

UNIVERSIDADE FEDERAL DE MATO GROSSO DO SUL
CÂMPUS DE CHAPADÃO DO SUL
PROGRAMA DE PÓS-GRADUAÇÃO EM AGRONOMIA

LAILA RODRIGUES CIRQUEIRA

**PREDIÇÃO DE DANOS MECÂNICOS, VIGOR E VIABILIDADE EM
SEMENTES DE SOJA ARMAZENADA UTILIZANDO MODELOS DE
APRENDIZADO DE MÁQUINA**

CHAPADÃO DO SUL – MS

2022

UNIVERSIDADE FEDERAL DE MATO GROSSO DO SUL
CÂMPUS DE CHAPADÃO DO SUL
PROGRAMA DE PÓS-GRADUAÇÃO EM AGRONOMIA

LAILA RODRIGUES CIRQUEIRA

**PREDIÇÃO DE DANOS MECÂNICOS, VIGOR E VIABILIDADE EM
SEMENTES DE SOJA ARMAZENADA UTILIZANDO MODELOS DE
APRENDIZADO DE MÁQUINA**

Orientador: Prof. Dr. Paulo Carteri Coradi

Co-Orientadora: Profa. Dra. Larissa Pereira Ribeiro Teodoro

Dissertação apresentada à Universidade Federal de Mato Grosso do Sul, como requisito para obtenção do título de Mestre em Agronomia, área de concentração: Produção Vegetal.

CHAPADÃO DO SUL – MS

2022



PROGRAMA DE PÓS-GRADUAÇÃO EM AGRONOMIA

CERTIFICADO DE APROVAÇÃO

DISCENTE: Laila Rodrigues Cirqueira

ORIENTADOR: Dr. Paulo Carteri Coradi

TÍTULO: Predição de danos mecânicos em sementes de soja armazenadas em diferentes ambientes utilizando Aprendizado de Máquina

AVALIADORES:

Prof. Dr. Paulo Carteri Coradi

Prof. Dr. Paulo Eduardo Teodoro

Prof. Dr. Eduardo Pradi Vendruscolo

Chapadão do Sul, 08 de novembro de 2022.



Documento assinado eletronicamente por **Paulo Eduardo Teodoro, Professor do Magisterio Superior**, em 08/11/2022, às 14:06, conforme horário oficial de Mato Grosso do Sul, com fundamento no art. 6º, § 1º, do [Decreto nº 8.539, de 8 de outubro de 2015](#).



Documento assinado eletronicamente por **Eduardo Pradi Vendruscolo, Usuário Externo**, em 08/11/2022, às 14:11, conforme horário oficial de Mato Grosso do Sul, com fundamento no art. 6º, § 1º, do [Decreto nº 8.539, de 8 de outubro de 2015](#).



Documento assinado eletronicamente por **Paulo Carteri Coradi, Usuário Externo**, em 08/11/2022, às 14:20, conforme horário oficial de Mato Grosso do Sul, com fundamento no art. 6º, § 1º, do [Decreto nº 8.539, de 8 de outubro de 2015](#).

AGRADECIMENTOS

Ao Pai Celestial por me guiar e dar forças para alcançar a conclusão deste trabalho.

Aos meus pais, minha irmã (*in memoriam*), meu marido e todos familiares que acompanharam todo esforço e dedicação para a conclusão desta dissertação.

Ao meu Orientador, Prof. Dr. Paulo Carteri Coradi, por sua dedicação, paciência e disponibilidade durante todo o percurso deste curso.

Aos docentes do Programa de Pós-Graduação em Agronomia, que contribuíram com ensinamentos e atenção sempre que requisitados, em especial à professora e co-orientadora Dra. Larissa Pereira Ribeiro Teodoro pela colaboração e auxílio durante o curso.

À Universidade Federal de Mato Grosso do Sul e ao Programa de Pós-Graduação em Agronomia pela oportunidade de ser aluna do programa de mestrado, agregando conhecimento e experiência profissional.

LISTA DE FIGURAS

Figura 1. Esquema experimental.....	12
Figura 2. Fluxograma e análises experimentais.	15
Figura 3. Boxplots da avaliação de dano mecânico (DM1) para o coeficiente de correlação de Pearson (r) e o erro médio absoluto (MAE) para diferentes modelos de aprendizado de máquina (AM). Regressão linear múltipla (RL), redes neurais artificiais (RNA), algoritmo M5 de Quinlan (M5P) e floresta aleatória (FA).	16
Figura 4. Boxplots da avaliação de dano mecânico (DM2) para o coeficiente de correlação de Pearson (r) e o erro médio absoluto (MAE) para diferentes modelos de aprendizado de máquina (AM). Regressão linear múltipla (RL), redes neurais artificiais (RNA), algoritmo M5 de Quinlan (M5P) e floresta aleatória (FA).	17
Figura 5. Boxplots da avaliação de umidade 1 (U1) para o coeficiente de correlação de Pearson (r) e o erro médio absoluto (MAE) para diferentes modelos de aprendizado de máquina (AM). Regressão linear múltipla (RL), redes neurais artificiais (RNA), algoritmo M5 de Quinlan (M5P) e floresta aleatória (FA).	18
Figura 6. Boxplots da avaliação de umidade 2 (U2) para o coeficiente de correlação de Pearson (r) e o erro médio absoluto (MAE) para diferentes modelos de aprendizado de máquina (AM). Regressão linear múltipla (RL), redes neurais artificiais (RNA), algoritmo M5 de Quinlan (M5P) e floresta aleatória (FA).	19
Figura 7. Boxplots da avaliação vigor (VG) para o coeficiente de correlação de Pearson (r) e o erro médio absoluto (MAE) para os diferentes modelos de aprendizado de máquina (AM). Regressão linear múltipla (RL), redes neurais artificiais (RNA), algoritmo M5 de Quinlan (M5P) e floresta aleatória (FA).	20
Figura 8. Boxplots da avaliação de viabilidade (VB) para o coeficiente de correlação de Pearson (r) e o erro médio absoluto (MAE) para os diferentes modelos de aprendizado de máquina (AM). Regressão linear múltipla (RL), redes neurais artificiais (RNA), algoritmo M5 de Quinlan (M5P) e floresta aleatória (FA).	21

LISTA DE TABELAS

Tabela 1. Níveis de viabilidade das sementes de soja em função dos diferentes danos mecânicos e umidades.....	13
--	----

PREDIÇÃO DE DANOS MECÂNICOS, VIGOR E VIABILIDADE EM SEMENTES DE SOJA ARMAZENADA UTILIZANDO MODELOS DE APRENDIZADO DE MÁQUINA

RESUMO: A inteligência artificial tem sido amplamente aplicada na predição de dados para melhor tomada de decisão e otimização dos processos. Na pós-colheita, o controle de fatores bióticos e abióticos é fundamental para a conservação da qualidade das sementes na etapa de armazenagem. O teste de tetrazólio tem sido utilizado para avaliação da qualidade de sementes, porém, com diversas limitações que podem levar a erros de avaliação. Assim, o objetivo deste estudo foi identificar o melhor modelo de aprendizado de máquina para predição de danos mecânicos, vigor e viabilidade de sementes de soja durante o armazenamento, em função de diferentes condições (10, 15 e 25 °C), embalagens (com revestimento e sem revestimento) e tempos de armazenamento (0, 3, 6, 9 e 12 meses). Os modelos algoritmo M5 de Quinlan (M5P) e floresta aleatória (FA) apresentaram os melhores desempenhos para predição do vigor e viabilidade das sementes, enquanto que o modelo de redes neurais artificiais melhor predisse os resultados de danos mecânicos nas sementes. O coeficiente de correlação de Pearson (r) e o erro médio absoluto (MAE) foram de baixa magnitude na predição de danos mecânicos e umidades das sementes. Na predição do vigor e da viabilidade das sementes de soja, verificaram-se maiores resultados para r e menores para a MAE. Os modelos de algoritmo M5 de Quinlan (M5P) e floresta aleatória (FA) são aqueles que melhor preveem os resultados de qualidade de sementes de soja, com uma análise mais simplificada e ágil para determinação do vigor e viabilidade das sementes de soja no armazenamento.

Palavras-chaves: Embalagens de armazenamento. Inteligência Artificial. Pós-colheita. Temperatura de armazenamento. Tempo de armazenamento. Teste de tetrazólio.

PREDICTION OF MECHANICAL DAMAGE, VIGOR AND VIABILITY IN STORED SOY SEEDS USING MACHINE LEARNING MODELS

ABSTRACT: Artificial intelligence has been widely applied in data prediction for better decision making and process optimization. In the post-harvest, the control of biotic and abiotic factors is fundamental for the conservation of seed quality in the storage stage. The tetrazolium test has been used to evaluate seed quality, however, with several limitations that can lead to evaluation errors. Thus, the aim of this study was to identify the best machine learning model for predicting mechanical damage, vigor and viability of soybean seeds during storage, depending on different conditions (10, 15 and 25 °C), packaging (with coating and uncoated) and storage times (0, 3, 6, 9 and 12 months). The Quinlan algorithm M5 (M5P) and random forest (FA) models showed the best performance for predicting seed vigor and viability, while the artificial neural network model best predicted the results of mechanical damage to seeds. The Pearson correlation coefficient (r) and mean absolute error (MAE) were of low magnitude in predicting mechanical damage and seed moisture. In predicting the vigor and viability of soybean seeds, higher results were found for r and lower results for MAE. The Quinlan algorithm M5 (M5P) and random forest (FA) models are those that best predict soybean seed quality results, with a more simplified and agile analysis for determining the vigor and viability of soybean seeds in storage.

Keywords: Storage packaging. Artificial Intelligence. Post-harvest. Storage temperature. Storage time. Tetrazolium test.

SUMÁRIO

INTRODUÇÃO	10
MATERIAL E MÉTODOS	11
Descrição e delineamento experimental.....	11
Análises de qualidade de sementes de soja	13
Modelos de Aprendizado de Máquina (AM)	14
Análises estatísticas.....	15
RESULTADOS	16
DISCUSSÃO.....	21
CONCLUSÕES.....	22
REFERÊNCIAS BIBLIOGRÁFICAS	23

INTRODUÇÃO

A qualidade das sementes pode sofrer alterações em função do aumento da sua atividade metabólica e respiração, proporcionada pelas condições do ambiente de armazenamento, tecnologias de embalagens e tempo de armazenamento. De acordo com Oliveira et al. (2021), os danos físicos provocados em sementes podem indicar redução da viabilidade, vigor e até a morte da semente (Rocha et al., 2017).

O teste de tetrazólio tem sido uma análise importante e eficiente para avaliação de qualidade de sementes, como índices de vigor e viabilidade (Oliveira et al., 2021). Porém, o teste de tetrazólio tem apresentado algumas limitações, como a necessidade de treinamento e conhecimento avançados sobre a ciência e tecnologia das sementes para interpretação dos resultados, com possibilidade de suscetíveis erros (Coradi et al., 2020). Muitas vezes as análises de qualidade de sementes geram uma quantidade de informações que impossibilita uma análise rápida e eficaz a curto prazo. Portanto, resultados errôneos podem implicar em prejuízos econômicos para as unidades de beneficiamento de sementes (André et al., 2022).

Ao longo dos anos, a inteligência artificial, mais precisamente os modelos de aprendizado de máquina (AM), estão sendo introduzidos aos meios de produção agrícola (Benos et al., 2021). Por meio do reconhecimento de padrões se tem aumentado a eficiência e a otimização dos processos agrícolas ao utilizar a máquina para fazer recomendações ou tomar decisões a partir de algoritmos e um conjunto significativo de dados (Helm et al., 2020; Batarseh et al., 2021). Os métodos de AM têm sido bastante utilizados para prever a qualidade das sementes de soja (André et al., 2022), e classificar com alta acurácia sementes de milho (Xu et al., 2021). Assim como classificar as sementes de trigo de acordo com as propriedades de rendimento ao utilizar métodos como árvore de decisão, máquina de vetores de suporte e k-vizinho (Baryshev et al., 2020). Além de avaliar a taxa de germinação e qualidade de sementes através dos modelos de floresta aleatória, árvore de decisão e redes neurais com acurácia de até 98% (Škrubej et al., 2015; Ropelewska e Piecko, 2022).

Para minimizar as lacunas ocasionadas por análises convencionais de sementes, dependentes de interpretações pessoais, as técnicas de AM podem ser alternativas para analisar a qualidade de sementes de soja armazenadas, podendo ser utilizadas como uma ferramenta de apoio para tomadas de decisões sobre condições, formas e tempos de armazenamento, para manutenção da qualidade e redução de perdas de sementes de soja. Desta forma, o objetivo deste estudo foi identificar o melhor modelo de aprendizado de máquina para predição de danos mecânicos, vigor e viabilidade de sementes de soja durante o armazenamento.

MATERIAL E MÉTODOS

Descrição e delineamento experimental

Inicialmente, foram removidas as impurezas e matérias estranhas dos lotes de sementes de soja, com auxílio de uma máquina de ar e peneira da marca LC 160 (Kepler Weber, Rio Grande do Sul, Brasil). Na sequência, a soja foi submetida à secagem até a obtenção de teores de 12% de umidade (b.u.), utilizando silos-secadores com fluxo de ar radial e na temperatura de 40 °C (Silos Roma, Paraná, Brasil). Após a secagem, as sementes foram submetidas à classificação, utilizando um separador espiral (Akyurek Technology, Mersin, Turquia) e uma mesa assimétrica modelo SDS-80 (Silomax, Paraná, Brasil), a fim de padronizar o tamanho e a massa das sementes. As avaliações experimentais ocorreram na etapa de armazenagem, em um delineamento inteiramente casualizado (DIC), em um fatorial (3 x 2 x 5), sendo três temperaturas de armazenagem (10, 15 e 25 °C), duas embalagens (com revestimento e sem revestimento) e cinco tempos de armazenagem (0, 3, 6, 9 e 12 meses) (Figura 1).

As sementes foram armazenadas em sacos de rafia com e sem revestimento (polipropileno) em ambientes climatizados, com controle de temperatura. As embalagens utilizadas foram sacos de rafia confeccionados de 20 cm (largura) x 30 cm (altura) x 0,25 cm (espessura), com revestimento de polipropileno de alta densidade. A temperatura da massa das sementes foi monitorada com termohigrômetros digital modelo Logbox, RHT-LCD (Novus Electronic Products Company, Rio Grande do Sul, Brasil). A cada três meses de armazenagem, as sementes de soja foram amostradas para avaliação da qualidade.

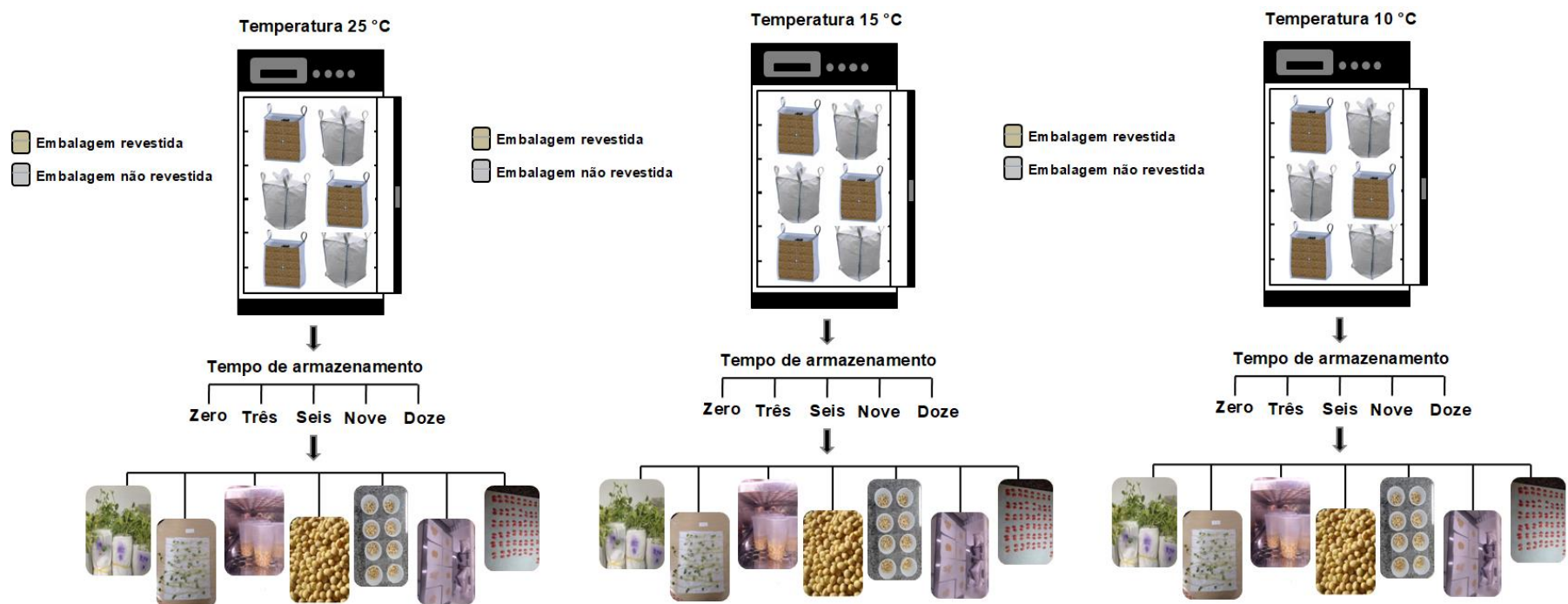


Figura 1. Esquema experimental

Análises de qualidade de sementes de soja

Utilizou-se uma estufa de circulação de ar forçado com capacidade de 220 L (Tecnal, São Paulo, Brasil) e temperatura do ar controlada a $105\text{ }^{\circ}\text{C} \pm 1\text{ }^{\circ}\text{C}$. Quatro amostras de 50 g de sementes de cada tratamento foram pesadas em balança modelo B13200H (Shimadzu, São Paulo - SP, Brasil) e colocadas na estufa por 24 h. Após, as amostras foram retiradas e colocadas em um dessecador (Tecnal, Piracicaba - SP, Brasil) com sílica para o resfriamento das sementes. Pela diferença de peso inicial e final das sementes determinaram-se os teores de água das sementes (% b.u.) (Brasil, 2009). A massa específica aparente das sementes foi determinada com auxílio de um béquer com volume conhecido de 1 L e uma balança de precisão. Calculou-se a massa específica aparente das sementes, através da relação massa e volume da amostra. Foram realizadas quatro repetições da análise para cada tratamento (Brasil, 2009).

Para a determinação dos testes de germinação (viabilidade) e vigor foram utilizadas quatro subamostras de 50 sementes de cada unidade experimental. As sementes foram distribuídas em rolos de papel toalha (Germitest), umedecidas com água destilada em aproximadamente 2,5 vezes a massa seca do papel. Após este processo, os rolos de papel com as sementes foram colocadas em um germinador do modelo Mangesdor (Tecnal, São Paulo, Brasil), condicionado à temperatura de $25\text{ }^{\circ}\text{C} \pm 2\text{ }^{\circ}\text{C}$. No quinto dia após o início dos testes avaliou-se o vigor das sementes, e após o oitavo dia realizou-se a avaliação da viabilidade das sementes pelo teste de germinação, conforme as Regras para Análise de Sementes (Brasil, 2009). Consideraram-se como sementes viáveis, aquelas capazes de produzir plântulas normais. Os níveis de viabilidade das sementes de soja em função dos danos mecânicos, umidades, vigor e viabilidade estão apresentados na Tabela 1, de acordo com as recomendações de França Neto et al. (1988).

Tabela 1. Níveis de viabilidade das sementes de soja em função dos diferentes danos mecânicos e umidades

Análises	Níveis de viabilidade das sementes*
Danos mecânicos 1 (DM1)	1 a 8
Danos mecânicos 2 (DM2)	6 a 8
Umidade 1 (U1)	1 a 8
Umidade 2 (U2)	6 a 8
Vigor (VG)	1 a 3
Viabilidade (VB)	1 a 6

Níveis: 1 – viável e mais alto vigor, 2 – viável e alto vigor, 3 – viável e vigor médio, 4 – viável e vigor baixo, 5 – viável e vigor muito baixo, 6 – não viável, 7 – não viável, 8 – semente morta.

No teste de tetrazólio foram utilizadas quatro subamostras de 50 sementes de cada unidade experimental. As sementes foram pré-umedecidas em papel Germitest por 16 h a 25 °C e depois imersos em solução de tetrazólio 0,075%, na qual foram mantidos por 3 h a 35 °C. Após esse período, as sementes foram lavadas em água corrente e seu vigor, viabilidade e danos por umidade (Kong et al., 2008; Mylona et al., 2012) foram avaliados de acordo com a metodologia estabelecida por França Neto e Krzyzanowski (1999).

Modelos de Aprendizado de Máquina (AM)

Os dados foram analisados no software Weka versão 3.9.5. utilizando os seguintes modelos: regressão linear múltipla (RL), redes neurais artificiais (RNA), algoritmo M5 de Quinlan (M5P) e floresta aleatória (FA) (Figura 2).

O modelo RL foi utilizado como modelo controle. A RNA testada consiste em uma camada oculta única formada por um número de neurônios que é igual ao número de atributos, mais o número de classes, todos divididos por 2 (Egmont-Petersen et al., 2002). O modelo M5P testado é uma reconstrução do algoritmo M5 de Quinlan que se baseia na árvore de decisão convencional com a adição de uma função de regressão linear aos nós das folhas (Blaifi et al., 2018). O modelo FA é capaz de produzir várias árvores de predição para o mesmo conjunto de dados e usar um esquema de votação entre todas essas árvores aprendidas para prever novos valores (Belgiu e Drăgu, 2016).

As análises foram realizadas utilizando a configuração padrão do software para todos os modelos testados (Bouckaert et al., 2010) em uma CPU Intel® Core™ i5-3317U com 4 Gb de RAM, com validação cruzada aleatória estratificada de dobras com 10 repetições. Os dados de entrada para cada modelo para a predição das variáveis foram: três diferentes temperaturas de armazenamento (T) (10, 15 e 25 °C), duas embalagens (E) (com revestimento e sem revestimento) e cinco tempos de armazenamento (TA) (0, 3, 6, 9 e 12 meses) para predição das variáveis dano mecânico 1 (DM1), dano mecânico 2 (DM2), umidade 1 (U1), umidade 2 (U2), vigor (VG) e viabilidade (VB).

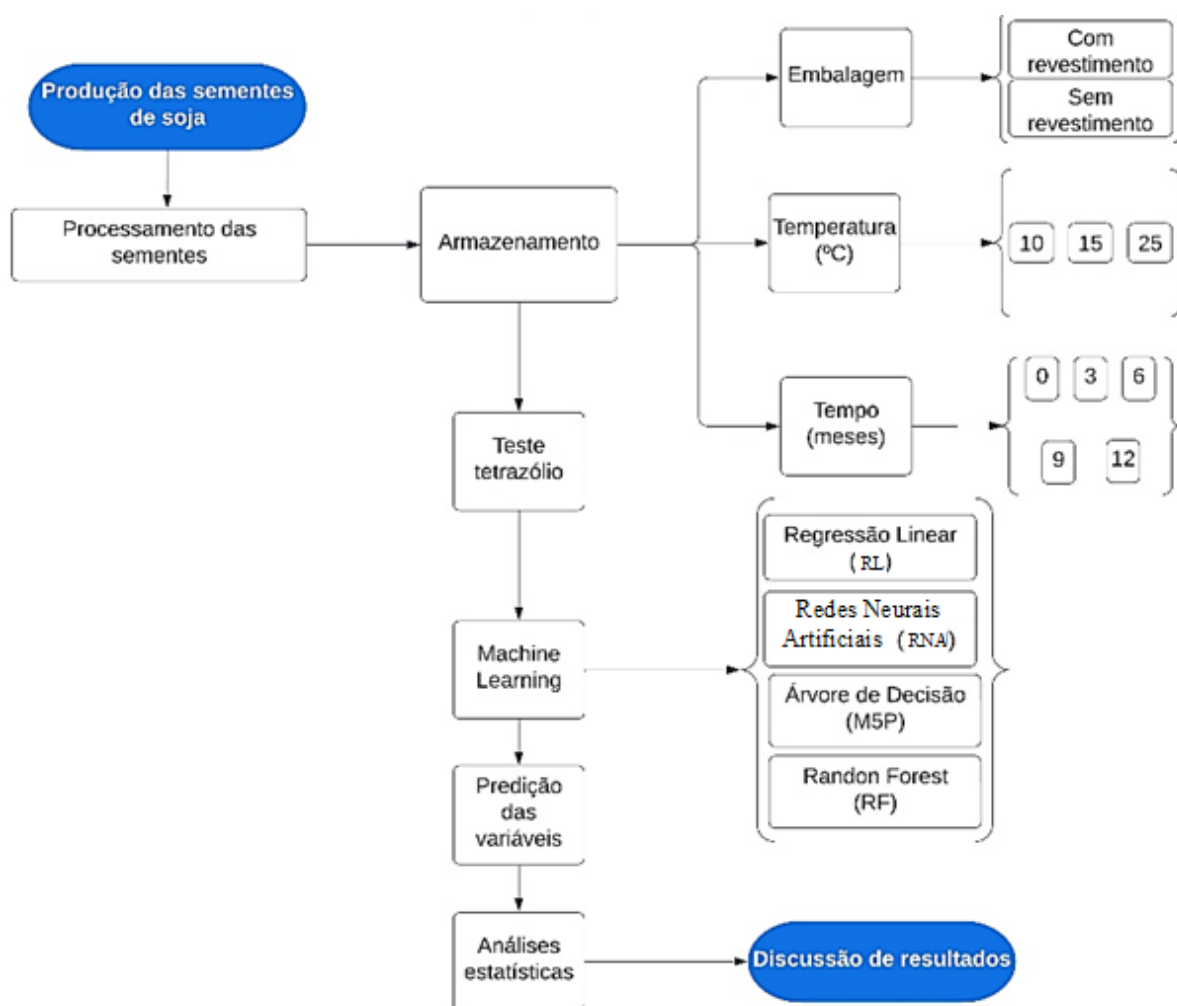


Figura 2. Fluxograma e análises experimentais.

Análises estatísticas

Foram obtidos o coeficiente de correlação (r) e o erro aparente médio (MAE) na predição da qualidade das sementes de soja armazenadas. Em seguida, fez-se a análise de variância, adotando-se o delineamento inteiramente ao acaso (DIC). Para o DIC foram considerados os modelos de aprendizado de máquina (RNAs, M5P e FA) mais a regressão linear múltipla (RL) para comparação. Adotou-se 10 repetições para cada modelo. Para o agrupamento das médias foi adotado o teste de Scott-knott a 5% de probabilidade. A partir das médias foram gerados gráficos Boxplot para cada variável (r e MAE). Utilizou-se para as análises o software Rbio (Bhering, 2017) e o software R (Team, 2018) com os pacotes ExpDes.pt e ggplot2.

RESULTADOS

As condições, embalagens e tempo de armazenamento influenciaram nos danos mecânicos (DM1) das sementes, em níveis de viabilidade de 1 a 8 (Figura 3). Neste estudo, o coeficiente de correlação de Pearson ($r = 0,58$) encontrados na análise de DM1 para os modelos preditivos não apresentaram diferenças estatísticas entre eles. O erro médio absoluto (MAE = 11,98) indicou que o modelo de redes neurais artificiais (RNA) foi o único que estatisticamente diferiu dos demais. O melhor ajuste dos dados foi obtido pelo modelo floresta aleatória (FA) (Figura 3).

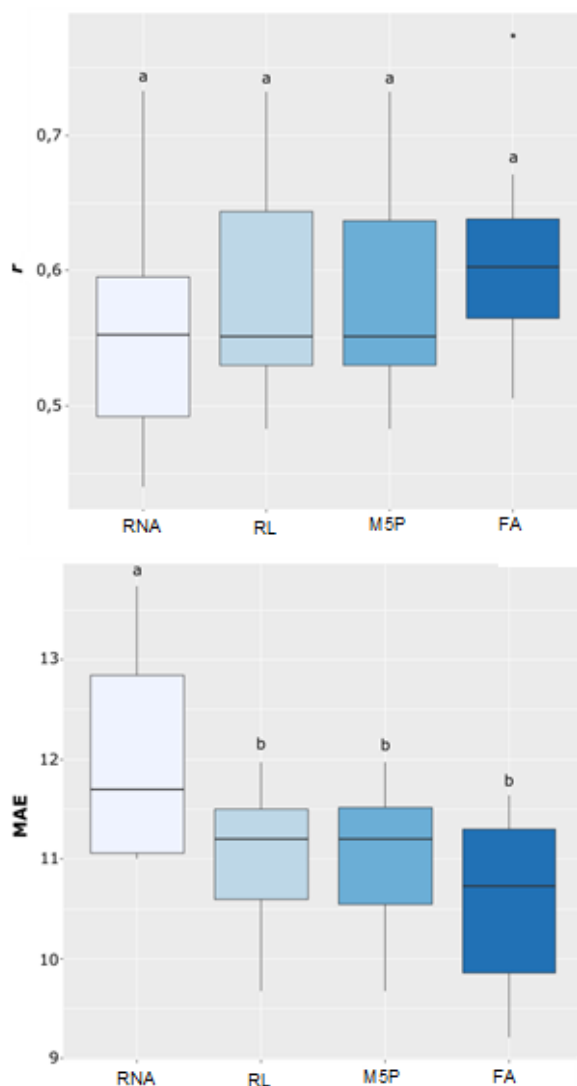


Figura 3. Boxplots da avaliação de dano mecânico (DM1) para o coeficiente de correlação de Pearson (r) e o erro médio absoluto (MAE) para diferentes modelos de aprendizado de máquina (AM). Regressão linear múltipla (RL), redes neurais artificiais (RNA), algoritmo M5 de Quinlan (M5P) e floresta aleatória (FA).

Os resultados foram similares na análise de dano mecânico 2 (DM2), para níveis de viabilidade entre 6 a 8, quando não se verificou diferenças estatísticas entre os modelos de predição. Além disto, os valores de r foram muito baixos ($r = 0,21$), indicando uma baixa precisão (Figura 4). Ainda assim, os modelos de FA e o algoritmo M5 de Quinlan (M5P) ($MAE = 3,0$) foram aqueles que melhor ajustaram os dados.

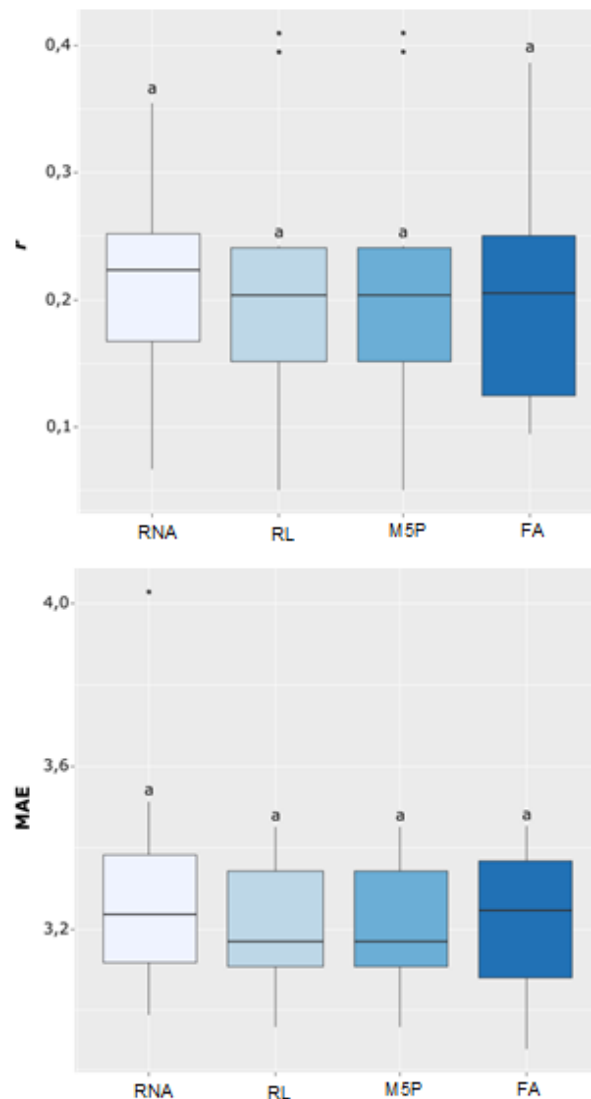


Figura 4. Boxplots da avaliação de dano mecânico (DM2) para o coeficiente de correlação de Pearson (r) e o erro médio absoluto (MAE) para diferentes modelos de aprendizado de máquina (AM). Regressão linear múltipla (RL), redes neurais artificiais (RNA), algoritmo M5 de Quinlan (M5P) e floresta aleatória (FA).

Na análise de umidade 1 (Figura 5.5), os modelos M5P, FA e RNA tiveram as maiores médias de r , variando de 0,45 e 0,41, enquanto que o modelo de RL alcançou um $r = 0,32$. Os

resultados de MAE não apresentaram diferenças estatísticas significativas entre os modelos, variando de 19,53 a 18,04.

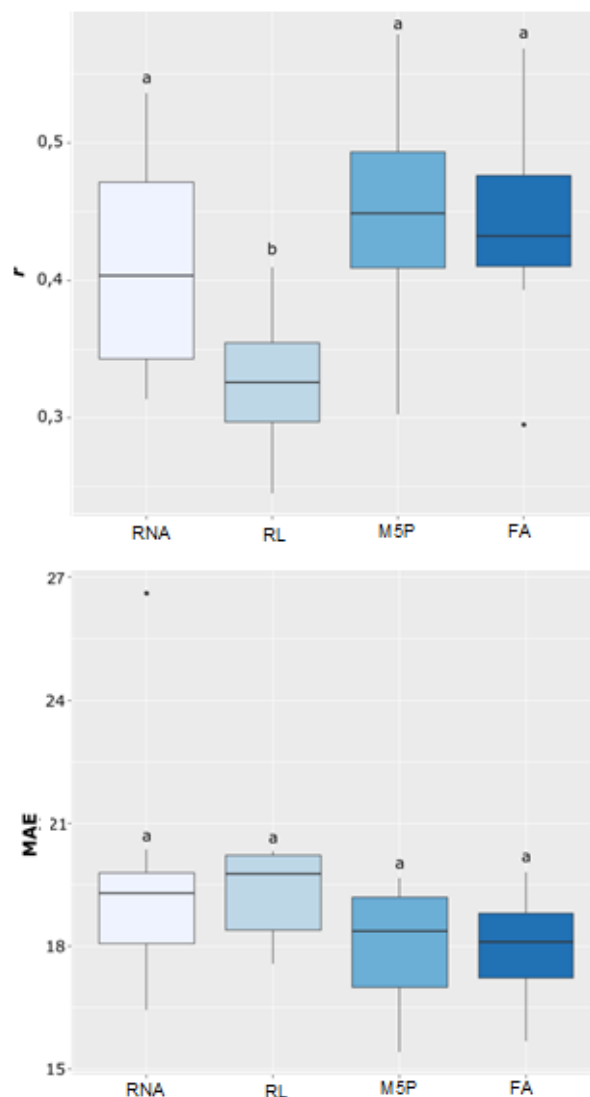


Figura 5. Boxplots da avaliação de umidade 1 (U1) para o coeficiente de correlação de Pearson (r) e o erro médio absoluto (MAE) para diferentes modelos de aprendizado de máquina (AM). Regressão linear múltipla (RL), redes neurais artificiais (RNA), algoritmo M5 de Quinlan (M5P) e floresta aleatória (FA).

Para a variável umidade 2 (Figura 6) os valores de r foram satisfatórios para os modelos FA, M5P e RNA, não apresentando diferenças estatísticas entre eles e variando entre 0,87 e 0,85. Para a MAE, os resultados médios foram de 3,37 para os modelos FA, M5P e RNA, enquanto que o modelo RL o valor da MAE chegou a 6,17. Pode-se afirmar, que o modelo de floresta aleatória foi o modelo que melhor predisse os resultados de umidade 2.

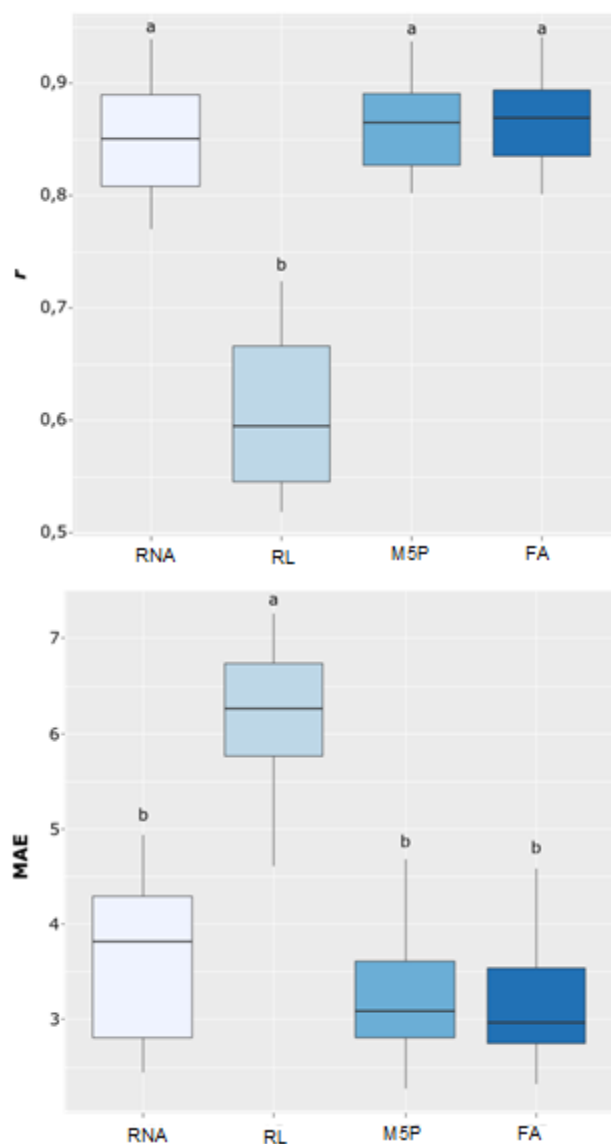


Figura 6. Boxplots da avaliação de umidade 2 (U2) para o coeficiente de correlação de Pearson (r) e o erro médio absoluto (MAE) para diferentes modelos de aprendizado de máquina (AM). Regressão linear múltipla (RL), redes neurais artificiais (RNA), algoritmo M5 de Quinlan (M5P) e floresta aleatória (FA).

Na análise de vigor, os modelos M5P e FA alcançaram os melhores resultados de r (0,75), enquanto que o modelo de RNA obteve um $r = 0,74$ e o modelo de RL um $r = 0,52$. Os resultados da MAE foram mais baixos (10,0) para os modelos de M5P e FA, não diferindo-se estatisticamente (Figura 7).

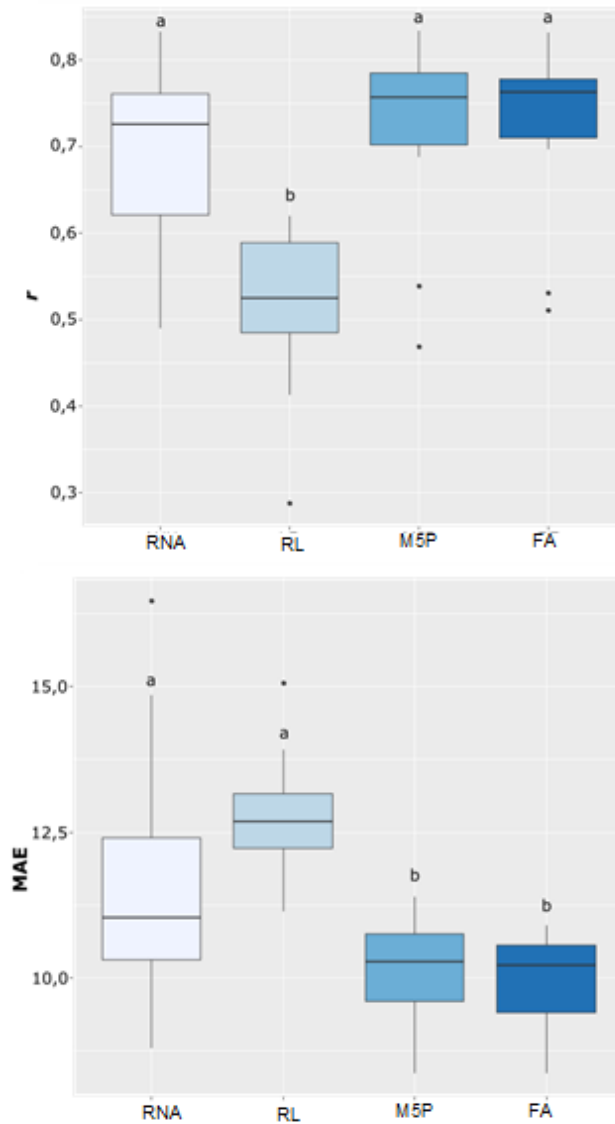


Figura 7. Boxplots da avaliação vigor (VG) para o coeficiente de correlação de Pearson (r) e o erro médio absoluto (MAE) para os diferentes modelos de aprendizado de máquina (AM). Regressão linear múltipla (RL), redes neurais artificiais (RNA), algoritmo M5 de Quinlan (M5P) e floresta aleatória (FA).

Na análise de viabilidade das sementes de soja, verificou-se que os modelos FA, M5P e RNA apresentaram os resultados mais satisfatórios ($r = 0,80$), porém não houve diferença estatística entre eles. Considerando o erro médio absoluto da variável, verificou-se que os modelos FA e M5P apresentaram os menores erros 4,92 e 4,94, respectivamente (Figura 8). A quantidade de informações geradas pelo número elevado de análises puderam ser melhor preditas e interpretadas pelo modelo de FA.

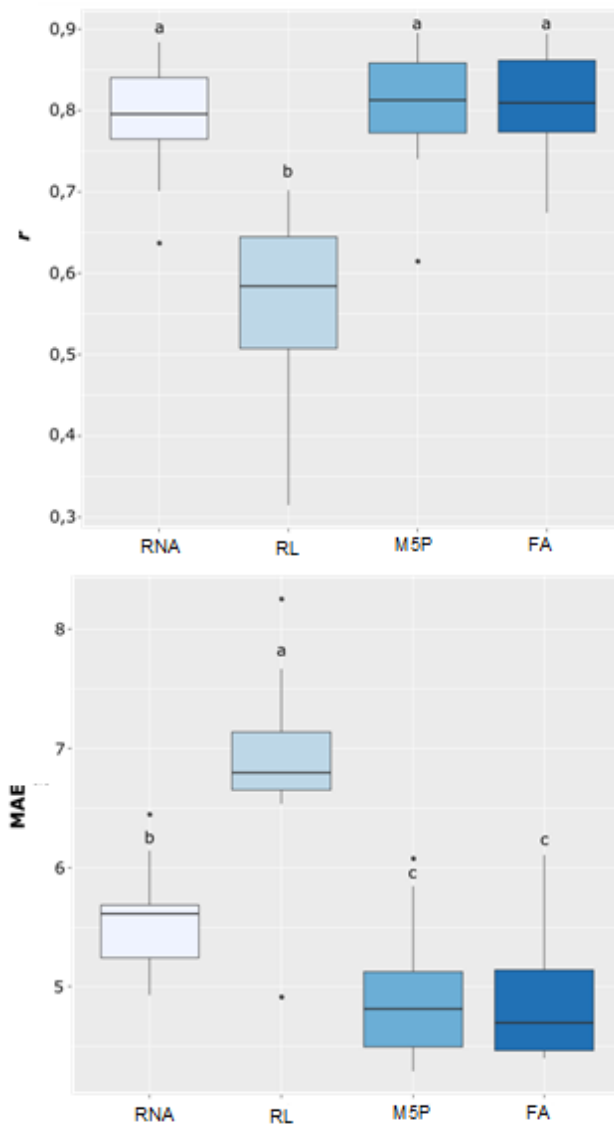


Figura 8. Boxplots da avaliação de viabilidade (VB) para o coeficiente de correlação de Pearson (r) e o erro médio absoluto (MAE) para os diferentes modelos de aprendizado de máquina (AM). Regressão linear múltipla (RL), redes neurais artificiais (RNA), algoritmo M5 de Quinlan (M5P) e floresta aleatória (FA).

DISCUSSÃO

Sabe-se que os altos níveis de danos mecânicos nas sementes afetam o potencial fisiológico, diminuindo vigor e a viabilidade da soja (Oliveira et al., 2021; Neve et al., 2016). Conforme o avanço tecnológico e o impacto da viabilidade na determinação da qualidade das sementes (KUSUMANINGRUM, et al., 2018), e a necessidade de análises mais rápidas e economicamente viáveis, o aprendizado de máquina está se tornando uma alternativa aos métodos convencionais (BAEK et al., 2019).

Assim, pesquisas recentes demonstraram a eficácia dos modelos de aprendizado de máquina (AM) na predição da viabilidade, vigor e velocidade de germinação de sementes de diferentes culturas (Medeiros et al., 2020). Pereira et al. (2020), André et al (2022) e Gadotti et al. (2022) alcançaram resultados satisfatórios utilizando algoritmos de Aprendizado de Máquina (AM) para prever e classificar sementes de soja, porém os modelos que melhor desempenharam a predição foram diferentes em função das condições de processamento e armazenamento.

Conforme estudo realizado por Ropelewska e Piecko (2022), o método redes neurais artificiais (RNA) teve alta acurácia para determinar viabilidade das sementes. Essa variação também ocorreu neste estudo realizado com sementes de soja, porém os algoritmos de M5 de Quinlan (M5P) e floresta aleatória (FA) foram os mais recomendados em função dos menores erros estatísticos (r e MAE), considerando as diferentes condições de temperatura, embalagens e tempo de armazenamento (Coradi et al., 2020).

Estudos recentes apontam uma precisão superior a 85% quando utilizado o aprendizado de máquina para prever a viabilidade e o vigor em diferentes variedades de sementes de arroz (JIN et al., 2022). Fan et al. (2020) ao combinar o algoritmo de projeções sucessivas com o algoritmo floresta aleatória obteve acurácia de 88,5% ao predizer e distinguir as sementes de trigo em vigorosas e não viáveis.

Este trabalho, foi motivado pela verificação do desempenho de diferentes técnicas de aprendizado de máquina para avaliar a predição de variáveis que representam a qualidade fisiológica das sementes de soja armazenadas. As técnicas de aprendizado de máquina superaram os resultados do modelo de controle proposto. Entre os algoritmos utilizados, árvore de decisão e floresta aleatória resultaram em melhores predições. A técnica de redes neurais artificiais apresentou maiores erros quando comparado com floresta aleatória e o algoritmo M5 de Quinlan (M5P). Para trabalhos futuros pretendemos testar outros algoritmos.

CONCLUSÕES

Os modelos de algoritmo M5 de Quinlan (M5P) e floresta aleatória (FA) preveem melhor os resultados de qualidade das sementes de soja durante o armazenamento. A identificação do melhor modelo de predição da qualidade das sementes de soja armazenadas proporciona uma redução de erros que poderiam ocorrer comumente nas análises laboratoriais baseadas em estimativas visuais.

REFERÊNCIAS BIBLIOGRÁFICAS

ANDRÉ, G. S. et al. Predicting the quality of soybean seeds stored in different environments and packaging using machine learning. **Scientific Reports**, v. 12, n. 1, p. 1-13, 2022.

<https://doi.org/10.1038/s41598-022-12863-5>

BAEK, I. et al. Rapid measurement of soybean seed viability using Kernel-based multispectral image analysis. **Sensors**, v. 19, n. 2, p. 271, 2019.

<https://doi.org/10.3390/s19020271>

BARYSHEV, D.D. et al. Comparison of Machine Learning Methods for Solving the Problem of Wheat Seeds Classification by Yield Properties. **Russian Agricultural Sciences**. v. 46, p. 410–41, 2020.

<https://doi.org/10.3103/S1068367420040047>

BATARSEH, F.A. et al. A survey on artificial intelligence assurance. **Journal of Big Data**, v.8, n. 60, 2021.

<https://doi.org/10.1186/s40537-021-00445-7>

BENOS, L. et al. Machine Learning in Agriculture: A Comprehensive Updated Review. **Sensors**, v. 21, n. 11, 2021.

<https://doi.org/10.3390/s21113758>

BELGIU, M.; DRĂGUT, 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.

<https://doi.org/10.1016/j.isprsjprs.2016.01.011>

BHERING, L. L. Rbio: A tool for biometric and statistical analysis using the R platform. **Crop Breeding and Applied Biotechnology**, v. 17, p. 187-190, 2017.

<https://doi.org/10.1590/1984-70332017v17n2s29>

BLAIFI, S. A. et al. M5P model tree based fast fuzzy maximum power point tracker. **Solar Energy**, v. 163, p. 405-424, 2018.

<https://doi.org/10.1016/j.solener.2018.01.071>

BOUCKAERT, R. R. et al. **Weka manual for version 3-6-0**. University of Waikato, Hamilton, New Zealand, 2, 2008.

<https://www.academia.edu/download/57056677/WekaManual-3-6-2.pdf>

BRASIL. Ministry of agriculture, Livestock and supply. Normative instruction No. 06, of February 16, 2009. Official gazette of the federative republic of Brazil. Brasília, Brazil: Executive Branch, 2009.

CORADI, P. C. et al. Evaluation of coatings for application in rafa big bags in conditioned storage of soybean cultivars in seed processing units. **Journals Plos One**. v. 15, 2020.

<https://doi.org/10.1371/journal.pone.0242522>

EGMONT-PETERSEN, M. et al. Image processing with neural networks—a review. **Pattern recognition**, v. 35, n. 10, p. 2279-2301, 2002.

[https://doi.org/10.1016/S0031-3203\(01\)00178-9](https://doi.org/10.1016/S0031-3203(01)00178-9)

FRANÇA NETO, J. B.; KRZYZANOWSKI, F. C. Metodologia do teste de tetrazólio em semente de soja. **Embrapa Soja**, Londrina. ISSN 2176-2937; n.406, 2018.

FRANÇA NETO, J. B.; KRZYZANOWSKI, F. C. Tetrazolium: an important test for physiological seed quality evaluation. **Journal of Seed Science**. Viçosa, v. 41, n. 3, jul/set, 2019.

<https://doi.org/10.1590/2317-1545v41n3223104>

GADOTTI, G. I. et al. Machine learning for soybean seeds lots classification. **Engenharia Agrícola**, v. 42, 2022.

<https://doi.org/10.1590/1809-4430-Eng.Agric.v42nepe20210101/2022>

HELM, J. M. et al. Machine Learning and Artificial Intelligence: Definitions, Applications, and Future Directions. **Current Reviews in Musculoskeletal Medicine**, v.13, p. 69-76, 2020.

<https://doi.org/10.1007/s12178-020-09600-8>

KONG, F. et al. Changes of soybean quality during storage as related to soymilk and tofu making. **Journal of Food Science**, v. 73, n. 3, p. S134-S144, 2008.

<https://doi.org/10.1111/j.1750-3841.2007.00652.x>

KUSUMANINGRUM, D. et al. Non-destructive technique for determining the viability of soybean (*Glycine max*) seeds using FT-NIR spectroscopy. **Journal of the Science of Food and Agriculture**, v. 98, n. 5, p. 1734-1742, 2018.

<https://doi.org/10.1002/jsfa.8646>

MEDEIROS, A. D. de et al. Interactive machine learning for soybean seed and seedling quality classification. **Scientific Reports**, v. 10, p. 1-10, 2020

<https://doi.org/10.34117/bjdv6n8-725>

MYLONA, K. et al. Relationship between environmental factors, dry matter loss and mycotoxin levels in stored wheat and maize infected with *Fusarium* species. **Food Additives & Contaminants: Part A**, v. 29, n. 7, p. 1118-1128, 2012.

<https://doi.org/10.1080/19440049.2012.672340>

NEVE, J. M. G. et al. Quality of Soybean Seeds with High Mechanical Damage Index after Processing and Storage. **Revista Brasileira de Engenharia Agrícola e Ambiental**, v. 20 n. 11, 2016.

<https://doi.org/10.1590/1807-1929/agriambi.v20n11p1025-1030>

OLIVEIRA, G. R. F. et al. Treatment of Soybean Seeds with Mechanical Damage: Effects on Their Physiological Potential. *Journal of Seed Science* 43 (2021): **Journal of Seed Science**, v. 43, 2021.

<https://doi.org/10.1590/2317-1545v43247404>

PEREIRA, D. F. et al. Assessing Active Learning Strategies to Improve the Quality Control of the Soybean Seed Vigor. **IEEE Transactions on Industrial Electronics**, v. 68, n. 2, p. 1675-1683, 2020.

<https://doi.org/10.1109/TIE.2020.2969106>

TEAM, R. C. R: A language and environment for statistical computing, 2018.

ROCHA, G. C. et al. Physiological quality of treated and stored soybean seeds. **Cient. Mult. Journal**, v.4, p. 50-65, 2017.

<https://doi.org/10.29247/2358-260X.2017v4i1.p50-65>

ROPELEWSKA, E.; PIECKO, J. Discrimination of tomato seeds belonging to different cultivars using machine learning. **Technol European Food Research and Technology**, v. 248, p. 685–705, 2022.

<https://doi.org/10.1007/s00217-021-03920-w>

ŠKRUBEJ, U. et al. Assessment of germination rate of the tomato seeds using image processing and machine learning. **European Journal of Horticultural Science**, v. 80, n. 2, p. 68-75, 2015.
<http://dx.doi.org/10.17660/eJHS.2015/80.2.4>

XU, P. et al. Varietal classification of maize seeds using computer vision and machine learning techniques. **Journal of Food Process Engineering**, v. 44, n. 11, p. e13846, 2021.

<https://doi.org/10.1111/jfpe.13846>