

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

GEOVANE DA SILVA ANDRÉ

**PREDIÇÃO DA QUALIDADE DE SEMENTES DE SOJA
ARMAZENADAS EM DIFERENTES AMBIENTES E EMBALAGENS
USANDO APRENDIZADO DE MÁQUINA**

CHAPADÃO DO SUL – MS

2021

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

GEOVANE DA SILVA ANDRÉ

**PREDIÇÃO DA QUALIDADE DE SEMENTES DE SOJA
ARMAZENADAS EM DIFERENTES AMBIENTES E EMBALAGENS
USANDO APRENDIZADO DE MÁQUINA**

Orientador: Prof. Dr Paulo Carteri Coradi

Co-Orientadora: Prof. Dra Larissa Pereira Ribeiro Teodoro

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

CHAPADÃO DO SUL – MS

2021



Serviço Público Federal
Ministério da Educação

Fundação Universidade Federal de Mato Grosso do Sul



PROGRAMA DE PÓS-GRADUAÇÃO EM AGRONOMIA

CERTIFICADO DE APROVAÇÃO

DISCENTE: Geovane da Silva André

ORIENTADOR: Dr. Paulo Carteri Coradi

TÍTULO: Predição da qualidade de sementes de soja armazenadas em diferentes ambientes e embalagens usando aprendizado de máquina

AVALIADORES:

Prof. Dr. Paulo Carteri Coradi

Prof. Dr. Paulo Eduardo Teodoro

Prof. Dr. Alan Mario Zuffo

Chapadão do Sul, 30 de novembro de 2021.



Documento assinado eletronicamente por **Paulo Carteri Coradi, Usuário Externo**, em 30/11/2021, às 10:58, 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 Eduardo Teodoro, Professor do Magisterio Superior**, em 30/11/2021, às 10:59, 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 **Alan Mario Zuffo, Professor do Magisterio Superior - Visitante**, em 30/11/2021, às 11:00, 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](#).

DEDICATÓRIA

*À minha mãe, pai, irmã e noiva
por serem os maiores
incentivadores das realizações
dos meus objetivos.*

DEDICO

AGRADECIMENTOS

Com gratidão, agradeço primeiramente a Deus - fonte de toda sabedoria e poder, pela conclusão deste trabalho.

À minha noiva que acompanhou de perto todo meu esforço e empenho para a conclusão desta dissertação. À meus pais e minha irmã que mesmo de longe ficaram na torcida para meu melhor desempenho no curso.

Ao meu Orientador, Prof. Dr Paulo Carteri Coradi, pela oportunidade de trabalharmos juntos, pelos ensinamentos, por seu exemplo profissional de dedicação à pesquisa, ensino e, sobretudo, pela disponibilidade e paciência em todos os momentos.

Aos docentes do Programa de Pós-Graduação em Agronomia, pelo apoio e pela disponibilidade que sempre demonstraram, em especial à professora e co-orientadora Dra. Larissa Pereira Ribeiro Teodoro e ao professor Dr. Paulo Eduardo Teodoro por todo o auxílio que deram a mim durante o curso.

Aos membros da banca examinadora, Prof. Dr. Paulo Eduardo Teodoro e Prof. Dr. Alan Mario Zuffo pela disponibilidade e contribuição na melhoria do trabalho.

À Universidade Federal de Mato Grosso do Sul (UFMS) e ao Programa de Pós-Graduação em Agronomia pela oportunidade e estrutura, por ter me acolhido como aluno de mestrado e por todo o aprendizado que me proporcionou.

A todos profissionais da saúde, aos cientistas, pesquisadores, e a outros tantos profissionais que incansavelmente trabalharam para que o quanto antes tivéssemos disponível a vacina contra o Covid-19.

EPÍGRAFE

“Descobrir consiste em olhar para o que todo mundo está vendo e pensar uma coisa diferente”.

(Roger Von Oech)

LISTA DE FIGURAS

Figura 1. Etapas de desenvolvimento da pesquisa.	16
Figura 2. Valores médios e dispersão para as variáveis coeficiente de correlação de Pearson (r) e erro médio absoluto (EMA) entre os valores observados e estimados do teor de água em sementes de soja por diferentes modelos de aprendizagem de máquina e entradas.	22
Figura 3. Valores médios e dispersão para as variáveis coeficiente de correlação de Pearson (r) e erro médio absoluto (EMA) entre os valores observados e estimados da massa específica em sementes de soja por diferentes modelos de aprendizagem de máquina e entradas.....	24
Figura 4. Valores médios e dispersão para as variáveis coeficiente de correlação de Pearson (r) e erro médio absoluto (EMA) entre os valores observados e estimados da condutividade elétrica em sementes de soja por diferentes modelos de aprendizagem de máquina e entradas.	26
Figura 5. Valores médios e dispersão para as variáveis coeficiente de correlação de Pearson (r) e erro médio absoluto (EMA) entre os valores observados e estimados da germinação em sementes de soja por diferentes modelos de aprendizagem de máquina e entradas.	28
Figura 6. Valores médios e dispersão para as variáveis coeficiente de correlação de Pearson (r) e erro médio absoluto (EMA) entre os valores observados e estimados do vigor em sementes de soja por diferentes modelos de aprendizagem de máquina e entradas.	30

LISTA DE TABELAS

Tabela 1. Projeto experimental e agrupamento de ambientes de armazenamento.....	19
Tabela 2. P-valor da análise de variância para o coeficiente de correlação de Pearson (r) entre os valores observados e estimados do teor de água (TA), massa específica (ME), condutividade elétrica (CE), germinação (GERM) e vigor em sementes de soja por diferentes modelos de aprendizagem de máquina e inputs.	20
Tabela 3. Desdobramento da interação significativa entre modelo x entrada para o coeficiente de correlação de Pearson (r) entre os valores observados e estimados do teor de água em sementes de soja por diferentes modelos de aprendizagem de máquina e inputs.	21
Tabela 4. Desdobramento da interação significativa entre modelo x entrada para o erro médio absoluto (EMA) entre os valores observados e estimados do teor de água em sementes de soja por diferentes modelos de aprendizagem de máquina e inputs.	21
Tabela 5. Agrupamento de médias para as variáveis coeficiente de correlação de Pearson (r) e erro médio absoluto (EMA) entre os valores observados e estimados da massa específica em sementes de soja por diferentes modelos de aprendizagem.	23
Tabela 6. Agrupamento de médias para as variáveis coeficiente de correlação de Pearson (r) e erro médio absoluto (EMA) entre os valores observados e estimados da massa específica em sementes de soja em função de diferentes inputs.	23
Tabela 7. Agrupamento de médias para as variáveis coeficiente de correlação de Pearson (r) entre os valores observados e estimados da condutividade elétrica em sementes de soja por diferentes modelos de aprendizagem.	24
Tabela 8. Agrupamento de médias para as variáveis coeficiente de correlação de Pearson (r) entre os valores observados e estimados da condutividade elétrica em sementes de soja em função de diferentes inputs.....	25
Tabela 9. Desdobramento da interação significativa entre modelo x entrada para o erro médio absoluto (EMA) entre os valores observados e estimados da condutividade elétrica em sementes de soja por diferentes modelos de aprendizagem de máquina e inputs.	25
Tabela 10. Desdobramento da interação significativa entre modelo x entrada para o coeficiente de correlação de Pearson (r) entre os valores observados e estimados da germinação em sementes de soja por diferentes modelos de aprendizagem de máquina e inputs.	26
Tabela 11. Desdobramento da interação significativa entre modelo x entrada para o erro médio absoluto (EMA) entre os valores observados e estimados da germinação em sementes de soja por diferentes modelos de aprendizagem de máquina e inputs.....	27

Tabela 12. Agrupamento de médias para as variáveis coeficiente de correlação de Pearson (r) e erro médio absoluto (EMA) entre os valores observados e estimados do vigor em sementes de soja por diferentes modelos de aprendizagem..... 28

Tabela 13. Agrupamento de médias para as variáveis coeficiente de correlação de Pearson (r) e erro médio absoluto (EMA) entre os valores observados e estimados do vigor em sementes de soja em função de diferentes inputs..... 29

PREDIÇÃO DA QUALIDADE DE SEMENTES DE SOJA ARMAZENADAS EM DIFERENTES AMBIENTES E EMBALAGENS USANDO APRENDIZADO DE MÁQUINA

RESUMO

O armazenamento como etapa do pós-colheita é responsável por postergar a deterioração e manter a qualidade das sementes, e durante este tempo algumas variáveis tem impacto direto sobre o material, como o tempo de armazenamento, temperatura e embalagem. Análises tradicionais de qualidade de sementes são processos que, por vezes, requerem vários testes em laboratórios para obter resultados confiáveis. Portanto, ao utilizar técnicas de aprendizado pode maximizar o tempo, e assim colaborar com as melhores tomadas de decisão nos processos de armazenagem de sementes. Este trabalho analisou a viabilidade de aplicar algoritmos de aprendizado de máquina para prever a qualidade fisiológica das sementes de soja. Para isso foram testadas combinações diferentes de entradas (temperatura de armazenamento (C), embalagem (E), tempo (T), C+E, C+T, E+T e C+E+T) e modelos de aprendizagem de máquina (regressão linear, redes neurais artificiais, árvore de decisão, M5P e floresta aleatória). Para cada combinação, foram obtidos os valores do coeficiente de Person(r) e o erro absoluto médio entre os valores observados e preditos com 10 repetições (folds). Estes valores foram submetidos a uma análise de variância e agrupamento de médias pelo teste de Scott-Knott considerando um fatorial 5x7 (modelos de aprendizado de máquina x entradas) em delineamento inteiramente casualizado. Os resultados dessa pesquisa indicam a possibilidade de prever variáveis relacionadas a qualidade de sementes utilizando informações sobre o ambiente de armazenamento como entrada no modelo de floresta aleatória. A abordagem proposta se destaca em termos de custo e rapidez quando comparado aos métodos de análises rotineiramente utilizados.

Palavras-chaves: Inteligência computacional. Processamento de sementes. Qualidade fisiológica de sementes. *Glyxine max*.

ABSTRACT

Storage as a post-harvest stage is responsible for postponing deterioration and maintaining seed quality, and during this time some variables have a direct impact on the material, such as storage time, temperature and packaging. Traditional seed quality analyzes are processes that sometimes test several in laboratories to obtain results. Therefore, using learning techniques can maximize time, and thus collaborate with better decision making in storage processes. This work analyzed the feasibility of applying machine learning algorithms to predict the physiological quality of soybean seeds. For this, different inputs were tested (storage temperature (C), packaging (E), time (T), C + E, C + T, E + T and C + E + T) and machine learning models (linear regression, artificial neural networks, decision tree, M5P and random forest). For each combination, the Person coefficient (r) values and the mean absolute error between the observed and predicted values with 10 repetitions (folds) were captured. These values were collected in an analysis of variation and grouping of means by the Scott-Knott test considering a 5x7 factorial (machine processing models x inputs) in a completely randomized design. The results of this research indicate the possibility of predicting variables related to seed quality using information about the storage environment as input to the random forest model. The proposed approach stands out in terms of cost and speed when compared to routinely used analysis methods.

Keywords: Computational intelligence. Seed processing. Physiological quality of seeds. *Glyxine max*.

SUMÁRIO

ABSTRACT	11
1.INTRODUÇÃO	13
2. MATERIAL E MÉTODOS.....	14
2.1 Caracterização do experimento	14
2.2 Obtenção dos lotes de sementes da soja	16
2.3 Avaliação da qualidade física e fisiológica da semente de soja	17
2.4 Modelos de aprendizagem de máquina	19
2.5 Análises estatísticas	20
3. RESULTADOS.....	20
3.1 Análise de variância	20
3.2 Teor de água	21
3.3 Massa específica.....	22
3.4 Condutividade elétrica.....	24
3.5 Germinação.....	26
3.6 Vigor	28
4. DISCUSSÃO	30
5. CONCLUSÃO	33
6. REFERÊNCIAS	34

1. INTRODUÇÃO

Na pós-colheita, a etapa de armazenamento tem como objetivo a conservação da qualidade das sementes (Baoua et al., 2014; Coradi e Lemes, 2019). Porém, as variações dos teores de água das sementes, da forma, do ambiente e do tempo de armazenamento podem influenciar sobre a atividade metabólica e qualidade fisiológica das sementes (Kong et al., 2008; Silva et al., 2010; Mylona et al., 2012). Para reduzir a atividade metabólica das sementes sugere-se controlar a temperatura e a umidade relativa do ambiente de armazenamento, para que as sementes permaneçam em equilíbrio higroscópico com teores de água próximo à 12% (b.u.), considerado uma umidade segura (Lima et al., 2014; Sarath et al., 2016; Coradi e Lemes, 2018; Ebone et al., 2020).

Atualmente a utilização de tecnologias de resfriamento artificial na conservação de sementes no armazenamento (Alencar et al., 2008; Demito e Afonso, 2009; Mylona e Magan, 2011; Ferreira et al., 2017) tem se mostrado uma estratégia viável para a conservação das sementes. A manutenção das sementes à baixas temperaturas, associada a uma condição controlada de umidade relativa do ar pode proporcionar uma condição favorável de armazenamento. Entretanto, os custos com refrigeração e o deslocamento e a espera para a semeadura dos lotes de sementes após armazenamento, ainda assim poderiam comprometer a qualidade. O uso de embalagens herméticas ou semi-herméticas, ao invés da refrigeração artificial poderia manter a qualidade fisiológica das sementes (Coles et al., 2003; Yildirim, 2011; Yildirim et al., 2018).

Assim, no intuito de obter informações mais precisas sobre a qualidade de sementes na etapa de armazenamento, sugere-se a aplicação de algoritmos computacionais preditivos, quanto a deterioração, perdas de massa específica e percentuais de germinação em função das condições e tempo de armazenamento. Neste contexto o uso de Aprendizado de Máquina (AM) tem oferecido capacidade de processamento, análise e interpretações dos dados (Moreti et al., 2021). Quando devidamente modelados, os dados sob alguma técnica de AM podem oferecer as respostas com menor tempo, quando comparado aos modelos estatísticos de regressão.

Florestas Aleatórias (FA) é uma técnica de AM utilizada com êxito na previsão de safras (Jeong et al., 2016; Ramos et al., 2020). Ainda dentro do campo das agrárias, este método apresentou-se como um método eficaz e de maior facilidade de uso nas análises de predição do rendimento de safras (milho, trigo e batata) quando comparado aos modelos de

regressão linear múltiplas (Jeong et al., 2016). Redes Neurais Artificiais (RNA) é outro método que pode ser treinado a partir de dados relativos as entradas e saídas correspondentes de acordo com Pazoki e Pazoki (2011). As RNAs são ferramentas úteis para análises e interpretação de dados complexos de segurança alimentar, previsões físicas, química, como também em aplicações de agricultura de precisão (Goyal, 2014). Durante os últimos anos, pesquisas têm investigado os resultados da utilização de métodos de AM para a classificação dentro do contexto de problemas agrícolas, como por exemplo (previsão do teor de nitrogênio (Osco et al., 2019); prever disponibilidade de PAHs em solos corrigidos; classificação de sementes (Hussain e Ajaz, 2015); prever a redução do fósforo em águas residuais de usinas de arroz (Kumar e Deswal, 2020); classificação de proteínas de armazenamento (Radhika e Rao, 2014).

A fim de preencher lacunas aonde a estatística convencional não pode gerar resultados satisfatórios de predição. A adoção de modelagem de dados com técnicas de aprendizado de máquina poderia tornar-se uma alternativa viável para avaliação da qualidade de sementes de soja armazenada, ao invés dos testes demorados em laboratórios. Assim, este trabalho visou analisar a performance de algoritmos de AM utilizando como entrada combinações das variáveis de acondicionamento (temperatura, embalagens) e tempo de armazenamento para a predição da qualidade física e fisiológica de sementes de soja armazenadas.

2. MATERIAL E MÉTODOS

2.1 Caracterização do experimento

A pesquisa foi realizada no laboratório de Informática e no Laboratório de Sementes da Universidade Federal de Mato Grosso do Sul, Campus de Chapadão do Sul, juntamente com a colaboração do Laboratório de Pós-colheita da Universidade Federal de Santa Maria, Campus Cachoeira do Sul. Na figura 1 é apresentada a organização de como a pesquisa foi desenvolvida.

Foi utilizado o delineamento experimental inteiramente casualizado com três fatores ($2 \times 3 \times 5$). Duas embalagens (saco de rafia e saco de rafia com revestimento de polietileno). Três ambientes de armazenamento (25 °C, 15 °C e 10 °C). Cinco tempos de armazenamento momentos de avaliação (0 dias, 90 dias, 180 dias, 270 dias e 365 dias). A cada três meses, três

pacotes (ou seja, três repetições) de cada tratamento foram amostrados para fazer avaliações de qualidade. Após esse procedimento a embalagem foi descartada.

Os sacos de rafia foram feitos de material de polipropileno de 20 cm (largura) x 30 cm (altura) x 0,25 cm. O revestimento de polietileno utilizado para armazenar os grãos nos sacos de rafia tinha dimensões de 20 cm (largura) x 30 cm (altura) x 0,1 cm (espessura de alta densidade) sendo produzido pela empresa especializada em embalagens de alimentos (Videplast Company). As embalagens de polietileno foram constituídas por material de resina termoplástica parcialmente cristalina e flexível obtida através da polimerização do etileno, possuindo baixa densidade, alta tenacidade, boa resistência ao impacto, flexibilidade, fácil processabilidade, propriedades elétricas e estabilidade, e baixa permeabilidade à água. É formado por compostos orgânicos polares e pode ser alterado pela temperatura ambiente.

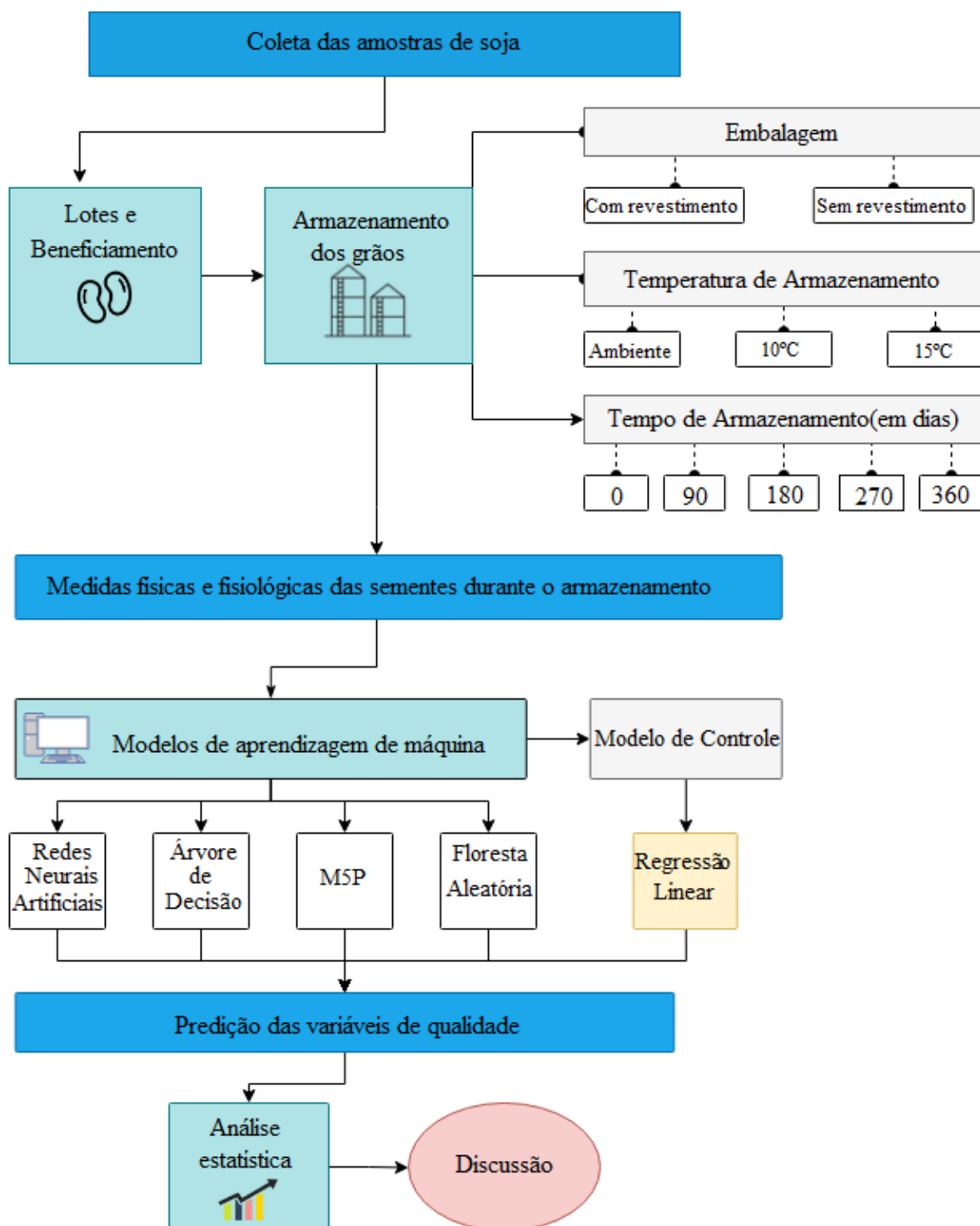


Figura 1. Etapas de desenvolvimento da pesquisa.

2.2 Obtenção dos lotes de sementes da soja

Este trabalho coletou amostras de soja de uma propriedade rural no município de Chapadão do Céu no estado de Goiás produtora de sementes. As sementes passaram previamente por um processo de limpeza antes do armazenamento para a retirada impurezas

que vira a existir. Em seguida foram submetidas à secagem sob temperatura de 40 °C em silos de ar radial até atingir o teor de água igual a 12% (b.u.), e para a padronização de acordo com peso e tamanho passaram por uma etapa de beneficiamento em um separador espiral e em uma mesa dessimétrica. Para a definição dos lotes, foi utilizado um calador manual para a retirada de 9 kg de amostras de sementes para o propósito do experimento. Os lotes com os bags revestidos de rafia contendo cada amostra foram armazenados em um ambiente climatizado com controle de temperatura.

2.3 Avaliação da qualidade física e fisiológica da semente de soja

Afim de analisar o impacto das condições de armazenamento na qualidade das sementes de soja reuniu três condições – Embalagem, Tempo de Armazenamento e Temperatura de Armazenamento, para delinear os ambientes de armazenamento, como apresentado na Tabela 1. Durante o tempo de armazenamento coletas de dados a respeito dos sementes eram realizadas para posterior estudo da qualidade física e fisiológica. Semanalmente a temperatura da massa das sementes eram monitoradas com um termohigrômetro digital (Novus[®], modelo Logbox-RHT-LCD).

Atendendo as recomendações da Regra para Análise de Sementes (Brasil, 2009) o teor de água foi obtido a partir de quatro repetições usando estufa de circulação forçada de ar (220 L, Tecnal Company) a 105 °C $\pm$ 1 °C, por 24 horas. O material utilizado na etapa anterior posteriormente é colocado em dessecador para resfriamento (5 L, Tecnal Company) e em seguida pesado (Shimadzu, modelo B13200H). Deste modo o teor de água é representado em porcentagem (bu) pela diferença entre o peso da amostra inicial e a final.

Para a determinação da massa específica aparente das sementes foi utilizado um becker de 150 ml e uma balança de precisão. Quatro repetições foram realizadas sob a relação de massa/volume (BRASIL, 2009). O teste de vigor foi conduzido a partir da realização do teste de tetrazólio, submetidas a um pré-umedecimento em papel “Germitest” por um período de 16 horas a 25°C, e após colocadas em uma solução de tereazólio a 0,075% por três horas em uma ambiente sem luz e com uma temperatura de 35°C. Após isso as sementes foram avaliadas seguindo a metodologia de França-Neto (1999).

A medição da condutividade elétrica foi realizada por meio da separação de quatro repetições 25 sementes de cada tratamento. Pesadas devidamente em uma balança de precisão de 0,001 g e em seguida colocadas em um recipiente com 75 mL de água destilada para que ficassem por 24 horas em uma incubadora BOD a 25 °C. Após este período foram feitas medições com um condutímetro digital (Digimed CD-21). Para a definição do teor de água foi

realizado a pesagem de 10g em duas repetições e colocadas em uma estufa a $105^{\circ}\text{C} \pm 1^{\circ}\text{C}$ durante 24 h. Em seguida deixadas em um dessecador para seu resfriamento. O índice de teor de água foi obtido a partir da diferença da massa na amostra inicial com a amostra final.

As realizações dos testes de germinação foram conduzidas utilizando quatro repetições de 50 sementes de cada tratamento. As sementes foram colocadas em papel toalha (Germitest) umedecido com 2.5 vezes a massa do papel, logo após o material foi colocado em um germinador (Mangesdorf) com temperatura ajustada em $25^{\circ}\text{C} \pm 2^{\circ}\text{C}$. Para avaliar os resultados realizou-se a verificação das plântulas normais e anormais, sementes duras, dormentes e mortas (Brasil, 2009).

2.4 Modelos de aprendizagem de máquina

Os modelos testados foram: Redes Neurais Artificiais (RNA), Árvore de Decisão (AD), algoritmo M5P, Floresta Aleatória (FA) e Regressão Linear (RL). 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 (Egmount-Petersen, 2002). O modelo AD é uma adaptação do classificador C4.5 que pode ser usado em problemas de regressão com uma etapa de poda adicional baseada em uma estratégia de redução de erro (Snousy et al., 2011). O modelo RL foi utilizado como modelo controle por servir para prever os comportamentos entre variáveis que tenham uma boa correlação, além de ser um modelo amplamente utilizado na estatística. O modelo M5P 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 et al., 2016).

A predição das variáveis teor de água (TA), massa específica (ME), condutividade elétrica (CE), germinação (GERM) e vigor em sementes de soja foi realizada pelos cinco modelos de aprendizado de máquina (ML) em uma validação cruzada estratificada aleatória de 10 vezes com 10 repetições (100 execuções para cada modelo). Foram considerados diferentes inputs para cada modelo na predição destas variáveis: temperatura de armazenamento (C), embalagem (E), tempo (T), C+E, C+T, E+T e C+E+T. As estatísticas utilizadas para verificar a qualidade do ajuste foram: coeficiente de correlação de Pearson (r) entre os valores observados e os preditos por cada modelo e o erro médio absoluto (EMA) dos valores preditos em relação aos observados.

As análises de aprendizagem de máquina foram realizadas com software Weka 3.9.4 utilizando a configuração padrão para todos os modelos testados (Bouckaert et al., 2010), em uma CPU Intel® Core™ i5 com 4 Gb de RAM.

Tabela 1. Projeto experimental e agrupamento de ambientes de armazenamento

Embalagem	Temperatura de Armazenamento (°C)	Tempo de Armazenamento Meses
Com revestimento	25	0
Com revestimento	25	3
Com revestimento	25	6
Com revestimento	25	9
Com revestimento	25	12
Com revestimento	15	0
Com revestimento	15	3
Com revestimento	15	6
Com revestimento	15	9
Com revestimento	15	12
Com revestimento	10	0
Com revestimento	10	3
Com revestimento	10	6
Com revestimento	10	9
Com revestimento	10	12
Sem revestimento	25	0
Sem revestimento	25	3
Sem revestimento	25	6
Sem revestimento	25	9
Sem revestimento	25	12
Sem revestimento	15	0
Sem revestimento	15	3
Sem revestimento	15	6
Sem revestimento	15	9
Sem revestimento	15	12
Sem revestimento	10	0
Sem revestimento	10	3
Sem revestimento	10	6
Sem revestimento	10	9
Sem revestimento	10	12

2.5 Análises estatísticas

Após a obtenção das estatísticas coeficiente de correlação (r) e do erro médio absoluto (EMA) foi realizada análise de variância considerando um esquema fatorial duplo (5x7). Neste fatorial foram considerados cinco modelos, sendo quatro de aprendizado de máquina (M5P, RNA, FA e AD) e um de regressão linear; e sete combinações de entrada C, T, E, C+E, C+T, E+T e C+E+T. Os resultados foram obtidos a partir de 10 repetições (folds). O r varia entre 0 e 1, e sua proximidade de 1 indica que o modelo está explicando melhor a variabilidade dos dados da amostra. O EMA espera um resultado inverso aos do coeficiente de correlação, neste caso, por ser usado para analisar o erro entre os valores preditos pelo modelo e os esperados, quanto menor os valores, mais próximo o modelo está das saídas observadas. A escolha das médias foram agrupadas pelo teste de Scott-knott a 5% de probabilidade. Foram construídos gráficos de barras para cada variável (r e EMA) considerando os modelos e inputs testados. Essas análises foram realizadas com software R R_Core_Team (2018) com auxílio dos pacotes ExpDes.pt e ggplot2.

3. RESULTADOS

3.1 Análise de variância

É possível observar que houve interação significativa (p -valor<0,05) entre esses fatores para o r e EMA do teor de água e da germinação, EMA da condutividade elétrica (Tabela 2). Para a r da massa específica houve efeito significativo apenas para os inputs, enquanto para o EMA houve significância dos efeitos isolados de M e I. Para o r da variável condutividade elétrica e r e EMA do vigor também houve significância dos efeitos isolados de M e I.

Tabela 2. P-valor da análise de variância para o coeficiente de correlação de Pearson (r) entre os valores observados e estimados do teor de água (TA), massa específica (ME), condutividade elétrica (CE), germinação (GERM) e vigor em sementes de soja por diferentes modelos de aprendizagem de máquina e inputs.

Fontes de variação	TA		ME		CE		GERM		Vigor	
	r	EMA	r	EMA	r	EMA	r	EMA	r	EMA
Modelo (M)	<0,00	<0,00	0,99	0,00	0,03	<0,00	<0,00	<0,00	0,02	0,00
Input (I)	<0,00	<0,00	0,00	0,00	0,00	<0,00	<0,00	<0,00	0,00	0,00
MxI	<0,00	<0,00	1,00	1,00	0,43	<0,00	<0,00	<0,00	0,40	0,64

3.2 Teor de água

O Para os inputs C, C+E, E, E+T e T não houve diferença entre os modelos testados. Contudo para os inputs C+E+T e C+T os modelos RNA, AD, M5P e FA apresentaram as maiores médias em relação a RL (Tabela 3). Ao analisar os inputs dentro de cada modelo, é possível observar que independente do modelo a configuração C+E+T proporcionou as maiores médias de r, todavia, para o modelo RL não diferiu do input C + T.

Tabela 3. Desdobramento da interação significativa entre modelo x entrada para o coeficiente de correlação de Pearson (r) entre os valores observados e estimados do teor de água em sementes de soja por diferentes modelos de aprendizagem de máquina e inputs.

Modelo	C	C+E	C+E+T	C+T	E	E+T	T
RNA	0,36 aE	0,43 aD	0,94 aA	0,86 aB	0,10 aF	0,63 aC	0,63 aC
AD	0,36 aE	0,43 aD	0,95 aA	0,87 aB	0,10 aF	0,63 aC	0,63 aC
RL	0,36 aC	0,37 aC	0,72 bA	0,72 bA	0,10 aD	0,63 aB	0,63 aB
M5P	0,36 aE	0,43 aD	0,94 aA	0,87 aB	0,10 aF	0,63 aC	0,63 aC
FA	0,36 aE	0,43 aD	0,95 aA	0,87 aB	0,10 aF	0,63 aC	0,63 aC

Médias seguidas por letras minúsculas iguais na mesma coluna e maiúsculas iguais na mesma linha não diferem entre si pelo teste de Scott-Knott a 5% de probabilidade.

Para os inputs C, C+E, E, o modelo RNA apresentou as maiores médias. Para o input C+E+T o modelo RL apresentou a maior média (Tabela 4). Para o input C+T os modelos RNA e RL apresentaram as maiores médias. Para os inputs E+T e T não houve diferenças estatísticas entre os modelos testados. É importante destacar que o EMA se comporta de forma contrária ao r e baixos valores dessa estatística refletem maior proximidade entre os valores observados e estimados. Ao analisar os inputs dentro de cada modelo, é possível observar que para os modelo analisados a configuração C+E+T proporcionou as menores médias de EMA. Mas, para o modelo RL não diferiu dos Inputs C+T, E+T e T.

Tabela 4. Desdobramento da interação significativa entre modelo x entrada para o erro médio absoluto (EMA) entre os valores observados e estimados do teor de água em sementes de soja por diferentes modelos de aprendizagem de máquina e inputs.

Modelo	C	C+E	C+E+T	C+T	E	E+T	T
RNA	1,26 aA	1,22 aA	0,41 bD	0,67 aC	1,33 aA	0,92 aB	0,92 aB
AD	1,07 bA	1,09 bA	0,30 bD	0,53 bC	1,11 bA	0,80 aB	0,81 aB
RL	1,07 bA	1,09 bA	0,73 aB	0,73 aB	1,11 bA	0,81 aB	0,81 aB
M5P	1,07 bA	1,09 bA	0,32 bD	0,53 bC	1,11 bA	0,81 aB	0,81 aB
FA	1,07 bA	1,09 bA	0,30 bD	0,53 bC	1,11 bA	0,81 aB	0,81 aB

Médias seguidas por letras minúsculas iguais na mesma coluna e maiúsculas iguais na mesma linha não diferem entre si pelo teste de Scott-Knott a 5% de probabilidade.

A Figura 2 demonstra graficamente os resultados das Tabelas anteriores onde é possível observar que modelos RNA, AD, M5P e FA quando associados aos inputs C+E+T e C+T proporcionaram os maiores valores de r e menores valores de EMA.

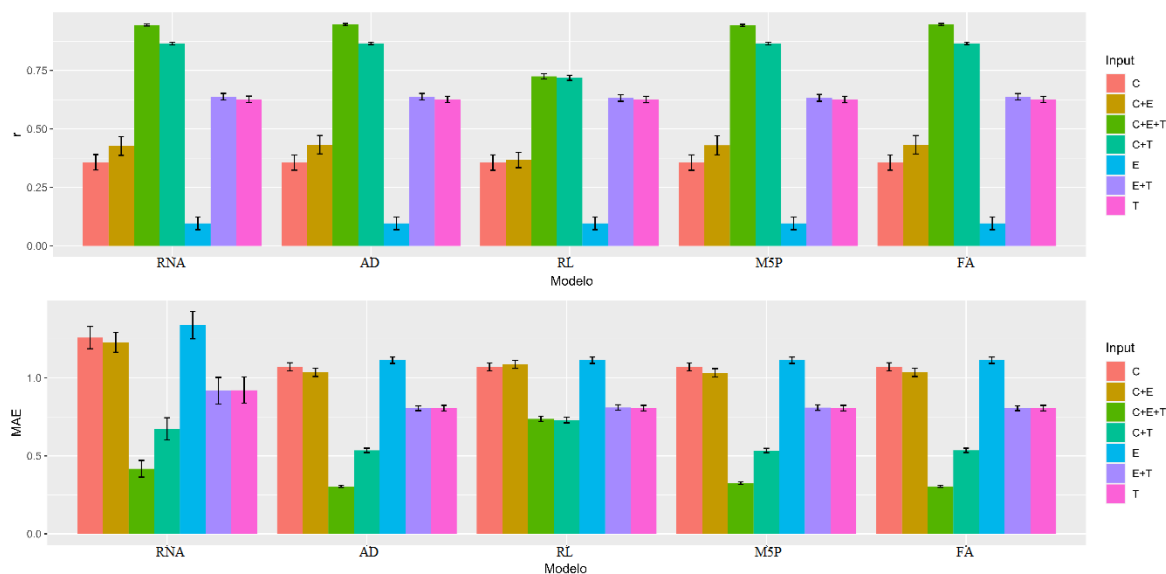


Figura 2. Valores médios e dispersão para as variáveis coeficiente de correlação de Pearson (r) e erro médio absoluto (EMA) entre os valores observados e estimados do teor de água em sementes de soja por diferentes modelos de aprendizagem de máquina e entradas.

3.3 Massa específica

As estatísticas obtidas para a variável massa específica (r e EMA) não apresentaram interação significativa entre modelos e inputs conforme visto na Tabela 5. Ao analisar os modelos testados, é possível observar que eles não diferiram entre si para o r . Contudo, o modelo RNA apresentou a maior média de EMA em relação aos demais, o que indica que esse modelo superestima os valores de massa específica. Com relação aos inputs testados, é possível observar na Tabela 6 que C+E+T, C+T, E+T e T apresentaram as maiores médias de r e as menores médias de EMA.

Tabela 5. Agrupamento de médias para as variáveis coeficiente de correlação de Pearson (r) e erro médio absoluto (EMA) entre os valores observados e estimados da massa específica em sementes de soja por diferentes modelos de aprendizagem.

Modelo	r	EMA
RNA	0,35 a	17,12 a
AD	0,35 a	10,45 b
RL	0,35 a	10,45 b
M5P	0,36 a	10,45 b
FA	0,35 a	10,44 b

Médias seguidas por letras iguais na mesma coluna não diferem entre si pelo teste de Scott-Knott a 5% de probabilidade.

Tabela 6. Agrupamento de médias para as variáveis coeficiente de correlação de Pearson (r) e erro médio absoluto (EMA) entre os valores observados e estimados da massa específica em sementes de soja em função de diferentes inputs.

Input	r	EMA
C	0,00 b	13,86 a
C+E	-0,02 b	13,88 a
C+E+T	0,63 a	10,47 b
C+T	0,63 a	10,31b
E	-0,02 b	13,88 a
E+T	0,63 a	10,47 b
T	0,63 a	9,59 b

Médias seguidas por letras iguais na mesma coluna não diferem entre si pelo teste de Scott-Knott a 5% de probabilidade.

A Figura 3 demonstra graficamente os resultados das Tabelas anteriores onde é possível observar que modelos AD, M5P e FA quando associados aos inputs C+E+T, C+T, E+T e T proporcionaram os maiores valores de r e menores valores de EMA. É importante destacar que embora o modelo RNA apresente bons resultados de r com os inputs supracitados, esse modelo apresentou altos valores de EMA para todos os inputs. Ademais, nenhum dos modelos tanto considerando r, como o EMA tiveram resultados diferentes estatisticamente do modelo de RL.

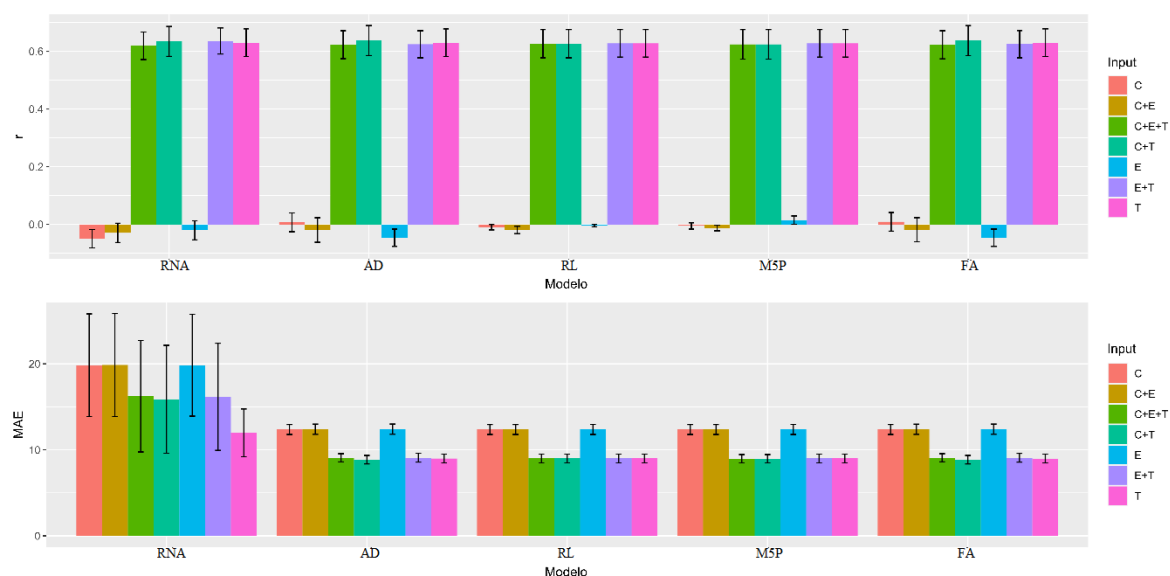


Figura 3. Valores médios e dispersão para as variáveis coeficiente de correlação de Pearson (r) e erro médio absoluto (EMA) entre os valores observados e estimados da massa específica em sementes de soja por diferentes modelos de aprendizagem de máquina e entradas.

3.4 Condutividade elétrica

Estatisticamente não foi observada diferença significativa para os modelos analisados conforme é apresentado na Tabela 7, porém ainda sim o valor de r para o modelo RL é menor se comparado com todos os outros modelos. A cerca dos inputs testados para a condutividade elétrica observa-se na Tabela 8 que C+E+T e C+T foram as entradas que tiveram as maiores médias de r, e o input isolado E apresentou o menor r médio.

Tabela 7. Agrupamento de médias para as variáveis coeficiente de correlação de Pearson (r) entre os valores observados e estimados da condutividade elétrica em sementes de soja por diferentes modelos de aprendizagem.

Modelo	R
RNA	0,41 a
AD	0,42 a
RL	0,38 b
MSP	0,42 a
FA	0,42 a

Médias seguidas por letras iguais na mesma coluna não diferem entre si pelo teste de Scott-Knott a 5% de probabilidade.

Tabela 8. Agrupamento de médias para as variáveis coeficiente de correlação de Pearson (r) entre os valores observados e estimados da condutividade elétrica em sementes de soja em função de diferentes inputs.

Input	R
C	0,32 c
C+E	0,34 c
C+E+T	0,65 a
C+T	0,63 a
E	0,03 d
E+T	0,45 b
T	0,44 b

Médias seguidas por letras iguais na mesma coluna não diferem entre si pelo teste de Scott-Knott a 5% de probabilidade.

A Tabela 9 apresenta o desdobramento da interação significativa para a variável condutividade elétrica entre o modelo x entrada para o EMA observado e predito. Para os inputs C, E, E+T e T os valores de EMA não diferem entre os modelos testados, já para C+E+T e C+T houve menores médias utilizando os modelos AD, M5P e FA.

Tabela 9. Desdobramento da interação significativa entre modelo x entrada para o erro médio absoluto (EMA) entre os valores observados e estimados da condutividade elétrica em sementes de soja por diferentes modelos de aprendizagem de máquina e inputs.

Modelo	C	C+E	C+E+T	C+T	E	E+T	T
RNA	29,91 aA	30,43 aA	25,13 aB	25,24 aB	31,12 aA	30,23 aA	30,35 aA
AD	28,25 aA	28,33 aA	21,67 bB	21,94 bB	29,93 aA	26,61 aA	26,80 aA
RL	28,25 aA	28,37 aA	25,47 aB	25,39 aB	29,95 aA	26,81 aB	26,77 aB
M5P	28,25 aA	1,03 bC	21,60 bB	22,01 bB	29,95 aA	26,78 aA	26,77 aA
FA	28,26 aA	28,34 aA	21,67 bB	21,95 bB	29,94 aA	26,61 aA	26,81 aA

Médias seguidas por letras minúsculas iguais na mesma coluna e maiúsculas iguais na mesma linha não diferem entre si pelo teste de Scott-Knott a 5% de probabilidade.

O modelo RNA para as entradas já mencionadas possui valores de r melhores, porém o EMA não chega a ser estatisticamente diferente do modelo RL.

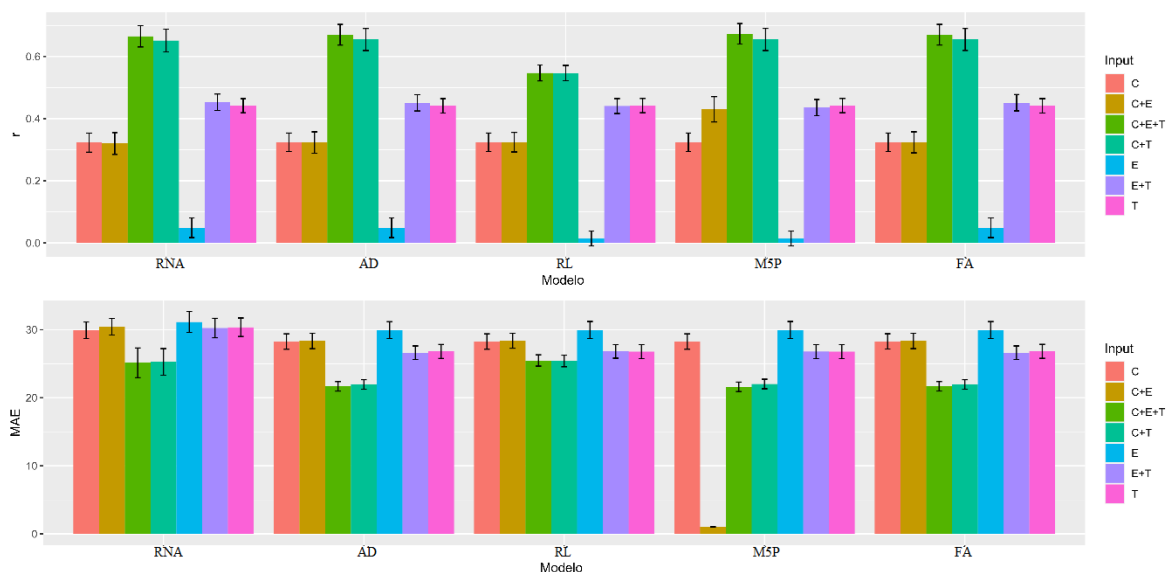


Figura 4. Valores médios e dispersão para as variáveis coeficiente de correlação de Pearson (r) e erro médio absoluto (EMA) entre os valores observados e estimados da condutividade elétrica em sementes de soja por diferentes modelos de aprendizagem de máquina e entradas.

3.5 Germinação

A análise dos resultados dos modelos testados mostra que usando os inputs C, C+E, E, e T não apresenta variação significativa (Tabela 10). As maiores médias ficaram com os inputs C+E+T e C+T utilizando os modelos RNA, AD, M5P e FA, e esta mesma média também é vista com o modelo AD com o input E+T.

Tabela 10. Desdobramento da interação significativa entre modelo x entrada para o coeficiente de correlação de Pearson (r) entre os valores observados e estimados da germinação em sementes de soja por diferentes modelos de aprendizagem de máquina e inputs.

Modelo	C	C+E	C+E+T	C+T	E	E+T	T
RNA	0,33 aB	0,33 aB	0,67 aA	0,65 aA	0,03 aC	0,36 bB	0,35 aB
AD	0,33 aB	0,33 aB	0,67 aA	0,65 aA	0,03 aC	0,65 aA	0,35 aB
RL	0,33 aB	0,33 aB	0,48 bA	0,48 bA	-0,02 aC	0,35 bB	0,35 aB
M5P	0,33 aB	0,32 aB	0,66 aA	0,65 aA	-0,02 aC	0,36 bB	0,35 aB
FA	0,33 aB	0,33 aB	0,67 aA	0,65 aA	0,03 aC	0,36 bB	0,35 aB

Médias seguidas por letras minúsculas iguais na mesma coluna e maiúsculas iguais na mesma linha não diferem entre si pelo teste de Scott-Knott a 5% de probabilidade.

A Tabela 11 apresenta o desdobramento da interação significativa com o modelo x entrada para o EMA, os valores observados e estimados de germinação das sementes

Para os inputs C, C+E, E e T as médias foram consideradas altas, assim como para o input E+T que apenas no modelo AD apresentou uma média comparada as menores. Com os inputs C+E+T e C+T verifica-se um melhor desempenho para o modelo RL, como também para o modelo RNA com C+T de entrada, já para as outras condições são observadas os menores valores de EMA.

Tabela 11. Desdobramento da interação significativa entre modelo x entrada para o erro médio absoluto (EMA) entre os valores observados e estimados da germinação em sementes de soja por diferentes modelos de aprendizagem de máquina e inputs.

Modelo	C	C+E	C+E+T	C+T	E	E+T	T
RNA	13,33 Aa	13,65 bA	9,77 aC	11,61aB	13,70 aA	14,71 aA	12,16 aB
AD	11,75 aA	11,76 bA	8,95 aB	9,10 bB	12,67 aA	9,10 cB	11,89 aA
RL	11,77 aA	11,79 bA	11,26 aA	11,25 aA	12,67 aA	11,89 bA	11,87 aA
M5P	11,77 aA	11,84 bA	9,05 aB	9,21 bB	12,67 aA	11,89 bA	11,87 aA
FA	11,76 aB	17,01 aA	8,95 aC	9,10 bC	12,68aB	11,92 bB	11,89 aB

Médias seguidas por letras minúsculas iguais na mesma coluna e maiúsculas iguais na mesma linha não diferem entre si pelo teste de Scott-Knott a 5% de probabilidade.

A Figura 5 coloca graficamente os resultados das tabelas de germinação, aonde considerando as entradas C+E+T e C+T tiveram bons resultados quando apresentados aos modelos RNA, AD, M5P e FA. Já para a entrada E+T maior média de r e menor EMA são vistos como o modelo AD.

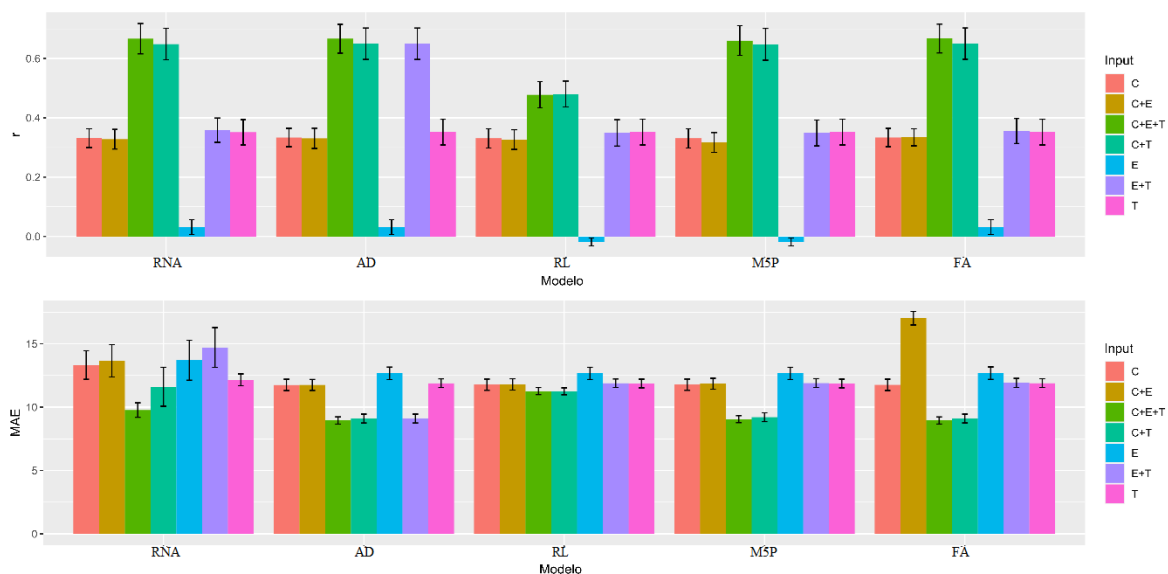


Figura 5. Valores médios e dispersão para as variáveis coeficiente de correlação de Pearson (r) e erro médio absoluto (EMA) entre os valores observados e estimados da germinação em sementes de soja por diferentes modelos de aprendizagem de máquina e entradas.

3.6 Vigor

As estatísticas obtidas para a variável vigor (r e EMA) não apresentaram interação significativa entre modelos e inputs. Ao analisar na Tabela 12 com os modelos testados é possível observar que eles não diferiram entre si para o r. Contudo, o modelo RNA apresentou a maior média de EMA em relação aos demais, o que indica que esse modelo superestima os valores de massa específica.

Tabela 12. Agrupamento de médias para as variáveis coeficiente de correlação de Pearson (r) e erro médio absoluto (EMA) entre os valores observados e estimados do vigor em sementes de soja por diferentes modelos de aprendizagem.

Modelo	r	EMA
RNA	0,44 a	17,33 a
AD	0,44 a	15,46 c
RL	0,39 b	16,20 b
M5P	0,44 a	15,51 c
FA	0,43 a	15,46 c

Médias seguidas por letras iguais na mesma coluna não diferem entre si pelo teste de Scott-Knott a 5% de probabilidade.

Com relação aos inputs testados é possível observar na Tabela 13 que C+E+T, C+T, E+T e T apresentaram as maiores médias de r e as menores médias de EMA.

Tabela 13. Agrupamento de médias para as variáveis coeficiente de correlação de Pearson (r) e erro médio absoluto (EMA) entre os valores observados e estimados do vigor em sementes de soja em função de diferentes inputs.

Input	R	EMA
C	0,34 c	17,23 b
C+E	0,33 c	17,27 b
C+E+T	0,68 a	13,09 d
C+T	0,68 a	13,24 d
E	0,01 d	18,56 a
E+T	0,47 a	16,29 c
T	0,47 a	26,28 c

Médias seguidas por letras iguais na mesma coluna não diferem entre si pelo teste de Scott-Knott a 5% de probabilidade.

A Figura 6 demonstra graficamente os resultados das tabelas anteriores aonde é possível observar que modelos AD, M5P e FA quando associados aos inputs C+E+T, C+T, E+T e T proporcionaram os maiores valores de r e menores valores de EMA. É importante destacar que embora o modelo RNA apresente bons resultados de r com os inputs supracitados, esse modelo apresentou altos valores de EMA para todos os inputs. Ademais, nenhum dos modelos tanto considerando r, como o EMA tiveram resultados diferentes estatisticamente do modelo de RL.

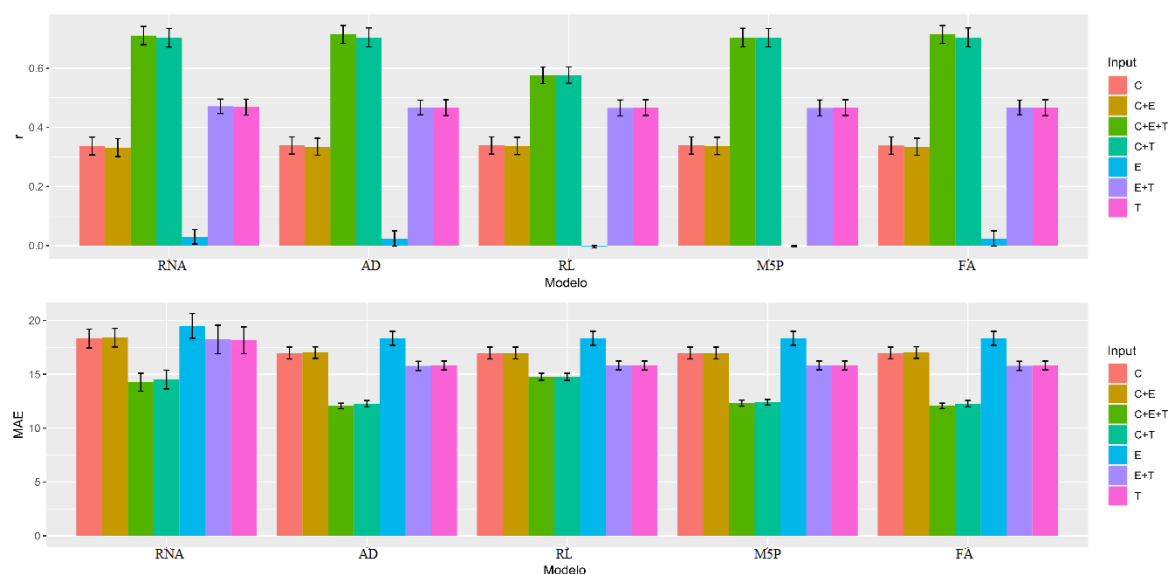


Figura 6. Valores médios e dispersão para as variáveis coeficiente de correlação de Pearson (r) e erro médio absoluto (EMA) entre os valores observados e estimados do vigor em sementes de soja por diferentes modelos de aprendizagem de máquina e entradas.

4. DISCUSSÃO

As sementes são armazenadas em um teor de água ideal, porém os processos biológicos continuam a acontecendo no decorrer do tempo de armazenagem, e processos de respiração são capazes de aumentar o teor de água. Para a estimativa de teor de água as entradas que combinaram C+E+T e a combinação C+T foram as que tiveram bons resultados. A explicação para esta combinação de variáveis pode ser pelo motivo de até atingir o equilíbrio higroscópico as sementes são influenciadas pelo tempo de armazenamento (Carneiro et al., 2005, Ribeiro et al., 2008).

Ademais, as condições de temperatura também colaboram diretamente no teor de água, reduções foram observadas em sementes de soja armazenadas em temperatura ambiente. Juvino et al. (2014), também observaram uma amplitude maior de teores de água em ambientes de temperatura não controlada, em comparação ao climatizado em 18°C. Porém, quando submetidas a temperaturas baixas não ocorre variações tão significativas nos teores de água (Bragantini, 2005). As entradas tiveram comportamentos diferentes entre os modelos de AM e o modelo de regressão linear. No entanto isso vale apenas para as entradas que de fato se mostraram como alternativas viáveis de predição. Na análise do teor de água duas

combinações de entrada obtiveram resultados em que os modelos de AM foram superiores aos vistos na RL. Os resultados de EMA dos algoritmos de AM tendo como *inputs* C+T+E e C+T variaram entre 0,30 – 0,41, enquanto isso a RL obteve 0,73. E uma boa precisão na classificação do teor de água é vista quando analisado o coeficiente de correlação de Person (r). Nesta uma forte correlação para a combinação C+E+T em que todos os modelos de AM tem valores de r acima ou igual a 0,94, já para a regressão o r observado foi de 0,72.

As entradas nos modelos que melhor conseguiram explicar a massa específica das sementes de soja armazenada foram C+T+E, C+T, E+T e T. O r para estes *inputs* foi de 0,63 e o EMA entre 9,59 - 10,47. Estes resultados mostram que o tempo de armazenamento foi a condição presente em todas as combinações de entradas que melhor previu os níveis de ME. Um estudo conduzido por Alencar et al. (2009), verificou que a ME aparente é alterada de acordo com as condições de temperatura e tempo de armazenamento. O estudo ainda observou a diminuição da massa específica aparente a partir de 180 dias pode estar relacionada ao aumento da atividade metabólica dos grãos e aumento de teor de água, portanto dando condições para a manifestação de agentes externos. Para a predição da massa específica embora os modelos não apresentaram diferenças entre si para o coeficiente de Person, o EMA = 17,12 para o modelo RNA indica maior precisão para este tipo de problema.

Para a predição dos dados referentes a condutividade elétrica, tal como visto no teor de água, os *inputs* que melhor representaram a predição foram C+E+T com $r = 0,65$ e $r = 0,63$ para C+T. Os modelos AD, M5P e FA apresentaram os mais baixos EMA (21,67) considerando a utilização das três entradas ao mesmo tempo. Desempenho próximo é visto para as entradas C+T em que o EMA encontra-se entre 21,95 e 22,01. Efeitos significativos sobre a condutividade elétrica é observada por Alencar et al., 2008 em que a interação entre o teor de água, temperatura e tempo de armazenamento provoca alteração nos níveis de condutividade elétrica das sementes.

Carvalho et al. (2016) observaram que o aumento mais expressivo da condutividade de sementes de soja se deu após 180 dias de armazenamento. E considerando que o teor de água é influenciado diretamente pelo tempo e temperatura do ambiente armazenado, é esperado que a predição da variável condutividade elétrica também tenha influência destas condições de entradas. Além disso, a depender da condição de permeabilidade da embalagem e do tempo de armazenamento, as condições são propensas ao aumento da condutividade elétrica em virtude da maior deterioração dos grãos (Carvalho et al., 2016).

Em análise dos dados de correlação de Pearson (r) para a predição da germinação os *inputs* C+E+T e C+T obtiveram valores entre 0,65 e 0,67 para todos os modelos de aprendizado. O modelo de regressão linear para esta variável, em ambas condições de entrada, teve valores abaixo com $r=0,48$. Tanto a temperatura quanto o tempo de armazenamento, acrescidos de teor de água altos provocam nas sementes um aumento da taxa respiratória, o que agrava a deterioração, que por consequência provoca desgastes fisiológicos e diminuem índices de germinação e vigor das sementes (UllmRNA et al., 2016).

A influência da temperatura na germinação é observada por Demito e Afonso (2009), em que sementes de soja resfriadas artificialmente por 140 dias obtiveram porcentagem de germinação superior àquelas que passaram por processos de resfriamento. O desempenho dos modelos se repete quando analisado o EMA. O modelo de RL obteve o valor de EMA (11,26) para a combinação de entrada C+E+T, isso foi pouco acima dos valores vistos nos modelos de AM. Neste caso, tanto o algoritmo de AD e FA apresentaram os menores EMA (8,95), isso tanto para C+E+T, quanto para C+T.

Ferreira et al. (2017) analisaram as consequências do armazenamento em baixas temperaturas nas sementes. Ocorreram reduções no metabolismo e nos ataques de microorganismos patogênicos, e uma das consequência é a manutenção do vigor. Na predição dos valores de vigor os algoritmos de maneira geral não tiveram um bom desempenho. O modelo de regressão alcançou um valor de $r = 0,39$, e o modelos de AM foram pouco melhores, não ultrapassando r igual a 0,44. Seguindo o mesmo padrão, o desempenho dos modelos na predição do vigor não chega a ter diferenças significativas que justifique a escolha de um em detrimento de outro considerando somente o desempenho, embora o modelo RNA obteve EMA = 17,33.

Os demais modelos tiveram uma taxa de erro um pouco menor do que a de 12,20 observada na regressão linear. As entradas C+E+T e C+T apresentaram um coeficiente de Person no igual a 0,68. E as entradas E+T e T, mesmo sem diferença significativa, tiveram um r igual 0,47. No entanto a escolha das duas primeiras combinações justifica-se quando analisado os valores dos erros médio absoluto melhores. O EMA para C+E+T foi de 13,09, e C+T no valor de 13,24.

Neste trabalho, a motivação foi verificar o desempenho de diferentes técnicas de aprendizado de máquina e de combinações de entrada para avaliar a predição de variáveis que representam a qualidade fisiológica das sementess de soja armazenadas. No geral as técnicas de AM superam os resultados do modelo de controle proposto. Porém dentro do grupo dos modelos de AM os algoritmos de Árvore de Decisão e Floresta Aleatória foram que

proporcionaram melhores previsões. A técnica de Rede Neural Artificial apresentou problemas em relação às taxas maiores de erros. A combinação das três variáveis como entrada no modelo foi uma boa combinação, no entanto foi visto que a combinação de temperatura e tempo de armazenamento também foi capaz de representar as condições necessárias para que os modelos pudessem executar as saídas de previsões de qualidade fisiológica.

A embalagem utilizada como variável de entrada nos modelos de AM não teve grande influência na previsão das condições de qualidade fisiológica das sementes de soja para este trabalho. A justificativa para isso pode ser explicada pelo fato de que embora as embalagens sejam responsáveis por também contribuir na redução do processo de deterioração e manutenção do vigor, o efeito das embalagens fica mais evidente se considerar temperaturas mais altas de armazenamento e a partir de 270 dias (Lima, 2020).

Os resultados da pesquisa colocam em evidência a possibilidade de alcançar bons resultados utilizando apenas duas variáveis de entrada, no caso tempo de armazenamento e temperatura de armazenamento. Para os modelos de AM, há um direcionamento para o uso do algoritmo de floresta aleatória para a previsão das variáveis de qualidade das sementes de soja. A construção desta abordagem permite que os resultados conseguidos sejam menos suscetíveis aos problemas de overfitting que os outros algoritmos estão sujeitos, e assim adequando-se melhor em conjuntos com menor quantidade de dados.

5. CONCLUSÃO

Para a previsão dos índices de qualidade das sementes de soja durante o período de armazenamento a combinação das variáveis de entrada temperatura de armazenamento e tempo de armazenamento foi a que melhor conseguiu prever as variações. Os resultados dessa pesquisa indicam que o algoritmo floresta aleatória foi aquele que melhor previu os índices de qualidade fisiológica das sementes durante o período de armazenamento. A abordagem proposta se destaca em termos de custo e rapidez quando comparado aos métodos de análises rotineiramente utilizados. Deste modo tornando os processos mais robustos e com baixo custo operacional, quando comparados com estratégias de análises laboratoriais tradicionalmente utilizadas. Esta pesquisa abre caminho para que se possa utilizar tais métodos para a previsão de índices de qualidade de outras sementes. Ainda como consequências da utilização do AM é o evidente auxílio na tomada das melhores decisões dentro de ambiente de armazenamento, e assim colaborar na redução das perdas às quais as sementes estão suscetíveis durante todo o processo de armazenagem.

6 REFERÊNCIAS

ALENCAR, E. R.; FARONI, L. R. D.; LACERDA FILHO, A. F.; FERREIRA, L. F.; MENEGHITTI, M. R. Qualidade dos grãos de soja em função das condições de armazenamento. *Engenharia na Agricultura, Viçosa*, v. 16, n. 2, p. 155-166, 2008.

ALENCAR, E.R.; FARONI, L.R.D.; LACERDA FILHO, A.F.; PETERNELLI, L.A.; COSTA, A.R. Qualidade dos grãos de soja armazenados em diferentes condições. *Revista Brasileira de Engenharia Agrícola e Ambiental*, v. 13, n. 5, p. 606-613, 2009.

ALENCAR, ERNANDES R. DE et al. Quality of soy bean grains stored under different conditions. *Revista Brasileira de Engenharia Agrícola e Ambiental*, v. 13, n.5 pp. 606-613. 2009.

BAOUA I.B.; AMADOU L.; OUSMANE B.; BARIBUTSA D.; MURDOCK L.L.; PICS bags for post-harvest storage of maize grain in West Africa. *Journal of Stored Products Reserch* 58, 20-28, 2014.

BELGIU, M.; DRAGUT, L. Random forest in remote sensing: A review of applications and future directions. *ISPRS Journal of Photogrammetry and Remote Sensing*, v.114, 24–31. 2016.

BLAIFI, S. A., MOULAHOU, S., BENKERCHA, R., TAGHEZOUIT, B., SAIM, A. M5P model tree based fast fuzzy maximum power point tracker. *Solar Energy*, v.163, 405-424. 2018.

BOUCKAERT, R. FRANK, E. HALL, M. KIRKBY, R. REUTEMRNA, P. SEEWALD, A. WEKA manual for version 3.7.1, 2010.

BRAGANTINI, C. Alguns aspectos do armazenamento de sementes e grãos de feijão Santo Antonio do Goiás, Embrapa Arroz e Feijão. 28 p. (Documento, 187), 2005.

BRASIL. Ministério da Agricultura, Pecuária e Abastecimento. Regras para análise de sementes. Ministério da Agricultura, Pecuária e Abastecimento. Secretaria de Defesa Agropecuária. Brasília: MAPA/ACS. 395p., 2009.

CARNEIRO, L.M.T.A.; BIAGI, J.D.; FREITAS, J.G. DE; CARNEIRO, M.C.; FELÍCIO, J.C. Diferentes épocas de colheita, secagem e armazenamento na qualidade de grãos de trigo comum e duro. *Bragantia, Campinas*, v. 64, n.1, p.127-137, 2005.

CARVALHO, E. R.; OLIVEIRA, J. A.; MAVAIEIE, D. P. R.; SILVA. H. W.; LOPES, C. G. M. Pre-packing cooling and types of packages in maintaining physiological quality of soybean seeds during storage. *Journal of Seed Science, Londrina*, v. 38, n. 2, p. 129-139, 2016.

COLES, R.; MCDOWELL, D.; KIRWAN, M.J. Food packaging technology. Boca Raton, Florida, 2003.

CORADI P.C.; LEMES A.F.C.; Experimental prototype of silo-dryer-aerator of grains using Computational Fluid Dynamics (CFD) system. *Acta Scientiarum Technology*, v.41, 2019.

CORADI. P.C.; LEMES, A.F.C.; Experimental silo-dryer-aerator for the storage of soybean grains. *Revista Brasileira de Engenharia Agrícola e Ambiental*, v. 22, 279-285, 2018.

DEMITO, A.; AFONSO, A.D.L. Qualidade das sementes de soja resfriadas artificialmente. *Engenharia na Agricultura*, v.17, n.1, p.7-14, 2009.

EBONE L.A.; CAVERZAN A.; TAGLIARI A.; CHIOMENTO J.L.T.; SILVEIRA D.C.; CHAVARRIA G. SOYBEAN SEED VIGOR: Uniformity and Growth as Key Factors to Improve Yield. *Agronomy*, v.10: p.1–15, 2020.

EGMONT-PETERSEN, M.; DE RIDDER, D.; HANDELS, H. Image processing with neural networks a review. *Pattern Recognition*. v. 35, p. 2279–2301, 2002.

FERREIRA F.C.; VILLELA F.A.; MENEGHELLO G.E.; SOARES V.N. Cooling of soybean seeds and physiological quality during storage. *Journal of Seed Science*, v. 39, p. 385–392, 2017.

FRANÇA-NETO, J. B. et al. Determinação da qualidade fisiológica dos grãos de soja colhidos na safra 2015/16. *Embrapa Soja-Artigo em anais de congresso (ALICE)*. Londrina, Reunião de pesquisa de soja, v. 36, p. 249-251, 2017.

GOYAL, S. Artificial Neural Networks in Fruits: A Comprehensive Review. *International Journal of Image, Graphics and Signal Processing*. v. 6. p. 53-63, 2014.

HUSSAIN, L.; AJAZ, R. Seed Classification using Machine Learning Techniques. *Journal of Multidisciplinary Engineering Science and Technology (JMEST)*. v. 02. p. 1098-1102, 2015.

JEONG, J.H.; RESOP, J.P.; MUELLER, N.D.; FLEISHER, D.H.; YUN, K.; BUTLER, E.E.; et al. Random Forests for Global and Regional Crop Yield Predictions. *PLoS ONE*11(6), e0156571, 2016.

JUVINO, A. N. K. et al. Vigor da cultivar BMX Potência RR de soja durante o beneficiamento e períodos de armazenamento. *Revista Brasileira de Engenharia Agrícola e Ambiental-Agriambi*, Campina Grande, v. 18, n. 8, p.844-850, 2014.

KONG F.; CHANG S.K.; LIU Z.; WILSON L.A., Changes of soybean quality during storage as related to soymilk and tofu making. *Journal of Food Science*, v. 73, p. 134–144, 2008.

KUMAR S.; DESWAL S.; Estimation of phosphorus reduction from wastewater by artificial neural network, random forest and M5P model tree approaches. India. Pollution, v. 6, p. 427–438, 2020.

LIMA, R. E. Ambientes de armazenamento de sementes de cultivares de soja. Universidade Federal de Mato Grosso do Sul, 2020.

LIMA. D.C.; DUTRA, A.S.; CAMILO, J.M.; Physiological quality of sesame seeds during storage. Revista Ciência Agronômica, v. 5, n.1, p.138-145, 2014.

MORETI, M. P.; OLIVEIRA, T.; SARTORI, R.; CAETANO, W. Inteligência artificial no agronegócio e os desafios para a proteção da propriedade intelectual. Cadernos de Prospecção, v. 14, p. 60, 2021.

MYLONA K.; MAGAN N.; Fusarium langsethiae: Storage environment influences dry matter losses and T2 and HT-2 toxin contamination of oats. Journal Stored Products Research, v. 47, p. 321–327. 2011.

MYLONA K.; SULYOK M.; MAGAN N.; Relationship between environmental factors, dry matter loss and mycotoxin levels in stored wheat and maize infected with Fusarium species. Food Additives & Contaminants, v. 29, p. 1118–1128, 2012.

OSCO, L.P.; PAULA, A.; RAMOS, M.; PEREIRA, D.R.; AKEMI, E.; MORIYA, S.; MATSUBARA, E.T.; Predicting canopy nitrogen content in citrus-trees using random forest algorithm associated to spectral vegetation indices from UAV-imagery. Remote Sensing 11 v. 24, p. 2925–2942, 2019.

PAZOKI, A.; PAZOKI, Z.; Classification system for rain fed wheat grain cultivars using artificial neural network. African Journal of Biotechnology. v. 10, 2011.

R CORE TEAM. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. 2018. URL <https://www.R-project.org/>.

RADHIKA, V.; V. RAO. Computational approaches for the classification of seed storage proteins. Journal of Food Science and Technology, v. 52, p. 4246-4255, 2014.

RAMOS, A.P.M.; OSCO, L.P.; FURUYA, D.E.G; GONÇALVES, W.N.; CORDEIRO, D.C.; PEREIRA, L.R.T.; JUNIOR, C.A.S.; SILVA, G.F.C.; LI, J.; BAIO, F.H.R.; JUNIOR, J.M.; TEODORO, P.E.; PISTORI, H. A random forest ranking approach to predict yield in maize with uav-based vegetation spectral indices. Computers and Electronics in Agriculture, 178, 2020.

RIBEIRO, N.D. et al. Períodos de semeadura e condições de armazenamento na qualidade de cozimento de grãos de feijão. Ciência Rural, v. 38, p. 936-941, 2008.

SARATH K.L.L.; GONELI A.L.D.; FILHO C.P.H.; MASETTO T.E.; OBA G.C.; Physiological potential of peanut seeds submitted to drying and storage. *Journal of Seed Science*, v. 38, p. 233–240, 2016.

SILVA, F. S. DA; PORTO, A.G.; PASCUALI, L.C.; SILVA, F.T.C. DA. Viabilidade do armazenamento de sementes em diferentes entes embalagens para pequenas propriedades rurais. *Revista de Ciências Agroambientais*, v.8, p.45-56, 2010

SNOUSY, M.B.A.; EL-DEEB, H.M.; BADRAN, K.; KHLIL, I.A.A. Suite of decision tree-based classification algorithms on cancer geneexpression data. *Egyptian Informatics Journal*. v. 12, p.73–82, 2011.

ULLMRNA, R. et al. Higroscopicidade das sementes de sorgo-sacarino. *Engenharia Agrícola*, Jaboticabal, v.36, n.3, p. 515-524, 2016.

YILDIRIM S. Active packaging for food biopreservation. In: C Lacroix, editor. *Protective cultures, antimicrobial metabolites and bacteriophages for food and beverage biopreservation*. Cambridge, England. Woodhead Publishing Ltd. p 460– 89, 2011.

YILDIRIM, S.; RÖCKER, B.; PETTERSEN, M.K.; NILSEN-NYGAARD, J.; AYHAN, Z.; RUTKAITE, R.; RADUSIN, T.; SUMINSKA, P.; MARCOS, B.; COMA, V. Active Packaging Applications for Food. *Comprehensive Reviews in Food Science and Food Safety*, v. 17, p.165-199, 2018.