



UNIVERSIDADE ESTADUAL DE FEIRA DE SANTANA
PROGRAMA DE PÓS-GRADUAÇÃO EM RECURSOS GENÉTICOS VEGETAIS

MARIA APARECIDA RODRIGUES FERREIRA

**AVALIAÇÃO E RECOMENDAÇÃO DE ESPECTRÔMETRO -
VIS-NIR PORTÁTIL PARA O MONITORAMENTO DA
QUALIDADE FISIOLÓGICA DE SEMENTES FLORESTAIS
NATIVAS**

Feira de Santana- Bahia

2025

MARIA APARECIDA RODRIGUES FERREIRA

**AVALIAÇÃO E RECOMENDAÇÃO DE ESPECTRÔMETRO -
VIS-NIR PORTÁTIL PARA O MONITORAMENTO DA
QUALIDADE FISIOLÓGICA DE SEMENTES FLORESTAIS
NATIVAS**

Tese apresentada ao Programa de Pós-Graduação em Recursos
Genéticos Vegetais, da Universidade Estadual de Feira de Santana
para obtenção do título de Doutora em Recursos Genéticos Vegetais.

Orientadora: Profa. Dra. Bárbara França Dantas

Coorientador: Prof. Dr. Sérgio Tonetto de Freitas

Feira de Santana- Bahia

2025

Ficha catalográfica - Biblioteca Central Julieta Carteadó – UEFS

Ferreira, Maria Aparecida Rodrigues

F442a Avaliação e recomendação de espectrômetro-Vis-NIR portátil para o monitoramento da qualidade fisiológica de sementes florestais nativas / Maria Aparecida Rodrigues Ferreira. - 2025.
88f. : il.

Orientadora: Barbara França Dantas

Coorientador: Sérgio Tonetto de Freitas

Tese (doutorado) - Universidade Estadual de Feira de Santana.
Programa de Pós-Graduação em Recursos Genéticos Vegetais, 2025.

I. Espectroscopia. 2. Vigor de sementes. 3. Teor de água. 4. *Machine Learning*. 5. Predição. I. Dantas, Barbara França, orient. II. Freitas, Sérgio Tonetto de, coorient. III. Universidade Estadual de Feira de Santana.
Programa de Pós-Graduação em Recursos Genéticos Vegetais. IV. Título.

CDU: 543.42:581.48

BANCA EXAMINADORA

Lucas Kennedy Silva Lima

(Universidade Estadual de Feira de Santana)



Documento assinado digitalmente
LUCAS KENNEDY SILVA LIMA
Data: 30/04/2025 09:15:38-0300
Verifique em <https://validar.iti.gov.br>

Karla dos Santos Melo de Sousa

(Universidade Federal do Vale do São Francisco)



Documento assinado digitalmente
KARLA DOS SANTOS MELO DE SOUSA
Data: 02/05/2025 16:30:17-0300
Verifique em <https://validar.iti.gov.br>

Gizele Ingrid Gadotti

(Universidade Federal de Pelotas)



Documento assinado digitalmente
GIZELE INGRID GADOTTI
Data: 08/05/2025 16:08:14-0300
Verifique em <https://validar.iti.gov.br>

Elza Alves Corrêa

(Universidade Estadual Paulista)



Documento assinado digitalmente
ELZA ALVES CORREA
Data: 09/05/2025 09:30:19-0300
Verifique em <https://validar.iti.gov.br>

Sérgio Tonetto de Freitas

(Embrapa semiárido)

Coorientador e presidente da banca



Documento assinado digitalmente
SERGIO TONETTO DE FREITAS
Data: 12/05/2025 16:24:12-0300
Verifique em <https://validar.iti.gov.br>

Feira de Santana –BA

2025

A Deus, minha fonte de força, sabedoria e paz, que me guiou em cada passo dessa jornada. A Ele, toda honra e glória. E à minha família, pelo apoio incondicional.

AGRADECIMENTO

Sou imensamente grata a Deus, que com Sua mão firme me guiou e fortaleceu, permitindo que eu seguisse em frente, sem hesitar. Agradeço por Sua infinita bondade e misericórdia, por tudo o que Ele faz por mim. Foi por Sua graça que cheguei até aqui, e é por ela que seguirei, firme na certeza de que estás comigo, hoje e sempre, meu ABBA.

Agradeço ao meu querido esposo Leonardo, que esteve ao meu lado, apoiando-me em todos os momentos e sempre disposto a ajudar. Muito obrigada, meu amor.

Aos meus familiares, que sempre torceram por minhas conquistas e me proporcionaram amor e apoio incondicionais, amo vocês de coração.

Aos amigos que estiveram presentes durante toda essa caminhada: Monica, Elisangela, Ianca, Dalila, Maiza, Rebeca, Elaine, Karina, Aline, Cássia e Cinthia. Agradeço pelas vibrações positivas, orações e por cada gesto de carinho.

Aos colegas do laboratório LASESA, Raquel, Jailton, Oli, Jamile, Williane e Marcos, vocês foram essenciais para que essa jornada fosse possível. Sou grata por todo o suporte e aprendizado.

À minha orientadora, Dra. Bárbara França Dantas, agradeço profundamente por sua competência, pela abertura de ideias e pela contribuição valiosa na construção deste trabalho. Minha gratidão é imensa.

Ao meu coorientador, Dr. Sérgio Tonetto de Freitas, que me acompanhou desde a graduação como orientador e agora como coorientador, minha eterna gratidão pelo seu apoio contínuo e por toda a orientação ao longo dessa trajetória acadêmica.

A Gizele Ingrid Gadotti, que com suas dicas e sugestões contribuiu para a melhoria deste trabalho. Agradeço imensamente pelo seu auxílio.

A Jasciane, que sempre foi prestativa e disposta a tirar dúvidas e ajudar no desenvolvimento deste trabalho. Minha sincera gratidão por sua dedicação e apoio.

Agradeço à Empresa Brasileira de Pesquisa Agropecuária (Embrapa Semiárido), especialmente à Dra. Bárbara França Dantas, pelo apoio, pela disponibilização de equipamentos e pelo suporte técnico essencial para a execução de todas as etapas deste trabalho.

Aos meus colegas de turma do programa de recursos genéticos vegetais, com um agradecimento especial à Sandra e à Syntia, minha sincera gratidão.

Aos docentes da Universidade Estadual de Feira de Santana (UEFS) e ao Programa de Pós-Graduação em Recursos Genéticos Vegetais (PPGRGV), sou grata pelo suporte constante e por contribuírem para o meu crescimento profissional e intelectual durante essa importante fase.

Agradeço à Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) - Código de Financiamento 001, pelo apoio financeiro imprescindível para a realização deste trabalho.

Deem graças ao Senhor, porque ele é bom; o seu amor dura para sempre (Salmos 107:1).

RESUMO

O uso da espectroscopia na região do visível e infravermelho próximo (Vis-NIR) para avaliar a qualidade de sementes possibilita que, antes e durante a coleta, seja obtida uma análise inicial das características das sementes, evitando a destruição do material como ocorre no método padrão. O Vis-NIR, aliado a técnicas de *machine learning* (ML), contribui para desenvolver modelos que permitam avaliar a qualidade das sementes de espécies florestais nativas do bioma Caatinga. Este estudo teve como objetivo desenvolver modelos de calibração precisos com algoritmos de aprendizado de máquina para avaliar a qualidade das sementes de duas leguminosas florestais nativas do bioma Caatinga, utilizando um espectrômetro Vis-NIR portátil. Foram formados sublotos de sementes com diferentes teores de água (TA) a partir dos lotes originais, ajustando o teor por meio de hidratação ou desidratação controlada. As sementes foram colocadas sobre tela em caixas plásticas tipo "gerbox" com 40 mL de água (hidratação) ou sílica em gel (desidratação) por períodos de 24 ou 72 horas, gerando sublotos com diferentes TA. A avaliação espectral do TA das sementes (TA_{NIR}) foi feita por leitura individual com o espectrômetro portátil Vis-NIR. Após a coleta dos dados espectrais, o TA de referência (TA_{REF}) foi medido com um método tradicional adaptado para sementes individuais. Os dados espectrais originais foram utilizados, e um filtro de amostragem foi aplicado para ajuste dos modelos. No segundo trabalho, as leituras espectrais das sementes foram seguidas pelo teste de germinação em ambas as espécies, com 25 sementes e quatro repetições por lote. Após a germinação, classificou-se as sementes quanto à capacidade de germinação e formação de plântulas (normais, anormais ou sem formação). Para determinar o vigor, as leituras também foram feitas individualmente, com as sementes incubadas em B.O.D. a 30 °C, sendo o vigor classificado em alto, baixo ou nulo. Foram avaliados quatro algoritmos discriminativos: *Multilayer Perceptron* (MP), *Random Forest* (RF), J48 e *Support Vector Machine*. O algoritmo *Gaussian Processes* foi utilizado apenas no primeiro trabalho, enquanto o J48 foi aplicado exclusivamente no segundo trabalho. O MP se destacou na predição do teor de água (TA) das sementes de *Anadenanthera colubrina* e *Cenostigma pyramidale* com os dados originais. Por outro lado, o RF apresentou resultados superiores após a aplicação do filtro de amostragem. No segundo trabalho, *A. colubrina* e *C. pyramidale*, os algoritmos *Random Forest* (RF) e J48 apresentaram resultados superiores na classificação de germinação, plântulas e vigor. Esses métodos se destacaram ao classificar as sementes germinadas, as que não produziram plântulas e as de alto vigor, com as maiores taxas de acerto em ambas as espécies. A combinação de Vis-NIR com aprendizado de máquina mostrou-se eficaz na discriminação entre as classes de germinação e vigor, sendo necessário, no entanto, incluir mais dados para aprimorar a precisão dos modelos.

Palavra chaves: espectroscopia; vigor de sementes; teor de água; *machine learning*; predição.

ABSTRACT

The use of visible and near-infrared (Vis-NIR) spectroscopy to assess seed quality allows an initial analysis of seed characteristics to be obtained before and during seed collection, avoiding the destruction of the material as occurs in the standard method. Vis-NIR, combined with machine learning (ML) techniques, contributes to the development of models that allow the assessment of seed quality of native forest species of the Caatinga biome. This study aimed to develop accurate calibration models with machine learning algorithms to assess the seed quality of two native forest legumes of the Caatinga biome, using a portable Vis-NIR spectrometer. Sublots of seeds with different water contents (WC) were formed from the original lots, adjusting the content through controlled hydration or dehydration. The seeds were placed on a screen in plastic boxes type "gerbox" with 40 mL of water (hydration) or silica gel (dehydration) for periods of 24 or 72 hours, generating sublots with different WC. The spectral evaluation of the WC of the seeds (WC_{NIR}) was performed by individual reading with the portable Vis-NIR spectrometer. After collecting the spectral data, the reference WC (WC_{REF}) was measured with a traditional method adapted for individual seeds. The original spectral data were used, and a sampling filter was applied to adjust the models. In the second study, the spectral readings of the seeds were followed by the germination test in both species, with 25 seeds and four replicates per lot. After germination, the seeds were classified according to their germination capacity and seedling formation (normal, abnormal or without formation). To determine the vigor, the readings were also performed individually, with the seeds incubated in B.O.D. at 30 °C, with vigor classified as high, low or null. Four discriminative algorithms were evaluated: Multilayer Perceptron (MP), Random Forest (RF), J48 and Support Vector Machine. The Gaussian Processes algorithm was used only in the first study, while J48 was applied exclusively in the second study. MP stood out in predicting the WC of *Anadenanthera colubrina* and *Cenostigma pyramidale* seeds with the original data. On the other hand, RF obtained better results after applying the sampling filter. In the second study, for *A. colubrina* and *C. pyramidale*, the Random Forest (RF) and J48 algorithms presented the best results in the classification of germination, seedlings and vigor. These methods stood out in classifying germinated seeds, those that did not produce seedlings and those with high vigor, with the highest accuracy rates in both species. The combination of Vis-NIR with machine learning proved to be effective in discriminating between germination and vigor classes, although it is necessary to include more data to improve the accuracy of the models.

Keywords: spectroscopy; seed vigor; water content; machine learning; prediction.

SUMÁRIO

1 INTRODUÇÃO	13
2 REVISÃO DE LITERATURA	14
2.1 Angico-de-carço (<i>Anadenanthera colubrina</i> (Vell.) Brenan)	14
2.2 Catingueira-verdadeira (<i>Cenostigma pyramidale</i> (Tul.) Gagnon & G.P.Lewis)	15
3. Avaliação da qualidade das sementes	16
3.1 Teor de água	16
3.2 Teste de germinação	17
3.3 Teste de vigor	18
4. Espectroscopia do Infravermelho Próximo (Vis-NIR)	19
4.1 Quimiometria	21
5. Aprendizagem de máquina (<i>Machine learning</i> - ML)	23
5.1 Subamostragem aleatória k-fold	23
5.2 Validação cruzada k-fold	23
5.3 Desequilíbrio de classes	24
5.4 Supervisionado: Regressão e classificação	25
5.5 Árvore de decisão	25
5.5.1 <i>Random Forest</i> e J48	26
5.6 <i>Support Vector Machine</i> (SVM)	27
5.7 Rede neural artificial (RNA)	27
5.8 Redes <i>Perceptrons</i> de Múltiplas Camadas (MLP)	28
5.9 <i>Gaussian Processes</i> (GP)	29
6 REFERÊNCIAS	30
Capítulo 1: Predição não destrutiva do teor de água de sementes florestais nativas da caatinga utilizando espectroscopia Vis-NIR e aprendizagem de máquina.....	38
1 INTRODUÇÃO	40
2 MATERIAL E MÉTODOS.....	42
2.1 Coleta e armazenamento das sementes avaliadas.....	42
2.2 Obtenção de lotes com diferentes teores de água.....	42
2.3 Aquisição dos espectros Vis-NIR	43
2.4 Construção dos modelos.....	43
2.5 Métricas de regressão	44
3 RESULTADOS.....	44

3.1 Valores de referência (TA_{REF}): utilizados para o treinamento e teste:	44
3.2 Espectros de refletância.....	45
3.3 Espectros de refletância.....	46
3.4 Modelos com dados brutos.....	46
3.5 Modelos com Resample	47
4 DISCUSSÃO.....	48
5 CONCLUSÃO	52
6 REFERÊNCIAS	52
Capítulo 2: Uso do espectrômetro Vis-NIR portátil e <i>machine learning</i> na determinação do vigor de sementes florestais.....	59
1 INTRODUÇÃO.....	61
2 MATERIAL E MÉTODOS.....	62
2.1 Material vegetal.....	63
2.2 Obtenção dos dados espectrais Vis-NIR	63
2.3 Teste de Germinação.....	63
2.3.1 Classificação da germinação	64
2.3.2 Classificação de plântulas	64
2.3.3 Avaliação de plântulas	65
2.4 Avaliação do vigor	66
2.4.1 Classificação do vigor das sementes	66
2.5 Construção dos modelos de calibração	66
3 Métricas das acurácias.....	67
3.1 Cálculos das acurácias.....	67
4 RESULTADOS.....	69
5 DISCUSSÃO.....	78
6 REFERÊNCIAS	84

1 INTRODUÇÃO

A conservação de sementes florestais desempenha um papel importante na preservação, enriquecimento e restauração de ambientes degradados. As alterações climáticas e a pressão agrícola aumentaram a vulnerabilidade da desertificação das florestas tropicais secas (MARENGO, 2008; MACINNIS-NG et al., 2021). Neste contexto, a conservação da biodiversidade vegetal é imprescindível, principalmente em populações localizadas em zonas críticas que são ameaçadas pelas alterações climáticas e pressão agrícola (MARENGO, 2014).

A avaliação tradicional da qualidade das sementes armazenadas depende de testes destrutivos para avaliar o vigor, a germinação e o teor de água. Estes testes geralmente são mais demorados, necessitam de pessoas especializada, não sendo convenientes para ações de restauração (GOMES et al., 2019). Desse modo, métodos não destrutivos rápidos e precisos podem permitir a tomada de decisões eficientes em todas as atividades de conservação de sementes nativas, desde a colheita e armazenamento até o plantio em ações de restauração.

A espectrometria aliada à aprendizagem de máquina pode ser um método prático para avaliar a qualidade de sementes florestais nativas (TIGABU et al., 2020). Essa técnica pode proporcionar para processo de restauração mais agilidade, consequentemente a manutenção de um volume menor de sementes armazenadas (MUKASA et al., 2019).

Para estimar a qualidade das sementes é necessário em algumas análises o uso de reagentes, ou a utilização de outros materiais para realização das avaliações. Assim, a utilização da tecnologia Vis-NIR pode facilitar as análises de qualidade de sementes, tornando estas mais rápidas, práticas e precisa (CHENG et al., 2023; WANG et al., 2023).

Na literatura não existem trabalhos voltados para uso da tecnologia Vis-NIR em sementes espécies florestais do bioma Caatinga. Nesse sentido, fez-se necessário realizar estudos sobre o uso desta tecnologia para a avaliação da qualidade de sementes de *Anadenanthera colubrina* (angico-de-carço) e *Cenostigma pyramidale* (catingueira-verdadeira).

A espectroscopia na região do infravermelho próximo (NIR, do inglês *Near Infrared Spectroscopy*) é um método analítico proposto para a avaliação da qualidade de diversos produtos agrícola, tendo as vantagens de ser preciso, rápido e não necessitar de preparo de amostras ou uso de reagentes químicos (JHA et al., 2014; FERREIRA, 2015; MARQUES et al., 2016; MEDEIRO, 2021). Contudo, o uso dessa tecnologia exige a calibração e validação do espectrômetro NIR para atributos de qualidade das sementes. Para isso é necessário o

conhecimento sobre esse método (BETEMPS et al., 2013; MARQUES, 2017; RIBEIRO et al., 2021).

A espectrometria (NIR) em combinação com os métodos de aprendizagem de máquina (*machine learning*) podem estimar o teor de água e a composição química de materiais biológicos com base na absorção de radiação infravermelha próxima por ligações entre átomos leves (BRERETON, 2015). O método de aprendizagem de máquina faz parte do ramo da inteligência artificial e é uma técnica que vem sendo bastante utilizada em diferentes áreas do conhecimento em conjunto com espectroscopia na região do infravermelho próximo para calibração e predição de conjuntos de dados com uso de diferentes algoritmos (SOUZA et al., 2012; YU et al., 2021; FERREIRA et al., 2022).

Desta forma, a hipótese deste trabalho é que a utilização da espectroscopia Vis-NIR pode viabilizar a determinação rápida e precisa da qualidade de sementes de espécies nativas florestais para fins de conservação e restauração. A técnica Vis-NIR pode ser utilizada para determinar o teor de água, germinação e vigor com mais rapidez e precisão. Assim, o objetivo deste trabalho foi avaliar a qualidade fisiológica de sementes de *Anadenanthera colubrina* e *Cenostigma pyramidale* com um espectrômetro Vis-NIR portátil em conjunto com as técnicas de aprendizagem de máquina.

2 REVISÃO DE LITERATURA

O bioma Caatinga é constituído por árvores e arbustos espinhentos, plantas suculentas espinhosas e por plantas herbáceas que se desenvolvem após o período de chuva (CARDOSO e QUEIROZ et al, 2006; LIMA et al., 2018). A família Fabaceae exibe grande diversidade, com 293 espécies em 77 gêneros, das quais 144 espécies são endêmicas (QUEIROZ et al., 2006). Entre essas, são encontrados nesse bioma duas espécies de grande importância para a população do semiárido nordestino, *Anadenanthera colubrina* e *Cenostigma pyramidale*.

2.1 Angico-de-carço (*Anadenanthera colubrina* (Vell.) Brenan)

Anadenanthera colubrina (Vell.) Brenan é conhecida comumente como angico-de-carço. Uma espécie presente em abundância na vegetação de Caatinga, no Cerrado e Mata Atlântica (CAMPOS FILHO e SARTORELLI, 2015; NASCIMENTO et al., 2019). Essa espécie é encontrada nas regiões Nordeste, Centro-Oeste, Sudeste e no Paraná, entre 100 e 1.100 m de altitude (PAREYN et al., 2018). *A. colubrina* pertence à família Fabaceae, é uma espécie arbórea pioneira e de fácil identificação em razão ao seu elevado porte e tronco (tem entre 10 e 20 m de altura) com casca descamante, apresentando uma madeira de cor

avermelhada, as sementes são escuras, brilhantes, orbicular, achatada, com ala estreita, com até 15 mm de comprimento (CARVALHO, 2002).

As sementes não apresentam dormência e germinam com facilidade, essa espécie apresenta também produção vigorosa e tem resistência à dessecação de suas plantas, por causa da presença de órgão de reservas (PIRES-NETO et al., 2016).

Os produtores rurais da região semiárida cultivam essa espécie, pois a planta apresenta grande importância ecológica, utilizadas para fins medicinais, madeira para manutenção de cercas, produção de forragem para alimentação animal, contribui no aumento do teor de matéria orgânica quando incorporada ao solo e pode atuar como fonte de nitrogênio a outras plantas consorciadas (SILVA et al., 2012).

2.2 Catingueira-verdadeira (*Cenostigma pyramidale* (Tul.) Gagnon & G.P.Lewis

Cenostigma pyramidale (catingueira-verdadeira) pertence à família da Fabaceae, essa espécie teve origem no bioma Caatinga, é conhecida por diferentes nomes: canela-de-velho, catingueira-verdadeira, pau-de-rato, catingueira-de-porco, catingueira, catingueira-de-mulata, catingueiro-das-folhas largas e mussitaíba (MATIAS et al., 2017). No período de seca perde as folhas para impedir a perda de água por evapotranspiração e quando começam as primeiras chuvas brotam e frutificam (OLIVEIRA et al., 2011). A *C. pyramidale* é uma espécie endêmica do Brasil que tem uma grande distribuição no Nordeste brasileiro, compreendendo os estados do Piauí, Ceará, Rio Grande do Norte, Paraíba, Pernambuco, Alagoas, Sergipe, Bahia e Maranhão, Sudeste (Minas Gerais) e podendo ocorrer no Centro-Oeste (Goiás) (GAEM, 2020).

Cenostigma pyramidale é uma árvore de porte médio, sem espinhos, com 4 a 6 m de altura, por vezes alcançando 12 m (LIMA et al., 2018). Os frutos apresentam deiscência explosiva, propagando sementes para distante da árvore matriz. Desse modo, a época mais recomendada para fazer a colheita é quando os frutos começam a se abrir ou mudar a sua coloração de verde-amarelo para marrom, sendo recomendado realizar a colheita no período em que as vagens exibem coloração marrom. A semente apresenta forma elipsóide, cor castanho-amarelada e superfície lisa, polida e dura (MATIAS et al., 2019; MENDONÇA et al., 2016; SILVA; DANTAS, 2017). *C. pyramidale* tem grande potencial econômico, sendo útil para produção de lenha, forragem, uso medicinal e sobretudo por suas propriedades extrativistas, suas folhas e cascas são utilizadas como alternativa para o tratamento de

infecções respiratórias e estomacais (DANTAS et al., 2009; ANTUNES et al., 2014; SANTOS et al., 2020).

3. Avaliação da qualidade das sementes

A qualidade de um lote é caracterizada pelo conjunto de atributos genéticos, fisiológicos, sanitários e físicos que determinam o seu valor para semeadura, destacando-se as características fisiológicas da qualidade das sementes, diretamente responsável pelo seu desempenho após o armazenamento. A germinação e vigor obtido em laboratório, permitem a comparação entre os lotes de sementes e avaliar a probabilidade de sucesso com sua aquisição e utilização (NETO e KRZYZANOSKI, 2009; MARCOS-FILHO, 2011).

O conhecimento sobre a qualidade das sementes é importante para evitar perdas no campo, nesse sentido, a coleta deve acontecer quando as sementes atingirem a maturação fisiológica, ocasião na qual apresentam elevado percentual de germinação, maior vigor e alto potencial para o armazenamento, por isso, é essencial determinar o momento em que a semente atinge a maturação fisiológica, dado que o tempo da colheita varia em função da espécie, do ano e de árvore para árvore, sempre havendo a necessidade de acompanhar o estágio de maturação para indicar o momento da colheita e a qualidade das sementes (NOGUEIRA e MEDEIROS, 2007).

3.1 Teor de água

As sementes são constituídas de água e matéria seca, sendo que nesta se encontram os carboidratos, proteínas, óleos, vitaminas e nutrientes (CARVALHO e NAKAGAWA, 2012). O conhecimento sobre as propriedades físicas das sementes, como o teor de água é primordial para assertividade do momento ideal da colheita e do tempo de armazenamento, a fim de obtê-las com qualidade apropriada (NETO et al., 2014; KRELING et al., 2022). Além disso, a manutenção da qualidade de sementes depende de outros fatores ambientais como a temperatura e umidade relativa do ar (TRESENA et al., 2010). O teor de água corresponde ao ponto de equilíbrio e progride com elevação da umidade relativa do ar ou vice-versa (MARCOS FILHO, 2015). Em condições ambientais inadequadas, o teor de água das sementes pode aumentar e afetar negativamente a qualidade fisiológica (GOLDFARB et al., 2008).

A adequada diagnose da quantidade de água deve estar entre as propriedades de análises de sementes ou de um lote de sementes, certamente proporcionando um melhor aproveitamento das mesmas, evitando perdas desnecessárias (PIÑA-RODRIGUES et al.,

2015). O conteúdo de água presente nas sementes é indispensável, já que tem grande relevância para sua conservação e manejo, visto que a atividade fisiológica de sementes está diretamente relacionada com o teor de água (LIMA JR et al., 2010).

Na literatura, são mencionados diferentes tipos de água. A água do tipo 1 refere-se à água de constituição, que está ligada aos constituintes químicos da semente por meio de uma força de -150 MPa. Esse tipo de água está associado à matriz celular. A água do tipo 2, com potencial hídrico entre -170 a -11 MPa, corresponde a níveis de umidade entre 8% e 22% (base úmida). Já a água do tipo 3, com potencial de -11 a -4 MPa, está presente em um teor de umidade de 22% a 33% (base úmida), dependendo da espécie. Os dois primeiros tipos de água estão adsorvidos às moléculas presentes nas sementes e necessitam de maior pressão de vapor para serem removidos (OLIVEIRA et al., 2021).

A água dos tipos 2 e 3 é a mais relevante para a secagem comercial, pois a maioria dos processos de secagem de sementes ocorre nessas faixas de umidade, sendo sua remoção promovida pelo gradiente de umidade relativa e temperatura do ar de secagem. Na faixa de potencial hídrico entre -4 e -2 MPa, o teor de água nas sementes varia de 33% a 55%, correspondendo à água do tipo 4. Já a água do tipo 5 está associada a níveis de umidade acima de 55% (base úmida). As águas dos tipos 4 e 5 são facilmente removidas, sendo recomendado o uso de temperaturas mais baixas e maior velocidade do ar durante a secagem (OLIVEIRA et al., 2021).

3.2 Teste de germinação

O teste de germinação é primordial para avaliação da qualidade de sementes, visto que facilita a avaliação do potencial de germinação, podendo também, em conjunto com outras ferramentas, ser indicador de possíveis falhas ocorridas no processo de produção, devido aos eventos fenológicos ou nas etapas de colheita, beneficiamento e armazenamento (PIÑA-RODRIGUES et al., 2015).

O teste de germinação é realizado na rotina dos laboratórios de análises de sementes, sendo importante que seja conduzido em condições ideais, considerando a disponibilidade de água e aeração, temperatura, luminosidade, características do substrato, com objetivo de se obter máxima porcentagem de germinação (FOWLER e MARTINS 2001; SILVA et al., 2019). Os principais objetivos do teste de germinação visam obter informações para determinar o valor da qualidade das sementes para semeadura e a comparação de diferentes lotes.

Fisiologista e tecnologista de sementes definem a germinação com conceitos parcialmente distintos a depender da finalidade: A germinação é definida como a continuidade do desenvolvimento do eixo embrionário, tem início com a hidratação e encerra com a emissão da raiz principal. Outra definição, considerar que a germinação de sementes é a emissão da radícula ou grau aceitável de desenvolvimento em suas estruturas essenciais resultando em obtenção de uma plântula normal sob condições favoráveis no campo (CARVALHO e NAKAGAWA; 2012; MARCOS FILHO, 2015). Quando é realizada a avaliação é necessário fazer uma interpretação do teste de germinação que tem a finalidade de indicar potencial germinativo das sementes de um lote. Em um lote é importante definir o número de plântulas normais, anormais, sementes mortas, duras e dormentes, essas características citadas indicam o vigor de um determinado lote (BRASIL, 2013).

3.3 Teste de vigor

O vigor é definido como a soma de atributos que confere a semente potencial para germinar, emergir e formar plântulas normais e saudáveis com rapidez, sob condições de grande diversidade ambiental (KRZYZANOWSKI e FRANÇA-NETO, 2001; CARVALHO e NAKAGAWA, 2012; PIÑA-RODRIGUES et al., 2015). Para avaliar o potencial fisiológico de lotes de sementes, considerando seu impacto direto no estabelecimento do estande, o que pode influenciar, de forma direta ou indireta, a produção de uma cultura, em conjunto com outros fatores (MARCOS-FILHO, 2015). Segundo o mesmo autor, o vigor relaciona um conjunto de características que poderiam ser consideradas como atributos independentes, como velocidade de germinação, crescimento da plântula, a habilidade para germinar sob condições de estresse. Desta forma, o teste de vigor avalia os fatores bioquímicos e fisiológicos relacionado ao potencial de desenvolvimento de um lote (KRZYZANOWSKI e FRANÇA-NETO, 2001).

O vigor das sementes está relacionado com sua deterioração, quanto mais deteriorada menos vigorosas serão as sementes. Os testes de vigor podem ser divididos em fisiológico, bioquímicos e físicos. Os testes fisiológicos são aqueles que avaliam aspectos da germinação ou do crescimento das plântulas, tanto em condições favoráveis quanto sob estresse, estimando as reações fisiológicas associadas à perda de vigor. Já os testes bioquímicos analisam reações químicas ou processos, como a atividade enzimática e respiratória, que estão presumidamente relacionados à capacidade germinativa das sementes, esses testes avaliam o comportamento bioquímico relacionado à perda de vigor das sementes, enquanto os testes

físicos observam o comportamento das sementes em condições de estresse (PINA-RODRIGUES et al., 2015; MARCOS- FILHO, 2015).

Os testes de vigor podem ser classificados como diretos e indiretos. Os testes diretos simulam condições de campo, onde as sementes demonstram a capacidade de emergir sob estresse simulado. Segundo Carvalho e Nakagawa (2012), os testes diretos mais estudados incluem o teste a frio, a velocidade de emergência em campo, a população inicial, o peso da matéria verde, o peso da matéria seca e o crescimento das plântulas. Já os testes indiretos, que abrangem principalmente testes bioquímicos, destacam-se pela condutividade elétrica, teste de aldeídos voláteis e teste de tetrazólio, e do teor de ácidos graxos (VIEIRA e CARVALHO, 1994, PIÑA-RODRIGUES et al., 2015).

4. Espectroscopia do Infravermelho Próximo (Vis-NIR)

A origem da espectroscopia de infravermelho próximo ocorreu no início do século dezenove, com os primeiros trabalhos do pesquisador Frederick William Herschel, astrônomo e músico que escreveu dois artigos descrevendo o comportamento do espectro da radiação solar em relação aos efeitos de temperatura (AGELET e HURBURGH, 2010; MUNIZ et al., 2012). Em 1800 Herschel, na execução do seu estudo, separou as faixas espectrais relacionada à região visível utilizando um termômetro de bulbo enegrecido e prisma de vidro para separar a radiação solar em diferentes comprimentos de onda, determinou que abaixo do vermelho, onde não existia mais luz visível era a região que apresentava máxima temperatura. Essa região em seguida foi chamada de infravermelho, usando o prefixo grego “infra” que significa “abaixo” (VALDERRAMA, 2005).

A espectrometria Vis-NIR foi utilizada pela primeira vez em aplicações agrícolas por Noris (1964) para medir a umidade em grãos. Desde então, tem sido utilizada para análises rápidas, demonstrando que a técnica pode oferecer uma estimativa confiável, especialmente de umidade, proteínas e teor de gordura de uma grande variedade de produtos agrícolas e alimentícios (DAVIES e GRANT, 1987). Outros trabalhos demonstram que a espectrometria Vis-NIR tem um grande potencial para monitorar diferentes parâmetros de qualidade em alimentos (HUANG et al., 2014; TEYE et al., 2019). Equipamentos Vis-NIR portáteis têm utilidade para análise de controle de qualidade, *online* ou *in situ*, e apresentam vantagens como baixo custo, tamanho reduzido, robustez e simplicidade de análise (OLIVEIRA et al., 2020).

Na literatura são reportados diversos trabalhos sobre o uso da espectroscopia na região

do visível e infravermelho próximo (Vis-NIR) para avaliação da qualidade de sementes cultivadas. No estudo de Mayrink et al. (2020) avaliou-se diferentes níveis de vigor de sementes individuais de algodão com a espectroscopia Vis-NIR. Na pesquisa de Fan et al. (2020), foi avaliado o vigor de sementes individuais de trigo, desse modo gerando conhecimento sobre a qualidade dos grãos. Em outro trabalho com sementes de soja, foi utilizada a espectroscopia Vis-NIR juntamente com algoritmos de seleção de comprimento de onda para determinar o teor de água (GUO et al., 2023).

Os estudos observados na literatura com utilização da espectroscopia Vis-NIR para avaliar sementes florestais nativas, foram de Mukasa et al. (2019) e Tigabu et al. (2020), onde a espectroscopia Vis-NIR foi utilizada para a classificação do vigor de sementes de *Chamaecyparis obtusa* (Siebold & Zucc.) Endl. e *Cunninghamia lanceolata* (Lamb.) Hook. Em outros estudos de Tigabu et al. (2004) e Tigabu et al. (2007) foram desenvolvidos modelos para a avaliação e predição da germinação, bem como para a classificação de sementes infestadas e não infestadas por insetos. Foi evidenciado poucos trabalhos com espécies florestais brasileiras para a avaliação da qualidade fisiológica com o uso da tecnologia Vis-NIR. No estudo de Granja et al. (2021), sementes de *Phyllanthus niruri*, foram avaliadas por meio da espectroscopia Vis-NIR para identificar a correlação da coloração com o processo de maturação fisiológica.

A espectroscopia é compreendida como a ciência associada ao conhecimento do desenvolvimento da radiação eletromagnética e da sua interação com a matéria. O espectro eletromagnético contém uma ampla faixa de energia, promovendo várias formas de interação da radiação com a matéria (TIBOLA et al., 2018). A espectroscopia Vis-NIR, consegue identificar as impressões digitais espectrais que o material orgânico presente na amostra possui. Essa interação permite mostrar que cada substância tem resposta intrinsecamente distinta das outras. Logo pode-se entender que cada substância guarda uma impressão digital individual, em razão da relação de absorção muito forte de sobretons e combinação de banda de vibrações moleculares, sendo oriundo de diversos grupos funcionais orgânicos, podendo ser citado como exemplo C-H que está inserido aos grupos metileno, carbonil e aromáticos, O-H que pertence a hidroxila e N-H que pertence à amida primária, terciária e quaternária e também aos sais minerais e complexo orgânicos (VALDERMA, 2005; XIABO et al., 2010; SOUZA, 2016).

A espectroscopia Vis-NIR abrange a faixa do espectro eletromagnético com comprimentos de onda entre 750 e 2500 nm, observando que essa faixa de energia é bastante

para impulsionar as moléculas de seu estado vibracional fundamental para outros estados excitados (PASQUINI, 2003; CHAGAS, 2006; NICOLAI et al., 2007; GALLO, 2008; TIBOLA et al., 2018; PASQUINI, 2018).

Segundo Bruning et al. (2020), quando a luz atinge a superfície de um material vegetal, ela pode ser refletida, absorvida ou transmitida. A extensão desses processos depende do comprimento de onda da luz, do ângulo de incidência, da textura da superfície, da estrutura celular e da composição química do material. A transmitância refere-se à quantidade de luz que atravessa o material, representando a fração da intensidade luminosa original que passa e emerge do outro lado. A absorbância indica a quantidade de luz absorvida pelo material e é calculada como o logaritmo negativo da transmitância. A reflectância, por outro lado, mede a quantidade de luz refletida, ou seja, a fração da luz incidente que retorna para o ambiente de origem (DURMUS e DAVIS, 2015; TAKEDA et al., 2016; LEI et al., 2017).

Parte da luz absorvida por uma amostra resulta da vibração e estiramento de ligações moleculares específicas (COATES, 2000). O espectro de uma amostra pode ser separado em três faixas distintas de comprimentos de onda, determinadas pelos principais fatores que influenciam a reflectância em cada região: o Visível (VIS 350–700 nm), infravermelho próximo NIR 700–1100 nm) e Infravermelho de ondas curtas (SWIR 1.100–2.500 nm) permitindo observar fenômenos de transferência de energia eletromagnética em ligações químicas dos grupos C–H, N–H, O–H e S–H, que são comumente encontrados em moléculas orgânicas (CEN e HE, 2007; DA-WEN, 2009; TIBOLA et al., 2018; BRUNING et al., 2020).

4.1 Quimiometria

A quimiometria pode ser definida como uma área multidisciplinar do conhecimento, onde a matemática e a estatística são utilizadas para tratamentos de dados relativos aos processos químicos, de natureza multivariada, de forma a extrair uma maior quantidade de informações e melhores resultados analíticos, como por exemplo informações resultantes da espectroscopia, cromatografia, entre outras (Ferrão et al., 2004). Dessa forma, a quimiometria ocorreu em razão da necessidade de uma ferramenta que tivesse a capacidade de analisar dados químicos, originados a partir da interação da matéria com diferentes radiações eletromagnéticas (VALDERMA, 2005; FERREIRA, 2015). O uso de espectrômetros na caracterização da qualidade de amostras depende dos processos de calibração e validação dos equipamentos para esta finalidade (PEIRIS et al., 1999; MORAES, 2018). O processo de calibração envolve modelagem, onde os modelos desenvolvidos são capazes de relacionar as variáveis X, valores de medida instrumental obtidos através das informações fornecidas pelos

espectros, com valores de Y, obtidos pelos procedimentos de análises de referências para aquele tipo de determinação (FERREIRA et al., 1999; COSTA, 2018). No processo de validação, análises de referência são comparadas com as análises realizadas com os modelos desenvolvidos para determinação do erro analítico dos modelos. Após os processos de calibração e validação, a espectroscopia pode ser empregue nas análises de sementes de forma rápida, prática e sem a necessidade de preparação das amostras e uso de reagentes (BETEMPS et al., 2011). Além destes benefícios, a espectroscopia Vis-NIR permite a avaliação de diversos parâmetros de qualidade simultaneamente (NILSSON, 2012). Esta tecnologia tem sido empregada com êxito na avaliação de qualidade de sementes de soja, algodão e milho (BAZONI et al., 2017; SIMEONE et al., 2019; MAYRINK et al., 2020).

Para alguns autores o espectro de infravermelho próximo (Vis-NIR) da água é caracterizado por bandas de absorção intensas, que se devem às vibrações das ligações O-H na molécula de água. As principais ocorrem nos seguintes comprimentos de onda: 970 nm, relacionada à primeira sobreposição da vibração de estiramento assimétrico da molécula de água. 1190 nm associada à segunda sobreposição das vibrações de estiramento das ligações (LIU e GUO, 2015; SERSON et al., 2020; ZHANG e GUO, 2020; WANG et al., 2021; GUO et al., 2023). Outro autor relatou que os comprimentos de onda utilizados para quantificar grupos O-H estão nas regiões entre 800 a 1100 nm, 1100 a 2500 nm e 2500 a 4000 nm (WORKMAN JR, 1996).

O espectro NIR das proteínas exhibe bandas de absorção relacionadas principalmente aos sobretons e combinações das vibrações das ligações C-H, N-H e O-H (GRIFFITHS, 1983; PRETSCH et al., 2000). No trabalho de Shao et al. (2011), foi revelado que a faixa espectral associada às proteínas em grãos de arroz abrange de 1100 a 2500 nm. Nessa região, as bandas de absorção estão principalmente relacionadas aos sobretons e combinações das vibrações das ligações C-H, N-H e O-H, características das proteínas. Utilizando medições na faixa do infravermelho próximo (NIR) de 700 a 3500 nm, os autores investigaram como o NIR pode ser utilizado para distinguir entre grupos N-H e O-H, tanto os envolvidos em ligações de hidrogênio quanto os não envolvidos. As substâncias analisadas incluem metanol, anilina e poli- γ -benzil-L-glutamato (ELLIOTT e AMBROSE, 1950). Outras pesquisas indicam que as bandas em 605 nm e 540 nm estão associadas à quarta e quinta harmônica do amido, respectivamente, enquanto o pico em 643 nm está relacionado à terceira e quarta harmônica das proteínas (CHENG et al., 2023; WANG et al., 2023).

5. Aprendizagem de máquina (*Machine learning*- ML)

Machine learning (ML) pode ser definido como classe de métodos para análises automatizada de dados, os quais são capazes de identificar padrões que podem ser usados para prever informações desconhecidas ou para auxiliar nos processos de tomada de decisão sob incerteza (XAVIER e CHEN, 2022; CHENG et al., 2023), em outras palavras, o método ML é um campo da inteligência artificial que está concentrada no estudo de algoritmos que aprendem com os dados para melhorar seu desempenho em uma tarefa específica. Por exemplo, na tarefa de classificação e regressão revelam padrões do conjunto de dados da amostra de estudo que podem ser extremamente complexos. O ML deve ser usado em duas situações: primeiro para tratar problemas que os métodos tradicionais não conseguem de maneira aproximada e em segundo, para otimizar a solução de problemas tratáveis, porém de maneira melhor e mais robusta, rápida e econômica (MURPHY, 2012; VÁZQUEZ-MARRUFO, 2023; SCHLEDER e FAZZIO, 2021).

A inteligência artificial e o aprendizado de máquina podem impulsionar o desenvolvimento sustentável na agricultura, ajudando a criar modelos preditivos que apoiam diversos segmentos do setor e promovem a otimização de recursos. Com a crescente acessibilidade de inovações tecnológicas no campo, abre-se a oportunidade de integrar a análise da qualidade das sementes a processos tecnológicos (GADOTTI et al., 2022). Para a construção do modelo, são utilizados métodos como a subamostragem aleatória k-fold ou a validação cruzada k-fold.

5.1 Subamostragem aleatória k-fold

Na subamostragem aleatória k-fold, o método de retenção única é repetido k vezes, de modo que k pares de conjuntos de treinamento e conjuntos de teste, são gerados. O algoritmo de aprendizagem é aplicado a cada conjunto de treinamento, e o modelo resultante é então aplicado ao conjunto de teste correspondente. O desempenho é calculado como a média dos resultados obtidos em cada um dos k conjuntos de teste. Vale destacar que os pares de conjuntos de treinamento e teste são disjuntos, ou seja, não compartilham nenhum elemento. (BERRAR, 2019).

5.2 Validação cruzada k-fold

A validação cruzada k-fold é um método em que o conjunto de dados é dividido em k subconjuntos disjuntos e de tamanho aproximadamente igual, sem sobreposição entre os conjuntos de teste. O modelo é treinado em subconjuntos e avaliado no subconjunto restante,

repetindo o processo até que todos os subconjuntos sejam usados como conjunto de validação. O desempenho final é a média das avaliações. Esse método é amplamente utilizado para ajustar hiperparâmetros, prevenir *overfitting*, comparar algoritmos e estimar o erro de generalização de modelos (BERRAR, 2019; GHOJOGH e CROWLEY, 2019).

5.3 Desequilíbrio de classes

A distribuição de classes refere-se ao processo de categorização de objetos rotulados de forma equilibrada. Esse agrupamento pode ser feito com base na quantidade igual de eventos nos conjuntos de dados ou com base em características ou atributos semelhantes desses eventos. Em muitos casos, a proporção de objetos rotulados é semelhante entre as classes, mas isso nem sempre ocorre. Por exemplo, quando a quantidade de uma determinada classe (a classe majoritária) é significativamente maior que a de outra (a classe minoritária) em um conjunto de dados, ocorre um problema de desequilíbrio de classes (SUN et al., 2009; KRAWCZYK et al., 2014).

Em mineração de dados, o desequilíbrio de classes é um problema frequente e seus efeitos são bastante prejudiciais, já que torna desafiadora a identificação de eventos raros, atividades anômalas e padrões incomuns. Uma das principais dificuldades envolve a classificação incorreta desses eventos raros, o que resulta em uma significativa perda de desempenho e aumento de custos. Para reduzir o impacto do desequilíbrio de classes, o principal foco deve estar no aprendizado de conjuntos de dados desbalanceados (LÓPEZ et al., 2013; HAIXIANG et al., 2017).

O foco principal no estudo de conjuntos de dados desequilibrados é a identificação de fatores como a presença de classes minoritárias, eventos raros ou a proporção entre as classes majoritárias e minoritárias. Alguns aspectos importantes sobre o desequilíbrio de dados incluem o fato de que ele é avaliado em relação a um conjunto de dados específico. Ou seja, se a distribuição de treinamento segue uma distribuição padrão para a classificação e presume-se que essa distribuição está correta, o conjunto de dados é considerado balanceado. No entanto, se o desempenho da classificação não for satisfatório, a situação se torna mais complexa. Nesse caso, pode-se tentar balancear os dados artificialmente para mitigar o problema, mas essa abordagem não elimina a questão por completo. Outro ponto a ser considerado é a proporção relativa dos eventos dentro dos conjuntos de dados, que desempenha um papel importante na avaliação do equilíbrio (AL-GHRAIBAH et al., 2015).

5.4 Supervisionado: Regressão e classificação

Segundo Schleder e Fazzio (2021) existem categorias de problemas de ML de acordo com os tipos de entrada e saídas, referindo-se à aprendizagem supervisionada. Na aprendizagem supervisionada, o intuito é determinar a função que leva as entradas e saídas alvos, sendo que dessa forma os dados são rotulados (x, y) e são conhecidos como conjunto de treinamento (DÍAZ et al., 2021). O método *machine learning* usa vários tipos de algoritmo para obter modelos. Os algoritmos de aprendizagem são escolhidos de modo particular, sendo essa escolha uma etapa muito importante para construção do modelo, tem uma variedade de opções disponíveis, cada uma é adequada para um determinado conjunto de dados (TANG et al., 2021).

A regressão é uma técnica de análise de dados utilizada para apoiar na tomada de decisões sobre variáveis quantitativas. Trata-se de um método de mineração de dados voltado para a previsão de valores contínuos ou intervalos numéricos. Esse método de previsão utiliza uma equação de regressão que envolve duas variáveis: a variável preditora (a ser determinada) e a variável de resposta (os valores que se deseja prever) (SAGAR et al., 2021). Alguns exemplos são: Regressão Linear Simples, a regressão linear múltipla, redes neurais artificiais, o SVM (*Support Vector Machines*). Os modelos de classificação ordenam os dados de entrada em categorias. É aconselhado que os dados sejam categorizados, separados e identificados em grupos ou classes específicas. Neste caso, reconhece que a resposta da variável Y é qualitativa. Exemplos de algoritmos utilizados: Regressão logística, o SVM (*Support Vector Machines*), o *Random Forests* e J48 (Schleder e Fazzio, 2021; MATOS, 2021).

O propósito principal dos algoritmos de regressão e classificação é aprender a função que se aproxima melhor possível a distribuição dos dados disponíveis para o tratamento dos modelos. Então os algoritmos têm como objetivo minimizar o erro entre o valor predito pelo modelo aproximado e os valores de referência usados para treinamento (SCHEIDER e FAZZIO, 2021).

5.5 Árvore de decisão

A árvore de decisão é utilizada para dados de regressão e classificação. Os dados de entrada têm funções de vetor de atributos e uma decisão como valor de saída que pode ser classificada como: sim ou não (SATO et al., 2013). Esse algoritmo é representado graficamente por ramos, semelhante a uma árvore, com sentido invertido, o nó raiz é o

primeiro nó da árvore, no topo da estrutura e é o principal nó da árvore de decisão (Figura 1). Os nós filhos ou internos estão conectados por ramos e estão localizados abaixo do nó raiz e são nós de decisão também. Cada um contém um teste sobre a variável independente e os resultados desse teste formam os ramos da árvore. Os nós folhas, nas extremidades da árvore, representam valores de predição para variável dependente (POZZER, 2006; MEIRA et al., 2008; LAURETTO, 2010).

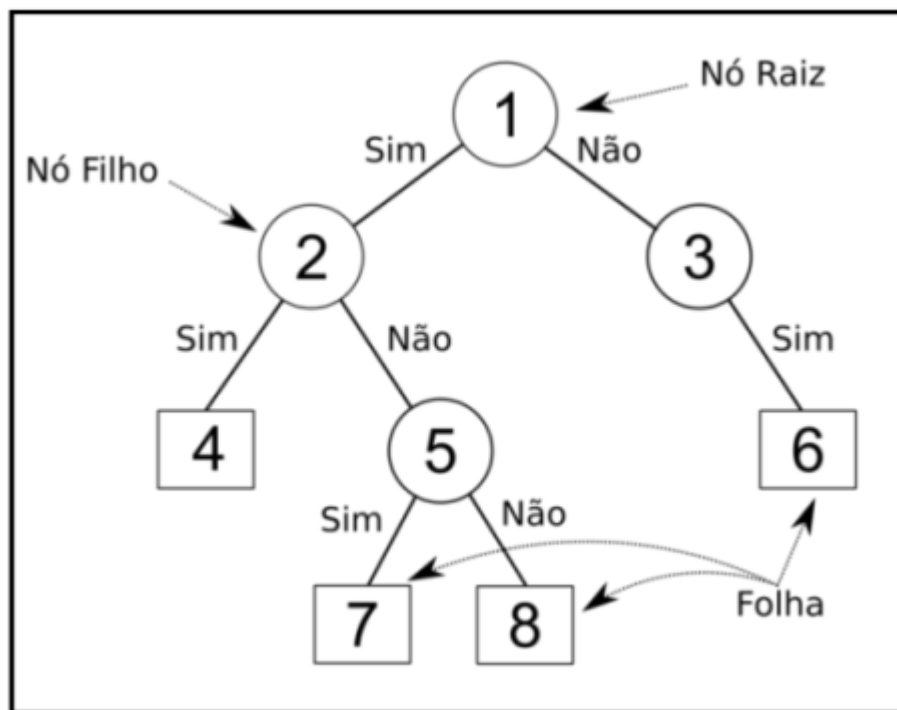


Figura 1. Estrutura geral de uma árvore de decisão

Fonte: Sato et al. (2013).

5.5.1 *Random Forest* e J48

O *Random Forest* ou floresta aleatória é um algoritmo constituído por várias árvores de decisão, essas árvores são geradas como classificadores fracos com intuito de dar origem a um classificador forte, para atingir a melhor performance, os testes destes algoritmos irão testar várias árvores, e desse modo gerar uma que melhor se ajuste ao conjunto dos modelos de treinados (MARIN et al., 2021). Selecionando aleatoriamente a partir do conjunto original, contendo todos os atributos e que possui um tipo de amostragem chamada de bootstrap, que possibilita melhor análise dos dados (CHEN et al., 2015). Segundo Feng et al. (2015) esse método de aprendizado de *ensemble* é utilizado em tarefas de classificação, regressão, funcionando por meio da construção de múltiplas árvores de decisão na fase de treinamento.

Para realizar uma previsão, o *Random Forest* considera a classe mais frequente (no caso de classificação) ou calcula a média das previsões (em casos de regressão) entre todas as árvores individuais.

O algoritmo J48 é utilizado para dados de classificação e trabalha com atributos discretos e contínuos, é um dos algoritmos mais utilizados de árvores de decisão. Esse algoritmo é compreendido por produzir um modelo classificador a partir de um conjunto de registro que já existe para que esse modelo seja utilizado para classificar outros exemplos de classes desconhecidas (LIBRELOTTO e MOZZAQUATRO, 2014). O algoritmo irá fazer a escolha do atributo que melhor particionar o conjunto em subconjuntos de dados, esse particionamento é definido como ganho de informações, depois de ser determinado um atributo com maior ganho, o mesmo atributo é selecionado para tomar a decisão e assim será repetido o procedimento nos subconjuntos até que seja criada uma árvore de decisão (RAMYA et al., 2015). No estudo realizado por Revathi e Hemalatha (2011) diversos algoritmos de aprendizado de máquina foram utilizados para classificar a qualidade de sementes de algodão. Após a comparação entre eles, o algoritmo J48 apresentou o melhor desempenho

5.6 Support Vector Machine (SVM)

É um algoritmo que pode ser usado para aplicações lineares e não lineares, baseado no princípio de encontrar o melhor limite (ou hiperplano). As principais características desse algoritmo é o uso de *kernels*, que são funções matemáticas que podem ser usadas para transformar os dados em um espaço de dimensão superior. Isso facilita que o SVM trabalhe com dados não lineares e faça previsões precisas. Esse algoritmo também possui dois parâmetros importantes: C e gama. C é um parâmetro de regularização que controla o compromisso entre maximizar a margem e minimizar o erro de regressão, e o gama é um parâmetro que controla a influência de cada exemplo de treinamento na função de decisão, ou seja, é importante sempre ajustar as funções do algoritmo (WANG et al., 2023; SARKAR, 2023).

5.7 Rede neural artificial (RNA)

É uma técnica computacional que adquirem conhecimento através da experiência, deste modo RNA é capaz de reconhecer padrões, ou seja, possui a capacidade de aprender por meio de exemplos e generalizar as informações aprendidas, criando um modelo não linear,

distribuídos em uma ou mais camadas e interligada por um grande número de conexões. Isso faz sua aplicação ser bastante eficiente para diversos tipos de análises (PONGHIRAN et al., 2019).

5.8 Redes *Perceptrons* de Múltiplas Camadas (MLP)

A estrutura de uma *Multilayer Perceptron* é formada por um conjunto de neurônios dispostos em camadas, contendo uma camada de entrada, uma ou mais camadas intermediárias, e uma camada de saída. Cada um dos neurônios da camada de entrada está conectado a todos os neurônios da camada intermediária. Da mesma forma, cada neurônio da camada intermediária está conectado com todos os neurônios da camada de saída (COUTINHO et al, 2016).

A camada de entrada da Figura 2 possui nós ($X_1...X_n$) como entrada dos dados, valores numéricos dependentes do conjunto de dados. Nenhum cálculo é realizado na camada de entrada, portanto seus valores inalterados são inseridos na camada oculta. A primeira camada intermediária, possui a mesma quantidade de nós da camada de entrada ($X_1...X_n$) e sua saída depende das saídas da camada de entrada bem como a saída dos pesos ($W_i...W_n$) associados às arestas (conexões), que são iniciados com valores iniciais aleatórios. Cada nó (neurônio) da segunda camada oculta combina os valores descritos pela camada anterior conectados a ele e executa cálculos semelhantes. A camada de saída recebe os valores da última camada oculta anterior e o resultado dos valores calculados (Y_i) atua como resultado da MPL (SOARES et al., 2015; SILVA e GRACIOSO, 2020).

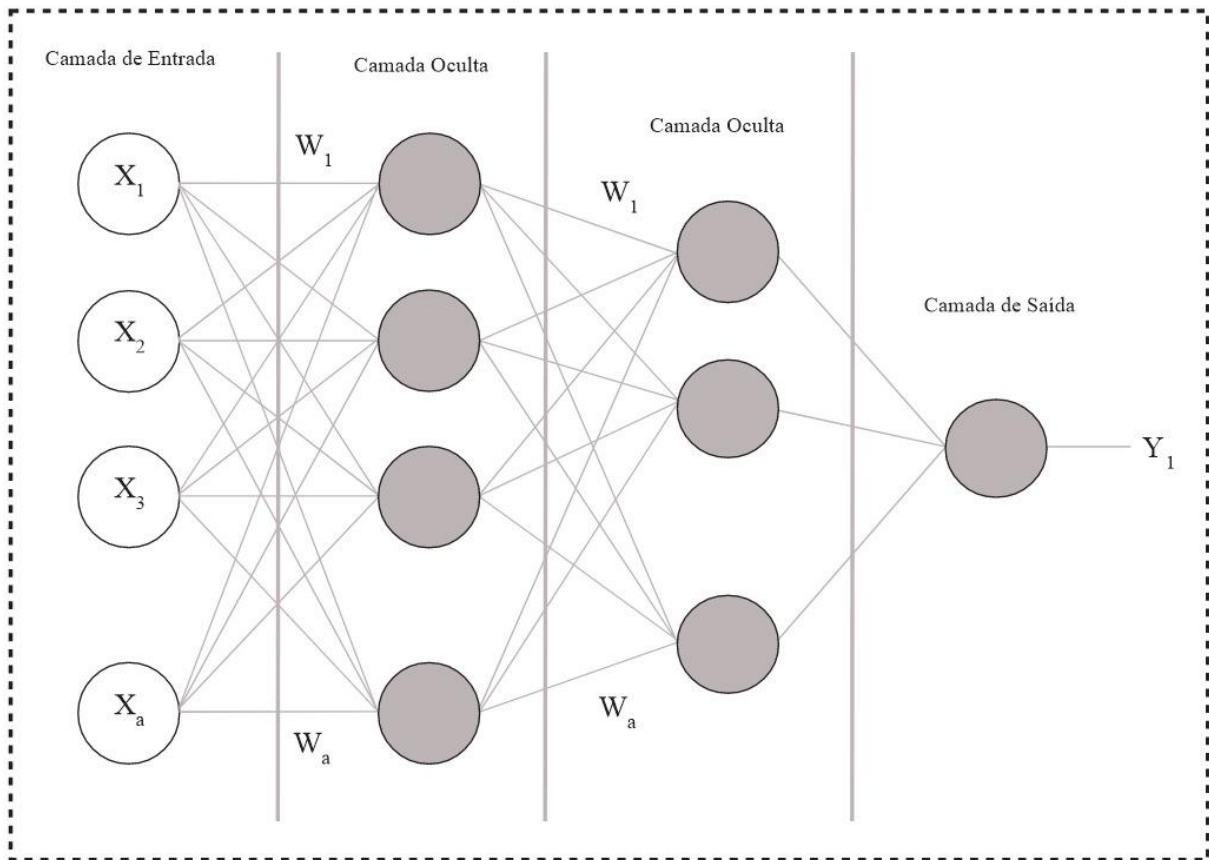


Figura 2. Rede MLP com duas camadas intermediárias

Fonte: adaptado por Silva e Gracioso (2020).

5.9 Gaussian Processes (GP)

Esse algoritmo é utilizado para modelar sistemas complexos não lineares e multimodais, bem como para lidar com amostras menores e espaços de entrada de alta dimensão. Sua principal vantagem é que esse algoritmo fornece uma forma baseada em princípios para quantificar e propagar a incerteza, que é evidente em aplicações na qual os dados são ruidosos ou sistemas subjacentes não são bem compreendidos.

Outra vantagem principal é que os GP é um método de aprendizado de máquina não paramétrico que o nível de expressividade aumenta com o tamanho do conjunto (RELAÑO et al., 2023; MUÑOZ et al., 2023). No estudo de Zibaei-Rad et al. (2024) sobre revestimentos comestíveis à base de mucilagem de semente, o algoritmo GP foi utilizado para prever diversos parâmetros, incluindo perda de peso, sólidos solúveis totais, teor de gordura, pH, teor

de carboidratos solúveis, viabilidade, compostos fenólicos totais, atividade antioxidante, presença de mofo e levedura, clorofilas totais, carotenoides totais, além de características sensoriais como cor, odor e aceitação geral. Os resultados indicaram uma excelente concordância entre os dados reais e os valores previstos.

6 REFERÊNCIAS

- AGELET, L. E; HURBURGH, C. R. A tutorial on near infrared spectroscopy and its calibration. **Critical Reviews in Analytical Chemistry**, v. 40, n. 4, p. 246-260, 2010.
- AL-GHRAIBAH, A et al. A study of feature selection of magnetogram complexity features in an imbalanced solar flare prediction data-set. In: **2015 IEEE international conference on data mining workshop (ICDMW)**. IEEE, p. 557-564, 2015.
- ANTUNES, C. G. C. et al. Desenvolvimento de mudas de catingueira em diferentes substratos e níveis de luminosidade. **Cerne**, v. 20, n. 1, p. 55- 60. 2014.
- BAZONI, C. H V. et al. Near-infrared spectroscopy as a rapid method for evaluation physicochemical changes of stored soybeans. **Journal of stored products research**, v. 73, p. 1-6, 2017.
- BERRAR, D. Cross-validation. 2019. Disponível em https://dberrar.github.io/papers/Berrar_EBCB_2nd_edition_Cross-validation_preprint.pdf. Acesso em: 23 de dez.2024
- BETEMPS, D. L. et al. Espectroscopia do visível e infravermelho próximo (vis/nir) na avaliação da qualidade de mangas Tommy Atkins. **Revista Brasileira de Fruticultura**, v. 33, p. 306-313, 2011.
- BORGHETTI, F.; FERREIRA, A. G. Interpretação de resultados de germinação. **Germinação: do básico ao aplicado**. Porto Alegre: Artmed, p. 209-222, 2004.
- BRASIL, Ministério da agricultura e Reformas Agrárias. **Regras para análises de sementes**. Brasília. DF: SNDA, 2009.365p. https://www.gov.br/agricultura/ptbr/assuntos/insumos299agropecuarios/arquivos-publicacoes-insumos/2946_regras_analise_sementes.pdf Acesso em: 05 ago, 2022.
- BRASIL. Ministério da Agricultura, Pecuária e Abastecimento. Instruções para análises de sementes florestais Brasília, DF, 2013. Disponível em: <https://www.gov.br/agricultura/pt-br/assuntos/lfd/a/arquivos-publicacoes-laboratorio/florestal_documento_pdf-ilovepdf-compressed.pdf>. Acesso em: 05 ago, 2022.
- BRERETON, Richard G. Pattern recognition in chemometrics. **Chemometrics and Intelligent Laboratory Systems**, v. 149, p. 90-96, 2015.
- BRUNING, B et al. Approaches, applications, and future directions for hyperspectral vegetation studies: an emphasis on yield-limiting factors in wheat. **The Plant Phenome Journal**, v. 3, n. 1, p. e20007, 2020.
- CAMPOS FILHO E. M, SARTORELLI P.A. R. **Guia de árvores com valor econômico**. 1ª. ed. São Paulo: Agroicone; 2015. 139p.
- CARDOSO, D. B. O. S.; QUEIROZ, L. P. Diversidade de Leguminosae nas Caatingas de Tucano, Bahia: implicações para a fitogeografia do semiárido do Nordeste do Brasil. **Rodriguésia**, v.58, n. 2. p. 379-391, 2007.

- CARVALHO, P. E. R. **Angico-Branco**. 2002 Disponível em: <https://www.infoteca.cnptia.embrapa.br/bitstream/doc/306306/1/CT0056.pdf>. Acesso em: 03 de fev. 2021.
- CHAGAS, I. P. **Desenvolvimento de um fotômetro portátil NIR para determinação do teor de Água no Álcool combustível e do teor de etanol na Gasolina**. 2006. 151 f. (Tese de doutorado) - Universidade Estadual de Campinas, São Paulo, 2006.
- CHEN, H. et al. Diagnosis of colorectal cancer by near-infrared optical fiber spectroscopy and random forest. **Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy**, v. 135, p. 185-191, 2015.
- CHENG, Y. et al. Machine learning for metabolic pathway optimization: A review. **Computational and Structural Biotechnology Journal**, v.21, p. 2381-2393, 2023.
- COATES, J et al. Interpretation of infrared spectra, a practical approach. **Encyclopedia of analytical chemistry**, v. 12, p. 10815-10837, 2000.
- CHENG, T. et al. Hyperspectral and imagery integrated analysis for vegetable seed vigor detection. **Infrared Physics & Technology**, v. 131, p. 104605, 2023. <https://doi.org/10.1016/j.infrared.2023.104605>
- COSTA, R.C. **Metodologias analíticas para a agroindústria de frutas empregando infravermelho próximo e quimiometria**. 2018. 109f. (Tese de doutorado) - Universidade federal do Rio grande do Norte, Natal, 2018.
- COUTINHO, E. R. et al. Utilização de técnicas de inteligência computacional na predição de dados meteorológicos. **Revista Brasileira de Meteorologia**, v. 31, n. 1, p. 24-36, 2016.
- DANTAS, B. F. et al. Taxas de crescimento de mudas de catingueira submetidas a diferentes substratos e sombreamentos. **Revista árvore**, v. 33, n. 3, p. 413-423, 2009.
- DAVIES, A. M. C; GRANT, Review: Near infra-red analysis of food. **International Journal of Food Science & Technology**, v. 22, n. 3, p. 191-207, 1987.
- DÍAZ, G. et al. Co-training for visual object recognition based on self-supervised models using a cross-entropy regularization. **Entropy**, v. 23, n. 4, p. 423, 2021.
- DURMUS, D.; DAVIS, W. Optimising light source spectrum for object reflectance. **Optics express**, v. 23, n. 11, p. A456-A464, 2015.
- ELLIOTT, A.; AMBROSE, E. J. Evidence of chain folding in polypeptides and proteins. **Discussions of the Faraday Society**, v. 9, p. 246-251, 1950.
- FAN, Y. et al. Individual wheat kernels vigor assessment based on NIR spectroscopy coupled with machine learning methodologies. **Infrared Physics & Technology**, Yangling, Shaanx, v. 105, p. 103213, 2020.
- FENG, Z et al. A Random Forest-based ensemble method for activity recognition. In: **2015 37th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)**. IEEE, 2015. p. 5074-5077.
- FERRÃO, M.F. et al. Determinação simultânea dos teores de cinza e proteína em farinha de trigo empregando NIR-PLS e DRIFT-PLS. **Ciência e Tecnologia de Alimentos**, v. 24, n. 3, p. 333-340, 2004.
- FERREIRA, M. et al. Quimiometria I: calibração multivariada, um tutorial. **Química Nova**, Campinas, v. 22, n. 5, p. 724-731, 1999.
- FERREIRA, M.M. C. **Quimiometria- Conceitos, Métodos e Aplicações**. 1 ed. Campinas, SP: Editora da Unicamp, 2015.493p.
- FERREIRA, I. J.S. et al. Determination of quality and ripening stages of ‘pacovan’bananas using vis-NIR spectroscopy and machine learning. **Engenharia Agrícola**, v. 42, p. e20210160, 2022
- FOWLER, J.A.P; MARTINS E.G. Manejo de sementes de espécies florestais. Colombo, **Embrapa**. 71 p, 2001.

GADOTTI, G. I. et al. Prediction of ranking of lots of corn seeds by artificial intelligence. **Engenharia Agrícola**, v. 42, n. 4, p. e20210005, 2022.

GAEM, P. H. *Cenostigma* in Flora e Funga do Brasil. 2020. Jardim Botânico do Rio de Janeiro. Disponível em: <https://floradobrasil.jbrj.gov.br/FB606076>. Acesso em: 01 ago. 2023.

GALLO, L.V. **Determinação de parâmetros de qualidade em bananas utilizando espectroscopia no infravermelho e calibração multivariada**. 2008. 63f. (Dissertação de mestrado) - Universidade Estadual de Campinas. Campinas, São Paulo.

GHOJOGH, B; CROWLEY, M. The theory behind overfitting, cross validation, regularization, bagging, and boosting: tutorial. **arXiv preprint arXiv:1905.12787**, 2019.

GOLDFARB, M et al. Teor de água limite para criopreservação das sementes de pinhão manso (*Jatropha curcas* L.). **Revista Brasileira de Produtos Agroindustriais**, v. 10, n. 2, p. 121-129, 2008.

GOMES, S. E. V. et al. Influence of current and future climate on the seed germination of *Cenostigma microphyllum* (Mart. ex G. Don) E. Gagnon & GP Lewis. **Folia Geobotanica**, v. 54, p. 19-28, 2019.

GRANJA, J. A. A. et al. Aspectos morfológicos, germinação e análise do espectro de infravermelho próximo (NIR) de *Phyllanthus niruri* em diferentes colorações da semente. **Revista Brasileira de Plantas Medicinais**, v. 20, p. 141-149, 2018.

GRIFFITHS, P. R. Fourier transform infrared spectrometry. **Science**, v. 222, n. 4621, p. 297-302, 1983.

GUO, Z. et al. Application of visible-near-infrared hyperspectral imaging technology coupled with wavelength selection algorithm for rapid determination of moisture content of soybean seeds. **Journal of Food Composition and Analysis**, v. 116, p. 105048, 2023.

HUANG, X. et al. Rapid measurement of total polyphenols content in cocoa beans by data fusion of NIR spectroscopy and electronic tongue. **Analytical Methods**, v. 6, n. 14, p. 5008-5015, 2014.

JHA, S.N. et al. Nondestructive prediction of maturity of mango using near infrared spectroscopy. **Journal of Food Engineering**, v.124, p.152 -157, 2014.

KRAWCZYK, B. Weighted one-class classification for different types of minority class examples in imbalanced data. In: **2014 IEEE Symposium on Computational Intelligence and Data Mining (CIDM)**. IEEE, 2014. p. 337-344.

KRELING, B.E. et al. Propriedades físicas, composição e teor de água em grãos. In: SILVA-MATOS, R.R.S.S. et al. **Desenvolvimento rural e processos sociais nas ciências agrárias**. Ponta Grossa: Atena, 2022. p.270-280.

KRZYŻANOWSKI, F.C. and FRANÇA-NETO, J.B. Vigor de Sementes. [Seed Vigour]. **Informativo ABRATES**, Londrina, v.11, p.81-84, 2001.

LEI, N et al. Suomi NPP VIIRS solar diffuser screen transmittance model and its applications. **Applied Optics**, v. 56, n. 31, p. 8676-8686, 2017.

LIBRELOTTO, S. R; MOZZAQUATRO, P. M. Análise dos algoritmos de mineração J48 e Apriori aplicados na detecção de indicadores da qualidade de vida e saúde. **Revista Interdisciplinar de Ensino, Pesquisa e Extensão-RevInt**, v. 1, n. 1, 2014.

LIMA JR, M.J. Determinação do grau de umidade. in: LIMA JR., M.J.; FIGLIOLIA, M.B. PIÑA-RODRIGUES, F.C.M.; GENTIL, D.F.O. SOUZA, M.M.; SILVA, V.S. **Manual de procedimentos para análise de sementes florestais Manaus: UFAM**, 2010. p.39-53.

LIMA, C. R. et al. Phenology of *Poincianella pyramidalis* (Tul.) L. P. Queiroz and its relationship with temporal distribution of rainfall the Brazilian semi-arid region. **Ciência Florestal**, v. 28, n. 3, p. 1035–1048, 2018.

LIU, D; GUO, W. Identification of kiwifruits treated with exogenous plant growth regulator using near-infrared hyperspectral reflectance imaging. **Food Analytical Methods**, v. 8, n.1, p. 164-172, 2015.

LÓPEZ, V et al. An insight into classification with imbalanced data: Empirical results and current trends on using data intrinsic characteristics. **Information sciences**, v. 250, p. 113-141, 2013.

MACINNIS-NG, Cate et al. Climate-change impacts exacerbate conservation threats in island systems: New Zealand as a case study. **Frontiers in Ecology and the Environment**, v. 19, n. 4, p. 216-224, 2021.

MARCOS FILHO, J. Seed vigor testing: an overview of the past, present and future perspective. **Scientia agricola**, v. 72, n.4, p. 363-374, 2015.

MARENGO, J. A. Água e mudanças climáticas. **Estudos avançados**, v. 22, p. 83-96, 2008.

MARENGO, J. A. O futuro clima do Brasil. **Revista USP**, n. 103, p. 25-32, 2014.

MARIN, M. C. et al. Caracterização e classificação do tráfego da darknet com modelos baseados em árvores de decisão. In: Anais do XXXIX Simpósio Brasileiro de Redes de Computadores e Sistemas Distribuídos. SBC.Juiz de Fora, MG. 2021, p. 127-140.

MARQUES, E. J.N. et al. Rapid and non-destructive determination of quality parameters in the ‘Tommy Atkins’ mango using a novel handheld near infrared spectrometer. **Food Chemistry**, v.197. p.1207-1214, 2016.

MARQUES, J. N. **Desenvolvimento de métodos analíticos não destrutivos para avaliação da qualidade de frutas utilizando espectroscopia e imagem hiperespectral na região do infravermelho próximo**. 2017. 185 f. (Tese de doutorado) - Universidade Federal de Pernambuco. Recife, 2017.

MATIAS, J. R.; SILVA, F. F. S.; DANTAS, B. F. Catingueira-verdadeira *Poincianella pyramidalis* [Tul.]. LP Queiroz. **Associação brasileira de tecnologias de sementes. Comitê técnico de sementes florestais, Embrapa semiárido-nota técnica**, n. 6, 2017.

MATIAS, J.R. et al. Catingueira-verdadeira *Poincianella pyramidalis* [Tul.] L.P.Queiroz; **informativo abrates**.v.29, n. 1, 2, 3,p. 28-33, 2019.

MATOS, G. R. **Machine learning aplicado à gestão de activos físicos industriais**. 89 f. (dissertação de mestrado), Instituto Superior de Engenharia de Lisboa, 2021.

MEDEIROS, M.L.S. **Aplicação de espectroscopia no infravermelho próximo e imagens hiperespectrais para quantificar o teor de óleo e classificar sementes do gênero:BRASSICA=Application of Near Infrared spectroscopy and hyperspectral imaging to quantify the oil content and classify Brassica seeds**. 2021. 88f, (Dissertação de Mestrado) -Universidade Estadual de Campinas, São Paulo, 2021.

MEIRA, C. A. A et al. Análise da epidemia da ferrugem do cafeeiro com árvore de decisão. **Tropical Plant Pathology**, v. 33, n.2, p. 114-124, 2008.

MENDONÇA, A. V. R. et al. Morfologia de frutos e sementes e germinação de *Poincianella pyramidalis* (Tul.) LP Queiroz, comb. Nov. **Ciência Florestal**, v. 26, n. 2, p. 375-387, 2016.

MORAES, F. P. D. **Abordagem quimiométrica e avaliação físico-química, bioativa e biológica in vitro da acerola (*Malpighia emarginata*) in natura e liofilizada**. 2018. 171 f. (Tese de doutorado) - Universidade Federal do Rio Grande do Norte, Natal, 2018.

MUKASA, P. et al. Determination of viability of Retinispora (Hinoki cypress) seeds using FT-NIR spectroscopy. **Infrared Physics & Technology**, v. 28, p. 62-68, 2020.

MUÑOZ, J. et al. Gaussian processes and Fast Marching Square based informative path planning. **Engineering Applications of Artificial Intelligence**, v. 121, p. 106054, 2023.

MURPHY, K.P. **Machine Learning: A Probabilistic Perspective**. Califonia: Cambridge, MIT press, 2012.1049p.

NASCIMENTO, J.P. B.et al. Angico *Anadenanthera colubrina* varo cebil (Vell.) Brenan. **Informativo ABRATES**, v.29, n.1,2019.

NETO, J. B. F; KRZYZANOSKI, F.C. **Qualidade de sementes**. Disponível em<<https://www.alice.cnptia.embrapa.br/bitstream/doc/578559/1/Progressosnoconceitod equalidadedesementesnoBrasil.pdf>>. Acesso em: 30 jul. 2023.

NICOLAI, B. M. et al. Nondestructive measurement of fruit and vegetable quality by means of NIR spectroscopy: A review. **Postharvest biology and technology**, v. 46, n. 2, p. 99-118, 2007.

NOGUEIRA, A. C.; MEDEIROS, AC de S. Coleta de sementes florestais nativas. **Embrapa Florestas Circular Técnica (INFOTECA-E)**. Disponível em:<https://www.infoteca.cnptia.embrapa.br/bitstream/doc/293783/1/Circular144.pdf>. Acesso em: 03. mar. 2021.

OLIVEIRA, J. A., ROSA; S. D. V. F; Carvalho, E. R. Água nas sementes. In: OLIVEIRA, J. A. (Org.). **Secagem de sementes**. UFLA.2021. 67-93.

OLIVEIRA, L. M. et al. Qualidade fisiológica de sementes de *caesalpinia pyramidalis* tul: durante o armazenamento. **Revista brasileira de sementes**, v. 33, n. 2, p. 289-298, 2011.

OLIVEIRA, M. M. et al. Portable near-infrared spectroscopy for rapid authentication of adulterated paprika powder. **Journal of Food Composition and Analysis**, v. 87, p. 103403, 2020.

PAREYN, F.G.C. et al *Anadenanthera Columbrina* (Angico). 2018. Disponível em: <https://www.infoteca.cnptia.embrapa.br/infoteca/bitstream/doc/1103448/1/LivroNordeste740745.2018.pdf> . 2018. Acesso em 20 fev.2021.

PASQUINI, C. Near infrared spectroscopy: A mature analytical technique with new perspectives—A review. **Analytica Chimica Acta**, Campinas v. 1026, p. 8-36, 2018.

PASQUINI, C. Near infrared spectroscopy: fundamentals, practical aspects and analytical applications. **Journal of the Brazilian chemical society**, Campinas v. 14, n. 2, p. 198-219, 2003.

PEIRIS, K. H. S. et al. Spatial variability of soluble solids or dry-matter content within individual fruits, bulbs, or tubers: implications for the development and use of NIR spectrometric techniques. **HortScience**, v. 34, n. 1, p. 114-118, 1999.

PONGHIRAN, W. et al Reinforcement learning with low-complexity liquid state machines. **Frontiers in Neuroscience**, v. 13, p. 465263, 2019.

POZZER, C. T. **Aprendizado por árvores de decisão**.2006.<https://silo.tips/download/aprendizado-por-arvores-de-decisao>. Acesso em: 18 out. 2023.

QUEIROZ L.P. et al. **Towards greater knowledge of the Brazilian Semi-arid Biodiversity**. Brasília: Ministério da Ciência e Tecnologia, 2006, 140p.

RAMYA, M. C. et al. A predictive model construction for mulberry crop productivity. **Procedia Computer Science**, v. 45, p. 156-165, 2015.

RELAÑO, C. et al. Gaussian process regression for forward and inverse kinematics of a soft robotic arm. **Engineering Applications of Artificial Intelligence**, v. 126, p. 107174, 2023.

REVATHI, P; HEMALATHA, M. Categorize the quality of cotton seeds based on the different germination of the cotton using machine knowledge approach. **International Journal of Advanced Science and Technology**, v. 36, p. 9-14, 2011.

RIBEIRO, J. P. O. et al. FT-NIR and linear discriminant analysis to classify chickpea seeds produced with harvest aid chemicals. **Food Chemistry**, v.342, p.128324, 2021.

SAGAR, P et al. A forecasting method with efficient selection of variables in multivariate data sets. **International Journal of Information Technology**, v. 13, n.3, p. 1039-1046, 2021.

SANTANA, D.G; RANAL, M.A. **Análise da Germinação: um enfoque estatístico**. Brasília: Editora UNB, 2004. 228p.

SANTOS, W. S. et al. Produção de lenha e forragem de *Poincianella pyramidalis* (Tul.) LP Queiroz submetida à poda anual. **Ciência Florestal**, v. 30, n. 1, p. 89-103, 2020.

SHAO, Y et al. Infrared spectroscopy and chemometrics for the starch and protein prediction in irradiated rice. **Food chemistry**, v. 126, n. 4, p. 1856-1861, 2011.

SARKAR, S. et al. Soybean seed composition prediction from standing crops using PlanetScope satellite imagery and machine learning. **ISPRS Journal of Photogrammetry and Remote Sensing**, v. 204, p. 257-274, 2023.

SATO, L. Y. et al. Análise comparativa de algoritmos de árvore de decisão do sistema WEKA para classificação do uso e cobertura da terra. In: Anais do **XVI Simpósio Brasileiro de Sensoriamento Remoto**. 2013. Foz do Iguaçu, PR, 2013 p. 2353-2360, 2013.

Secagem de sementes. Lavras. Editora UFLA. 2021.p,69-70.

SERSON, W; et al. Development of Whole and Ground Seed Near-Infrared Spectroscopy Calibrations for Oil, Protein, Moisture, and Fatty Acids in *Salvia hispanica*. **Journal of the American Oil Chemists' Society** v.97, n.1, p.3-13,2020.

SILVA F. S. et al. Dynamics of traditional knowledge of medicinal plants in a rural community in the Brazilian semi-arid region. *Braz J Pharmacogn*. V. 21, n.3, p.382–391, 2011.

SILVA NETA, M.L. **Recomendação de espectrômetros NIR portáteis para monitoramento da qualidade de mangas ‘keitt’ produzidas no submédio do vale do são francisco** 2019. 76f. (Dissertação de mestrado) - Universidade Federal de Sergipe, Petrolina, 2019.

SILVA, E.M. et al. Diagnóstico do uso de leguminosas em propriedades rurais no município de Aparecida-PB. **Revista Verde de Agroecologia e Desenvolvimento Sustentável**, v. 7, n. 3, p. 212-217, 2012.

SILVA, J. N. et al. Testes de envelhecimento acelerado e condutividade elétrica para Sementes de Espécies Florestais Nativas: Uma breve revisão. **Meio Ambiente (Brasil)**, v. 1, n. 2, p.24-30, 2020.

SILVA, L.B. et al. Anatomia e densidade básica da madeira de *Caesalpinia pyramidalis* Tul. (Fabaceae), espécie endêmica da caatinga do Nordeste do Brasil. **Acta botânica brasílica**, v. 23, n. 2, p. 436-445, 2009.

SILVA, M.V; GRACIOSO, A. C. N. R. Modelo de previsão de demanda para ATMS utilizando rede neural artificial do tipo multilayer perceptron. In: **Colloquium Exactarum**. v.12, n.4, p. 54-62,2020.

SOARES, F. C. et al. Predição da produtividade da cultura do milho utilizando rede neural artificial. **Ciência Rural**, v. 45, n. 11, p. 1987-1993, 2015.

SOUZA, J. S. **Desenvolvimento de modelos de calibração multivariada em espectroscopia de infravermelho próximo para ácidos graxos em amostras de carne bovina**. 2016. 134f (Dissertação de mestrado) - Universidade Federal de São Carlos, São Carlos, São Paulo, 2016.

SOUZA, D.M. et al. Aplicação de técnicas multivariadas e inteligência artificial na análise de espectros de infravermelho para determinação de matéria orgânica em amostras de solos. **Química Nova**, v. 35, n.9, p. 1738-1745, 2012.

SUN, Y et al. Classification of imbalanced data: A review. **International journal of pattern recognition and artificial intelligence**, v. 23, n. 04, p. 687-719, 2009.

TAKEDA, S et al. Inter-reflection compensation of immersive projection display by spatio-temporal screen reflectance modulation. **IEEE transactions on visualization and computer graphics**, v. 22, n. 4, p. 1424-1431, 2016.

TANG, J. et al. Exploring AdaBoost and Random Forests machine learning approaches for infrared pathology on unbalanced data sets. **Analyst**, v. 146, n. 19, p. 5880-5891, 2021.

TEYE, E. et al. Innovative and rapid analysis for rice authenticity using hand-held NIR spectrometry and chemometrics. **Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy**, v. 217, p.147-154, 2019.

TIBOLA, C. S. et al. **Espectroscopia no infravermelho próximo para avaliar indicadores de qualidade tecnológica e contaminantes em grãos**. 1. ed. Brasília: Embrapa, 2018. 159 p.

TIGABU, M. et al. Application of near-infrared spectroscopy for the detection of internal insect infestation in *Picea abies* seed lots. **Canadian Journal of Forest Research**, v. 34, n. 1, p. 76-84, 2004.

TIGABU, M. et al. Germination of *Juniperus procera* seeds in response to stratification and smoke treatments, and detection of insect-damaged seeds with VIS+ NIR spectroscopy. **New Forests**, v. 33, p. 155-169, 2007.

TIGABU, M. et al. Rapid and non-destructive evaluation of seed quality of Chinese fir by near infrared spectroscopy and multivariate discriminant analysis. **New Forests**, v. 51, p. 395-408, 2020.

TRESENA, N. L. et al. Determinação do teor de água limite para criopreservação das sementes de ipê amarelo (*Tabebuia chrysotrica* (Mart. Ex. DC.) Standl.). **Cerne**, v. 16, p. 171-175, 2010.

VALDERRAMA, P. **Avaliação de figuras de mérito em calibração multivariada na determinação de parâmetros de controle de qualidade em indústria alcooleira por espectroscopia no infravermelho próximo**. 2005. 135 f. (Dissertação de mestrado) - Universidade Estadual de Campinas, Campinas- SP, 2005.

VÁZQUEZ-MARRUFO, M. et al. A systematic review of the application of machine-learning algorithms in multiple sclerosis. **Neurología (English Edition)**, v.38 p. 577-590, 2023.

WANG Z et al. Application of long-wave near infrared hyperspectral imaging for determination of moisture content of single maize seed. **Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy** v.254, p.119666,2021.

WANG, Z. et al. Assessment of protein content and insect infestation of maize seeds based on on-line near-infrared spectroscopy and machine learning. **Computers and Electronics in Agriculture**, v. 211, p. 107969, 2023.

WANG, Z. et al. Assessment of protein content and insect infestation of maize seeds based on on-line near-infrared spectroscopy and machine learning. **Computers and Electronics in Agriculture**, v. 211, p. 107969, 2023.

WORKMAN JR, J. J. Interpretive spectroscopy for near infrared. **Applied Spectroscopy Reviews**, v. 31, n. 3, p. 251-320, 1996.

XAVIER, B. A.; CHEN, Po-Hao. Natural language processing for imaging protocol assignment: machine learning for multiclass classification of abdominal CT protocols using indication text data. **Journal of Digital Imaging**, v. 35, n. 5, p. 1120-1130, 2022.

YU, Y. et al. Cross target attributes and sample types quantitative analysis modeling of near-infrared spectroscopy based on instance transfer learning. **Measurement**, v. 177, p. 109340, 2021.

ZHANG Y.; GUO, W. Moisture content detection of maize seed based on visible/near-infrared and near-infrared hyperspectral imaging technology. **International Journal of Food Science & Technology**, v.55, n.2 631-640,2020.

ZIBAEI-RAD, A et al. Probiotic-loaded seed mucilage-based edible coatings for fresh pistachio fruit preservation: an experimental and modeling study. **Scientific Reports**, v. 14, n. 1, p. 509, 2024

CAPÍTULO 1

**PREDIÇÃO NÃO DESTRUTIVA DO TEOR DE ÁGUA DE SEMENTES
FLORESTAIS NATIVAS DA CAATINGA UTILIZANDO ESPECTROSCOPIA VIS-
NIR E APRENDIZAGEM DE MÁQUINA**

Capítulo 1: Predição não destrutiva do teor de água de sementes florestais nativas da caatinga utilizando espectroscopia Vis-NIR e aprendizagem de máquina

Resumo: A determinação do teor de água (TA) é crucial para a escolha dos métodos mais adequados à conservação das sementes. O uso do espectrômetro Vis-NIR permite a avaliação do TA sem causar danos ao material, em contraste com os métodos convencionais. Este estudo teve como objetivo desenvolver modelos baseados em algoritmos de Aprendizagem de Máquina (AM) para estimar o TA de sementes de duas leguminosas florestais nativas, utilizando um espectrômetro Vis-NIR. Foram formados sublotes com diferentes teores de água, e as sementes foram analisadas espectralmente. Foi aplicado o filtro de reamostragem e testados quatro algoritmos discriminativos. Para o TA das sementes de *Anadenanthera colubrina* o algoritmo *Random Forest* (RF) teve o melhor ajuste com $R^2 = 0.35$ e 0.86 de RMSE= 7,24 e 3,32 e MAE=5,08 e 2,09. Para o TA das sementes de *Cenostigma pyramidale*, o algoritmo *Multilayer Perceptron* (MP) apresentou o melhor desempenho com $R^2 = 0,90$ e $0,95$ e obteve menores erros. Com aplicação do *resample* o RF demonstrou melhor ajuste para o TA das duas espécies R^2 acima de $0,70$ e menores erros. Logo o Vis-NIR, associado ao uso de algoritmos de AM, pode ser aplicado para predição do TA de sementes florestais nativas.

Palavras chaves: inovação, tecnologia, umidade, *Anadenanthera colubrina* e *Cenostigma pyramidale*

Chapter 1: Non-destructive prediction of the water content of native caatinga forest seeds using machine learning and Vis-NIR spectroscopy

Abstract: The determination of water content (WC) is crucial for choosing the most appropriate methods for seed conservation. The use of the Vis-NIR spectrometer allows the evaluation of WC without causing damage to the material, in contrast to conventional methods. This study aimed to develop models based on Machine Learning (ML) algorithms to estimate the WC of seeds of two native forest legumes, using a Vis-NIR spectrometer. Sublots with different water contents were formed, and the seeds were spectrally analyzed. The resampling filter was applied and four discriminative algorithms were tested. For the WC of *Anadenanthera colubrina* seeds, the Random Forest (RF) algorithm had the best fit with $R^2 = 0.35$ and 0.86 , $RMSE = 7.24$ and 3.32 and $MAE = 5.08$ and 2.09 . For the TA of *Cenostigma pyramidale* seeds, the Multilayer Perceptron (MP) algorithm presented the best performance with $R^2 = 0.90$ and 0.95 and obtained smaller errors. With the application of resample, the RF demonstrated a better fit for the TA of both species, R^2 above 0.70 and smaller errors. Therefore, Vis-NIR, associated with the use of ML algorithms, can be applied to predict the TA of native forest seeds.

Keywords: innovation, technology, humidity, *Anadenanthera colubrina* e *Cenostigma pyramidale*

1 INTRODUÇÃO

O conhecimento do teor de água (TA) das sementes é essencial para orientar os processos de pós-colheita, como secagem e armazenamento, garantindo a qualidade física, fisiológica e sanitária. Além disso, o teor de água é um importante indicador da qualidade das sementes, influenciando tanto o tempo de armazenamento quanto a taxa de germinação (NERY et al., 2004; XU et al., 2019). Controlar o teor de água nas sementes é essencial para prevenir danos durante o processo de secagem, minimizar riscos de deterioração e garantir um armazenamento adequado. Além disso, esse controle contribui para o manejo e transporte mais seguros, favorecendo a germinação eficiente e o desenvolvimento saudável das plantas. (RABANERA et al., 2021; REZAEI et al., 2023).

O teor de água elevado pode intensificar a respiração das sementes, o que, por sua vez, acelera o processo de deterioração, levando à perda da atividade enzimática e ao comprometimento das membranas celulares, resultando na redução do vigor das sementes (KONG et al., 2015; BONJOVANI e BARBEDO, 2019). Por outro lado, quando o teor de água das sementes é muito baixo, há um aumento no risco de deterioração. Isso ocorre devido à desidratação excessiva, que pode afetar componentes celulares vitais, como membranas, enzimas e até o material genético. Esse ressecamento intenso compromete a estrutura celular, reduzindo a viabilidade e dificultando a germinação. Além disso, sementes excessivamente secas ficam mais vulneráveis a danos físicos, como rachaduras, durante o transporte e o armazenamento (BALOUCHI et al., 2017; AMORIM et al., 2021; ZHANG et al., 2022).

Diante disso, o monitoramento periódico e adequado do teor de água (TA) ao longo de todas as etapas, da colheita ao processamento, armazenamento, comercialização e utilização na semeadura direta ou na produção de mudas, é fundamental. Esse acompanhamento permite identificar e prevenir problemas que poderiam comprometer a qualidade das sementes em qualquer fase do manejo (PINA-RODRIGUES et al., 2015; ZHANG E GUO, 2019). Ressalta-se que o TA apropriado para o armazenamento de algumas espécies florestais da Caatinga pode variar entre 13% e 5% (DANTAS et al., 2009; LIMA et al., 2012; NASCIMENTO et al., 2018).

A metodologia padrão utilizada para determinar o teor de água (TA) das sementes consiste na secagem em estufa a 105 ± 3 °C por um período de 24 horas, conforme estabelecido pelas Regras para Análise de Sementes (BRASIL, 2013) e pela ISTA (2025). Esse método é amplamente recomendado para diferentes tipos de sementes, por ser reconhecido como referência internacional.

No entanto, apesar de sua padronização, diversos fatores podem comprometer a precisão dos resultados obtidos, como: o tamanho das sementes, erros na pesagem antes ou após a secagem, variações na temperatura da estufa, além de possíveis reações químicas, como oxidação ou volatilização de compostos, que podem ocorrer durante o processo. Outro aspecto importante é que o método é considerado destrutivo, ou seja, as sementes não podem ser reutilizadas após a análise, o que representa uma perda de amostras. Além disso, trata-se de um procedimento relativamente lento, o que pode ser uma limitação em contextos que exigem maior agilidade.

A padronização dos procedimentos para detectar o teor de água é essencial para garantir a qualidade das sementes de espécies florestais nativas. A espectroscopia na região Vis-NIR tem se destacado como uma alternativa eficiente, pois permite a análise rápida e precisa do teor de água sem os processos tradicionais de secagem ou pesagem. Esse método não invasivo utiliza a interação da luz infravermelha com as moléculas de água, fornecendo dados em tempo real, desde que o equipamento seja adequadamente calibrado para as características específicas das sementes analisadas.

Desse modo, o uso da espectroscopia na região do visível e infravermelho próximo (Vis-NIR) surge como uma alternativa rápida, prática e precisa para determinar o teor de água das amostras, sem prejudicar as sementes (WANG et al., 2021). A espectroscopia Vis-NIR é uma tecnologia não destrutiva, não invasiva, que exige pouco preparo da amostra e dispõe de um instrumental simples, o que permite seu transporte a longas distâncias (MUNIZ et al., 2012; GUO et al., 2022). Além disso, a utilização dessa técnica para avaliar o teor de água das sementes permite que, antes e durante a colheita, já se tenha o conhecimento inicial dessa característica, evitando a destruição das amostras como ocorre no método padrão. Essa abordagem pode otimizar as análises, reduzindo o tempo necessário para a avaliação de sementes florestais, além de preservar o material. No entanto, para se obter resultados promissores com a espectroscopia Vis-NIR, é necessário desenvolver um modelo de calibração validado, o que assegura a capacidade preditiva do espectrômetro.

A utilização de algoritmos de *machine learning* (ML) pode contribuir para a identificação de um modelo de calibração que possibilite a determinação do teor de água de espécies florestais (ZHANG et al., 2018; SERSON et al., 2019; LARIOS et al., 2020). O desenvolvimento de trabalhos que utilizam o espectrômetro Vis-NIR com auxílio de diferentes algoritmos para avaliar o TA das sementes de espécies florestais pode auxiliar o

processo de tomada de decisão, na coleta, armazenamento e conservação das sementes até os processos de restauração.

A principal hipótese deste estudo é que um método não destrutivo, baseado em espectroscopia Vis-NIR e algoritmos de aprendizado de máquina, é capaz de avaliar com precisão o teor de água nas sementes de duas espécies de leguminosas florestais nativas do bioma Caatinga.

Objetivou-se desenvolver um método não destrutivo para avaliar o teor de água de sementes de duas espécies de leguminosas florestais nativas do bioma Caatinga utilizando espectroscopia Vis-NIR e algoritmos de aprendizagem de máquinas.

2 MATERIAL E MÉTODOS

2.1 Coleta e armazenamento das sementes avaliadas

Foram utilizadas sementes de *Anadenanthera colubrina* (Vell.) Brenan e *Cenostigma pyramidale* (Tul.) Gagnon & G.P. Lewis provenientes de diferentes populações e coletadas em diferentes anos, armazenadas em câmara fria a 10 ± 1 °C e $20\pm 5\%$ de umidade relativa até a montagem do experimento. As sementes de *A. colubrina* foram coletadas nos distritos Jutaí, em Lagoa Grande ($8^{\circ} 33' 35.7''$ S, $40^{\circ} 12' 1.9''$ W) (nos anos de 2017, 2019 e 2022), Orocó ($08^{\circ} 33' 12''$ S, $39^{\circ} 27' 48''$ W) (2022) e Sertânia ($08^{\circ} 04' 25''$ S, $37^{\circ} 15' 52''$ W) (2020) em Pernambuco, assim como no município de Juazeiro, BA ($09^{\circ} 24' 42''$ S- $40^{\circ} 29' 55''$ W) (2021). As sementes de *C. pyramidale* foram coletadas nos povoados de Salitre ($9^{\circ} 31' 15''$ S, $40^{\circ} 15' 15''$ W) (2015, 2016) e Massaroca ($9^{\circ} 51' 55''$ S - $40^{\circ} 16' 57''$ O) (nos anos de 2015, 2016 e 2017), no município de Juazeiro, BA (2022).

2.2 Obtenção de lotes com diferentes teores de água

A partir dos lotes de sementes armazenadas, foram formados sublotes com diferentes teores de água por meio de processos controlados de hidratação ou desidratação parcial. Para isso, foram utilizadas de *A. colubrina* e sementes de *C. pyramidale*, dispostas sobre telas em caixas gerbox contendo 40 mL de água (para hidratação) ou sílica gel (para desidratação), durante períodos de 24 e 72 horas. Com esse procedimento, foram obtidos seis sublotes de sementes de *A. colubrina* e cinco sublotes de *C. pyramidale*, cada um com teores de água distintos, determinados pelo método de referência com secagem em estufa a 105 °C por 24 horas (BRASIL, 2009) (Tabela 1).

Tabela 1. Lotes e sublotes do teor de água das sementes de *Anadenanthera colubrina* e *Cenostigma pyramidale*

Teor de água (%)											
<i>A. colubrina</i>						<i>C. pyramidale</i>					
Desidratação			Hidratação			Desidratação			Hidratação		
Controle						Controle					
Lotes	72h	24h		72h	24h	Lotes	72h	24h		72h	24h
J-2017	5,14	5,50	12,25	33,86	15,77	M- 2015	4,74	5,3	10,33	26,71	16,05
J-2019	4,97	9,65	12,39	35,66	19,13	M-2016	5,13	5,7	10,80	30,97	13,80
N- 2020	5,94	5,56	10,50	37,19	26,90	M-2017	5,30	5,9	10,50	26,71	12,63
N- 2021	5,72	5,59	10,50	29,05	25,09	S- 2015	4,07	6,3	10,54	32,24	20,88
J - 2022	4,65	5,75	7,40	38,89	26,29	S-2016	4,85	6,4	9,68	27,30	16,61
O-2022	4,86	9,94	10,38	37,49	23,63						

2.3 Aquisição dos espectros Vis-NIR

Foram utilizadas, para a avaliação individual com Vis-NIR (TA_{NIR}) e análises de referência (TA_{REF}) 2.312 sementes de *A. colubrina* e 2.907 sementes de *C. pyramidale* (Tabela 2). Os dados espectrais foram coletados com o espectrômetro portátil F-750 Produce Quality Meter (Felix Instruments, Estados Unidos), que permite o registro de dados na faixa espectral de 300 a 1100 nm. Os dados espectrais (TA_{NIR}) foram coletados posicionando cada semente sobre uma placa de Teflon® (politetrafluoroetileno – PTFE). Em seguida, foi realizada a análise do teor de água a partir do método tradicional (BRASIL, 2009) adaptado para sementes individuais, os valores obtidos nessa etapa foram empregados como valores de referência de teor de água das sementes (TA_{REF}) para construção dos modelos, assim, para cada semente há um espectro (TA_{NIR}) e um valor correspondente de teor de água (TA_{REF}). As sementes foram pesadas individualmente em balança analítica, numeradas, colocadas separadamente em bandeja de alumínio e levadas para a estufa a 105 °C durante 24 h e em seguida pesadas individualmente para obter o peso seco e cálculo do TA_{REF} .

2.4 Construção dos modelos

Os modelos foram desenvolvidos utilizando o software WEKA 3.9 (Grupo de *Machine Learning*, Universidade de Waikato). Inicialmente, os dados espectrais, salvos em formato Excel, foram convertidos para os formatos *Comma-Separated Value* (CSV) e *Attribute Relation File Format* (ARFF), compatíveis com o software. Os dados foram divididos em dois conjuntos: um de treinamento (70%), utilizado para construir e treinar o

modelo, e outro de teste (30%), destinado a avaliar seu desempenho. A distribuição foi realizada de forma aleatória (ZHANG et al., 2020; DEMSASH, 2023) (Tabela 2). Os modelos foram desenvolvidos a partir dos dados espectrais (TA_{NIR}) com os resultados das análises de referência (TA_{REF}) obtidos para cada semente. Para a construção dos modelos foi selecionada a faixa espectral 400 nm a 1100 nm, utilizando os dados originais (absorbância). Além disso, aplicado o filtro Reamostrar (Resample) aos espectros, conforme Gadotti et al. (2022 a, b). Os algoritmos discriminativos dos grupos functions (Support Vector Machine (SVM), Gaussian Processes e Multilayer Perceptron) e trees (Random Forest) foram aplicados. Os modelos foram avaliados por meio de validação interna no conjunto de treinamento, utilizando o método de validação cruzada com 10 folds, e por validação externa, empregando os dados do conjunto de teste.

2.5 Métricas de regressão

O desempenho dos modelos foi determinado por métricas de regressão que tem objetivo de prever valores contínuos. O coeficiente de determinação (R^2) (1), a raiz do erro quadrático médio (RMSE) (2) e erro médio absoluto (MAE) (3) (Eqs. 1- 3). R^2 varia entre 0 e 1, no qual 0 (zero) indica que o modelo não explica variação, 1 (um) indica que o modelo explica toda variação observada (Navidi, 2012). O RMSE é um parâmetro utilizado para medir o desvio entre os valores previstos e reais. O MAE calcula a média da diferença entre o valor real com o predito, sendo que quanto menor o valor dessas métricas, maior a precisão da predição do modelo (MARQUES et al., 2016; LI et al., 2018; SUN et al., 2019; WANG et al., 2021; FERREIRA et al., 2022; ZHANG et al., 2022; SOUZA et al., 2022; GUO et al., 2023).

$$(Eq.1) \quad R^2 = \frac{\sum_{t=1}^T (\hat{Y}_t - \bar{Y})^2}{\sum_{t=1}^T (Y_t - \bar{Y})^2} \quad (Eq.2) \quad RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - p_i)^2} \quad (Eq.3) \quad MAE = \frac{1}{n} \sum_{i=1}^n |y_i - p_i|$$

3 RESULTADOS

3.1 Valores de referência (TA_{REF}): utilizados para o treinamento e teste:

Os valores de referências utilizados para treinamento e teste estão expostos na Tabela 2. O teor de água observado nos sublotos das sementes de *A. colubrina* variaram entre 2 e 37,2% com média de 12,21% para o conjunto de treinamento e para o teste a variação foi 2,31 e 37,63% com média de 13,59%. Nas sementes de *C. pyramidale* a variação foi 2,05 a

38,91%, como média 13,15% para o conjunto de treinamento e para o teste a variação foi de 2,04 e 13,45% com média 13,45%.

Tabela 2: Divisão do conjunto de treinamento e teste do teor de água (T.A) das sementes de diferentes lotes de *Anadenanthera colubrina* e *Cenostigma pyramidale*.

<i>Anadenanthera colubrina</i>					
Conjunto de amostras	Nº de amostras	Max (%)	Min (%)	Média (%)	DP
Treinamento	1618	37,26	2,00	12,21	8,92
Teste	694	37,63	2,31	13,59	9,43
Total do conjunto completo	2312				
<i>Cenostigma pyramidale</i>					
Treinamento	2035	38,91	2,05	13,15	9,06
Teste	872	38,97	2,04	13,45	9,10
Total do conjunto completo	2907				

Abreviaturas: Max: TA máximo da semente; Min: AT mínimo de sementes; DP: desvio padrão

3.2 Espectros de refletância

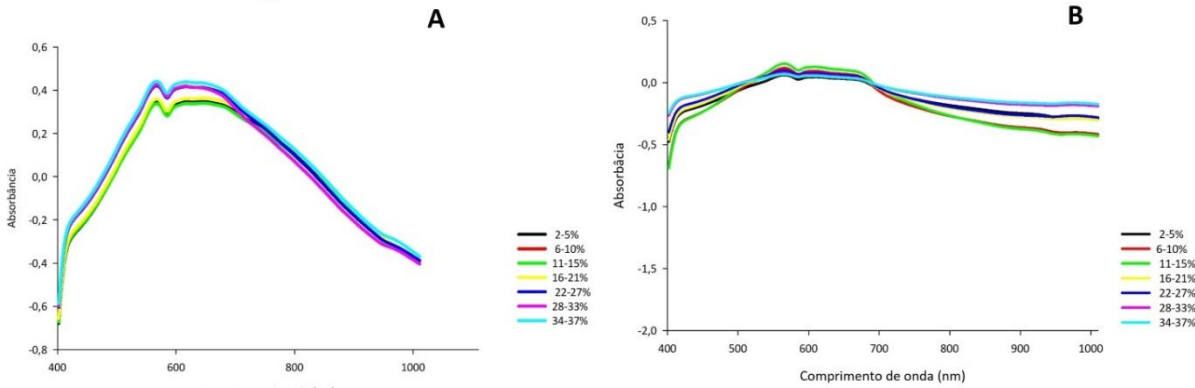


Figura 1. Espectros médios (originais) do teor de água (T.A) de sementes de *Anadenanthera colubrinae* (A) e *Cenostigma pyramidale* (B).

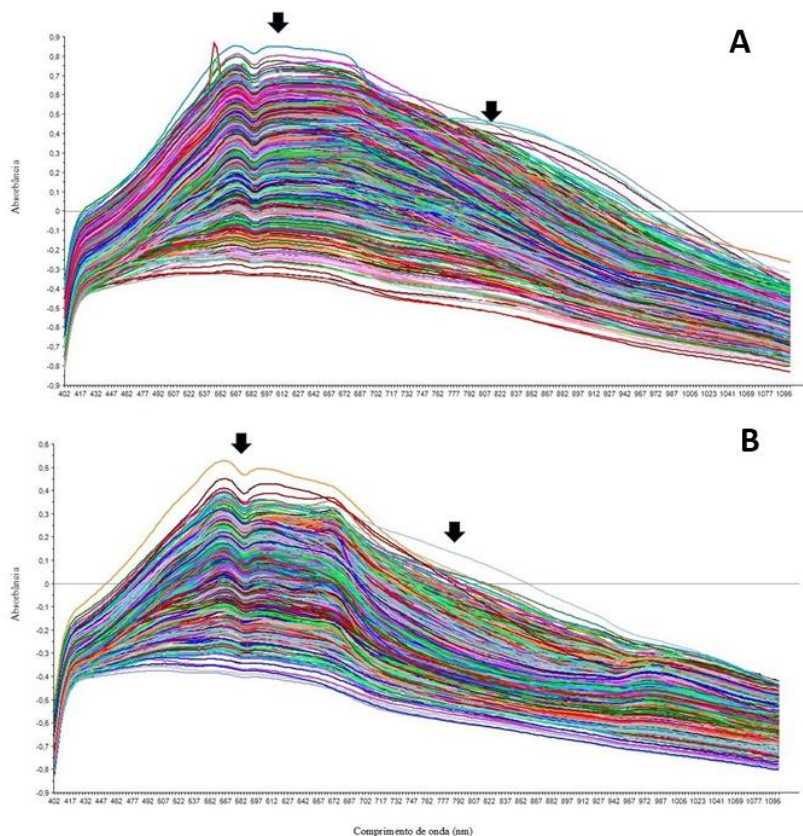


Figura 2: Todos os Espectros do teor de água (T.A) de sementes de *Anadenanthera colubrina* (A) *Cenostigma pyramidale* (B)

3.3 Espectros de refletância

Os espectros de absorvância obtidos do teor de água de sementes de *A. colubrina* e *C. pyramidale* com níveis diferentes de teor de água na faixa de 400 a 1100 nm estão apresentados os espectros médios (Figura 1A e B) e todos os espectros (Figura 2A e B). Os picos de absorção observados para o teor de água das sementes das duas espécies estão entre 500 nm a 720 nm.

3.4 Modelos com dados brutos

O algoritmo RF apresentou um bom desempenho preditivo para os dados originais do teor de água das sementes de *Anadenanthera colubrina*, com valores de R^2 de 0,35 e 0,86, RMSE de 7,24 e 3,32, e MAE de 5,08 e 2,09, nos conjuntos de treinamento e teste, respectivamente. Apesar dos baixos ajustes observados nos dados de treinamento, o desempenho no conjunto de teste foi significativamente superior. Para o teor de água das sementes de *C. pyramidale* o algoritmo MP apresentou o melhor desempenho com um $R^2=0,90$ e 0,95, RMSE=2,83 e 2,05 e MAE=1,68 e 1,18 no conjunto de teste e treinamento (Tabela 3).

Tabela 3. Modelos de regressão (dados originais) do teor de água (T.A) de sementes de *Anadenanthera colubrina* e *Cenostigma pyramidale*.

Treinamento				Dados originais	Teste		
<i>Anadenanthera colubrina</i>					<i>Anadenanthera colubrina</i>		
Modelos	R ²	RMSE	MAE	Modelos	R ²	RMSE	MAE
RF	0,35	7,24	5,08	RF	0,86	3,32	2,09
SVM	0,51	6,20	3,82	SVM	0,46	6,78	4,11
GP	0,49	6,35	4,60	GP	0,44	6,68	4,70
MP	0,49	6,70	4,82	MP	0,56	7,56	6,14
<i>Cenostigma pyramidale</i>					<i>Cenostigma pyramidale</i>		
Modelos	R ²	RMSE	MAE	Modelos	R ²	RMSE	MAE
RF	0,59	5,76	4,13	RF	0,60	5,73	3,97
SVM	0,75	4,49	2,98	SVM	0,77	4,33	2,82
GP	0,54	6,10	4,78	GP	0,57	6,09	4,72
MP	0,90	2,83	1,68	MP	0,95	2,05	1,18

3.5 Modelos com Resample

Com a aplicação do filtro *resample*, observou-se um ajuste superior em todos os dados das duas espécies, devido à correção do viés presente nos conjuntos de dados, que é comum em situações de desequilíbrio entre as classes. As sementes *A. colubrina*, demonstrou um melhor ajuste para todos os algoritmos de forma geral, porém com destaque para os algoritmos RF que apresentou no conjunto de treinamento e teste R²= 0,70 e 0,94, RMSE= 4,66 e 2,15, MAE=2,54 e 0,99, respectivamente. Para o teor de água das sementes de *C. pyramidale* os algoritmos RF no conjunto de treinamento e teste obtiveram um R²= 0,82 e 0,98, RMSE=3,72 e 1,31 e MAE= 2,09 e 0,73, respectivamente (Tabela 4).

Tabela 4. Modelos de regressão (filtro resample) do teor de água (T.A) de sementes de *Anadenanthera colubrina* e *Cenostigma pyramidale*.

Treinamento				Teste			
<i>Anadenanthera colubrina</i>				<i>Anadenanthera colubrina</i>			
Modelos	R ²	RMSE	MAE	Modelos	R ²	RMSE	MAE
RF	0,70	4,66	2,54	RF	0,94	2,15	0,99
SVM	0,57	5,62	3,82	SVM	0,65	4,88	3,43
GP	0,47	6,31	4,63	GP	0,49	6,12	4,53
MP	0,77	5,23	3,50	MP	0,72	4,49	3,01
<i>Cenostigma pyramidale</i>				<i>Cenostigma pyramidale</i>			
Modelos	R ²	RMSE	MAE	Modelos	R ²	RMSE	MAE
RF	0,82	3,72	2,09	RF	0,98	1,31	0,73
SVM	0,60	5,65	4,07	SVM	0,64	5,45	3,93
GP	0,56	6,00	4,71	GP	0,56	6,01	4,74
MP	0,77	4,37	3,04	MP	0,84	3,45	2,21

RF: Ramdon Forest; SVM: Support Vector Machine; GP: Gaussian Processes; MP: Multilayer Perceptron. R: determination coefficient; RMSE: Root Mean Square Error; MAE: mean absolute error.

4 DISCUSSÃO

Neste estudo, os baixos valores de umidade observados (aproximadamente 2%) podem ser atribuídos ao fato de que as sementes foram submetidas à secagem com sílica em gel. Essa abordagem foi adotada com o intuito de simular condições típicas de ambientes semiáridos ou outras regiões com baixa umidade. Da mesma forma, as sementes que foram umedecidas em água apresentaram valores mais elevados, cerca de 38%, o que pode simular condições de ambientes com alta umidade. Dessa forma, buscou-se promover uma maior variação nos dados, o que possibilitaria uma predição mais precisa dos modelos.

A variação dos teores de água das sementes das duas espécies, que não foram submetidas a nenhum tratamento, mostrou resultados condizentes com outros estudos. Na literatura, é indicado que as sementes de *A. colubrina* e *C. pyramidale* apresentam teores de água variando entre 6% e 10% (DANTAS et al., 2009; BISBO et al., 2017; CASTRO et al., 2017).

A espectroscopia Vis-NIR é adequada para medição do teor de água, visto que o tom do grupo OH da água e as bandas de combinação são destacadas nesta região do espectro, sendo a água um forte absorvedor na região NIR (SUNDARAM et al., 2012). Nessa região, a radiação é absorvida de forma eficiente pela água em determinados comprimentos de onda, sendo um constituinte químico de grande importância para os órgãos vegetais, participando de forma efetiva na formação, desenvolvimento e maturação das sementes. Além disso, o

espectro de infravermelho próximo é essencialmente composto por um grande conjunto de tons harmônicos e bandas combinadas (NICOLAI et al., 2007).

Neste estudo, foram observados fortes picos de absorção entre 500 nm e 720 nm, possivelmente caracterizando os compostos químicos das duas espécies analisadas. Além disso, leves picos foram identificados em todos os espectros nas regiões entre 790 nm e 970 nm. A absorbância em torno de 960 nm e 970 nm pode estar associada ao efeito do segundo tom de carboidratos e água, conforme descrito por Liu e Guo (2015) e Xu et al. (2019). Os picos de absorção de água das sementes parecem diferir entre diferentes espécies. Trabalhos com semente de pepino, chia, soja e milho, os picos do teor de água foram observados em torno de 960 1200 nm (SERSON et al., 2019; ZHANG e GUO, 2020; WANG et al., 2021; GUO et al., 2023). No trabalho de Sundaram et al. (2009) foi observado que os picos de absorção de teor de água das sementes de amendoim estão em torno de 1134 nm-1200 nm que também está relacionado ao teor de água e carboidratos das sementes.

Essa variação pode estar associada aos fatores que influenciam as bandas de absorção de água, a exemplo do soluto na água, monômero orgânico, assim como ligações de hidrogênio, com as quais as ligações de água estão fortemente associadas. Desse modo, nas sementes individuais das duas espécies do presente estudo pode conter fortes ligações de hidrogênio, o que pode ter influenciado a absorção NIR (SUNDARAM et al., 2009). Desta forma, as amostras com alto teor de água são fortemente dominadas pela assinatura da água nas bandas de absorção. Estas características permitem o uso da espectroscopia NIR na medição da água em ampla faixa de concentrações em amostras sólidas. Os espectros NIR possuem bandas de absorções que estão relacionadas a três ligações químicas: CH, geralmente associada a óleos e gorduras, OH, encontrada na água e as ligações de NH, identificadas nas proteínas (GUO et al., 2023).

O teor de água é um indicativo significativo da viabilidade, germinação e conservação das sementes, uma vez que impacta diretamente o metabolismo e a longevidade dessas. Além disso, esse parâmetro é essencial na análise da qualidade das sementes, ajudando a determinar o estágio ideal de maturação e as condições ideais para seu armazenamento (MARCOS-FILHO, 2015; XU et al., et al., 2019). Nos estádios iniciais de maturação as sementes apresentam alto teor de água, reduzindo nos estádios finais, ainda assim permanece suficientemente elevada para permitir a transferência da massa seca das plantas para as sementes (LIMA et al., 2012; Granja et al., 2018). Durante a maturação das sementes, a qualidade fisiológica aumenta, com alterações no teor de água, acúmulo de massa seca,

potencial de germinação e vigor, sendo um período essencial para o desenvolvimento de sementes viáveis (ARAÚJO et al., 2023; BAI et al., 2023).

Para os dados brutos, os algoritmos RF e MP demonstraram melhor desempenho preditivo para o teor de água das sementes *A. colubrina* e *C. pyramidale*, apresentando R^2 acima de 80% e baixos RMSE e MAE, o que é o esperado para ter uma boa predição do modelo (ZHANG et al., 2022). O MP é um algoritmo de rede neural que é bastante utilizado para treinamento de modelos em diversas áreas da pesquisa, assim esses algoritmos têm chamado atenção, pois as suas respostas são por meio de classificação de multicamadas, em que, ele aplica os dados brutos escondido e organiza para simular conexões de um cérebro humano (LIMA et al., 2006; MATHIAS et al., 2016; SCHLEDER e FAZZIO et al., 2021; SUMAYLI, 2023).

Estudos comprovam que o algoritmo MP tem um bom desempenho para análises da qualidade de sementes, onde o MP teve um bom desempenho preditivo com $R = 0,99$ e $RMSE = 0.0003$ para avaliar o TA de sementes de trigo sob diferentes temperaturas (KASHANINEJAD et al., 2009). Em outro estudo de Rajabi-Sarkhan et al. (2023), o algoritmo MP foi utilizado para classificar a viabilidade das sementes de amendoim em conjunto com a espectroscopia NIR, o qual apresentou um bom desempenho com a acurácia de 98%. Contudo, esse algoritmo tem característica de ser muito demorado para realizar a modelagem dos dados, o que não foi diferente para os dados do TA das duas espécies do presente estudo.

Para os dados com aplicação do *resample*, o algoritmo RF demonstrou um desempenho superior para os modelos do teor de água das sementes de *A. colubrina* e *C. pyramidale*. A aplicação do *resample* aos espectros é útil para equilibrar os dados e foi inicialmente utilizado para evitar viés do algoritmo e melhorar seu desempenho (GADOTTI et al., 2022 a,b). Com isso apresentaram modelos com desempenho superior, reduzindo os valores de RMSE e MAE. Desse modo, o RF apresentou um R^2 acima de 90%, assim como também um RMSE e MAE baixos, mostrando um bom desempenho. A RF cria várias árvores de decisão para prever com mais eficiência os resultados, esse algoritmo calcula a média das previsões de muitas árvores individuais (SCHONLAU e ZOU, 2020).

O RF é algoritmo bastante utilizado em trabalhos com sementes cultivadas, este algoritmo mostrou um bom desempenho para predição do teor de água das sementes de soja com uso do Vis-NIR e para classificação do vigor das sementes de trigo em conjunto com as técnicas de quimiometria (MEDEIROS et al., 2020; FAN et al., 2020; WANG et al., 2020;

WANG et al., 2021). Outro trabalho com espécies de pinhões e aplicação de espectroscopia mais aprendizagem de máquina mostrou desempenho eficiente dos modelos (HUANG et al., 2022).

O filtro *Resample* foi utilizado com o intuito de reduzir os erros nos dados do teor de água das sementes das duas espécies. Com a aplicação desse filtro, foi obtido um melhor ajuste em comparação aos dados brutos; ainda assim, de forma geral, os valores de RMSE e MAE permaneceram levemente elevados. A principal premissa pode ser devido os lotes das sementes apresentarem alta variabilidade genética e morfológica.

É necessário ressaltar que o trabalho foi realizado com espécies florestais, que não passaram por um melhoramento genético, desse modo, com grande variabilidade genética e morfológica dentro do mesmo lote, os estudos da literatura apresentam boa predição dos modelos para as espécies agrícolas, isso pode estar relacionado a diversidade genética que é mais reduzida, assim como sementes com características morfológicas mais similares (Amorim et al., 2021; WANG et al., 2021; GUO et al., 2022). A ocorrência de alguns erros também pode estar relacionada ao número reduzido de sementes utilizadas, o que possivelmente limitou a precisão da modelagem. Esse resultado indica que a ampliação da amostragem pode contribuir para uma maior robustez dos modelos, ainda que os resultados obtidos neste estudo tenham sido satisfatórios. Para trabalhos futuros utilizando aprendizagem de máquina e espectroscopia do infravermelho próximo é essencial a implementação de bancos de dados com sementes florestais nativas da caatinga, para colaborar com pesquisas com esse foco.

A aplicação da espectroscopia Vis-NIR na avaliação da qualidade de sementes de espécies florestais ainda é limitada. No entanto, nos últimos anos, alguns trabalhos avaliaram o teor de água e o teor de óleo das sementes de *Camellia gauchowensis* e *Camellia semiserrata* (ZHANG et al., 2018), teor de água, teor de óleo e de proteínas das sementes de *Salvia hispânica* (SERSON et al., 2019) e o teor de água de sementes e mudas de pinheiros (LESTANDER e GELADI, 2003; LI et al., 2023). Ainda, um estudo relatou precisão na aplicação da espectroscopia em grãos *Torreya grandis* (espécie florestal da china) para determinar o teor de água em dados espectrais brutos (XIANG et al., 2023).

As avaliações periódicas após o armazenamento do TA são necessárias já que o TA tem correlação direta com a longevidade e viabilidade durante o período de armazenamento (XU et al., 2020). Os modelos desenvolvidos para determinação do TA de sementes florestais nativas podem ser uma alternativa para monitorar as mudanças que ocorrem no TA durante a

conservação das sementes, a espectroscopia na região do visível e infravermelho próximo em conjunto com as técnicas de AM podem ser aplicadas para a caracterização das propriedades físicas e químicas de sementes florestais, então com utilização do espectrômetro as perdas podem ser minimizadas e as avaliações são executadas com mais rapidez. A aplicação desse método ainda pode ser útil para previsão do ponto ideal de maturidade das sementes, pois para definir o ponto de colheita é necessário identificar o TA (LIMA et al., 2012). Então, quando é realizada a avaliação do TA são evitadas as perdas após a colheita, dessa forma mantém qualidade ideal de sobrevivência das sementes para sua conservação (COSTA, 2009; SENA et al., 2022).

5 CONCLUSÃO

A aplicação da espectroscopia na região do visível e infravermelho próximo (Vis-NIR), associada a algoritmos de Aprendizado de Máquina (AM), demonstrou ser uma abordagem eficiente para a predição do teor de água das sementes de *A. colubrina* e *C. pyramidale*.

Os quatro algoritmos testados apresentaram bom desempenho preditivo; no entanto, os algoritmos MP e RF se destacaram com os melhores ajustes quando os dados foram analisados sem o pré-tratamento com o filtro *resample*. Por outro lado, quando o pré-tratamento *resample* foi aplicado, apenas o algoritmo RF manteve desempenho superior, indicando que esse tipo de filtragem contribui significativamente para a melhoria da capacidade preditiva dos modelos, ao reduzir o viés dos dados.

Dessa forma, os resultados obtidos neste estudo reforçam o potencial da combinação entre espectroscopia Vis-NIR e algoritmos de AM como uma alternativa rápida, precisa e não destrutiva para a estimativa do teor de água em sementes florestais nativas.

6 REFERÊNCIAS

- ALEIXANDRE-TUDO, J. L. et al. Chemometric compositional analysis of phenolic compounds in fermenting samples and wines using different infrared spectroscopy techniques. **Talanta**, v. 176, p. 526-536, 2018. <https://doi.org/10.1016/j.talanta.2017.08.065>
- AMORIM, I. P; SOUZA, J. G; BARBEDO, Claudio José. Análise comparativa das taxas de deterioração de sementes ortodoxas de *Erythrina speciosa* e recalcitrantes de *Eugenia* spp. **Iheringia, Série Botânica.**, v. 76, 2021. <http://doi.org/10.21826/2446-82312021v76e2021013>
- ARAÚJO, M. E. S et al. Physiological maturity of *Cenostigma tocaninum* Ducke (Fabaceae) seeds. **Journal of Seed Science**, v. 45, p. e202345039, 2023. <https://doi.org/10.1590/2317-1545v45275434>

BALOUCHI, H.; BALADI, S.; MORADI, A.; DEHNAVI, M. M. The influence of temperature and moisture content on seed longevity of two genotypes of *Linum usitatissimum*. *Seed Science and Technology*, v.45, n.1, p.130-138, 2017. <https://doi.org/10.15258/sst.2017.45.1.08>

BAI, B. et al. SeedTransNet: a directional translational network revealing regulatory patterns during seed maturation and germination. *Journal of Experimental Botany*.v.74, n.9, p.2416–2432, 2023. <https://doi.org/10.1093/jxb/erac394>

BERNARDY, R et al. Identifying soybean genotypes with artificial intelligence and near infrared reflectance. *Ciência e Agrotecnologia*, v. 48, p. e005224, 2024. <https://doi.org/10.1590/1413-7054202448005224>

BONJOVANI, M.R., BARBEDO, C.J. Respiration and deterioration of *Inga vera* ssp. *affinis* embryos stored at different temperatures. *Journal of Seed Science* v.41, n.1, p.044-053, 2019. <https://doi.org/10.1590/2317-1545v41n1193955>

BRASIL. Ministério da agricultura, pecuária e abastecimento. **Instruções para análise de sementes florestais**. 2013. 98p.<https://www.gov.br/agricultura/pt-br/assuntos/insumos-agropecuarios/insumos-agricolas/sementes-e-mudas/publicacoes-sementes-e-mudas/instrucoes-para-analise-de-sementes-de-especies-florestais/view>

BRASIL. Ministério da Agricultura e Reforma Agrária. **Regras para Análises de Sementes**. Brasília, DF: CLAV: DNDV: SNDA: MA, 2009.[https://www.gov.br/agricultura/ptbr/assuntos/insumos299 agropecuarios/arquivos-publicacoes-insumos/2946_regras_analise sementes.pdf](https://www.gov.br/agricultura/ptbr/assuntos/insumos299%20agropecuarios/arquivos-publicacoes-insumos/2946_regras_analise%20sementes.pdf)

CASTRO, L.E. et al. Physiological, cellular and molecular aspects of the desiccation tolerance in *Anadenanthera colubrina* seeds during germination. *Brazilian Journal of Biology*, v. 77, p. 774-780, 2017. <https://doi.org/10.1590/1519-6984.00616>

CHEN, X. et al. C. Quality evaluation of decoction pieces of *Rhizoma Atractylodis Macrocephalae* by near infrared spectroscopy coupled with chemometrics. *Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy*, v. 221, p.117169 ,2019. <https://doi.org/10.1016/j.saa.2019.117169>

COSTA, C.J. **Armazenamento e conservação de sementes de espécies do Cerrado**. 2009. Disponível em:<<https://www.infoteca.cnptia.embrapa.br/bitstream/doc/664379/1/doc265.pdf>. Acesso> em: 15 mar.2023.

DANTAS, B. F. et al. Taxas de crescimento de mudas de catingueira submetidas a diferentes substratos e sombreamentos. *Revista árvore*, v. 33, p. 413-423, 2009. <https://doi.org/10.1590/S0100-67622009000300003>

DEMSASH, A. W. “Using best performance machine learning algorithm to predict child death before celebrating their fifth birthday.” *Informatics in Medicine Unlocked*, v.40, p101298, 2023. <https://doi.org/10.1016/j.imu.2023.101298>

DINIZ, A. Y. S. et al. Biometria de frutos, sementes e plântulas de *Cenostigma tocantinum* Ducke. *Research, Society and Development*, v. 11.n.11, p.e460111133773-e460111133773,2022: <http://dx.doi.org/10.33448/rsd-v11i11.33773>

FAN, Y et al. Individual wheat kernels vigor assessment based on NIR spectroscopy coupled with machine learning methodologies. *Infrared Physics & Technology*, v. 105, p. 103213, 2020. <https://doi.org/10.1016/j.infrared.2020.103213>

FERREIRA, I. J.S. et al. Determination of quality and ripening stages of ‘pacovan’bananas using vis-NIR spectroscopy and machine learning. *Engenharia Agrícola*, v. 42, p. e20210160, 2022. <https://doi.org/10.1590/1809-4430-Eng.Agric.v42nepe20210160/2022>

GADOTTI G.I et al. Prediction of ranking of lots of corn seeds by artificial intelligence. *Engenharia Agrícola*, v.42, n.4 p. e20210005,2022b. <http://dx.doi.org/10.1590/1809-4430-Eng.Agric.v42n4e20210005/2022>

GADOTTI, G. I et al. Machine learning for soybean seeds lots classification. **Engenharia Agrícola**, v.42, n. spe ep.20210101, 2022a. <http://dx.doi.org/10.1590/1809-4430-Eng.Agric.v42nepe20210101/2022>

GRANJA, J A. A. et al. Aspectos morfológicos, germinação e análise do espectro de infravermelho próximo (NIR) de *Phyllanthus niruri* em diferentes colorações da semente. **Revista Brasileira de Plantas Mediciniais**, v. 20, p. 141-149, 2018.

GUERRA-LÓPEZ, J. R; GÜIDA, J. A; VÉDOVA, D.C.O. Infrared and Raman studies on renal stones: the use of second derivative infrared spectra. **Urological research**, v. 38, p. 383-390, 2010. <https://doi.org/10.1007/s00240-010-0305-2>

GUO, Z. et al. Application of visible-near-infrared hyperspectral imaging technology coupled with wavelength selection algorithm for rapid determination of moisture content of soybean seeds. **Journal of Food Composition and Analysis**, v. 116, p.105048, 2023. <https://doi.org/10.1016/j.jfca.2022.105048>

HUANG, B et al. Applications of machine learning in pine nuts classification. **Scientific Reports**, v. 12, n. 1, p. 8799, 2022. <https://doi.org/10.1038/s41598-022-12754-9>

KONG, L. et al. Antioxidant response and related gene expression in aged oat seed. **Frontiers in plant science**, v. 6, p. 158, 2015. <https://doi.org/10.3389/fpls.2015.00158>

KASHANINEJAD, M. et al. Modeling of wheat soaking using two artificiais neurais networks (MLP and RBF). **Journal of food engineering**, v.91 n. 4, p. 602-607,2009. <https://doi.org/10.1016/j.jfoodeng.2008.10.012>

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, 2020. <https://doi.org/10.1039/D0AY01238F>

LARIOS, G. S. et al. Laser-Induced Breakdown Spectroscopy as a Powerful Tool for Distinguishing High- and Low-Vigor Soybean Seed Lots. **Food Analytical Methods**. v. 13, n. 9, p. 1691-1698, 2020. <https://doi.org/10.1007/s12161-020-01790-8>

LESTANDER, T.A.; GELADI, Paul. NIR spectroscopic measurement of moisture content in Scots pine seeds. **Analyst**, v. 128, n. 4, p. 389-396, 2003. <https://doi.org/10.1039/B300234A>

LI, Y. et al. Detection of Moisture Content of *Pinus massoniana* Lamb. Seedling Leaf Based on NIR Spectroscopy with a Multi-Learner Model. **Forests**, v. 14, n. 5, p. 883, 2023. <https://doi.org/10.3390/f14050883>

LIMA, A. D. M. et al. Inference of oil content in petroleum waxes by artificiais neurais networks. **IFAC Proceedings** v.39, n.2, p.759-764, 2006. <https://doi.org/10.3182/20060402-4-BR-2902.00759>

LIMA, C. R. et al. Physiological maturity of fruits and seeds of *Poincianella pyramidalis* (Tul.) L.P. Queiroz. **Revista Brasileira de Sementes**. v. 34, n. 2, p. 231-240, 2012. <http://dx.doi.org/10.1590/S0101-3122201200020000>

LIMA, L.B.; MARCOS-FILHO, J. Condicionamento fisiológico de sementes de pepino e germinação sob diferentes temperaturas. **Revista Brasileira de Sementes**, v.32, n.1, p.138-147,2010.http://www.scielo.br/scielo.php?script=sci_pdf&pid=S010131222010000100017&lng=en&nrm=iso&tlng=pt

LIU, D; GUO, W. Identification of kiwifruits treated with exogenous plant growth regulator using near-infrared hyperspectral reflectance imaging. **Food Analytical Methods**, v. 8, n.1, p. 164-172, 2015.

MARQUES, E. J. N. et al. Rapid and nondestructive determination of quality parameters in the _Tommy Atkins_ mango using a novel handheld near infrared spectrometer. **Food Chemistry**, v.197, p.1207-1214, 2016. <https://doi.org/10.1016/j.foodchem.2015.11.080>

MARQUES, E. J. N.; FREITAS, S. T. Performance of new low-cost handheld NIR spectrometers for nondestructive analysis of umbu (*Spondias tuberosa* Arruda) quality. **Food chemistry**, v. 323, p. 126820, 2020. <https://doi.org/10.1016/j.foodchem.2020.126820>

MATHIAS, I. M. et al. BRNeural-Artificial Neural Networks Simulator with Topology Multilayer Perceptron Using the Encog Framework. **IEEE Latin America Transactions**, v.14, n.1, 309-313, 2016. <https://doi.org/10.1109/TLA.2016.7430094>

MEDEIROS, A. D. 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, 2020. <https://doi.org/10.3390/s20154319>

NASCIMENTO, J.P.B; BISPO, J. S.; DANTAS, B. F. **Angico *Anadenanthera colubrina* var. cebil (Vell.) Brenan**. 2018.
Disponível em: <<https://www.alice.cnptia.embrapa.br/alice/bitstream/doc/1131693/1/Angico-Anadenanthera-colubrina-2018.pdf>. Acesso em: 10 dez. 2023.

NAVIDI, W. **Probabilidade E Estatística Para Ciências Exatas**, Brasil: Amgh Editora, 2012, 616p

NERY, M.C. et al. Determinação do grau de umidade de sementes de ipê-do-cerrado *Tabebuia ochracea* ((Cham.) Standl.) pelos métodos de estufa e forno de microondas. **Ciência e Agrotecnologia**, v. 28, n.6 p. 1299-1305, 2004. <https://doi.org/10.1590/S1413-70542004000600011>

PINA-RODRIGUES, F. C. M.; FIGLIOSA, M. B.; SILVA, A. da. **Sementes florestais tropicais: da ecologia à produção**. Londrina: ABRATES, 2015.

QIAO, Q.; YUNUSA-KALTUNGO, A.; EDWARDS, R.E. Feature selection strategy for machine learning methods in building energy consumption prediction. **Energy Reports**, v. 8, p. 13621-13654, 2022. <https://doi.org/10.1016/j.egyr.2022.10.125>

RABANERA, J. D. et al. Rapid and Non-destructive measurement of moisture content of peanut (*Arachis hypogaea* L.) kernel using a near-infrared hyperspectral imaging technique. **Journal of Food Measurement and Characterization**, v. 15, p. 3069-3078, 2021. <https://doi.org/10.1007/s11694-021-00894-x>

RAJABI-SARKHANI, M. et al. Identifying Optimal Wavelengths from Visible–Near-Infrared Spectroscopy Using Metaheuristic Algorithms to Assess Peanut Seed Viability. **Agronomy**, v.13, n.12, p.2939, 2023. <https://doi.org/10.3390/agronomy13122939>

REZAEI, S. et al. Contrasting seed moisture sorption behaviour between two species and the implication for seed longevity. **Seed Science Research**, v. 33, n. 2, p. 126-134, 2023.4. <https://doi.org/10.1017/S0960258523000156>

SARKAR, S. et al. Soybean seed composition prediction from standing crops using PlanetScope satellite imagery and machine learning. **ISPRS Journal of Photogrammetry and Remote Sensing**, v. 204, p.257-274,2023. <https://doi.org/10.1016/j.isprsjprs.2023.09.010>

SCHLEDER, G. R.; FAZZIO, A. Machine Learning na Física, Química, e Ciência de Materiais: Descoberta e Design de Materiais. **Revista Brasileira de Ensino de Física**, v. 43 n.1, p. e20200407, 2021. <https://doi.org/10.1590/1806-9126-RBEF-2020-0407>

SCHONLAU, M.; ZOU, R. Y. The random forest algorithm for statistical learning. **The Stata Journal**, v. 20, n. 1, p. 3-29, 2020. <https://doi.org/10.1177/1536867X20909688>

SENA, D. V. D. A. et al. Ponto de maturidade fisiológica de sementes de *Sideroxylon obtusifolium* [(Roem. & Schult.) TD Penn.]. **Ciência Florestal**, v.32, n.3, p. 1106-1124, 2022. <https://doi.org/10.5902/1980509834217>

SERSON, W. et al. Development of Whole and Ground Seed Near-Infrared Spectroscopy Calibrations for Oil, Protein, Moisture, and Fatty Acids in *Salvia hispanica*. **Journal of the American Oil Chemists' Society**, v. 97, n. 1, p. 3-13, 2020. <https://doi.org/10.1002/aocs.12300>

SHAFIEE, S.; MINAEI, S. Combined data mining/NIR spectroscopy for purity assessment of lime juice. **Infrared Physics & Technology**, v. 91, p. 193-199, 2018. <https://doi.org/10.1016/j.infrared.2018.04.012>

SOUSA, L. M. S. et al. *Poincianella pyramidalis* (Tul) LP Queiroz: A review on traditional uses, phytochemistry and biological-pharmacological activities. **Journal of ethnopharmacology**, v. 264, p. 113181, 2021. <https://doi.org/10.1016/j.jep.2020.113181>

SOUZA, C. C. et al. Aplicação de inteligência artificial na previsão de produções de açúcar e etanol, no período de 2020 até 2025, nas macrorregiões Centro-Sul e Norte-Nordeste do Brasil. **Revista Brasileira de Geografia Física**, v.15, n.5, p. 2200-2211,2022.

SUMAYLI, A. Development of advanced machine learning models for optimization of methyl ester biofuel production from papaya oil: Gaussian process regression (GPR), multilayer perceptron (MLP), and K-nearest neighbor (KNN) regression models. **Arabian Journal of Chemistry**, v. 16, n. 7, p. 104833, 2023. <https://doi.org/10.1016/j.arabjc.2023.104833>

SUN, J. et al. Detection of moisture content in peanut kernels using hyperspectral imaging technology coupled with chemometrics. **Journal of Food Process Engineering**, v. 42, n. 7, p. e13263, 2019. <https://doi.org/10.1111/jfpe.13263>

SUNDARAM, J et al. Application of NIR reflectance spectroscopy on determination of moisture content of in-shell peanuts: A non-destructive analysis. **Transactions of the ASABE**, v. 53, n. 1, p. 183-189, 2010. <https://doi.org/10.13031/2013.29485>

SUNDARAM, J. et al. Sensing of moisture content of in-shell peanuts by NIR reflectance spectroscopy. **Journal of sensor technology**, v. 2012, 2012. <http://www.scirp.org/journal/PaperInformation.aspx?PaperID=17910>

SUNDARAM, J; KANDALA, C. V.; BUTTS, C. L. Application of near infrared spectroscopy to peanut grading and quality analysis: overview. **Sensing and Instrumentation for Food Quality and Safety**, v. 3, p. 156-164, 2009. <https://doi.org/10.1007/s11694-009-9081-5>

WANG, Z. et al. Determination of moisture content of single maize seed by using long-wave near-infrared hyperspectral imaging (LWNIR) Coupled with UVE-SPA combination variable selection method. **IEEE Access**, v.8, p. 195229-195239, 2020. <https://doi.org/10.1109/ACCESS.2020.3033582>

WANG, Z. et al. Application of long-wave near infrared hyperspectral imaging for determination of moisture content of single maize seed. **Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy**, v.254, p.119666. <https://doi.org/10.1016/j.saa.2021.119666>

XIANG, J. et al. sustainable way to determine the water content in *Torreya grandis* kernels based on near-infrared spectroscopy. **Sustainability**, v. 15, n. 16, p. 12423, 2023. <https://doi.org/10.3390/su151612423>

Xu, Y. et al. Rapid prediction and visualization of moisture content in single cucumber (*Cucumis sativus* L.) seed using hyperspectral imaging technology. **Infrared Physics & Technology**, v102, p.103034, 2019. <https://doi.org/10.1016/j.infrared.2019.103034>

XU, Y.D. et al. Near-Infrared spectroscopy combined with multivariate calibration to predict the yield of sesame oil produced by traditional aqueous extraction process. **Journal of Food Quality**, v. 2017, p.1-5, 2017. <https://doi.org/10.1155/2017/2515476>

ZHANG, L. et al. Moisture detection of single corn seed based on hyperspectral imaging and deep learning. **Infrared Physics & Technology**, p.125, p.104279, 2022. <https://doi.org/10.1016/j.infrared.2022.104279>

ZHANG, L. et al. Identification of seed maize fields with high spatial resolution and multiple spectral remote sensing using random forest classifier. **Remote Sensing**, v. 12, n. 3, p. 362, 2020. <https://doi.org/10.3390/rs12030362>

ZHANG, Y. et al. Rapid determination of the oil and moisture contents in *Camellia gauchowensis* Chang and *Camellia semiserrata* Chi seeds kernels by near-infrared reflectance

spectroscopy. **Molecules**, v. 23, n. 9, p. 2332, 2018.

<https://doi.org/10.3390/molecules23092332>

ZHANG, Y; GUO, W. Moisture content detection of maize seed based on visible/near-infrared and near-infrared hyperspectral imaging technology. **International Journal of Food Science & Technology**, v. 55, n. 2, p. 631-640, 2020. <https://doi.org/10.1111/ijfs.14317>

CAPÍTULO 2

USO DO ESPECTRÔMETRO VIS-NIR PORTÁTIL E *MACHINE LEARNING* NA DETERMINAÇÃO DO VIGOR DE SEMENTES FLORESTAIS

Capítulo 2: Uso do espectrômetro Vis-NIR portátil e *machine learning* na determinação do vigor de sementes florestais

Resumo: Germinação e vigor são critérios essenciais para avaliar a qualidade das sementes florestais, refletindo na qualidade das mudas. A espectroscopia Vis-NIR é uma alternativa precisa para análises de sementes de espécies florestais nativas. Este estudo teve como objetivo desenvolver modelos para avaliar a germinação e o vigor de sementes e plântulas de duas leguminosas do bioma Caatinga, usando um espectrômetro portátil Vis-NIR e algoritmos de aprendizado de máquina. Foram usadas sementes de *Anadenanthera colubrina* e *Cenostigma pyramidale*, coletadas de diferentes populações e anos. Os espectros foram obtidos com o espectrômetro F-750 Produce Quality Meter e, as sementes foram classificadas quanto à germinação e as plântulas quanto ao vigor (alto, baixo e nulo). Os modelos de calibração foram desenvolvidos com quatro algoritmos discriminativos, *Random Forest* (RF) J48, *SupportVector Machine* (SVM) e *Multilayer Perceptron* (MP). Para os resultados superiores de classificação da germinação, plântulas e vigor foram alcançados com os algoritmos *Random Forest* (RF) e J48. Esses algoritmos apresentaram melhor desempenho na classificação das sementes germinadas, das sementes que não produziram plântulas e das sementes com alto vigor, alcançando as maiores taxas de acerto nessas categorias em ambas as espécies. A combinação de Vis-NIR com algoritmos de aprendizado de máquina mostrou-se eficaz na discriminação das classes de germinação e vigor das sementes. No entanto, a inclusão de mais dados nos modelos é essencial para aprimorar a precisão das análises.

Palavras chaves: espectroscopia, vigor, conservação, *Random Forest*, J48

Chapter 2: Use of portable Vis-NIR spectrometer and machine learning in determining forest seed vigor

Abstract: Germination and vigor are essential components for assessing the quality of forest seeds, reflecting on the quality of seedlings. Vis-NIR spectroscopy offers a precise alternative for analyzing seeds of native forest species. The aim of this study was to use a portable Vis-NIR spectrometer and machine learning algorithms to develop models for assessing the germination and vigor of seeds and seedlings of two legumes from the Caatinga biome. Seeds of *Anadenanthera colubrina* and *Cenostigma pyramidale* were collected from different populations and years. The spectra were obtained using the F-750 Produce Quality Meter spectrometer. Following the germination test, the seeds were classified according to their germination and seedlings according to their vigor (high, low, and zero). The calibration models were developed using four discriminative algorithms. For *A. colubrina* and *C. pyramidale*, the Random Forest (RF) and J48 algorithms provided the best classification for germination, seedling, and vigor. These algorithms performed better in the classification of germinated seeds, seeds that did not produce seedlings, and seeds with high vigor, attaining the highest accuracy rates in these categories for both species. The integration of Vis-NIR with machine learning algorithms was found to be an effective approach to discriminate seed germination and vigor classes. However, the inclusion of additional data in the models is crucial to improve the precision of the analyses.

Keywords: spectroscopy, vigor, conservation, *Random Forest*, J48

1 INTRODUÇÃO

As sementes desempenham um papel crucial na restauração florestal e na preservação de espécies, uma vez que são fundamentais para a formação de novos ecossistemas (GASPERIN et al., 2021; FREMOUT et al., 2022). O vigor é um dos principais indicadores da qualidade fisiológica das sementes e é essencial para avaliar essa qualidade. Esse atributo influencia a rapidez e a uniformidade da germinação, além da viabilidade das mudas tanto em campo quanto durante o armazenamento (MARCOS-FILHO, 2015; ROSSI et al., 2017). Em outras palavras, relaciona-se a porcentagem e uniformidade de germinação e de emergência com rapidez e desenvolvimento de plântulas normais, sob ampla diversidade de condições ambientais (KRZYZANOWSKI e FRANÇA-NETO, 2001; CHEN et al., 2022).

O vigor das sementes é uma característica fisiológica complexa que atinge seu máximo potencial na fase final da maturação, mas tende a diminuir após a colheita e com o tempo de armazenamento (CHEYED, 2020; ZHAO et al., 2021; REED et al., 2022). Testes de vigor permitem avaliar, de maneira direta ou indireta, as informações bioquímicas e fisiológicas que indicam o potencial de desempenho de um lote, oferecendo uma diferenciação mais significativa em comparação ao teste de germinação (MARCOS-FILHO, 2015; FINCH-SAVAGE e BASSEL, 2016). Dessa forma, o vigor pode ser medido em conjunto com a germinação ou de maneira independente.

As técnicas desenvolvidas para realizar a análise do vigor de sementes de acordo com as Regras para análises de sementes (BRASIL, 2009), baseadas nas regras internacionais para teste de sementes (ISTA, 2025), são os testes de germinação, teste de velocidade de germinação, teste de envelhecimento acelerado, condutividade elétrica, emergência em campo e teste de tetrazólio. Todos esses métodos são destrutivos e requerem treinamento especializado, experiência e tempo são necessários; além disso, é preciso uma grande quantidade de sementes para avaliação. Por serem mais demorados, esses processos podem dificultar a conservação de espécies e a tomada de decisões.

Portanto, a adaptação de um método rápido, prático, preciso e não destrutivo para avaliar a viabilidade de bancos de sementes pode ser uma alternativa valiosa.

A utilização da espectroscopia Vis-NIR em conjunto com as técnicas de modelagem com *machine learning* são métodos bastante utilizados para avaliar a qualidade de produtos agrícolas em geral (WANG et al., 2022). Uma vantagem dessa técnica é que ela não é destrutiva, o que possibilita a obtenção de informações sobre a amostra, conservando-as intactas. A espectroscopia é considerada um método prático, rápido e preciso para avaliar a

viabilidade de sementes de plantas cultivadas, podendo também ser aplicada a sementes de espécies florestais nativas (TIGABU et al., 2020; FENG et al., 2021).

As sementes de espécies florestais nativas são fundamentais para a recuperação de ecossistemas degradados, pois ajudam na restauração da biodiversidade e na funcionalidade dos ambientes naturais. Adaptadas às condições locais, elas preservam recursos genéticos e promovem a sustentabilidade ambiental (SCHMIDT et al., 2018). A espectroscopia oferece uma forma rápida e precisa de avaliar a qualidade dessas sementes (XIA et al., 2019), garantindo o uso apenas das viáveis e vigorosas e, assim, aumentando as chances de sucesso em projetos de conservação e restauração de áreas degradadas.

De acordo com a literatura recente (FAN et al., 2020). A espectroscopia Vis-NIR, associada a algoritmos de aprendizagem de máquinas, pode ser utilizada para monitorar de forma não destrutiva a germinação e o vigor de sementes (FAN et al., 2020). Levando-se em conta que a avaliação da qualidade das sementes depende de testes destrutivos, para medir germinação e vigor de sementes florestais, buscamos avaliar a exequibilidade da utilização da espectroscopia na região do visível infravermelho próximo (Vis-NIR) como uma ferramenta para apoiar a análise da qualidade fisiológica de sementes florestais nativas, oferecendo uma abordagem mais eficiente para determinar esses parâmetros da qualidade de sementes.

A hipótese central deste estudo é que a espectroscopia Vis-NIR, associada a algoritmos de máquinas de aprendizagem, pode ser utilizada para monitorar de forma rápida, prática, precisa e não destrutiva a germinação e o vigor de sementes de espécies de florestas secas. Dessa forma, este estudo teve como objetivo desenvolver modelos de calibração precisos para avaliar a germinação e o vigor de sementes e plântulas de duas leguminosas florestais nativas de Florestas Secas, utilizando um espectrômetro portátil Vis-NIR e algoritmos de aprendizado de máquina

2 MATERIAL E MÉTODOS

2.1 Material vegetal

Sementes de *Anadenanthera colubrina* (Vell.) Brenan e *Cenostigma pyramidale* (Tul.) Gagnon & G.P. Lewis provenientes de diferentes populações do nordeste do Brasil. O clima local é semiárido e a vegetação é Floresta Tropical Sazonalmente Seca, perturbada pela ação humana. As sementes foram coletadas em diferentes períodos entre 2015 a 2023, armazenadas em câmara fria a $10\pm 1^{\circ}\text{C}$ e $20\pm 5\%$ de umidade relativa até serem submetidas às avaliações. As sementes de *A. colubrina* foram colhidas em 2017, 2019, 2022 e 2023 no distrito de Jutaí, em Lagoa Grande-PE ($8^{\circ} 33' 35.7''$ S, $40^{\circ} 12' 1.9''$ W); em 2020 em Sertânia-PE ($08^{\circ} 04' 25''$ S, $37^{\circ} 15' 52''$ W) e em 2021 no município de Juazeiro-BA ($09^{\circ} 24' 42''$ S, $40^{\circ} 29' 55''$ W). As sementes de *C. pyramidale* foram colhidas no povoado de Salitre ($9^{\circ} 31'$ S, $40^{\circ} 15'$ W) (2015, 2016) e Massaroca ($9^{\circ} 51' 55''$ S, $40^{\circ} 16' 57''$ O) (2015, 2016 2017), no município de Juazeiro- BA (2022) e Sertânia-PE ($08^{\circ} 04' 25''$ S, $37^{\circ} 15' 52''$ W) (2019 e 2023).

2.2 Obtenção dos dados espectrais Vis-NIR

Os espectros foram coletados com o espectrômetro portátil F-750 Produce Quality Meter (Felix Instruments, EUA), que permite o registro de dados na faixa espectral de 300 nm a 1100 nm, com intervalo de aquisição de dados de 3 nm. A avaliação foi realizada individualmente para cada semente, utilizando o espectrômetro, sendo as mesmas também submetidas às demais avaliações de referência (classificação da germinação, plântulas e vigor das sementes). Foram utilizados 1300 e 1700 sementes e plântulas, das quais 700 e 400 sementes destinadas à análise de vigor de *A. colubrina* e *C. pyramidale*, respectivamente (Tabela 1).

. Para a coleta dos dados espectrais, cada semente foi colocada em uma placa de Teflon® (politetrafluoroetileno – PTFE), sendo cada espectro de cada semente representado pela média de 4 leituras. Após a coleta dos dados espectrais, cada semente foi submetida às análises de referência (BRASIL, 2013).

2.3 Teste de Germinação

A análise de referência de germinação é essencial para a interpretação precisa dos espectros, pois fornece dados fundamentais que possibilitam a identificação de padrões específicos nos espectros das amostras. No teste de germinação, foram utilizadas 1300 e 1700 sementes de *A. colubrina* e *C. pyramidale*, respectivamente, para cada um dos seis lotes das duas espécies (Tabela 1). Antes da realização do teste de germinação foram realizadas as

leituras das sementes individuais com o espectrômetro Vis-NIR. Após a coleta dos espectros individuais de cada semente, estas foram distribuídas em rolos de papel *germitest*, umedecidos com água destilada na proporção de 2,5 vezes o peso do papel seco. Os rolos foram embalados em sacos plásticos para diminuir a evaporação de água do papel e posteriormente incubados na câmara de germinação do tipo B.O.D. (*Biochemical Oxygen Demand*) à temperatura constante de $30\pm 1^{\circ}\text{C}$ por dez dias para *A. colubrina* e quatorze dias para *C. pyramidale* (BRASIL, 2013).

2.3.1 Classificação da germinação

Para classificar a ausência e presença de germinação das sementes, essas foram divididas em duas categorias: Sim (sementes germinadas) e Não (sementes não germinadas). As sementes que apresentaram comprimento de raiz principal maior ou igual a 2 mm foram consideradas germinadas (Figura 1).

2.3.2 Classificação de plântulas

As plântulas foram classificadas em três categorias, conforme as instruções para sementes de espécies florestais (BRASIL, 2013). Normal, que se refere a plântulas intactas e vigorosas, com potencial para se desenvolverem em plantas adultas saudáveis; anormal, que são plântulas sem potencial para crescimento e que não se tornam plantas adultas normais; e sementes não germinadas, que incluem aquelas duras ou mortas, incapazes de gerar plântulas (Figura 1).

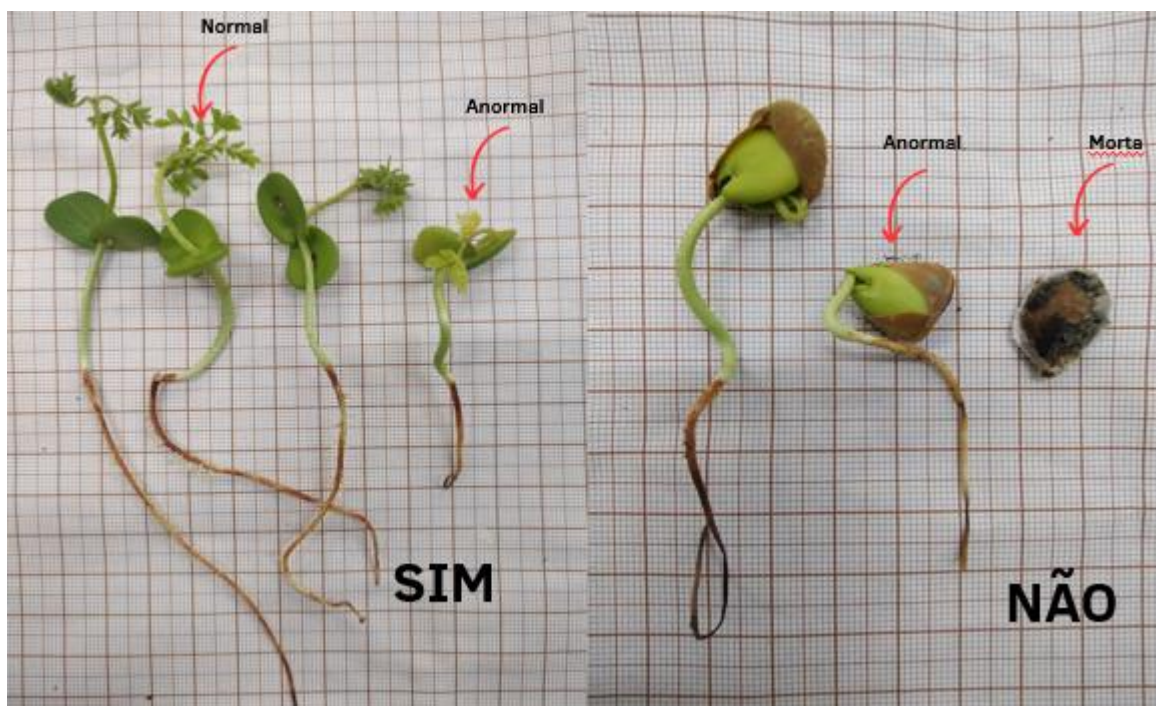


Figura 1: Classificação das sementes germinadas (sim) e não germinada (não); plântulas normais, anormais e sem produção de plântula (mortas) de *A. colubrina*.

2.3.3 Avaliação de plântulas

Dez plântulas normais de cada amostra por repetição foram separadas para a avaliação do crescimento. O comprimento das plântulas foi medido com uma régua milimetrada, e os resultados médios foram expressos em centímetros. No período de dez dias para *A. colubrina* e quatorze dias para *C. pyramidale*.

A biomassa seca das plântulas foi individualmente determinada após a avaliação do comprimento. A parte aérea e o sistema radicular foram mantidos, enquanto os cotilédones das plântulas foram descartados. As plântulas foram então colocadas individualmente em sacos de papel e submetidas à secagem em estufa com circulação forçada de ar a 65 °C por 72 horas. A biomassa seca das plântulas foi obtida por meio de pesagem em uma balança analítica com precisão de 0,001 g, com resultados médios expressos em gramas.

2.4 Avaliação do vigor

Em outro experimento, foi realizado o teste de vigor com 700 e 400 sementes de *A. colubrina* e *C. pyramidale* respectivamente, seguido da coleta dos dados espectrais Vis-NIR. As sementes foram dispostas em caixas Gerbox, contendo duas folhas, e incubadas em uma câmara de germinação tipo B.O.D. (*Biochemical Oxygen Demand*) à temperatura constante de $30\pm 1^\circ\text{C}$. Para a classificação quanto ao vigor, todas as sementes foram inspecionadas visualmente e avaliadas diariamente ao longo de um período de dez dias para *A. colubrina* e quatorze dias para *C. pyramidale* (adaptado de Brasil, 2009).

2.4.1 Classificação do vigor das sementes

O vigor das sementes foi classificado em três categorias com base na emissão da raiz principal, sendo estas nomeadas alto vigor, para sementes que germinam em até três dias; baixo vigor, para aquelas que germinam entre quatro e dez dias; e nulo vigor, para sementes que não germinam em mais de dez ou quatorze dias, sendo consideradas sem vigor. Para as sementes de *A. colubrina*, foram consideradas apenas as classificações de alto e nulo vigor, devido ao fato de as sementes germinarem rapidamente ou não germinarem.

2.5 Construção dos modelos de calibração

Os modelos de calibração foram desenvolvidos correlacionando os dados espectrais de sementes individuais com os resultados das análises de germinação, crescimento de plântulas e vigor das sementes. Foi utilizada, para maior precisão dos modelos, a faixa espectral de 400 nm a 1000 nm. Além disso, foi aplicado um filtro de *resample* a todos os dados espectrais obtidos, visando aprimorar o ajuste dos modelos e retirar qualquer enviesamento (GADOTTI et al. 2022 a, b).

Os dados espectrais foram carregados no software WEKA 3.9 (*Machine Learning Group, University of Waikato*) e foram divididos aleatoriamente em um conjunto de treinamento (70%) e um conjunto de teste (30%), utilizando o filtro aleatório do Weka (ZHANG et al., 2020; DEMSASH, 2023) (Tabela 1). Foram testados os algoritmos discriminativos dos grupos Trees -Random Forest (RF) e J48- e Function -Support Vector Machine (SVM) e Multilayer Perceptron (MP). O desempenho foi avaliado através de validação interna, validação cruzada de 10 dobras e validação externa (LARIOS et al., 2020).

Para comparar dois métodos de modelagem, foi realizada uma validação cruzada k-fold (10 dobras) com o conjunto completo de sementes das duas espécies, utilizando apenas os algoritmos RF e J48, que apresentaram desempenhos superiores (Tabela 1).

Tabela 1. Conjunto de treinamento e teste de sementes germinadas (Sim) e não germinadas (Não), plântulas (normal, anormal e sem produção de plântula) e sementes com alto, baixo e nulo vigor de *A. colubrina* e *C. pyramidale*.

<i>A. colubrina</i>			<i>C. pyramidale</i>		
Classes	Treinamento	Teste	Classes	Treinamento	Teste
Sim	524	207	Sim	494	209
Não	386	183	Não	696	301
Normal	402	153	Morta	697	300
Anormal	119	52	Anormal	224	91
Morta	389	185	Normal	269	119
Total	910	390	Total	1190	510
Total do conjunto completo			1700		
Classes	Treinamento	Teste	Classes	Treinamento	Teste
Alto	263	120	Alto	116	48
Nulo	227	90	Baixo	102	47
Total	490	210	Nulo	62	25
Total do conjunto completo			400		

3 Métricas das acurácias

Os dados foram avaliados com base na matriz de confusão com os valores de: (I) Verdadeiros positivos (VP): classificação correta da classe positiva; (II) Falsos positivos (FP): erro em que o modelo previu a classe positiva quando o valor real era classe negativa (Erro Tipo I); (III) Verdadeiros negativos (VN): classificação correta da classe negativa; (IV)

Falsos negativos (FN): erro em que o modelo previu a classe negativa quando o valor real era classe positiva (erro tipo II) (RODRIGUES, 2018; RODRIGUES, 2019).

3.1 Cálculos das acurácias

Acurácia indica uma performance geral do modelo, entre todas as classificações. Tem como objetivo avaliar corretamente o percentual de acertos e é representada por:

$$Acurácia = \frac{VP+VN}{VN+FP+FN} \text{ (Eq.1);}$$

Onde,

VP: verdadeiro positivo.

VN: verdadeiro negativo.

FP: falso positivo.

FN: falso negativo.

Precisão é utilizada em uma situação em que os falsos positivos podem causar prejuízos para o modelo, sendo representada por: $Precisão = \frac{VP}{VP+FP}$ (Eq.2);

Onde,

VP: verdadeiro positivo.

FP: falso positivo.

F-measure é média harmônica entre precisão e sensibilidade e fornece equilíbrio entre essas duas métricas, levando em consideração os falsos positivos e falsos negativos: $F1 = \frac{Precisão \times sensibilidade}{Precisão + sensibilidade}$ (Eq.3);

Matthews correlation coefficient (MCC) é uma medida de qualidade de classificações binárias que pode ser usada mesmo se as classes possuírem tamanhos diferentes: $MCC = \frac{VP \times VN - FP \times FN}{\sqrt{(VP+FP)(VP+FN)(VN+FP)(VN+FN)}}$ (Eq.4);

Onde,

VP: verdadeiro positivo.

VN: verdadeiro negativo.

FP: falso positivo.

FN: falso negativo.

O Coeficiente Kappa pode ser definido como uma medida de associação usada para descrever e testar o nível de concordância na classificação entre dois ou mais conjuntos de dados:

$$Kappa = 2 \times \frac{VP \times VN - FP \times FN}{VP \times FN + VP \times FP + 2 \times VP \times VN + FN^2 + FN \times VN + FP^2 + FP \times VN} \quad (Eq.5);$$

VP: verdadeiro positivo.

VN: verdadeiro negativo.

FP: falso positivo.

FN: falso negativo.

Curva de características de operação do receptor (Roc Área), mostra o desempenho do modelo de classificação em todos os limites de classificação, é representada por: taxa de verdadeiro positivo (VPR) e taxa de falso positivo (FPR). As equações estão abaixo:

$$VPR = \frac{VP}{VP+FN}$$

Onde,

VP: verdadeiro positivo.

VN: verdadeiro negativo.

FN: falso negativo.

(Eq.6); $FPR = \frac{FP}{FP+VN}$ (Eq.7).

Onde,

FP: falso positivo

VN: verdadeiro negativo.

FP: falso positivo.

4 RESULTADOS

Os lotes avaliados foram heterogêneos apresentando tanto germinação quanto vigor diferentes entre eles. Os lotes 2020 e 2023 de *A. colubrina* e 2019 e 2022 de *C. pyramidale* apresentaram alto vigor. Os lotes das sementes de *A. colubrina* colhidos em 2019 apresentaram germinação de 11%, enquanto os lotes de *A. colubrina* colhidos em 2017 e os de *C. pyramidale* colhidos em 2015 e 2016 apresentaram germinação nula, indicando morte das sementes (Tabela 2).

Tabela 2. Resultados dos testes de médias da germinação, comprimento e biomassa seca de plântulas, de *A. colubrina* e *C. pyramidale*.

<i>Anadenanthera colubrina</i>			
Lotes	Germinação (%)	Comprimento(cm)	Biomassa (g)
2017	0	0	0
2019	11d	0	0
2020	99 a	11.9 b	0.325 c
2021	85 b	14.6 a	0.468 a
2022	66 c	14.0 a	0.421 ab
2023	97 a	16.8 a	0.392 bc
<i>Cenostigma pyramidale</i>			
2015	0	0	0
2016	0	0	0
2017	57 b	10.7 b	0.371 a
2019	89a	12.2 b	0.474 a
2022	92a	15.3 a	0.433 a
2023	74 ab	10.6 b	0.329 a

Tukey a 5% de probabilidade

Os picos de absorção nos espectros das classes de germinação, plântulas e vigor das sementes de ambas as espécies situam entre 500 nm e 700 nm (Figura 1).

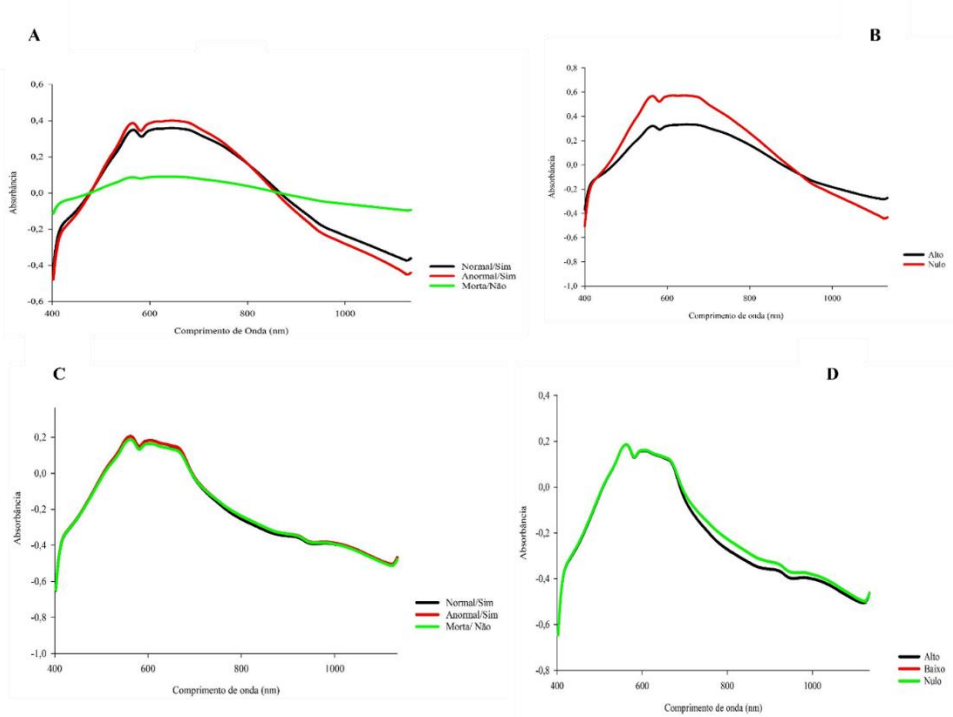


Figura 2. Espectros médios (originais) da germinação de sementes e plântulas (A) e vigor de sementes (B) de *A. colubrina* e da germinação de sementes e plântulas (C) e vigor de sementes (D) de *C. pyramidale*

Os resultados da classificação da germinação das sementes de *A. colubrina* e *C. pyramidale* indicam eficiência do algoritmo RF, alcançando valores acima de 80% para acurácia, precisão, F-measure, MCC e ROC área, e kappa acima de 80% (*A. colubrina*) e 50% (*C. pyramidale*) no conjunto de treinamento. No conjunto de teste, os resultados foram superiores, com todas as métricas superiores a 95%. Além disso, o algoritmo J48 também foi eficiente para *A. colubrina*, no conjunto de treinamento, com acurácia, precisão, F-measure e ROC área apresentando valores acima de 85%, e Kappa e MCC acima de 75%. Para *C. pyramidale*, no conjunto de treinamento, alcançou acurácia, precisão, F-Measure e Área ROC com valores acima de 70%, e Kappa e MCC acima de 45%. No conjunto de testes, foram observados valores acima de 90% para todas as métricas em ambas as espécies (Tabela 3).

Tabela 3. Parâmetros de Predição da Germinação das Sementes de *Anadenanthera colubrina* e *Cenostigma pyramidale* utilizando espectrômetro Vis-NIR portátil e algoritmos de aprendizado de máquina.

<i>Anadenanthera colubrina</i>						
Treinamento						
Algoritmo	Acurácia (%)	Precisão (%)	F-Measure (%)	Kappa (%)	MCC (%)	ROC Área (%)
R F	91,6	91,7	91,6	82,0	82,9	95,5
SVM	87,9	83,9	90,3	74,0	76,0	86,2
J48	89,1	89,1	89,1	77,0	77,7	90,4
MP	88,5	88,9	88,4	76,0	76,7	89,4
Teste						
R F	97,6	97,7	97,7	95,0	95,4	99,8
SVM	87,6	89,6	87,4	74,0	77,0	83,4
J48	96,1	96,2	96,2	92,0	92,3	98,8
MP	91,5	92,3	91,5	82,0	83,7	90,2
<i>Cenostigma pyramidale</i>						
Treinamento						
Algoritmo	Acurácia (%)	Precisão (%)	F-Measure (%)	Kappa (%)	MCC (%)	ROC Área (%)
R F	79,7	79,7	79,5	57,0	57,9	88,8
SVM	70,5	70,3	69,9	37,0	38,2	63,9
J48	75,2	75,2	75,3	48,0	49,0	73,0
MP	75,9	75,9	75,9	50,0	80,5	79,4
Teste						
R F	100	100	100	100	100	100
SVM	73,1	72,9	72,8	43,0	43,6	71,2
J48	95,4	95,7	95,5	90,0	90,8	98,6
MP	84,1	84,1	84,1	67,0	67,1	89,9

RF: Random Forest; SVM: Support Vector Machine; MP: Multilayer Perceptron

Na matriz de confusão, no conjunto de treinamento, o algoritmo RF classificou corretamente 94% e 70% das sementes germinadas de *A. colubrina* e *C. pyramidale*, respectivamente. As sementes não germinadas foram classificadas corretamente em 87% dos

resultados para ambas as espécies. No conjunto de teste, todas as categorias de sementes de *A. colubrina* e de *C. pyramidale*, foram classificadas corretamente em mais de 97%. O algoritmo J48 classificou corretamente, no conjunto de treinamento, 92% e 69% das sementes germinadas e 84% e 80% das sementes não germinadas de *A. colubrina* e *C. pyramidale*, respectivamente. No conjunto de teste, 97% e 90% das sementes germinadas e 96% e 99% das sementes não germinadas de *A. colubrina* e *C. pyramidale*, respectivamente (Tabela 4).

Tabela 4. Matriz de confusão do conjunto de treinamento e teste para classificação de germinação das sementes de *Anadenanthera colubrina* e de *Cenostigma pyramidale*, considerando as classes de sementes germinadas (sim) e sementes não germinadas (não).

<i>Anadenanthera colubrina</i>				
Algoritmo Random Forest				
	Treinamento		Teste	
Classe	Sim	Não	Sim	Não
Sim	495	29	203	4
Não	47	339	5	178
Algoritmo J48				
Sim	484	40	200	7
Não	59	327	8	175
<i>Cenostigma pyramidale</i>				
Algoritmo Random Forest				
	Treinamento		Teste	
Classe	Sim	Não	Sim	Não
Sim	345	149	209	0
Não	92	604	0	301
Algoritmo J48				
Sim	342	152	188	21
Não	142	554	2	299

Os resultados de classificação das plântulas de *A. colubrina* e *C. pyramidale* evidenciaram o bom desempenho do algoritmo RF. No conjunto de treinamento, o RF obteve acurácia, precisão, F-measure e ROC área acima de 75%, e Kappa e MCC superiores a 75% (para *A. colubrina*) e 55% (para *C. pyramidale*). No conjunto de teste, os resultados foram superiores, com métricas acima de 95%, em ambas as espécies. O algoritmo J48 foi eficiente para as duas espécies. As plântulas de *A. colubrina* obtiveram, no conjunto de treinamento, acurácia, precisão, F-measure e ROC área acima de 80%, e Kappa e MCC acima de 70%. As plântulas de *C. pyramidale* alcançaram, no conjunto de treinamento, acurácia, precisão, F-measure e ROC área acima de 70%, e Kappa e MCC acima de 45%. No conjunto de teste, foram observados valores acima de 90% em ambas as espécies (Tabela 5).

Tabela 5. Parâmetros de predição da classificação de plântulas (normais, anormais e mortas) de *Anadenanthera colubrina* e de *Cenostigma pyramidale* utilizando espectrômetro Vis-NIR portátil e algoritmos de aprendizado de máquina.

<i>Anadenanthera colubrina</i>						
Treinamento						
Algoritmo	Acurácia (%)	Precisão (%)	F-Measure (%)	Kappa (%)	MCC (%)	ROC Área (%)
R F	85,3	85,5	85,3	76,0	76,3	93,7
SVM	77,03	79,3	75,9	60,0	63,8	80,2
J48	81,6	81,7	81,6	70,0	70,6	86,4
MP	79,5	79,3	79,2	65,0	66,9	88,3
Teste						
R F	97,1	97,8	97,8	95,0	95,9	99,8
SVM	77,6	81,5	79,9	62,0	66,9	82,1
J48	94,3	94,4	94,3	90,0	90,8	97,8
MP	87,1	87,3	87,1	78,0	79,5	94,8
<i>Cenostigma pyramidale</i>						
Treinamento						
Algoritmo	Acurácia (%)	Precisão (%)	F-Measure (%)	Kappa (%)	MCC (%)	ROC Área (%)
R F	77,2	77,5	75,9	56,0	58,4	87,7
SVM	62,5	54,6	52,5	16,0	23,2	47,4
J48	70,0	70,0	70,0	47,0	48,6	74,7
MP	68,3	66,5	67,1	41,0	44,6	78,7
Teste						
R F	100	100	100	100	100	100
SVM	65,2	64,1	56,5	23,0	32,1	60,2
J48	95,6	95,7	95,6	92,0	92,7	98,7
MP	78,0	77,3	77,4	59,0	62,2	81,7

RF: Ramdon Forest; SVM: Support Vector Machine; MP: Multilayer Perceptron

Na matriz de confusão, no conjunto de treinamento, o algoritmo RF classificou corretamente plântulas normais em 89% e 57%, plântulas anormais em 66% e 49%, sem produção de plântulas em 88% 94% de *A. colubrina* e *C. pyramidale*, respectivamente no conjunto de teste, todas as categorias de plântulas de *A. colubrina* e de *C. pyramidale*, foram classificadas corretamente em mais de 97%. O algoritmo J48 classificou corretamente, no conjunto de treinamento, 82% e 60% das plântulas normais; 61% e 50% das plântulas anormais; e 87% e 80% das sementes sem produção de plântulas de *A. colubrina* e *C. pyramidale*, respectivamente. No conjunto de teste, todas as categorias de plântulas de *A. colubrina* e de *C. pyramidale*, foram classificadas corretamente em mais de 85% (Tabela 6).

Tabela 6. Matriz de confusão do conjunto de treinamento e teste para a classificação das sementes e plântulas de *Anadenanthera colubrina* e de *Cenostigma pyramidale* considerando as classes de plântulas (normais, anormais e mortas).

<i>Anadenanthera colubrina</i>						
Algoritmo Random Forest						
	Treinamento			Teste		
Classe	Normal	Anormal	Morta	Normal	Anormal	Morta
Normal	357	19	26	148	2	3
Anormal	35	79	5	1	51	0
Morta	42	6	341	3	2	180
Algoritmo J48						
Normal	332	37	33	145	2	6
Anormal	32	73	14	4	46	2
Morta	43	8	338	6	2	177
<i>Cenostigma pyramidale</i>						
Algoritmo Random Forest						
	Treinamento			Teste		
Classe	Normal	Anormal	Morta	Normal	Anormal	Morta
Normal	156	10	103	119	0	0
Anormal	22	111	91	0	91	0
Morta	31	14	652	0	0	300
Algoritmo J48						
Normal	162	41	66	114	4	1
Anormal	36	112	76	2	79	10
Morta	71	66	560	2	0	295

Os resultados da classificação do vigor das sementes de *A. colubrina* e *C. pyramidale* mostraram que os algoritmos RF e J48 apresentaram bom desempenho. O RF obteve, no conjunto de treinamento, acurácia, precisão, F-mesure e ROC área acima de 90% para *A. colubrina* e 65% para *C. pyramidale*, com Kappa e MCC superiores a 80% para *A. colubrina* e 45% para *C. pyramidale*. No conjunto de teste, o algoritmo atingiu 100% de predição para ambas as espécies. O J48 obteve, no treinamento, acurácia, precisão, F-mesure e área ROC acima de 85% para *A. colubrina* e 60% para *C. pyramidale*, com Kappa e MCC acima de 75% para *A. colubrina* e 35% para *C. pyramidale*. No conjunto de teste, os resultados foram superiores a 90% para ambas as espécies (Tabela 7).

Tabela 7. Predição do vigor das sementes (alto e nulo vigor) de *Anadenanthera colubrina* e de *Cenostigma pyramidale*, utilizando espectrômetro Vis- NIR portátil e algoritmos de aprendizagem de máquina

<i>Anadenanthera colubrina</i>						
Treinamento						
Algoritmo	Acurácia (%)	Precisão (%)	F-Measure (%)	Kappa (%)	MCC (%)	ROC Área (%)
R F	93,06	93,1	93,6	86,0	86,1	97,4
SVM	86,3	87,7	86,1	73,0	73,6	81,4
J48	88,3	88,4	88,3	76,0	76,6	89,9
MP	80,6	83,3	79,9	60,0	63,2	76,4
Teste						
R F	100	100	100	100	100	100
SVM	90,4	91,5	90,3	80,0	81,4	89,0
J48	99,5	99,5	99,5	99,0	99,0	99,0
MP	83,8	84,5	83,5	66,0	67,2	79,3
<i>Cenostigma pyramidale</i>						
Treinamento						
Algoritmo	Acurácia (%)	Precisão (%)	F-Measure (%)	Kappa (%)	MCC (%)	ROC Área (%)
R F	65,7	65,5	65,6	46,0	46,9	81,3
SVM	62,8	62,9	62,8	42,0	42,4	71,1
J48	60,0	60,6	60,4	39,0	39,4	70,4
MP	57,8	57,1	57,3	34,0	35,0	72,7
Teste						
R F	100	100	100	100	100	100
SVM	89,1	0,90	89,3	83,0	84,0	91,3
J48	94,1	94,1	94,1	90,0	91,1	97,1
MP	87,5	88,4	87,3	80,0	81,0	93,8

RF: Ramdon Forest; SVM: Support Vector Machine; MP: Multilayer Perceptron

Na matriz de confusão, no conjunto de treinamento, o algoritmo RF classificou corretamente 95% e 76% das sementes de *A. colubrina* e *C. pyramidale* com alto vigor, respectivamente, 91% e 58% das sementes com vigor nulo para *A. colubrina* e *C. pyramidale*, respectivamente, e 58% das sementes com baixo vigor, apenas para *C. pyramidale*. No conjunto de teste, todas as categorias de vigor de *A. colubrina* e de *C. pyramidale*, foram classificadas corretamente. O algoritmo J48 classificou corretamente no conjunto de treinamento 92% e 68% das sementes de *A. colubrina* e *C. pyramidale* com alto vigor, respectivamente, e 84% e 54% das sementes com vigor nulo para *A. colubrina* e *C. pyramidale*, respectivamente, e 54% das sementes com baixo vigor, apenas para *C. pyramidale*. No conjunto de teste, todas as categorias de sementes de *A. colubrina* e de *C. pyramidale*, foram corretamente classificadas em mais de 85% (Tabela 8).

Tabela 8. Matriz de confusão do conjunto de treinamento e teste para classificação do vigor das sementes de *Anadenanthera colubrina* e de *Cenostigma pyramidale*, considerando as classes de vigor das sementes (alto, baixo e nulo).

<i>Anadenanthera colubrina</i>						
Algoritmo Random Forest						
Treinamento			Teste			
Classes	Alto	Nulo	Alto	Nulo		
Alto	250	13	120	0		
Nulo	21	206	0	90		
Algoritmo J48						
Alto	242	21	120	0		
Nulo	36	191	1	89		
<i>Cenostigma pyramidale</i>						
Algoritmo Random Forest						
Treinamento			Teste			
Classes	Alto	Baixo	Nulo	Alto	Baixo	Nulo
Alto	86	21	9	48	0	0
Baixo	14	62	26	0	47	0
Nulo	9	17	36	0	0	25
Algoritmo J48						
Alto	79	25	12	46	0	2
Baixo	23	56	23	2	45	0
Nulo	13	15	34	1	2	22

Os resultados do conjunto de dados completo, referente à germinação das sementes, com a aplicação dos algoritmos RF e J48, obtiveram acurácia, precisão, F-measure e área ROC acima de 85%, com Kappa e MCC superiores a 60% para *A. colubrina* e *C. pyramidale* (Tabela 9). Na matriz de confusão, o algoritmo RF classificou corretamente as sementes germinadas de *A. colubrina* e *C. pyramidale* com 95% e 80% de acerto, respectivamente. Já as sementes não germinadas foram classificadas com 90% e 97% de acerto, respectivamente. O algoritmo J48 obteve 92% de acerto na classificação das sementes germinadas de *A. colubrina* e 77% para as de *C. pyramidale*. Quanto às sementes não germinadas, o algoritmo teve 85% de acerto para *A. colubrina* e 83% para *C. pyramidale* (Tabela 10).

Os resultados da classificação das plântulas de *A. colubrina* e *C. pyramidale* mostraram que os algoritmos RF e J48 apresentaram bom desempenho. Os dois algoritmos obtiveram acurácia, precisão, F-measure e área ROC acima de 75%, com Kappa e MCC superiores a 70% (*A. colubrina*) e 50% (*C. pyramidale*) (Tabela 9). Na matriz de confusão, o algoritmo RF classificou corretamente as plântulas normais em 91% e 67%; das plântulas anormais em 68% e 61%; sem produção de plântulas em 91% e 94% das espécies de *A. colubrina* e *C. pyramidale*, respectivamente. Por sua vez, o algoritmo J48 classificou corretamente as plântulas normais em 84% e 64%; das plântulas anormais em 55% e 60%; sem produção de plântulas em 90% e 85%, das espécies de *A. colubrina* e *C. pyramidale*, respectivamente (Tabela 10).

Os resultados da classificação do vigor das sementes de *A. colubrina* e *C. pyramidale*, com a aplicação do algoritmo RF, foram superiores a 85% para *A. colubrina* e 60% para *C. pyramidale* em todas as métricas. Já para o algoritmo J48, os valores ficaram acima de 80% para *A. colubrina*, enquanto para *C. pyramidale*, a acurácia, precisão, F-measure e ROC área ficaram acima de 65%, com os índices Kappa e MCC superiores a 50% (Tabela 9). Na matriz de confusão, o algoritmo RF classificou corretamente as sementes com alto vigor em 90% e 78%, e vigor nulo em 90% e 71%, para as espécies *A. colubrina* e *C. pyramidale*, respectivamente; e com baixo vigor, 76%, somente para as sementes de *C. pyramidale*. O algoritmo J48 classificou corretamente as sementes com alto vigor em 94% e 67%, e vigor nulo em 91% e 65%, para as sementes de *A. colubrina* e *C. pyramidale*, respectivamente; e com baixo vigor, 72%, apenas para *C. pyramidale* (Tabela 10).

Tabela 9. Modelos de validação cruzada da germinação, vigor das sementes e desenvolvimento de plântulas de *Anadenanthera colubrina* e de *Cenostigma pyramidale*, utilizando espectrômetro Vis- NIR portátil e algoritmos de aprendizagem de máquina.

Algoritmo	Acurácia (%)	Precisão (%)	F-Measure (%)	Kappa (%)	MCC (%)	ROC Área (%)
<i>Anadenanthera colubrina</i>						
Germinação						
R F	92,8	92,8	92,8	85,3	85,4	97,3
J48	89,3	89,5	89,3	78,1	78,1	92,7
Plântula						
R F	88,3	88,3	88,3	80,0	81,3	96,5
J48	83,07	82,7	82,8	71,6	72,6	88,4
Vigor						
R F	95,0	95,0	95,0	90,0	89,9	98,2
J48	92,4	92,4	92,4	84,6	84,6	93,4
<i>Cenostigma pyramidale</i>						
Germinação						
R F	85,9	85,9	85,9	70,9	71,0	93,5
J48	81,0	81,0	81,0	71,0	61,0	82,4
Plântula						
R F	82,3	82,3	81,7	67,4	68,7	89,1
J48	75,4	75,1	75,3	56,6	57,6	80,5
Vigor						
R F	75,7	75,9	62,7	75,7	62,6	89,4
J48	68,5	68,5	68,5	51,7	51,6	78,8

RF: Random Forest

Tabela 10. Matriz de confusão dos modelos de validação cruzada da germinação, vigor das sementes e desenvolvimento de plântulas de *Anadenanthera colubrina* e de *Cenostigma pyramidale*. Considerando as classificações da germinação de sementes (Sim e Não), plântulas (normais, anormais e morta) e do vigor de sementes (alto, baixo e nulo).

Anadenanthera colubrina										
Germinação			Plântula				Vigor			
Algoritmo Random Forest										
Classes	sim	Não	Classes	Normal	Anormal	Morta	Classe	Alto	Nulo	
Sim	706	38	Normal	523	23	29	Alto	377	14	
Não	55	501	Anormal	45	112	7	Nulo	21	288	
			Morta	37	10	514				
Algoritmo J48										
Sim	685	59	Normal	485	33	57	Alto	367	24	
Não	80	476	Anormal	49	91	24	Nulo	29	280	
			Morta	42	15	504				
Cenostigma pyramidale										
Algoritmo Random Forest										
Classe	Sim	Não	Classe	Normal	Anormal	Morta	Classe	Alto	Baixo	Nulo
Sim	574	141	Normal	269	17	107	Alto	110	25	6
Não	98	847	Anormal	30	198	92	Baixo	25	123	13
			Morta	30	24	933	Nulo	16	12	70
Algoritmo J48										
Sim	555	160	Normal	251	49	93	Alto	94	32	15
Não	163	822	Anormal	38	191	91	Baixo	31	116	14
			Morta	79	67	841	Nulo	20	14	64

5 DISCUSSÃO

Após o aprofundamento nas análises da bibliografia científica atual, verificou-se que este é o primeiro estudo a avaliar a germinação e o vigor de sementes florestais nativas do bioma Caatinga, utilizando a tecnologia Vis-NIR em combinação com algoritmos de aprendizado de máquina. Apesar desta tecnologia já ter sido testada para avaliar a qualidade de sementes de espécies cultivadas (AMBROSE et al., 2016; AMANAH et al., 2020; ANDRADE., et al., 2020; FENG et al., 2021), no entanto, os padrões morfológicos, bioquímicos e fisiológicos das sementes de espécies comerciais apresentam maior homogeneidade em comparação às sementes de espécies florestais.

O bom desempenho e precisão dessa tecnologia para monitorar a qualidade de sementes de espécies florestais nativas, o monitoramento da germinação e vigor das sementes representam um avanço significativo na análise de sementes, pois esta tecnologia oferece resultados rápidos, precisos e não destrutivos sobre o potencial fisiológico de diferentes lotes de sementes. A escolha dessas espécies para os estudos teve como propósito evidenciar que

essa tecnologia pode ser utilizada em sementes de espécies florestais nativas, facilitando as análises laboratoriais, bem como tomadas de decisões rápidas sobre ações de conservação ou plantio.

O estudo de Wang et al. (2021) identificou que os comprimentos de onda selecionados se concentraram em faixas específicas, como 550 nm e 780 nm, que são biologicamente relevantes para avaliar o vigor das sementes. Segundo os autores, a faixa de 500 nm a 780 nm pertence ao espectro visível e pode refletir as variações de cor entre diferentes amostras. As sementes de *A. colubrina* e *C. pyramidale* produziram plântulas normais e anormais (sementes germinadas), apresentando picos de absorção em torno de 500 nm a 700 nm. No entanto, não foi possível observar um forte pico de absorção espectral na classe sem produção de plântulas (sementes não germinadas). As classes com alto vigor e vigor nulo mostraram picos de absorção semelhantes, sem discriminação espectral para essas classes na espécie *A. colubrina*. Os espectros médios das sementes e plântulas de *C. pyramidale* exibiram padrões semelhantes nas curvas espectrais de todas as classes de germinação e vigor, com picos de absorção entre 500 nm e 700 nm.

Esses espectros são sensíveis a diferentes componentes químicos presentes nas sementes, como água, proteínas e carboidratos. Pesquisas anteriores indicam que as bandas em 605 nm e 540 nm correspondem à quarta e quinta harmônica do amido, enquanto o pico em 643 nm está relacionado à terceira e quarta harmônica das proteínas (CHENG et al., 2023; WANG et al., 2023).

Além disso, outros comprimentos de onda foram associados às ligações químicas C-H, O-H e N-H, relacionadas a água, carboidratos e proteínas nas sementes de milho. O estudo de Peng et al. (2017) demonstrou que a espectroscopia Vis-NIR pode ser aplicada para analisar o teor dos componentes principais das sementes de milho, como proteínas, amidos e gorduras.

Ao longo do desenvolvimento das sementes, há um acúmulo de reservas ricas em carboidratos, como açúcares, amido e celulose. Esses componentes se concentram nas fases intermediárias e finais da maturação, sendo essenciais para a nutrição do embrião durante a germinação, além de promover o vigor e o crescimento das plântulas, assim como as proteínas (ERBAŞ et al., 2016).

As proteínas exercem papéis importantes para o crescimento e desenvolvimento do eixo embrionário, assim como também para a formação e emergência de plântulas no campo. Quando as sementes passam pelo processo de deterioração acontece a desnaturação e redução do conteúdo e síntese de proteínas, sendo a degradação proteica um dos mecanismos

indicados como causa na redução da viabilidade das sementes (PINHEIRO et al., 2023). Para Marcos Filho (2015), a maior evidência do início da morte de sementes é o declínio da velocidade de germinação de sementes viáveis, seguida por uma redução no tamanho das plântulas e aumento na incidência de plântulas anormais. Segundo o mesmo autor a baixa porcentagem de germinação é associada a sinais precoce de desorganização da membrana, assim se tem a ocorrência de anormalidade nas plântulas que é atribuída à morte significativa de tecidos em diferentes partes das sementes, principalmente em tecidos meristemáticos.

Sementes de alto vigor conseguem resistir a longos períodos de armazenamento devido à regulação de genes na planta mãe, como aqueles responsáveis pelas proteínas LEA, que inibem a germinação durante o desenvolvimento das sementes (WU et al., 2017). Segundo Ratajczak et al. (2015), as sementes são uma adaptação importante das plantas para a reprodução, mas também sofrem envelhecimento, o que reduz a capacidade de germinação e impede o desenvolvimento de novas plantas. Durante o armazenamento, as sementes sofrem deterioração, com alterações como o acúmulo de ROS, peroxidação de lipídios, danos nas membranas celulares, inativação de enzimas, redução do metabolismo energético e degradação do DNA (KONG et al., 2015). Radicais livres gerados no metabolismo do oxigênio oxidam diretamente componentes celulares e atuam como sinais moleculares de estresse, resultando na diminuição da viabilidade das sementes (TIAN et al., 2019).

Alguns estudos sugerem que ocorra a discriminação do perfil espectral da germinação e do vigor de sementes através dos compostos químicos (DANESHVAR et al., 2015; AL-AMERY et al., 2018; CHENG et al., 2023). Lohumi et al. (2013), em sua pesquisa sobre a avaliação de sementes de melancia utilizando NIR, mostraram que é possível distinguir entre sementes que germinam e aquelas que não germinam, com base nos componentes químicos presentes nas sementes. Nesse contexto, os espectros de reflectância observados nas bandas de absorção neste estudo podem estar relacionados às características de proteínas, carboidratos e lipídios.

Neste estudo, observou-se que a aplicação do filtro *resample*, juntamente com a utilização de diferentes algoritmos, promoveu um bom desempenho preditivo para os modelos de germinação e vigor das sementes de *A. colubrina* e *C. pyramidale*. Esse filtro foi aplicado para remover o enviesamento, conforme descrito por Gadotti et al. (2022 a, b). O enviesamento ocorreu devido à maior quantidade de sementes germinadas em comparação com as amostras de sementes não germinadas.

De maneira geral, os modelos de classificação para germinação, plântulas, vigor de sementes de *A. colubrina* e *C. pyramidale*, utilizando os algoritmos RF e J48, apresentaram um desempenho razoável nas métricas, tanto no conjunto de treinamento quanto no conjunto de teste. Os resultados do conjunto completo sugerem que os modelos com os algoritmos RF e J48 apresentaram um bom desempenho para a classificação da germinação, plântulas e vigor das sementes das duas espécies, com valores elevados nas métricas. Embora os modelos tenham demonstrado boa capacidade de classificação, é importante ressaltar que o aumento na quantidade de sementes utilizadas poderia aprimorar a acurácia das predições em todas as classes.

Pesquisas com NIRS relacionadas ao vigor de sementes são diversas e focam na classificação entre sementes viáveis ou não viáveis para diversas espécies de sementes de *Pinus sylvestris* (LESTANDER e ODÉN, 2002), sementes de *Citrullus lanatus* (LOHUMI et al., 2013), sementes de *Spinacia oleracea* L. e *Ricinus cummunis* L. (OLESEN, et al., 2011; OLSEN et al., 2015), grãos de milho híbridos comerciais (*Zea mays* L.) (AMBROSE et al., 2016) e *Cunninghamia lanceolata* (Lamb) Hook) (TIGABU et al., 2020), assim como também alguns trabalhos relacionaram o NIRS com a classificação de sementes com alto, médio e baixo vigor, obtendo resultados significativos (ZHANG et al., 2020; SILVA, et al., 2024).

No estudo de Fan et al. (2020), a espectroscopia Vis-NIR foi aplicada em conjunto com um algoritmo de aprendizado de máquina para classificar diferentes níveis de vigor das sementes de trigo. Os resultados demonstraram que essa combinação de Vis-NIR e aprendizado de máquina pode ser promissora para estimar com alta precisão a qualidade das sementes.

A avaliação do teste de germinação e vigor são os métodos mais comumente utilizados para avaliar a qualidade fisiológica das sementes, servindo como critério para aceitar ou rejeitar lotes (KISHK e ELBATRAWY, 2023). Vale ressaltar que, para as sementes analisadas neste estudo, esse teste exige um período de duas a quatro semanas para a obtenção dos resultados (ANDRADE et al., 2020). Assim, a implementação de novas metodologias capazes de avaliar o potencial fisiológico das sementes e realizar uma pré-seleção de lotes de baixa e alta qualidade pode agilizar os processos de tomada de decisão sobre o destino desses lotes.

Ao analisar as matrizes de confusão dos dados particionados de ambas as espécies, foi possível verificar que os algoritmos utilizados apresentaram altas taxas de acerto na

classificação das sementes germinadas e não germinadas, tanto no conjunto de treinamento quanto no de teste. Este bom desempenho sugere que os algoritmos são eficazes na distinção entre sementes germinadas e não germinadas, independentemente do conjunto de dados.

No que diz respeito à classificação das plântulas, os resultados foram um pouco diferentes. Nos dados de treinamento, os algoritmos obtiveram mais acertos na categoria de plântulas não produzidas (mortas), enquanto, no conjunto de teste, apresentaram resultados superiores em todas as categorias de plântulas. Isso indica que, embora o desempenho tenha sido bom no treinamento, os algoritmos conseguiram generalizar melhor para novas amostras no teste, alcançando uma taxa de acerto consistente em todas as classes de plântulas.

Em relação ao vigor das sementes, foi possível observar que, no conjunto de treinamento, dois algoritmos se destacaram ao acertarem com mais precisão as classes de alto vigor das sementes de ambas as espécies. Esse resultado é relevante, pois demonstra que os algoritmos foram eficazes em identificar sementes com alto vigor, um fator importante para prever a qualidade das sementes. Já no conjunto de teste, esses mesmos algoritmos mantiveram seu bom desempenho, com altas taxas de acerto em todas as classes de vigor das sementes, confirmando sua capacidade de generalizar e apresentar bons resultados mesmo quando avaliados com novos dados.

Elevados valores de germinação e vigor são fundamentais para garantir o bom estabelecimento das plântulas, o que, por sua vez, contribui para a uniformidade do estande e das plantas no campo (FINCH-SAVAGE e BASSEL, 2016).

Na classificação do vigor das sementes de ambas as espécies, a categoria de alto vigor registrou o maior número de acertos, a razão disso é que alguns lotes apresentam uma alta porcentagem de germinação. Para a classificação da germinação e das plântulas de *C. pyramidale*, observou-se um maior número de acertos para as sementes que não germinaram, resultando em uma baixa produção de plântulas. Isso ocorreu porque alguns lotes utilizados nos modelos apresentaram baixa qualidade. No trabalho de Mayrink et al. (2020) foi utilizada a espectroscopia Vis-NIR para determinar as classes de sementes de algodão com baixa e alta qualidade obtendo resultados satisfatórios na classificação dos modelos.

Observou-se um maior desequilíbrio nas classes de plântulas anormais em ambas as espécies. Esse desequilíbrio ocorreu devido ao fato de que as classes com maior taxa de erro continham um número limitado de plântulas dessa categoria, o que comprometeu a precisão da classificação pelos algoritmos. Em todas as classes de ambas as espécies, foi observado que o conjunto de treinamento apresentou mais erros, os quais foram superados no conjunto

de teste, que obteve maior número de acertos. No estudo de Bernardy et al. (2023) sobre a classificação de lotes de sementes com o uso de algoritmos de aprendizado de máquina, foram observados erros em uma das classes no conjunto de treinamento. No entanto, a maioria dos dados de teste foi classificada corretamente. Isso indica que a técnica de treinamento utilizada foi adequada para a tarefa.

Ao realizar a validação cruzada K-fold em todo o conjunto de dados, observou-se, nas matrizes de confusão, que os modelos se ajustaram com precisão, resultando em taxas de acerto mais altas nas classificações de sementes germinadas para ambas as espécies. Esse resultado indica que os algoritmos foram eficazes na identificação correta das sementes que germinaram. Além disso, no caso das sementes que não produziram plântulas, os dois algoritmos mais eficazes se destacaram, alcançando os maiores índices de acerto. Esse bom desempenho na distinção entre sementes germinadas e não germinadas sugere uma alta precisão dos modelos na classificação das diferentes condições das sementes. De forma semelhante, para as sementes de alto e nulo vigor, os algoritmos também demonstraram um bom desempenho, com elevados índices de acerto nas duas categorias de vigor. Isso reflete a capacidade dos modelos de identificar corretamente os diferentes níveis de vigor das sementes em ambas as espécies, evidenciando a robustez e a confiabilidade dos algoritmos aplicados nesse tipo de análise.

Este estudo demonstrou que, de maneira geral, tanto as classes com particionamento quanto a classificação sem particionamento (conjuntos de dados completos) apresentaram precisões eficientes nas classes avaliadas. Além disso, os resultados indicaram que os algoritmos RF e J48 possuem a capacidade de prever os modelos com boa precisão.

Segundo Zhang et al. (2018), a classificação do vigor das plântulas é fundamental para avaliar rapidamente o potencial fisiológico dos lotes de sementes durante os testes de germinação. Identificar sementes que produzem plântulas saudáveis é importante, pois possibilita uma avaliação precisa e apoiar decisões estratégicas sobre o destino dos lotes. Dessa forma, tecnologias que contribuam para esse processo são extremamente valiosas para o setor de sementes, promovendo melhorias na qualidade e segurança das sementes.

Os resultados deste estudo indicam que é viável prever a germinação, o vigor das plântulas e o vigor das sementes usando técnicas de Vis-NIR e aprendizado de máquina para as duas espécies avaliadas. No entanto, é necessário um maior número de amostras de sementes para melhorar a precisão dos modelos de todas as classes. Portanto, é importante implementar medidas, como a criação de bancos de dados de sementes florestais nativas do

bioma Caatinga, para apoiar pesquisas que utilizam essas técnicas. Com um maior número de sementes, a confiabilidade das predições pode ser significativamente aumentada.

A implementação de técnicas de aprendizagem de máquina é fundamental na área de sementes, pois os testes necessários para avaliar sua qualidade geram um grande volume de dados, tornando impossível para a capacidade humana processá-los de forma rápida e eficiente em um laboratório de controle de qualidade, especialmente em prazos curtos (PINHEIRO et al., 2021). Essas dificuldades tornam-se ainda maiores para sementes de espécies nativas, como as avaliadas neste estudo pertencentes ao bioma Caatinga.

Conclusão

A espectroscopia Vis-NIR aliada ao aprendizado de máquina é uma alternativa eficaz para distinguir as classes de germinação, plântulas e vigor das sementes de *A. colubrina* e *C. pyramidale*,

Os algoritmos RF e J48 são recomendados para a predição e determinação da germinação, do desenvolvimento de plântulas e do vigor das sementes de florestas secas.

Apesar dos resultados promissores, é fundamental incluir mais amostras de sementes para aumentar a precisão e robustez dos modelos de calibração, visando o uso dessa tecnologia na rotina de análise de sementes.

6 REFERÊNCIAS

AL-AMERY, M. et al. Near-infrared spectroscopy used to predict soybean seed germination and vigour. **Seed Science Research**, v. 28, n. 3, p. 245-252, 2018.

<https://doi.org/10.1017/S0960258518000119>

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, 2020.

<https://doi.org/10.1016/j.infrared.2020.103477>

AMBROSE, A et al. Comparative nondestructive measurement of corn seed viability using Fourier transform near-infrared (FT-NIR) and Raman spectroscopy. **Sensors and Actuators B: Chemical**, v. 224, p. 500-506, 2016. <https://doi.org/10.1016/j.snb.2015.10.082>

ANDRADE, G. C. et al. 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, 2020. <https://doi.org/10.1016/j.heliyon.2020.e03477>

BRASIL. Ministério da Agricultura, Pecuária e Abastecimento. Instruções para análises de sementes florestais Brasília, DF, 2013. Disponível em: <https://www.gov.br/agricultura/pt-br/assuntos/lfd/a/arquivos-publicacoes-laboratorio/florestal_documento_pdf-ilovepdf-compressed.pdf>. Acesso em: 28 ago, 2023.

BRASIL. Ministério da Agricultura e Reforma Agrária. **Regras para Análises de Sementes**. Brasília,DF:CLAV:DNDV:SNDA:MA,2009.Diponivelem:<https://www.gov.br/agricultura/pt-br/assuntos/insumos299_agropecuarios/arquivos-publicacoes-insumos/2946_regras_analise_sementes.pdf

BERNARDY, R. et al. FITTING Data Mining Settings for Ranking Seed Lots. **Engenharia Agrícola**, v. 43, n. 2, p. e20220193, 2023. <https://doi.org/10.1590/1809-4430-Eng.Agric.v43n2e20220193/2023>

CHEN, Bing-Xian et al. Identification of metabolomic biomarkers of seed vigor and aging in hybrid rice. **Rice**, v. 15, n. 1, p. 7, 2022. <https://doi.org/10.1186/s12284-022-00552-w>

CHENG, T. et al. Hyperspectral and imagery integrated analysis for vegetable seed vigor detection. **Infrared Physics & Technology**, v. 131, p. 104605, 2023. <https://doi.org/10.1016/j.infrared.2023.104605>

CHEYED, S. H. Effect of storage method and period on vitality and vigour of seed wheat. **Indian Journal of Ecology**, v.47, n.10, p.2731, 2020. https://www.researchgate.net/publication/340732516_Effect_of_Storage_Method_and_Period_on_Vitality_and_Vigour_of_Seed_Wheat_Manuscript_Number_S-324_NAAS_Rating_496

DANESHVAR, A. et al. Single seed Near Infrared Spectroscopy discriminates viable and non-viable seeds of *Juniperus polycarpus*. **Silva Fennica**, v. 49, n. 5, 2015. <https://doi.org/10.14214/sf.1334>

ERBAŞ, S et al. Mobilization of seed reserves during germination and early seedling growth of two sunflower cultivars. **Journal of Applied Botany and Food Quality**, v. 89, p217-222, 2016. <http://dx.doi.org/10.5073/JABFQ.2016.089.028>

FAN, Y. et al. Individual wheat kernels vigor assessment based on NIR spectroscopy coupled with machine learning methodologies. **Infrared Physics & Technology**, v. 105, p. 103213, 2020. <https://doi.org/10.1016/j.infrared.2020.103213>

FENG, L. et al. Assessment of rice seed vigour using selected frequencies of electrical impedance spectroscopy. **Biosystems Engineering**, v. 209, p. 53-63, 2021. <https://doi.org/10.1016/j.biosystemseng.2021.06.011>

FINCH-SAVAGE, W E. et al. Seed vigour and crop establishment: extending performance beyond adaptation. **Journal of experimental botany**, v. 67, n. 3, p. 567-591, 2016. <https://doi.org/10.1093/jxb/erv490>

FREMOUT, T. et al. Diversity for Restoration (D4R): Guiding the selection of tree species and seed sources for climate-resilient restoration of tropical forest landscapes. **Journal of Applied Ecology**, v. 59, n. 3, p. 664-679, 2022. <https://doi.org/10.1111/1365-2664.14079>

GADOTTI, G. I et al. Machine learning for soybean seeds lots classification. **Engenharia Agrícola**, v.42, n. spe ep.20210101, 2022a. <http://dx.doi.org/10.1590/1809-4430-Eng.Agric.v42nepe20210101/2022>

GADOTTI G.I et al. Prediction of ranking of lots of corn seeds by artificial intelligence. **Engenharia Agrícola**, v.42, n.4 p. e20210005, 2022b. <http://dx.doi.org/10.1590/1809-4430-Eng.Agric.v42n4e20210005/2022>

GASPERINI, C. et al. Edge effects on the realised soil seed bank along microclimatic gradients in temperate European forests. **Science of the Total Environment**, v. 798, p. 149373, 2021. <https://doi.org/10.1016/j.scitotenv.2021.149373>

JIA, Z. et al. A novel approach using multispectral imaging for rapid development of seed pellet formulations to mitigate drought stress in alfalfa. **Computers and Electronics in Agriculture**, v. 212, p. 108136, 2023. <https://doi.org/10.1016/j.compag.2023.108136>

KONG, L. et al. Antioxidant response and related gene expression in aged oat seed. **Frontiers in plant science**, v. 6, p. 158, 2015. <https://doi.org/10.3389/fpls.2015.00158>

KRZYZANOWSKI, F.C; FRANÇA-NETO, J.B. Vigor de Sementes. [Seed Vigour]. **Informativo ABRATES** (Londrina), 11, 81-84.2001

KISHK, A; ELBATRAWY, W. S. The relationship between seed vigor tests and field emergence of wheat lots. **Egyptian Journal of Agricultural Research**, v. 101, n. 4, p. 1054-1061, 2023. https://ejar.journals.ekb.eg/article_323542.html

LEITE, D.R.A. et al. Método de Aprendizagem de Máquina para Classificação da intensidade do desvio vocal utilizando Random Forest. **Journal of Health Informatics**, v. 12, 2020. <https://jhi.sbis.org.br/index.php/jhi-sbis/article/view/814>

LESTANDER T.A; ODÉN PC. Separation of viable and non-viable filled Scots pine seeds by differentiating between drying rates using single seed near infrared transmittance spectroscopy. **Seed Science and Technology** v.30, 383–392,2002.

LOHUMI, S.et al. Nondestructive evaluation for the viability of watermelon (*citrullus lanatus*) seeds using fourier transform near infrared spectroscopy. **Journal of Biosystems Engineering**, v. 38, n. 4, p. 312-317, 2013. <https://doi.org/10.5307/JBE.2013.38.4.312>

MARCOS FILHO, J. Seed vigor testing: an overview of the past, present and future perspective. *Scientia agricola*, v. 72, p. 363-374, 2015. <https://doi.org/10.1590/0103-9016-2015-0007>

MAYRINCK, L. G et al. Use of near infrared spectroscopy in cotton seeds physiological quality evaluation. **Journal of Seed Science**, v. 42, p. e202042016, 2020. <https://doi.org/10.1590/2317-1545v42227169>

MUKASA, P. et al. Determination of viability of *Retinispora (Hinoki cypress)* seeds using FT-NIR spectroscopy. **Infrared Physics & Technology**, v. 98, p. 62-68, 2019. <https://doi.org/10.1016/j.infrared.2019.02.008>

OLESEN, M. H. et al. Classification of viable and non-viable spinach (*Spinacia oleracea* L.) seeds by single seed near infrared spectroscopy and extended canonical variates analysis. **Journal of Near Infrared Spectroscopy**, v. 19, n. 3, p. 171-180, 2011. <https://opg.optica.org/jnirs/abstract.cfm?URI=jnirs-19-3-171>

OLESEN, M. H. et al. Viability prediction of *Ricinus cummunis* L. seeds using multispectral imaging. **Sensors**, v. 15, n. 2, p. 4592-4604, 2015. <https://doi.org/10.3390/s150204592>

PENG, Y; LIU, H. Determination of rice amylose content by near- infrared spectroscopy. **Agric. Sci. Technol.** v.18, n.4, p.729–732, 2017. <https://www.proquest.com/openview/1f2715314a5225fe23c60afe59d12aa9/1?pq-origsite=gscholar&cbl=1596357>

PINHEIRO, M.R et al. Inteligência artificial na agricultura com aplicabilidade no setor sementeiro. **Diversitas Journal**, v. 6, n. 3, p. 2996-3012, 2021. https://doi.org/10.48017/Diversitas_Journal-v6i3-1857

PINHEIRO, D.T et al. Oxidative stress, protein metabolism, and physiological potential of soybean seeds under weathering deterioration in the pre-harvest phase. **Acta Scientiarum. Agronomy**, v. 45, p. e56910, 2023. <https://doi.org/10.4025/actasciagron.v45i1.56910>

RATAJCZAK, E. et al. The production, localization and spreading of reactive oxygen species contributes to the low vitality of long-term stored common beech (*Fagus sylvatica* L.) seeds. **Journal of Plant Physiology**, v. 174, p. 147-156, 2015. <https://doi.org/10.1016/j.jplph.2014.08.021>

RODRIGUES, V. Entenda o que é AUC e ROC nos modelos de Machine Learning. 2018. Disponível em< <https://medium.com/bio-data-blog/entenda-o-que-%C3%A9-auc-e-roc-nos-modelos-de-machine-learning-8191fb4df772>> Acesso em 05 nov. 2023.

RODRIGUES, V. Métricas de Avaliação: acurácia, precisão, recall... quais as diferenças? 2019. Disponível em <<https://vitorborbarodrigues.medium.com/m%C3%A9tricas-de-avalia%C3%A7%C3%A3o-acur%C3%A1cia-precis%C3%A3o-recall-quais-as-diferen%C3%A7as-c8f05e0a513c>> Acesso em 05 nov. 2023.

ROSSI, R.F. et al. Vigor de sementes, população de plantas e desempenho agrônomo de soja. **Revista de Ciências Agrárias Amazonian Journal of Agricultural and Environmental Sciences**, v. 60, n. 3, p. 215-222, 2017.

REED, R et al. Imtiyaz. Seed germination and vigor: ensuring crop sustainability in a changing climate. **Heredity**, v. 128, n. 6, p. 450-459, 2022. <https://doi.org/10.1038/s41437-022-00497-2>

SILVA, M. F. et al. Near infrared spectroscopy for the classification of vigor level of soybean seed. **Revista Ciência Agronômica**, v. 55, p. e20238703, 2024. <https://doi.org/10.5935/1806-6690.20240005>

SCHMIDT, I. B et al. Community-based native seed production for restoration in Brazil—the role of science and policy. **Plant Biology**, v. 21, n. 3, p. 389-397, 2019. <https://doi.org/10.1111/plb.12842>

TIAN, P. P et al. Effect of artificial aging on wheat quality deterioration during storage. **Journal of Stored Products Research**, v. 80, p. 50-56, 2019. <https://doi.org/10.1016/j.jspr.2018.11.009>

TIGABU, M. et al. Rapid and non-destructive evaluation of seed quality of Chinese fir by near infrared spectroscopy and multivariate discriminant analysis. **New Forests**, v. 51, p. 395-408, 2020. <https://doi.org/10.1007/s11056-019-09735-8>

WANG, Y et al. A review of the application of near-infrared spectroscopy (NIRS) in forestry. **Applied Spectroscopy Reviews**, v. 57, n. 4, p. 300-317, 2022. <https://doi.org/10.1080/05704928.2021.1875481>

WANG, Y. et al. Application of hyperspectral imaging assisted with integrated deep learning approaches in identifying geographical origins and predicting nutrient contents of Coix seeds. **Food Chemistry**, v. 404, p. 134503, 2023. <https://doi.org/10.1016/j.foodchem.2022.134503>

WANG, Y. et al. Discriminant analysis and comparison of corn seed vigor based on multiband spectrum. **Computers and Electronics in Agriculture**, v. 190, p. 106444, 2021. <https://doi.org/10.1016/j.compag.2021.106444>

WU, X et al. Genetic modification for improving seed vigor is transitioning from model plants to crop plants. **Frontiers in Plant Science**, v. 8, p. 8, 2017. <https://doi.org/10.3389/fpls.2017.00008>

XIA, Y. et al. Recent advances in emerging techniques for non-destructive detection of seed viability: A review. **Artificial Intelligence in Agriculture**, v. 1, p. 35-47, 2019. <https://doi.org/10.1016/j.aiia.2019.05.001>

ZHANG, H. et al. Proteome analysis of poplar seed vigor. **PloS one**, v. 10, n. 7, p. e0132509, 2015. <https://doi.org/10.1371/journal.pone.0132509>

ZHANG, L. et al. Hyperspectral imaging technology combined with deep forest model to identify frost-damaged rice seeds. **Spectrochimica acta part A: molecular and biomolecular spectroscopy**, v. 229, p. 117973, 2020. <https://doi.org/10.1016/j.saa.2019.117973>

ZHANG, T. et al. Non-destructive analysis of germination percentage, germination energy and simple vigour index on wheat seeds during storage by Vis/NIR and SWIR hyperspectral imaging. **Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy**, v. 239, p. 118488, 2020. <https://doi.org/10.1016/j.saa.2020.118488>

ZHANG, T. et al. A reliable methodology for determining seed viability by using hyperspectral data from two sides of wheat seeds. **Sensors**, v.18, n.3, p.813,2018.
<https://doi.org/10.3390/s18030813>

ZHAO, J et al. Advances in the identification of quantitative trait loci and genes involved in seed vigor in rice. **Frontiers in Plant Science**, v. 12, p. 659307, 2021.
<https://doi.org/10.3389/fpls.2021.659307>