

UNIVERSIDADE FEDERAL DE MATO GROSSO DO SUL  
CÂMPUS DE CHAPADÃO DO SUL  
CURSO DE GRADUAÇÃO EM AGRONOMIA

GABRIELA SOUZA OLIVEIRA

**APRENDIZAGEM DE MÁQUINA E SENSORIAMENTO HIPERESPECTRAL  
NA PREDIÇÃO DA QUALIDADE FISIOLÓGICA DE SEMENTES DE SOJA**

CHAPADÃO DO SUL – MS

2024

UNIVERSIDADE FEDERAL DE MATO GROSSO DO SUL  
CÂMPUS DE CHAPADÃO DO SUL  
CURSO DE GRADUAÇÃO EM AGRONOMIA

**APRENDIZAGEM DE MÁQUINA E SENSORIAMENTO HIPERESPECTRAL  
NA PREDIÇÃO DA QUALIDADE FISIOLÓGICA DE SEMENTES DE SOJA**

Trabalho de Conclusão de Curso  
apresentado à Universidade Federal de  
Mato Grosso do Sul, como parte dos  
requisitos para obtenção do título de  
Engenheira Agrônoma.

ORIENTADORA: Prof. Dr.(a). Larissa Pereira Ribeiro Teodoro

CHAPADÃO DO SUL-MS

2024

06/12/2024, 05:03

SEI/UFMS - 5250185 - Certificado



Serviço Público Federal  
Ministério da Educação  
Fundação Universidade Federal de Mato Grosso do Sul



### CERTIFICADO DE APROVAÇÃO

AUTORA: **GABRIELA SOUZA OLIVEIRA.**

ORIENTADORA: **Profa. Dra. Larissa Pereira Ribeiro Teodoro**

Aprovada pela Banca Examinadora como parte das exigências do Componente Curricular Não Disciplinar TCC, para obtenção do grau de BACHARELA EM AGRONOMIA, pelo curso de Bacharelado em Agronomia da Universidade Federal de Mato Grosso do Sul, Câmpus de Chapadão do Sul.

**Profa. Dra. Larissa Pereira Ribeiro Teodoro**  
Presidente da Banca Examinadora e Orientador

**Profa. Dra. Ana Carina da Silva Cândido Seron**  
Membro da Banca Examinadora

**Prof. Dr. Paulo Eduardo Teodoro**  
Membro da Banca Examinadora

Chapadão do Sul, 21 de novembro de 2024.

NOTA  
MÁXIMA  
NO MEC

UFMS  
É 10!!!



Documento assinado eletronicamente por **Paulo Eduardo Teodoro, Professor do Magisterio Superior**, em 21/11/2024, às 16:20, conforme horário oficial de Mato Grosso do Sul, com fundamento no § 3º do art. 4º do [Decreto nº 10.543, de 13 de novembro de 2020](#).

NOTA  
MÁXIMA  
NO MEC

UFMS  
É 10!!!



Documento assinado eletronicamente por **Larissa Pereira Ribeiro, Professora do Magistério Superior**, em 21/11/2024, às 16:20, conforme horário oficial de Mato Grosso do Sul, com fundamento no § 3º do art. 4º do [Decreto nº 10.543, de 13 de novembro de 2020](#).

NOTA  
MÁXIMA  
NO MEC

UFMS  
É 10!!!



Documento assinado eletronicamente por **Ana Carina da Silva Candido, Professora do Magistério Superior**, em 21/11/2024, às 16:58, conforme horário oficial de Mato Grosso do Sul, com fundamento no § 3º do art. 4º do [Decreto nº 10.543, de 13 de novembro de 2020](#).

## **DEDICATÓRIA**

Dedico esse trabalho primeiramente a Deus, que esteve presente em toda minha trajetória acadêmica, me dando graças e sabedoria para enfrentar cada momento e por sempre agir no ordinário cada dia da minha vida.

Para todas as pessoas que me apoiaram e contribuíram de inúmeras formas, como também dedico a mim mesma por confiar em Deus, por cada esforço e por sempre buscar fazer as coisas com excelência.

## AGRADECIMENTOS

Á A Deus por ter planos bem maiores e melhores que os meus, por abençoar minha vida e cuidar de cada detalhe com muito amor, por me fazer lembrar, em meio às dificuldades, de ser forte e corajosa. Toda honra e glória seja dada a Ele.

À minha orientadora, Profa. Dra. Larissa Pereira Ribeiro Teodoro por me apoiar, incentivar e impulsionar a ser melhor a cada dia, acreditando no meu potencial. Uma das minhas inspirações como professora e pesquisadora.

Aos meus pais, Manoel Flávio de Oliveira e Cristiane De Oliveira Souza, por me incentivarem desde criança a ter responsabilidade, buscar sempre o estudo e principalmente por me derem asas para voar, confiando e me apoiando. Mesmo longe de casa, sempre vou honrar vocês, fazendo cada trabalho com máxima excelência.

Aos meus irmãos, Mikaela e Jose Guilherme, por serem meus maiores admiradores e com isso me impulsionarem a ser sempre um bom exemplo a seguir. Com todo carinho, me fez me manter forte longe de casa.

Ao meu namorado, Thomas Forlin, por cada conselho, por sempre ser um ombro amigo, por não medir esforços para que eu esteja bem e feliz. Obrigada por fazer com que eu me veja com mais amor, e não com muitas cobranças.

Aos meus tios, Sinésio e Liliane, que me receberam em sua casa, oferecendo todo apoio para que eu pudesse concluir a faculdade. Obrigada por cada carinho e incentivo, sempre serão minha inspiração como pessoas. As minhas primas, Lívia e Lorena, por cada momento de descontração em meio ao caos.

Aos meus melhores amigos, Gleciene, Marcos Eduardo e Lauani, por fazer essa trajetória leve e com muita risada. Vocês são muito especiais para mim.

Ao meu grupo de estudo GEMP, em especial à Dthenifer Cordeiro Santana, que desde o início me acolheu com muito carinho, apoio em cada parte desse trabalho e sempre com conselhos e ajudas preciosas. Obrigada pela amizade que construímos, sempre será uma inspiração para mim de dedicação e amor pela pesquisa.

De modo geral, gostaria de expressar minha gratidão à Universidade Federal de Mato Grosso do Sul (UFMS/CPCS) pelos recursos disponibilizados durante minha formação e a realização da pesquisa. Agradeço também aos professores pelos conhecimentos que compartilharam comigo, que foram essenciais para minha formação e desenvolvimento.

## **PREDIÇÃO DE GERMINAÇÃO DE SEMENTE DE SOJA UTILIZANDO APRENDIZAGEM DE MÁQUINA.**

**RESUMO-** Sementes de soja com qualidade apresentam características genéticas, físicas e fisiológicas que interferem diretamente no rendimento da cultura. A utilização de sensores hiperspectrais, juntamente com a aprendizagem de máquina (ML), pode facilitar e otimizar o tempo para os testes de germinação. Com isso, os objetivos desse trabalho foram: i) avaliar se as refletâncias das folhas e sementes são eficazes na predição da qualidade fisiológica de sementes de soja, utilizando algoritmos de ML e ii) identificar qual algoritmo proporciona melhor acurácia nessa predição. Trinta e dois genótipos de soja foram avaliados em experimento à. As folhas e sementes foram analisadas com um sensor hiperspectral, equipamento capaz de medir o espectro na faixa de 350 a 2500 nm. Os dados foram submetidos à análise de ML com dois tipos de entradas: variáveis espectrais de folha e semente, bem como a variável a ser predita (output) que foram teste de germinação (GERM), condutividade elétrica (CE), primeira contagem de germinação (PCG), teste de tetrazólio vigoroso (TZV) e teste de tetrazólio viáveis (VIAB). Essa predição foi desempenhada por meio de validação cruzada estratificada com k-fold = 10 e dez repetições (100 execuções para cada modelo). Todos os parâmetros dos modelos foram estabelecidos de acordo com a configuração default do software Weka 3.8.5. Os modelos de ML utilizados na predição foram Redes Neurais Artificiais; árvores de decisão REPTree e M5P; Floresta aleatória (RF); Máquina de vetor suporte (SVM) e ZeroR, sendo este último utilizado como algoritmo controle. Os algoritmos que obtiveram maior acurácia de germinação foram o M5P, SVM e RF para ambos os inputs. A refletância de sementes apresentou maior acurácia e menores valores nas métricas de acurácia MAE e RMSE. A determinação pode ser realizada também com a refletância foliar, que apresentou resultados satisfatórios, indicando que este conjunto de entrada pode ser utilizado como uma estratégia de predição antecipada e a campo da qualidade fisiológica de sementes de soja.

**Palavra-chave:** árvore de decisão, inteligência computacional, germinação de sementes, *Glycine max (L.) Merrill*.

## SOYBEAN SEED GERMINATION PREDICTION USING MACHINE LEARNING.

**Abstract-** Quality soybean seeds have genetic, physical and environmental characteristics that directly affect crop yield. The use of hyperspectral sensors, together with machine learning (ML), can facilitate and optimize the time for germination tests. Therefore, the objectives of this work were: i) to evaluate whether leaf and seed reflectances are effective in predicting the physiological quality of soybean seeds, using ML algorithms and ii) to identify which algorithm provides better accuracy in this prediction. Thirty-two soybean genotypes were evaluated in a spectrophotometric experiment. The leaves and seeds were analyzed with a hyperspectral sensor, an equipment capable of measuring the spectrum in the range of 350 to 2500 nm. The data were subjected to ML analysis with two types of inputs: leaf and seed spectral variables, as well as the variable to be predicted (output) which were germination test (GERM), electrical conductivity (EC), first germination count (PCG), vigorous tetrazolium test (TZV) and viable tetrazolium test (VIAB). This prediction was performed using stratified cross-validation with  $k$ -fold = 10 and ten repetitions (100 runs for each model). All model parameters were set according to the default configuration of the Weka 3.8.5 software. The ML models used in the prediction were Artificial Neural Networks; REPTree and M5P decision trees; Random Forest (RF); Support Vector Machine (SVM) and ZeroR, the latter being used as a control algorithm. The algorithms that obtained the highest germination accuracy were M5P, SVM and RF for both inputs. Seed reflectance showed higher accuracy and lower values in the MAE and RMSE accuracy metrics. The determination can also be performed with leaf reflectance, which presented satisfactory results, indicating that this input set can be used as a strategy for early and field prediction of the physiological quality of soybean seeds.

**Keyword:** decision tree, computational intelligence, seed germination, *Glycine max (L.) Merril.*

## Sumário

<b>1. INTRODUÇÃO.....</b>	<b>8</b>
<b>2. MATERIAL E MÉTODOS .....</b>	<b>10</b>
<b>2.1 Condução do experimento.....</b>	<b>10</b>
<b>2.2 Análises espectrais .....</b>	<b>11</b>
<b>2.3 Variáveis fisiológicas de sementes .....</b>	<b>11</b>
<b>2.4 Aprendizagem de máquina .....</b>	<b>12</b>
<b>3. RESULTADOS .....</b>	<b>14</b>
<b>4. DISCUSSÃO .....</b>	<b>19</b>
<b>5. CONCLUSÃO.....</b>	<b>22</b>
<b>6. REFERÊNCIA BIBLIOGRAFICA .....</b>	<b>23</b>



## 1. INTRODUÇÃO

A produção de semente de qualidade que apresente características desejadas, como atributos genéticos, físicos e fisiológicos, não reflete apenas diretamente sobre o rendimento das culturas, mas também na viabilidade de manter um sistema de exploração viável financeiramente (SOUZA; SMIDERLE, 2024A). Uma das particularidades das características fisiológicas, a qualidade da semente, verifica-se através de testes de germinação e vigor. A germinação é essencial para obter um estabelecimento da cultura, crescimento e desenvolvimento das plantas a campo (MANGENA, 2021). O vigor relaciona-se à capacidade das sementes de gerar plântulas normais de forma rápida e uniforme, mesmo diante de condições desfavoráveis de ambiente e armazenamento (MARCOS FILHO, 2015). Contudo, as análises de qualidade das sementes são realizadas através de testes destrutivos, subjetivos e que demandam tempo e mão de obra. Para facilitar esses testes, tem se implementado o sensoriamento remoto, que diminui erros humanos e não são destrutivos.

Os sensores hiperespectrais são equipamentos capazes de realizar avaliações de biomassa, produtividade e degradação do solo, visto que permitem a obtenção de informações biofísicas e bioquímicas das plantas (SAHOO; RAY; MANJUNATH, 2015). Particularmente nas plantas, as modificações espectrais estão relacionadas às alterações no aparato fotossintético e seu desempenho (PINTO et al., 2020). A utilização de técnicas como a reflectância em dados hiperespectrais permite a captura de informações pelos comprimentos de onda que o sensor detecta (VOLLMANN et al., 2022). Uma das principais qualidades desse método é que ele não é destrutivo e não tem modificação de acordo com a luminosidade do ambiente, o que permite reduzir erros dispostos pela luz difusa. Com isso, a utilização de reflectância para a previsão de germinação de sementes se apresenta um método promissor para adquirir informações acuradas mais rápidas do que os métodos tradicionais.

A combinação dos sensores hiperespectrais com a inteligência computacional, como métodos de aprendizagem de máquina (ML), auxilia a obter resultados em menor tempo. Os conjuntos de dados possuem alta dimensionalidade, e ainda há relações não-lineares entre as variáveis espectrais e as variáveis agronômicas. Isso faz com que a modelagem de predição ou classificação tenha baixa acurácia ou até mesmo seja inviável

utilizando métodos tradicionais, como a regressão. Diante disso, a inteligência computacional é promissora para auxiliar a processar esses dados. Porém, o desempenho de cada algoritmo é distinto com cada conjunto de dados, indicando a necessidade de testar diferentes algoritmos para encontrar o melhor classificador (SANTANA et al., 2021).

Diante disso, a pesquisa teve como objetivos: i) avaliar se as refletâncias das folhas e sementes são eficazes na predição de qualidade fisiológica de sementes de soja, utilizando algoritmos de ML e ii) identificar qual algoritmo proporciona melhor acurácia nessa predição.

## 2. MATERIAL E MÉTODOS

### 2.1 Condução do experimento

O experimento foi conduzido na Universidade Federal de Mato Grosso do Sul, campus Chapadão do Sul (18°41'33''S, 52°40'45''W, com 810 m de altitude), na safra de 2022/23. O clima da região é Tropical de Savana (Aw), segundo a classificação de Köppen e Geiger. O solo em questão possui classificação de Latossolo Vermelho Distrófico Argiloso, com as seguintes características na camada de 0 – 0,20 m: pH (H<sub>2</sub>O) = 6,2; Al trocável (cmolc dm<sup>-3</sup>) = 0,0; Ca+Mg (cmolc dm<sup>-3</sup>) = 4,31; P (mg dm<sup>-3</sup>) = 41,3; K (cmolc dm<sup>-3</sup>) = 0,2; Matéria orgânica (g dm<sup>-3</sup>) = 19,74; V (%) = 45; m (%) = 0,0; Soma de bases (cmolc dm<sup>-3</sup>) = 2,3; CTC (cmolc dm<sup>-3</sup>) = 5,1.

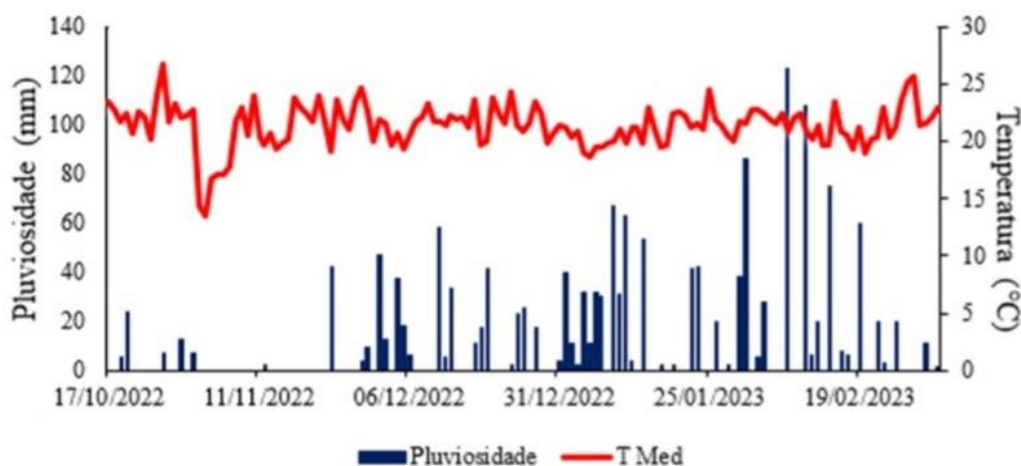


Figura 1. Temperatura média e pluviosidade durante o período experimental.

Na pesquisa, 32 genótipos de soja foram analisados, plantados em outubro de 2022, distribuídos em três repetições nos blocos casualizados. As parcelas foram estruturadas por quatro fileiras de um metro de extensão, com espaçamento de 0,45 m e densidade de 15 sementes por metro.

Antes da semeadura, o solo foi preparado de maneira tradicional, utilizando aração e nivelamento. As sementes de soja foram tratadas com uma mistura de fungicida (Piraclostrobina + Metil Tiofanato) e inseticida (Fipronil), aplicados na proporção de 200 mL de produto a cada 100 kg de sementes. Além disso, foi realizada inoculação com *Bradyrhizobium* no sulco de plantio para promover a fixação biológica de nitrogênio, de acordo com a recomendação do fabricante. Ao crescimento da cultura, foram feitos tratamentos culturais com fungicidas, inseticidas e herbicidas conforme a demanda.

## **2.2 Análises espectrais**

As leituras espectrais de três amostras foliares de cada parcela foram realizadas após 60 dias da emergência (DAE) para o laboratório. Após a colheita, as sementes foram levadas para o laboratório e acondicionadas em placas de Petri para realizar as leituras espectrais. A análise nas folhas e sementes foi realizada com espectrorradiômetro (FIELDSPEC 4 JR., ANALYTICAL SPECTRAL DEVICES, BOULDER, EUA). O sensor é capaz de medir o espectro na faixa de 350 a 2500 nm. Utilizou-se o equipamento ASD Plant Probe para a leitura das folhas, que é destinado a realizar medições espectrais em materiais sólidos. Para a calibração do equipamento, utilizou-se uma placa branca que reflete 100% da luz, composta de sulfato de bário. O equipamento foi calibrado após o encerramento das leituras das amostras de cada bloco, concluindo quatro calibrações.

O sensor manteve-se conectado ao computador para os registros de cada leitura por intermédio de um software próprio do sensor chamado RS3, que registra as leituras executadas pelo equipamento. Seguente a isso, os arquivos são importados para outro software, o ViewSpectroPro, para extração dos dados em extensão.txt, contribuindo para uma análise estatística mais simples de ser executada.

## **2.3 Variáveis fisiológicas de sementes**

Com a maturação dos genótipos, as sementes foram colhidas das quatro linhas laterais de cada parcela, para a execução dos testes para análise de sementes. Amostras de sementes de cada parcela de teste foram colocadas em placas de Petri para coletar dados espectrais e essas mesmas sementes foram também realizados os testes das variáveis fisiológicas.

No início, o método da estufa foi usado para avaliar o teor de água das sementes. Duas subamostras de aproximadamente 4,0 gramas de sementes foram retiradas de cada genótipo e foram colocadas em estufa de circulação de ar forçado durante 24 horas a 105 graus Celsius (mais ou menos 3) (BRASIL, 2009). Os resultados foram registrados em porcentagem de base úmida. Ao adicionar 40 ml de água destilada às caixas de poliestireno gerbox que contém uma grade para não ter o contato direto da água com a semente, foram condicionadas com o objetivo de aumentar o teor de água para 12 a 13%. Durante 24 horas, as caixas gerbox foram mantidas no germinador a 25 °C. Esse

condicionamento foi necessário para garantir que a probabilidade de germinar de cada semente fosse a mesma.

O teste de germinação (GERM) foi realizado utilizando quatro subamostras com 50 sementes de cada genótipo distribuídas sobre papel germitest, umedecido anteriormente em 2,5 vezes o seu peso com água e posteriormente mantidas em germinador à 25 °C (BRASIL,2009). A contagem de germinação foi realizada ao oitavo dia, e os resultados foram expressos em porcentagem de plântulas normais. A primeira contagem de germinação (PCG) foi realizada junto com o teste de germinação. Ao quinto dia do teste, foi realizado a contagem de plântulas normais, sendo os resultados expressos em porcentagem de plântulas normais.

Para o teste de condutividade elétrica (CE) foram pesadas 25 sementes em balança de precisão (0,0001 g) e colocadas para embeber em copos plásticos com 75 mL de água destilada sendo para embeber, mantidas no germinador a 25 °C por 24 horas (VIEIRA; KRZYZANOWSKI, 1999). As leituras foram realizadas com condutivímetro DIGMED DM-31, com valores expressos em  $\mu\text{S cm g}$  de sementes.

O teste de tetrazólio foi preparado em quatro subamostras de 25 sementes que primeiramente foram submetidas a uma pré-embebição em papel germitest umedecido 2,5 vezes seu peso, mantendo-se no germinador a 25 °C por 16 horas. Ao término desse período, as sementes foram embebidas em solução 2,3,5-cloreto de trifeniltetrazólio a 0,075 % ,colocadas em B.O.D. no escuro, à temperatura de 35 °C por quatro horas. Após esse procedimento, as sementes foram avaliadas e classificadas individualmente em vigorosas (TZV) e viáveis (VIAB), de acordo com metodologia proposta por (COSTA ET AL., 1998).

## **2.4 Aprendizagem de máquina**

Após a coleta de dados espectrais e fisiológicos das sementes, os dados foram tabelados e analisados estatisticamente. Os dados foram submetidos à análise de ML testando duas configurações de entrada (inputs): variáveis espectrais de folha (i) e semente (ii). As variáveis preditas (outputs) foram GERM, CE, PCG, TZV e VIAB. Essa predição foi desempenhada por meio de validação cruzada estratificada com  $k\text{-fold} = 10$  e dez repetições (100 execuções para cada modelo). Todos os parâmetros dos modelos foram estabelecidos de acordo com a configuração default do software Weka 3.8.5. Os modelos de aprendizagem de máquina utilizados na predição foram redes neurais artificiais do tipo Perceptron Multicamadas (MP) (EGMONT-PETERSEN et al., 2002);

árvore de decisão REPTree (SNOUSY et al., 2011); Árvore de decisão M5P (BLAIFI et al., 2018); Floresta aleatória, (BELGIU; DRĂGUȚ, 2016); Máquina de vetor suporte (NALEPA; KAWULOK, 2019) e ZeroR, sendo este último um algoritmo de predição padrão do Weka.

Coeficiente de correlação ( $r$ ), erro absoluto médio (MAE) e raiz quadrada do erro médio (RMSE) foram usados para avaliar o desempenho dos modelos de predição testados. Os algoritmos passaram por uma análise de variância para verificar a significância dos efeitos de inputs, modelos de ML e a interação entre ambos. Quando significantes, foram criados boxplots com as médias de  $r$ , MAE e RMSE, que foram agrupados a um nível de probabilidade de 5% usando o teste de Scott-Knott (Scott & Knott, 1974). Os pacotes do software R ExpDes.pt e ggplot2 foram usados para criar boxplots e agrupamentos de médias.

### 3. RESULTADOS

Utilizando o input folha, observaram-se maiores médias de  $(r)$  com M5P (0,426), RF (0,453) e SVM (0,462). As maiores correlações entre valores preditos e observados  $(r)$  para GERM foram alcançadas pelos algoritmos M5P, RF e SVM, com médias entre 0,565 a 0,575, quando a entrada utilizada foi sementes. Analisando as entradas, a que teve maior acurácia em todos os algoritmos foi sementes. (Figura 2).

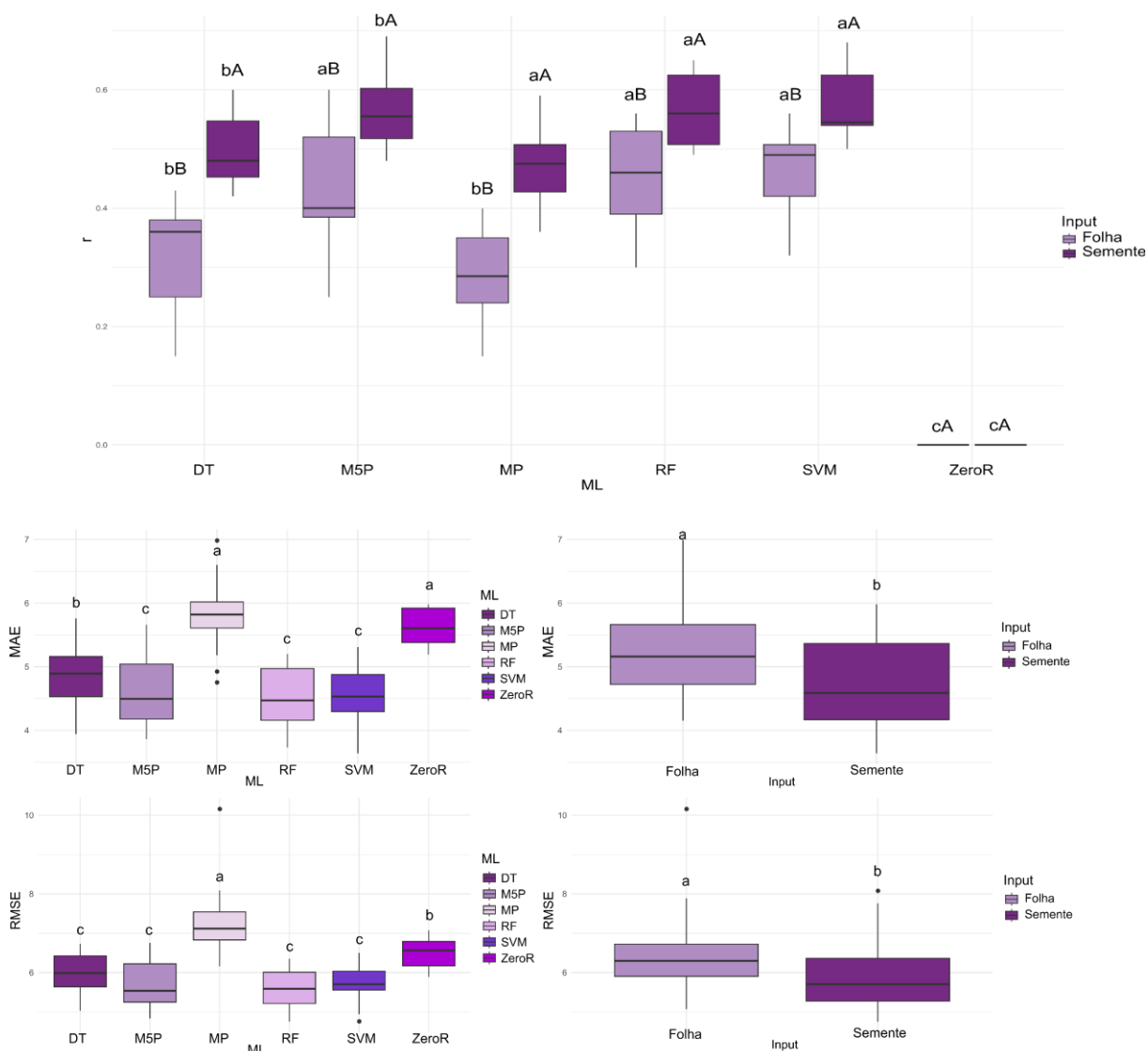


Figura 2. Boxplot para as médias do erro médio absoluto (MAE) e da raiz quadrada do erro médio (RMSE) na predição de germinação (GERM) de sementes de soja utilizando diferentes configurações de entrada e algoritmos de aprendizagem de máquina.

As médias com as mesmas letras maiúsculas para as diferentes entradas e as letras minúsculas iguais para os diferentes algoritmos de ML não diferem pelo teste de Scott-Knott a 5% de probabilidade.

Considerando o MAE, os algoritmos com melhor desempenho foram M5P, RF e SVM e o input que teve menor erro foi utilizando reflectância de semente. Na métrica de acurácia RMSE, os algoritmos que tiveram menor erro foram o DT, M5P, RF e SVM. A entrada com menor erro nessa métrica de acurácia foi utilizando reflectância de sementes.

Nas métricas de acurácia na predição de condutividade elétrica (CE), o algoritmo que apresentou maior acurácia foi M5P, com média de 0,506, e SVM, com 0,492, quando a entrada foi folha. Quando a entrada foi semente, o melhor algoritmo foi SVM com r médio de 0,658, mas para os algoritmos MP, RF, SVM e ZeroR a entrada de melhor desempenho é a semente (Figura 3). O SVM apresentou os menores valores de MAE (0,0145) e RMSE (0,0255), sendo, portanto, o algoritmo mais acurado na predição de CE.

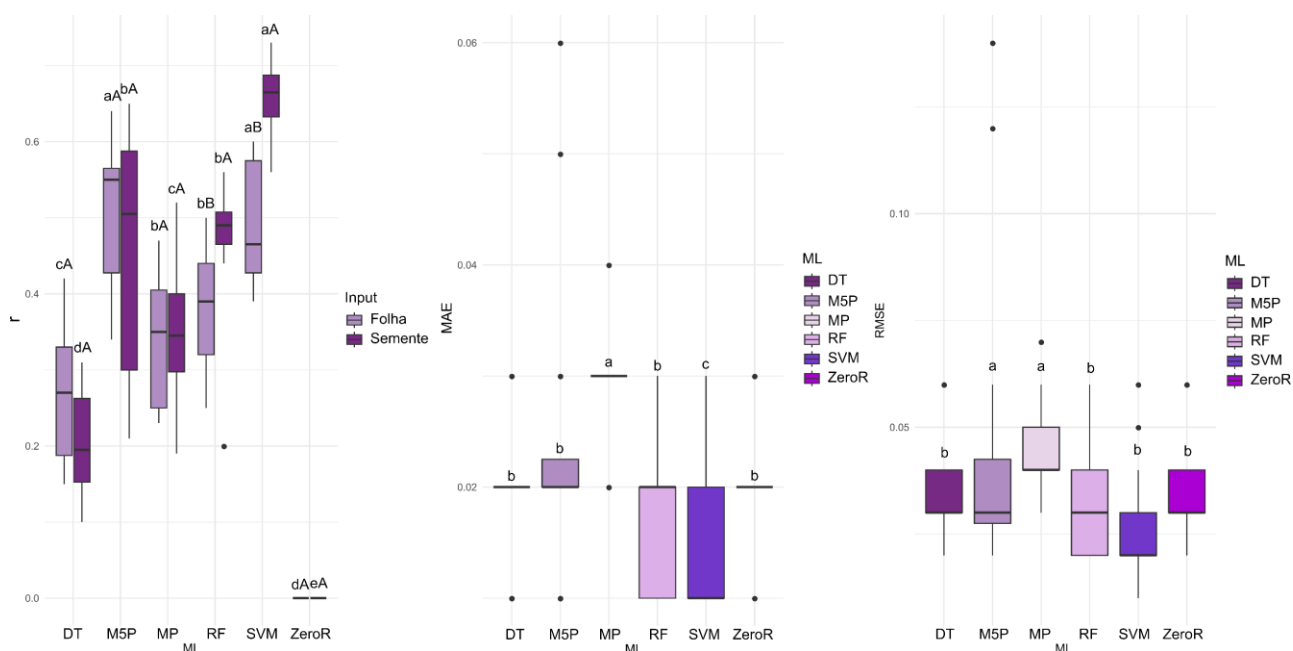


Figura 3. Boxplot para as médias do erro médio absoluto (MAE) e da raiz quadrada do erro médio (RMSE) na predição da condutividade elétrica (CE) de sementes de soja utilizando diferentes configurações de entrada e algoritmos de aprendizagem de máquina. As médias com as mesmas letras maiúsculas para as diferentes entradas e as letras minúsculas iguais para os diferentes algoritmos de ML não diferem pelo teste de Scott-Knott a 5% de probabilidade.



Nas métricas de acurácia na predição de primeira contagem de germinação (PCG), o de maior precisão foi M5P, apresentando média de 0,720 para a entrada folha. Utilizando a entrada semente, o melhor foi M5P, RF e SVM, com médias em torno de 0,735 a 0,777. O input mais assertivo em todos os algoritmos testados foi semente (Figura 4).

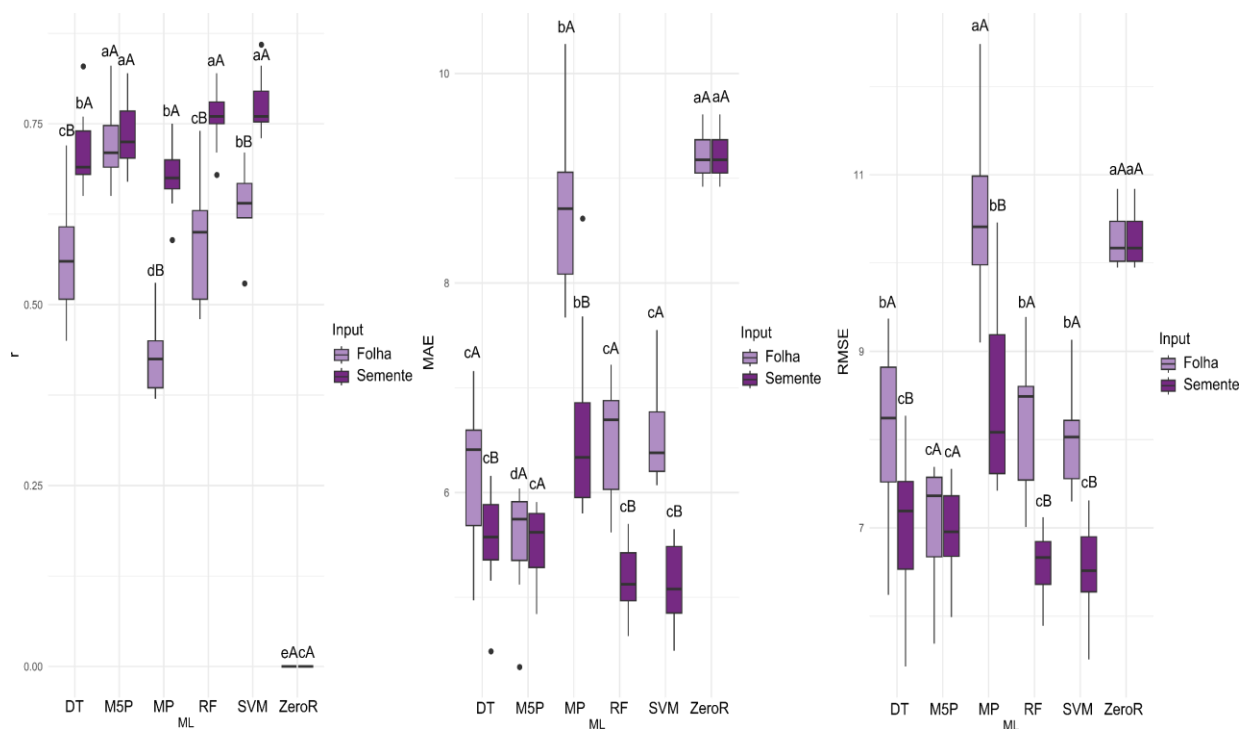


Figura 4. Boxplot para as médias do erro médio absoluto (MAE) e da raiz quadrada do erro médio (RMSE) na predição da primeira contagem de germinação (PCG) de sementes de soja utilizando diferentes configurações de entrada e algoritmos de aprendizagem de máquina

As médias com as mesmas letras maiúsculas para as diferentes entradas e as letras minúsculas iguais para os diferentes algoritmos de ML não diferem pelo teste de Scott-Knott a 5% de probabilidade.

Para o input folha, o algoritmo M5P proporcionou os menores valores de MAE. Quando o input foi semente, os menores valores de MAE foram alcançados pelos algoritmos M5P DT, M5P, RF e SVM. Para o algoritmo ZeroR, as médias dos inputs desse fator são estatisticamente iguais, diferentemente dos outros níveis testados, onde a entrada semente foi a que apresentou menor erro. Quando utilizando a refletância foliar como conjunto de entrada, o M5P resultou em menor RMSE. Já quando as variáveis de

entrada foram as reflectâncias da semente, os algoritmos DT, M5P, RF e SVM proporcionaram os menores valores de RMSE.

Na predição de tetrazólio vigor (TZVG), o de melhor desempenho foi o algoritmo RF com aproximadamente 0,25. Considerando as variáveis, a com maior acurácia foi semente (Figura 5). O M5P apresentou os menores valores na métrica de acurácia MAE. Para as variáveis de entrada refletância foliar, os menores valores de RMSE foram com o algoritmo ZeroR. Quando as variáveis de entrada foram refletância de semente, os menores valores na RMSE foram com o algoritmo RF (8.355), já para os algoritmos ZeroR, RF e SVM, o input semente teve menor erro, diferentemente do MP que o input folha obteve menor erro.

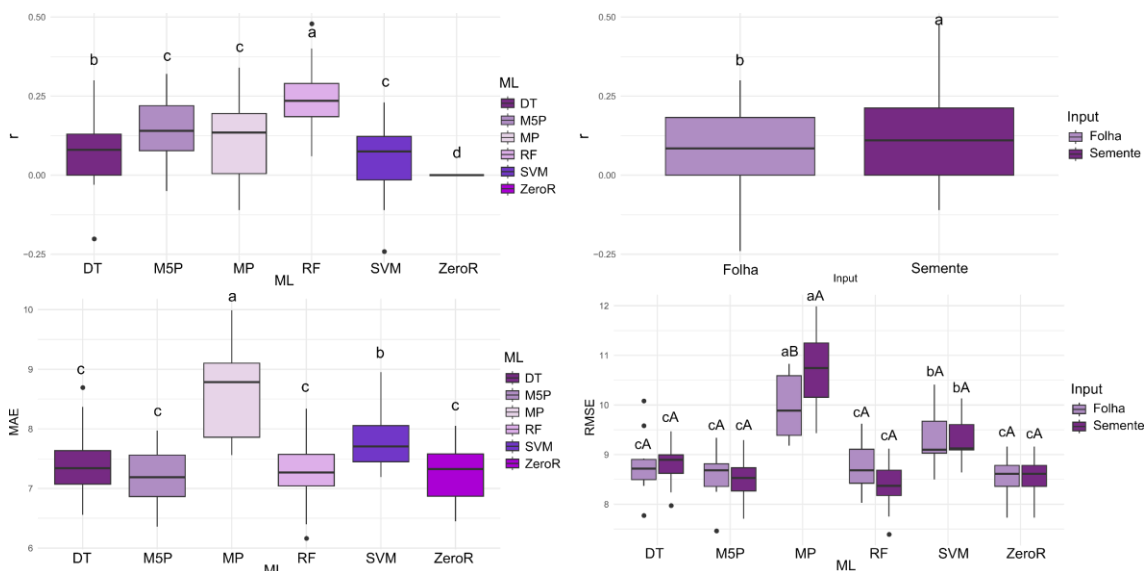


Figura 5. Boxplot para as médias do erro médio absoluto (MAE) e da raiz quadrada do erro médio (RMSE) na predição de tetrazólio vigor (TZVG) de semente de soja utilizando diferentes configurações de entrada e algoritmos de aprendizagem de máquina. As médias com as mesmas letras maiúsculas para as diferentes entradas e as letras minúsculas iguais para os diferentes algoritmos de ML não diferem pelo teste de Scott-Knott a 5% de probabilidade.

As maiores correlações entre valores preditos e observados ( $r$ ) para VIAB foram alcançadas pelos algoritmos SVM, MP, RF e M5P. Considerando o input semente, os melhores desempenhos foram obtidos pelos algoritmos M5P, RF e SVM, cujas médias de  $r$  foram acima de 0,5. O input com maior acurácia foi semente para todos os algoritmos

analisados, com exceção de ZeroR, que apresentou performance estatisticamente igual para as duas entradas (Figura 6). Considerando o MAE, os algoritmos com melhor desempenho foram M5P, RF e SVM e o input que teve menor erro foi utilizando reflectância de semente. Para a métrica de acurácia RMSE considerando como entrada a reflectância foliar, os algoritmos DT, M5P, ZeroR tiveram os menores valores. Com a reflectância de semente, os valores com menores valores foram M5P, RF e SVM. Similarmente às outras métricas, as menores médias de erro foram alcançadas utilizando o input sementes

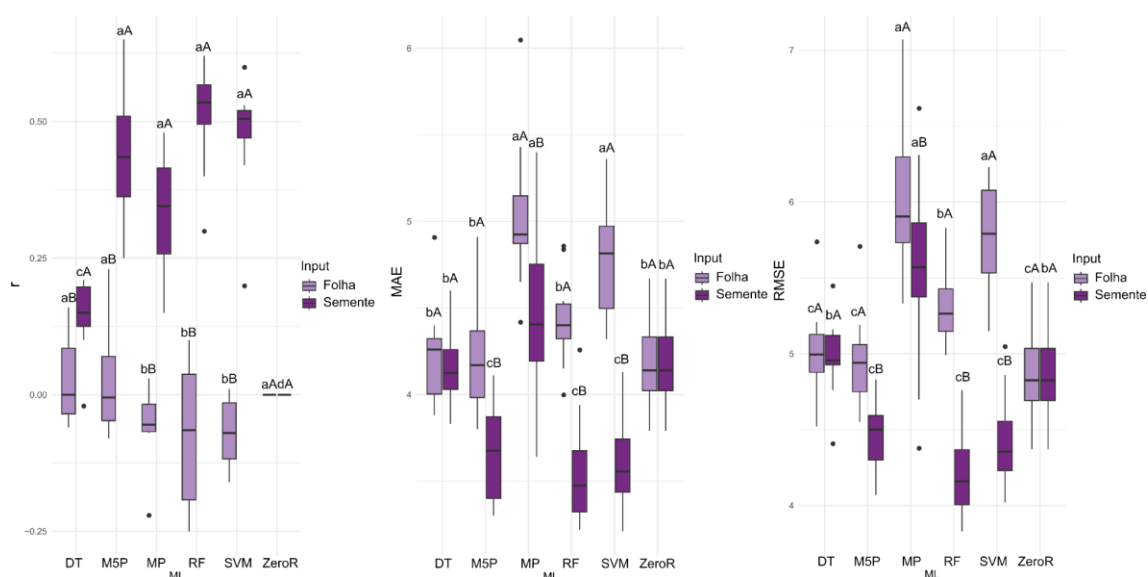


Figura 6. Boxplot para as médias do erro médio absoluto (MAE) e da raiz quadrada do erro médio (RMSE) na predição de tetrazólio viabilidade (VIAB) de semente de soja utilizando diferentes configurações de entrada e algoritmos de aprendizagem de máquina. As médias com as mesmas letras maiúsculas para as diferentes entradas e as letras minúsculas iguais para os diferentes algoritmos de ML não diferem pelo teste de Scott-Knott a 5% de probabilidade.

Observou-se que dos algoritmos analisados, o com maior assertividade foi a Árvore de Decisão M5P, que também se destacou apresentando menores valores nas métricas de acurácia MAE e RMSE. O input que apresentou menores valores e consequentemente maior acurácia foi semente.

#### 4. DISCUSSÃO

A produção de sementes tem como objetivo final estabelecer um campo com uma população adequada de plantas, que possuem potencial genético, fisiológico, como a germinação, e sanitário para maximizar o rendimento e evitar a propagação de pragas, patógenos e plantas invasoras (SOUZA; SMIDERLE, 2024). A capacidade de prever a germinação é essencial para promover práticas agrícolas sustentáveis (SANDHIYA, et al. (2024). Permite aos agricultores planejar melhor o plantio, otimizar o uso de recursos como água e nutrientes, evitar replantio, evitar deterioração da semente, e assim proporcionar um estande saudável e mais resistente a condições adversas. Experimentos de germinação em larga escala são desafiadores e sujeitos a erros humanos, o que ressalta a necessidade de métodos automatizados como o ML aprendizagem de máquina (COLMER et al., 2020).

O sensoriamento remoto conta como ferramenta os sensores hiperespectrais, que possibilitam a obtenção de centenas de bandas espectrais, reproduzindo os espectros de reflectância de alvos, possibilitando sua identificação (KUMAR et al., 2002). Os materiais apresentam assinaturas espectrais, cada parte vegetal estudada tem formação de curvas características, uma vez que o conteúdo de clorofila e compostos orgânicos podem sofrer alterações em diferentes partes da planta (PRABHAKAR; PRASAD; RAO, 2012). Essa ferramenta tem facilitado os testes de germinação, apresentando um método não destrutivo e eficiente para essa identificação sem ter os riscos de erros humanos.

O algoritmo M5P une uma árvore de decisão tradicional com a capacidade de realizar regressão linear nos nós, estimando assim o valor da classe das instâncias que chegam até a folha (DANG et al. SINGH, 2021). Em Kisi et al. (2021), a árvore de decisão M5P foi utilizada para a modelagem da evapotranspiração de referência, onde foi considerada melhorada. Para Maabreh et al., 2023, o algoritmo é reconhecido como uma abordagem sólida e eficaz de séries temporais que permitiu a possibilidade de prever, por análises dos solos, que a pastagem perturbada tem maior quantidade de erosão entre sulcos. Rivandi et al. (2022) mostraram através dos seus estudos que o algoritmo apresenta mais precisão devido à velocidade e menor custo do que o modelo de regressão para a previsão (90%) dos pesos de estigma seco e da flor de açafraão para a comercialização.

A RF é a alteração das árvores, é uma técnica de aprendizado em conjunto que cria uma floresta composta por várias árvores de decisão geradas a partir de amostras

bootstrap de um conjunto de dados de treinamento (DANG ET AL, 2021) . Observado pelos autores Shahare et al. (2024), obteve-se acurácia de 93% no treinamento e 83% no teste ao prever os níveis de fertilidade dos solos, mostrando uma maneira confiável de melhorar a produtividade das plantações. Na pesquisa dos autores Kumar et al., 2024 concluíram que o algoritmo SVM obteve os melhores resultados (98,5 %) para a classificação da demanda da água e economia de água doce que pode ser feita por utilizar a irrigação inteligente.

O input que proporcionou maior acurácia para os algoritmos testados e menores valores de MAE e RMSE foi utilizando as reflectâncias das sementes. As sementes apresentam características de qualidade genética, física, fisiológica e sanitária, proporcionando assim a certeza de um alto desempenho agrônomo. O potencial fisiológico é avaliado através da viabilidade e vigor, que são indicadores que medem a habilidade da semente de germinar de maneira rápida e uniforme (WEN et al., 2018). A germinação é o termo utilizado para analisar a viabilidade das sementes, observando a emergência da raiz primária e o desenvolvimento de uma plântula saudável, com raízes e parte aérea adequadamente formadas (ZHANG et al., 2023).

A maior acurácia proporcionada por este input pode ser explicada pela estreita relação entre as reservas contidas na semente e a sua capacidade de germinação (MOLES; WESTOBY, 2004). Características ligam significativamente os carotenoides e clorofilas com a qualidade da semente, mesmo que influenciada por fatores genéticos (GEBREGZIABHER et al., 2022). Quando em estresse, a resposta de defesa é a produção de carotenoides e clorofilas (XIE et al., 2019). Conseqüentemente, as sementes que refletiram mais tiveram maior estresse, relatando menor qualidade fisiológica, indicando menor germinação, uma vez que essas características estão relacionadas. Logo, as que refletiram menos são as de maior qualidade fisiológica por não terem em grandes quantidades a clorofila e carotenoides, conseguindo assim expressar a viabilidade e vigor, fatores que aumentam a taxa de germinação.

A viabilidade das sementes é caracterizada pela sua capacidade de germinar quando satisfaz todas as necessidades de umidade e temperatura. Por outro lado, o vigor refere-se à aptidão da semente de germinar e produzir plântulas saudáveis de maneira rápida, mesmo em condições ambientais desfavoráveis (XIA et al., 2019). Diante dessas informações, a acurácia do input semente se apresenta maior porque os atributos essenciais para a germinação estão inseridos na sua formação.

Pesquisas realizada pela autora Santana, 2022 conseguiu classificar as características de produção da soja através da refletância das folhas. Oliveira, 2023 classificou a qualidade fisiológica das sementes a partir da refletância das folhas. Porém esse trabalho conseguiu implementar uma informação importante e inovadora para os produtores de sementes, evidenciando que a acurácia de predição das variáveis fisiológicas utilizando folha como entrada apresentaram resultados satisfatórios para todas as variáveis fisiológicas, com exceção de tetrazólio viabilidade (VIAB), quando utilizando o algoritmo M5P. Tais resultados mostram que utilizar folha como input é uma estratégia promissora para uma predição antecipada da qualidade fisiológica das sementes de soja.

Predizer a qualidade fisiológica da semente por meio da refletância foliar é uma estratégia muito interessante porque conseguimos obter informações sobre germinação, vigor antes mesmo da colheita. Futuros trabalhos inclusive podem utilizar sensores acoplados em veículo aéreo não tripulado com os Drones para obter essa informação foliar antes do final do ciclo da cultura. A variável de entrada refletância foliar pode proporcionar uma acurácia adequada quando se utiliza o algoritmo M5P, que apresentou maiores valores de  $r$  e menores valores de MAE e RMSE para predição de Teste de germinação e Primeira contagem de germinação. As folhas têm um papel fundamental na produção e transporte de fotoassimilados para as sementes (ZANDALINAS et al, 2018). A planta-mãe deixa de enviar os fotoassimilados quando a semente se encontra em maturidade fisiológica, o embrião está completamente formado pelo processo de embriogênese, neste momento também a viabilidade e vigor estão em alta capacidade (MONTIEL-SÁNCHEZ et al., 2021). Então, uma folha saudável é capaz de proporcionar uma planta com sementes bem desenvolvidas e que garantam a propagação da espécie.

As perspectivas para futuros estudos são que experimentos sejam realizados em um maior número de safras para obter maiores conjuntos de dados, assim podendo encontrar modelos acurados, a fim de melhorar a eficiência dos testes de germinação, diminuindo tempo e mão de obra.

## 5. CONCLUSÃO

O input semente proporcionou maior correlação entre valores observados e preditos e menores valores de erro. O algoritmo que obteve maior acurácia de predição de germinação foi a Árvore de decisão M5P para ambas as entradas. A acurácia de predição das variáveis fisiológicas utilizando folha como entrada apresentaram resultados satisfatórios para todas as variáveis fisiológicas, com exceção de tetrazólio viabilidade (VIAB), quando utilizando o algoritmo M5P. Tais resultados mostram que utilizar folha como input é uma estratégia promissora para uma predição antecipada da qualidade fisiológica.

## 6. REFERÊNCIAS BIBLIOGRAFICA

- BELGIU, M.; DRĂGUȚ, L. Random Forest in remote sensing: A review of applications and future directions. *ISPRS Journal of Photogrammetry and Remote Sensing*, v. 114, p. 24–31, 2016.
- BLAIFI, S.; Moulahoum S.; Benkercha R.; Taghezouit B.; Saim A.; M5P model tree based fast fuzzy maximum power point tracker. *Solar Energy*, v. 163, p. 405–424, 2018.
- BRASIL. MINISTÉRIO DA AGRICULTURA, P. e Abastecimento. S. de D. A. Regras para análise de sementes. [s.l.: s.n.].
- COLMER, J.; O'NEILL C.; WELLS, R.; BOSTROM, A.; REYNOLDS D.; WEBSDALE, D.; SHIRALAGI, G.; LU, W.; LOU, Q.; LE CORNU, T.; BALL, J.; RENEMA, J.; FLORES, A. G.; BENJAMINS, R.; PENFIELD, S.; ZHOU, J. SeedGerm: a cost-effective phenotyping platform for automated seed imaging and machine-learning based phenotypic analysis of crop seed germination. *New Phytologist*, v. 228, n. 2, p. 778–793, 1 out. 2020.
- COSTA, N. P.; FRANÇA NETO, J.B.; KRZYŻANOWSKI, F.C.; HENNING, A.A.; PEREIRA, J.E.; Valiação de metodologia alternativa para o teste de tetrazólio para sementes de soja. *Scientia Agricola*, v. 55, 1998.
- DANG, S. K.; SINGH, K. Predicting tensile-shear strength of nugget using M5P model tree and random forest: An analysis. *Computers in Industry*, v. 124, p. 103345, 2021.
- EGMONT-PETERSEN, M.; DE RIDDER, D.; HANDELS, H. Image processing with neural networks—a review. *Pattern Recognition*, v. 35, n. 10, p. 2279–2301, 2002.
- GEBREGZIABHER, B. S.; ZHANG, S.; AZAM, M.; QI, J.; AGYENIM-BOATENG, K. G.; FENG, Y.; LIU, Y.; LI, J.; LI, B.; SUN, J. Natural variation and geographical distribution of seed carotenoids and chlorophylls in 1167 Chinese soybean accessions. *Journal of Integrative Agriculture*, [s. l.], 2022.
- KISI, O. KESHTEGAR B.; ZOUNEMAT-K.M.; HEDDAM S.; TRUNG N.; modeling reference evapotranspiration using a novel regression-based method: radial basis M5 model tree. *Theoretical and Applied Climatology*, v. 145, n. 1, p. 639–659, 2021.
- KUMAR, G. K. Internet of things sensors and support vector machine integrated intelligent irrigation system for agriculture industry. *Discover Sustainability*, v. 5, n. 1, p. 6, 11 jan. 2024.



KUMAR, L.; SCHMIDT, K.; DURY, S.; SKIDMORE, A. Imaging spectrometry and vegetation science. *Imaging spectrometry: basic principles and prospective applications*, [s. l.], p. 111–155, 2001.

MAABREH, H. G. WAHEEB K.; RYADH A.; ABDULGHANIS. B.; HAMOODAH Z.J., JASIM N.Y.; ALAJEELI F.; AL MANSOR A.H.O.; ANDREEVICH M.; Application of M5 algorithm of decision tree in simulation and investigation of effective factors of erosion in rangelands and forests. *Caspian Journal of Environmental Sciences*, v. 21, n. 3, p. 533–541, 2023.

MANGENA, P. Analysis of correlation between seed vigour, germination and multiple shoot induction in soybean (*Glycine max L. Merr.*). *Heliyon*, v. 7, n. 9, p. e07913, set. 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, ago. 2015.

MOLES, A. T.; WESTOBY, M. Seedling survival and seed size: a synthesis of the literature. *Journal of Ecology*, v. 92, n. 3, p. 372–383, 1 jun. 2004.

MONTIEL, S, M.; GARCÍA, C.T.; GÓMEZ, M.A.; GARCÍA, H.S.; CANO, M. P.; In vitro gastrointestinal stability, bioaccessibility and potential biological activities of betalains and phenolic compounds in cactus berry fruits (*Myrtillocactus geometrizans*). *Food Chemistry*, v. 342, p. 128087, 2021.

NALEPA, J.; KAWULOK, M. Selecting training sets for support vector machines: a review. *Artificial Intelligence Review*, v. 52, n. 2, p. 857–900, 2019.

OLIVEIRA, C.I. Novas ferramentas para avaliar a qualidade fisiológica de sementes de soja. Tese (Doutorado em Genética, Melhoramento Vegetal e propagação de plantas.), Faculdade de Engenharia, Universidade Estadual Paulista “Júlio de Mesquita Filho”, Ilha Solteira, 2023.

PINTO, J.; POWELL, S.; PETERSON, R.; ROSALEN, D.; FERNANDES, O. Detection of defoliation injury in peanut with hyperspectral proximal remote Sensing. *Remote Sensing*, [s. l.], v. 12, n. 22, p. 3828, 2020.

PRABHAKAR, M.; PRASAD, Y. G.; RAO, M. N. Remote Sensing of Biotic Stress in Crop Plants and Its Applications for Pest Management. Em: VENKATESWARLU, B. et al. (Eds.). *Crop Stress and its Management: Perspectives and Strategies*. Dordrecht: Springer Netherlands, 2012. p. 517–545.

- RIVANDI&IRM, M.; GHASEMNEZHAD, A.; GHORBANIK.; HEMMATI&LRMK.; ABHARI &LRM, A. Prediction of Saffron Flower and Stigma Yield Based on the Physical and Chemical Properties of Water and Soil Using Linear Multivariate Regression Models and M5 Decision Tree. *Journal of Saffron Research*, v. 9, n. 2, p. 352–367, 2022.
- SAHOO, R. N.; RAY, S. S.; MANJUNATH, K. R. Hyperspectral remote sensing of agriculture. *Current Science*, v. 108, n. 5, p. 848–859, 2015.
- SANDHIYA, M.; VISVÉSH, B.; UGENDRABABU, M.; TINISHA, A. Varietal Seed Classification and Seed Germination Prediction System. 2024 Second International Conference on Emerging Trends in Information Technology and Engineering (ICETITEIEEE), 2024. p. 1-5.
- SANTANA, D. C.; COTRIM, M. F.; FLORES, M. S.; ROJO BAIO, F. H.; SHIRATSUCHI, L. S.; SILVA JUNIOR, C. A. Da; TEODORO, L. P. R.; TEODORO, P. E. UAV-based multispectral sensor to measure variations in corn as a function of nitrogen topdressing. *Remote Sensing Applications: Society and Environment*, [s. l.], v. 23, 2021
- SANTANA, D. C. Fenotipagem de alta precisão de caracteres industriais de soja usando aprendizagem de máquina e diferentes sensores espectrais. 2023. Tese (Doutorado em Produção vegetal), Faculdade de Engenharia, Universidade Estadual Paulista “Júlio de Mesquita Filho”, Ilha Solteira, 2023.
- SHAHARE, Y. R.; SINGH, M. P.; SINGH, S. P.; SINGH, P.; DIWAKAR, M. ASUR: Agriculture Soil Fertility Assessment Using Random Forest Classifier and Regressor. *Procedia Computer Science*, v. 235, p. 1732–1741, 2024.
- SNOUSY, M.; EL-DEEB, H.; BADRAN, K. KHIL, I. Suite of decision tree-based classification algorithms on cancer gene expression data. *Egyptian Informatics Journal*, v. 12, n. 2, p. 73–82, 2011.
- SOUZA, A. G.; SMIDERLE, O. J. DETERIORAÇÃO E CARACTERÍSTICAS DE QUALIDADE E VIGOR DE SEMENTES. Em: *Ciências Agrárias: tecnologia, sustentabilidade e inovação - Volume 1*. [s.l.] Editora Científica Digital, 2024a. p. 128–143.
- SOUZA, A. G.; SMIDERLE, O. J. DETERIORAÇÃO E CARACTERÍSTICAS DE QUALIDADE E VIGOR DE SEMENTES. Em: *Ciências Agrárias: tecnologia, sustentabilidade e inovação - Volume 1*. [s.l.] Editora Científica Digital, 2024b. p. 128–143.

VOLLMANN, J.; RISCHBECK, P.; PACHNER, M.; ĐORĐEVIĆ, V.; MANSCHADI, A. M. High-throughput screening of soybean di-nitrogen fixation and seed nitrogen content using spectral sensing. *Computers and Electronics in Agriculture*, [s. l.], v. 199, p. 107169, 2022.

WEN, D.; HOU, H.; MENG, A.; MENG, J.; XIE, L.; ZHANG, C. Rapid evaluation of seed vigor by the absolute content of protein in seed within the same crop. *Scientific Reports*, [s. l.], v. 8, n. 1, 2018.

XIA, Y.; XU, Y.; LI, J.; ZHANG, C.; FAN, S. Recent advances in emerging techniques for non-destructive detection of seed viability: A review. *Artificial Intelligence in Agriculture*, [s. l.], v. 1, p. 35–47, 2019.

XIE, X.; HE, Z.; CHEN, N.; TANG, Z.; WANG, Q.; CAI, Y. The roles of environmental factors in regulation of oxidative stress in plant. *BioMed research international*, [s. l.], v. 2019, 2019.

ZANDALINAS, S. I.; MITTLER, R.; BALFAGÓN, D.; ARBONA, V.; GÓMEZ-CADENAS, A. *Plant adaptations to the combination of drought and high temperatures*, Blackwell Publishing Ltd, 2018.

ZHANG, T.; LU, L.; YANG, N.; FISK, I. D.; WEI, W.; WANG, L.; LI, J.; SUN, Q.; ZENG, R. Integration of hyperspectral imaging, non-targeted metabolomics and machine learning for vigour prediction of naturally and accelerated aged sweetcorn seeds. *Food Control*, [s. l.], p. 109930, 2023.