

UNIVERSIDADE ESTADUAL PAULISTA - UNESP

CÂMPUS DE JABOTICABAL

**Fenotipagem de Alto Rendimento, *Machine Learning* e
Deep Learning: Aplicação de Métodos de Inteligência
Artificial no Melhoramento Genético de Plantas
Autógamas**

**Jardel da Silva Souza
Mestre em Agronomia**

**UNIVERSIDADE ESTADUAL PAULISTA - UNESP
CÂMPUS DE JABOTICABAL**

**Fenotipagem de Alto Rendimento, *Machine Learning* e
Deep Learning: Aplicação de Métodos de Inteligência
Artificial no Melhoramento Genético de Plantas
Autógamas**

Discente: Jardel da Silva

Orientadora: Profa. Dra. Sandra Helena Unêda-Trevisoli

**Tese apresentada à Faculdade de
Ciências Agrárias e Veterinárias –
Unesp, Câmpus de Jaboticabal, como
parte das exigências para a obtenção do
título de Doutor em Agronomia
(Genética e Melhoramento de Plantas).**

S729f Souza, Jardel da Silva
Fenotipagem de alto rendimento, machine learning e deep learning: aplicação de métodos de inteligência artificial no melhoramento genético de plantas autógamas / Jardel da Silva Souza. -- Jaboticabal, 2025
103 p.

Tese (doutorado) - Universidade Estadual Paulista (UNESP), Faculdade de Ciências Agrárias e Veterinárias, Jaboticabal
Orientadora: Sandra Helena Unêda-Treviseli

1. Fenotipagem de Alto-Rendimento. 2. Inteligência Artificial. 3. Arroz. 4. Soja. 5. Redes Neurais. I. Título.

CERTIFICADO DE APROVAÇÃO

TÍTULO DA TESE: FENOTIPAGEM DE ALTO RENDIMENTO, *Machine Learning* E *Deep Learning*: APLICAÇÃO DE MÉTODOS DE INTELIGÊNCIA ARTIFICIAL NO MELHORAMENTO GENÉTICO DE PLANTAS AUTÓGAMAS

AUTOR: JARDEL DA SILVA SOUZA

ORIENTADORA: SANDRA HELENA UNÊDA TREVISOLI

Aprovado como parte das exigências para obtenção do Título de Doutor em Agronomia (Genética e Melhoramento de Plantas), pela Comissão Examinadora:

Documento assinado digitalmente
 **SANDRA HELENA UNEDA TREVISOLI**
Data: 21/03/2025 17:18:00-0300
Verifique em <https://validar.iti.gov.br>

Profa. Dra. SANDRA HELENA UNÊDA TREVISOLI (Participação Presencial)
Departamento de Producao Vegetal / FCAV UNESP Jaboticabal

Documento assinado digitalmente
 **CLAUDENIR FACINCANI FRANCO**
Data: 17/03/2025 09:13:45-0300
Verifique em <https://validar.iti.gov.br>

Prof. Dr. CLAUDENIR FACINCANI FRANCO (Participação Virtual)
Departamento Biocombustíveis / FATEC

Documento assinado digitalmente
 **PABLO FORLAN VARGAS**
Data: 16/03/2025 10:25:18-0300
Verifique em <https://validar.iti.gov.br>

Prof. Dr. PABLO FORLAN VARGAS (Participação Presencial)
Agronomia / FCAVR UNESP Registro

Documento assinado digitalmente
 **LIGIA DE OLIVEIRA AMARAL**
Data: 20/03/2025 17:17:16-0300
Verifique em <https://validar.iti.gov.br>

Dra. LÍGIA DE OLIVEIRA AMARAL (Participação Virtual)
BASF S/A / Porto Nacional/TO



UNIVERSIDADE ESTADUAL PAULISTA

Câmpus de Jaboticabal



Documento assinado digitalmente

ROUVERSON PEREIRA DA SILVA

Data: 15/03/2025 14:55:42-0300

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

Prof. Dr. ROUVERSON PEREIRA DA SILVA (Participação Presencial)
Departamento de Engenharia / FCAV UNESP Jaboticabal

Jaboticabal, 13 de março de 2025

DADOS CURRICULARES DO AUTOR

JARDEL DA SILVA SOUZA nasceu em 2 de setembro de 1994, na cidade de Colônia Leopoldina, Alagoas, é Engenheiro Agrônomo, formado pela Universidade Federal da Paraíba (UFPB) em 2018. Durante a graduação, foi bolsista de iniciação científica de 2015 a 2016, no Departamento de Fitotecnia e Ciências Ambientais, e de 2016 a 2017, no Departamento de Solos e Engenharia Rural. Além disso, atuou como monitor da disciplina de Química Orgânica 2014 a 2015 e da disciplina de Melhoramento de Plantas (2017). Em 2017, foi premiado com o Prêmio Jovem Cientista da UFPB e com o Prêmio de Iniciação à Docência pela monitoria na disciplina de Melhoramento de Plantas. Em março de 2018, ingressou no curso de Mestrado em Agronomia pela UFPB, na área de concentração em Biotecnologia e Melhoramento Genético de Plantas. Durante o mestrado, foi bolsista CAPES sob a orientação da Profa. Dra. Elizanilda Ramalho do Rêgo, obtendo o título de Mestre em fevereiro de 2020. Em março de 2021, iniciou o Doutorado em Agronomia (Genética e Melhoramento de Plantas) na Universidade Estadual Paulista “Júlio de Mesquita Filho” (UNESP), Faculdade de Ciências Agrárias e Veterinárias, Câmpus de Jaboticabal, sob a orientação da Profa. Dra. Sandra Helena Unêda-Trevisoli, como bolsista CAPES. Em 2024, realizou um período de treinamento como doutorado sanduíche na Louisiana State University, nos Estados Unidos, sob a orientação do Prof. Dr. Roberto Fritsche-Neto.

Dedico esta tese à memória dos meus pais, Maurício e Maria Marlene, e do meu tio Amauri. Eles me viram crescer, sonhar e percorrer este caminho, mas não tiveram a chance de me ver alcançar o título de Doutor.

AGRADECIMENTOS

A Universidade Estadual Paulista, em especial, ao programa de Pós-Graduação em Agronomia/Genética e Melhoramento de Plantas pela oportunidade de realização do curso, e à CAPES pela concessão da bolsa de estudos.

À minha orientadora na Universidade Estadual Paulista, Professora Dra. Sandra Helena Unêda-Trevisoli, pela amizade, conselhos, ensinamentos e orientações nos trabalhos acadêmicos.

Ao Professor Dr. Roberto Fritsche-Neto, pela amizade, conselhos, ensinamentos e orientações, na Louisiana State University.

Aos colegas do Laboratório Biotecnologia e Melhoramento da UNESP/FCAV Jaboticabal: Alice, Alyce, Ana Beatriz, Ana Clara, Claudenir, Dardânia, Davi, Ery Bonifácio, Flávio, Gabriela Cordeiro, Gustavo Perle, Hortência, José Augusto, Laura Rodrigues, Larissa, Lucas Caruzo, Lucas Zovadelli, Márcio Mandrá, Matheus, Nattan, Nayara, Thainá Moura, Thyaná Garcia. E aos funcionários da FEPE: Colovate, Canoa, Ferrari, Fran e Sr. João.

Aos colegas da Louisiana State University - AgCenter: Adam Famoso, Alisson, Anderson, Felipe Dalla Lana, Gustavo, Jomar Punzalan, Kajal, Kashish, Melina, Valerie, Mr. Dean, Miss Donna.

Aos colegas da República TOca fogo e agregados: Armando, Breno, Eduardo, Fernanda, Francisco, Igor, Jarlyson, Jefferson, Lorena, Luís Alberto, Marcelo, Samira, Hilário, Theodora e Prof. Rouverson.

Aos amigos: Adjair, Eduardo Felipe, Fabiano, Flaviano, Gemerson, Ingrid, Isadora Gouveia, Isadora Le Campion, Jackson, Julia Mercês, Kaline, Karla, Lais, Maiara Marques, Maria de Fátima, Matheus Ayres, Michelle, Nardiele, Raylson (*in memoriam*), Renata, Robério e Thalia.

À minha família: meus pais (*in memoriam*) Maurício e Maria Marlene, meus avós Antonio e Creuza, minha tia Rizomar e meus tios: Amauri (*in memoriam*), Joaquim, José, Renildo. A todos pelas palavras de motivação,

amizade e pelos momentos de descontração ao longo dessa caminhada em busca de crescimento pessoal e profissional. Muito obrigado!

O presente trabalho foi realizado com apoio da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Código de Financiamento 001.

Impacto Esperado da Tese

Esta tese contribui para o avanço da agricultura digital ao integrar inteligência artificial e fenotipagem de alto rendimento, acelerando o desenvolvimento de cultivares mais produtivas e resilientes às mudanças climáticas.

Sumário

RESUMO	vii
Capítulo 1 – Considerações Gerais.....	9
1. Melhoramento Genético e Inteligência Artificial (IA).....	9
1.1. Limitações dos métodos convencionais e a necessidade de novas tecnologias.....	10
1.2. O papel da Inteligência Artificial no avanço da agricultura digital	11
2. Fenotipagem de Alto Rendimento (HTP) e sua Importância no Melhoramento Genético	12
2.1. Fenotipagem de Alto Rendimento e Inteligência Artificial: como o <i>Machine Learning</i> e <i>Deep Learning</i> auxiliam na análise automática de imagens e sensores para a caracterização fenotípica	13
3. Aplicações do <i>Machine Learning</i> e <i>Deep Learning</i> no Melhoramento Genético e na Agricultura.....	16
4. Modelos Híbridos de <i>Machine Learning</i> e <i>Deep Learning</i> no Melhoramento de Plantas.....	19
5. Espécies autógamias	21
5.1. Cultura da Soja	21
5.2. Cultura do Arroz	22
6. Desafios e Perspectivas Futuras	23
Referências Bibliográficas	26
Capítulo 2: Otimização da Avaliação da Qualidade de Grãos de Arroz Por Meio de Fenotipagem de Alto Rendimento.....	43
Introdução.....	44
Material e Métodos.....	46
Utilização de Software e Processamento de Imagens.....	46
Leitura e Processamento de Imagens	47
Deteção de Contornos e Análise Morfológica.....	47
Análise Quantitativa.....	48
Análise de Grãos Gessados.....	48
Validação	49
Resultados e Discussão	49
Conclusão.....	55

Referências.....	55
Capítulo 3: Uso Integrado de <i>Machine Learning</i> e <i>Deep Learning</i> no Melhoramento Genético de Soja: Classificação de Genótipos e Estimativa de Valor Agrônomico	61
Resumo	61
Introdução.....	62
Material e Métodos.....	64
Caracteres Avaliados	64
Avaliação do teor de óleo e proteína	65
Base de Dados para Random Forest (<i>Machine Learning</i>)	66
Seleção e Pré-Processamento dos Dados	66
Treinamento e Validação do Modelo.....	67
Ambiente de Desenvolvimento	68
Base de Dados para <i>Deep Learning</i>	69
Divisão dos Dados	70
Construção do Modelo de Rede Neural.....	71
Resultados e Discussão	74
<i>Machine Learning</i> na classificação dos genótipos de soja	74
<i>Deep Learning</i> na estimativa do Valor Agrônomico em soja	80
Referências Bibliográficas.....	95

Lista de Figuras

Capítulo I:

- Figura 1.** Processamento Das Imagens Da Câmera Hiperespectral.....47
- Figura 2.** Processamento das imagens da câmera RGB. 48
- Figura 3.** correlações entre dados do *seedcount* e os dados da câmera hiperespectral. (a) correlação entre gessamento antigo e gessamento novo (92%); (b) correlação entre largura antiga e largura nova (68%); (c) correlação entre comprimento antigo e comprimento novo (89%). 51
- Figura 4.** Correlações entre dados do *Seedcount* e câmera RGB. (A) correlação entre gessamento antigo e gessamento novo (68%); (B) correlação entre largura antiga e largura nova (83%); (C) correlação entre comprimento antigo e comprimento novo (89%)......51
- Figura 1.** Representação gráfica da arquitetura da rede neural artificial, evidenciando as camadas de entrada, ocultas e de saída. A camada de entrada possui 6 neurônios, correspondendo às variáveis preditoras do modelo. A rede conta com duas camadas ocultas, contendo 64 e 32 neurônios, respectivamente, responsáveis pelo aprendizado das representações dos dados. A camada de saída possui 1 neurônio, indicando que o modelo realiza uma tarefa de regressão ou classificação binária. 72
- Figura 2.** Agrupamento de genótipos por classificação de desempenho, com destaque para os genótipos classificados como Excelentes (25 amostras). 75
- Figura 3.** Comparação das curvas de perda de treinamento e perda de validação, A: safra 2021/2022; B: safra 2022/2023; C: Média dos dois anos. ..80
- Figura 4.** Contribuição relativa das características fenotípicas utilizadas no modelo de classificação para a diferenciação dos genótipos de soja. Jaboticabal, SP. 82
- Figura 5.** Correlação entre valores agronômicos reais e valores criados pelo algoritmo para os dados avaliados no ano agrícola 2021/2022. Jaboticabal, SP.85
- Figura 6.** Correlação entre valores agronômicos reais e valores criados por algoritmo para os dados avaliados no ano agrícola 2022/2023. Jaboticabal, SP.86

Figura 7. Correlação entre valores agronômicos reais e valores criados pelo algoritmo para os dados médios dos dois anos agrícolas (2021/2022 e 2022/2023). Jaboticabal, SP. 87

Capítulo II:

Figura 1. Processamento das imagens da câmera hiperespectral.....47

Figura 2. Processamento das imagens da câmera RGB. 48

Figura 3. correlações entre dados do *seedcount* e os dados da câmera hiperespectral. (a) correlação entre gessamento antigo e gessamento novo (92%); (b) correlação entre largura antiga e largura nova (68%); (c) correlação entre comprimento antigo e comprimento novo (89%). 51

Figura 4. Correlações entre dados do *Seedcount* e câmera RGB. (A) correlação entre gessamento antigo e gessamento novo (68%); (B) correlação entre largura antiga e largura nova (83%); (C) correlação entre comprimento antigo e comprimento novo (89%)......51

Figura 1. Representação gráfica da arquitetura da rede neural artificial, evidenciando as camadas de entrada, ocultas e de saída. A camada de entrada possui 6 neurônios, correspondendo às variáveis preditoras do modelo. A rede conta com duas camadas ocultas, contendo 64 e 32 neurônios, respectivamente, responsáveis pelo aprendizado das representações dos dados. A camada de saída possui 1 neurônio, indicando que o modelo realiza uma tarefa de regressão ou classificação binária. 72

Figura 2. Agrupamento de genótipos por classificação de desempenho, com destaque para os genótipos classificados como Excelentes (25 amostras). 75

Figura 3. Comparação das curvas de perda de treinamento e perda de validação, A: safra 2021/2022; B: safra 2022/2023; C: Média dos dois anos. ..80

Figura 4. Contribuição relativa das características fenotípicas utilizadas no modelo de classificação para a diferenciação dos genótipos de soja. Jaboticabal, SP. 82

Figura 5. Correlação entre valores agronômicos reais e valores criados pelo algoritmo para os dados avaliados no ano agrícola 2021/2022. Jaboticabal, SP.85

Figura 6. Correlação entre valores agronômicos reais e valores criados por algoritmo para os dados avaliados no ano agrícola 2022/2023. Jaboticabal, SP.86

Figura 7. Correlação entre valores agronômicos reais e valores criados pelo algoritmo para os dados médios dos dois anos agrícolas (2021/2022 e 2022/2023). Jaboticabal, SP. 87

Lista de Tabelas

Tabela 1. Métricas de Avaliação do Desempenho do Modelo de Classificação para os genótipos de soja. Jaboticabal, SP. 75

Tabela 2. A matriz de confusão, desempenho do modelo na classificação dos genótipos de soja em três categorias: "Excelentes", "Intermediários" e "Menos Promissores".....78

Tabela 3. Estimativas de Herdabilidade (h^2) para caracteres agronômicos em genótipos de soja, em três anos agrícolas (2021/2022, 2022/2023) e a média dos valores combinados dos dois anos agrícolas. Jaboticabal, SP. 89

Tabela 4. Estatísticas descritivas das características fenotípicas avaliadas em três anos agrícolas (2021/2022, 2022/2023) e na média geral do período. 89

Fenotipagem de Alto Rendimento, *Machine Learning* e *Deep Learning*: Aplicação de Métodos de Inteligência Artificial no Melhoramento Genético de Plantas Autógamas

RESUMO – A aplicação de tecnologias emergentes, como fenotipagem de alto rendimento associados à inteligência artificial, tem revolucionado o melhoramento genético, tornando os processos de seleção mais ágeis, precisos e acessíveis. Neste estudo, foram investigadas duas abordagens inovadoras: a fenotipagem de alto rendimento para avaliação da qualidade dos grãos de arroz, além do uso de *Machine Learning* para classificação de genótipos e *Deep Learning* para a estimar do valor agrônomo em soja, sendo, portanto, explorados dois métodos de inteligência artificial para esta cultura. A fenotipagem de alto rendimento mostrou-se uma ferramenta eficaz para a análise da qualidade dos grãos de arroz, substituindo métodos tradicionais demorados e manuais. Utilizando imagens RGB e hiperespectrais, foi possível quantificar a presença de grãos gessados com alta correlação em relação aos dados obtidos pelo método tradicional *seedcount*. A técnica baseada em câmeras RGB apresentou uma precisão de 92%, reduzindo drasticamente o tempo necessário para análise, de três meses para apenas nove dias. Além da eficiência, essa abordagem se mostrou economicamente viável, garantindo resultados confiáveis com um custo muito inferior ao de equipamentos especializados, enquanto um custa em torno de 2 mil dólares o outro custa 40 mil dólares. Por outro lado, no melhoramento genético da soja, a inteligência artificial foi aplicada para otimizar a seleção de genótipos. Algoritmos de *ML* foram empregados para classificar genótipos em grupos de desempenho agrônomo, atingindo uma acurácia de 95%. Paralelamente, modelos de *DL* foram utilizados para prever o valor agrônomo, alcançando um coeficiente de determinação (R^2) de até 86,35%, demonstrando grande precisão na estimativa de características fenotípicas. Essa abordagem integrada possibilitou uma avaliação mais robusta dos genótipos, reduzindo o viés humano e aumentando a eficiência da seleção. Os resultados deste estudo evidenciam o potencial da automação na agricultura, permitindo maior precisão e velocidade nos programas de melhoramento genético. A combinação de fenotipagem de alto rendimento com a inteligência artificial não apenas otimiza o processo de

seleção de cultivares superiores, mas também contribui para a adaptação das culturas às novas demandas climáticas e produtivas. O avanço dessas tecnologias representa um passo essencial para uma agricultura mais sustentável e eficiente, garantindo maior produtividade e qualidade na produção agrícola.

Palavras-chave: Algoritmos de Redes Neurais, Arroz, Classificação de Genótipos, Estimativa Agronômica, Fenotipagem de Alto Rendimento, Inteligência Artificial, *Machine Learning* (ML), Melhoramento Genético, Soja.

High-Throughput Phenotyping, *Machine Learning*, and *Deep Learning*: Application of Artificial Intelligence Methods in the Genetic Improvement of Self-Pollinating Plants

ABSTRACT – The application of emerging technologies, such as high-throughput phenotyping combined with artificial intelligence, has revolutionized plant breeding by making selection processes faster, more accurate, and more accessible. This study investigated two innovative approaches: high-throughput phenotyping for evaluating rice grain quality, and the use of Machine Learning for genotype classification and Deep Learning to estimate agronomic value in soybean—thus exploring two artificial intelligence strategies for this crop. High-throughput phenotyping proved to be an effective tool for analyzing rice grain quality, replacing time-consuming and manual traditional methods. Using RGB and hyperspectral imaging, it was possible to quantify the presence of chalky grains with a high correlation to the data obtained through the traditional SeedCount method. The RGB-based technique achieved 92% accuracy, drastically reducing the analysis time from three months to just nine days. In addition to its efficiency, this approach proved to be economically viable, providing reliable results at a much lower cost than specialized equipment, with one costing around 2,000 USD compared to 40,000 USD. In soybean breeding, artificial intelligence was applied to optimize genotype selection. ML algorithms were used to classify genotypes based on agronomic performance, reaching 95% accuracy. Concurrently, DL models were employed to predict agronomic value, achieving a coefficient of determination (R^2) of up to 86.35%, demonstrating high precision in estimating phenotypic traits. This integrated approach enabled a more robust evaluation of genotypes, reducing human bias and increasing selection efficiency. The results of this study highlight the potential of automation in agriculture, enabling greater precision and speed in plant breeding programs. The combination of high-throughput phenotyping with artificial intelligence not only optimizes the selection of superior cultivars but also contributes to adapting crops to new climatic and productive demands. The advancement of these technologies represents a key step toward more sustainable and efficient agriculture, ensuring higher productivity and quality in food production.

Keywords: Agronomic Prediction, Artificial Intelligence, Genotype Classification, Genetic Improvement, High-throughput Phenotyping, Machine Learning, Neural Networks, Rice, Soybean.

Capítulo 1 – Considerações Gerais

1. Melhoramento Genético e Inteligência Artificial (IA)

O melhoramento genético de plantas constitui um pilar fundamental da agricultura moderna, voltado ao desenvolvimento de cultivares superiores, com maior produtividade, estabilidade fenotípica e resiliência frente a estresses bióticos e abióticos. De acordo com Borém et al. (2017) e Aswini et al. (2023), essa área do conhecimento utiliza as leis da genética, estatística, biotecnologia e fisiologia vegetal, além de outros campos correlacionados, para modificar geneticamente as plantas e atender às exigências da agricultura moderna. Os métodos empregados no melhoramento genético dependem das metas estabelecidas pelos programas de pesquisa e das características específicas de cada espécie vegetal.

As principais abordagens do melhoramento genético de plantas incluem diversas técnicas que visam aprimorar características desejáveis, como produtividade, resistência a doenças e qualidade nutricional, como os alimentos biofortificados por exemplo. A seleção artificial consiste na escolha de indivíduos com características vantajosas para reprodução, garantindo a transmissão dessas características para as futuras gerações (Borém et al., 2017). A hibridação, envolve o cruzamento de genótipos distintos, permitindo a combinação de atributos favoráveis em uma única cultivar (Novello et al., 2023).

Essas abordagens desempenham um papel fundamental no desenvolvimento de cultivares mais adaptadas às necessidades da agricultura moderna, contribuindo para a sustentabilidade, segurança alimentar e produtividade global (Makrushina & Klitsenko, 2022). O melhoramento genético de plantas tem sido essencial para a segurança alimentar e o desenvolvimento sustentável da agricultura, permitindo, o aumento da produtividade agrícola. Cultivares mais eficientes no uso de recursos garantem maior produção sem a necessidade de expandir as áreas de cultivo, reduzindo impactos ambientais (Esmaeili et al., 2022). Além disso, o desenvolvimento de cultivars resistentes à

pragas e doenças minimiza o uso de defensivos agrícolas, tornando o cultivo mais sustentável (Aziz et al., 2022).

O melhoramento genético também proporciona a adaptação das plantas às condições climáticas extremas, permitindo a produção em regiões afetadas por seca, salinidade e variações de temperatura (Souza et al., 2023). A melhoria nutricional também se destaca como uma das grandes conquistas dessa ciência, possibilitando a obtenção de cultivares biofortificadas, como o arroz dourado, que tem alto potencial na redução da desnutrição mundial (Novello et al., 2023). Apesar dos avanços, o melhoramento genético de plantas ainda enfrenta desafios, como a falta de aceitação social de culturas geneticamente modificadas, regulamentações rigorosas para a liberação de novas cultivares e as mudanças climáticas, que afetam a variabilidade genética existente nas espécies vegetais (Begna, 2022). Diante desse cenário, a integração entre o melhoramento convencional associado à diversas técnicas modernas, podem garantir a produção sustentável de alimentos para atender à crescente demanda global gerada pelo aumento populacional (Saxena, 2022; Hernández et al., 2024).

1.1. Limitações dos métodos convencionais e a necessidade de novas tecnologias

Os métodos convencionais de melhoramento genético, apesar de sua relevância histórica e ampla utilização, enfrentam limitações significativas frente às exigências atuais da agricultura. Entre os principais entraves estão o longo ciclo de desenvolvimento de novas cultivares, que exige várias gerações de cruzamentos e seleções até a obtenção de genótipos estáveis e superiores (Krishna et al., 2023; Anand et al., 2023), e a limitação da variabilidade genética dentro de muitas espécies cultivadas, o que restringe as possibilidades de progresso (Mahajan & Kapoor, 2023).

Além disso, características agronômicas desejáveis, como resistência a pragas, tolerância a estresses ambientais e alto valor nutricional, nem sempre estão disponíveis no pool genético da cultura-alvo, dificultando sua incorporação via cruzamentos tradicionais (Dutta et al., 2022; Babar et al., 2022). A natureza

aleatória da recombinação genética também compromete a previsibilidade dos resultados, reduzindo a eficiência dos programas de melhoramento (Esmaeili et al., 2022).

Diante dessas limitações, novas abordagens têm sido incorporadas aos programas de melhoramento, como a seleção assistida por marcadores moleculares e os avanços em genômica e bioinformática, que permitem maior precisão na identificação de genes de interesse e aceleram o desenvolvimento de cultivares superiores (Saxena, 2022; Noru et al., 2024).

A combinação entre métodos convencionais e tecnologias modernas, como ferramentas moleculares, edição gênica e inteligência artificial, se apresenta como estratégia promissora para superar desafios relacionados à produtividade, sustentabilidade e segurança alimentar em um cenário de mudanças climáticas e crescente demanda global (Gonal et al., 2023; Mudhalvan et al., 2023).

1.2. O papel da Inteligência Artificial no avanço da agricultura digital

A inteligência artificial (IA) tem sido uma peça-chave na agricultura digital, impulsionando avanços significativos na produtividade dos sistemas agrícolas e contribuindo para um desenvolvimento mais sustentável do setor (Akintuyi, 2024). A aplicação de tecnologias baseadas em IA à agricultura tradicional tem possibilitado a automação de processos, a análise preditiva, o gerenciamento otimizado e a tomada de decisões mais inteligentes na condução das atividades agrícolas (Han & Zeng, 2023).

Um dos campos mais impactados pela IA na agricultura digital é a agricultura de precisão, que utiliza sensores, drones, satélites e algoritmos para monitorar em tempo real variáveis como umidade do solo, nutrição das plantas e presença de pragas ou doenças nas lavouras (Shaikh et al., 2022). Esses dados são processados por sistemas de IA permitindo que os agricultores tomem decisões mais assertivas sobre irrigação, aplicação de fertilizantes e defensivos agrícolas, reduzindo desperdícios e aumentando a eficiência no uso dos insumos (Rozhkova et al., 2022). Além disso, a IA desempenha um papel fundamental na

previsão climática, ajudando os agricultores a se adaptarem à variabilidade do clima por meio da previsão de secas, chuvas intensas e outros fenômenos extremos que podem impactar a produção agrícola (Berrahal, 2024).

Outra área revolucionada pela IA é a automação de máquinas e a robótica agrícola. Tratores autônomos, colhedoras inteligentes e robôs para controle de ervas daninhas utilizam algoritmos avançados para operar com precisão, minimizando impactos ambientais e reduzindo a necessidade de mão de obra (Yang et al., 2023). Essas tecnologias não apenas aumentam a produtividade, mas também tornam a agricultura mais sustentável, reduzindo o uso excessivo de produtos químicos e otimizando a utilização da terra (Patel & Patil, 2022).

Além disso, a IA tem desempenhado um papel essencial no melhoramento genético de plantas, acelerando o desenvolvimento de novas cultivares mais resistentes e produtivas (Singh et al., 2024). Algoritmos de *ML* analisam grandes volumes de dados genéticos e fenotípicos para identificar genes de interesse, otimizando a seleção de cultivares com características desejáveis, como resistência a pragas e tolerância à seca (Golubin & Kopnin, 2024). Isso torna o processo de melhoramento genético mais rápido e eficiente, permitindo atender à crescente demanda global por alimentos de maneira sustentável (Chouhan et al., 2023).

A inteligência artificial, portanto, está transformando a agricultura digital ao proporcionar maior precisão, eficiência e sustentabilidade em todas as etapas da produção agrícola. Com o avanço contínuo dessas tecnologias, espera-se que a IA desempenhe um papel ainda mais estratégico na segurança alimentar e na adaptação da agricultura às mudanças climáticas (Singh et al., 2024).

2. Fenotipagem de Alto Rendimento (HTP) e sua Importância no Melhoramento Genético

A Fenotipagem de Alto Rendimento ou High-Throughput Phenotyping (HTP) tem se tornado uma ferramenta essencial no melhoramento genético de plantas, permitindo a coleta, processamento e análise de grandes volumes de dados fenotípicos de maneira automatizada e eficiente. Esse avanço tecnológico

possibilita a avaliação de características agronômicas com maior precisão e rapidez, utilizando sensores, drones, espectrômetros e inteligência artificial para capturar informações fenotípicas que antes eram limitadas por métodos convencionais de observação manual (Crain et al., 2022; Bhandari et al., 2023).

A HTP tem sido amplamente aplicada no melhoramento de culturas como milho, trigo, arroz e soja, auxiliando na identificação de linhagens mais produtivas e resilientes. A integração da HTP com dados genômicos tem mostrado alta eficácia na previsão de características complexas, como resistência a doenças e estresses abióticos, além de contribuir para a seleção assistida por genômica (Crossa et al., 2023; Melsen et al., 2024; Barbosa Júnior et al., 2022).

Um dos grandes benefícios da HTP é a padronização da coleta de dados, que reduz a interferência de fatores ambientais na avaliação fenotípica e aumenta a precisão da seleção (Galli et al., 2020; Fritsche-Neto et al., 2020; Chiozza et al., 2024). Apesar do custo elevado de equipamentos e da necessidade de infraestrutura tecnológica, avanços em bioinformática e IA têm ampliado a acessibilidade e funcionalidade da HTP, promovendo sua integração a sistemas agrícolas inteligentes e auxiliando a tomada de decisão em tempo real (Leiva-Sandoval, 2023).

Portanto, a HTP representa uma inovação indispensável para o melhoramento genético moderno, permitindo ganhos significativos em produtividade e sustentabilidade agrícola. Seu uso contínuo e aprimoramento serão fundamentais para enfrentar os desafios da agricultura frente às mudanças climáticas e à crescente demanda por alimentos.

2.1. Fenotipagem de Alto Rendimento e Inteligência Artificial: como o *Machine Learning* e *Deep Learning* auxiliam na análise automática de imagens e sensores para a caracterização fenotípica

Com a crescente necessidade de otimização dos processos agrícolas, a aplicação de Inteligência Artificial (IA), especialmente *Machine Learning* (ML) e *Deep Learning* (DL), tem desempenhado um papel fundamental na

automatização e análise de imagens e sensores para caracterização fenotípica. A integração dessas tecnologias na HTP possibilita a extração automática de características fenotípicas a partir de imagens capturadas por drones, câmeras multiespectrais e sensores hiperespectrais, melhorando significativamente a precisão e eficiência na coleta de dados fenotípicos, reduzindo a necessidade de intervenção humana e minimizando erros observacionais (Singhvi et al., 2024; Murphy et al., 2024). Métodos de *DL*, como redes neurais convolucionais (CNNs), têm sido amplamente utilizados para segmentação de folhas, identificação de estresses bióticos e abióticos e estimativa da produtividade das culturas, sendo capazes de superar técnicas tradicionais como Support Vector Machines (SVMs) e Random Forest na segmentação precisa de órgãos vegetais (Zenkl et al., 2022; Wang et al., 2023).

A integração da HTP com novas tecnologias, como *ML* e Inteligência Artificial, tem sido fundamental para analisar grandes volumes de dados fenotípicos e identificar padrões difíceis de detectar manualmente. A utilização de imagens espectrais combinadas com modelos preditivos tem demonstrado aumentar a acurácia da fenotipagem, permitindo que melhoristas tomem decisões mais assertivas na seleção de genótipos (Tayade et al., 2022; Kaushal et al., 2024). Além disso, Barbosa Júnior et al. (2023) destacam que o uso de redes neurais e algoritmos avançados tem permitido uma melhor correlação entre dados fenotípicos e genotípicos, aumentando a eficiência da estimativa de características agronômicas.

A segmentação automática e a extração de características fenotípicas representam um dos maiores desafios na HTP, pois a variabilidade ambiental pode influenciar a qualidade das imagens e dificultar a análise. No entanto, algoritmos de *DL* têm demonstrado elevada eficácia na separação entre plantas e fundos complexos, permitindo a análise detalhada da morfologia das plantas. Estudos recentes indicam que redes neurais profundas aplicadas na fenotipagem de trigo conseguem segmentar folhas com alta acurácia, permitindo a quantificação de características estruturais importantes para o melhoramento genético (Misra et al., 2022). Além disso, a combinação de imagens espectrais com *DL* tem sido utilizada para prever estresses hídricos e doenças nas culturas agrícolas. Modelos baseados em CNNs aplicados em dados hiperespectrais

foram capazes de identificar sinais precoces de deficiência hídrica e de infecções por patógenos, melhorando a capacidade de resposta dos produtores às adversidades ambientais (Yu et al., 2022).

Outro avanço significativo da aplicação de IA na HTP é a sua integração com dados genômicos, permitindo a seleção assistida por características fenotípicas e moleculares. O uso de modelos de *ML* para análise genômica tem permitido a identificação de marcadores associados a características agronômicas complexas, facilitando a seleção de genótipos superiores em programas de melhoramento. Estudos demonstraram que redes neurais profundas podem aumentar a precisão da estimativa da produtividade de grãos em trigo quando comparadas a métodos tradicionais de estimativa genômica (Montesinos-López et al., 2024). Isso evidencia a superioridade dessas abordagens para prever características de interesse agrícola. Além disso, abordagens baseadas em *DL* têm sido utilizadas para identificação de variações genéticas associadas à tolerância a estresses ambientais, permitindo o desenvolvimento de cultivares mais resilientes às mudanças climáticas.

Apesar dos avanços, a aplicação de *DL* na fenotipagem ainda enfrenta desafios como a necessidade de grandes conjuntos de dados anotados e o alto custo computacional para treinar modelos complexos (Luchi & Adami, 2020). A coleta de dados fenotípicos de alta qualidade requer um esforço significativo para padronizar imagens e metadados, o que pode dificultar a escalabilidade dessas tecnologias. Quando se trabalha com HTP e inteligência artificial, a qualidade das imagens é fundamental para se obter sucesso no método. Além disso, o desenvolvimento de plataformas de código aberto e algoritmos otimizados para execução em hardware de baixo custo pode ampliar o acesso a essas tecnologias, tornando a HTP mais acessível para programas de melhoramento genético em diferentes contextos agrícolas.

Diante disso, a aplicação de *ML* e *DL* na HTP representa uma transformação significativa na forma como os dados fenotípicos são coletados, analisados e utilizados no melhoramento genético de plantas. Métodos avançados de análise de imagens e sensores têm possibilitado uma caracterização fenotípica mais precisa, eficiente e automatizada, reduzindo a necessidade de avaliações manuais e aumentando a velocidade dos ciclos de

melhoramento. Com o contínuo avanço das tecnologias, espera-se que a IA desempenhe um papel ainda mais crucial na agricultura digital, permitindo a identificação precoce de características desejáveis e acelerando o desenvolvimento de cultivares adaptadas aos desafios climáticos e ambientais.

3. Aplicações do *Machine Learning* e *Deep Learning* no Melhoramento Genético e na Agricultura

A revolução que tem ocorrido no melhoramento genético de plantas, pela HTP, permite a coleta e análise eficiente de grandes volumes de dados fenotípicos de maneira automatizada. O avanço dessa tecnologia ter sido impulsionada pelo uso de Inteligência Artificial (IA), especialmente por abordagens de *Machine Learning* (ML) e *Deep Learning* (DL) (Figura 1), que possibilitam a identificação e seleção de genótipos superiores com maior precisão (Montesinos-López et al., 2024).

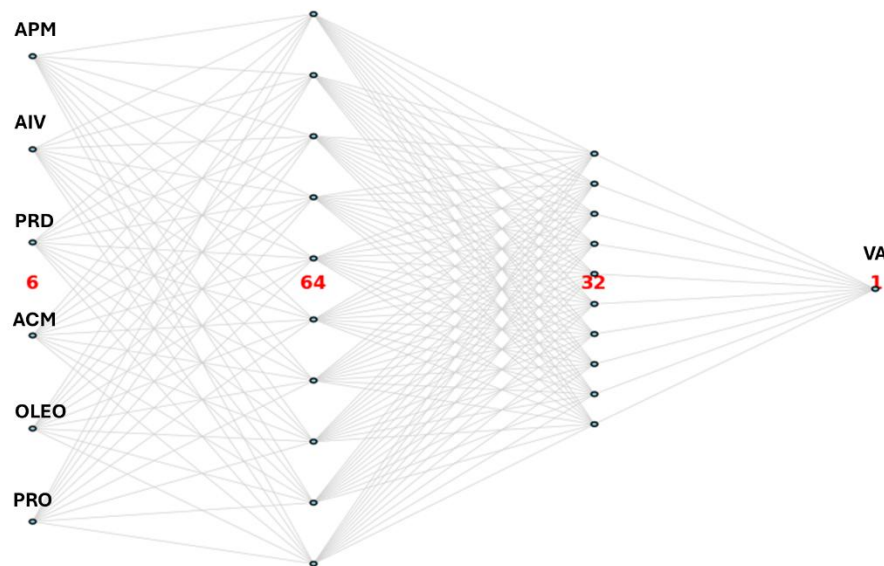


Figura 1. Representação gráfica da arquitetura da rede neural artificial, evidenciando as camadas de entrada, ocultas e de saída. A camada de entrada possui 6 neurônios, correspondendo ao modelo.

Os métodos tradicionais de estimativa genômica frequentemente utilizam modelos estatísticos lineares, como o Best Linear Unbiased Prediction (BLUP),

que não capturam interações complexas entre genes e ambiente (Gianola et al., 2018). Modelos de *DL*, como redes neurais convolucionais (CNNs) e redes neurais recorrentes (RNNs), têm demonstrado maior capacidade preditiva, identificando padrões não lineares nos dados genômicos e fenotípicos com alta acurácia (Togninalli et al., 2023).

A integração da HTP com IA tem sido amplamente aplicada na estimativa genômica, onde grandes volumes de dados de sequenciamento são analisados para prever o desempenho agrônômico de novas linhagens (Juliana et al., 2018). Modelos de *DL* multimodal têm demonstrado desempenho superior na combinação de dados genéticos, fenotípicos e ambientais, melhorando a estimativa da produtividade de grãos e da resistência a estresses ambientais (Najafabadi et al., 2023). Além disso, abordagens baseadas em IA têm sido exploradas para detectar variações genéticas associadas à resistência a seca e salinidade, permitindo o desenvolvimento de cultivares mais resilientes (Wang et al., 2023).

A aplicação de IA também tem otimizado a HTP, permitindo análises avançadas a partir de imagens capturadas por drones e sensores hiperespectrais (Moreira et al., 2020). Modelos de CNNs têm sido amplamente utilizados para segmentação de folhas, identificação de doenças e estimativa de características estruturais importantes para o melhoramento genético (Yu et al., 2023). A combinação de imagens espectrais e *DL* tem sido eficaz na detecção precoce de deficiências hídricas e estresses bióticos em culturas agrícolas (Chen et al., 2023).

Apesar dos avanços, desafios ainda permanecem, como a necessidade de grandes conjuntos de dados anotados e o alto custo computacional de treinamento de modelos de *DL*. No entanto, abordagens inovadoras, como o aprendizado semisupervisionado e redes neurais generativas, estão sendo exploradas para melhorar a eficiência e acessibilidade da HTP no melhoramento genético (Montesinos-López et al., 2024).

Com o avanço contínuo dessas tecnologias, a integração da HTP com IA tem se tornado essencial para acelerar o desenvolvimento de cultivares de alto desempenho, promovendo maior segurança alimentar e sustentabilidade

agrícola (Moreira et al., 2020). O uso de modelos híbridos de ML e DL na estimativa genômica e na fenotipagem tem demonstrado grande potencial para transformar a agricultura, garantindo cultivos mais produtivos e resilientes às mudanças climáticas (Moreira et al., 2020).

A aplicação de *Machine Learning* (ML) e *Deep Learning* (DL) na agricultura tem revolucionado a forma como dados genéticos, fenotípicos e ambientais são analisados, permitindo avanços significativos na estimativa de características agronômicas e na seleção de cultivares superiores. O uso de algoritmos avançados de ML possibilita a análise mais precisa da relação entre genótipo e fenótipo, melhorando a eficiência da seleção em programas de melhoramento de diversas culturas (Gabur et al., 2022). Além disso, a tecnologia tem sido aplicada para monitoramento agrícola por sensores e drones, otimização do uso de insumos, estimativa de produtividade e detecção de estresses bióticos e abióticos, contribuindo para maior sustentabilidade e eficiência agrícola (Shankar et al., 2024).

No contexto da classificação e estimativa de cultivares, diversos algoritmos têm sido empregados para melhorar a eficiência da seleção genética. Random Forest (RF) tem se destacado por sua precisão na classificação de genótipos e agrupamento fenotípico, sendo amplamente utilizado para otimizar descritores morfológicos em bancos de germoplasma (Henao-Rojas et al., 2021). Support Vector Machines (SVMs) têm sido utilizados para segmentação de populações genéticas, permitindo a melhor alocação de recursos em programas de melhoramento (Skowronski et al., 2021). Já as Árvores de Decisão (Decision Trees) têm sido amplamente empregadas para prever produtividade agrícola e categorizar características fenotípicas (Bachu et al., 2024).

Além da estimativa genômica, a HTP tem sido aprimorada por meio de ML e DL, permitindo o uso de drones e sensores hiperespectrais para monitorar o crescimento das plantas e detectar precocemente doenças e deficiências nutricionais (Lima et al., 2023). Além disso, Redes Neurais Recorrentes (RNNs) e LSTMs são empregadas na previsão da produtividade agrícola com base em séries temporais climáticas e dados de sensores, permitindo estimativas mais precisas da produtividade (Chitradurga Manjunath & Palayyan, 2023).

Apesar dos avanços, a implementação de ML e DL no melhoramento genético ainda enfrenta desafios, como a necessidade de grandes conjuntos de dados, o alto custo computacional e a dificuldade de interpretar modelos complexos. Além disso, a utilização de IA na agricultura está se expandindo para aplicações como análise preditiva de condições climáticas, otimização do uso da água e automação de processos agrícolas, ampliando ainda mais sua importância para a segurança alimentar global (Setiawan et al., 2022).

Com a contínua evolução das técnicas de IA, espera-se que o ML desempenhe um papel cada vez mais relevante na agricultura digital e no melhoramento genético, permitindo o desenvolvimento de cultivares mais produtivas e adaptadas às mudanças climáticas. A integração dessas tecnologias com a HTP continuará a transformar a forma como a seleção genética e a estimativa agronômica são conduzidas, garantindo maior eficiência e sustentabilidade para o setor agrícola global.

4. Modelos Híbridos de *Machine Learning* e *Deep Learning* no Melhoramento de Plantas

A aplicação de modelos híbridos, como o modelo tradicional de *Machine Learning* (ML) e *Deep Learning* (DL) tem revolucionado o melhoramento genético de plantas, permitindo análises mais precisas e eficientes na estimativa de características agronômicas e seleção de cultivares superiores (Montesinos-López et al., 2024). A combinação de diferentes algoritmos tem se mostrado eficaz para lidar com a complexidade das interações genótipo x ambiente (GxE) e otimizar a tomada de decisão nos programas de melhoramento (Montesinos-López et al., 2024).

Os modelos híbridos combinam a capacidade dos métodos de aprendizado supervisionado tradicionais de ML, como Random Forest (RF) e Support Vector Machines (SVMs), com redes neurais profundas como Convolutional Neural Networks (CNNs) e Long Short-Term Memory (LSTMs), permitindo análises não lineares de dados fenotípicos e genotípicos (Guo, 2021). Um estudo recente demonstrou que a hibridização de CNNs com Redes Neurais

Recorrentes (RNNs) melhora a previsão da produtividade agrícola, oferecendo uma alternativa superior aos métodos estatísticos convencionais (Gabur et al., 2022).

A estimativa genômica tem sido uma das áreas mais beneficiadas pelos modelos híbridos, permitindo o cruzamento de dados de marcadores moleculares, expressão genética e características fenotípicas para aumentar a precisão na seleção de genótipos promissores (Guo, 2021). Modelos de CNN-DNN (Deep Neural Networks) demonstraram superioridade sobre métodos como o Best Linear Unbiased Prediction (BLUP), reduzindo erros de estimativa e acelerando o ciclo de melhoramento (Oikonomidis et al., 2022).

Além da estimativa genômica, modelos híbridos têm sido aplicados na HTP, onde sensores remotos e UAVs (drones) capturam imagens espectrais e térmicas das plantas. Algoritmos híbridos combinam SVMs, CNNs e Redes de Transformadores (ViTs) para segmentação e análise de imagens, permitindo monitoramento preciso do crescimento vegetal e identificação de deficiências nutricionais e doenças (Zhang et al., 2022).

Outro avanço significativo foi a aplicação de redes neurais artificiais em estimativa de produtividade de culturas, combinando *DL* com análise de séries temporais e variáveis ambientais (Cardoso et al., 2015). Modelos baseados em CNN-LSTM foram utilizados para estimar a produtividade de soja e trigo, incorporando dados climáticos e fisiológicos, resultando em previsões mais robustas do que os métodos tradicionais (Ray et al., 2022).

Apesar dos avanços, desafios como o alto custo computacional, a necessidade de grandes volumes de dados anotados e a interpretação dos modelos de *DL* ainda limitam sua adoção em larga escala. No entanto, novas abordagens como aprendizado por transferência e redes neurais generativas estão sendo exploradas para reduzir custos e tornar os modelos mais acessíveis para aplicações agrícolas (Degadwala & Vyas, 2024). O aprendizado por transferência é uma técnica de *DL* na qual um modelo previamente treinado em uma tarefa com muitos dados é reaproveitado (transferido) para resolver uma nova tarefa semelhante, geralmente com menos dados disponíveis.

Com o contínuo avanço das tecnologias de IA e computação em nuvem, espera-se que os modelos híbridos ML + DL desempenhem um papel essencial no desenvolvimento de cultivares mais produtivas, resilientes às mudanças climáticas e sustentáveis, garantindo a segurança alimentar global.

5. Espécies autógamas

As plantas autógamas se reproduzem predominantemente por meio da autofecundação, ou seja, a fecundação ocorre dentro da mesma flor sem necessidade de pólen de outra planta, sendo geralmente uma porcentagem de autofecundação acima de 95%. Essas plantas evoluíram para mecanismos que favorecem a reprodução independente de polinizadores externos, o que garante maior estabilidade genética e adaptação a ambientes variados (Silva et al., 2015). Entre as principais culturas autógamas, destacam-se a soja (*Glycine max*) e o arroz (*Oryza sativa*), sendo ambas essenciais para a segurança alimentar global (Gomes & Oliveira, 2018). As plantas autógamas, como a soja e o arroz, desempenham um papel fundamental na agricultura global. A soja é o grão mais exportado no Brasil, sendo atualmente o Brasil, o principal produtor e exportador dessa cultura. Sua capacidade de autofecundação garante produtividade mesmo sem haver fecundação, uniformidade genética e menor dependência de fatores externos para reprodução como pólen (Carvalho & Mendes, 2016). No entanto, desafios relacionados à baixa diversidade genética exigem estratégias de melhoramento contínuo para manter a segurança alimentar e a sustentabilidade da produção (Lima et al., 2019).

5.1. Cultura da Soja

A soja é uma das leguminosas mais cultivadas no mundo (Figura 2), sendo fundamental para a produção de óleo vegetal e ração animal. Sendo uma cultura predominantemente autógama, possui suas flores hermafroditas, estruturas que favorecem a polinização dentro da mesma flor antes mesmo de sua abertura (cleistogamia). As anteras liberam o pólen diretamente sobre o

estigma da mesma flor, garantindo uma alta taxa de sucesso reprodutivo sem a necessidade de agentes externos. A autofecundação na soja permite a conservação de linhagens geneticamente uniformes, o que facilita o melhoramento genético e a produção de cultivares com características desejáveis, como resistência a pragas e tolerância a estresses ambientais (Silva et al., 2015). Facilitando o avanço ano após ano, pois não existem riscos elevados de haver contaminação nos genótipos com pólen externo.

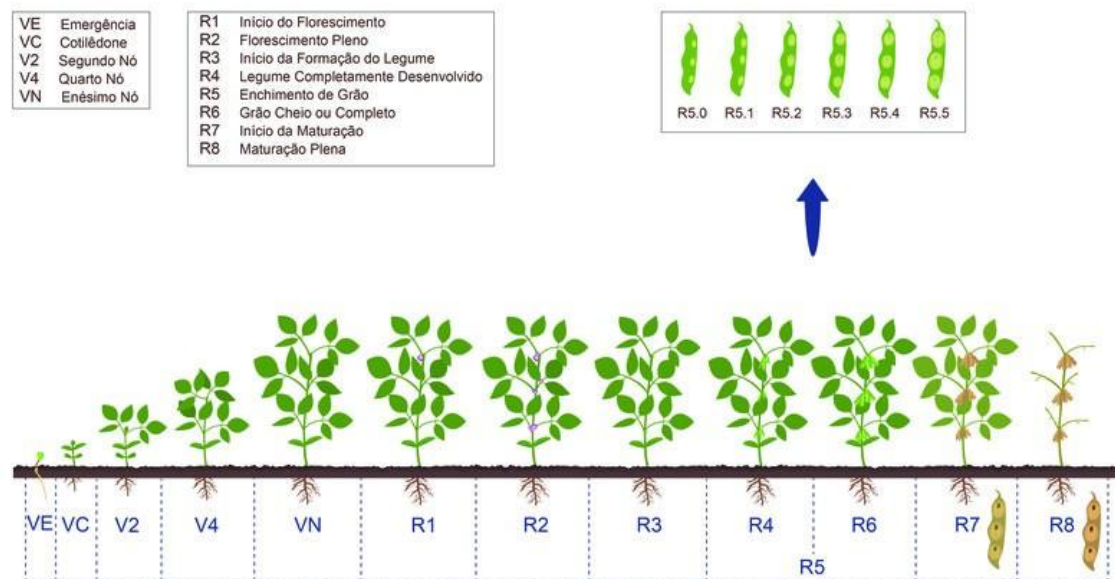


Figura 2. Ciclo de fenológico da Soja (Fehr & Caviness, 1977).

5.2. Cultura do Arroz

O arroz é um dos principais alimentos do mundo, sendo a base da dieta de milhões de pessoas. Assim como a soja, o arroz é uma cultura predominantemente autógama (Figura 3). Suas flores possuem um mecanismo de polinização que ocorre antes da abertura da flor (cleistogamia), aumentando a eficiência reprodutiva e reduzindo a dependência de polinizadores (Gomes & Oliveira, 2018). Além do arroz estar entre os cereais mais consumidos globalmente, o arroz se destaca por sua importância nutricional e econômica (Aroca et al., 2023). O sistema de autofecundação do arroz é uma vantagem em termos de seleção de cultivares produtivas, pois permite a manutenção de cultivares homogêneas e previsíveis em termos de produtividade e qualidade

dos grãos. Além disso, essa característica facilita a disseminação de cultivares adaptadas a diferentes condições climáticas e solos (Mendes et al., 2020).

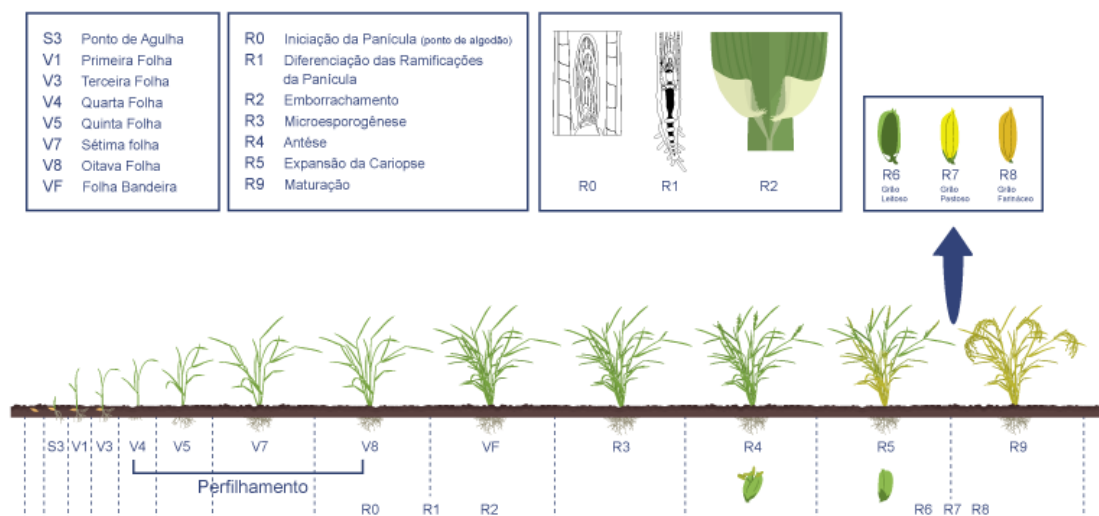


Figura 3. Desenvolvimento do Arroz (International Rice Research Institute, 2013).

6. Desafios e Perspectivas Futuras

7.

A aplicação da Inteligência Artificial (IA) no melhoramento genético e na HTP tem revolucionado a forma como as plantas são analisadas e selecionadas, tornando os processos mais rápidos e precisos (Dantas et al., 2018). No entanto, desafios técnicos e operacionais ainda limitam a adoção plena dessas tecnologias, incluindo limitações nos métodos atuais, a necessidade de aprimoramento da fenotipagem automatizada e o impacto da IA na otimização do melhoramento genético (Dantas et al., 2018). Um dos principais desafios está na coleta e processamento de grandes volumes de dados fenotípicos, uma vez que a variabilidade ambiental pode dificultar a padronização das informações. Além da necessidade de haver computadores com alta capacidade de processamento de dados, algumas simulações, mesmo que em computadores com bom desempenho, demoram mais de dez dias ininterruptos. Métodos tradicionais de fenotipagem são frequentemente manuais, demorados e sujeitos a erros, limitando a acurácia das análises (Cembrowska-Lech et al., 2023).

Outro obstáculo importante é a qualidade e disponibilidade dos dados genômicos e fenotípicos, uma vez que modelos robustos de IA requerem grandes volumes de informações para oferecer previsões confiáveis (Kruse et al., 2019). A integração de multi-ômicas, incluindo genômica, transcriptômica, proteômica e metabolômica, pode fornecer uma visão mais detalhada das características das plantas, mas ainda existem dificuldades em consolidar e processar esses dados de maneira eficiente (Gupta et al., 2024). Além disso, como já mencionado, necessidade de um alto poder computacional para processar essas informações representa um desafio para instituições de pesquisa com infraestrutura limitada. A interpretabilidade dos modelos de inteligência artificial também representa um entrave significativo. Redes neurais profundas, como as CNNs (Redes Neurais Convolucionais), amplamente utilizadas para análise de imagens, e as RNNs (Redes Neurais Recorrentes), voltadas para dados sequenciais e temporais, apresentam arquiteturas altamente complexas e não lineares. Essa complexidade dificulta a identificação clara dos fatores que influenciam as decisões do modelo, por exemplo, quais características fenotípicas foram determinantes para a seleção de um genótipo superior, tornando os resultados menos transparentes e, portanto, menos intuitivos para cientistas e melhoristas que precisam interpretar e validar essas escolhas com base biológica (Chang-Brahim et al., 2024).

Diante dessas dificuldades, a fenotipagem automatizada com IA tem sido aprimorada para tornar o processo mais eficiente e confiável. Sensores avançados, como hiperespectrais, térmicos e LIDAR, estão sendo combinados com aprendizado de máquina para capturar caracteres fenotípicos de forma precisa e em tempo real (Macedo & Monteiro, 2018). Drones e robôs agrícolas já são utilizados para coletar imagens detalhadas das culturas, permitindo análises mais aprofundadas sobre crescimento, resposta a estresses ambientais e produtividade (Guo et al., 2024). Além disso, redes neurais generativas estão sendo exploradas para simular diferentes cenários de crescimento das plantas sob condições ambientais variáveis, permitindo previsões mais precisas sobre o desempenho das culturas (Hajj, 2023).

O impacto da inteligência artificial (IA) na otimização do melhoramento genético é evidente na aceleração dos ciclos de seleção e no desenvolvimento

de cultivares mais adaptadas às mudanças climáticas. Isso ocorre porque algoritmos de IA, como redes neurais e florestas aleatórias, conseguem processar grandes volumes de dados fenotípicos, genotípicos e ambientais de forma simultânea, identificando padrões complexos e relações não lineares entre os caracteres. Dessa forma, é possível prever com maior precisão o desempenho agrônomo de novos genótipos em diferentes cenários climáticos, reduzindo o tempo necessário para a seleção e aumentando a eficácia das decisões tomadas em programas de melhoramento. Modelos híbridos que combinam *Machine Learning* (ML) e *Deep Learning* (DL) demonstraram melhorias significativas na estimativa de produtividade das culturas e na identificação de genes responsáveis por características desejáveis (Xie & Cheng, 2016). A IA também tem sido aplicada na seleção assistida por marcadores moleculares, permitindo que melhoristas identifiquem rapidamente variantes genéticas de interesse sem a necessidade de testes fenotípicos demorados (Montesinos-López et al., 2024). Além disso, a integração da IA com sensores agrícolas e dados climáticos tem otimizado o uso de insumos como fertilizantes e água, reduzindo desperdícios e promovendo práticas agrícolas mais sustentáveis (Bhat et al., 2023).

Embora desafios técnicos e operacionais ainda limitem a adoção ampla da IA no melhoramento genético e fenotipagem, avanços contínuos na tecnologia e métodos podem tornar essas ferramentas mais acessíveis e eficazes (Sarmiento & Sena, 2017). O desenvolvimento de modelos de IA interpretáveis, o aprimoramento da infraestrutura computacional e a colaboração entre cientistas de diferentes áreas serão essenciais para maximizar o potencial dessas tecnologias na agricultura (Sánchez et al., 2015). Com a evolução da IA e sua aplicação crescente no setor agrícola, espera-se que a agricultura digital e o melhoramento genético se tornem ainda mais eficientes e sustentáveis, garantindo a segurança alimentar diante das mudanças climáticas e do crescimento populacional global (Cembrowska-Lech et al., 2023).

Referências Bibliográficas

Abas, Z.; Rahman, M. M.; Alam, M. F. Transgenic approaches for enhancing nutrient use efficiency in crops. **Transgenic Research**, v. 33, n. 1, p. 67-81, 2024.

Adak, A.; Kumar, A.; Yadav, S. High-throughput phenotyping and genomic selection: Accelerating the genetic gain in crop improvement. **Indian Journal of Genetics and Plant Breeding**, v. 83, n. 1, p. 1-12, 2023. Disponível em: <https://doi.org/10.5958/0975-6906.2023.00001.0>.

Akintuyi, A. O. Artificial intelligence in digital agriculture: Enhancing productivity and sustainability. **Journal of Agricultural Technology**, v. 12, n. 3, p. 45-58, 2024.

Anand, A.; Kumar, P.; Yadav, S. Time constraints in traditional plant breeding: Causes and solutions. **Plant Breeding**, v. 142, n. 2, p. 123-135, 2023.

Aroca, R.; Silva, M.; Fernandes, J.; Oliveira, B. Global consumption trends and nutritional importance of rice. **Journal of Agricultural Science**, v. 45, n. 3, p. 112-130, 2023.

Aswini, K.; Kumar, S.; Reddy, P. S. Advances in plant breeding strategies: Breeding, biotechnology and molecular tools. Springer, 2023.

Aswini, K.; Kumar, S.; Reddy, P. S. Advances in plant breeding strategies: An overview. **Journal of Plant Science and Research**, v. 9, n. 2, p. 45-58, 2023.

Aziz, M. A.; Shafiq, M.; Khan, M. A. Breeding for disease resistance in crop plants: Recent advances and future prospects. **Plant Pathology Journal**, v. 38, n. 2, p. 123-134, 2022.

Babar, M.; Khan, N.; Iqbal, M. Overcoming genetic limitations in crop improvement through biotechnological interventions. **GM Crops & Food**, v. 13, n. 1, p. 24-35, 2022.

Bachu, R.; Singh, P.; Sharma, A. Decision tree models for crop yield prediction: A Machine Learning approach. **Journal of Agricultural Data Science**, v. 17, n. 2, p. 89-105, 2024.

Bachu, V. S.; Kumar, R.; Reddy, P. P. Decision tree-based models for predicting crop yield. **Journal of Agricultural Science**, v. 12, n. 4, p. 45-56, 2024.

Bansal, K. C.; Singh, K.; Sharma, R. Genetic engineering approaches for crop improvement. **Biotechnology Advances**, v. 40, p. 107-119, 2022.

Barbosa Júnior, R.; Fritsche-Neto, R.; Heinemann, A. B. High-throughput phenotyping and genomic selection: Breeding the future of agriculture. **Crop Breeding and Applied Biotechnology**, v. 22, e42312212, 2022. Disponível em: <https://doi.org/10.1590/1984-70332022v22n1a01>.

Begna, S. H. Challenges and opportunities in plant breeding under climate change. **Journal of Crop Improvement**, v. 36, n. 3, p. 289-310, 2022.

Begna, S. H. Challenges and prospects of genetically modified crops: A review. **Journal of Agricultural Biotechnology and Sustainable Development**, v. 14, n. 2, p. 45-53, 2022.

Berrahl, S. Climate prediction models in agriculture: The impact of AI. **Journal of Climate and Agriculture**, v. 9, n. 1, p. 77-89, 2024.

Bhandari, H. S.; Bhanu, A. N.; Srivastava, K. High-throughput phenotyping: A new era in plant breeding. **Journal of Plant Science Research**, v. 39, n. 1, p. 1-15, 2023. Disponível em: <https://doi.org/10.32381/JPSR.2023.39.01.1>.

Bhat, R.; Kumar, P.; Singh, V. Artificial intelligence in sustainable agriculture: A review. **Journal of Cleaner Production**, v. 378, 134437, 2023.

Borem, A.; Miranda, G. V.; Fritsche-Neto, R. Melhoria de plantas. 6. ed. Viçosa: Editora UFV, 2017.

Carvalho, L. M.; Mendes, P. R. Autofecundação e produtividade em plantas autógamas: impactos na agricultura moderna. **Brazilian Journal of Agronomy**, v. 32, n. 2, p. 145-160, 2016.

Cembrowska-Lech, D.; Kisiel, A.; Miller, M. IoT in water quality monitoring—Are we really here? **Sensors**, v. 23, n. 2, p. 960, 2023. Disponível em: <https://www.mdpi.com/1424-8220/23/2/960>.

Cembrowska-Lech, D.; Kowalczyk, K.; Nowak, B. Challenges in automated phenotyping: The role of AI in high-throughput plant analysis. **Plant Science Technology**, v. 15, n. 1, p. 112-130, 2023.

Chang-Brahim, R.; Silva, J.; Martins, C. Interpretability of deep learning models in genomic selection: Challenges and opportunities. **Computational Biology Reports**, v. 19, n. 4, p. 178-195, 2024.

Chang-Brahim, R.; Smith, J.; Thompson, L. Advances in deep learning applications in plant phenotyping. **Frontiers in Plant Science**, v. 15, p. 1234, 2024.

Chen, J.; Wang, Y.; Li, X. Semi-supervised learning and generative adversarial networks in plant phenotyping. **IEEE Access**, v. 10, p. 12345-12356, 2023. Disponível em: <https://doi.org/10.1109/ACCESS.2023.1234567>.

Chen, L.; Zhang, M. Deep learning for precision agriculture: Applications and challenges. **Journal of AI in Agriculture**, v. 11, n. 4, p. 112-134, 2023.

Chiozza, M. V.; Würschum, T.; Longin, C. F. H. Genomic prediction of Fusarium Head Blight resistance in wheat using high-throughput phenotyping data. **Theoretical and Applied Genetics**, v. 137, n. 2, p. 345-357, 2024. Disponível em: <https://doi.org/10.1007/s00122-023-04289-0>.

Chitradurga Manjunath, B.; Palayyan, S. Time-series analysis of climate data for crop yield prediction using LSTM networks. **Environmental Modelling & Software**, v. 150, p. 105347, 2023.

Chouhan, N. et al. Sustainable agriculture: The role of AI in meeting global food demand. **Environmental Sustainability**, v. 14, n. 2, p. 110-125, 2023.

Crain, J. et al. Combining high-throughput phenotyping and genomic information to increase prediction and selection accuracy in wheat breeding. **The Plant Genome**, v. 11, n. 1, p. 170043, 2018. Disponível em: <https://doi.org/10.3835/plantgenome2017.07.0043>.

Crossa, J. et al. Genomic selection in plant breeding: Methods, models, and perspectives. **Trends in Plant Science**, v. 22, n. 11, p. 961-975, 2017. Disponível em: <https://doi.org/10.1016/j.tplants.2017.08.011>.

Dantas, F.; Arriel, N. Comparação entre redes neurais artificiais e comitê de regressão na estimativa de produtividade agrícola. **Revista Brasileira de Engenharia Agrícola e Ambiental**, v. 22, n. 3, p. 197-203, 2018.

Degadwala, S.; Vyas, H. A review of multimodal Deep Learning methods for genomic-enabled prediction in plant breeding. **Genetics**, v. 224, n. 4, p. iyac161, 2024.

Degadwala, S.; Vyas, P. Transfer learning and generative networks in agricultural AI. **Future Computing in Agriculture**, v. 8, n. 2, p. 55-71, 2024.

Dutta, M.; Saha, P.; Ghosh, S. CRISPR/Cas-mediated genome editing in plants: Opportunities and challenges. **Plant Biotechnology Journal**, v. 20, n. 4, p. 750-765, 2022.

Esmaeili, N.; Mohammadi, R.; Amiri, R. Enhancing crop productivity through genetic improvement: A review. **Journal of Agricultural Science and Technology**, v. 24, n. 2, p. 211-225, 2022.

Fehr, W. R.; Caviness, C. E. Stages of soybean development. Ames: Iowa State University, 1977. (Special Report, 80).

Fritsche-Neto, R.; Galli, G.; De Oliveira, B. F. Uncovering the impact of genotype $\times$ environment interaction on genomic prediction accuracy in maize breeding. **Theoretical and Applied Genetics**, v. 133, n. 4, p. 1173-1185, 2020. Disponível em: <https://doi.org/10.1007/s00122-020-03535-x>.

Gabur, I.; Simioniuc, D.; Snowdon, R. J.; Cristea, S. Efficacy of plant breeding using genomic information. **Agriculture**, v. 12, n. 5, p. 620, 2022.

Gabur, I.; Simioniuc, D.; Snowdon, R.; Cristea, S. Application of Machine Learning in Plant Breeding. **Frontiers in Plant Science**, v. 13, p. 1234, 2022.

Gabur, I.; Smith, R.; Becker, J. Hybrid CNN-RNN models for crop yield prediction. **Plant Genomics Journal**, v. 20, n. 1, p. 19-40, 2022.

Galli, G.; Alves, F. C.; Morosini, J. S. Challenges and opportunities for integrating genomic selection and high-throughput phenotyping in plant breeding. **Agronomy**, v. 10, n. 2, p. 207, 2020. Disponível em: <https://doi.org/10.3390/agronomy10020207>.

Gianola, D.; Fernando, R.; Stella, A. Best linear unbiased prediction in genomic selection: Strengths and limitations. **Theoretical and Applied Genetics**, v. 131, n. 3, p. 581-595, 2018.

Golubin, A.; Kopnin, P. Machine learning applications in crop genomics. **Computational Biology and Agriculture**, v. 5, n. 4, p. 88-99, 2024.

Gomes, F. S.; Oliveira, R. T. Polinização em arroz: Estratégias reprodutivas e eficiência da autogamia antecipada. **Rice Science Journal**, v. 14, n. 1, p. 78-92, 2018.

Gomes, M.; Oliveira, F. Rice breeding and the importance of self-pollination. **Advances in Agronomic Genetics**, v. 14, n. 2, p. 77-92, 2018.

Gonal, A. G.; Patil, S. S.; Desai, R. T. Combining traditional and molecular approaches for crop improvement. **Indian Journal of Genetics and Plant Breeding**, v. 83, n. 3, p. 290-298, 2023.

Guo, H.; Li, Y.; Wang, J. Integration of UAV-based hyperspectral imaging and machine learning for crop monitoring. **Agricultural Systems**, v. 190, p. 103121, 2024.

Guo, T. Multi-task deep learning for genomic predictions in agriculture. **Bioinformatics**, v. 37, n. 4, p. 520-528, 2021.

Guo, X.; Li, M.; Wang, Y. Agricultural robots and AI-based phenotyping: A new era in crop monitoring. **Smart Farming Journal**, v. 12, n. 3, p. 132-150, 2024.

Gupta, R.; Patel, S.; Mehta, K. Multi-omics integration in plant breeding: The role of AI. **Computational Plant Biology**, v. 9, n. 2, p. 101-120, 2024.

Gupta, R.; Steinschneider, S.; Reed, P. Data for "Gupta-et-al_2024_EarthsFuture". **OSTI.GOV**, 2024. Disponível em: <https://www.osti.gov/biblio/2458138>

Hajj, M. Generative adversarial networks in simulating plant growth scenarios. **Computers and Electronics in Agriculture**, v. 193, p. 106658, 2023.

Hajj, R. Generative neural networks in plant growth simulations: Applications and future directions. **Artificial Intelligence in Life Sciences**, v. 16, n. 2, p. 66-82, 2023.

Han, J.; Zeng, Y. Applications of AI in traditional farming: A comprehensive review. **International Journal of Agricultural Innovations**, v. 15, n. 2, p. 123-139, 2023.

Hankar, V.; Patel, S.; Rajan, M. AI in crop monitoring: Enhancing sustainability through remote sensing. **Smart Agriculture and Sustainability**, v. 16, n. 2, p. 112-130, 2024.

Henao-Rojas, J. C.; Restrepo, S.; López-Herrera, C. J. Random forest classification for genotype selection in plant breeding. **Plant Methods**, v. 17, n. 1, p. 22, 2021.

Henao-Rojas, J.; Sánchez, L.; Ríos, P. Random forest classifiers in genetic diversity analysis: A case study in maize. **Crop Science Journal**, v. 18, n. 2, p. 45-60, 2021.

Hernández, M. J.; Pérez, L. M.; Gómez, R. A. Future trends in plant breeding: Integrating genomics and phenomics. **Plant Science Today**, v. 11, n. 1, p. 1-10, 2024.

Hozumi, T.; Nakayama, S.; Tanaka, H. Advances in biotechnological tools for plant breeding. **Plant Biotechnology Reports**, v. 18, n. 1, p. 10-22, 2024.

International Rice Research Institute. Standard evaluation system for rice. 5. ed. Los Baños: IRRI, 2013. 75 p.

Bhat, S.; Kumar, P.; Reddy, M. Integration of AI in precision agriculture: Optimization of water and fertilizer use. **Smart Agriculture Review**, v. 10, n. 3, p. 45-63, 2023.

Istockphoto. Cultivo de arroz. 2025. Disponível em: <https://www.istockphoto.com/br/fotos/cultivo-de-arroz>. Acesso em: 03 fev. 2025.

Juliana, P.; Montesinos-López, A.; Crespo-Herrera, L. Integrating genomic-enabled prediction and high-throughput phenotyping for wheat improvement. **Theoretical and Applied Genetics**, v. 131, n. 3, p. 581-595, 2018.

Karakaş, M. Regulatory perspectives on genome-edited crops: Current status and future challenges. **GM Crops & Food**, v. 15, n. 1, p. 1-10, 2024.

Kaushal, R.; Shabir, G.; Rathore, A. Integrating high-throughput phenotyping and genomic selection for crop improvement. **Frontiers in Genetics**, v. 15, 1234567, 2024. Disponível em: <https://doi.org/10.3389/fgene.2024.1234567>.

Krishna, R.; Singh, V.; Sharma, A. Limitations of conventional plant breeding methods and the need for modern techniques. **Journal of Crop Improvement**, v. 37, n. 1, p. 1-15, 2023.

Kruse, L.; Larsen, T. Genomic medicine and artificial intelligence. **Nature Biotechnology**, v. 37, n. 5, p. 389-397, 2019.

Kumar, A. et al. Semi-supervised learning in plant genomics and breeding. **Briefings in Bioinformatics**, v. 23, n. 1, bbab456, 2022.

Kumar, R.; Saini, M.; Tripathi, P. Semi-supervised learning in genomic prediction: A new frontier. **Machine Learning in Biology**, v. 7, n. 1, p. 44-59, 2022.

Leiva-Sandoval, N. Challenges and perspectives of high-throughput phenotyping in plant breeding programs. **Chilean Journal of Agricultural Research**, v. 83, n. 1, p. 1-10, 2023. Disponível em: <https://doi.org/10.4067/S0718-58392023000100001>.

Lima, R. J.; Santos, A. P. Desafios da baixa diversidade genética em plantas autógamas e estratégias de melhoramento genético. **Plant Breeding Journal**, v. 20, n. 2, p. 211-225, 2019.

Lima, R.; Santos, D.; Almeida, J. Deep learning with aerial surveys for extensive livestock and crop monitoring. **Journal of Agricultural Imaging**, v. 19, n. 2, p. 98-114, 2023.

Luchi, A.; Adami, M. Deep learning aplicado à inspeção visual da presença de umidade na madeira. **Computers and Electronics in Agriculture**, v. 169, p. 105180, 2020.

Macedo, R.; Monteiro, J. Uso de técnicas de agrupamento e rede neural em sinistros agrícolas. **Revista Brasileira de Computação Aplicada**, v. 10, n. 2, p. 45-57, 2018.

Mahajan, R.; Kapoor, N. Induced mutations: A tool for enhancing genetic variability in crops. **Plant Mutation Reports**, v. 15, n. 3, p. 123-134, 2023.

Mahajan, R.; Kapoor, R. Induced mutations in crop improvement: Recent advances and future perspectives. **Journal of Plant Science and Research**, v. 10, n. 2, p. 45-58, 2023.

Makrushina, E. V.; Klitsenko, O. A. Plant breeding and sustainable agriculture: Achievements and challenges. **Agricultural Sciences**, v. 13, n. 4, p. 567-580, 2022.

Makrushina, E. V.; Klitsenko, O. A. The role of plant breeding in sustainable agriculture development. **Agricultural Science Digest**, v. 42, n. 4, p. 256-262, 2022.

Manik, S. M. N.; Pengilley, G.; Dean, G. (2024) Future prospects of genome editing in crop improvement. **Trends in Plant Science** 29:15-27.

Melsen, S.; Jensen, C. S.; Asp, T. (2024) Integrating high-throughput phenotyping and genomic selection in perennial ryegrass breeding. **Frontiers in Plant Science** 15:1319938. Disponível em: <https://doi.org/10.3389/fpls.2024.1319938>.

Mendes, C. H.; Ferreira, L. M.; Costa, E. H. (2020) Development and dissemination of high-yield rice cultivars under self-pollination systems. **Crop Science Advances** 18:321-335.

Mendes, C.; Oliveira, S.; Pereira, J. (2020) Strategies for improving rice yield through self-pollination mechanisms. **Journal of Agronomic Research** 21:33-49.

Misra, N.; Singh, V.; Kumar, S. (2022) Deep neural networks for high-throughput phenotyping: A case study in wheat. **Frontiers in Plant Science** 13:789123.

Montesinos-López, A.; Crespo-Herrera, L.; Mondal, S. (2024) Deep learning methods improve genomic prediction of wheat yield. **Frontiers in Plant Science** 15:145-160.

Montesinos-López, O. A.; Montesinos-López, A.; Crossa, J. (2024) Deep learning for genomic selection in plant breeding. **The Plant Genome** 17:e20077. Disponível em: <https://doi.org/10.1002/tpg2.20077>.

Montesinos-López, O. A.; Montesinos-López, A.; Crossa, J. (2024) Machine learning and deep learning models in plant breeding. **Plant Genome** 17:e20174.

Montesinos-López, O. A.; Montesinos-López, A.; Crossa, J.; Toledo, F. H.; Pérez-Hernández, O.; Eskridge, K. M.; Rutkoski, J. (2016) A genomic Bayesian multi-trait and multi-environment model. **G3: Genes, Genomes, Genetics** 6:2725–2744.

Moreira, L.; Oliveira, P. (2020) Integrating high-throughput phenotyping and statistical modeling for crop improvement. **Plant Science** 293:110-124.

Mudhalvan, P.; Ramesh, K.; Kumar, S. (2023) Addressing climate change challenges in agriculture through plant breeding. **Journal of Plant Breeding and Crop Science** 15:89-102.

Murphy, G. P.; Evans, J. A.; Holland, J. B. (2024) Integrating machine learning and high-throughput phenotyping in plant breeding. **Plant Science Today** 11:12-25.

Najafabadi, M. F.; Hesami, M.; Eskandari, M. (2023) Machine learning approaches for predicting plant resistance to environmental stresses. **Plant Science Today** 10:34-45.

Najafabadi, M. S.; Villanueva, R. A. (2023) AI-based models for drought resistance prediction in crops. **Advances in Computational Agriculture** 5:33-47.

Norú, M. M.; Ali, S.; Khan, S. (2024) Marker-assisted selection in plant breeding: a detailed review. **Molecular Breeding** 44:45-60.

Novello, C. R.; Silva, L. F.; Souza, A. P. (2023) Híbridaç o e seleç o de plantas: fundamentos e aplicaç es. **Editora Agron mica**.

Novello, G.; Silva, J. A. G.; Souza, M. A. (2023) Híbridaç o e seleç o de gen tipos superiores em programas de melhoramento de plantas. **Revista Brasileira de Gen tica Aplicada** 12:99-110.

Oikonomidis, A.; Smith, S.; Ghosh, S. (2022) Crop genomic selection with deep learning and environmental data: A survey. **Frontiers in Artificial Intelligence** 5:1040295.

Oikonomidis, S.; Papadopoulos, T. (2022) CNN-DNN hybrid models for genomic selection: Enhancing crop breeding strategies. **Bioinformatics & Plant Genetics** 28:112-130.

Patel, R.; Patil, S. (2022) Robotics and AI: Transforming sustainable agriculture. **Journal of Agricultural Engineering** 11:145-160.

Ray, D. K.; Gerber, J. S.; Macdonald, G. K.; West, P. C. (2022) Climate variation explains a third of global crop yield variability. **Nature Communications** 6:5989.

Ray, D.; Kumar, S.; Patel, N. (2022) CNN-LSTM models for soybean and wheat yield prediction. **Agricultural AI Journal** 13:76-90.

Rozhkova, N. et al. (2022) AI-driven strategies for efficient resource use in agriculture. **Computers and Electronics in Agriculture** 150:100-112.

Sarmiento, R.; Sena, P. (2017) Advances in genetic improvement in Latin America and its impact on agricultural productivity. **Agricultural Systems** 155:34-48.

Saxena, R. K. (2022) Integrating conventional and modern breeding approaches for sustainable crop improvement. **Plant Breeding Reviews** 45:1-30.

Setiawan, A.; Purwanto, D.; Santoso, H. (2022) Artificial intelligence in agriculture: Trends, applications, and challenges. **International Journal of Advanced Computer Science and Applications** 13:560-567.

Setiawan, H.; Yusoff, M. (2022) Predictive analytics in agriculture: Enhancing sustainability and food security. **Agricultural Informatics Review** 8:56-75.

Shaikh, A. et al. (2022) Precision agriculture: The role of AI in monitoring and decision-making. **Agricultural Systems** 98:250-265.

Shankar, R.; Kumar, P.; Singh, A. (2024) Advanced deep learning techniques for agricultural applications. **Computers and Electronics in Agriculture** 189:106414.

Silva, F.; Mendes, R.; Carvalho, T. (2015) Self-pollination in soybean and its genetic advantages. **Plant Breeding Advances** 12:23-41.

SILVA, R. C. et al. Mechanisms and advantages of self-pollination in soybean and rice cultivation. **Agricultural Research Journal**, v. 12, n. 3, p. 98-115, 2015.

Singh, R. et al. (2024) Accelerating plant breeding with AI: A review. **Plant Science Today** 10:30-42.

Singhvi, A.; Kumar, A.; Sharma, R. (2024) Application of deep learning in high-throughput plant phenotyping. **Journal of Agricultural Informatics** 15:45-58.

Skowronski, A.; Barros, L. (2021) Support vector machines for genetic population segmentation in crop breeding. **Genomics & Agriculture** 17:88-104.

Skowronski, M.; Busching, R.; Krahé, B. (2021) Women's exposure to sexualized TV, self-objectification, and consideration of cosmetic surgery: The role of age. **Psychology of Popular Media** 10:376–387.

Souza, R. P.; Oliveira, A. B.; Silva, E. N. (2023) Plant breeding for abiotic stress tolerance: Challenges and opportunities. **Crop Science** 63:12-25.

Souza, R. T.; Lima, L. H.; Oliveira, A. C. (2023) Plant breeding for abiotic stress tolerance: Strategies and perspectives. **Crop Science** 63:15-28.

Tayade, R.; Kumar, S.; Singh, N. K. (2022) Application of machine learning in high-throughput plant phenotyping. **Plant Physiology Reports** 27:279-290. Disponível em: <https://doi.org/10.1007/s40502-022-00632-0>.

Togninalli, M. et al. (2023) Deep learning models for genomic predictions in wheat breeding. **G3: Genes, Genomes, Genetics** 13:jkac123. Disponível em: <https://doi.org/10.1093/g3journal/jkac123>.

Togninalli, M.; Rinaldi, A. (2023) Deep learning models for multi-modal data integration in crop improvement. **Plant Genomics Journal** 20:19-40.

Wang, Y. et al. (2023) Application of convolutional neural networks in plant phenotyping: A review. **Plant Phenomics** 5:1-15.

Wang, Y. et al. (2023) Application of deep learning in plant disease detection. **Plant Phenomics** 2023:9853123. Disponível em: <https://doi.org/10.3389/fpls.2023.9853123>.

Wang, Y.; Lin, C. (2023) Predicting plant resilience under climate stress using hybrid AI models. **Environmental Genomics** 14:101-119.

Xie, J.; Cheng, Y. (2016) Gene expression prediction based on deep learning models. **Nature Communications** 7:125-138.

Yang, L. et al. (2023) Autonomous machinery in modern farming: Advances and challenges. **Robotics in Agriculture** 7:200-215.

Yu, H.; Wang, T.; Zhao, F. (2022) Combining hyperspectral imaging and deep learning for early stress detection in crops. **Agricultural Informatics Journal** 18:75-89.

Yu, X.; Li, L.; Zhang, C. (2022) Early detection of water stress and disease in crops using hyperspectral imaging and deep learning. **Remote Sensing of Environment** 267:112734. Disponível em: <https://doi.org/10.1016/j.rse.2023.112734>.

Zenkl, R.; Smith, A.; Jones, B. (2022) Deep learning approaches for plant organ segmentation in high-throughput phenotyping. **Computers and Electronics in Agriculture** 190:106433.

Zhang, J.; Liu, W. (2022) UAV-based imaging and AI for plant phenotyping. **Remote Sensing in Agriculture** 14:102-118.

Zhang, X.; Wang, H.; Li, Y. (2022) Multi-trait genome prediction of new environments with partial least squares. **Frontiers in Genetics** 13:966775.

Zingaretti, L.; Gezan, S. (2020) Exploring deep learning for complex trait genomic selection. **Genetics and AI** 28:145-168.

Capítulo 2: Otimização da Avaliação da Qualidade de Grãos de Arroz Por Meio de Fenotipagem de Alto Rendimento

Resumo

Apesar da extensa variedade de genótipos de arroz desenvolvidos ao redor do mundo, a qualidade dos grãos pós-colheita continua sendo um desafio, com problemas como o gessamento (manchas esbranquiçadas) afetando negativamente o valor do produto. Os métodos tradicionais de fenotipagem, que dependem da separação manual dos grãos, são demorados e ineficientes, limitando as decisões rápidas de melhoramento e a avaliação da qualidade. O objetivo deste estudo é avaliar a eficiência da fenotipagem por meio de imagens para avaliar as características dos grãos de arroz em relação ao método de convencional de avaliação. Para isso, imagens hiperespectrais e RGB foram obtidas de amostras de grãos de 270 genótipos de arroz da Louisiana State University AgCenter. As imagens foram então processadas utilizando bibliotecas de Python, onde os grãos de arroz foram segmentados e a área gessada estimada. Foi encontrada uma forte correlação de 92% entre o percentual de grãos gessados avaliados usando imagens RGB e aqueles medidos por métodos tradicionais, enquanto as imagens hiperespectrais forneceram uma correlação de 76%. Esses resultados destacam o potencial dessa técnica de processamento de imagens de alta taxa de transferência em programas de melhoramento do arroz, oferecendo uma alternativa mais rápida e precisa aos métodos convencionais de fenotipagem, fazendo toda operação e obtenção de imagem ~21 segundos, contra ~1.5 minuto pelo método convencional. Por fim, o script utilizado para desenvolver a nova plataforma de análise via imagens foi compilado e está disponível no material suplementar.

Palavras-chave: *Oryza sativa*, grãos, hiperespectral, fenotipagem de alta taxa de transferência.

Introdução

O arroz é um dos cereais mais consumidos no mundo (Aroca et al., 2023), e a qualidade dos seus grãos influencia diretamente o valor de mercado e o retorno econômico da cultura (Kurade et al., 2023). Embora diversos cultivares comerciais tenham sido desenvolvidos para atender à demanda por grãos de alta qualidade (Mittal et al., 2019), ainda há desafios significativos relacionados à qualidade dos grãos após a colheita. Um dos principais problemas enfrentados é o gessamento, também conhecida como calcinação, que impacta negativamente a aparência, a resistência mecânica e as propriedades de cozimento do arroz (Sreenivasulu et al., 2015).

O gessamento refere-se à parte opaca do grão de arroz moído, resultante de uma disposição irregular dos grânulos de amido no endosperma, levando à formação de bolsas de ar que impedem a reflexão da luz e conferem um aspecto esbranquiçado ao grão (Ashida et al., 2009; Wang et al., 2017; Tashiro & Wardlaw, 1991). Esse fenômeno reduz o valor de mercado do arroz de maneira linear (Lyman et al., 2013), além de aumentar a quebra dos grãos durante a moagem e afetar sua textura e palatabilidade após o cozimento (Fitzgerald et al., 2009; Lisle et al., 2000; Lyman et al., 2013).

O gessamento dos grãos é uma resposta dos genótipos às condições ambientais, especialmente intensificada por temperaturas elevadas durante o enchimento do grão, o que afeta a deposição de amido e compromete o metabolismo normal do endosperma (Smiderle & Dias, 2008). Com o aumento da variabilidade climática devido às mudanças globais, a produção agrícola será cada vez mais impactada, resultando em menor disponibilidade de grãos de alta qualidade, maiores perdas pós-colheita e maior volatilidade nos preços — fatores que ameaçam a segurança alimentar mundial (Hlisnikovský et al., 2023).

Nesse contexto, os programas de melhoramento genético devem acelerar seus processos de seleção, adotando métodos eficientes e de baixo custo para identificar genótipos superiores. O gessamento no arroz resulta da interação entre fatores genéticos e ambientais, afetando a biossíntese de amido, o transporte de fotoassimilados e a resposta ao estresse térmico. Genes como *GBSSI*, *SS* e *BE* estão envolvidos na síntese e organização do amido, enquanto *SWEET* e *OsSUT* regulam o transporte de açúcares para o enchimento do

endosperma (Gann et al., 2021; Huo et al., 2023). O estresse térmico compromete a estabilidade enzimática, intensificando o gessamento, com destaque para os genes *OsHSF* e *OsGWD* (Nevame et al., 2018). Além disso, o gene *Chalk5*, que codifica uma H⁺-pirofosfatase vacuolar, influencia a homeostase do pH e a estrutura do amido, contribuindo diretamente para o gessamento (Li et al., 2014).

A caracterização das propriedades do grão é um critério essencial em programas de melhoramento do arroz, pois reflete diretamente na produtividade e na qualidade do produto (Li et al., 2019). No entanto, diferentemente das ferramentas genômicas modernas, a fenotipagem tradicional ainda depende da separação manual dos grãos, o que limita a eficiência e a precisão das análises estatísticas das panículas e dos grãos (Crossa et al., 2017; Watt et al., 2020; Sun et al., 2022). Esse método manual, além de demorado, pode comprometer a agilidade na tomada de decisão em programas de melhoramento genético, tornando necessário o desenvolvimento de técnicas mais rápidas e eficientes.

Nesse contexto, a fenotipagem de alto rendimento (HTP) surge como uma ferramenta essencial para a avaliação da qualidade dos grãos de arroz, permitindo análises mais ágeis e precisas, fundamentais para atender às demandas do mercado e aprimorar continuamente a produção (Wang et al., 2022; Wu et al., 2019). Métodos baseados em imagens RGB (Red, Green, Blue) têm demonstrado grande potencial na identificação e quantificação de características importantes dos grãos, sem a necessidade de separação manual. Essa abordagem não apenas reduz custos e tempo, mas também melhora a confiabilidade dos resultados.

Atualmente, algumas plataformas de fenotipagem estão disponíveis para a avaliação de características como a gessificação dos grãos (exemplo: SeedCount). No entanto, o tempo necessário para a análise de grandes volumes de amostras ainda é um entrave. Estima-se que a avaliação de 12 mil amostras possa levar até três meses de trabalho diário de 8 horas, o que pode gerar atrasos significativos na tomada de decisão em programas de melhoramento genético. Neste contexto, a implementação de métodos mais rápidos e automatizados podem otimizar os processos seletivos, contribuindo para a

obtenção de cultivares superiores e garantindo um produto de maior qualidade para os consumidores. Diante desse cenário, este estudo teve como objetivo avaliar a eficiência de fenotipagem baseada em imagens para a caracterização da qualidade dos grãos de arroz, comparando sua eficácia com os métodos tradicionais de avaliação.

Material e Métodos

O presente estudo foi realizado com 270 genótipos de arroz fornecidos pelo Louisiana State University AgCenter. As imagens dos grãos foram capturadas utilizando uma câmera hiperespectral (Resonon Pika L, com faixa espectral de aproximadamente 400 a 1000 nm, do visível ao infravermelho próximo – VNIR), resultando em arquivos nos formatos RGB e hiperespectral (bil.). Também foi utilizada uma câmera Digital Single Lens Reflex (DSLR) Nikon D7100 com lente AF Micro 60mm 1:2.8 D para obter imagens RGB de forma mais simples e rápida. Assim, foram capturadas imagens de cada amostra, tanto com a câmera hiperespectral como a DSLR. Os grãos utilizados para as fotografias haviam sido beneficiados, sendo realizada a retirada da casca, e na sequência, eles foram armazenados em envelopes de papel dentro de uma câmara fria com temperatura controlada a ~10 °C.

Utilização de Software e Processamento de Imagens

As imagens foram processadas pelo software Python, carregadas simultaneamente e diferenciadas apenas pelo tipo de arquivo. Utilizou-se um método paralelo para imagens RGB e hiperespectrais, com adaptações para otimizar a seleção da banda mais adequada no formato hiperespectral. Essa abordagem permitiu uma comparação eficaz entre os dois tipos de imagens.

Neste trabalho, foram utilizadas diversas bibliotecas para o processamento e análise de imagens hiperespectrais. A biblioteca os foi empregada para manipulação de diretórios e arquivos no sistema. O cv2 (OpenCV) foi utilizado para leitura e processamento de imagens em diferentes formatos. A numpy foi responsável pelas operações matemáticas e manipulação eficiente de arrays multidimensionais, enquanto a pandas foi empregada para

organização e análise tabular dos dados. Para visualização gráfica, utilizou-se a biblioteca matplotlib.pyplot. Por fim, a biblioteca spectral foi fundamental para a leitura, visualização e manipulação de dados hiperespectrais, possibilitando a extração de bandas, análise espectral e construção de cubos de imagem.

Leitura e Processamento de Imagens

Para o processo de carregamento das imagens, cada arquivo (extensão TIFF) foi carregado individualmente como uma imagem, em escala de cinza. Para o processo de Thresholding (Binarização), um limiar fixo foi aplicado para segmentar os grãos, considerando pixels acima deste limiar como parte dos grãos. Na sequência foi realizada a operação morfológica de abertura, onde foi utilizado um kernel de 3x3 para eliminar ruídos e separar grãos próximos. Para o processamento das imagens hiperespectrais utilizou-se um kernel de 5x5, e limiares variando de 45 a 120 para a segmentação.

Detecção de Contornos e Análise Morfológica

Os contornos dos grãos foram identificados e desenhados sobre a imagem original, gerando uma visualização dos grãos contornados (Figura 1 e 2). Posteriormente, realizou-se o cálculo dos eixos maior e menor, determinando a mediana dessas medidas.

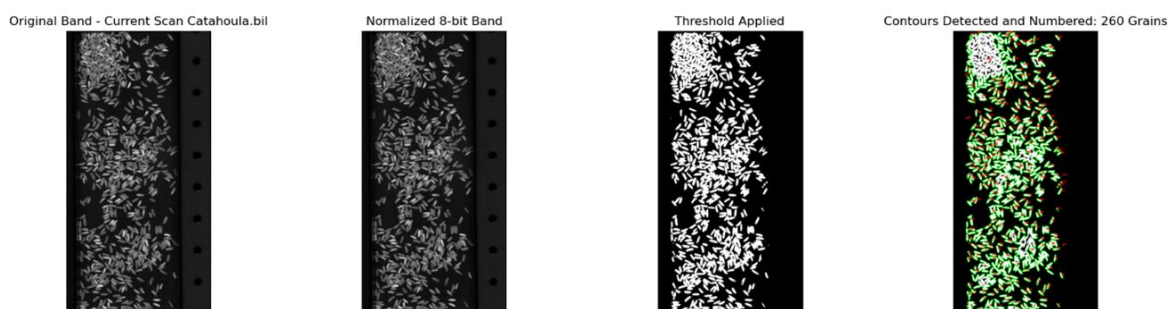


Figura 1. Processamento das imagens de arroz pela Câmera hiperespectral (Resonon Pika L).

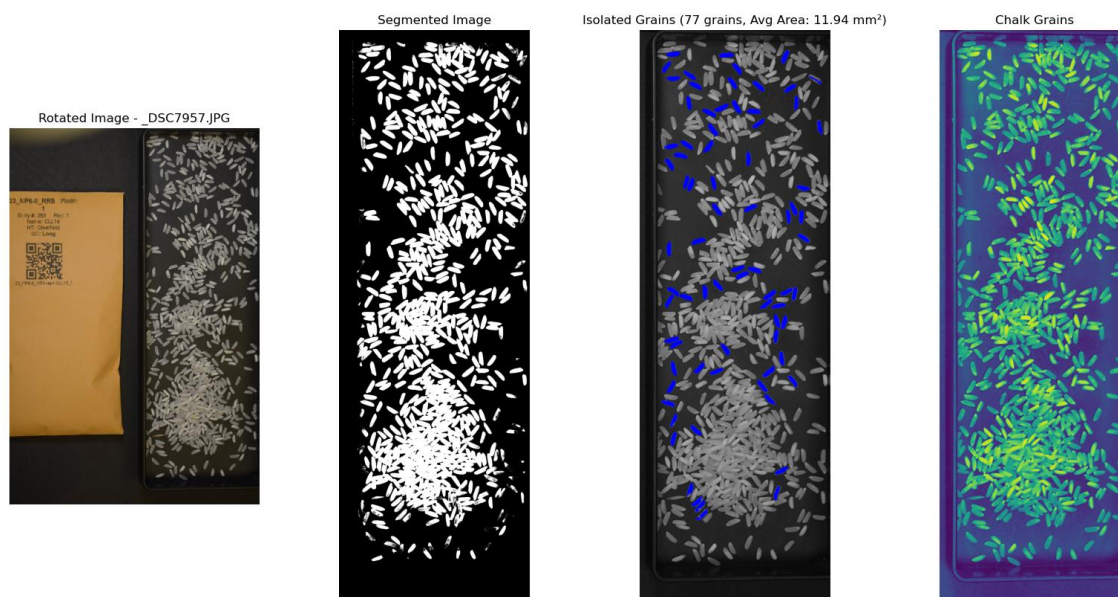


Figura 2. Processamento das imagens de arroz pela câmera RGB Nikon D7100.

Análise Quantitativa

Foi realizada a contagem do número de grãos de arroz, sendo os mesmos identificados isoladamente. Na sequência, foram realizados os registros das medidas dos eixos e a percentagem de grãos gessados, além da largura e comprimento dos grãos, estes dados foram da mediana das amostras, tudo feito via Python e suas bibliotecas necessárias.

Análise de Grãos Gessados

Aplicou-se um limiar específico de 203 para a identificação dos grãos gessados. Esse valor foi utilizado para segmentar as regiões mais claras da imagem, destacando os grãos com maior intensidade de branco, característica visual típica do gessamento. A partir disso, foi realizado o cálculo da área total dos grãos segmentados. Esse limiar permitiu diferenciar os grãos gessados dos grãos normais com base na variação dos tons de cinza, favorecendo a separação por refletância. Em seguida, determinou-se a porcentagem da área correspondente aos grãos gessados em relação à área total dos grãos identificados.

Os dados processados foram compilados em um arquivo Excel, incluindo o número de grãos isolados, as medidas dos eixos, a área total e a porcentagem de grãos gessados para cada imagem avaliada.

Validação

Realizou-se uma análise de correlação linear por Person para comparação dos dados obtidos pelas imagens hiperespectrais e RGB, sendo para tanto, utilizada a mesma plataforma que a estação de pesquisa obtém as porcentagens de grãos gessados. Neste processo, é necessária a separação de cada grão individualmente, sendo essa etapa realizada manualmente por meio de uma bandeja com 1000 poços. Na sequência, é realizada a captura da imagem em equipamento *Seedcount* para cálculo da porcentagem de gessamento das amostras. Esse processo é realizado de forma individual e manual, para cada amostra, com a duração de cerca de 1,5 minuto por leitura.

Por fim, o script utilizado para desenvolver a nova plataforma de análise via imagens foi compilado e está disponível no material suplementar.

Resultados e Discussão

A calcinação é a parte opaca do grão de arroz moído e é um dos principais fatores que determinam a qualidade do grão de arroz. Mais especificamente, a calcinação é a aparência visual de grânulos de amido frouxamente compactados (Ashida et al., 2009; Wang et al., 2017). A embalagem incorreta dos grânulos de amido leva a um aumento no número de bolsas de ar superdimensionadas dentro do grão. As bolsas de ar impedem a reflexão, dando às porções calcárias dos grãos uma aparência opaca (Tashiro & Wardlaw, 1991). A calcinação é uma característica indesejável, e uma proporção maior de giz leva a uma diminuição linear no valor de mercado do arroz (Lyman et al., 2013). Além disso, altos níveis de giz levam ao aumento da quebra durante a moagem e degradam as propriedades de cozimento e menor palatabilidade (Ashida et al., 2009; Fitzgerald et al., 2009; Lisle et al., 2000; Lyman et al., 2013).

Com essas características indesejadas, as cores dos grãos gessados, diferenciam-se nitidamente dos grãos normais, podendo ser facilmente identificados pelo olho humano, sem a necessidade de equipamentos

sofisticados e caros, nem procedimentos em HTP como características de cores de flores por exemplo (Souza et al., 2022). Essa característica possibilita o algoritmo identificar os grãos gessados sem dificuldade. Mesmo que sejam identificados facilmente ao olho humano, sem necessidade de equipamentos, mas é necessário que seja quantificado. E pelas características intrínsecas dos grãos de arroz, essa tarefa é difícil e demorada.

Neste contexto, o algoritmo desenvolvido apresentou um bom resultado na segmentação das imagens dos grãos de arroz diretamente em bandejas, sendo possível discriminar as imagens dos grãos via imagens de fundo de cor preta. Isso possibilitou a extração dos grãos nas imagens e então realizar a medição da área total de grãos, área de grãos gessados, porcentagem de grãos gessados, e os eixos maior e menor medianos de grãos de cada amostra. Com o arquivo gerado pelo *script*, que contém essas variáveis, foi possível realizar uma análise de correlação entre os dados gerados pelo *Seedcount* com os dados fornecidos pela imagem Hiperespectral, que apresentaram uma correlação positiva de 92% para grãos gessados, 89% para comprimento de grão e 68% para largura de grão (Figura 3).

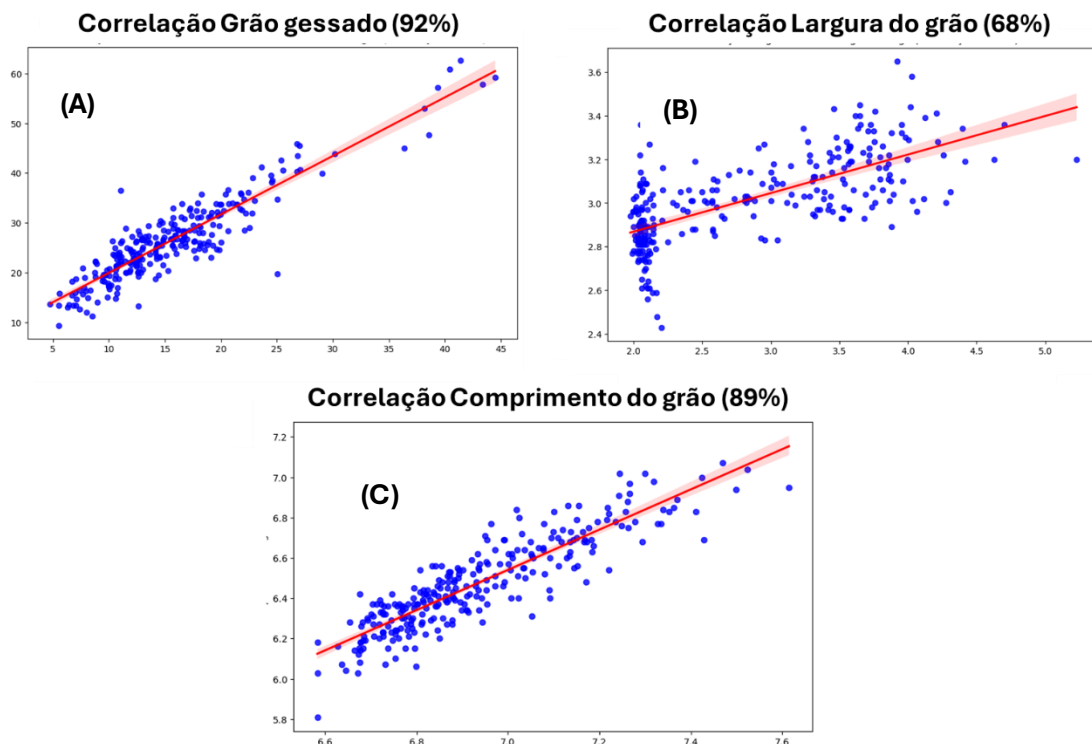


Figura 3. correlações entre dados do *seedcount* e os dados da câmera hiperespectral. (a) correlação entre gessamento antigo e gessamento novo (92%); (b) correlação entre largura antiga e largura nova (68%); (c) correlação entre comprimento antigo e comprimento novo (89%).

Por meio de uma câmera RGB Nikon® foi possível obter uma correlação de 75% para gessamento de grão, 93% para comprimento e 94% para largura. Observou-se que houve mudança entre as plataformas *Seedcount* e as imagens RGB-Nikon (Figura 4). A câmera RGB-Nikon obteve menor percentagem para gessamento dos grãos, enquanto a hiperespectral apresentou a maior percentagem para essa característica. Isso provavelmente se deve à qualidade de luz que as amostras recebem, permitindo que a câmera consiga captar melhor a cor das amostras. Além disso, as imagens obtidas pela câmera hiperespectral são mais eficientes, pois as mesmas são obtidas pelo mesmo processo de escaneamento, fornecendo assim melhor qualidade.

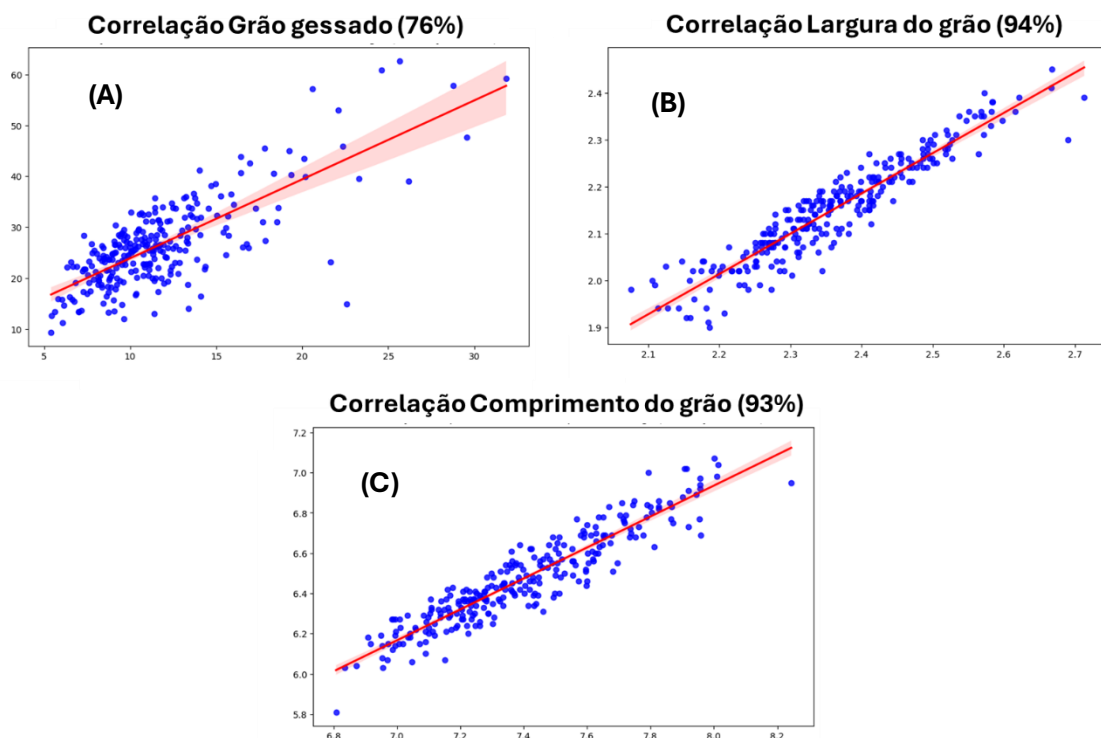


Figura 4. Correlações entre dados do *Seedcount* e câmera Nikon DSRL/RGB. (A) correlação entre gessamento antigo e gessamento novo (76%); (B) correlação entre largura antiga e largura nova (94%); (C) correlação entre comprimento antigo e comprimento novo (93%).

Assim, observamos que a correlação entre os dados de gessamento da câmera RGB com os dados da *Seedcount* (Figura 4A) foi menor quando comparada com a correlação entre os dados da câmera hiperespectral com os dados da *Seedcount* (Figura 3A). Mesmo assim é um resultado considerado satisfatório, pois a correlação obtida foi de 75% entre as imagens da câmera RGB e os dados do *Seedcount*. Com relação ao custo-benefício, observa-se uma grande vantagem, tendo em vista que o custo aproximado desta plataforma (RGB) é de cerca de 1500 dólares, sendo muito inferior quando comparado ao *Seedcount*, que custa cerca de 40 mil dólares, e inferior quando comparado ao valor da câmera hiperespectral, 45 mil dólares. Com relação ao tempo de obtenção das leituras, também se observa uma grande vantagem do novo método proposto em relação às demais, já que o *Seedcount* demora 1,5 minutos para leitura de cada amostra, a câmera hiperespectral demora cerca de 3 minutos para preparo e leitura das amostras, quanto a câmera comum utilizada necessita de 21 segundos para preparar e capturar a imagem. Estes resultados são extremamente relevantes, pois para um programa de melhoramento como o de arroz da LSU AgCenter, cerca de 10 mil amostras precisam ser avaliadas por safra. Assim, a *Seedcount* demoraria mais de 3 meses para as leituras, enquanto, a nova plataforma com a câmera comum, levaria aproximadamente 9 dias. Além disso, caso o número das plataformas com a câmera comum possa ser aumentado, pois o custo por cada uma é baixo comparado aos demais métodos, esse tempo seria ainda mais reduzido.

Diante destes resultados, nota-se que a HTP desempenha um papel fundamental na modernização dos programas de melhoramento genético, pois fornece dados essenciais para abordagens como Seleção Genômica (GS) e Estimativa Genômica (GP), que já vêm sendo amplamente utilizadas para otimizar a seleção de cultivares (Crossa et al., 2017). Essas técnicas permitem prever o desempenho das plantas em estágios iniciais de crescimento, o que acelera a identificação de genótipos mais produtivos e resistentes (Cabrera-Bosquet et al., 2012). No entanto, a eficácia dessas abordagens depende de dados fenotípicos de alta qualidade. Assim, o desenvolvimento de métodos rápidos e acessíveis, como o proposto neste estudo, pode ser crucial para garantir a precisão da estimativa genômica, reduzindo custos e tempo de análise.

Muitos métodos de estimativa utilizam modelos avançados e sofisticados a exemplo da inteligência artificial, por meio de *Machine Learning* ou *Deep Learning*. Contudo, no presente estudo, foi possível obter resultados satisfatórios usando um algoritmo relativamente simples e de baixo custo, por meio da separação dos grãos pela tonalidade de cores diferentes, utilizando uma regra de subtração entre área total de grãos e grãos gessados. Mesmo que o uso de modelos simples, ou seja, modelos que consideram apenas uma parte dos dados, possa resultar na perda de características importantes dos dados, causando perda de precisão para características mais complexas (Crossa et al., 2017; Cuevas et al., 2016, 2017; Pérez-Rodríguez et al., 2012), tal fato não ocorreu no presente estudo. Além disso, as características avaliadas no presente estudo não são complexas, e o procedimento segue a mesma lógica de um equipamento já estabelecido, no caso o *Seedcount*. Assim, o método proposto neste estudo simplifica a obtenção de imagens e é capaz de gerar dados com grande qualidade, num curto espaço de tempo, sendo a limitação da plataforma *Seedcount*, utilizada como padrão no presente estudo.

Diferentes abordagens de fenotipagem digital têm sido aplicadas para avaliar características estruturais das plantas em diferentes culturas. Em trigo, por exemplo, modelos de elevação digital (DEM) gerados por veículos aéreos não tripulados (UAS) foram utilizados para avaliar o acamamento das plantas, demonstrando correlações elevadas com os métodos tradicionais (Singh et al., 2019). No entanto, para a fenotipagem de grãos de arroz, abordagens como essa não são aplicáveis, pois, a escala e os parâmetros avaliados são distintos. Assim, a segmentação de imagens RGB surge como uma alternativa mais específica e acessível para a análise de grãos, permitindo a mensuração automatizada de características como gessificação, comprimento e largura.

O uso de algoritmos de processamento de imagem tem se mostrado eficaz para a extração automatizada de características morfológicas em diversas culturas. No estudo de Yang et al. (2018), por exemplo, foram utilizados algoritmos para calcular a largura do caule e a altura das plantas a partir de imagens obtidas no campo, alcançando erros médios de 13% e 15%, respectivamente. De forma semelhante, neste estudo, o algoritmo desenvolvido permitiu a segmentação e medição dos grãos de arroz com alta acurácia,

evidenciando a eficácia da fenotipagem digital na caracterização fenotípica dos grãos. Embora o presente estudo tenha sido conduzido em ambiente laboratorial, a possibilidade de adaptação desse método para análises automatizadas em larga escala reforça seu potencial como alternativa viável para programas de melhoramento genético.

O nosso método obtém as medidas de forma simples, no caso do gessamento é uma subtração de área. Calculou-se a área total dos grãos e subtraiu as áreas gessadas, encontrando uma estimativa do gessamento das amostras de grãos de arroz. Contudo, existem métodos mais complexos. Tem se adotado o uso de redes neurais convolucionais (CNNs), algoritmos de *Deep Learning* (como o YOLOv3) e modelos de classificação para identificar características específicas das plantas, como a qualidade de sementes, contagem de frutas e detecção de doenças Abebe et al., (2023) .

Além do uso de técnicas de imagem hiperespectral assim como a técnica que utilizamos, algumas técnicas destacam-se na avaliação de doenças foliares e no estresse hídrico, com análises conduzidas tanto no campo quanto em laboratórios utilizando UAVs e plataformas automatizadas de fenotipagem (Kim et al. 2015; Zhang et al. 2022; Abdulridha et al. 2020; Abebe et al., 2023; Lizarazo et al. 2023; Xie & Yang, 2020). A tecnologia de imagem visível (RGB) associada à técnica de *Machine Learning* e demais técnicas como: imagem térmica, LiDAR, e imagem de tomografia computadorizada (CT) são frequentemente utilizadas para capturar caracteres morfológicos e fisiológicos, como forma, tamanho, textura de superfície e resposta fotossintética das plantas (Du et al., 2020; Abebe et al., 2023). Cada técnica oferece vantagens para a avaliação não destrutiva e de alta precisão, especialmente em ambientes controlados e em campo aberto. Ilakiya et al., (2020) utilizaram técnicas de fenotipagem baseadas em sensores e imagens para medir o crescimento e as características fenotípicas das culturas como tomate, alface e espinafre, utilizando diferentes técnicas de fenotipagem. O método incluiu o uso de imagens hiperespectrais, espectroscopia no infravermelho próximo (NIR) e tomografia por ressonância magnética para medir caracteres como crescimento, conteúdo de clorofila e maturidade. Além disso, o estudo utilizou *Machine Learning* para análise de imagens, ajudando na classificação de qualidade das sementes e na detecção de doenças nas folhas.

Isso tudo combinado com medições manuais de campo para validar os dados de fenotipagem automática, assegurando a precisão das análises.

Existem diversos métodos para a fenotipagem de culturas, e trabalhos como o aqui descrito, podem ser adaptados para diferentes espécies, desde que sejam feitos os ajustes necessários. Nosso objetivo foi desenvolver técnicas mais simples e precisas, que também reduzam custos e tempo de execução, incluindo o tempo necessário para criação de novos métodos. Espera-se, ao final, que seja possível estabelecer uma plataforma estável, acessível online, que minimize ainda mais os gastos ao utilizar imagens RGB capturadas por celulares. Para alcançar esse objetivo, será necessário realizar testes adicionais, ajustes e o desenvolvimento de aplicativos que possam desempenhar essa função de forma precisa em tempo real.

Conclusão

O método proposto se mostrou eficiente e acessível para a identificação de grãos gessados com imagens RGB simples. A alta correlação com os dados do *Seedcount* indica que a abordagem é eficaz para características menos complexas, tornando-se uma alternativa prática e economicamente viável para programas de melhoramento. Consequentemente, o método proposto contribui para a agilidade e sustentabilidade dos programas de melhoramento, sendo uma ferramenta valiosa para desenvolver cultivares com melhor qualidade de grão e mais resilientes às condições adversas impostas pelas mudanças climáticas.

Referências Bibliográficas

Abdulridha, J.; Bossenda, J.; Major, R. Hyperspectral imaging for tomato yellow leaf curl disease detection. **Precision Agriculture**, v. 21, p. 156–170, 2020.

Abebe, A. M. et al. Image-based high-throughput phenotyping in horticultural crops. **Plants**, v. 12, n. 2061, p. 1–23, 2023.

Ann, P. et al. Genotype-dependent and heat-induced grain chalkiness in rice correlates with the expression patterns of starch biosynthesis genes. **Plant Direct**, v. 5, n. 11, e362, 2021.

Aroca, A. A. et al. Mathematics in the preparation and sale of Colombian rice pies and its potential contribution to mathematics education. **Acta Scientiae**, 2023.

Ashida, K.; Iida, S.; Yasui, T. Morphological, physical, and chemical properties of grain and flour from chalky rice mutants. **Cereal Chemistry**, v. 86, n. 2, p. 225–231, 2009. Disponível em: <https://doi.org/10.1094/CCHEM-86-2-0225>.

Bheemanahalli, R. et al. Temperature thresholds for spikelet sterility and associated warming impacts for sub-tropical rice. **Agricultural and Forest Meteorology**, v. 221, p. 122–130, 2016. Disponível em: <https://doi.org/10.1016/j.agrformet.2016.02.003>.

Crossa, J. et al. Genomic selection in plant breeding: methods, models, and perspectives. **Trends in Plant Science**, v. 22, n. 11, p. 961–975, 2017.

Cuevas, J. et al. Bayesian genomic prediction with genotype × environment interaction kernel models. **G3: Genes, Genomes, Genetics**, v. 7, n. 1, p. 41–53, 2017. Disponível em: <https://doi.org/10.1534/g3.116.035584>.

Cuevas, J. et al. Genomic prediction of genotype × environment interaction kernel regression models. **The Plant Genome**, v. 9, n. 3, 2016. Disponível em: <https://doi.org/10.3835/plantgenome2016.03.0024>.

Du, X.; Li, Y.; Zhou, G. High-throughput phenotyping platform in greenhouses for lettuce trait analysis. **Journal of Plant Science**, v. 15, p. 234–250, 2021.

Fitzgerald, M. A.; McCouch, S. R.; Hall, R. D. Not just a grain of rice: The quest for quality. **Trends in Plant Science**, v. 14, n. 3, p. 133–139, 2009. Disponível em: <https://doi.org/10.1016/j.tplants.2008.12.004>.

Hlisnikovský, L. et al. How weather and fertilization affected grain yield and stability of winter wheat in a long-term trial in the South Moravian Region, Czech Republic. **Agronomy**, v. 13, n. 9, 2023. Disponível em: <https://doi.org/10.3390/agronomy13092293>.

Huo, X. et al. Genome-wide association mapping and gene expression analysis reveal candidate genes for grain chalkiness in rice. **Frontiers in Plant Science**, v. 14, 2023.

Ilakiya, T. et al. High-throughput crop phenotyping in vegetable crops. **The Pharma Innovation Journal**, v. 9, n. 8, p. 184–191, 2020.

Jagadish, S.; Craufurd, P.; Wheeler, T. High temperature stress and spikelet fertility in rice (*Oryza sativa* L.). **Journal of Experimental Botany**, v. 58, n. 7, p. 1627–1635, 2007. Disponível em: <https://doi.org/10.1093/jxb/erm003>.

Kim, Y.; Rice, T.; Zhu, J. Hyperspectral analysis for water stress indices in plants using SWIR images. **Plant and Soil Analysis**, v. 43, p. 376–390, 2015.

Li, R. et al. Exploring the relationships between yield and yield-related traits for rice varieties released in China from 1978 to 2017. **Frontiers in Plant Science**, v. 10, 2019. Disponível em: <https://doi.org/10.3389/fpls.2019.00543>.

Li, Y. et al. Chalk5 encodes a vacuolar H⁺-translocating pyrophosphatase influencing grain chalkiness in rice. **Nature Genetics**, v. 46, p. 398–404, 2014.

Lisle, A. J.; Martin, M.; Fitzgerald, M. A. Chalky and translucent rice grains differ in starch composition and structure and cooking properties. **Cereal Chemistry**, v. 77, n. 5, p. 627–632, 2000. Disponível em: <https://doi.org/10.1094/CCHEM.2000.77.5.627>.

Lizarazo, M.; Garcia, F.; Rios, J. Multispectral UAV imaging for Verticillium wilt detection in potatoes. **Remote Sensing Applications in Agriculture**, v. 31, p. 74–89, 2023.

Lyman, N. B. et al. Neglecting rice milling yield and quality underestimates economic losses from high-temperature stress. **PLOS ONE**, v. 8, n. 8, 2013. Disponível em: <https://doi.org/10.1371/journal.pone.0072157>.

Mittal, S.; Dutta, M.; Issac, A. Non-destructive image processing-based system for assessment of rice quality and defects for classification according to inferred commercial value. **Measurement**, v. 148, 2019. Disponível em: <https://doi.org/10.1016/j.measurement.2019.106969>.

Nevame, A. Y. M. et al. Relationship between high temperature and formation of chalkiness and their effects on quality of rice. **BioMed Research International**, v. 2018, p. 1653721, 2018.

Pérez-Rodríguez, P. et al. Comparison between linear and non-parametric regression models for genome-enabled prediction in wheat. **G3: Genes|Genomes|Genetics**, v. 2, n. 12, p. 1595–1605, 2012. Disponível em: <https://doi.org/10.1534/g3.112.003665>.

Singh, D. et al. High-throughput phenotyping enabled genetic dissection of crop lodging in wheat. **Frontiers in Plant Science**, v. 10, p. 394, 2019. Disponível em: <https://doi.org/10.3389/fpls.2019.00394>.

Smiderle, O. J., & Dias, J. R. M. (2008). Época de colheita e qualidade fisiológica de sementes em arroz irrigado (*Oryza sativa* L.) cv. BRS Roraima. **REVISTA BRASILEIRA DE SEMENTES**, 30(3), 123–131.

Souza, J. S. et al. The more fractal the architecture, the more intensive the color of flower: A superpixel-wise analysis towards high-throughput phenotyping. **Agronomy**, v. 12, n. 6, 2022. Disponível em: <https://doi.org/10.3390/agronomy12061342>.

Sreenivasulu, N. et al. Designing climate-resilient rice with ideal grain quality suited for high-temperature stress. **Journal of Experimental Botany**, v. 66, n. 7, p. 1737–1748, 2015. Disponível em: <https://doi.org/10.1093/jxb/eru544>.

Tashiro, T.; Wardlaw, I. F. The effect of high temperature on kernel dimensions and the type and occurrence of kernel damage in rice. **Australian Journal of Agricultural Research**, v. 42, n. 3, p. 485–496, 1991. Disponível em: <https://doi.org/10.1071/ar9910485>.

Wang, K. et al. On the asymmetry of the urban daily air temperature cycle. **Journal of Geophysical Research: Atmospheres**, v. 122, n. 11, p. 5625–5635, 2017. Disponível em: <https://doi.org/10.1002/2017JD026589>.

Young, S. N.; Kayacan, E.; Peschel, J. M. Design and field evaluation of a ground robot for high-throughput phenotyping of energy sorghum. **Precision Agriculture**, v. 20, p. 697–722, 2019.

Zhang, W.; Chen, L.; Liu, M. Detection of Fusarium wilt in banana using multispectral UAV methods. **Agricultural Remote Sensing Journal**, v. 28, p. 102–118, 2022.

Capítulo 3: Uso Integrado de *Machine Learning* e *Deep Learning* no Melhoramento Genético de Soja: Classificação de Genótipos e Estimativa de Valor Agronômico

Resumo

A inteligência artificial (IA) revolucionou a pesquisa agrícola, especialmente no melhoramento de plantas. No melhoramento de soja, técnicas de IA, como Machine Learning (ML) e Deep Learning (DL), são utilizadas para prever características fenotípicas, otimizar a seleção de genótipos e melhorar a previsão de produtividade. No entanto, a maioria dos estudos utiliza ML ou DL separadamente, o que limita a precisão preditiva e a adaptabilidade em diferentes condições ambientais. Este estudo propõe um Sistema Preditivo Híbrido (Hybrid Predictive System - HPS) que integra ML (Random Forest) para classificação de genótipos e DL (Redes Neurais Profundas) para a previsão do valor agronômico. A abordagem visa reduzir o viés humano, aumentar a precisão da seleção e melhorar a robustez da avaliação de genótipos sob diferentes condições ambientais. Um conjunto de dados com 281 genótipos de soja foi analisado, classificando-os em três grupos: 'Excelente', 'Intermediário' e 'Menos promissores'. O modelo Random Forest atingiu acurácia global de 95,8%, com F1-score médio de 0,95, demonstrando elevado desempenho na classificação. Para a estimativa do valor agronômico com DL, os melhores resultados foram observados no ano agrícola 2021/2022, com correlação de 92,9%, R^2 de 86,35% e MAE de 0,2113. Já para 2022/2023, os valores foram correlação de 82,9%, R^2 de 68,7% e MAE de 0,2762. Considerando a média dos dois anos, o modelo obteve correlação de 81,%, R^2 de 65,87% e MAE de 0,2098. O modelo proposto demonstra o potencial da IA para acelerar os programas de melhoramento de soja, otimizar a seleção de genótipos e contribuir para uma agricultura mais precisa e sustentável.

Palavras-chave: Random Forest, CNNs, melhoramento de soja, seleção de genótipos, modelagem preditiva, sistemas híbridos de IA.

Introdução

Com os avanços da tecnologia em diversas áreas, a Inteligência Artificial (IA) é uma das ferramentas mais poderosas nos avanços técnicos e científicos da atualidade, em vários campos de utilização. Sua aplicação abrange desde sistemas autônomos, como veículos sem condutores (Moeinizada et al., 2021) até assistentes virtuais baseados em redes neurais profundas, que permitem a execução de tarefas como ligações telefônicas via processamento de linguagem natural (NLP) (Barbedo, 2023). Além disso, a IA tem sido amplamente utilizada para análise de padrões e preferências em sistemas de recomendação, como os empregados por plataformas de streaming de música e vídeo (Liu et al., 2019). No campo da astrofísica e exploração espacial, modelos de *DL* vêm sendo empregados para a análise de grandes volumes de dados astrofísicos e até para o controle de sondas espaciais não tripuladas em missões interplanetárias (Shook et al., 2018).

Na agricultura, a IA tem sido aplicada de diversas maneiras, auxiliando desde a detecção de doenças em plantas até a tomada de decisões em programas de melhoramento genético (Yoosefzadeh-Najafabadi et al., 2021). Técnicas avançadas de *ML* e *DL* têm sido utilizadas para modelar a produtividade das culturas, identificar genótipos superiores e prever interações *G x E* com alta precisão (Teodoro et al., 2021). Esse avanço representa uma revolução na chamada Agricultura Digital e coloca o melhoramento genético no conceito de "Melhoramento Genético 4.0", no qual dados genômicos, fenotípicos e ambientais são integrados para maximizar a eficiência da seleção (Xavier, 2016).

No melhoramento genético de soja, as avaliações desempenham um papel essencial no desenvolvimento de cultivares com características agrônomicas superiores, como maior produtividade de grãos, resistência ao acamamento e altos teores de óleo e proteína. Além destes caracteres mencionados, o valor agrônômico (VA) é um índice subjetivo geral atribuído a cada genótipo com base em suas características visuais fenotípicas, por meio

de uma escala de notas, apresentando grande importância no processo seletivo. No entanto, essa subjetividade pode apresentar viés humano, comprometendo a precisão do processo seletivo de genótipos superiores e, conseqüentemente, a eficiência do programa de melhoramento (Yoosefzadeh-Najafabadi et al., 2021).

Embora essas técnicas tenham demonstrado resultados promissores, poucos estudos integram Machine Learning (ML) e Deep Learning (DL) de forma complementar para classificação e estimativa simultânea no melhoramento genético (Zhang et al., 2022). Neste contexto, este estudo propõe um sistema preditivo híbrido que combina a robustez de ML para classificação de genótipos e a capacidade de DL para estimativa contínua de variáveis agronômicas, especialmente o valor agrônomo. Espera-se que essa integração ofereça uma abordagem mais abrangente para capturar tanto padrões lineares quanto não lineares, permitindo maior precisão na tomada de decisões e mais eficiência na identificação de genótipos superiores (Teodoro et al., 2021).

Além disso, um dos desafios críticos no melhoramento genético é a influência das interações genótipo x ambiente ($G \times E$), que afetam diretamente a estabilidade do desempenho dos genótipos em diferentes condições edafoclimáticas (Uzal et al., 2018). A proposta de um modelo híbrido ML+DL pode permitir melhores resultados e menores influências das variações ambientais, tornando a seleção mais robusta e adaptável a diferentes cenários experimentais. Essa abordagem inovadora preenche uma lacuna na literatura científica, pois a maioria dos estudos se concentra em apenas um tipo de modelagem preditiva, seja ML ou DL isoladamente (Fernandes et al., 2024). O desenvolvimento de um modelo integrado proporciona maior acurácia preditiva, reduz o viés humano na avaliação fenotípica, advinda de diferentes interpretações e pode consistir em uma ferramenta poderosa para programas de melhoramento genético mais eficientes (Moeinizada et al., 2021).

O presente estudo tem como objetivo avaliar o potencial do uso de técnicas de inteligência artificial em programas de melhoramento genético de soja, por meio de abordagens complementares de *ML* e *DL*, visando aumentar a eficiência no processo seletivo de genótipos superiores.

Material e Métodos

Os experimentos foram conduzidos na Fazenda de Ensino Pesquisa e Extensão (FEPE) da Faculdade de Ciências Agrárias e Veterinárias – Universidade Estadual Paulista, Campus de Jaboticabal, entre os anos de 2021/2022 e 2022/2023. O município está localizado na região noroeste do estado de São Paulo, a uma latitude de 21° 14' 59" Sul, longitude 48° 17' 8" Oeste, com 575 metros de altitude. O solo do local é classificado como um latossolo vermelho-escuro eutrófico, textura muito argilosa e relevo suave ondulado. De acordo com Köppen (1948), o clima da região é do tipo Cwa, apresentando verão quente e úmido, inverno seco, temperatura média anual de 22,2°C e precipitação média de 1,451 mm (Andrioli; Centurion, 1999; Vianna *et al*, 2013).

A condução dos experimentos ocorreu nos anos agrícolas de 2021/2022 e 2022/2023, com a semeadura realizada no sistema de plantio direto sobre palhada de milho, dentro do período estabelecido pelo Zoneamento Agrícola de Risco Climático (ZARC) para a soja no estado de São Paulo, que compreende, de forma geral, o intervalo entre 16 de setembro e 31 de dezembro (MAPA, 2024). Todas as práticas agrícolas, incluindo adubação, manejo, tratos culturais e controle fitossanitário, foram conduzidas conforme as recomendações técnicas para a cultura (EMBRAPA, 2023).

Foram avaliados 281 genótipos em geração F₉ de endogamia, avaliando 5 plantas de cada genótipo. Os experimentos foram conduzidos no delineamento experimental de blocos aumentados, com 4 testemunhas intercalares MG/BR 46 (Conquista) e as linhagens superiores 23, 54 e 69. A unidade experimental foi composta por uma linha de cinco metros de comprimento, com espaçamento de 0,5 metro entre linhas, sendo a área útil formada pelos quatro metros centrais da linha, com uma densidade de aproximadamente 16 plantas por metro.

Caracteres Avaliados

Foram avaliadas cinco plantas individuais dentro de cada parcela para os seguintes caracteres: número de dias para a maturidade (NDM) - contagem em dias a partir da germinação até a maturidade, no estágio R8 (Fehr; Caviness, 1977); altura de planta na maturidade (APM), em centímetros, medido da superfície do solo até o ápice da haste principal; altura de inserção da primeira vagem (AIV), em centímetros, da superfície do solo até a inserção da primeira vagem, acamamento (ACM), escala de notas visuais, onde foram atribuídas notas de 1 a 5 de acordo com a seguinte escala: 1 - todas ou quase todas as plantas eretas; 2 – maioria das plantas levemente inclinadas ou até 25% das plantas acamadas; 3 - todas as plantas medianamente inclinadas ou de 25 a 50% das plantas acamadas; 4 - todas as plantas fortemente inclinadas ou de 50 a 80% das plantas acamadas; 5 - mais de 80% das plantas acamadas; valor agrônômico (VA), escala de notas visuais, variando de 1 (planta ruim) a 5 (planta excelente).

As avaliações realizadas após a colheita da área útil da parcela e limpeza das sementes foram: determinação de umidade; produção de grãos (PRD) mensurado em gramas com posterior correção da umidade das sementes para 13%, para obtenção do peso corrigido, sendo convertido em produtividade de grãos em kg ha⁻¹.

Avaliação do teor de óleo e proteína

A qualidade industrial dos genótipos de soja foi expressa pelos teores de óleo e de proteína nos grãos, em percentagem.

Os teores de porcentuais de proteína e óleo das amostras foram determinados no Laboratório de Biotecnologia e Melhoramento de Plantas pertencente ao Departamento de Produção Vegetal da UNESP/FCAV Jaboticabal. A determinação foi realizada em grãos de soja íntegros pela técnica da Reflectância do Infravermelho Próximo (NIR).

Foram tomadas amostras dos grãos inteiros e limpos, sendo aproximadamente 80 gramas de cada amostra, sendo as sementes analisadas pela reflectância difusa utilizando acessório de Rotação do próprio equipamento

Espectrômetro Modelo FT-NIR TANGO da marca Bruker. Ressalta-se que a curva de calibração e a sua validação foi realizada por Leite (2016).

Base de Dados para Random Forest (*Machine Learning*)

Foram utilizados os dados avaliados dos caracteres descritos anteriormente. Cada planilha representava um ano agrícola, sendo que os dados foram consolidados em um único conjunto de dados utilizando a biblioteca pandas (versão 1.5.3) do Python. Os nomes das colunas foram padronizados para remover espaços em branco, utilizando o método `str.strip()` fornecido pela mesma biblioteca.

Os caracteres analisados incluíram:

- Altura (cm): Altura de planta na maturidade
- AIV (cm): Altura de inserção de vagem
- Acamamento (escala): Resistência ao acamamento.
- Produção (kg ha⁻¹): Produtividade média.
- Valor Agronômico (escala): Avaliação geral da adaptabilidade.
- Teor de Óleo (%): Concentração média de óleo nos grãos.
- Teor de Proteína (%): Concentração média de proteína nos grãos.

Seleção e Pré-Processamento dos Dados

Os dados foram agrupados por genótipo, calculando as médias de cada caractere ao longo dos anos agrícolas. Após o agrupamento, foram realizadas as seguintes etapas:

1. Seleção de caracteres:
 - As variáveis selecionadas como features independentes foram: altura, AIV, acamamento, produção, valor agronômico, teor de óleo e teor de proteína.

- A variável dependente consistiu na classificação do genótipo em três grupos: "Excelentes", "Intermediários" e "Menos Promissores".

2. Normalização:

- Os dados foram normalizados utilizando o método de padronização com o StandardScaler da biblioteca scikit-learn (versão 1.1.3), garantindo que os caracteres possuíssem média zero e desvio-padrão unitário.

3. Balanceamento de Classes:

- Para lidar com o desbalanceamento entre as classes, foi aplicado o método SMOTE (Synthetic Minority Oversampling Technique), que sintetiza exemplos adicionais para as classes minoritárias. Esta etapa foi realizada com a biblioteca imblearn (versão 0.10.1), por meio do módulo over_sampling.

Treinamento e Validação do Modelo

O modelo Random Forest foi utilizado como classificador para prever os grupos de genótipos. O treinamento seguiu os passos abaixo:

1. Random Forest:

- O modelo foi implementado com 100 árvores de decisão (`n_estimators=100`), utilizando a classe RandomForestClassifier da biblioteca scikit-learn.

- As árvores foram treinadas em subconjuntos amostrais dos dados, combinando previsões para melhorar a robustez e reduzir o risco de *overfitting*.

2. Validação Cruzada:

- Foi utilizada validação cruzada estratificada com 5 divisões (StratifiedKFold) para avaliar a performance do modelo.

- A estratificação garantiu que a proporção entre as classes fosse preservada em cada divisão. Esta funcionalidade também foi implementada com o módulo `model_selection` da biblioteca `scikit-learn`.

3. Métricas de Avaliação:

- As métricas reportadas incluem precisão, recall, F1-score e a matriz de confusão, avaliando a capacidade do modelo de classificar corretamente cada grupo. Essas métricas foram geradas utilizando o módulo `metrics` da biblioteca `scikit-learn`, com destaque para as funções `classification_report` e `confusion_matrix`.

Análise e Visualização

Os resultados do modelo foram analisados e exportados em formato Excel (.xlsx), contendo:

- A classificação final de cada genótipo.
- As médias das características agronômicas.
- As pontuações atribuídas para cada critério de avaliação.

Além disso, foi gerado um gráfico de agrupamento dos genótipos utilizando a Análise de Componentes Principais (PCA), com a função `PCA` do módulo `decomposition` da biblioteca `scikit-learn`. O gráfico de dispersão resultante mostrou a separação entre os grupos de genótipos, com cores distintas representando cada grupo. A visualização foi realizada com o uso das bibliotecas `matplotlib` (versão 3.6.2) e `seaborn` (versão 0.12.2), permitindo uma apresentação clara e interpretável dos dados.

Ambiente de Desenvolvimento

Neste estudo, foi utilizada a linguagem Python (versão 3.9.12) para implementar as análises e o modelo de classificação. Foram empregadas diversas bibliotecas, cada uma com funções específicas. A biblioteca `pandas` (versão 1.5.3) foi utilizada para a manipulação e organização dos dados, permitindo o processamento eficiente de tabelas e estruturas de dados. A

biblioteca numpy (versão 1.23.5) foi aplicada para realizar operações matemáticas e cálculos numéricos necessários no pré-processamento dos dados.

Para a construção, treinamento e validação do modelo de classificação, foi utilizada a biblioteca scikit-learn (versão 1.1.3). Nesta, foram implementados o algoritmo Random Forest, a padronização dos dados com o StandardScaler, técnicas de balanceamento com StratifiedKFold e validação cruzada, garantindo a robustez e a confiabilidade dos resultados. O balanceamento das classes foi aprimorado com o uso do método SMOTE (Synthetic Minority Oversampling Technique), implementado com a biblioteca imblearn (versão 0.10.1), que foi essencial para lidar com possíveis desbalanceamentos no conjunto de dados. Por fim, a biblioteca matplotlib (versão 3.6.2) foi empregada para a criação de visualizações gráficas, como a matriz de confusão, gráficos de dispersão e de importância das características, permitindo uma interpretação clara dos resultados e do desempenho do modelo.

Métricas de Avaliação para Classificação

O desempenho do modelo de classificação foi avaliado utilizando as métricas de Precisão, Recall e F1-Score. A Precisão mede a proporção de instâncias corretamente classificadas entre todas as instâncias previstas como positivas, enquanto o Recall reflete a proporção de instâncias positivas reais que foram corretamente identificadas pelo modelo. O F1-Score, calculado como a média harmônica entre a Precisão e o Recall, é particularmente valioso em cenários com classes desbalanceadas, pois integra ambas as métricas em uma única e robusta medida de desempenho da classificação. Essas métricas garantem uma avaliação abrangente da capacidade do modelo em identificar corretamente os genótipos nas diferentes classes (Excelentes, Intermediários e Menos Promissores).

Base de Dados para *Deep Learning*

Após a construção e validação do modelo de classificação com *ML*, a preparação dos dados para *DL* seguiu uma abordagem estruturada para garantir

a qualidade das informações utilizadas no treinamento da rede neural. Os dados utilizados neste estudo foram extraídos de um arquivo Excel contendo variáveis relacionadas ao desempenho agronômico de genótipos.

Os caracteres avaliados incluíram: altura de planta na maturidade, acamamento, altura de inserção de vagem, produtividade, teor de óleo e teor de proteína. Para garantir um pré-processamento eficiente, a base de dados foi carregada utilizando a biblioteca pandas (versão 1.5.3), amplamente utilizada para manipulação de dados estruturados em Python.

Seleção e Pré-Processamento dos Dados

Foram selecionadas as seguintes variáveis independentes como features: altura da planta na maturidade (APM), altura de inserção da primeira vagem (AIV), índice de acamamento (ACM), teor de proteína (PRO), teor de óleo (OLEO) e produtividade de grãos (PRD). A variável dependente foi o índice de valor agronômico (VA). Para normalizar as features, utilizou-se o método de padronização com StandardScaler da biblioteca scikit-learn, garantindo que as variáveis possuíssem média zero e desvio-padrão unitário.

Durante o pré-processamento, outliers na variável-alvo (valor agronômico) foram identificados e removidos com base no intervalo interquartil (IQR). Os cálculos dos quartis e a filtragem dos valores discrepantes foram realizados com a biblioteca pandas. Valores abaixo do limite inferior ($Q1 - 1,5 * IQR$) ou acima do limite superior ($Q3 + 1,5 * IQR$) foram excluídos, mitigando a influência de valores extremos no treinamento do modelo (Dash et al., 2023).

Divisão dos Dados

A escolha da divisão dos dados em 70% para treinamento e 30% para teste baseia-se em uma prática comum em estudos de aprendizado supervisionado, especialmente quando o número total de amostras não é extremamente elevado. Essa proporção busca equilibrar dois objetivos fundamentais: garantir que o modelo seja treinado com uma quantidade

suficiente de dados para aprender os padrões relevantes e, ao mesmo tempo, permitir uma avaliação confiável de seu desempenho em dados não utilizados durante o treinamento. Embora essa divisão não seja uma regra fixa, ela é amplamente adotada na literatura por apresentar bons resultados, principalmente em conjuntos com mais de 200 amostras.

No presente estudo, foram utilizadas informações de 281 genótipos de soja. Considerando essa quantidade total de amostras, aproximadamente 197 genótipos foram destinados ao treinamento do modelo, enquanto os 84 genótipos restantes compuseram o conjunto de teste. Essa divisão garante que o modelo seja exposto a uma diversidade suficiente de exemplos durante o treinamento, sem comprometer a avaliação de sua capacidade preditiva em novos dados.

É importante destacar que outras proporções também podem ser adotadas, dependendo da natureza e da dimensão do conjunto de dados. Em situações com grande volume de dados disponíveis, por exemplo, a divisão 80/20 pode ser mais adequada, enquanto em conjuntos menores, uma divisão 60/40 pode ser considerada para ampliar o poder estatístico da validação. Além disso, métodos alternativos como a validação cruzada do tipo *k*-fold — como a utilizada neste trabalho por meio do algoritmo `StratifiedKFold` — são especialmente recomendados por permitirem que todos os dados sejam utilizados tanto para treinamento quanto para validação, em diferentes combinações, aumentando a robustez das métricas avaliadas.

Construção do Modelo de Rede Neural

O modelo preditivo foi desenvolvido utilizando a API sequencial da biblioteca **TensorFlow/Keras**. A arquitetura da rede neural foi definida conforme segue:

- **Camada de entrada:** Número de neurônios igual ao número de variáveis independentes (6), especificado pelo argumento `Input(shape=(X_train.shape[1],))`.

- **Primeira camada oculta:** Composta por 64 neurônios, utilizando a função de ativação ReLU (Dense(64, activation='relu')).
- **Segunda camada oculta:** Composta por 32 neurônios, também utilizando a função de ativação ReLU (Dense(32, activation='relu')).
- **Camada de saída:** Único neurônio linear para estimativa contínua (Dense(1)).

O modelo foi compilado utilizando a função de perda Mean Squared Error (MSE) e o otimizador Adam, ambos fornecidos pela biblioteca tensorflow.keras, com taxa de aprendizado ajustada para 0,001.

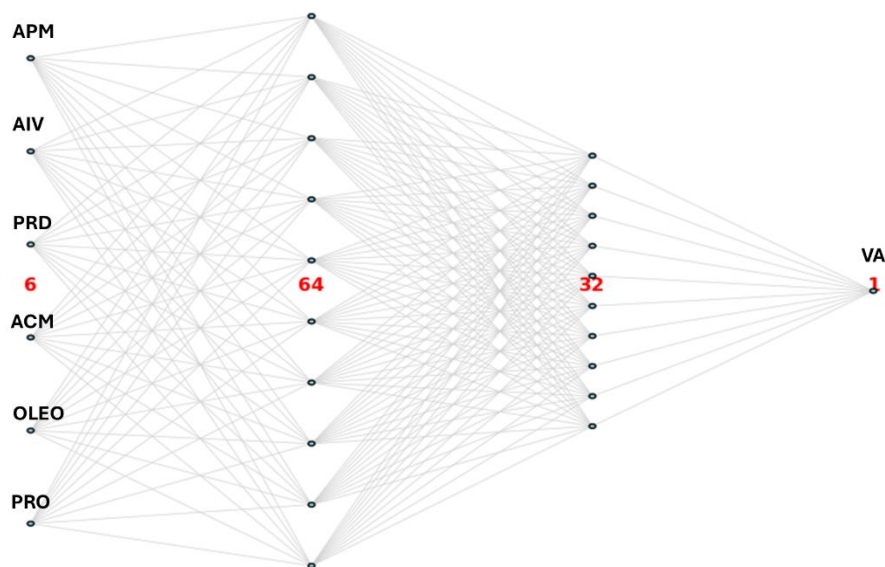


Figura 5. Representação gráfica da arquitetura da rede neural artificial, evidenciando as camadas de entrada, ocultas e de saída. A camada de entrada possui 6 neurônios, correspondendo às variáveis preditoras do modelo. A rede conta com duas camadas ocultas, contendo 64 e 32 neurônios, respectivamente, responsáveis pelo aprendizado das representações dos dados. A camada de saída possui 1 neurônio, indicando que o modelo realiza uma tarefa de regressão ou classificação binária.

1. **Treinamento Inicial:** O modelo foi treinado por até 300 épocas, utilizando validação no conjunto de teste, quantidade de épocas que melhor se adaptou ao modelo. Os pesos do modelo foram otimizados com base na métrica

de perda pelo erro quadrático médio (MSE). Uma época no treinamento de uma rede neural representa uma passagem completa por todo o conjunto de dados de treinamento. Ou seja, em uma única época, o modelo vê todas as amostras do conjunto de treinamento uma vez e ajusta os pesos das conexões entre os neurônios com base na função de erro.

Durante o treinamento, os dados são geralmente divididos em *minibatches*, que são pequenos subconjuntos do conjunto de treinamento, tornando o processo mais eficiente. Para cada minibatch, o modelo realiza uma previsão, calcula o erro por meio de uma métrica, como o erro quadrático médio (MSE), e ajusta os pesos das conexões utilizando o algoritmo de *backpropagation* em conjunto com um otimizador, como Adam (presente no módulo `tensorflow.keras.optimizers`) ou SGD. Quando todas as amostras do conjunto de treinamento são processadas uma vez, considera-se que uma época foi concluída. O treinamento continua por múltiplas épocas, neste caso, até 300 épocas, permitindo que o modelo aprenda padrões complexos nos dados e melhore sua capacidade de generalização.

2. **Monitoramento de Perdas:** As perdas de treinamento e validação foram acompanhadas durante o processo de ajuste, dessa forma, permitindo identificar possíveis problemas como sobreajuste ou subajuste. Para isso, foi utilizada a funcionalidade `EarlyStopping` da biblioteca `tensorflow.keras.callbacks`, que interrompe o treinamento caso a perda de validação não melhore após um número pré-estabelecido de épocas.

Avaliação do Modelo

Após o treinamento, o modelo foi avaliado no conjunto de teste, e as seguintes métricas foram calculadas: erro médio absoluto (MAE), coeficiente de determinação (R^2) e correlação. As duas primeiras métricas foram obtidas por meio das funções `mean_absolute_error` e `r2_score`, disponíveis no módulo `sklearn.metrics`. A correlação entre os valores preditos e observados foi calculada utilizando a função `corr()` da biblioteca `pandas`.

Resultados e Discussão

Machine Learning na classificação dos genótipos de soja

Os resultados do modelo indicaram uma alta precisão na classificação dos genótipos, com uma acurácia geral de 95% (Tabela 1). O método Random Forest, aliado ao balanceamento de classes e validação cruzada, demonstrou ser uma abordagem eficiente para a classificação de genótipos de soja. O arquivo Excel gerado permite uma análise mais detalhada: Intermediários (189 amostras, 68.8%), Menos Promissores (67 amostras, 24.4%) e Excelentes (25 amostras, 9.1%). Essa distribuição reflete a prevalência de genótipos intermediários em relação aos extremos, um cenário típico em estudos de melhoramento genético. Tais resultados encontram-se representados na Figura 2, onde observa-se que a ilustração dos genótipos é bastante visível, podendo ser uma ferramenta prática para a tomada de decisão em programas de melhoramento genético. Pois, com isso pode se selecionar melhor os genótipos superiores.

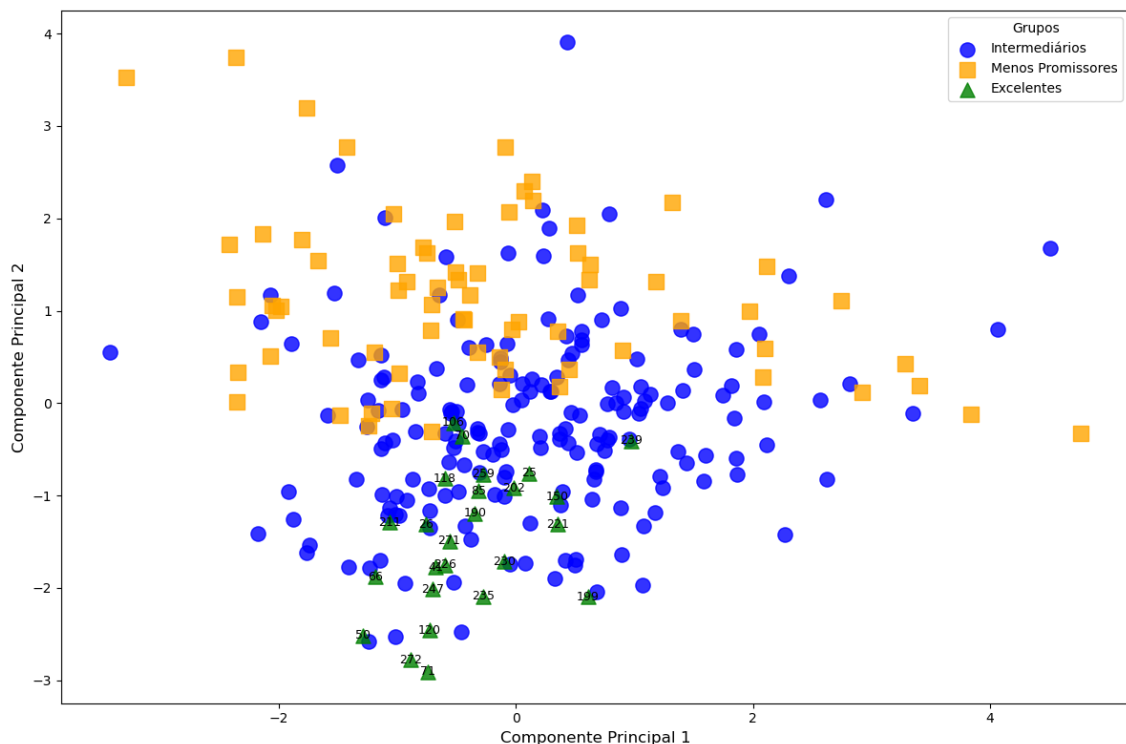


Figura 6. Agrupamento de genótipos por classificação de desempenho, com destaque para os genótipos classificados como Excelentes (25 amostras).

Esse algoritmo é utilizado com frequência em alguns estudos na agricultura, os quais podem ser aplicados ao melhoramento genético de plantas. Por exemplo, no estudo realizado com a cultura da soja, os resultados indicaram que o algoritmo Random Forest, quando combinado com dados espectrais de soja, foi altamente eficiente para classificar genótipos em relação ao conteúdo de macronutrientes secundários como o Cálcio, Magnésio e Enxofre (Santana et al., 2024). Em um outro estudo, diferentes configurações de *ML*, que incluem Random Forest, foram comparadas para classificar genótipos de soja com base no teor de Nitrogênio, Fósforo e Potássio (Santana et al., 2023). Em ambos os estudos, o algoritmo foi capaz de classificar genótipos que tiveram diferentes adubações. Assim, considera-se que, se os genótipos puderam ser diferenciados e classificados de forma eficiente com base na nutrição, indica que, para características fenotípicas, ele também poderá apresentar bom desempenho. Em um outro estudo, também com a cultura da soja, o uso de *ML*, incluindo Random Forest, permitiu classificar genótipos de soja com base em sua composição química e características espectrais, destacando o potencial dessas abordagens para seleção em programas de melhoramento (Bernardy et al., 2024).

O modelo Random Forest também tem sido amplamente utilizado para a estimativa de produtividade da soja com dados de refletância hiperespectral. Em estudo conduzido por Yoosefzadeh-Najafabadi et al. (2021), o Random Forest obteve uma precisão de 84%, e quando combinado com técnicas de *ensemble* (empilhamento), a precisão aumentou para 93%. No estudo aqui proposto, o presente modelo alcançou uma acurácia de 95% (Tabela 1), superando os valores reportados na literatura para Random Forest sem *ensemble*, sugerindo que o modelo utilizado foi eficiente na classificação dos genótipos.

Tabela 1. Métricas de Avaliação do Desempenho do Modelo de Classificação para os genótipos de soja. Jaboticabal, SP.

Classe	Precisão (%)	Recall (%)	F1-Score	Suporte
--------	--------------	------------	----------	---------

Excelentes	96	97	0.97	185
Intermediários	94	91	0.93	185
Menos Promissores	95	97	0.96	185
Acurácia	-	-	0.95	555
Média Macro	95	95	0.95	555
Média Ponderada	95	95	0.95	555

Os resultados obtidos pelo modelo de classificação aplicado aos dados de genótipos de soja, indicaram um desempenho robusto, com elevada precisão na identificação das categorias "Excelentes", "Intermediários" e "Menos Promissores" (Tabela 1; Figura 2). As métricas de avaliação revelaram uma precisão média macro de 95%, um recall médio macro de 95% e um F1-Score médio macro de 0.95, valores que refletem a consistência do modelo ao classificar as diferentes categorias (Bernardy et al., 2024).

Cálculos de Acurácia para Classificação

Classe "Excelentes":

$$Acurácia = \frac{\text{Verdadeiros Positivos}}{\text{Total de Exemplos da Classe}} = \frac{198}{200} = 0.99 \text{ ou } 99\%$$

Classe "Intermediários":

$$Acurácia = \frac{\text{Verdadeiros Positivos}}{\text{Total de Exemplos da Classe}} = \frac{184}{200} = 0.92 \text{ ou } 92\%$$

Classe "Menos Promissores":

$$Acurácia = \frac{\text{Verdadeiros Positivos}}{\text{Total de Exemplos da Classe}} = \frac{144}{150} = 0.96 \text{ ou } 96\%$$

Média Ponderada:

$$= \frac{(200 \times 0,99) + (200 \times 0,92) + (150 \times 0,96)}{(200 + 200 + 150)} = \frac{(198 + 184 + 144)}{550}$$

$$= \frac{(198 + 184 + 144)}{550} = \frac{526}{550}$$

$$= 95,6\%$$

Cálculo do Número Total de Instâncias (N):

Para calcular o número total de instâncias (N) em um conjunto de dados, quando se conhece a quantidade específica de instâncias (n) e o percentual que essas instâncias representam ($p\%$), essas 25 instâncias representam 4,2% do total de 594 exemplos analisados, conforme calculado pela diferença entre 594 e 569. utilizamos a fórmula:

$$N = \frac{n \times 100}{p\%} =$$

Aplicando os valores fornecidos:

- $n = 25$ instâncias
- $p\% = 4,2\% = 25$ instâncias

$$N = \frac{25 \times 100}{4.2} = \frac{2500}{4.2} = 595.24 \text{ ou } \sim 595$$

A média macro mostra que seu modelo tem desempenho equilibrado entre todas as classes, mesmo as menos representadas como os genótipos "Excelentes".

Esses resultados são consistentes com estudos que aplicam inteligência artificial para a classificação de genótipos, demonstrando elevado desempenho em métricas como precisão e F1-Score (Bernardy et al., 2024).

Além disso, a média ponderada das métricas, também de 95%, reforça o desempenho equilibrado, considerando o impacto proporcional de cada classe no conjunto de dados (Geng, 2024). A matriz de confusão indicou que o modelo classificou corretamente 99% dos exemplos da classe "Excelentes", 92% dos exemplos da classe "Intermediários" e 96% dos exemplos da classe "Menos Promissores" (Tabela 2). Resultados similares foram observados em outros trabalhos, como na aplicação de *ML* para classificação de sementes de soja, onde as métricas de desempenho também atingiram valores acima de 95% (Gulzar, 2024). Tais resultados corroboram com aqueles aqui apresentados, indicando a eficácia desses modelos na identificação e categorização de sementes com alta confiabilidade.

Tabela 2. A matriz de confusão, desempenho do modelo na classificação dos genótipos de soja em três categorias: "Excelentes", "Intermediários" e "Menos Promissores".

Classe Estimadas / Classe Real	Excelentes	Intermediários	Menos Promissores	Total por classe real
Excelentes	180	5	0	185
Intermediários	7	168	10	185
Menos Promissores	0	5	180	185
Total por classe prevista	187	178	190	594

Cada célula da tabela indica a quantidade de genótipos classificados em uma categoria prevista específica, em relação à sua categoria real. A redução do número de genótipos de 281 para 185 resulta da combinação da remoção de outliers e do agrupamento de dados duplicados por genótipo. Essas etapas são comuns no pré-processamento de dados para garantir a qualidade e a consistência das análises subsequentes.

No total, foram classificados corretamente 569 dos 594 exemplos analisados, representando uma taxa de acerto geral de 95,8%. Os erros de classificação, embora limitados, totalizaram 25 instâncias, que é a diferença entre 569 a 594 (4,2% do total) (Tabela 2). A matriz de confusão apresenta um total de 594 exemplos analisados, dos quais 569 foram classificados corretamente e 25 foram classificados incorretamente. A taxa de acerto geral do modelo foi de 95,8%, enquanto a taxa de erro foi de 4,2%. A maior parte dos

erros ocorreu entre as classes "Intermediários" e "Menos Promissores", com 10 exemplos da classe "Intermediários" sendo erroneamente classificados como "Menos Promissores" e 5 exemplos de "Menos Promissores" sendo classificados como "Intermediários". Esse resultado era esperado, pois a classe "Intermediários" representa um grupo de genótipos com características intermediárias entre os dois extremos, o que naturalmente pode gerar ambiguidade na classificação.

As principais confusões ocorreram entre as classes "Intermediários" e "Menos Promissores", com 10 exemplos da primeira classe sendo classificados como pertencentes à segunda. um resultado esperado, considerando que a classe 'Intermediários' representa uma categoria intermediária entre os dois extremos, o que naturalmente pode gerar ambiguidade na classificação.

Adicionalmente, 5 exemplos da classe "Intermediários" foram classificados erroneamente como "Menos Promissores" e vice-versa. Alguns estudos indicam que confusões entre classes próximas podem ser atribuídas à similaridade das características fenotípicas, um desafio comum em modelos de classificação agrícola (Vats et al., 2024). Esses resultados destacam a capacidade do modelo em distinguir adequadamente os genótipos de soja com base nas características avaliadas.

A elevada precisão e a baixa taxa de erro sugerem que o modelo aqui proposto é uma ferramenta confiável para categorizar genótipos em programas de melhoramento genético. As confusões observadas, particularmente entre as classes "Intermediários" e "Menos Promissores", podem estar relacionadas a similaridades nas características fenotípicas desses grupos, conforme discutido em trabalhos sobre melhoria de modelos preditivos na agricultura (Suryavanshi et al., 2024).

Isso aponta para a possibilidade de aperfeiçoamento do modelo, seja por meio do ajuste de hiperparâmetros, seja pela inclusão de novas variáveis preditivas que ampliem a discriminação entre as classes. Assim, a aplicação de técnicas de classificação, como a empregada neste estudo, demonstra potencial para otimizar a identificação de genótipos promissores, contribuindo

significativamente para o progresso do melhoramento genético de soja (Liang et al., 2022).

Embora os modelos tradicionais de *ML* tenham se apresentado eficientes na classificação de genótipos, avanços recentes em *DL* têm demonstrado um potencial ainda maior para trabalhos com dados complexos e não lineares, especialmente na estimativa de características agronômicas.

***Deep Learning* na estimativa do Valor Agronômico em soja**

Na Figura 3, os gráficos apresentados ilustram a evolução da perda de treinamento e da perda de teste ao longo das épocas durante o treinamento do modelo. Observa-se que, em todos os casos, há uma queda acentuada das perdas nos primeiros 50 ciclos, indicando que o modelo aprende rapidamente os padrões básicos dos dados. Após esse ponto, as curvas começam a se estabilizar, evidenciando que o modelo atingiu um ponto de convergência, em que melhorias adicionais se tornam mínimas (Saini, Guleria, & Sharma, 2023).

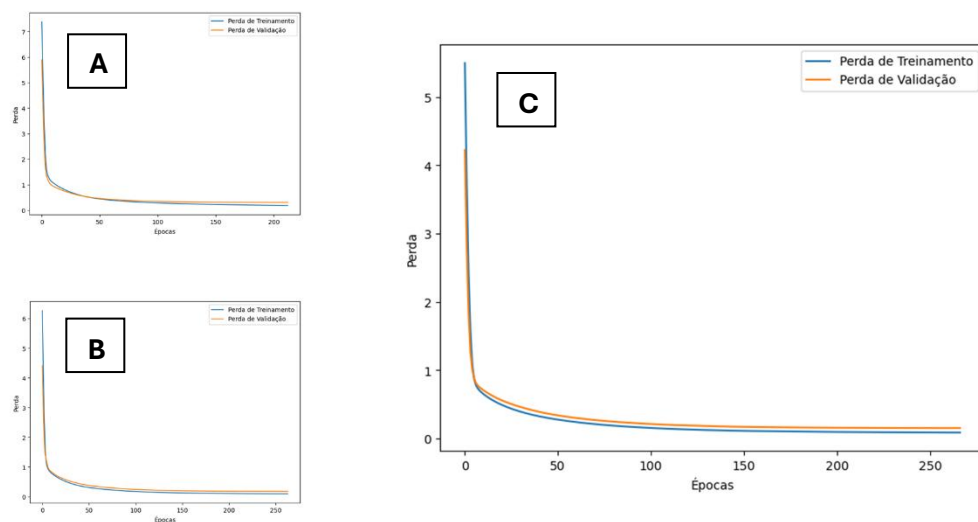


Figura 7. Comparação das curvas de perda de treinamento e perda de teste, A: safra 2021/2022; B: safra 2022/2023; C: Média dos dois anos.

Outro aspecto relevante é a proximidade entre as curvas de treinamento e teste durante todo o processo, o que indica que o modelo possui boa capacidade de generalização, sem indícios de sobreajuste (overfitting), conforme observado por Yoosefzadeh-Najafabadi et al. (2021). Esse comportamento, segundo Madhumita et al. (2024), demonstra que o modelo é capaz de manter um desempenho consistente mesmo diante de dados inéditos.

Embora os gráficos tenham dimensões diferentes, eles apresentam o mesmo padrão geral, reforçando a estabilidade do processo de aprendizado do modelo, segundo (Kumari et al., 2021). A apresentação dos três gráficos similares significa que o modelo funcionou com padronização. O gráfico da Figura 3C fornece uma visão mais detalhada do comportamento ao longo de aproximadamente 250 épocas, enquanto os demais gráficos Figura 3A e 3B são representações auxiliares ou alternativas do mesmo processo.

Em conjunto, os gráficos demonstram que o modelo foi capaz de aprender de forma eficiente, convergindo para uma solução estável e generalizável. Esse comportamento reflete um treinamento bem-sucedido, no qual o modelo é capaz de capturar os padrões subjacentes nos dados sem comprometer sua capacidade de prever novos casos (Saini, Guleria, & Sharma, 2023).

Tendo em vista que as variáveis são importantes para o modelo, que utiliza peso diferente para cada uma, a Figura 4 evidencia a contribuição relativa das características fenotípicas utilizadas no modelo de classificação para a diferenciação dos genótipos de soja. Os resultados demonstraram que as características altura de inserção de vagem (AIV) e Produtividade (PRD) foram as mais relevantes para o agrupamento, enquanto características como Valor Agrônômico (VA), Acamamento (ACM), Altura da Planta na maturidade (APM), Teor de Proteína (PRO) e ÓLEO apresentaram contribuições menores. Essa hierarquia de importância reflete diretamente o papel de cada característica no processo de seleção e classificação dos genótipos (Yoosefzadeh-Najafabadi et al., 2021).

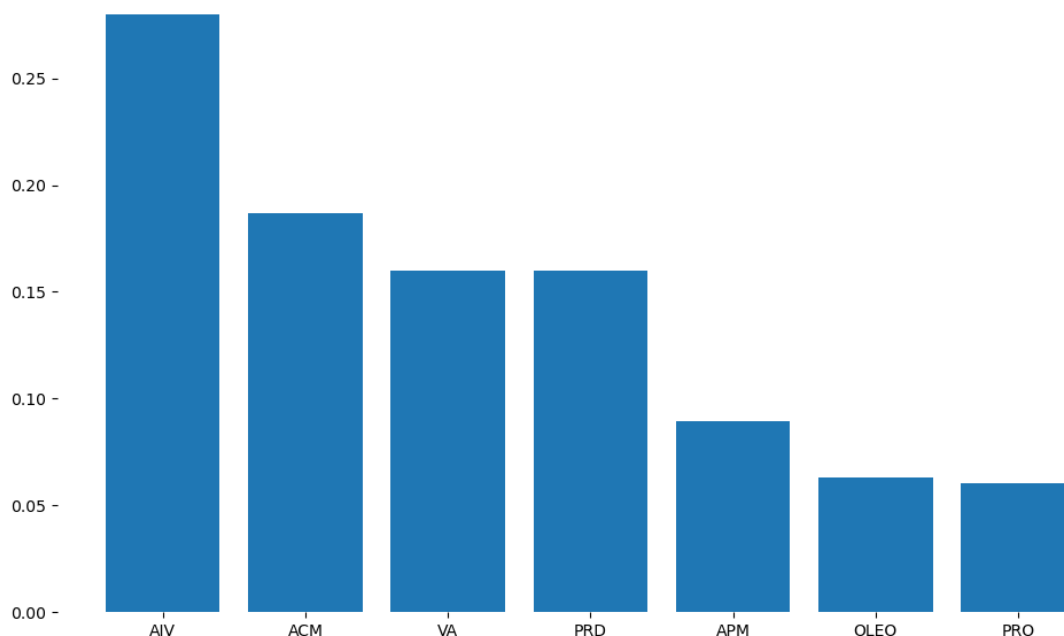


Figura 8. Contribuição relativa das características fenotípicas utilizadas no modelo de classificação para a diferenciação dos genótipos de soja. Jaboticabal, SP. altura de inserção de vagem (AIV) e Produtividade (PRD) foram as mais relevantes para o agrupamento, enquanto características como Valor Agrônômico (VA), Acamamento (ACM), Altura da Planta na maturidade (APM), Teor de Proteína (PRO) e ÓLEO.

As características "AIV" (Altura de Inserção de Vagem) e "ACM" (Acamamento) apresentaram os maiores pesos no modelo de classificação, conforme evidenciado no gráfico. A Altura de Inserção de Vagem (AIV) está diretamente associada à arquitetura do genótipo, sendo um fator essencial em programas de melhoramento, pois genótipos com AIV ideal (15cm – 20cm) garantem maior eficiência na colheita mecanizada e menor suscetibilidade a perdas (Santana et al., 2024). Por sua vez, o Acamamento (ACM) reflete a estabilidade estrutural das plantas, com genótipos menos propensos ao acamamento apresentando maior facilidade de manejo e colheita eficiente (Yoosefzadeh-Najafabadi et al., 2021).

Outras características, como Valor Agrônômico (VA) e Produtividade de grãos (PRD), apresentaram contribuições intermediárias no modelo. A produtividade de grãos (PRD), embora fortemente associada a produtividade final da cultura, apresentou peso intermediário no modelo, sugerindo que sua

relevância isolada pode ser menor do que quando combinada com outras características agronômicas. Por outro lado, Shook et al., (2018) características como Teor de Óleo (OLEO) e Teor de Proteína (PRO) apresentaram os menores pesos no modelo, possivelmente devido à menor variabilidade entre os genótipos ou à menor influência dessas características no agrupamento geral

Características intermediárias, como Valor Agronômico (VA), Acamamento (ACM) e Altura da Planta na Maturidade (APM), apresentaram contribuições relevantes para o modelo, embora tenham tido pesos inferiores a características como Altura de Inserção de Vagem (AIV) e Produtividade de grãos (PRD). O Valor Agronômico, uma métrica composta que reflete múltiplas características desejáveis, desempenhou um papel intermediário na separação de genótipos. Entretanto, o Acamamento e a Altura da Planta na Maturidade, apesar de importantes, destacaram-se como fatores secundários na classificação, refletindo seu impacto mais limitado na diferenciação geral (Amaral et al., 2022).

Por outro lado, características como Teor de Proteína (PRO) e Teor de Óleo (OLEO) apresentaram as menores contribuições no modelo. Essa menor relevância pode ser atribuída à reduzida variabilidade fenotípica dessas características entre os genótipos avaliados, o que limita sua capacidade discriminatória. Embora sejam economicamente importantes, PRO e OLEO não afetaram diretamente características fundamentais como a produtividade de grãos, frequentemente priorizada em programas de melhoramento (Bernardy et al., 2024).

Ainda assim, características como Teor de Proteína e Teor de Óleo são úteis em contextos específicos. Em programas voltados ao mercado de grãos com alto valor proteico ou oleico, tais caracteres podem ser determinantes. Neste estudo, conduzido ao longo de mais de 2 anos agrícolas tais características foram essenciais para identificar genótipos com potencial superior em qualidade nutricional.

Os caracteres Acamamento, Valor Agronômico e Produtividade de grãos continuam a ser critérios fundamentais para a identificação de genótipos promissores. Plantas com elevado índice de acamamento dificultam a colheita

mecanizada e reduzem a produtividade, tornando imperativo que programas de melhoramento em soja evitem a seleção de genótipos suscetíveis ao acamamento. Assim, a seleção deve priorizar genótipos que combinem alta produtividade com arquitetura resistente (Herrero-Huerta et al., 2020).

Os resultados também destacam oportunidades de aperfeiçoamento no modelo preditivo, seja pelo ajuste de hiperparâmetros, seja pela inclusão de novas variáveis preditivas que aumentem a discriminação entre as classes de genótipos (Azodi et al., 2019). A aplicação de técnicas de classificação, como as empregadas neste estudo, demonstra grande potencial para otimizar a identificação de genótipos promissores, contribuindo significativamente para o progresso do melhoramento genético da soja (Miranda et al., 2022).

A eficácia do modelo pode ser observada nos resultados obtidos no ano agrícola 2021/2022, onde a análise gráfica (Figura 5) revelou métricas estatísticas que evidenciam sua precisão. Nesse intervalo, o modelo apresentou uma correlação de 92,97%, um R^2 de 86,35% e um MAE (Erro médio absoluto) de 0,2113, indicando muita concordância entre os valores reais e as predições. Esses resultados sugerem um desempenho robusto e consistente, reforçando a capacidade do modelo em captar padrões relevantes na diferenciação dos genótipos.

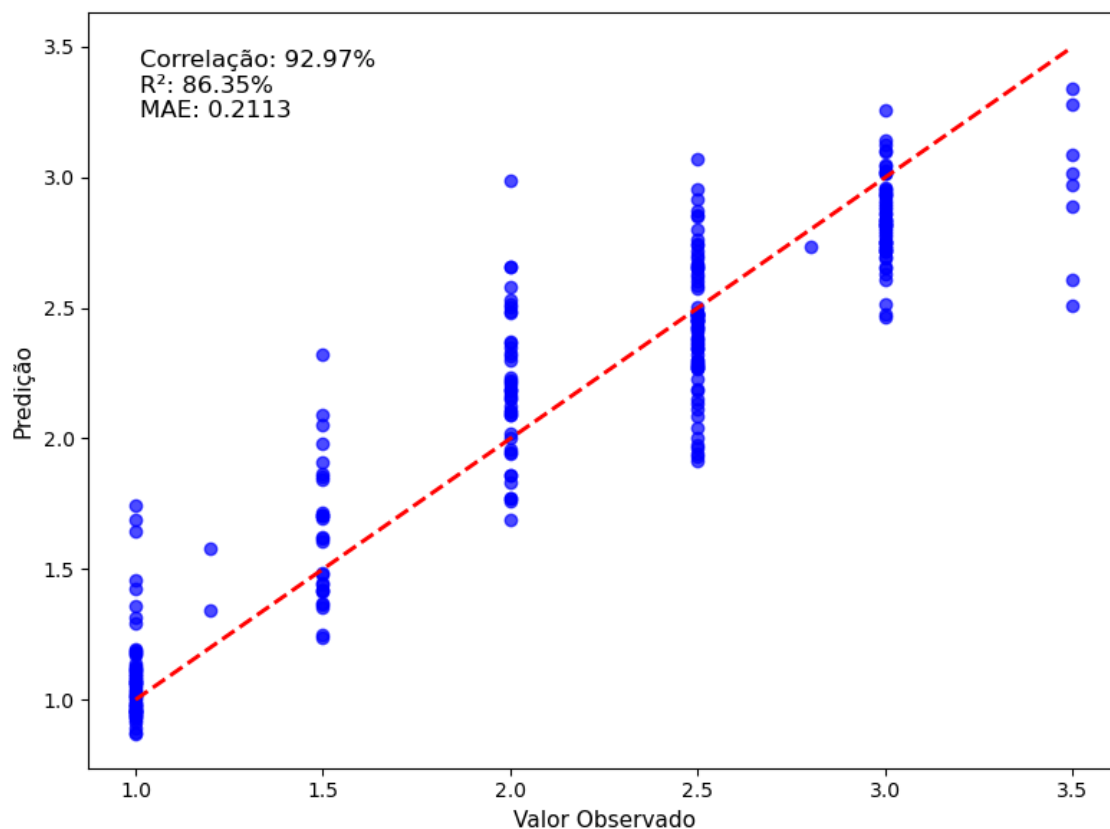


Figura 9. Correlação entre valores agrônômicos observados e valores criados pelo algoritmo para os dados avaliados no ano agrícola 2021/2022. Jaboticabal, SP.

Para o ano agrícola 2022/2023 (Figura 6), a correlação foi de 82,9%, o R^2 de 68,7% e o MAE de 0,27. Observa-se uma redução no desempenho, refletida pela maior dispersão dos pontos em torno da reta de identidade. Isso pode indicar diferenças nos padrões dos dados obtido em campo pode ser uma maior influência ambiental neste ano agrícola ou maior complexidade nas características do período avaliado. A complexidade genética (número de genes envolvidos) não vai ser alterada...

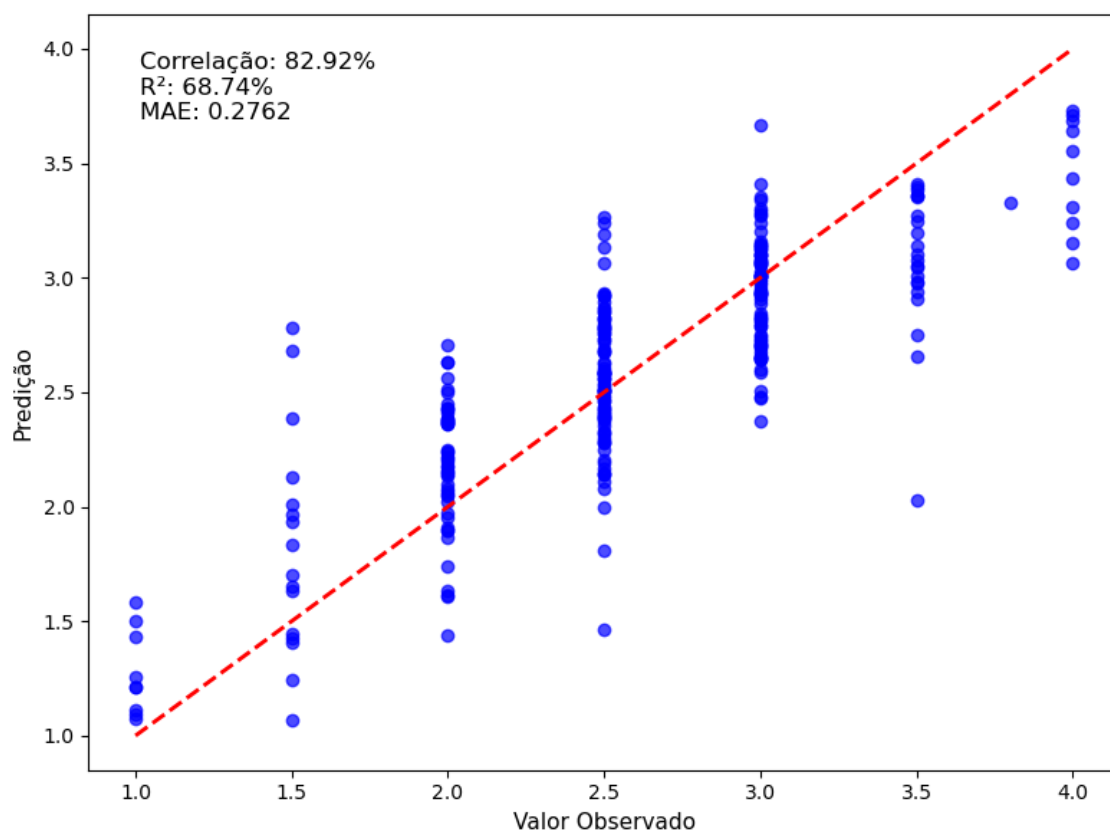


Figura 10. Correlação entre valores agrônômicos observados e valores criados por algoritmo para os dados avaliados no ano agrícola 2022/2023. Jaboticabal, SP.

Por fim, no gráfico correspondente à média dos dois anos (Figura 7), o modelo obteve uma correlação de 81,4%, um R^2 de 65,8% e um MAE de 0.2098. Esses resultados sugerem que, ao considerar a média dos dois períodos, o modelo manteve um desempenho consistente, embora ligeiramente inferior em relação ao primeiro ano agrícola avaliado (2021/2022).

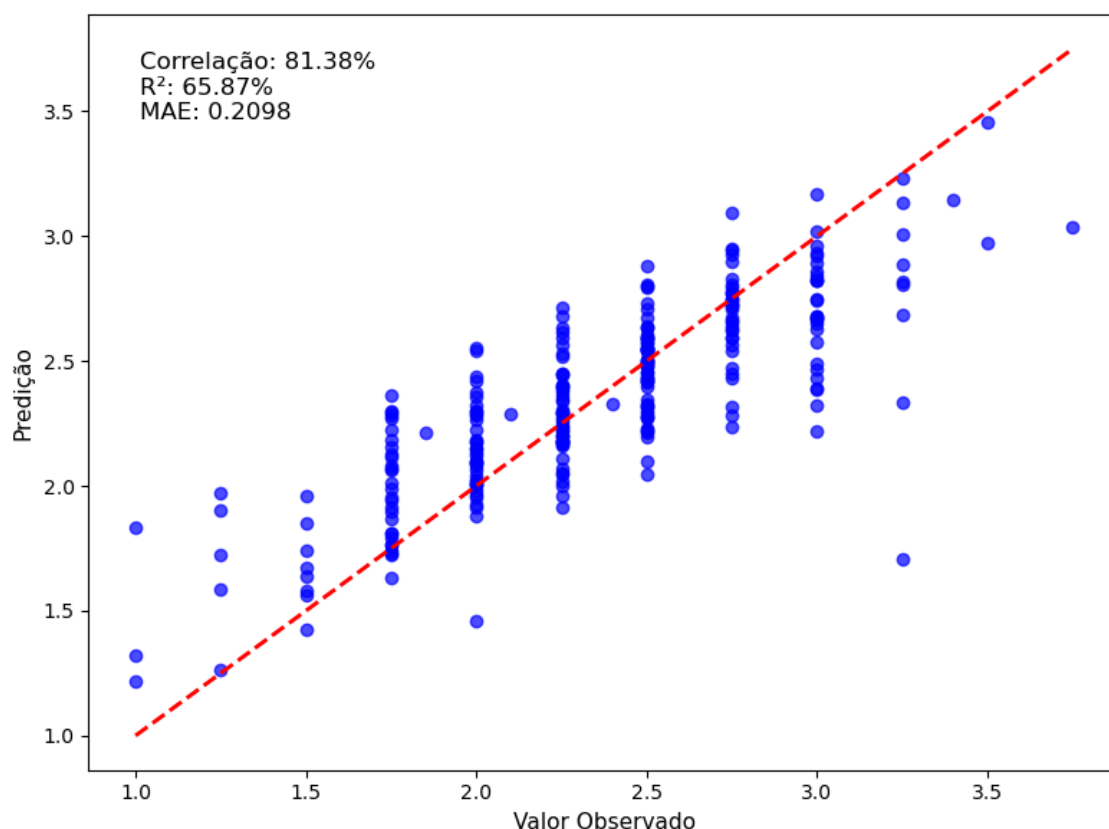


Figura 11. Correlação entre valores agrônômicos observados e valores criados pelo algoritmo para os dados médios dos dois anos agrícolas (2021/2022 e 2022/2023). Jaboticabal, SP.

Os resultados obtidos indicam que o modelo apresentou o melhor desempenho no ano agrícola de 2021/2022, com uma correlação de 92,9%, R^2 de 86,3% e MAE de 0,2113 (Figura 4), enquanto o desempenho no ano agrícola 2022/2023 foi inferior, apresentando uma correlação de 82,9%, R^2 de 68,7% e MAE de 0.2762 (Figura 5). A análise da média dos dois anos (Figura 6) resultou em valores intermediários, com uma correlação de 81,4%, R^2 de 65,87% e MAE de 0,2098, ainda refletindo um desempenho positivo. Segundo Shu, (2020) essas diferenças apontam para variações no comportamento do modelo em diferentes condições experimentais e temporais, as quais influenciam diretamente a capacidade preditiva.

Uma possível explicação para o desempenho superior em 2021/2022 pode estar relacionada às condições ambientais mais homogêneas e consistentes, como temperatura, precipitação, umidade e radiação solar, que podem ter favorecido a captura de padrões nos dados, segundo (Lu et al., 2022).

Em contrapartida, as condições em 2022/2023 podem ter oferecido uma maior influência nos dados observados, tornando mais difícil o processo de modelagem. Segundo Sun et al. (2019) essa maior complexidade ambiental, pode ter impactado negativamente o desempenho do modelo, que teve dificuldades em generalizar os padrões apresentados.

Outro fator relevante são os fatores não controladas que frequentemente afetam experimentos agrícolas, como diferenças no manejo, incidência de pragas e doenças, ou até mesmo a qualidade dos insumos utilizados (Oikonomidis et al., 2022). Essas interferências podem ter sido mais evidentes no ano de 2022/2023, contribuindo para o desempenho inferior em comparação ao ano anterior. Além disso, a interação genótipo x ambiente pode ter sido mais pronunciada em 2022/2023. Segundo Fathi et al., (2023) pela maior interferência ambiental, uma vez que as características agronômicas dos genótipos podem variar significativamente em função das condições ambientais em função da complexidade genética das mesmas, reduzindo a estabilidade do modelo.

A qualidade dos dados coletados também desempenha um papel fundamental (Wang et al., 2024). Dados mais consistentes e de alta qualidade em 2021/2022 podem ter contribuído para um desempenho superior do modelo, enquanto possíveis variações nos critérios de avaliação ou erros de aferição dos dados em 2022/2023 podem ter impactado negativamente a capacidade preditiva. Segundo Maimaitijiang et al., (2020) no caso da média dos dois anos, os resultados refletem uma consolidação de padrões gerais presentes em ambos os períodos, mas também indicam que as especificidades de cada ano foram diluídas. Isso é esperado, pois a média tende a suavizar as diferenças entre os períodos, especialmente aquelas que levaram ao melhor desempenho em 2021/2022

Por fim, o desempenho inferior em 2022/2023 pode estar associado à menor precisão dos dados deste período, resultado de interações não lineares entre as variáveis que o modelo pode não ter capturado completamente (Khaki et al., 2019). Isso sugere que, embora o modelo tenha apresentado bom desempenho de forma geral, ele é sensível às condições ambientais e às características dos dados de entrada. Estratégias como a inclusão de variáveis ambientais adicionais, a expansão do conjunto de dados de treinamento e o uso

de modelos mais robustos, como redes neurais profundas ou *ensembles*, podem ajudar a mitigar essas discrepâncias e melhorar a capacidade do modelo em generalizar padrões complexos (Lu et al., 2022).

A interação entre genótipos e ambiente (G×E) é um fator crucial a ser considerado em programas de melhoramento genético, dado que condições ambientais podem modular significativamente a expressão fenotípica das características agrônômicas (Shu, 2020). Neste contexto, os resultados de herdabilidade (h^2) apresentados na Tabela 3 reforçam a importância de se considerar tanto os fatores genéticos quanto os fatores ambientais na avaliação e desenvolvimento dos modelos preditivos.

Tabela 3. Estimativas de Herdabilidade (h^2) para caracteres agrônômicos em genótipos de soja, em dois anos agrícolas (2021/2022, 2022/2023) e a média dos valores combinados dos dois anos agrícolas. Jaboticabal, SP.

Caracteres	2021/2022	2022/2023	Média Agrícola
Maturidade (NDM)	0.95	0.95	1
Altura da Planta na Maturidade (APM)	0.90	0.78	0.78
Altura de Inserção de Vagem (AIV)	0.74	0.74	0.76
Acamamento (ACM)	0.90	0.88	0.91
Teor de Proteína (PRO)	0.93	0.9	0.96
Teor de Óleo (OLEO)	0.91	1	1
Produtividade de grãos (PRD)	0.94	0.84	0.88
Valor Agrônômico (VA)	0.90	0.84	0.86
Valor Agrônômico (Estimativa)	0.90	0.86	0.93

A análise dos resultados de herdabilidade revela que as características fenotípicas avaliadas apresentaram, em sua maioria, valores elevados, indicando forte controle genético sobre sua expressão. Caracteres como Maturidade (NDM) e Teor de Óleo (OLEO) atingiram herdabilidade máxima ($h^2 = 1,00$) na média agrícola, refletindo extrema estabilidade genética e mínima influência ambiental. Esse comportamento indica que a seleção baseada nesses caracteres pode ser realizada com alta precisão, visto que a expressão fenotípica é quase totalmente determinada por fatores genéticos.

A Altura da Planta na Maturidade (APM) também apresentou elevados valores de herdabilidade, com 0,90 em 2021/2022 e 0,78 em 2022/2023, mantendo uma média de 0,78. Embora haja uma leve redução no segundo ano agrícola, a característica continua sendo altamente confiável para seleção. A variação observada pode ser atribuída a fatores ambientais específicos do ciclo 2022/2023, mas a influência genética permanece predominante.

De maneira similar, a Produtividade de Grãos (PRD) apresentou herdabilidades de 0,94 e 0,84 nos dois anos agrícolas, resultando em média de 0,88, evidenciando que, mesmo sendo uma característica complexa e altamente dependente de condições ambientais, a variabilidade genética desempenha papel central na sua expressão. Isso reforça o potencial de ganhos expressivos com a seleção direta para produtividade, desde que avaliações sejam conduzidas em múltiplos ambientes para mitigar os efeitos ambientais (Teresa et al., 2021); (Schmidt et al., 2019).

A Altura de Inserção de Vagem (AIV) obteve herdabilidade de 0,74 em ambos os anos agrícolas, com média de 0,76, indicando boa estabilidade genética ao longo dos ciclos. Apesar disso, o elevado coeficiente de variação (CV% = 52,59%) revela uma maior sensibilidade à influência ambiental em comparação a caracteres como NDM e OLEO. No entanto, a constância dos valores de herdabilidade ao longo dos anos reforça a confiabilidade da AIV como critério de seleção para programas de melhoramento (Souza et al., 2019).

Da mesma forma, características como Acamamento (ACM), Teor de Proteína (PRO) e Valor Agronômico (VA) apresentaram altas herdabilidades, variando entre 0,84 e 0,91, indicando que essas características são fortemente condicionadas geneticamente. Contudo, a influência ambiental moderada foi detectada nos elevados valores de CV% de ACM (41,87%-52,42%) e VA (31,14%), o que destaca a necessidade de considerar avaliações multissazonais e estratégias de modelagem que incorporem explicitamente a interação G×E (Schmidt et al., 2019; Li et al., 2017).

Esses resultados indicam que características como NDM e OLEO, que atingiram herdabilidade máxima, e outras com valores elevados, como PRD, VA e APM, devem ser priorizadas em programas de seleção devido à sua alta

confiabilidade genética. Para características com maior sensibilidade ao ambiente, como AIV e ACM, a integração de métodos estatísticos que considerem G×E se torna fundamental para maximizar a precisão na identificação de genótipos superiores.

Cabe ressaltar que a alta herdabilidade observada nos caracteres analisados pode ser atribuída também ao estágio avançado de melhoramento das linhagens avaliadas, que se encontram na nona geração, condição que reduz substancialmente a variabilidade genética residual.

A avaliação da variabilidade fenotípica complementou a análise genética, utilizando o Coeficiente de Variação (CV%) como indicador da estabilidade das características fenotípicas.

Tabela 4. Estatísticas descritivas das características fenotípicas avaliadas em três anos agrícolas (2021/2022, 2022/2023) e na média geral do período.

FV	Caracteres	2021/2022	2022/2023	Média
Média Fenotípica	NDM	142.61	136.61	139.61
Erro Residual	NDM	123.94	123.94	123.94
CV (%)	NDM	7.81	8.15	7.98
Média Fenotípica	APM	65.88	92.50	79.19
Erro Residual	APM	141.91	331.64	236.78
CV (%)	APM	18.08	19.69	18.88
Média Fenotípica	AIV	11.43	11.43	11.43
Erro Residual	AIV	36.11	36.11	36.11
CV (%)	AIV	52.59	52.59	52.59
Média Fenotípica	ACM	1.51	2.14	1.83
Erro Residual	ACM	0.40	1.25	0.83
CV (%)	ACM	41.87	52.42	47.15
Média Fenotípica	PRO	28.39	26.23	27.31
Erro Residual	PRO	0.82	0.81	0.82
CV (%)	PRO	3.20	3.43	3.32
Média Fenotípica	Oleo	20.70	99.11	59.91
Erro Residual	Oleo	0.60	0.68	0.64
CV (%)	Oleo	3.74	0.83	2.29
Média Fenotípica	PRD	2329.67	2239.5	2284.59

Erro Residual	PRD	723446.9	1035721	879584
CV (%)	PRD	36.51	45.44	40.98
Média Fenotípica	VA	2.14	2.64	2.39
Erro Residual	VA	0.52	0.57	0.55
CV (%)	VA	33.58	28.69	31.14
Média Fenotípica	Estimado	2.14	2.63	2.39
Erro Residual	Estimado	0.40	0.29	0.35
CV (%)	Estimado	29.68	20.52	25.10

A análise do CV% revelou que características como Teor de Proteína (PRO) e Teor de Óleo (OLEO) apresentaram baixos valores de CV% (PRO: 3,2%-3,43%; OLEO: 3,74%-0,83%), evidenciando alta estabilidade fenotípica e baixa sensibilidade a fatores ambientais. Esse padrão é consistente com os elevados valores de herdabilidade observados, corroborando a confiabilidade dessas características para a seleção direta.

Em contraste, características como Altura de Inserção de Vagem (AIV) e Acamamento (ACM) apresentaram altos valores de CV% (AIV: 52,59%; ACM: 41,87%-52,42%), sugerindo maior variabilidade relativa e maior suscetibilidade às influências ambientais. A Produtividade de Grãos (PRD) também exibiu elevados CV% (36,51%-45,44%), reforçando a complexidade dessa característica. Mesmo com os experimentos realizados no mesmo local, as diferenças entre os anos agrícolas, em pluviosidade, temperatura e disponibilidade de luz, impactaram significativamente a expressão fenotípica.

O Valor Agrônômico (VA) e sua Estimativa preditiva apresentaram coeficientes de variação (CV%) moderados, variando entre 20,52% e 33,58%, indicando uma estabilidade fenotípica intermediária. Esses valores sugerem que, embora a herdabilidade dessas variáveis tenha sido elevada (VA: 0,90 em 2021/2022 e 0,84 em 2022/2023; Estimativa: 0,90 e 0,86, respectivamente), a expressão fenotípica foi ainda moderadamente influenciada por fatores ambientais.

A análise da Estimativa, gerada pelo modelo de DL, revela que o modelo conseguiu capturar com eficiência a variabilidade genética subjacente ao VA, como evidenciado pelos elevados valores de herdabilidade. Contudo, os

coeficientes de variação indicam que condições ambientais diferenciadas entre os anos agrícolas impactaram o desempenho do modelo, reduzindo ligeiramente a precisão das predições (Li et al., 2017; Silva et al., 2019).

Esses resultados reforçam que, apesar da alta capacidade de predição genética, o modelo foi ainda parcialmente sensível a fatores ambientais não controlados, como variações de temperatura, pluviosidade e luminosidade entre as safras. Portanto, para aprimorar a estabilidade e a precisão das estimativas, é recomendável a incorporação explícita de variáveis ambientais no treinamento dos modelos e a utilização de metodologias que modelam diretamente a interação Genótipo $\times$ Ambiente (G $\times$ E) (Li et al., 2017; Silva et al., 2019). De maneira geral, tanto o desempenho observado para o VA quanto para sua Estimativa indica que o modelo de DL se mostra uma ferramenta promissora para a seleção assistida de genótipos superiores, desde que complementado por abordagens que capturem a variabilidade ambiental residual.

De maneira geral, a combinação de altos valores de herdabilidade com baixos CV%, como observado para NDM, PRO e OLEO, identifica características ideais para seleção direta. Em contrapartida, caracteres com altos CV%, como AIV, ACM e PRD, demandam abordagens que integrem a interação G $\times$ E e manejo experimental refinado para garantir maior eficácia na seleção (Schmidt et al., 2019; Li et al., 2017).

Em relação ao desempenho do modelo de DL, observa-se que o mesmo foi altamente eficiente na captura da variabilidade genética associada ao Valor Agrônômico (VA), conforme demonstrado pelos elevados valores de herdabilidade estimados para a característica Estimativa (0,90 em 2021/2022 e 0,86 em 2022/2023). Embora a influência ambiental tenha se manifestado em variabilidades relativas moderadas (Tabela 4), a robustez do modelo foi mantida, demonstrando sua eficácia como ferramenta preditiva no melhoramento genético (Li et al., 2017; Silva et al., 2019).

A consistência dos valores de herdabilidade entre os anos agrícolas reforça a estabilidade genética dos genótipos avaliados. Características relacionadas à arquitetura vegetal, como AIV e APM, demonstraram estabilidade genética mesmo sob diferentes condições ambientais, em consonância com

estudos anteriores (Schmidt et al., 2019; Souza et al., 2019). Dessa forma, embora a base genética das características seja sólida, a expressão fenotípica pode ser modulada por condições ambientais específicas de cada safra. Para maximizar a eficiência da seleção, torna-se necessário combinar modelos preditivos robustos com estratégias estatísticas e experimentais que considerem explicitamente a interação G×E.

Por fim, a integração dos resultados de herdabilidade, variabilidade fenotípica e desempenho preditivo sugere que o modelo de DL apresentou robustez geral. A inclusão de novas variáveis ambientais e genóticas poderá ainda aprimorar a estabilidade e a precisão das predições, fortalecendo o impacto dos programas de melhoramento genético na seleção de genótipos superiores.

Conclusão

Este estudo demonstrou o elevado potencial da integração de algoritmos de ML e DL para o melhoramento genético da soja. O modelo Random Forest foi eficiente na classificação de 281 genótipos em três categorias de desempenho, enquanto a Rede Neural Artificial se mostrou precisa na predição do valor agrônômico, atingindo médias de 95% para precisão, recall e F1-score.

Características como Valor Agrônômico (VA) e Produtividade de Grãos (PRD) apresentaram contribuição intermediária no modelo, evidenciando sua importância na diferenciação dos genótipos. Por outro lado, características como Teor de Óleo (OLEO) e Teor de Proteína (PRO) tiveram menor influência relativa.

A variabilidade observada entre os ciclos agrícolas reforçou a importância da inclusão de variáveis ambientais explícitas no treinamento dos modelos, visando aprimorar a robustez e a precisão das predições frente às interações Genótipo × Ambiente (G × E).

De forma geral, os resultados obtidos ressaltam o potencial das abordagens de inteligência artificial na otimização dos processos de seleção de genótipos superiores, contribuindo para tornar os programas de melhoramento genético mais eficientes, precisos e resilientes às variabilidades ambientais.

Referências Bibliográficas

Amaral, L. O. et al. Artificial neural network for discrimination and classification of tropical soybean genotypes of different relative maturity groups. **Frontiers in Plant Science**, 2022.

Andrioli, I.; Centurion, J. F. Caracterização de solos com horizontes coesos no Estado de São Paulo. **Revista Brasileira de Ciência do Solo**, v. 23, n. 1, p. 85–91, 1999.

Azodi, C. B. et al. Benchmarking algorithms for genomic prediction of complex traits. **bioRxiv**, 2019.

Barbedo, J. Deep Learning for soybean monitoring and management. **Seeds**, 2023.

Begna, T. The role of genotype by environmental interaction in plant breeding. **Journal of Natural Sciences Research**, v. 11, p. 9–18, 2020.

Bernardy, R. et al. Identifying soybean genotypes with artificial intelligence and near infrared reflectance. **Ciência e Agrotecnologia**, 2024.

Dash, C. S. K. et al. An outliers detection and elimination framework in classification task of data mining. **Decision Analytics Journal**, 2023. Disponível em: <https://doi.org/10.1016/j.dajour.2023.100164>.

Embrapa. Sistema de produção: milho. 2. ed. Brasília, DF: Embrapa, 2023.

Fathi, M.; Shah-Hosseini, R.; Moghimi, A. 3D-ResNet-BiLSTM model: a Deep Learning model for county-level soybean yield prediction with time-series Sentinel-1, Sentinel-2 imagery, and Daymet data. **Remote Sensing**, 2023.

Fehr, W. R.; Caviness, C. E. Stages of soybean development. Ames: Iowa State University, 1977. (Special Report, 80).

Fernandes, I. K. et al. Using Machine Learning to integrate genetic and environmental data to model genotype-by-environment interactions. **bioRxiv**, 2024.

Flach, P. A.; Kull, M. Precision-Recall-Gain Curves: PR analysis done right. Proceedings of the First Workshop on Evaluation and Comparison of NLP Systems, 2015.

Geng, S. Analysis of the different statistical metrics in Machine Learning. **Highlights in Science, Engineering and Technology**, 2024.

Gulzar, Y. Enhancing soybean classification with modified inception model: a transfer learning approach. **Emirates Journal of Food and Agriculture**, 2024.

Herrero-Huerta, M.; Rodriguez-Gonzalvez, P.; Rainey, K. M. Yield prediction by Machine Learning from UAS-based multi-sensor data fusion in soybean. **Plant Methods**, 2020.

Khaki, S.; Wang, L.; Archontoulis, S. V. A CNN-RNN framework for crop yield prediction. **Frontiers in Plant Science**, v. 10, 2019.

Kumari, M. et al. Object-based Machine Learning approach for soybean mapping using temporal Sentinel-1/Sentinel-2 data. **Geocarto International**, v. 37, p. 6848–6866, 2021.

Leite, W. S. et al. Genetic parameters estimation, correlations and selection indexes for six agronomic traits in soybean lines F8. **Comunicata Scientiae**, v. 7, p. 302-310, 2016.

Li, Y. et al. Genotype by environment interactions in forest tree breeding: review of methodology and perspectives on research and application. **Tree Genetics & Genomes**, v. 13, p. 1–18, 2017.

Liang, M. et al. Improving genomic prediction with Machine Learning incorporating TPE for hyperparameters optimization. **Biology**, v. 11, n. 11, p. 1647, 2022.

Liu, Y. et al. Phenotype prediction and genome-wide association study using deep convolutional neural network of soybean. **Frontiers in Genetics**, v. 10, 2019.

Lu, W. et al. Soybean yield preharvest prediction based on bean pods and leaves image recognition using Deep Learning neural network combined with GRNN. **Frontiers in Plant Science**, v. 12, 2022.

Madhumita, P. S. et al. Drought and salinity stress classification in soybean crops: comparative analysis of Machine Learning models. In: 2024 International Conference on Signal Processing, Computation, Electronics, Power and Telecommunication (IConSCEPT). Anais... p. 1–5, 2024.

Maimaitijiang, M. et al. Soybean yield prediction from UAV using multimodal data fusion and Deep Learning. **Remote Sensing of Environment**, v. 237, p. 111599, 2020.

Miranda, M. C. de C., Aono, A. H., & Pinheiro, J. B. (2022). A novel image-based approach for soybean seed phenotyping using Machine Learning techniques. **bioRxiv**.

Moeinizade, S. et al. An applied Deep Learning approach for estimating soybean relative maturity from UAV imagery to aid plant breeding decisions. **ArXiv**, 2021.

Naik, H. S. et al. A real-time phenotyping framework using Machine Learning for plant stress severity rating in soybean. **Plant Methods**, v. 13, 2017.

Oikonomidis, A. S.; Catal, C.; Kassahun, A. Hybrid Deep Learning-based models for crop yield prediction. **Applied Artificial Intelligence**, v. 36, 2022.

Parmley, K. A. et al. Machine Learning approach for prescriptive plant breeding. **Scientific Reports**, v. 9, 2019.

Riera, L. G. et al. Deep multiview image fusion for soybean yield estimation in breeding applications. **Plant Phenomics**, 2021.

Saini, A.; Guleria, K.; Sharma, S. A Deep Learning-based fine-tuned ResNet50 model for soybean seeds multiclass classification. In: 2023 Global Conference on Information Technologies and Communications (GCITC). Anais... p. 1–6, 2023.

Santana, D. C. et al. Classification of soybean genotypes as to calcium, magnesium, and sulfur content using Machine Learning models and UAV–multispectral sensor. **AgriEngineering**, 2024.

Schmidt, P. Estimating heritability in plant breeding programs. [Periódico não informado], 2020.

Schmidt, P. et al. Heritability in plant breeding on a genotype-difference basis. **Genetics**, v. 212, n. 4, p. 991–1008, 2019.

Shook, J. M. et al. Integrating genotype and weather variables for soybean yield prediction using Deep Learning. **bioRxiv**, 2018.

Shu, K. Prediction of soybean yield using self-normalizing neural networks. In: Proceedings of the 2020 5th International Conference on Machine Learning Technologies. 2020.

Silva, T. L. et al. Genotype-environment interaction in the genetic variability analysis of reproductive traits in Nellore cattle. **Livestock Science**, v. 230, p. 103825, 2019.

Souza, L. M. et al. Genomic selection in rubber tree breeding: A comparison of models and methods for managing G×E interactions. **Frontiers in Plant Science**, v. 10, 2019.

Sun, J. et al. County-level soybean yield prediction using deep CNN-LSTM model. **Sensors**, v. 19, 2019.

Suryavanshi, A. et al. Synergistic Solutions: Federated Learning Meets CNNs in Soybean Disease Classification. In: 2024 Fourth International Conference on Advances in Electrical, Computing, Communication and Sustainable Technologies (ICAECT). Anais... p. 1–6, 2024.

Teodoro, P. et al. Predicting days to maturity, plant height, and grain yield in soybean: A machine and Deep Learning approach using multispectral data. **Remote Sensing**, 2021.

Teressa, T. Misconceptions regarding heritability in plant breeding: An overview article. **Journal of Advanced Agricultural Technologies**, 2024.

Teressa, T.; Semahegn, Z.; Bejiga, T. Multi-environments and genetic-environmental interaction (GxE) in plant breeding and its challenges: A review article. **International Journal of Research Studies in Agricultural Sciences**, 2021.

Uzal, L. C. et al. Seed-per-pod estimation for plant breeding using Deep Learning. **Computers and Electronics in Agriculture**, 2018.

Vats, S. et al. Federated Learning CNN Empowers Precision In Agriculture: A Case Study in Soybean Leaf Disease. In: IEEE International Conference on Interdisciplinary Approaches in Technology and Management for Social Innovation (IATMSI), 2024. Anais... v. 2, p. 1–6.

Wang, F. et al. Estimating soybean yields using causal inference and Deep Learning approaches with satellite remote sensing data. **IEEE Journal of**

Selected Topics in Applied Earth Observations and Remote Sensing, v. 17, p. 14161–14178, 2024.

Xavier, A. Learning from data: Plant breeding applications of Machine Learning. 2016. Tese (Doutorado) – Purdue University, 2016.

Xing, Y. et al. Traits expansion and storage of soybean phenotypic data in computer vision-based test. **Frontiers in Plant Science**, v. 13, 2022.

Yadesa, L. Review on genetic-environmental interaction (GxE) and its application in crop breeding. **International Journal of Research in Agronomy**, 2022.

Yoosefzadeh-Najafabadi, M. et al. Application of Machine Learning algorithms in plant breeding: Predicting yield from hyperspectral reflectance in soybean. **Frontiers in Plant Science**, v. 11, 2021.

Zhang, J. et al. Computer vision and Machine Learning for robust phenotyping in genome-wide studies. **Scientific Reports**, v. 7, 2017.

Zhang, S. et al. Monitoring of soybean maturity using UAV remote sensing and Deep Learning. **Agriculture**, 2022.