAERCIO JUNIO DA SILVA

Data: 31/05/2023 12:38:39-0300

Verifique em <https://validar.iti.gov.br>

Laércio Junio da Silva
Orientador

*Ao meu avô e maior incentivador, José Soares,
dedico.*

AGRADECIMENTOS

À Universidade Federal de Viçosa e ao Departamento de Agronomia, pela minha formação e oportunidade de cursar o mestrado.

À Coordenação de Aperfeiçoamento de Pessoal de Nível Superior – Brasil (CAPES) – Código de Financiamento 001, pelo apoio neste trabalho.

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

Ao Professor Laércio Junio da Silva, pela oportunidade, orientação, paciência, suporte, confiança e amizade.

À Professora Denise Cunha Fernandes dos Santos Dias pelo apoio, inspiração, ensinamentos e amizade.

Ao coorientador Daniel Teixeira Pinheiro pelo voto de confiança, amizade, estímulo e parceria na condução deste trabalho.

Aos integrantes do GSem e amigos Bruno Noronha, Geovana Brito, Marcelo Limão, João Paulo Oliveira, Amanda Ramos, Tainara Maciel, Soryana Melo, Joyce Araújo, Wander Pereira, Wanderson Xavier, Gabriel Peris, Fernanda Mara e José Custódio, pelo companheirismo, ajuda na condução do trabalho e por tornar os dias mais divertidos e agradáveis.

Ao Felipe Quevedo pelo apoio, carinho, companheirismo, cuidado e incentivo em todos os momentos.

Às minhas irmãs, Livia, Luíza e Clara por toda cumplicidade, amor, paciência, parceria e incentivo.

À toda minha família, em especial à minha tia Arlete Soares pelo suporte, amparo e colo. Ao meu tio Geraldo César e primos Samara e Felipe por terem me apoiado com tanta generosidade em minha formação. Ao meu tio Airton Soares (*in memoriam*) pelo incentivo à curiosidade sobre o mundo e paixão pela vida.

Aos meus avós, que não estudaram, mas nunca mediram esforços para que eu o fizesse.

Aos meus pais pela dádiva da vida e por acreditarem em mim.

A todos que de alguma forma contribuíram para a execução desse trabalho.

Muito obrigada!

RESUMO

SOARES, Júlia Martins, M.Sc., Universidade Federal de Viçosa, março de 2023. **Espectroscopia no infravermelho próximo e métodos quimiométricos para classificação de sementes de soja quanto ao potencial fisiológico.** Orientador: Laércio Junio da Silva.

A espectroscopia no infravermelho próximo (NIR) consiste em uma ferramenta promissora para otimização das análises de sementes. A utilização do NIR, em combinação com métodos quimiométricos ou associados a novas técnicas, pode permitir a investigação dos níveis de qualidade de sementes, de forma rápida e assertiva. Este trabalho teve como objetivo investigar a viabilidade do NIR, associado a métodos quimiométricos, para classificar lotes de sementes de soja quanto ao potencial fisiológico. No primeiro experimento, foram utilizados 77 lotes de sementes de soja avaliados quanto ao potencial fisiológico e obtidos espectros NIR das amostras de sementes, com total de 12 leituras para cada lote. Os espectros originais foram submetidos a diferentes métodos de pré-processamento, sendo eles *Standard Normal Variate* (SNV), SNV + 1ª e 2ª derivadas; *Gap-segment derivative* (GSD); e Savitzky-Golay, pelas derivadas de primeiro e segundo grau, como também a combinação entre os métodos. Os lotes foram divididos em duas classes: alta qualidade ($\geq 90\%$ de germinação) e baixa qualidade ($< 90\%$ de germinação). Os espectros pré-processados foram utilizados para a construção de modelos de classificação por meio dos métodos *Partial Least Squares – Discriminant Analysis* (PLS-DA), *Naive Bayes* (NB), *Random Forest* (RF) e *Support Vector Machine* (SVM). No segundo experimento, foram utilizadas sementes de cinco genótipos, com três lotes de cada um. Foram realizados os testes de germinação, envelhecimento acelerado, condutividade elétrica, tetrazólio e análise das enzimas antioxidantes superóxido dismutase, catalase, peroxidases e peroxidase do ascorbato, além do conteúdo de malonaldeído, óleo e proteína. Os espectros NIR foram obtidos de sementes individuais, com total de 100 leituras para cada lote. Os espectros originais foram submetidos aos métodos de pré-processamento *Multiplicative Scatter Correction* (MSC), *Standard Normal Variate* (SNV), autoescalonamento, 1ª e 2ª derivadas de Savitzky-Golay e a combinação desses métodos. Os espectros pré-processados foram utilizados para a construção de modelos de classificação por meio do método PLS-DA. No primeiro experimento, o classificador PLS-DA apresentou maior acurácia de classificação. Os menores valores de acurácia foram obtidos para os modelos NB e RF. No segundo experimento, os lotes diferiram quanto ao potencial fisiológico e foram categorizados em alto, médio e baixo vigor. Os lotes de baixo vigor apresentaram maior conteúdo de malonaldeído, óleo e, em geral, menores

atividades das enzimas do sistema antioxidativo. Utilizando-se os espectros pré-processados por meio da 2ª derivada de Savitzky-Golay, foi possível obter um modelo de classificação com 98% de acurácia e 96% de coeficiente *kappa*, para os dados de validação. As regiões entre os comprimentos de ondas de 1.000-1.200 nm e 2.200-2.500 nm foram as mais importantes para distinguir os níveis de qualidade das sementes de soja. A espectroscopia NIR é uma alternativa não destrutiva, rápida e promissora, com grande potencial para ser utilizada no controle de qualidade visando à classificação de lotes de sementes de soja quanto ao potencial fisiológico.

Palavras-chave: *Glycine max* L. Merrill. Qualidade de sementes. Espectroscopia NIR. Quimiometria.

ABSTRACT

SOARES, Júlia Martins, M.Sc., Universidade Federal de Viçosa, March, 2023. **Near-infrared spectroscopy and chemometric methods for classifying soybean seeds according to physiological potential.** Adviser: Laércio Junio da Silva.

The Near-infrared (NIR) Spectroscopy is a promising tool for optimizing seed analysis. The use of NIR, in combination with chemometric methods or associated with new techniques, can allow the investigation of seed quality levels, quickly and assertively. This work aimed to investigate the viability of NIR, associated with chemometric methods, to classify soybean seed lots according to their physiological potential. In the first experiment, 77 soybean seed lots were evaluated for germination potential and NIR spectra were obtained from seed samples, with a total of 12 readings for each lot. The original spectra were submitted to different pre-processing methods, namely Standard Normal Variate (SNV), SNV + 1st and 2nd derivatives; Gap-segment derivative (GSD); and Savitzky-Golay, for the first- and second-degree derivatives, as well as the combination between these methods. Lots were divided into two classes: high quality ($\geq 90\%$ germination) and low quality ($< 90\%$ germination). The pre-processed spectra were used to build classification models using the Partial Least Squares – Discriminant Analysis (PLS-DA), Naive Bayes (NB), Random Forest (RF) and Support Vector Machine (SVM) methods. In the second experiment, seeds of five genotypes were used, with three lots of each. The seeds were evaluated for germination, accelerated aging, electrical conductivity and tetrazolium tests and analysis of the antioxidant enzymes superoxide dismutase, catalase, peroxidase and ascorbate peroxidase were carried out, as well as the content of malondialdehyde, oil and protein were determined. NIR spectra were obtained from individual seeds, with a total of 100 readings for each lot. The original spectra were submitted to Multiplicative Scatter Correction (MSC), Standard Normal Variate (SNV), autoscaling, 1st and 2nd derivatives of Savitzky-Golay and a combination of these methods. The pre-processed spectra were used to build a classification model using the PLS-DA method. In the first experiment, the PLS-DA classifier showed the highest classification accuracy. The lowest accuracy values were obtained for the NB and RF models. In the second experiment, the lots differed in terms of physiological potential and were categorized into high, medium, and low vigor. The low vigor seed lots showed a higher content of malondialdehyde, oil and, in general, lower activities of the enzymes of the antioxidative system. Using the pre-processed spectra using the 2nd derivative of Savitzky-Golay, it was possible to obtain a classification model with 98% accuracy and 96%

kappa coefficient for the validation data. The regions between the wavelengths of 1,000-1,200 nm and 2,200-2,500 nm were the most important to distinguish the quality levels of soybean seeds. NIR spectroscopy is a non-destructive, fast, and promising alternative, with great potential to be used in internal quality control in order to classify soybean seed lots according to their physiological potential.

Keywords: *Glycine max* L. Merrill. Seed quality. NIR spectroscopy. Chemometry.

SUMÁRIO

INTRODUÇÃO.....	10
REVISÃO DE LITERATURA	14
Avaliação do potencial fisiológico de sementes de soja.....	14
Deterioração de sementes	15
Espectroscopia no Infravermelho Próximo (NIR)	16
Pré-tratamentos para os dados espectrais.....	18
Modelos para classificação de dados NIR	19
MATERIAL E MÉTODOS.....	20
Experimento I - Métodos quimiométricos para classificação do potencial fisiológico de sementes de soja pela Espectroscopia no Infravermelho Próximo.....	20
Grau de umidade	21
Germinação	21
Aquisição dos espectros NIR.....	21
Algoritmos de pré-processamento	21
Desenvolvimento de modelos para classificação.....	22
Validação dos modelos	22
Experimento II -Associação da espectroscopia NIR com a qualidade fisiológica, composição bioquímica e atividade enzimática em sementes de soja.....	22
Grau de umidade	23
Germinação	23
Envelhecimento acelerado	23
Condutividade elétrica	23
Tetrazólio	23
Peroxidação de lipídeos	24
Atividade enzimática	24
Teor de óleo e proteína total	26
Espectroscopia no infravermelho próximo	26
Delineamento experimental e análise estatística	26
RESULTADOS E DISCUSSÃO.....	27
CONCLUSÕES	45
REFERÊNCIAS BIBLIOGRÁFICAS	45

INTRODUÇÃO

A soja [*Glycine max* (L.) Merrill] é uma das culturas agrícolas mais produzidas em todo o mundo, dada sua importância como fonte de óleo e proteína. É usada para alimentação humana e animal, com grande potencial de expandir os atuais usos industriais do óleo. Tamanha produção é resultado dos avanços no melhoramento genético com a incrementação da biotecnologia e adoção de culturas transgênicas, que mudaram a indústria agrícola e impactaram a produção e comercialização das cultivares. À vista disso, as prioridades dos programas de melhoramento da soja nos últimos anos se concentraram no aumento da produtividade, resistência e tolerância a pragas, doenças, herbicidas e na composição e qualidade das sementes (ANDERSON et al., 2019).

Para atender à alta demanda de soja em todo o mundo, o estabelecimento bem-sucedido das plantas no campo é primordial e depende diretamente da qualidade das sementes (BAEK et al., 2019). A utilização de sementes de alta qualidade culmina na germinação e emergência de plântulas em campo de maneira rápida e uniforme, o que torna possível o desenvolvimento de plantas com elevado potencial produtivo (EBONE et al., 2020).

A semente considerada de alta qualidade deve ter quatro atributos mínimos de qualidade garantidos: qualidade física, sendo ideal lotes sem a presença de material inerte; qualidade genética, que indica a ausência de misturas varietais no lote; qualidade sanitária, que atesta que o lote está livre de patógenos e sementes de plantas daninhas; e qualidade fisiológica, que indica a germinação e o vigor das sementes. Para a comercialização das sementes de soja, impõe-se que os lotes atinjam o mínimo de 80% de germinação, conforme o padrão estabelecido no território nacional (BRASIL, 2013).

A qualidade fisiológica é determinada pelo teste de germinação, que expressa o potencial máximo de germinação das sementes, e pelos testes de vigor, que são essenciais no controle de qualidade interno das empresas produtoras de sementes. O conceito de vigor engloba o conjunto de atributos da semente que permite obter um estande adequado, rápido e uniforme de plântulas sob uma ampla gama de condições ambientais (MARCOS-FILHO, 2015). Para a soja, os testes de vigor mais comumente utilizados são o envelhecimento acelerado, condutividade elétrica e tetrazólio.

O teste de envelhecimento acelerado estimula o processo de deterioração das sementes após submetê-las a um ambiente de alta temperatura e umidade relativa, por vezes é associado à capacidade de armazenamento e tem duração de até sete dias (MARCOS-FILHO, 2020). O teste de condutividade elétrica indica a integridade das membranas por meio da avaliação da lixiviação de íons e é um indicador de deterioração (VIEIRA; MARCOS-FILHO, 2020). O teste

de tetrazólio está fundamentado na caracterização dos tecidos vivos da semente e permite identificar danos pontuais, como danos mecânicos, por umidade e picada de percevejo por meio da avaliação de sementes individuais (FRANÇA-NETO, 2018).

A perda de qualidade das sementes reflete em uma série de alterações bioquímicas desencadeadas pela deterioração e pelo estresse oxidativo. A deterioração estimula a produção excessiva de espécies reativas de oxigênio (EROs), que em desequilíbrio com mecanismos antioxidantes, estimula a peroxidação e degradação de lipídios de membranas, oxidação de proteínas, danos a ácidos nucleicos e inibição enzimática, capazes de inviabilizar e até causar a morte das sementes (BEWLEY; NONOGAKI, 2017; EBONE et al., 2019).

Sementes oleaginosas como a soja possuem alta sensibilidade à deterioração e a análise da atividade de enzimas antioxidantes, bem como a quantificação de compostos como óleo e proteínas, por vezes são usadas como uma avaliação mais sensível do vigor de sementes e da capacidade de tolerância a deterioração (WILSON; McDONALD, 1986; WIJEWARDANA; REDDY; BELLALLOUI, 2019). Embora importante, análises bioquímicas são limitadas aos laboratórios de pesquisa e universidades por serem trabalhosas, caras, morosas necessitam de pessoas treinadas e podem fazer uso de reagentes tóxicos.

Devido ao alto volume de lotes de sementes recebidos pelas empresas, há o estímulo de avaliações da qualidade a partir de métodos mais rápidos, simples e que otimizem a tomada de decisão quanto à aprovação ou descarte dos lotes de sementes. Embora essas novas metodologias não substituam os testes padrões utilizados para o comércio, métodos de triagem não invasivos e de alto rendimento que permitam descartar lotes de baixa qualidade em uma rápida pré-seleção reduzem o volume de avaliações dos testes de germinação e vigor, resultando em economia de tempo, espaço e recursos.

Nesse sentido, a espectroscopia no infravermelho próximo (NIR) é uma técnica que tem conquistado espaço na avaliação da qualidade de sementes, por se mostrar uma técnica analítica simples, rápida, não destrutiva, não exige preparo prévio das amostras e não faz uso de reagentes tóxicos (BIANCHINI, 2021; REDDY et al., 2022). Esses métodos têm sido usados com sucesso para classificar sementes com base na capacidade e velocidade de germinação (MEDEIROS et al., 2020), sanidade (JAILLAIS et al., 2015) e viabilidade das sementes (OLESEN et al., 2015).

A técnica NIR é baseada na absorção de radiação eletromagnética em comprimentos de onda na faixa de 780 a 2.500 nm. Essa faixa fornece informações relacionadas à composição do material em estudo, de modo que a leitura dos espectros permite identificar a presença de grupos funcionais (C-H, N-H, O-H). Essa leitura depende da interação do composto com a

radiação eletromagnética, sendo absorvida por água, carboidratos, lipídios e proteínas (LARIOS et al., 2020).

Para a soja, o NIR tem sido usado principalmente para investigar o grão quanto ao teor de saponinas, ácidos graxos, flavonoides e antocianina (ZHANG et al., 2017; AMANAH et al., 2020). Também tem sido estudado como método para distinguir soja geneticamente modificada do material convencional (LEE; CHOUNG, 2011). Além disso, Lee et al. (2010) estudaram a técnica NIR para seleção de sementes individuais em F2 com base no teor de proteína, e comprovaram a eficiência do método na seleção de sementes para programas de melhoramento.

Apesar da técnica NIR ser bastante aplicada para a obtenção de informações a respeito da composição do grão, na área de tecnologia de sementes esta abordagem ainda é um tanto incomum. No mercado, o uso do NIR para classificação de lotes de sementes ainda é incipiente e são poucos os trabalhos que estudam a técnica em associação com a germinação e vigor. Considerando que a deterioração da semente altera sua composição e organização das estruturas celulares, sugere-se que a técnica NIR possa ser uma ferramenta eficiente para análise do potencial fisiológico das sementes.

Um dos entraves na aplicação da técnica NIR é o processamento dos dados e a escolha dos métodos de classificação. Os dados originais fornecidos pelo equipamento carecem da aplicação de pré-tratamentos a fim de reduzir a variabilidade e aprimorar a característica de interesse buscada nos espectros. Esses tratamentos minimizam o efeito do equipamento e ruído das amostras, de forma a melhorar a eficiência do modelo de calibração (RINNAN et al., 2009). Apesar de serem importantes, existe o risco de aplicar o tipo de pré-processamento inadequado ou um pré-processamento muito severo, que removerá informações importantes dos espectros. Por isso, é importante avaliar o efeito de diferentes métodos de pré-processamento dos dados fornecidos pelo NIR antes de definir o modelo final.

As técnicas de pré-processamento comumente usadas para tratamento dos espectros NIR são categorizadas em métodos de correção de dispersão e métodos derivativos. Como métodos de correção de dispersão, os métodos de transformação MSC (*Multiplicative Scatter Correction*), SNV (*Standard Normal Variate*) e normalização são os mais comuns. Essas técnicas são projetadas para reduzir a variabilidade física entre as amostras devido à dispersão. Todos eles também se ajustam para uma linha de base entre as amostras. Já os métodos derivativos são usados quando há necessidade de remover efeitos aditivos e multiplicativos nos espectros, em que a primeira derivada remove a linha de base, enquanto a segunda derivada remove a linha de base e a tendência linear. Usualmente são utilizadas a primeira e segunda derivada de Savitzky-Golay (SG), em que ambas as técnicas usam a suavização para não reduzir

muito a relação sinal-ruído nos espectros corrigidos (RINNAN et al., 2009). Além de cada um dos métodos mencionados, há a possibilidade de alterações nas janelas usadas para estimar a correção nos métodos derivativos, bem como a combinação de mais de um tipo de pré-processamento.

Estudos recentes que compreendem aplicações de diferentes modelos de classificação usando espectros de infravermelho, combinado com abordagens quimiométricas, têm sido descritos para sementes, obtendo-se alta precisão. Alguns dos métodos usados para modelagem de dados espectrais extraídos de sementes incluem PCA (*Principal Components Analysis*), CDA (*Canonical Discriminant Analysis*), *AdaBoost*, *Naive Bayes*, PLS-DA (*Partial Least Squares-Discriminant Analysis*), *Random Forest* e SVM (*Support Vector Machine*). Shrestha et al. (2016) classificaram sementes de tomate quanto a identidade e pureza varietal utilizando PCA, CDA e SVM, com acurácias variando de 94% a 100% para cinco cultivares, independentemente dos métodos quimiométricos utilizados. Kosmowski e Worku (2018) testaram modelos de classificação por meio de *AdaBoost*, *Naive Bayes*, PLS-DA, *Random Forest* e SVM para classificar sementes de cevada, grão de bico e sorgo, obtendo modelos com acurácia de 89%, 96% e 87%.

Após a escolha do modelo, a validação adequada é fundamental no processo de modelagem para assegurar que o modelo definido possa ser usado para classificar outras amostras e, sobretudo, para evitar previsões irrealistas e superestimadas da capacidade de classificar novos lotes (WESTAD; MARANI, 2015). O desempenho do modelo e capacidade de previsão pode ser avaliado usando validação cruzada ou validação de conjunto de teste e a eficiência mensurada pela acurácia e coeficiente *kappa*.

A precisão da classificação por meio da espectroscopia NIR pode ser influenciada pelo preparo da amostra, como o estado físico em sementes inteiras ou moídas, sementes individuais ou amostra, composição bioquímica, limpeza, cor, variação da temperatura e o modelo estatístico utilizado (AMANAH et al., 2020; PLANS et al., 2013; BAGCHI et al., 2016).

Neste trabalho propõe-se desenvolver modelos preditivos a partir de dados espectrais de sementes de 77 lotes de soja, de diferentes genótipos, que naturalmente variam em sua qualidade fisiológica. Posteriormente, o método proposto será associado aos testes de germinação, vigor e análises bioquímicas, avaliando-se a relação entre a qualidade das sementes e as regiões de comprimentos de onda que mais contribuíram para a classificação dos lotes.

REVISÃO DE LITERATURA

Avaliação do potencial fisiológico de sementes de soja

Nos laboratórios de análise de sementes, rotineiramente são conduzidos testes para avaliação da qualidade de sementes de soja. Nesse contexto, a introdução de novas técnicas modernas e rápidas, capazes de gerar resultados confiáveis para o controle interno de empresas tem grande potencial para avaliação da qualidade durante o processo de produção.

Comercialmente, a qualidade fisiológica de sementes é anunciada pelo teste de germinação. Este é um teste padrão e obrigatório para comercialização das sementes de soja e informa a porcentagem de plântulas normais observadas após a germinação em condições ideais de umidade e temperatura. A porcentagem de germinação fornece uma estimativa do potencial de um lote de sementes para germinação e estabelecimento de plântulas em condições favoráveis. A padronização do teste segue as Regras para Análise de Sementes (BRASIL, 2009) no território brasileiro, e a *International Seed Testing Association* (ISTA, 2020) para análises internacionais.

Embora não seja incomum atribuir o potencial fisiológico de sementes apenas ao teste de germinação, o vigor é um importante estimador da qualidade fisiológica. Esta não é uma propriedade que pode ser quantificada, mas descreve características como velocidade de germinação, capacidade de emergência em condições adversas e capacidade de germinação após o armazenamento (ISTA, 2017). Os objetivos básicos dos testes de vigor são diferenciar lotes com germinação semelhante, distinguir lotes de alto e baixo vigor e relacionar níveis de vigor à emergência e armazenabilidade (MARCOS-FILHO, 2020a). Apesar de não ser uma análise obrigatória para o comércio, as empresas sementeiras utilizam destes testes para um controle de qualidade interno mais efetivo.

Dentre os testes de vigor, os mais comumente utilizados para a cultura da soja são os testes de tetrazólio, condutividade elétrica e o envelhecimento acelerado. O teste de tetrazólio se baseia na atividade das enzimas desidrogenases, que catalisam as reações respiratórias na mitocôndria, reduzem o sal de tetrazólio nos tecidos vivos e resultam na coloração vermelha desses tecidos. Com base na coloração, propicia informações referentes à viabilidade e o vigor, além de fornecer o diagnóstico dos principais problemas que podem afetar o desempenho das sementes, como os danos causados por percevejo, danos mecânicos, durante a secagem e por umidade no campo (FRANÇA-NETO, 2020).

O teste de condutividade elétrica é validado pela ISTA para determinar o vigor de sementes de soja e caracteriza os lotes de acordo com a medição da quantidade de eletrólitos

liberados pelas sementes na solução de embebição. Sementes com potencial fisiológico inferior apresentam membranas menos organizadas e com redução da permeabilidade seletiva relacionada ao nível de deterioração da semente (VIEIRA; MARCOS-FILHO, 2020). A secagem das sementes durante a maturação pode causar a desestruturação das membranas celulares e, com o avanço da deterioração, a membrana perde a capacidade seletiva, que pode levar ao atraso na germinação. Por outro lado, sementes mais vigorosas possuem maior eficiência dos mecanismos de reparo e, conseqüentemente, menor lixiviação de eletrólitos na solução.

O teste de envelhecimento acelerado proporciona informações valiosas sobre diferenças no vigor entre lotes de sementes, utilizado principalmente para investigar o potencial de armazenamento e de emergência de plântulas em campo. Após submetidas às condições de alta temperatura e umidade relativa do ar, a taxa de deterioração das sementes é intensificada. Dessa forma, lotes de sementes que apresentam altas taxas de germinação após o envelhecimento são mais vigorosos (MARCOS-FILHO, 2020b). Este é um fator interessante para recomendação do ambiente de armazenamento do genótipo à determinada região agroclimática.

Os testes de vigor destacados são baseados na permeabilidade de membranas, atividade respiratória, velocidade de germinação e capacidade de gerar plântulas normais após estresse, cujos princípios foram baseados no processo de deterioração das sementes. Estes processos estão intimamente relacionados, uma vez que à medida que a deterioração avança, há redução no vigor (PINHEIRO et al., 2023).

Deterioração de sementes

A máxima qualidade fisiológica da semente é geralmente obtida no momento da maturidade fisiológica. Após a maturidade, são estimuladas atividades metabólicas e alterações bioquímicas graduais que causam deterioração de forma inevitável e irreversível que podem ser retardadas, mas não evitadas (DELOUCHE, 1979; SHELAR et al., 2008). A degradação das membranas é citada como a primeira manifestação da deterioração de sementes (DELOUCHE; BASKIN, 1973) e causa grande impacto principalmente às mitocôndrias, que são o centro de produção de energia das células (RATAJCZAK et al., 2019). O colapso das mitocôndrias prejudica mecanismos de energia e biossíntese de novas moléculas e culminam em germinação mais lenta, redução no potencial de armazenamento, menor crescimento e uniformidade das plântulas, maior sensibilidade a condições adversas, menor emergência, aumento de plântulas com aberrações morfológicas e, por fim, a completa deterioração com a perda do poder germinativo (MARCOS-FILHO, 2015; DELOUCHE; BASKIN, 1973).

Com o processo de deterioração da semente, há acúmulo de espécies reativas de oxigênio (EROs) na forma de ânion superóxido (O_2^-), peróxido de hidrogênio (H_2O_2), radical hidroxila ($OH^\cdot$), entre outros subprodutos formados a partir da redução incompleta do oxigênio (MITTLER, 2017). As EROs são moléculas naturais no metabolismo celular e participam de processos como sinalização e proteção contra estresses bióticos e abióticos, mas em concentrações elevadas são altamente tóxicas e causam danos às membranas celulares, proteínas, carboidratos, lipídios e, em casos mais graves, desencadeiam a morte das sementes (SHARMA et al., 2012).

Para minimizar os efeitos prejudiciais das EROs, as sementes dispõem de um sistema de defesa antioxidante, que inclui compostos não enzimáticos como o ascorbato, glutatona e flavonóides, bem como enzimas antioxidativas. Dentre elas, a superóxido dismutase (SOD), dismuta o radical superóxido a peróxido de hidrogênio e oxigênio. O peróxido de hidrogênio também é tóxico para as células e seu excesso é eliminado pela ação da catalase (CAT) e outras peroxidases, tais como ascorbato peroxidase (APX), peroxidase (POX), dentre outras (BOWLER et al., 1992; MITTLER, 2002). Portanto, o estresse oxidativo ocorre quando há o desequilíbrio entre a produção de EROs e a atuação do sistema de defesa antioxidante, de modo que o acúmulo de EROs sem aumento satisfatório da atividade do sistema antioxidante tem sido associado à perda da viabilidade das sementes (EBONE et al., 2019; SANO et al., 2016).

Sabe-se que sementes oleaginosas como a soja são mais sensíveis a deterioração pela formação de radicais livres e a consequente peroxidação de lipídios, cujos maiores danos têm sido associados a ruptura das membranas celulares, em especial nas mitocôndrias (FU; AHMED; DIEDERICHSEN, 2015; BAILLY et al., 2008). Um dos principais produtos da peroxidação lipídica é o malonaldeído (MDA). Assim, a quantificação do conteúdo de MDA é citada como de grande relevância em estudos de sementes para expressar a peroxidação de lipídios e a relação com a deterioração (GOEL; SHEORAN, 2003; XIN et al., 2014).

Dessa forma, o vigor das sementes é associado a capacidade de resistir aos efeitos deletérios de estresses que estimulam a deterioração. Os testes bioquímicos fornecem informações sobre a composição das sementes, conteúdo de moléculas tóxicas e são úteis para uma análise mais detalhada da deterioração (FU; AHMED; DIEDERICHSEN, 2015).

Espectroscopia no Infravermelho Próximo (NIR)

A espectroscopia no infravermelho próximo (NIR) tem sido introduzida nos últimos anos para análise química de alto rendimento e em larga escala na indústria para composição de grãos, farinhas, leites e carnes (ZHANG et al., 2022; BALABIN; SMIRNOV, 2011;

KADEMI; ULUSOY; HECER, 2018). Suas vantagens mais importantes são o curto tempo de análise, pouco ou nenhum preparo das amostras, além de não utilizar reagentes tóxicos.

A região eletromagnética do infravermelho próximo compreende a absorção da radiação em comprimentos de onda na faixa de 780 a 2.500 nm. O espectro NIR representa a radiação em uma frequência correspondente à vibração de uma ligação química absorvida, enquanto o restante da radiação é refletida ou transmitida sem interagir com outras ligações (WORKMAN; SHENK, 2015). A leitura espectral nesta faixa permite associar picos de absorção aos grupos funcionais C-H, N-H e O-H, associados à água e compostos orgânicos como carboidratos, lipídios e proteínas (LARIOS et al., 2020).

O uso de NIR com medições de refletância de amostras moídas foram adotadas como método de análises aprovados para medir proteína em cevada, aveia, trigo e soja (BAIANU; GUO, 2011). Leituras espectrais também foram capazes de separar tipos de óleos de sementes de canola, soja, girassol e cânhamo (OZAKI et al., 2006).

Especificamente para a soja, a espectroscopia NIR tem sido um instrumento valioso na quantificação de componentes químicos, com capacidade de determinar umidade, teor de proteínas, lipídios (FERREIRA et al., 2014), aminoácidos (KOVALENKO et al., 2006), antocianina (AMANAH, et al., 2020), saponinas, ácidos graxos e flavonoides (ZHANG et al., 2017). Em análises qualitativas, já foi possível determinar se grupos de sementes de soja eram transgênicas, resistentes a herbicida, ou não transgênicas (LEE; CHOUNG, 2011), separar classes de sementes danificadas por fungos de sementes saudáveis (WANG et al., 2007) e classificar sementes de alto vigor e sementes de baixo vigor submetidas ao envelhecimento artificial (SILVA, 2020).

Considerando que a técnica da espectroscopia no infravermelho próximo permite explorar diferenças na composição bioquímica, principalmente relacionadas à óleo, proteína e carboidratos, e sabendo que a deterioração da semente altera sua composição, sugere-se que a técnica NIR possa ser uma ferramenta eficiente para classificação qualitativa do potencial fisiológico das sementes.

O método para leitura espectral das sementes depende do objetivo da análise. A aquisição de espectros de sementes individuais permite mapear cada semente e é interessante para os programas de melhoramento genético, sendo possível verificar se houve contaminação genética ou selecionar os indivíduos que darão sequência ao programa com base no teor de proteínas (LEE et al., 2010), ou outros compostos. Já a análise de uma amostra a granel poderia ser indicada para caracterizar a qualidade do lote de sementes como um todo, com benefício de permitir descartar lotes de baixo potencial fisiológico e reduzir o tempo para tomada de decisão.

Pré-processamento para os dados espectrais

Como proposta de otimizar o processo de seleção de lotes de sementes quanto à qualidade fisiológica, espera-se que haja pouco ou nenhum preparo das amostras antes da triagem. A falta de homogeneidade na superfície da semente inteira, o tamanho e formato dificultam as análises por aumentar o espalhamento de luz durante as leituras utilizando o NIR. A luz emitida e em contato com a superfície da amostra emerge em vários ângulos e refletem efeitos aditivos e multiplicativos nos espectros de absorção (AGELET et al., 2012; RINNAN et al., 2009). Considerando um espectro base, percebe-se o efeito aditivo quando as leituras se deslocam de forma vertical e os efeitos multiplicativos quando há inclinação no espectro (DHANOA et al., 1994)

Os pré-processamentos mais utilizados para correção da dispersão da luz são o *Multiplicative Scatter Correction* (MSC) e o *Standard Normal Variate* (SNV). A correção pelo MSC é feita com base no espectro médio de todo o conjunto de dados, em que cada espectro é subtraído do espectro médio e os coeficientes de regressão linear fornecem as estimativas. A média dos espectros transformados é diferente de zero e a variância é relacionada ao espectro médio (DELWICHE, 1998; DHANOA et al., 1994).

A diferença para a correção por SNV é que esta é aplicada a cada espectro isolado, sem associação a todo o conjunto de dados. Por meio desta transformação, os espectros são centralizados e o valor de cada comprimento de onda dividido pelo desvio padrão da absorbância do espectro, de forma que todos os espectros têm média zero e variância um, sendo assim, independentes das características do conjunto de dados (DHANOA et al., 1994; RINNAN et al., 2009).

Enquanto o MSC e SNV removem a variação aditiva e multiplicativa da linha de base sem alterar a forma dos espectros, com os métodos derivativos há mudança inevitável na forma dos espectros (JONSSON; GABRIELSSON, 2009). A primeira derivada remove a linha de base, enquanto a segunda derivada remove a linha de base e a tendência linear. A primeira derivada é estimada pela diferença entre dois pontos no espectro, e a segunda derivada calcula a diferença entre dois pontos derivados pela primeira ordem (RINNAN et al., 2009).

O método de derivação proposto por Savitsky e Golay (1964) propõe incluir uma etapa de suavização a partir de configurações nas janelas dos espectros, que podem ser ajustadas sem reduzir muito o número de variáveis. Já a derivada do tipo *Gap-segment derivate* (GSD) separa dois segmentos de intervalo definido e calcula a média entre eles (HOPKINS, 2001).

Modelos para classificação de dados NIR

Para o desenvolvimento dos modelos de classificação dos lotes de sementes quanto ao vigor, há necessidade de reconhecimento de padrão ou agrupamento desenvolvidos com base nos espectros NIR da amostra (MORTENSEN et al., 2021). Os dados do NIR são altamente correlacionados e um método comum para minimizar a alta dimensionalidade é através da análise de componentes principais (PCA – *Principal Component Analysis*). A PCA também é interessante para visualizar previamente qualquer tendência nos dados antes da modelagem.

O PLS-DA é um método de classificação supervisionado, ou seja, possibilita determinar a qual classe pertence uma amostra a partir dos dados espectrais e a classe original dos lotes, que utiliza a técnica de regressão multivariada por mínimos quadrados parciais (PLS) (WOLD; SJÖSTRÖM; ERIKSSON, 2001). Os dados são utilizados para criar variáveis latentes para prever os valores de resposta dos espectros. Este método pode ser recomendado para operar conjuntos de dados em que o número de variáveis é muito maior que o número de amostras (ANDERSEN; BRO, 2010). O método PLS-DA já se mostrou eficiente para classificar dados espectrais relacionados à viabilidade de sementes de melão (KANDPAL et al., 2016), milho (ANDRADE; COELHO; UARROTA, 2020) soja (SILVA, 2020), lentilha (LIMÃO, 2022) e pimenta (SEO et al., 2016).

O algoritmo *Naive Bayes* (NB) considera as variáveis igualmente importantes e calcula a predição por meio dos princípios da probabilidade, relacionando a frequência dos preditores com as variáveis de classificação. O cálculo final considera a maior probabilidade para a classe. Calcula a probabilidade *a posteriori* multiplicando pela probabilidade *a priori* (BERRAR, 2019). O algoritmo NB combinado com o pré-processamento MSC foi considerado eficiente para identificar sementes de arroz danificadas pelo processo de secagem (ZHANG; RAO; JI, 2020), toxinas em sementes de amendoim (ZHANG et al., 2021) e variedades de cacau (CASTRO et al., 2022).

Random Forest (RF) é um método baseado em árvores de decisão, em que as folhas representam as classes e os ramos representam as regras de decisão que conduzem a amostra a uma classe específica (SANTANA; NETO; POPPI, 2019). Cada árvore vai prever uma resposta para cada amostra e a predição se dá pelo voto majoritário do conjunto de árvores (BREIMAN, 2001). A classificação de sementes por espectroscopia NIR utilizando *Random Forest* demonstrou eficiência para classificar sementes de trigo (SINGH et al., 2020), sementes de soja com danos mecânicos (WANG; HUANG; WANG, 2021) e germinação de *Urochloa brizantha* quando dados NIR foram combinados com dados de raios X (MEDEIROS et al., 2020).

Com relação a *Support Vector Machine* (SVM), este algoritmo ajusta um hiperplano entre as variáveis e, durante o treinamento, as variáveis são classificadas conforme estão na margem ou fora dela. A margem é definida pela largura em que o limite pode ser considerado antes de atingir um dos pontos de dados e o hiperplano é formado pelos pontos de dados que estão na margem, chamados de “vetores de suporte” (MAHADEVAN; SHAH, 2009). Este método foi utilizado para classificar sementes de arroz que passaram por estresse térmico (ZHANG; RAO; JI, 2020), variedades de sementes de milho doce (ZHOU et al., 2020) e milho comum (YANG et al., 2021).

A validação é uma parte importante para definir o modelo de classificação, de forma a garantir que ele seja eficiente para classificar outras sementes e evitar previsões irrealistas da capacidade de classificar novas amostras (WESTAD; MARINI, 2015).

Na validação cruzada, os dados totais são divididos em subconjuntos usados para treinar o modelo. Cada subconjunto é também segmentado e cada porção é testada, excluída, e a porção seguinte usada para testar o modelo com base nos segmentos restantes. Dessa forma, todas as amostras são usadas para calibrar o modelo. Parte dos dados é usado para calibrar o modelo pela validação cruzada (calibração) e testado com um outro conjunto de dados para validação (teste). A validação do conjunto de teste é considerada mais confiável porque utiliza um grupo de dados que não fez parte do desenvolvimento da calibração (MORTENSEN et al., 2021).

Esta é uma técnica de análise que poderia ser utilizada em programas de controle da qualidade de lotes de sementes. A prévia seleção de sementes de baixa qualidade e descarte das mesmas tornaria o processo de análise da qualidade muito mais rápido, eficiente, com economia de tempo, espaço e recursos.

MATERIAL E MÉTODOS

O experimento foi conduzido no Laboratório de Pesquisa de Sementes, do Departamento de Agronomia da Universidade Federal de Viçosa, *campus* Viçosa, MG.

Experimento I - Métodos quimiométricos para classificação do potencial fisiológico de lotes de sementes de soja por meio da espectroscopia no infravermelho próximo

Sementes comerciais de soja de 77 lotes de diferentes cultivares, produzidas na safra 2021/2022, foram avaliadas quanto ao grau de umidade, obtidos os espectros NIR das amostras de cada lote e determinada a germinação, conforme descrito abaixo:

Grau de umidade

Foi utilizado o medidor de umidade, modelo CM-500, fabricante Celm®. Foram utilizadas quatro repetições de cada lote em volume suficiente para preencher todo o equipamento e os resultados expressos em porcentagem. Devido a influência do teor de água na leitura espectral, foi considerado o limite máximo de variação entre o grau de umidade de $\pm 2\%$ entre os lotes.

Germinação

Foram utilizadas quatro repetições de 50 sementes distribuídas em rolo de papel toalha umedecido com quantidade de água equivalente a 2,5 vezes o peso de substrato seco. Os rolos foram mantidos em germinador a 25 °C e as avaliações da porcentagem de plântulas normais foram realizadas no quinto (primeira contagem de germinação) e oitavo (germinação) dia após o início do teste (BRASIL, 2009).

Aquisição dos espectros NIR

Foram selecionadas amostras aleatórias de sementes de cada lote, divididas em quatro repetições e as leituras realizadas em triplicata, totalizando 12 leituras por lote de semente. Cada amostra foi composta por aproximadamente 50 sementes, de maneira a cobrir toda a superfície de emissão da luz. Os espectros das amostras de sementes foram obtidos utilizando-se espectrômetro infravermelho com transformada de Fourier (Antaris II FT-NIR Analyzer; Thermo Scientific Co., Waltham, MA, EUA), que compreende a faixa de comprimento de onda de 1.000 a 2.500 nm. O software de suporte do espectrômetro TQ Analysis foi utilizado para leitura e registro dos espectros.



Algoritmos de pré-processamento

Os dados espectrais originais foram pré-processados usando métodos de correção de dispersão, sendo eles *Standard Normal Variate* (SNV); SNV + 1ª derivada e SNV + 2ª derivada; *Gap-segment derivative*; e Savitzky-Golay pelas derivadas de primeiro e segundo grau. Além

disso, foram testadas combinações entre os tipos de pré-processamentos e verificados os melhores métodos para utilizar no modelo de classificação.

Desenvolvimento de modelos para classificação

Classes: Os lotes de sementes de soja foram alocados em duas classes de acordo com os resultados do teste de germinação. A classe I foi formada pelos lotes que apresentaram germinação igual ou acima de 90%. A classe II foi formada pelos lotes com germinação abaixo de 90%. No total, 46,7% (n = 377 leituras espectrais) dos lotes foram alocados na classe I e 53,3% (n = 430 leituras espectrais) foram alocados na classe II. O limite de 90% de germinação foi escolhido para possibilitar o equilíbrio do número de lotes entre as classes, o que é importante para aumentar a eficiência dos modelos.

Métodos de classificação: Foram avaliadas as acurácias de quatro modelos de classificação utilizando os espectros obtidos por meio das diferentes técnicas de pré-processamento: *Partial Least Squares - Discriminant Analysis* (PLS-DA), *Naive Bayes* (NB), *Random Forest* (RF) e *Support Vector Machine* (SVM). A classificação foi realizada utilizando-se o software R (R Core Team, 2022), com a adaptação do *script* disponibilizado por Kosmowski e Worku (2018).

Validação dos modelos

Os modelos foram construídos utilizando-se 70% dos dados para treinamento (calibração) e os 30% restantes para validação. Além disso, validação cruzada de cinco vezes foi realizada para cada modelo no conjunto de calibração. Os modelos foram avaliados pela acurácia, ou seja, pela porcentagem de lotes classificados corretamente no conjunto de validação.

Experimento II - Associação da espectroscopia NIR com a qualidade fisiológica, composição bioquímica e atividade enzimática em sementes de soja

Foram utilizados 12 lotes de sementes de soja, oriundos de quatro diferentes genótipos e três lotes de cada. Os lotes foram obtidos de empresas produtoras de sementes e produzidos em diferentes áreas de multiplicação na safra 2021/2022. Todos os materiais são cultivares

transgênicas, com tecnologia Intacta RR2 IPRO. Foram avaliados o grau de umidade, porcentagem de germinação, testes de vigor e análises bioquímicas para caracterização dos lotes. Além disso, foram coletadas leituras espectrais NIR de cada lote e posteriormente testados os diferentes métodos de pré-processamento e classificação.

Grau de umidade

Foi utilizado o método da estufa, a 105 °C, por 24 h. Foram utilizadas quatro repetições de 20 de sementes cada. Os resultados foram expressos em porcentagem (BRASIL, 2009).

Germinação

Foram utilizadas quatro repetições de 50 sementes distribuídas em rolo de papel toalha umedecido com quantidade de água equivalente a 2,5 vezes o peso de substrato seco. Os rolos foram mantidos em germinador a 25 °C e as avaliações da porcentagem de plântulas normais foram realizadas no quinto (primeira contagem de germinação) e oitavo (germinação) dia após o início do teste (BRASIL, 2009).

Envelhecimento acelerado

Foram utilizadas quatro repetições de 50 sementes cada, distribuídas em camada única sobre tela metálica dentro de caixa tipo *gerbox*, contendo 40 ml de água destilada ao fundo. As caixas foram mantidas fechadas em BOD a 41 °C por 48 horas (MARCOS-FILHO, 2020). Após esse período, as sementes foram submetidas ao teste de germinação e a porcentagem de plântulas normais avaliada cinco dias após o início do teste de germinação.

Condutividade elétrica

Quatro repetições de 50 sementes foram pesadas e colocadas em copos plásticos contendo 75 ml de água destilada. Os copos foram mantidos em câmaras de germinação a 25 °C, por 24 horas (VIEIRA; MARCOS-FILHO, 2020). Após esse período, a condutividade elétrica da solução foi determinada em medidor de condutividade modelo Digimed DM-32. Os valores médios foram calculados e expressos em $\mu\text{S cm}^{-1} \text{ g}^{-1}$ de sementes.

Tetrazólio

Foram utilizadas 100 sementes de cada lote, divididos em duas repetições de 50 sementes cada. As sementes foram pré-condicionadas em papel-toalha previamente umedecido e mantidas em germinador a 25 °C por 16 horas. Em seguida, as amostras foram colocadas para

embeber e colorir na solução de cloreto de 2,3,5 trifenil tetrazólio a 0,075% e mantidas em BOD a 38 °C por duas horas (FRANÇA-NETO; KRZYZANOWSKI, 2020). Após esse período, as sementes foram lavadas e mantidas em água destilada para interromper o processo de coloração. A avaliação do teste seguiu a metodologia descrita por França-Neto e Krzyzanowski (2020), em que cada semente foi avaliada individualmente e qualificada em classes de 1 a 8 conforme o nível de dano, observando-se a ocorrência de danos mecânicos, danos por percevejo e deterioração por umidade. A porcentagem de sementes vigorosas foi obtida pelo somatório do número de sementes nas classes 1-3; de sementes viáveis nas classes de 1-5; e de sementes não viáveis nas classes de 6-8.

Peroxidação de lipídeos

A peroxidação de lipídeos foi determinada pelo conteúdo de malonaldeído (MDA), baseado na metodologia proposta por Cakmak e Horst (1991), adaptada para sementes. Para o preparo do extrato bruto, foram utilizados 150 mg de amostras de sementes liofilizadas e moídas, adicionadas de 1,8 ml de solução de ácido tricloroacético (TCA 1%. p/v). Em seguida, a solução foi centrifugada a 12.000 rpm por 15 minutos a 4 °C e 500 µl do sobrenadante foi coletado e adicionado 1,5 ml do meio de reação, solução de ácido tiobarbitúrico (TBA 0,5% em TCA 20%). As amostras e o branco permaneceram em banho maria a 90 °C por 20 minutos e foram novamente centrifugadas, por 4 minutos. Para leitura do branco, foi utilizado o meio de extração sem amostra, adicionada a solução do meio de reação. As leituras foram realizadas em espectrofotômetro nos comprimentos de onda 532 e 600 nm, em que a concentração de MDA foi calculada utilizando o coeficiente de extinção molar $155 \text{ mM}^{-1} \text{ cm}^{-1}$ e os resultados foram expressos em nmol g^{-1} de massa seca.

Atividade enzimática

Para o preparo das amostras, quatro repetições de 20 sementes foram embebidas por 16 horas, os tegumentos retirados e os embriões liofilizados e moídos em moinho de bolas por um minuto, até a obtenção de um pó fino. Os extratos enzimáticos foram obtidos a partir de subamostras de 150 mg, adicionados a 2 ml do meio de extração, constituído de tampão fosfato de potássio 0,1 M, a pH 6,8, ácido etilenodiaminotetracético (EDTA) 0,1 mM, fluoreto de fenilmetilsulfônico (PMSF) 1 mM e polivinilpolipirrolidona (PVPP) 1 % (p/v) (PEIXOTO et al., 1999). O homogeneizado foi centrifugado a 12.000 rpm, por 15 min, a 4 °C e o sobrenadante utilizado como extrato enzimático bruto. Todas as leituras de absorbância para determinação

das atividades enzimáticas foram operadas em espectrofotômetro modelo Thermo Scientific™ UV-Vis Genesys 105.

Superóxido dismutase (SOD): A atividade da SOD foi determinada pela adição de 50 µl do extrato enzimático a 2,95 µl de meio de reação. O meio de reação foi constituído de tampão fosfato de sódio 50 mM, a pH 7,8, contendo metionina 13 mM, azul de p-nitro tetrazólio (NBT) 75 µM, EDTA 0,1 mM e riboflavina 2 µM (DEL LONGO et al., 1993). O preparo dos meios foi conduzido em sala escura e para a reação, os tubos contendo as amostras foram mantidos em uma câmara de reação sob iluminação por 5 minutos. Após a exposição à luz, a iluminação foi interrompida e a formazana azul, produzida pela fotorredução do NBT, foi medida pela absorbância a 560 nm. Para o branco, a leitura foi realizada em uma amostra com o meio de reação igual ao anterior, porém sem exposição à luz e o valor foi subtraído da leitura das demais amostras (GIANNOPOLITIS e RIES, 1977). Uma unidade de SOD foi definida como a quantidade de enzima necessária para inibir em 50% a fotorredução do NBT (Beauchamp e Fridovich, 1971).

Catalase (CAT): Foi determinada pela adição de 50 µl do extrato enzimático bruto a 2,95 ml do meio de reação, constituído por tampão fosfato de potássio 50 mM, a pH 7,0 e H₂O₂ 12,5 mM (HAVIR e MCHALE, 1987). O espectrofotômetro foi zerado utilizando o meio de reação e a leitura realizada no comprimento de onda de 240 nm durante um minuto. O cálculo da atividade da catalase foi realizado com base na inclinação da reta da absorbância, ou seja, ΔAbs/min e o resultado expresso em µmol·min⁻¹·mg⁻¹ de proteína (ANDERSON; PRASAD; STEWART, 1995).

Ascorbato peroxidase (APX): Foi determinada pela adição de 50 µL do extrato enzimático bruto a 2,95 ml de meio de reação, constituído de tampão fosfato de potássio 50 mM, a pH 7,8, ácido ascórbico 0,25 mM, EDTA 0,1 mM e H₂O₂ 0,3 mM. O espectrofotômetro foi zerado com o meio de reação e a leitura realizada no comprimento de onda 290 nm durante um minuto. A absorbância foi calculada pelo decréscimo durante a leitura, ou seja, ΔAbs/min e o resultado expresso em nmol min⁻¹·mg⁻¹ proteína (NAKANO; ASADA, 1981).

Peroxidase (POX): Para determinação da atividade da peroxidase, 50 µL do extrato enzimático bruto foi adicionado a 2,95 ml do meio de reação, constituído por tampão fosfato de potássio 25 mM, a pH 6,8, piragalol 20 mM e H₂O₂ 20 mM (KAR e MISHRA, 1976). Para zerar o equipamento foi usado meio de reação e a leitura das amostras no comprimento de onda 420 nm. O aumento na absorbância foi calculado durante um minuto (ΔAbs/min) e a atividade

enzimática definida utilizando-se o coeficiente de extinção molar de $2,47 \text{ mM}^{-1}\text{cm}^{-1}$ (CHANCE e MAEHLEY, 1955) e expressa em $\mu\text{mol min}^{-1} \text{ mg}^{-1}$ proteína.

Determinação de proteínas solúveis: Os teores de proteínas solúveis dos extratos enzimáticos utilizados para a determinação das atividades enzimáticas foram determinados pelo método de Bradford (1976), utilizando albumina bovina (BSA) como padrão. 50 μl do extrato enzimático foram adicionados a 1,5 ml do reagente de Bradford, seguido de agitação. Após 20 minutos, foi realizada a leitura da absorbância da amostra no comprimento de onda de 595 nm.

Teor de óleo e proteína total

Os teores de óleo e proteínas totais foram determinados por espectroscopia no infravermelho próximo (NIR), por meio do equipamento FT-NIR (Thermo Scientific, modelo Antaris II), utilizando-se modelos previamente calibrados. Quatro amostras das sementes foram moídas e retiradas alíquotas de 10 g do pó para análise, na qual cada amostra foi lida em triplicata, totalizando 12 leituras por lote. Os resultados foram expressos em porcentagem.

Espectroscopia no infravermelho próximo

Para aquisição dos espectros, foram selecionadas amostras aleatórias de 100 sementes para cada lote e as leituras realizadas em sementes individuais, totalizando 1.500 espectros para análise. Cada espectro foi obtido na faixa entre 1.000 e 2.500 nm pelo espectrômetro Thermo Scientific, modelo Antaris II, em modo de reflectância (R).

Delineamento experimental e análise estatística

O experimento foi conduzido em delineamento inteiramente casualizado, sendo utilizados cinco genótipos e três lotes de cada, em quatro repetições para todos os testes fisiológicos. Os dados foram submetidos à análise de variância e as médias dos testes de germinação, vigor (PCG, EA, CE e TZ) e as determinações bioquímicas tiveram as médias comparadas pelo teste de Tukey, a 5% de probabilidade para cada genótipo.

Os lotes foram classificados conforme o desempenho fisiológico estabelecido pelos testes de vigor (alto, médio e baixo vigor). Os dados espectrais foram submetidos aos métodos quimiométricos de pré-processamento: *standard normal variate* (SNV), *multiplicative scatter correction* (MSC), primeira e segunda derivadas de Savitzky-Golay utilizando janela de 11 variáveis (SAVITZKY; GOLAY, 1964). Além disso, foram testadas combinações entre os dois métodos. Para cada pré-tratamento ou combinação, foi gerado um modelo de classificação por meio da PLS-DA (BARKER; RAYENS, 2003), em que 70% dos dados foram usados para

treinamento e 30% para validação. O desempenho dos modelos foi avaliado pela acurácia e *kappa* obtidos para os dados de treinamento e de validação. Ainda, foram identificadas as faixas de comprimento de onda que foram mais importantes para cada classe de vigor na construção do modelo de classificação. Todas as análises foram realizadas utilizando o software R (R Core Team, 2022).

RESULTADOS E DISCUSSÃO

Experimento I - Métodos quimiométricos para classificação do potencial fisiológico de sementes de soja pela Espectroscopia no Infravermelho Próximo

Os 77 lotes de sementes de soja foram avaliados quanto ao potencial de germinação, sendo que 36 lotes apresentaram valores iguais ou superiores a 90% e 41 lotes apresentaram germinação inferior a 90% (Figura 1). Assim, os lotes foram divididos em duas classes quanto ao potencial fisiológico, e, em seguida, foram obtidos os espectros NIR das amostras de cada um deles.

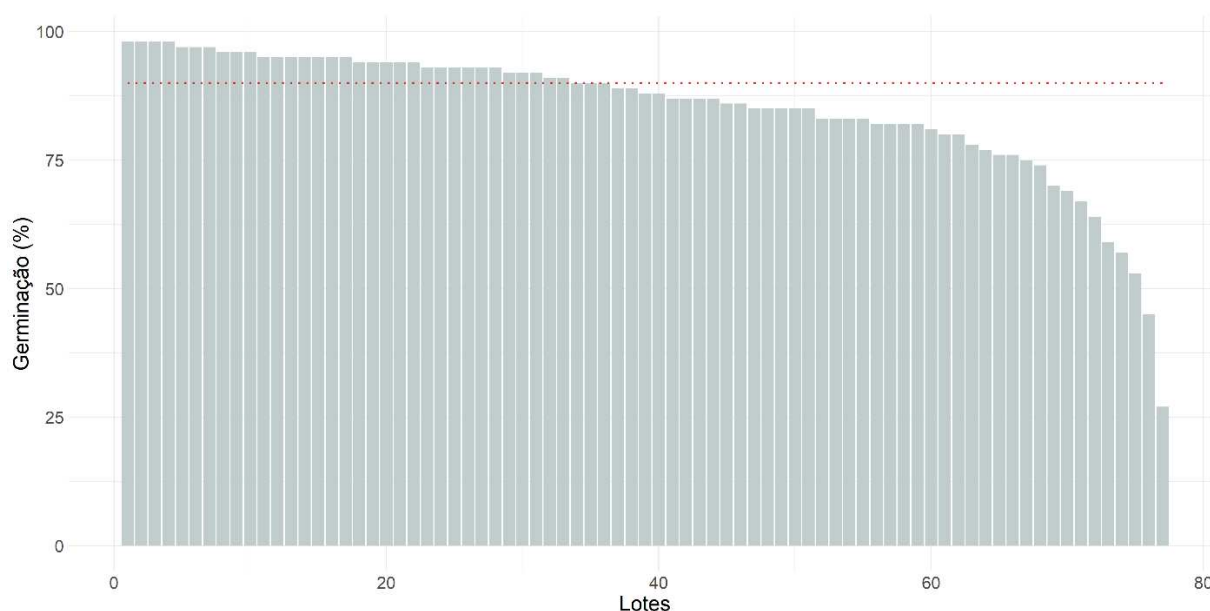


Figura 1. Germinação (%) dos 77 lotes de sementes de soja.

Por meio dos espectros brutos médios, foi possível identificar diferenças entre os picos de absorção das sementes das duas classes de potencial fisiológico. Portanto, este pode ser um indicativo de que diferenças na capacidade de germinação das sementes tem relação com a

composição ou alterações bioquímicas e pode ser avaliada pela leitura do espectro NIR. (Figura 2). O perfil espectral para ambas as classes foi semelhante, com picos de destaque em torno de 1.200 nm e 1.800 nm. Maiores valores de reflectância dos espectros brutos médios foram obtidos para os lotes de alta qualidade em quase toda a faixa espectral avaliada.

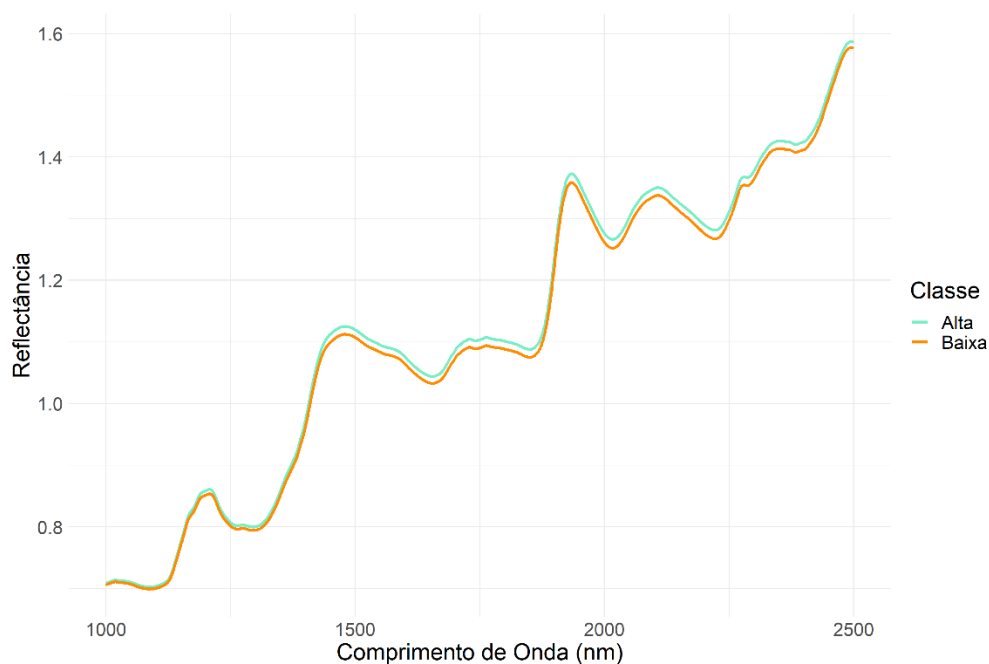


Figura 2. Espectros brutos médios de amostras de sementes de soja com diferentes níveis de qualidade fisiológica.

Foram testados diferentes tipos de pré-processamentos e modelos para classificação dos dois níveis de qualidade fisiológica das sementes, que demonstraram diferenças na eficiência de classificação, com variação da acurácia entre 51 e 78% (Figura 3). Na análise geral dos resultados, o classificador PLS-DA apresentou maiores valores de acurácias entre os métodos avaliados. O modelo SVM apresentou valores de acurácia satisfatórios para a maioria dos tipos de processamento, mas ainda inferior ao PLS-DA. Os menores valores de acurácia foram obtidos para os modelos *Naive Bayes* (NB) e *Random Forest* (RF).

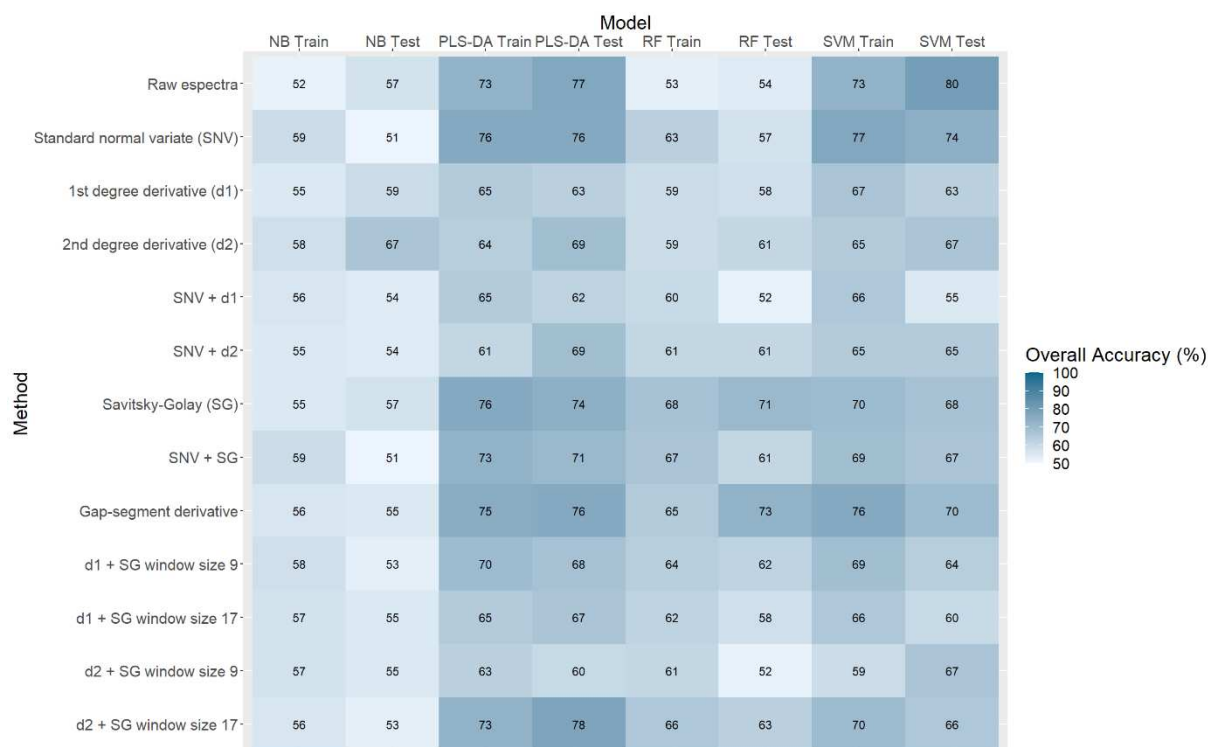


Figura 3. *Heatmap* da acurácia geral da classificação de sementes de soja em dois níveis de qualidade fisiológica (N = 77 lotes, n = 807 amostras) usando diferentes métodos de pré-processamento (linhas) e modelos (colunas) em espectros FT-NIR. NB - *Naïve Bayes*; PLS-DA - *Partial Least Squares – Discriminant Analysis*; RF - *Random Forest*; SVM – *Support Vector Machine*; Train - representa a calibração cruzada de 5 vezes no conjunto de treinamento das amostras (70%); Test - representa o conjunto de teste das amostras (30%).

Os modelos obtidos para os espectros pré-processados por SNV apresentaram alta acurácia utilizando PLS-DA e SVM (76% e 74% para os dados de teste, respectivamente). A correção dos espectros por SNV remove a variabilidade entre as amostras devido a dispersão da luz causada principalmente por efeitos físicos que induz variações nos espectros. A dispersão da luz nos estudos com NIR é muito influenciada pela diferença de tamanho das sementes e pela superfície esférica das amostras de sementes de soja.

Os métodos derivativos por Savitzky-Golay, GSD e a 2ª derivada de Savitzky-Golay com janela de 17 variáveis também tiveram maiores valores de acurácia em relação às demais técnicas e combinações. Métodos derivativos reduzem os ruídos com base na suavização dos dados pelo processo de convolução entre dois pontos no espectro. Dessa forma, o número de variáveis é reduzido e há destaque nas regiões de interesse. Com aumento da janela, o ruído tende a ser removido, mas se a janela for muito grande, picos de interesse também podem ser distorcidos e perder a importância na classificação. Por outro lado, as janelas menores podem reter mais ruídos (CHEN et al., 2004). Neste trabalho, o ajuste com janela de 17 apresentou maior acurácia que a janela de 9 para classificação dos lotes.

As derivadas e os métodos de correção de dispersão como o SNV removem efeitos diferentes e a combinação dos métodos pode ser benéfica para extrair informações dos espectros por remover maior quantidade de variáveis não relacionadas ao conteúdo de interesse. Apesar disso, neste trabalho foi observado que o SNV, combinado a quaisquer outros métodos de pré-processamentos, reduziu a acurácia quando comparada aos dados processados pelo SNV sozinho e removeu informações importantes para separação das classes de qualidade das sementes.

Em relação aos modelos de classificação, o PLS-DA foi superior aos outros para uma análise geral, visto que apresentou acurácia igual ou superior a 60% para todos os tipos de pré-processamentos, tanto nos dados de treinamento quanto nos dados de teste. Em segundo lugar, a classificação pelo modelo SVM apresentou valores de acurácia mais altos para os tipos de pré-processamentos SNV e GSD, seguido pela classificação por *Random Forest* com os pré-processamentos derivativos, Savitzky-Golay e GSD. Por último, o modelo *Naive Bayes* não apresentou bom desempenho na predição das classes para nenhum tipo de pré-processamento aplicado.

Em trabalhos com qualidade de sementes de outras espécies, o algoritmo PLS-DA também demonstrou melhor desempenho para os dados de espectroscopia. Medeiros et al. (2020) classificaram com maior precisão a capacidade de germinação de sementes de braquiária, com 82% de acurácia utilizando PLS-DA, enquanto SVM, *Random Forest* e *Naive Bayes* apresentaram acurácias de 79%, 77% e 69%, respectivamente.

Em modelo obtido utilizando-se o algoritmo *Random Forest*, foi observado maior capacidade para separação das classes de germinação de sementes de *Urochloa brizantha* para dados de espectroscopia NIR quando combinados com dados de raios X (MEDEIROS et al., 2020). Em identificação de danos por *Fusarium* em sementes de trigo, os modelos RF e NB foram mais eficientes que SVM (ZHANG et al., 2020). A utilização do SVM e NB também foi eficiente na classificação de sementes de arroz com diferentes graus de danos causados pelo calor (ZHANG; RAO; JI, 2020).

Embora por meio do SVM, RF e NB se tenha obtido eficiência na classificação de dados NIR de sementes em outros trabalhos, estes métodos não foram satisfatórios para classificar com segurança as amostras de sementes de soja (Figura 3). Diversos fatores podem ter contribuído para a diferença de resultados, como o equipamento utilizado para leitura espectral, o objetivo da classificação, a cultura, o tipo e volume de amostra, dentre outros.

Para o modelo PLS-DA em que foi utilizado o pré-tratamento 2ª derivada de SG, com janela de 17, os valores de acurácia, 78% e 73% para os dados de treinamento e validação,

respectivamente, estão entre os maiores obtidos. Assim, a partir do modelo obtido para os espectros foi construído o gráfico de importância geral das variáveis para a classificação das sementes em dois níveis de qualidade fisiológica (Figura 4). As faixas de comprimento de onda que apresentaram maior importância foram 1.000 - 1.200 nm, 1.800 - 1.900 nm e 2.250 - 2.300 nm.

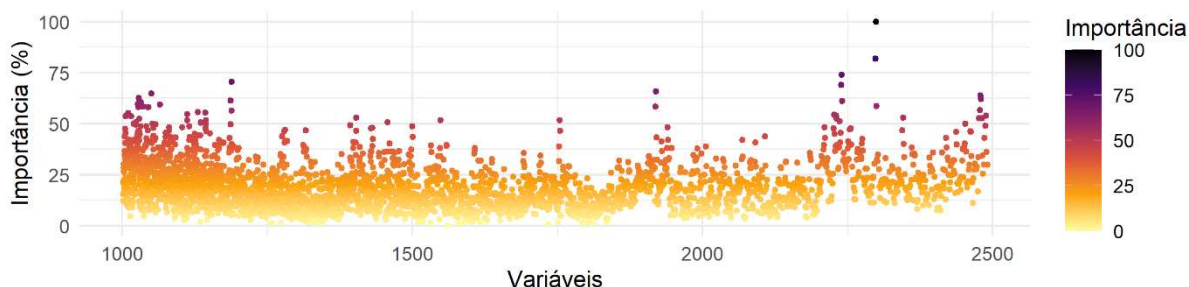


Figura 4. Importância geral das variáveis de comprimento de ondas para classificação via PLS-DA dos níveis de qualidade de sementes de soja. Os espectros foram pré-processados utilizando-se a 2ª derivada de Savitzky-Golay.

De acordo com os resultados obtidos no primeiro experimento, ficou evidenciado o potencial da espectroscopia no infravermelho próximo para a classificação de lotes de sementes de soja quanto a qualidade fisiológica.

O fato de as acurácias obtidas não terem ultrapassado o valor de 80% pode ser explicado por diversos fatores que compõem as amostras. Foram utilizados lotes que compreendem 75 diferentes genótipos de soja, ou seja, naturalmente possuem diferenças físicas e bioquímicas na composição das sementes. A diferença genética é uma variável que pode ser qualificada pelo NIR quando relacionadas ao conteúdo de proteínas, óleo, carboidratos e tem influência na leitura espectral.

A fim de validar o potencial da espectroscopia NIR para classificação dos lotes de sementes de soja e associar a leitura espectral com a composição bioquímica e a atividade enzimática, foi conduzido um segundo experimento utilizando lotes com diferentes níveis de qualidade fisiológica de um mesmo genótipo, sendo quatro genótipos diferentes. Os lotes foram submetidos à análise da qualidade fisiológica das sementes, quantificação do teor de óleo e proteína, avaliação da atividade de enzimas do sistema de defesa antioxidativo e espectroscopia NIR, utilizando a PLS-DA, modelo que apresentou maior acurácia no experimento 1.

Experimento II - Associação da espectroscopia NIR com a qualidade fisiológica, composição bioquímica e atividade enzimática em sementes de soja

Os 12 lotes de sementes de soja avaliados apresentaram diferenças no potencial fisiológico (Tabela 1). Foi observada variação de 74 a 95% entre o potencial de germinação dos lotes. A avaliação do vigor pela primeira contagem de germinação (PCG), condutividade elétrica (CE), envelhecimento acelerado (EA) e vigor pelo teste de tetrazólio (TZ) permitiram maior estratificação dos lotes. Dessa forma, os lotes apresentaram níveis contrastantes de qualidade fisiológica. A diferença entre a qualidade dos lotes dentro e entre os genótipos é um fator interessante para as análises subsequentes por fornecer maior representatividade dos dados. De acordo com a análise conjunta dos resultados de germinação e vigor, os lotes foram classificados como de alto, médio ou baixo vigor, de forma que para cada genótipo tivesse os três níveis de qualidade.

Tabela 1 – Caracterização inicial e médias dos parâmetros de qualidade fisiológica de 12 lotes de sementes de soja.

Genótipo	Lotes	PCG (%)	G (%)	EA (%)	CE ($\mu\text{S.cm}^{-1} \text{ g}^{-1}$)	TZ (%)	Classe de vigor
1	L1	91 a	94 a	71 a	127,49 a	95 a	Alto
	L2	82 b	86 b	77 a	141,05 a	83 b	Médio
	L3	75 b	81 b	74 a	133,88 a	79 b	Baixo
	CV (%)	4,32	3,64	4,93	10,17	3,39	
2	L1	92 a	95 a	87 a	125,21 a	87 a	Alto
	L2	91 a	94 a	86 a	93,00 b	84 a	Médio
	L3	80 b	87 b	84 a	99,78 ab	81 a	Baixo
	CV (%)	3,21	3,83	6,46	12,96	4,05	
3	L1	88 a	92 a	77 a	161,37 a	92 a	Alto
	L2	80 b	87 ab	82 a	158,31 a	78 b	Médio
	L3	75 b	85 b	68 b	176,39 a	74 c	Baixo
	CV (%)	4,07	3,38	4,02	13,4	2,47	
4	L1	90 a	92 a	77 a	166,21 b	86 a	Alto
	L2	85 a	89 a	76 a	212,83 a	81 a	Médio
	L3	69 b	74 b	70 a	237,20 a	70 b	Baixo
	CV (%)	4,78	4,26	6,01	9,47	3,51	

Letrais iguais em cada coluna e para cada genótipo não diferem entre si pelo teste de Tukey a 5% de probabilidade. Primeira contagem de germinação (PCG); germinação (G); envelhecimento acelerado (EA); condutividade elétrica (CE); e sementes vigorosas pelo teste de tetrazólio nas classes 1 a 3 (TZ).

A classe de vigor foi atribuída conforme os resultados obtidos nos testes para avaliação da qualidade, para cada genótipo e em análise conjunta dos testes. Quando não foi observada

diferença estatística entre os lotes, como por exemplo os lotes 2 e 3 do genótipo 1, foi considerado a tendência dos dados como critério para classificação.

Para a maioria dos lotes, o potencial de germinação apresentou diferença estatística condizente com a avaliação de sementes vigorosas pelo teste de tetrazólio, em que as maiores médias foram obtidas para os lotes de maior porcentagem de germinação (Tabela 1).

Lotes que apresentaram germinação estatisticamente semelhantes, quando submetidos aos testes de vigor, apresentaram diferenças significativas na qualidade. Foi observado baixo desempenho no teste de envelhecimento acelerado para a maioria dos lotes, com germinação abaixo de 80% e pouca variação dentro de cada genótipo. Isto indica que os lotes podem ser menos tolerantes ao armazenamento e apresentar menor desempenho no campo (MARCOS-FILHO, 2020).

O teste de germinação é conduzido sob condições ótimas para a semente e tende a expressar a capacidade máxima de germinação. Aliar os resultados deste teste com os dos testes de vigor é importante para estimar o potencial fisiológico das sementes em condições de campo e a resistência a estresses abióticos. Por esse motivo, a interpretação dos resultados com a finalidade de classificar lotes quanto ao vigor deve ser baseada em mais de um teste e em análise conjunta dos resultados.

A exemplo disto, o lote L2 do genótipo 3 apresentou bom desempenho no teste de germinação, mas obteve média inferior para o vigor pelo TZ. O lote L2 do genótipo 4, obteve alta germinação, mas com média alta no teste de CE. Elevados valores de condutividade elétrica indicam a presença de danos mecânicos e por umidade que levaram a maior lixiviação de íons, confirmada pelas médias inferiores no teste de tetrazólio. Além disso, a desestruturação e perda de integridade das membranas é atribuída ao processo de deterioração, com destaque para a peroxidação de lipídios, o que explica a alta sensibilidade à deterioração de sementes oleaginosas, como a soja (WILSON; MCDONALD, 1986).

A oxidação das membranas e compostos celulares, bem como alterações no metabolismo, são mecanismos importantes durante o processo de deterioração das sementes. Como forma de estimar o seu efeito deletério em lotes com diferentes níveis de vigor, foi quantificado o teor de malonaldeído (MDA) nas sementes, um dos subprodutos da peroxidação de lipídios (Figura 6).

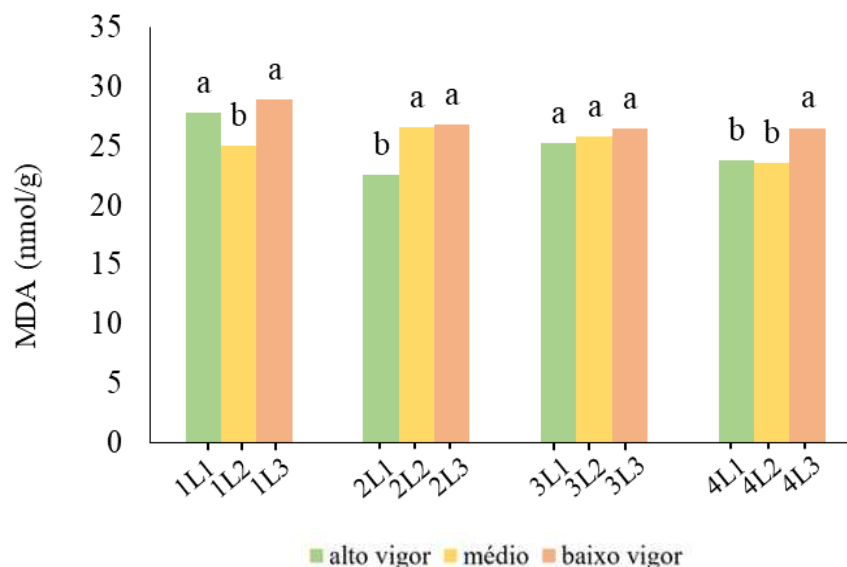


Figura 6 – Acúmulo de malonaldeído (MDA) em sementes de diferentes lotes e genótipos de soja. Médias seguidas pela mesma letra não diferem entre si pelo teste de Tukey a 5% de probabilidade, para cada genótipo.

Para os genótipos 2 e 4, os lotes de menor vigor apresentaram maiores médias no conteúdo de MDA. Para o genótipo 3 não houve diferença entre os lotes, enquanto para o genótipo 1, menor conteúdo de MDA foi obtido para o lote com qualidade intermediária. Assim, observa-se, em geral, que a baixa qualidade das sementes está associada com o acúmulo de MDA. Espécies reativas de oxigênio (EROs) são moléculas altamente reativas e tóxicas, com o potencial para danificar membranas celulares, ácidos nucleicos, proteínas, carboidratos e lipídios. A peroxidação de lipídios atua com grande impacto na estrutura e danos na membrana e pode interagir com as células para reduzir, ou mesmo eliminar suas funções, atuando com grande impacto na estrutura de membranas (RATAJCZAK et al., 2019; EBONE et al., 2019). O MDA é um dos principais biomarcadores do dano oxidativo por ser o principal subproduto liberado durante o processo de peroxidação lipídica (MIN et al., 2017). Por isso, muitos estudos têm como objetivo verificar o acúmulo de MDA como uma das consequências da deterioração de sementes, fortemente associado à diminuição da viabilidade e vigor.

As espécies reativas de oxigênio são removidas ou desintoxicadas por uma série de enzimas antioxidantes e o processo de remoção ocorre constantemente nas células para evitar alguns dos potenciais efeitos tóxicos causados pelas EROs (MITTLER, 2017). Os mecanismos antioxidantes têm a função de manter as EROs em níveis basais e não tóxicos e qualquer desequilíbrio no sistema pode ser catalisador de reações que estimulam o estresse oxidativo e,

consequentemente, reduzem a qualidade das sementes. No presente trabalho, foram observados comportamentos distintos na atividade das enzimas antioxidantes para cada genótipo e lote (Figura 7).

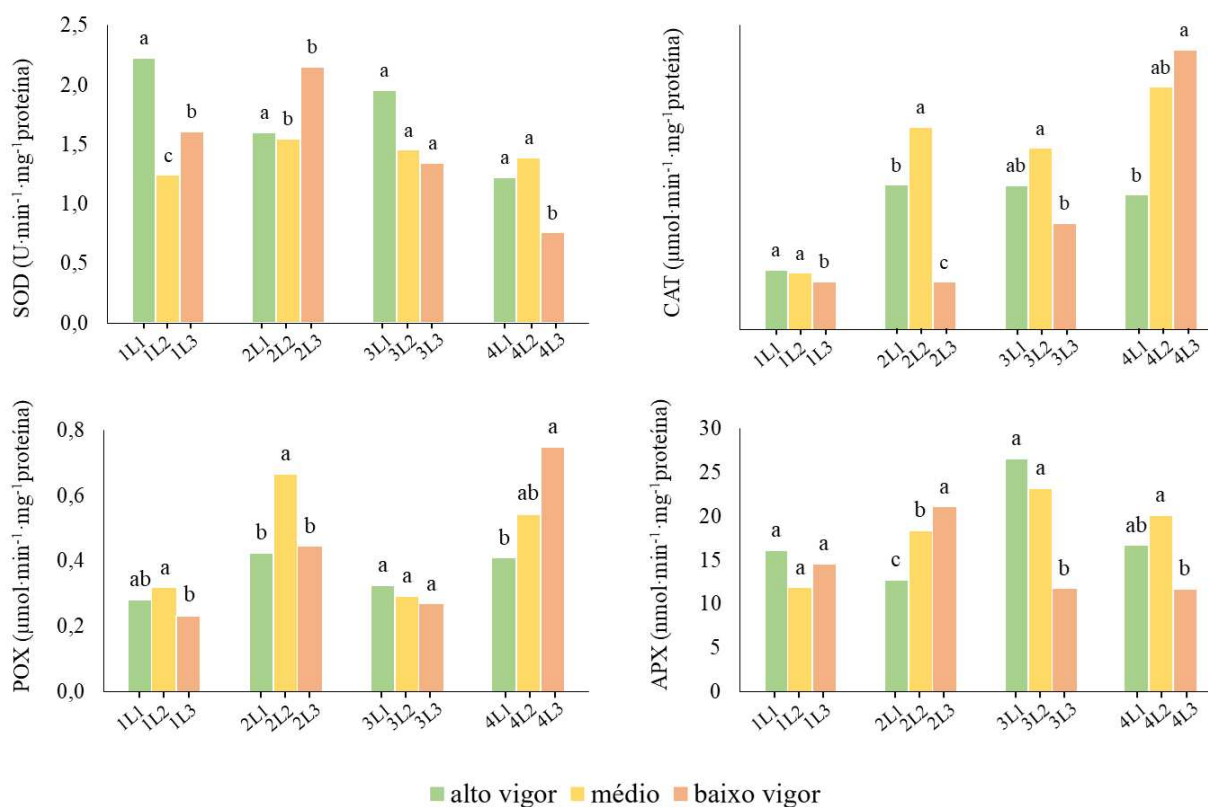


Figura 7 – Atividade das enzimas superóxido dismutase (SOD), catalase (CAT), peroxidase do ascorbato (APX) e peroxidases (POX) em diferentes genótipos e níveis de vigor de sementes de soja. Médias seguidas pela mesma letra não diferem entre si pelo teste de Tukey a 5% de probabilidade.

O lote de alto vigor do genótipo 1 apresentou maior atividade das enzimas SOD e CAT em relação ao lote de baixo vigor. Para o genótipo 2 foi observada maior atividade da SOD e APX para o lote de baixo vigor e alta atividade das enzimas CAT e POX para o lote com vigor intermediário. Para o genótipo 3 houve pouca variação estatística entre as medias das enzimas, com menor atividade da APX para o lote de baixo vigor. Apesar da pouca variação estatística entre os lotes desse genótipo, observa-se tendência de menor da atividade das enzimas no lote de menor nível de vigor. O genótipo 4 teve menor atividade da SOD para o lote de baixo vigor, mas maiores atividades da CAT e POX para o lote de baixo vigor em relação ao lote de alto vigor.

A redução da atividade de enzimas antioxidantes tem sido associada à redução do potencial fisiológico das sementes durante o processo de deterioração (PINHEIRO et al., 2023).

A menor atividade da SOD para os lotes de baixo vigor para genótipos 1, 3 e 4 corroboram com a classificação da qualidade dos lotes. Reduções na atividade da SOD são relatadas em sementes e plantas de soja submetidas a estresses salino, deterioração por umidade e toxidez por metais pesados (HASANUZZAMAN, et al., 2022; PINHEIRO, et al, 2023; MAO et al, 2018).

No presente trabalho, a propensão de menor atividade da SOD para lotes de menor vigor confirmam a peroxidação de lipídios como um dos fatores responsáveis pela redução na viabilidade das sementes, conforme os dados obtidos para o conteúdo de MDA (Figura 8). O genótipo 2 apresentou maior atividade da SOD para o lote de baixa qualidade. Neste caso, a maior atividade enzimática significa maior acúmulo de EROs nas sementes dos lotes de menor qualidade, mas níveis ainda pouco avançados de deterioração, já que todos os lotes desse genótipo apresentaram elevados potenciais de germinação e vigor (Tabela 1). A tendência é que com o avanço do processo de deterioração a qualidade das sementes reduza, associada com a redução na atividade enzimática.

A principal função da SOD é dismutar o ânion superóxido ($O_2^{\cdot-}$) em peróxido de hidrogênio (H_2O_2), que é menos reativo, porém tóxico quando presente em altas concentrações. Neste sentido, enzimas como a CAT, APX e POX fazem parte de um grupo de peroxidases que neutralizam o excesso de H_2O_2 por diferentes rotas (DAS; ROYCHOUDHURY, 2014).

s enzimas CAT e POX apresentadas nos lotes de baixo vigor do genótipo 4, e da APX no genótipo 2 sugerem que as sementes lidaram com o dano oxidativo aumentando a eliminação enzimática de H_2O_2 . Maior atividade da CAT já foi relatada em sementes de soja de baixo vigor infectadas por fungos dos gêneros *Diaporthe* e *Phomopsis* (PIZÁ et al., 2018). Em um trabalho comparando duas cultivares de soja que sofreram danos por percevejo, a atividade da POX também foi maior em um dos materiais, demonstrando efeito genético na atividade dessas enzimas (SABLJIC et al., 2020).

Os danos causados pelas espécies reativas de oxigênio (EROs) são extensos e os alvos incluem compostos como lipídios, proteínas e DNA. Essas moléculas danificam a integridade das membranas celulares e, em caso de excessos, levam à morte da semente. No entanto, as EROS desempenham papel duplo no metabolismo das sementes: são agentes deletérios, mas também acionam componentes de sinalização de estresse para evitar mais danos (DAS; ROYCHOUDHURY, 2014). Assim, o equilíbrio entre a atuação dos mecanismos antioxidantes e a concentração intracelular de EROs é importante para o equilíbrio redox da célula (MARCOS-FILHO, 2015; EBONE et al., 2019). Quando há aumento na concentração de EROs, em geral, observa-se aumento na atividade das enzimas antioxidantes, porém em níveis

elevados, as EROs levam à redução da atividade enzimática e, por consequência, levando ao processo de deterioração das sementes (EBONE et al., 2020; PINHEIRO et al., 2023).

Na figura 8 estão apresentados os dados de teores de proteína e óleo dos 12 lotes de sementes de soja avaliados. É possível observar que as proteínas solúveis (Figura 8A) foram as mais propensas a alterações em função do nível de qualidade das sementes. Os teores de proteínas totais demonstraram menor variação dentro de cada genótipo (Figura 8B).

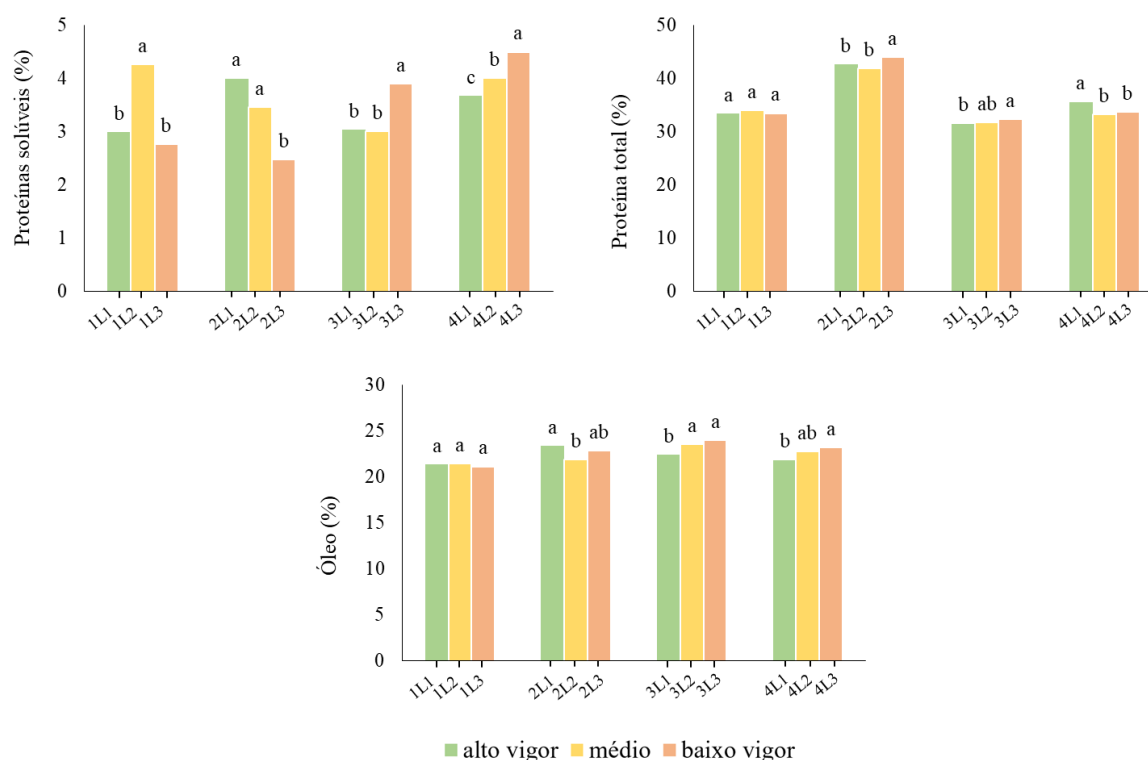


Figura 8 – Proteínas solúveis quantificadas pelo método de Bradford (A), porcentagem de proteína total (B) e teor de óleo (C) obtidos por espectroscopia no infravermelho (NIR) em diferentes genótipos e lotes de sementes de soja. Médias seguidas pela mesma letra não diferem entre si pelo teste de Tukey a 5% de probabilidade, para cada genótipo.

O conteúdo de proteínas solúveis mais baixo para os lotes de baixo vigor foi observado nas sementes dos genótipos 1 e 2. Os genótipos 3 e 4 apresentaram sementes com maior conteúdo de proteínas solúveis para os lotes de baixo vigor, o que pode estar relacionado com maior atividade enzimática observada para esses lotes (Figura 7). Os teores de óleo e proteína das sementes de soja são dominados por fatores genéticos, mas fortemente influenciados pelo ambiente, principalmente no período de enchimento dos grãos (ÁVILA et al., 2007).

O teor de óleo (Figura 8C) demonstrou variação entre os lotes para todos os genótipos, com exceção do genótipo 1. Para os genótipos 3 e 4, os lotes de baixo vigor apresentaram teores de óleo mais elevados. A oxidação de moléculas celulares como lipídios, membranas celulares,

DNA, RNA e proteínas, mediada por espécies reativas de oxigênio (EROS), afeta a qualidade e a longevidade das sementes (SANO et al., 2015).

Pela análise de componentes principais (PCA) (Figura 9), o diagrama de ordenação central (Figura 9A) apresenta a disposição dos 12 lotes de acordo com os respectivos grupos de qualidade. No círculo de correlações (Figura 9B), as variáveis relacionadas às características fisiológicas e bioquímicas das sementes evidenciam correlações negativas entre o conteúdo de MDA e condutividade elétrica (CE) com o teste de germinação e demais indicadores da qualidade fisiológica das sementes, o que confirma os resultados observados nas análises fisiológicas e bioquímicas das sementes.

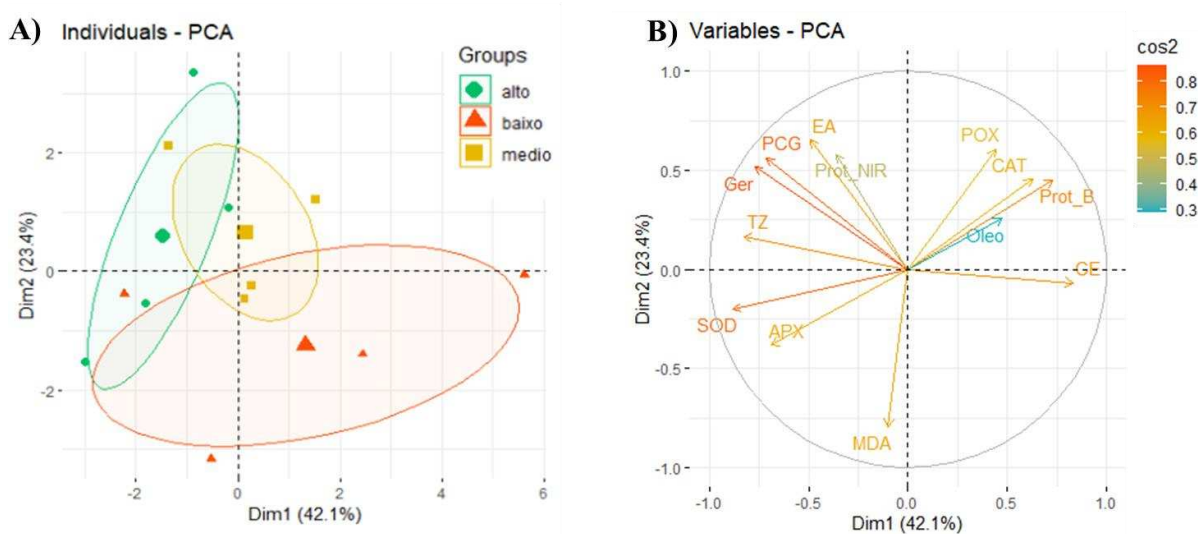


Figura 9. Análise de componentes principais (PCA) pela combinação das variáveis fisiológicas e bioquímicas dos 12 lotes de sementes de soja classificados quanto ao vigor. Diagrama de ordem (A) e círculo de correlação (B).

Os dois primeiros componentes explicaram 66% da variação total dos dados e nota-se separação nítida entre os grupos de qualidade, com agrupamento dos lotes de baixa qualidade dissociado dos outros lotes e poucos pontos de interseção entre os lotes de alta e qualidade intermediária. Os lotes localizados nos *scores* negativos do PC1 apresentaram maiores valores para o teste de germinação (G), primeira contagem de germinação (PCG), envelhecimento acelerado (EA), vigor pelo teste de tetrazólio (TZ) e menores valores de condutividade elétrica (CE) e conteúdo de malonaldeído (MDA).

A perda da qualidade fisiológica das sementes de soja é muitas vezes atribuída ao alto teor de óleo, principalmente associado à peroxidação de lipídios (SINGH et al., 2017; BEWLEY et al, 2013). Para os lotes estudados, foi observada pouca variação do teor de óleo, e a PCA mostrou menor contribuição dessa variável na variabilidade total dos dados.

Estes resultados corroboram com outros estudos que afirmam que a composição bioquímica da semente exerce influência sobre o processo de deterioração e afeta a qualidade e longevidade, assim como genótipos diferentes apresentam variação na composição e reflete na qualidade fisiológica e grau de tolerância a deterioração (SINGH et al, 2017; NAIK et al, 2019).

A Figura 10A mostra o total dos 1.200 espectros NIR brutos dos 12 lotes de sementes de soja, registrados na faixa entre 1.000 e 2.500 nm. A média dos dados originais obtidos pelos espectros NIR permitiu evidenciar uma sutil diferença entre os níveis de qualidade fisiológica das sementes (Figura 11B).

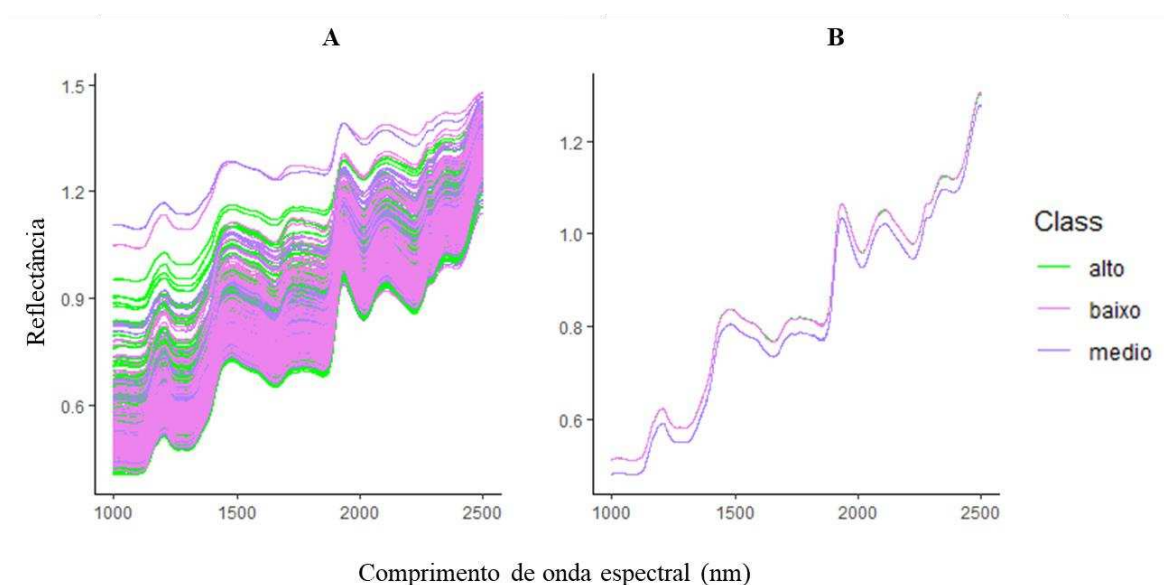


Figura 10. Espectros originais (A) e média dos espectros originais (B) das sementes por classe de vigor.

Observa-se nos espectros totais uma grande variação da linha de base ao longo de toda a faixa de leitura, que pode ter sido provocada pela desuniformidade das amostras, uma vez que foram utilizadas sementes inteiras e não um pó uniforme como é comumente utilizado na indústria para análise da qualidade de farinhas (ZHANG et al., 2022). A leitura também pode ter sido influenciada pela diferença de tamanho das sementes entre os genótipos, assim como a coloração do tegumento e hilo e até mesmo poeira na superfície da semente.

Algumas transformações quimiométricas nos dados espectrais têm sido úteis para reduzir o impacto dos ruídos das amostras e do equipamento. 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 (RINNAN et al., 2009).

Verificou-se, portanto, a necessidade de testar diferentes métodos de pré-processamento dos dados. Foram testados métodos de correção da dispersão, como *Multiplicative Scatter Correction* (MSC) ou *Standard Normal Variate* (SNV) e métodos derivativos (Savitzky-Golay) que reduzem a sobreposição de pico enquanto suaviza o sinal (AGELET; HURBURGH, 2014). 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* dos modelos obtidos para cada tipo de pré-processamento.

A acurácia é frequentemente utilizada para avaliação de problemas de classificação de aprendizado de máquina. Ela 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 (MANNING; RAGHAVAN; SCHÜTZE, 2009). 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. *Kappa* igual a 1 indica concordância perfeita, enquanto *kappa* igual a 0 indica concordância equivalente ao acaso (VIEIRA; GARRETT, 2005). Na Tabela 2 encontram-se os valores de acurácia e *kappa* obtidos para os modelos PLS-DA, obtidos a partir dos espectros NIR em diferentes estratégias de pré-processamento.

Tabela 2 – Resultados de acurácia e *kappa* para treinamento e validação (teste) dos diferentes métodos de pré-processamento obtidos pelos modelos de classificação via PLS-DA para lotes de sementes de soja com diferentes níveis de qualidade fisiológica.

Pré-processamento	Calibração (n=840)		Validação (n=360)	
	Acurácia	<i>Kappa</i>	Acurácia	<i>Kappa</i>
Espectros originais	0,5524	0,3286	0,5250	0,2875
SNV	0,5738	0,3607	0,4972	0,2458
MSC	0,5655	0,3482	0,4944	0,2417
1ª Derivada de SG	0,9357	0,9036	0,9194	0,8792
2ª Derivada de SG	0,9988	0,9928	0,9917	0,9875
Combinações				
MSC + 1ª Derivada de SG	0,9476	0,9214	0,9222	0,8833
MSC + 2ª Derivada de SG	0,9964	0,9946	0,9778	0,9667
SNV + 1ª Derivada de SG	0,9476	0,9214	0,9222	0,8833
SNV+ 2ª Derivada de SG	1,0000	1,0000	0,9917	0,9875

n = número de espectros usados para treinamento e/ou teste do modelo de classificação. SNV = *Standard Normal Variate* (variação normal padrão); MSC = *Multiplicative Scatter Correction* (correção do espalhamento multiplicativo); SG = Savitzky-Golay.

Em geral, os métodos de pré-processamento aplicados aos dados NIR melhoraram significativamente a capacidade discriminatória dos modelos em relação ao modelo construído com os dados brutos. Os melhores índices de classificação foram obtidos para os modelos construídos com a primeira e segunda derivada de Savitzky-Golay, e a combinação das derivadas com os métodos MSC e SNV.

Enquanto os espectros originais permitiram a classificação com acurácia de 53% e *kappa* de 29% para os dados de validação, a 2ª derivada de Savitzky-Golay e a combinação SNV + 2ª derivada de Savitzky-Golay atingiram *kappa* de 99% para ambos os métodos. 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 modelo teve boa capacidade de discriminar classes de vigor de sementes de soja.

Os resultados apresentados para os modelos de classificação PLS-DA obtidos com os espectros foram consistentes com outros estudos envolvendo a classificação de sementes por técnicas de espectroscopia. Medeiros et al. (2022) discriminaram espécies de Brássicas e forneceram uma estimativa do teor de óleo das sementes com 100% no desempenho do modelo PLS-DA com dados de espectrômetro NIR. A técnica também já foi utilizada para classificar sementes de grão-de-bico com alterações causadas por herbicidas e obteve 94% de acurácia quando os espectros foram suavizados pela primeira derivada do filtro Savitzky-Golay (RIBEIRO et al., 2021).

Para classificação da qualidade de sementes quanto ao potencial fisiológico, a suavização dos espectros pelas derivadas de Savitzky-Golay tem demonstrado alta capacidade em colaborar para o melhor desempenho do modelo de classificação. Silva (2020) conseguiu obter valores de acurácia muito próximos de 1 para separar sementes de soja de alto vigor, vigor intermediário e baixo vigor que foram envelhecidas artificialmente ou naturalmente. Além disso, lotes de sementes com diferentes níveis de vigor de lentilha e milho-doce foram classificados com 99% e 91% de acurácia, respectivamente, utilizando a primeira derivada do filtro Savitzky-Golay (LIMÃO, 2022; QUEIROZ, 2022).

Além da acurácia e *kappa*, a matriz de confusão é uma ferramenta muito utilizada para avaliações do desempenho de modelos de classificação por aprendizado de máquina. Ela exibe a distribuição das sementes em termos de suas classes atuais e de suas classes previstas e indica a qualidade e eficiência do modelo (Figura 13). A matriz de confusão dos dados após o pré-processamento pela segunda derivada de Savitzky-Golay indica que os erros foram principalmente associados ao vigor intermediário, tanto para os dados de calibração quanto para validação. Nos dados de treinamento e teste, apenas uma das amostras de alta qualidade foi

classificada com vigor médio; e nos dados de teste, duas amostras de vigor médio foram classificadas como de alto vigor. Assim, não foi observado erro na classificação dos lotes de menor qualidade.

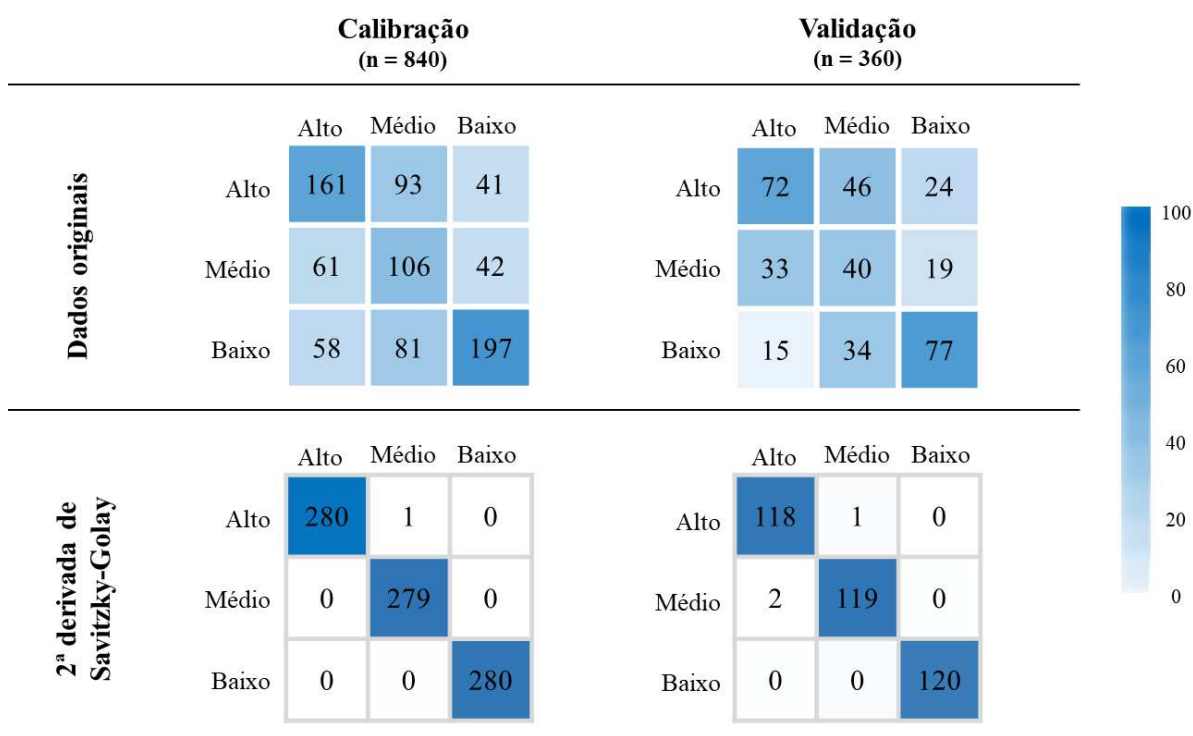


Figura 11. Matriz de confusão da classificação dos lotes de sementes de soja em níveis de vigor por meio do modelo PLS-DA utilizando-se os espectros originais e após o pré-processamento com a segunda derivada de Savitzky-Golay.

Os comprimentos de onda que mais contribuíram para a classificação das sementes em diferentes níveis de qualidade, determinados com base na importância das variáveis para separação das classes de vigor, estão descritos na Figura 14. Os picos de importância nas faixas de comprimento de onda correspondem a grupos funcionais que contribuíram para a classificação qualitativa dos lotes quanto à qualidade (ZHANG et al., 2020). Percebe-se que a faixa de comprimento de onda entre 1.000-1.250 nm e alguns pontos entre 2.200-2.500 nm foram os mais importantes para a classificação dos lotes de sementes de soja em alto vigor, vigor intermediário e baixo vigor. Destaca-se que essas faixas foram coincidentes com as obtidas no Experimento I (Figura 4), em que foram utilizados maior número de lotes. Assim, essas regiões estão relacionadas com compostos que sofreram alterações com o processo de deterioração das sementes, como os lipídios de membranas, óleo e proteínas.

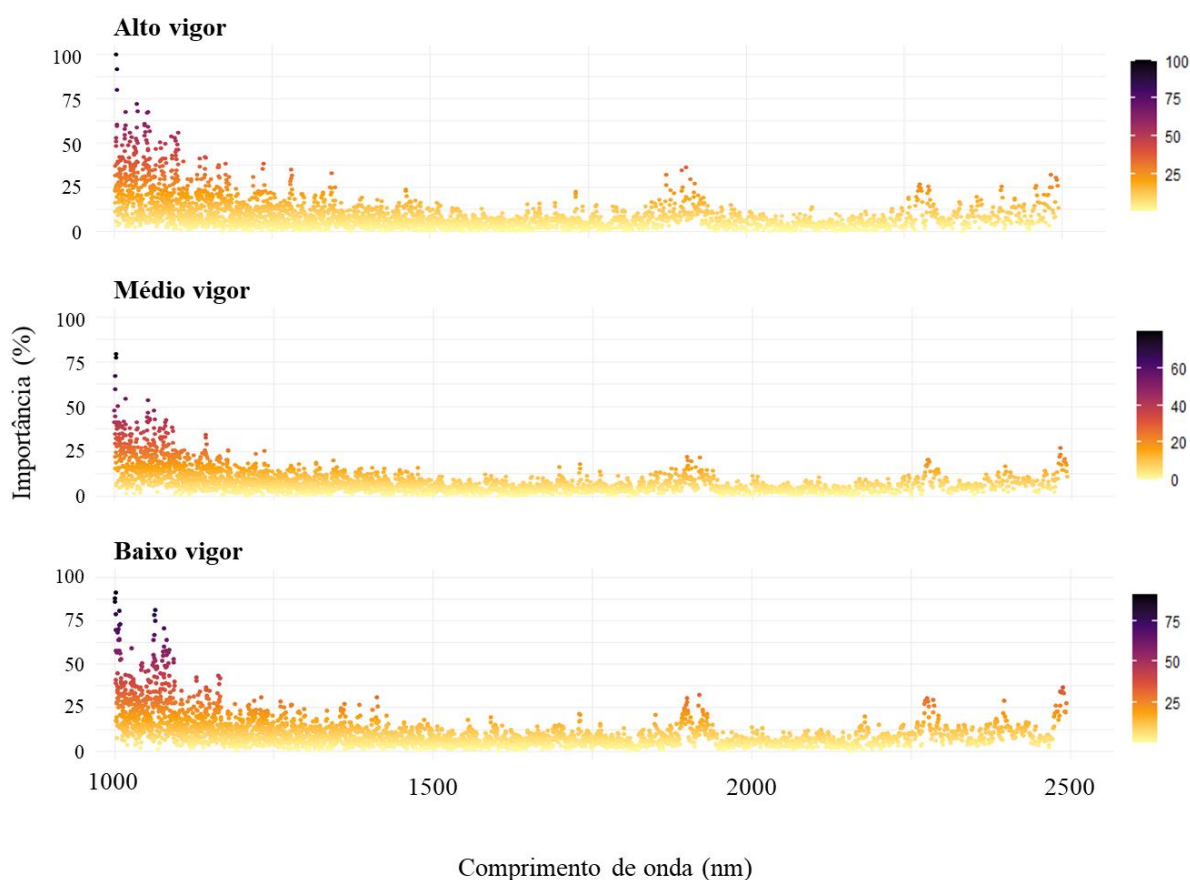


Figura 12. Importância das variáveis de comprimento de ondas para classificação via PLS-DA dos níveis de vigor de sementes de soja.

As bandas de absorção na faixa NIR estão associadas a grupos funcionais relacionados a água, proteínas, óleo e carboidratos, geralmente são amplas e se sobrepõem em várias partes da faixa espectral. Por exemplo, comprimentos de onda mais importantes com pico nas faixas 1.060-1.195 nm são associados à carboidrato, 1.100-1.185 nm à estrutura de proteínas e 930-1.390 nm ao teor de óleo (SHENK et al., 2007; AL-AMERY et al., 2018; MEDEIROS et al., 2022).

Apesar da importância menos expressiva, a faixa em torno de 2.250 nm apresentou comportamento similar para as três classes de vigor. Picos nos comprimentos de onda 2.347 nm são associados à absorbância de lipídios, 2.282 nm e 2.330 nm a carboidratos, 2.300 nm associada a banda de absorbância das proteínas, 2.340 ao grupo CH da celulose, 2.306 nm relacionado à proteína e ao óleo (XU et al., 2019). Picos de absorção em 1.002 nm, 1.920 nm e 2.282 nm também são associados com as características das moléculas de açúcares (OZAKI et al., 2006; SIMEONE et al., 2015).

As proteínas são fundamentais para o crescimento do eixo embrionário, formação da plântula e emergência em campo (ERBAŞ et al., 2016). Durante o processo de deterioração das sementes, há desnaturação e decréscimo do conteúdo e da síntese de proteínas, sendo a degradação de proteínas um dos mecanismos apontados como a causa da perda de viabilidade das sementes (MARCOS-FILHO, 2020; MCDONALD, 1999). Esse mecanismo de deterioração foi evidenciado nas sementes dos genótipos 1 e 2, em que os lotes de menor qualidade apresentaram redução no conteúdo de proteínas solúveis (Figura 9).

Ao contrário das proteínas, o teor de óleo se relaciona negativamente com a qualidade fisiológica das sementes de soja. A associação se dá principalmente pela deterioração por consequência da peroxidação de lipídios (SINGH et al., 2017; BEWLEY et al., 2013).

Influenciada pelo genótipo e ambiente, sabe-se que a composição bioquímica apresenta relação com a qualidade fisiológica e longevidade das sementes. Os efeitos deletérios do consumo de reservas também são evidentes e, embora nem sempre provoquem a morte, impactam negativamente no vigor.

As técnicas convencionais de análise bioquímicas são caras, demoradas e a seleção de comprimentos de onda importantes, relacionados às informações químicas de interesse, pode reduzir o tempo de processamento dos dados e fornecer modelos mais representativos e viáveis para aplicações industriais (ORRILLO et al., 2019).

Embora haja dificuldade na quantificação de compostos específicos em sementes pelo NIR devido a sobreposição de bandas espectrais associadas a mais de um composto, o modelo PLS-DA após pré-processamento derivativo dos espectros proposto em ambos os experimentos foi capaz de classificar lotes de sementes de soja pelo potencial fisiológico. Com base na resposta vibracional das ligações químicas à radiação NIR, foi possível identificar regiões do espectro eletromagnético que apresentam maior importância para distinguir os níveis de vigor de sementes de soja.

Esse trabalho demonstra que equipamentos a base de NIR podem ser calibrados para predição da qualidade fisiológica de sementes de soja e, embora essa metodologia não substitua os testes padrões utilizados para o comércio, a avaliação de lotes de sementes pela espectroscopia no infravermelho próximo pode ser utilizada para uma rápida pré-seleção e descarte de lotes de baixa qualidade. Além disso, com exceção do investimento com a aquisição do equipamento, o custo para coleta dos dados é mínimo. Incluir essa etapa no processo de produção pode reduzir o volume de avaliações dos testes de germinação e vigor em uma análise rápida, não invasiva e de alto rendimento, tendo como consequência a economia de tempo, espaço e recursos.

CONCLUSÕES

Modelos PLS-DA obtidos a partir de espectros NIR de amostras de sementes de soja, após o pré-processamento por método derivativo, permitem obter informações para discriminar as amostras com base nas diferenças em seus componentes bioquímicos e são uma alternativa para classificação de lotes de sementes em níveis de qualidade.

A espectroscopia no NIR combinada com quimiometria são ferramentas valiosas para triagem da qualidade de sementes de soja e fornecem informações que podem ser relacionadas à composição bioquímica das sementes.

REFERÊNCIAS BIBLIOGRÁFICAS

- AGELET, L. E.; AMSTRONG, P. R.; CLARIANA, I. R.; HURBURGH, C. R. Measurement of Single Soybean Seed Attributes by Near-Infrared Technologies. A Comparative Study. **Journal of Agricultural and Food Chemistry**, v. 60, n. 34, p. 8314–8322, 14 ago. 2012. <https://doi.org/10.1021/jf3012807>
- AGELET, L. E.; HURBURGH, C. R. Limitations and current applications of Near Infrared Spectroscopy for single seed analysis. **Talanta**, v. 121, p. 288–299, abr. 2014. <https://doi.org/10.1016/j.talanta.2013.12.038>
- AL-AMERY, M.; GENEVE, R. L.; SANCHES, M. F.; ARMSTRONG, P. R.; MAGHIRANG, E. B.; LEE, C.; VIEIRA, R. D.; HILDEBRAND, D. F. Near-infrared spectroscopy used to predict soybean seed germination and vigour. **Seed Science Research**, v. 28, p. 245-252, 2018.
- AMANA, H. Z. et al. Nondestructive measurement of anthocyanin in intact soybean seed using Fourier Transform Near-Infrared (FT-NIR) and Fourier Transform Infrared (FT-IR) spectroscopy. **Infrared Physics & Technology**, v. 111, p. 103477, dez. 2020. <https://doi.org/10.1016/j.infrared.2020.103477>
- ANDERSEN, C. M.; BRO, R. Variable selection in regression-a tutorial. **Journal of Chemometrics**, v. 24, n. 11-12, p. 728–737, nov. 2010. <https://doi.org/10.1002/cem.1360>
- ANDERSON, M. D.; PRASAD, T. K.; STEWART, C. R. Changes in Isozyme Profiles of Catalase, Peroxidase, and Glutathione Reductase during Acclimation to Chilling in Mesocotyls of Maize Seedlings. **Plant physiology**, v. 109, n. 4, p. 1247–1257, 1 dez. 1995.
- ANDRADE, G. C.; COELHO, C. M.M.; UARROTA, V. G. Modelling the vigour of maize seeds submitted to artificial accelerated ageing based on ATR-FTIR data and chemometric

- tools (PCA, HCA and PLS-DA). **Heliyon**, v. 6, n. 2, p. e03477, fev. 2020.
<https://doi.org/10.1016/j.heliyon.2020.e03477>
- ARMSTRONG, P. R. Rapid single-kernel NIR measurement of grain and oil-seed attributes. **Applied Engineering in Agriculture**, v. 22, p. 767-772, 2006.
<https://doi.org/10.13031/2013.21991>
- ÁVILA, M. R. et al. Componentes do rendimento, teores de isoflavonas, proteínas, óleo e qualidade de sementes de soja. **Revista Brasileira de Sementes**, v. 29, p. 111–127, 2007.
<https://doi.org/10.1590/S0101-31222007000300014>
- BAGCHI, T. B.; SHARMA, S.; CHATTOPADHYAY, K. Development of NIRS models to predict protein and amylose content of brown rice and proximate compositions of rice bran. **Food Chemistry**. Vol 191, 21-27, 2016. <https://doi.org/10.1016/j.foodchem.2015.05.038>
- BAIANU, I.; GUO, J. NIR Calibrations for Soybean Seeds and Soy Food Composition Analysis: Total Carbohydrates, Oil, Proteins and Water Contents. **Nature Precedings**, 16 nov. 2011.
- BAILLY, C.; EL-MAAROUF-BOUTEAU, H.; CORBINEAU, F. From intracellular signaling networks to cell death: the dual role of reactive oxygen species in seed physiology. **Comptes Rendus Biologies**, v.331, p.806–814, 2008.
- BALABIN, R. M.; SMIRNOV, S. V. Melamine detection by mid- and near-infrared (MIR/NIR) spectroscopy: A quick and sensitive method for dairy products analysis including liquid milk, infant formula, and milk powder. **Talanta**, v. 85, n. 1, p. 562–568, jul. 2011. <https://doi.org/10.1016/j.talanta.2011.04.026>
- BARKER, M.; RAYENS, W. Partial mínimos quadrados para discriminação. *Journal of Chemometrics*, v. 17, p. 166-173, 2003.
- BAZONI, C. H. V; 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, 1 set. 2017.
<https://doi.org/10.1016/j.jspr.2017.05.003>
- BERRAR, D. Bayes' Theorem and Nayve Bayes Classifier. *In*: RANGANATHAN, S; GRIBSKOV, M; NAKAI, K; SCHÖNBACH. **Encyclopedia of Bioinformatics and Computational Biology: ABC of Bioinformatics**. Tokio, Japão, Elsevier, 2019. p.403-412.
- BEWLEY, J. D.; BRADFORD, K. J.; HILHORST, H. W. M.; NONOGAKI, H. **Seeds: physiology of development, germination and dormancy**. 3° ed., New York: Springer. 2013, 407 p.
- BIANCHINI, V. M.; MASCARIN, G.M; SILVA, L.C.A.S; ARTHUR. V.; CARSTENSEN, J.M.; BOELT, B.; SILVA, C. B. Multispectral and X-ray images for characterization of *Jatropha curcas* L. seed quality. **Plant Methods** 17, 9, 2021.
<https://doi.org/10.1186/s13007-021-00709-6>

- BOWLER, C.; VAN MONTAGU, M.; INZÉ, D. Superoxide dismutase and stress tolerance. **Annual Review of Plant Physiology and Plant Molecular Biology**, v.43, p.83–116, 1992.
- BRADFORD, M. M. A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. **Analytical Biochemistry**, v. 72, n. 1–2, p. 248–254, 7 maio 1976.
- BRASIL. **Regras para análise de sementes**. Brasília, DF: MAPA/ACS, 2009.
- BRASIL. Ministério da Agricultura, Pecuária e Abastecimento. Padrões para produção e comercialização de sementes de soja (*Glycine max* (L.) Merrill). Instrução Normativa nº 45, de 17 de setembro de 2013. Diário Oficial da União, Brasília, DF, 20 set. 2013. Seção 1, p. 29.
- BREIMAN, L. Random Forests. **Machine Learning**, v. 45, n. 1, p. 5–32, 2001.
- CAKMAK, I.; HORST, W. J. Effect of aluminium on lipid peroxidation, superoxide dismutase, catalase, and peroxidase activities in root tips of soybean (*Glycine max*). **Physiologia Plantarum**, v. 83, n. 3, p. 463–468, 1 nov. 1991.
- CASTRO, W. et al. Amazonian cacao-clone nibs discrimination using NIR spectroscopy coupled to naïve Bayes classifier and a new waveband selection approach. **Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy**, v. 270, p. 120815, abr. 2022. <https://doi.org/10.1016/j.saa.2021.120815>
- CHANCE, B.; MAEHLY, A. C. [136] Assay of catalases and peroxidases. **Methods in Enzymology**, v. 2, p. 764–775, 1 jan. 1955.
- DAS, K.; ROYCHOUDHURY, A. Reactive oxygen species (ROS) and response of antioxidants as ROS-scavengers during environmental stress in plants. **Frontiers in Environmental Science**, v. 2, 2 dez. 2014. <https://doi.org/10.3389/fenvs.2014.00053>
- DEL LONGO, O. T. et al. Antioxidant Defences under Hyperoxygenic and Hyperosmotic Conditions in Leaves of Two Lines of Maize with Differential Sensitivity to Drought. **Plant and Cell Physiology**, v. 34, n. 7, p. 1023–1028, 1 out. 1993.
- DELOUCHE, J. C. Seed germination. In: **Proceedings of the Short Course for Seedsmen**, 360. Mississippi State: Mississippi State University, Seed Technology Laboratory, 1979. p. 59–76. Disponível em: <https://scholarsjunction.msstate.edu/seedsman-short-course/360>. Acesso em: 05 fev. 2023.
- DELOUCHE, J. C.; BASKIN, C. Accelerated aging techniques for predicting the relative storability of seed lots. **Seed Science and Technology**, v. 1, p. 427–552, 1973.
- DELWICHE, S. R. Protein Content of Single Kernels of Wheat by Near-Infrared Reflectance Spectroscopy. **Journal of Cereal Science**, v. 27, n. 3, p. 241–254, maio 1998.
- DHANOA, M. S. et al. The Link between Multiplicative Scatter Correction (MSC) and Standard Normal Variate (SNV) Transformations of NIR Spectra. **Journal of Near Infrared Spectroscopy**, v. 2, n. 1, p. 43–47, jan. 1994. <https://doi.org/10.1255/jnirs.30>

- EBONE, L. A. CAVERZAN, A.; SILVEIRA, D. C.; SIQUEIRA, L. O.; LÂNGARO, N. C.; CHIOMENTO, J. L. T.; CHAVARRIA, G.. Biochemical Profile of the Soybean Seed Embryonic Axis and Its Changes during Accelerated Aging. **Biology**, v. 9, n. 8, p. 186, 23 jul. 2020. <https://doi.org/10.3390/biology9080186>
- EBONE, L. A.; CAVERZAN, A.; CHAVARRIA, G. Physiologic alterations in orthodox seeds due to deterioration processes. **Plant Physiology and Biochemistry**, v. 145, p. 34–42, 1 dez. 2019.
- ERBAŞ, S.; TONGUÇ, M.; ŞANLI, A. Mobilization of seed reserves during germination and early seedling growth of two sunflower cultivars. **Journal of Applied Botany and Food Quality**, v. 89, 20 jun. 2016. <https://doi.org/10.5073/JABFQ.2016.089.028>
- FERREIRA, D. S. et al. Comparison and application of near-infrared (NIR) and mid-infrared (MIR) spectroscopy for determination of quality parameters in soybean samples. **Food Control**, v. 35, n. 1, p. 227–232, jan. 2014. <https://doi.org/10.1016/j.foodcont.2013.07.010>
- FRANÇA-NETO, J. B.; KRZYZANOWSKI, F. C. Metodologia para o teste de tetrazólio. Londrina: **Embrapa Soja**, 2018. 109 p. Disponível em: <https://ainfo.cnptia.embrapa.br/digital/bitstream/item/193315/1/Doc-406-OL.pdf>
- FRANÇA-NETO, J.B.; KRZYZANOWSKI, F. C. Teste de tetrazólio para determinação do vigor de sementes. 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.
- FU, Y.-B.; AHMED, Z.; DIEDERICHSEN, A. Towards a better monitoring of seed ageing underex situseed conservation. **Conservation Physiology**, v. 3, n. 1, p. cov026, 2015.
- GIANNOPOLITIS, C. N.; RIES, S. K. Superoxide dismutases: I. Occurrence in higher plants. **Plant physiology**, v. 59, n. 2, p. 309–14, 1 fev. 1977
- GOEL, A.; SHEORAN. Lipid peroxidation and peroxidescavenging enzymes in cotton seeds under natural ageing. **Biologia Plantarum**, v.46, n.3, p.429-434, 2003.
- HASANUZZAMAN, M.; RAIHAN, M. R. H; NOWROZ, F; FUJITA, M. Insight into the Mechanism of Salt-Induced Oxidative Stress Tolerance in Soybean by the Application of *Bacillus subtilis*: Coordinated Actions of Osmoregulation, Ion Homeostasis, Antioxidant Defense, and Methylglyoxal Detoxification. **Antioxidants**, v. 11, n. 10, p. 1856, 1 out. 2022. <https://doi.org/10.3390/antiox11101856>
- HAVIR, E. A.; MCHALE, N. A. Biochemical and developmental characterization of multiple forms of catalase in tobacco leaves. **Plant physiology**, v. 84, n. 2, p. 450–5, 1 jun. 1987.
- HEUNG, B.; BULMER, C. E.; SCHMIDT, M. G. Predictive soil parent material mapping at a regional-scale: A Random Forest approach. **Geoderma**, v. 214-215, p. 141–154, fev. 2014. <https://doi.org/10.1016/j.geoderma.2013.09.016>
- HOPKINS, D. W. What is a Norris Derivative? **NIR news**, v. 12, n. 3, p. 3–5, jun. 2001. <https://doi.org/10.1255/nirn.611>

- ISTA. **International Rules for Seed Testing**. Bassersdorf: International Seed Testing Association, 2017. 296 p.
- JAILLAS, B.; ROUMET, P.; PINSON-GADAIS, L.; BERTRAND, D. Detection of *Fusarium* head blight contamination in wheat kernels by multivariate imaging. **Food Control**. Vol. 54, 250-258, 2015. <https://doi.org/10.1016/j.foodcont.2015.01.048>
- JONSSON, H.; GABRIELSSON, J. Comprehensive Chemometrics: 2.11 - **Evaluation of Preprocessing Methods**. S.D. Brown, R. Tauler, and B. Walczak, Eds.Elsevier, pp. 199-206, 2009.
- KADEMI, H. I.; ULUSOY, B. H.; HECER, C. Applications of miniaturized and portable near infrared spectroscopy (NIRS) for inspection and control of meat and meat products. **Food Reviews International**, v. 35, n. 3, p. 201–220, 13 set. 2018. <https://doi.org/10.1080/87559129.2018.1514624>
- KANDPAL, L. M. et al. Near-infrared hyperspectral imaging system coupled with multivariate methods to predict viability and vigor in muskmelon seeds. **Sensors and Actuators B: Chemical**, v. 229, p. 534–544, jun. 2016. <https://doi.org/10.1016/j.snb.2016.02.015>
- KAR, M.; MISHRA, D. Catalase, Peroxidase, and Polyphenoloxidase Activities during Rice Leaf Senescence. **Plant physiology**, v. 57, n. 2, p. 315–9, 1 fev. 1976.
- KOSMOWSKI, F.; WORKU, T. Evaluation of a miniaturized NIR spectrometer for cultivar identification: The case of barley, chickpea and sorghum in Ethiopia. **PLOS ONE**, v. 13, n. 3, p. e0193620, 21 mar. 2018. <https://doi.org/10.1371/journal.pone.0193620>
- KOVALENKO, I. V.; RIPPKE, G. R.; HURBURGH, C. R. Determination of Amino Acid Composition of Soybeans (*Glycine max*) by Near-Infrared Spectroscopy. **Journal of Agricultural and Food Chemistry**, v. 54, n. 10, p. 3485–3491, 20 abr. 2006. <https://doi.org/10.1021/jf052570u>
- LANDIS, J.R. e KOCH, G.G. The measurement of observer agreement for categorical data. **Biometrics**, v.33, n.1, p. 159-174, 1977.
- LARIOS, G. et al. Soybean seed vigor discrimination by using infrared spectroscopy and machine learning algorithms. **Analytical Methods**, v. 12, n. 35, p. 4303–4309, 17 set. 2020. <https://doi.org/10.1039/D0AY01238F>
- LEE, J. H.; CHOUNG, M.-G. Nondestructive determination of herbicide-resistant genetically modified soybean seeds using near-infrared reflectance spectroscopy. **Food Chemistry**, v. 126, n. 1, p. 368–373, maio 2011. <https://doi.org/10.1016/j.foodchem.2010.10.106>
- LEE, J.-D.; SHANNON, J. G.; CHOUNG, M.-G. Selection for protein content in soybean from single F2 seed by near infrared reflectance spectroscopy. **Euphytica**, v. 172, n. 1, p. 117–123, 7 nov. 2009. <https://doi.org/10.1007/s10681-009-0067-5>
- LIMÃO, M. A. R. **Testes de vigor, atividade enzimática e espectroscopia no infravermelho próximo em sementes de lentilha (*Lens culinaris* Medik.)** Dissertação (mestrado) – Departamento de Agronomia, Universidade Federal de Viçosa, 2022. Acesso em 16 de

- janeiro de 2022: <<https://www.locus.ufv.br/bitstream/123456789/29541/1/texto%20completo.pdf>>
- MAHADEVAN, S.; SHAH, S. L. Fault detection and diagnosis in process data using one-class support vector machines. **Journal of Process Control**, v. 19, n. 10, p. 1627–1639, dez. 2009. <https://doi.org/10.1016/j.jprocont.2009.07.011>
- MAO, F. et al. The metal distribution and the change of physiological and biochemical process in soybean and mung bean plants under heavy metal stress. **International Journal of Phytoremediation**, v. 20, n. 11, p. 1113–1120, 29 ago. 2018. <https://doi.org/10.1080/15226514.2017.1365346>
- MANNING, C. D.; RAGHAVAN, P.; SCHÜTZE, H. **Introduction to information retrieval**. Cambridge: Cambridge University Press, Online edition 2009. 544 p. Disponível em: <<https://nlp.stanford.edu/IR-book/pdf/irbookonlinereading.pdf>>. Acesso em: 29 de jan. 2023.
- MARCOS-FILHO, J. **Fisiologia de sementes de plantas cultivadas**. Londrina: ABRATES, 2015. 660 p.
- MARCOS-FILHO, J. Teste de envelhecimento acelerado. 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.
- MCDONALD, M.B. Seed deterioration: physiology, repair and assessment. **Seed Science and Technology**, v.27, p.177-237, 1999.
- MEDEIROS, A. D. DE et al. Machine Learning for Seed Quality Classification: An Advanced Approach Using Merger Data from FT-NIR Spectroscopy and X-ray Imaging. **Sensors**, v. 20, n. 15, p. 4319, 1 jan. 2020. <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>
- MIN, C. W. et al. In-depth proteomic analysis of Glycine max seeds during controlled deterioration treatment reveals a shift in seed metabolism. **Journal of Proteomics**, v. 169, p. 125–135, 3 out. 2017. <https://doi.org/10.1016/j.jprot.2017.06.022>
- MITTLER, R. Oxidative stress, antioxidants and stress tolerance. **Trends in Plant Science**, v.7, p.405–410, 2002.
- MITTLER, R. ROS Are Good. **Trends in Plant Science**, v. 22, n. 1, p. 11–19, jan. 2017. <https://doi.org/10.1016/j.tplants.2016.08.002>
- 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**, 11, 301, 2021. <https://doi.org/10.3390/agriculture11040301>

- NAIK, S. M.; MADHUSUDAN, K.; MOTAGI, B. M.; MUGALI, S.; NADAF, H. L. Molecular characterization of seed longevity and associated characters using SSR markers in soybean [*Glycine max* (L.) Merrill]. **Journal of Pharmacognosy and Phytochemistry**, v. 8, p. 2357-2360, 2019.
- NAKANO, Y.; ASADA, K. Hydrogen Peroxide is Scavenged by Ascorbate-specific Peroxidase in Spinach Chloroplasts. **Plant and Cell Physiology**, v. 22, n. 5, p. 867–880, 1 ago. 1981.
- OLESEN, M.H.; NIKNESHAN, P.; SHRESTHA, S.; TADAYYON, A.; DELEURAN, L.C.; BOELT, B.; GISLUM, R. Viability Prediction of *Ricinus communis* L. Seeds Using Multispectral Imaging. **Sensors**, 15, 4592-4604, 2015. <https://doi.org/10.3390/s150204592>
- ORRILLO, I. CRUZ-TIRADO, J.P.; CARDENAS, A; ORUNA, M; CARNERO, A; BARBIN, D. F; SICHE, R. Hyperspectral imaging as a powerful tool for identification of papaya seeds in black pepper. **Food Control**, v. 101, p. 45–52, jul. 2019. <https://doi.org/10.1016/j.foodcont.2019.02.036>
- OSBORNE, B. G. Near-Infrared Spectroscopy in Food Analysis. **Encyclopedia of Analytical Chemistry**, p. 1-14, 2006. <https://doi.org/10.1071/ASEG2006ab128>
- OZAKI, Y.; MCCLURE, W. F.; CHRISTY, A. A. **Near-infrared spectroscopy in food science and technology**. New Jersey: Wiley-Interscience, 2006. 406 p.
- OZKUR, O; OZDEMIR, F; BOR, M; TURKAN, I. Physiochemical and antioxidant responses of the perennial xerophyte *Capparis ovata* Desf. to drought. **Environmental and Experimental Botany**, v. 66, n. 3, p. 487–492, set. 2009. <https://doi.org/10.1016/j.envexpbot.2009.04.003>
- RATAJCZAK, E; MALECKA, A; CIERESZKO, I; STASZAK, A, M. Mitochondria Are Important Determinants of the Aging of Seeds. **International Journal of Molecular Sciences**, v. 20, n. 7, p. 1568, 2019. <https://doi.org/10.3390/ijms20071568>
- PEIXOTO, P. H. P., CAMBRAIA, J., SANT'ANNA, 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**, v. 11, n. 3, p. 137–143, 1999.
- PINHEIRO, D. T; DIAS, D, C, F, S; SILVA, L. J; MARTINS, M, S; FINGER, F, L. Oxidative stress, protein metabolism, and physiological potential of soybean seeds under weathering deterioration in the pre-harvest phase. **Acta Scientiaum**, v. 45, 2023. <https://doi10.4025/actasciagron.v45i1.56910>
- PIZÁ, M. C. P. et al. Effects of non-thermal plasmas on seed-borne *Diaporthe/Phomopsis* complex and germination parameters of soybean seeds. **Innovative Food Science and Emerging Technologies**, v. 49, p. 82–91, 1 out. 2018. <https://doi.org/10.1016/j.ifset.2018.07.009>
- PLANS, M; SIMÓ, J; CASAÑAS, F; SABATÉ, J; SAONA, L. R. Characterization of common beans (*Phaseolus vulgaris* L.) by infrared spectroscopy: Comparison of MIR, FT-

- NIR and dispersive NIR using portable and benchtop instruments. **Food Research International**, 54, P. 1643-1651, 2013. <https://doi.org/10.1016/j.foodres.2013.09.003>
- REDDY, P.; GUTHRIDGE, K.M.; PANOZZO, J.; LUDLOW, E.J.; SPANGENBERG, G.C.; ROCHFORD, S.J. Near-Infrared Hyperspectral Imaging Pipelines for Pasture Seed Quality Evaluation: An Overview. **Sensors**, 22 (5), 2022. <https://doi.org/10.3390/s22051981>
- 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.
- RINNAN, Å.; BERG, F. VAN DEN; ENGELSEN, S. B. Review of the most common pre-processing techniques for near-infrared spectra. **TrAC Trends in Analytical Chemistry**, v. 28, n. 10, p. 1201–1222, nov. 2009. <https://doi.org/10.1016/j.trac.2009.07.007>
- SABLJIC, I. et al. Role of reactive oxygen species and isoflavonoids in soybean resistance to the attack of the southern green stink bug. **Peer J**, v. 8, p. e9956–e9956, 17 set. 2020. <https://doi.org/10.7717/peerj.9956>
- SAIRAM, R. K.; DESHMUKH, P. S.; SAXENA, D. C. Role of Antioxidant Systems in Wheat Genotypes Tolerance to Water Stress. **Biologia plantarum**, v. 41, n. 3, p. 387–394, 1 abr. 1998. <https://doi.org/10.1016/j.pla-phy.2009.11.003>
- SANO, N., RAJJOU, L., NORTH, H. M., DEBEAUJON, I., MARION-POLL, A., & SEO, M. Staying alive: Molecular aspects of seed longevity. **Plant and Cell Physiology**, 57, 4, p. 660-674, 2016. <https://doi.org/10.1093/pcp/pcv186>
- SANTANA, F. B.; NETO, W. B.; POPPI, R. J. Random forest as one-class classifier and infrared spectroscopy for food adulteration detection. **Food Chemistry**, v. 293, p. 323–332, 30 set. 2019. <https://doi.org/10.1016/j.foodchem.2019.04.073>
- SAVITZKY, ABRAHAM.; GOLAY, M. J. E. Smoothing and Differentiation of Data by Simplified Least Squares Procedures. **Analytical Chemistry**, v. 36, n. 8, p. 1627–1639, jul. 1964.
- SEO, Y.-W. et al. Non-Destructive Sorting Techniques for Viable Pepper (*Capsicum annuum* L.) Seeds Using Fourier Transform Near-Infrared and Raman Spectroscopy. **Journal of Biosystems Engineering**, v. 41, n. 1, p. 51–59, 2016. <https://doi.org/10.5307/JBE.2016.41.1.051>
- SHARMA, P; JHA, A. B; DUBEY, R. S; PESSARAKLI, M. Reactive Oxygen Species, Oxidative Damage, and Antioxidative Defense Mechanism in Plants under Stressful Conditions. *Journal of Botany*, v. 2012, p. 26, 2012. <https://doi.org/10.1155/2012/217037>
- SHRESTHA, S; DELEURAN, L. C; GISLUM, R. Classification of different tomato seed cultivars by multispectral visible-near infrared spectroscopy and chemometrics. **Journal of Spectral Imaging**, 5, 2016. <https://doi.org/10.1255/jsi.2016.a1>
- SHENK, J. S. Application of NIR spectroscopy to agricultural product. In: SHENK, J.S; WORKMAN, J.J; WESTERHAUS, M.O. **Handbook of Near-infrared Analysis**. CRC Press, 3^a ed. Boca Raton, p. 347-386, 2007.

- SILVA, M. F. **Armazenabilidade, envelhecimento e classificação do vigor de sementes de soja por espectroscopia do infravermelho próximo**. Tese (doutorado) – Departamento de Agronomia, Universidade Federal de Viçosa, outubro, 2020.
- SIMEONE, M. L. F; PARRELLA, R. A. da C; DAMASCENO, C. M. B; LEAL, M. C. B; PASQUINI, C. Quantificação de sacarose em caldo de sorgo sacarino por espectroscopia no infravermelho próximo (NIR). **Embrapa Milho e Sorgo**. Comunicado Técnico, 212. p.8, 2015. <https://www.infoteca.cnptia.embrapa.br/infoteca/handle/doc/1037783>
- SIMOVA-STOILOVA, L; VASEVA, I; GRIGOROVA, B; DEMIREVSKA, K; FELLER, U. Proteolytic activity and cysteine protease expression in wheat leaves under severe soil drought and recovery. **Plant Physiology and Biochemistry**, v. 48, n. 2, p. 200–206, 1 fev. 2010. <https://doi.org/10.1016/j.plaphy.2009.11.003>
- SINGH, J.; PAROHA, S.; MISHRA, R. P. Factors Affecting Oilseed Quality during Storage with Special Reference to Soybean (*Glycine max*) and Niger (*Guizotia abyssinica*) Seeds. **International Journal of Current Microbiology and Applied Sciences**, v. 6, n. 10, p. 2215-2226, 2017.
- SINGH, P. NAYYAR, A; SINGH, S; KAUR, A. Classification of wheat seeds using image processing and fuzzy clustered random forest. **International Journal of Agricultural Resources, Governance and Ecology**, v. 16, n. 2, p. 123, 2020. <https://doi.org/10.1504/IJARGE.2020.109048>
- CHEN, J.; JÖNSSON, P.; TAMURA, M.; GU, Z.; MATSUSHITA, B.; EKLUNDH, L. A simple method for reconstructing a high-quality NDVI time-series data set based on the Savitzky–Golay filter. **Remote Sensing of Environment**, v. 91, n. 3-4, p. 332–344, Jun. 2004 <https://doi.org/10.1016/j.rse.2004.03.014>
- TAJUDDIN, T; WATANABE, S; MASUDA, R; HARADA, K; KAWANO, S; Application of near Infrared Transmittance Spectroscopy to the Estimation of Protein and Lipid Contents in Single Seeds of Soybean Recombinant Inbred Lines for Quantitative Trait Loci Analysis. **Journal of Near Infrared Spectroscopy**, v. 10, n. 4, p. 315–325, out. 2002. <https://doi.org/10.1255/jnirs.348>
- TÜRKAN, İ; BOR, M; OZDEMIR, F; KOCA, H. Differential responses of lipid peroxidation and antioxidants in the leaves of drought-tolerant *P. acutifolius* Gray and drought-sensitive *P. vulgaris* L. subjected to polyethylene glycol mediated water stress. **Plant Science**, v. 168, n. 1, p. 223–231, jan. 2005. <https://doi.org/10.1016/j.plantsci.2004.07.032>
- VIEIRA, R. D. 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.
- VIERA, A. J.; GARRETT, J. M. Understanding interobserver agreement: the *kappa* statistic. **Family Medicine**, v. 37, n. 5, p. 360–363, 1 maio 2005.
- WANG, D; DOWELL, F. E; RAM, M. S; SCHAPAUGH, W. T. Classification of Fungal-Damaged Soybean Seeds Using Near-Infrared Spectroscopy. **International Journal of Food Properties**, v. 7, n. 1, p. 75–82, 31 dez. 2007. <https://doi.org/10.1081/JFP-120022981>

- WANG, L.; HUANG, Z.; WANG, R. Discrimination of cracked soybean seeds by near-infrared spectroscopy and random forest variable selection. **Infrared Physics & Technology**, p. 103731, mar. 2021. <https://doi.org/10.1016/j.infrared.2021.103731>
- WIJEWARDANA, C.; REDDY, K. R.; BELLALLOUI, N. Soybean seed physiology, quality, and chemical composition under soil moisture stress. **Food Chemistry**, v. 278, p. 92–100, abr. 2019. <https://doi.org/10.1016/j.foodchem.2018.11.035>
- WILLIAMS, P. C. Implementation of near-infrared technology. In **Near Infrared Technology in the Agricultural and Food Industries**, 2nd Ed., eds. P. Williams and K. Norris. St. Paul, Minn.: American Association of Cereal Chemists. 2001
- WILSON, D.O.; MCDONALD, M.B. The lipid peroxidation model of seed ageing. **Seed Science and Technology**, v.14, p. 269-300, 1986.
- WOLD, S.; SJÖSTRÖM, M.; ERIKSSON, L. PLS-regression: a basic tool of chemometrics. **Chemometrics and Intelligent Laboratory Systems**, v. 58, n. 2, p. 109–130, out. 2001. [https://doi.org/10.1016/S0169-7439\(01\)00155-1](https://doi.org/10.1016/S0169-7439(01)00155-1)
- WORKMAN, J.; SHENK, J. Understanding and Using the Near-Infrared Spectrum as an Analytical Method. **Near-Infrared Spectroscopy in Agriculture**, p. 1–10, 26 out. 2015. <https://doi.org/10.2134/agronmonogr44.c1>
- WESTAD, F. MARANI, F. Validation of chemometric models – A tutorial. **Analitica Chimica Acta**, 893, 14-24, 2015. <https://doi.org/10.1016/j.aca.2015.06.056>
- XIN, X. et al. Reduced mitochondrial and ascorbate–glutathione activity after artificial ageing in soybean seed. *Journal of Plant Physiology*, v. 171, n. 2, p. 140–147, jan. 2014. <https://doi.org/10.1016/j.jplph.2013.09.016>
- XU, J; NWAFOR, C. C; SHAH, N; ZHOU, Y; ZHANG, C. Identification of genetic variation in *Brassica napus* seeds for tocopherol content and composition using near-infrared spectroscopy technique. **Plant Breeding**, v. 138, n. 5, p. 624–634, 16 maio 2019. <https://doi.org/10.1111/pbr.12708>
- YANG, H.; IRUDAYARAJ, J.; PARADKAR, M. Discriminant analysis of edible oils and fats by FTIR, FT-NIR and FT-Raman spectroscopy. **Food Chemistry**, v. 93, n. 1, p. 25–32, nov. 2005. <https://doi.org/10.1016/j.foodchem.2004.08.039>
- YANG, S. et al. Discrimination of corn variety using Terahertz spectroscopy combined with chemometrics methods. **Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy**, v. 252, p. 119475, maio 2021. <https://doi.org/10.1016/j.saa.2021.119475>
- ZHANG, D. CHEN, G; ZHANG, H; et al. Integration of spectroscopy and image for identifying fusarium damage in wheat kernels. **Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy**, v. 236, p. 118344, ago. 2020. <https://doi.org/10.1016/j.saa.2020.118344>
- ZHANG, G. et al. Analysis of multiple soybean phytonutrients by near-infrared reflectance spectroscopy. **Analytical and Bioanalytical Chemistry**, v. 409, n. 14, p. 3515–3525, 19 abr. 2017. <https://doi.org/10.1007/s00216-017-0288-8>

- ZHANG, L.; RAO, Z.; JI, H. Hyperspectral imaging technology combined with multivariate data analysis to identify heat-damaged rice seeds. **Spectroscopy Letters**, v. 53, n. 3, p. 207–221, 12 fev. 2020. <https://doi.org/10.1080/00387010.2020.1726402>
- ZHANG, L. et al. Hyperspectral imaging technology combined with deep forest model to identify frost-damaged rice seeds. v. 229, p. 117973–117973, 15 mar. 2020. <https://doi.org/10.1016/j.saa.2019.117973>
- ZHANG, S. LI, Z; NA, J; YANG, Y; TANG, X. Identification of aflatoxin B1 in peanut using near-infrared spectroscopy combined with naive Bayes classifier. **Spectroscopy Letters**, v. 54, n. 5, p. 340–351, 28 maio 2021. <https://doi.org/10.1080/00387010.2021.1931792>
- ZHANG, S; LIU, S; SHEN, L; HE, L; LIU, A. Application of near-infrared spectroscopy for the nondestructive analysis of wheat flour: A review. **Current Research in Food Science**, v. 5, p. 1305–1312, 2022. <https://doi.org/10.1016/j.crfs.2022.08.006>
- ZHOU, Q. et al. Non-destructive discrimination of the variety of sweet maize seeds based on hyperspectral image coupled with wavelength selection algorithm. **Infrared Physics & Technology**, v. 109, p. 103418, set. 2020. <https://doi.org/10.1016/j.infrared.2020.103418>