



UNIVERSIDADE ESTADUAL DO MARANHÃO
CENTRO DE CIÊNCIAS AGRÁRIAS
PROGRAMA DE PÓS-GRADUAÇÃO
EM CIÊNCIAS AGRÁRIAS
CURSO DE MESTRADO EM
CIÊNCIAS AGRÁRIAS



LETICIA CARRAZANA MARTINEZ

**CLASSIFICAÇÃO DE CULTIVARES DE SOJA COM IMAGENS DE
SEMENTES E FOLHAS USANDO INTELIGENCIA ARTIFICIAL E
OTIMIZAÇÃO DO NÚMERO MÍNIMO DESSAS ESTRUTURAS A SER
FOTOGRAFADAS**

Balsas-MA
2026

LETICIA CARRAZANA MARTINEZ

**CLASSIFICAÇÃO DE CULTIVARES DE SOJA COM IMAGENS DE
SEMENTES E FOLHAS USANDO INTELIGENCIA ARTIFICIAL E
OTIMIZAÇÃO DO NÚMERO MÍNIMO DESSAS ESTRUTURAS A SER
FOTOGRAFADAS.**

Dissertação apresentada ao Programa de Pós- Graduação em Ciências Agrárias da Universidade Estadual do Maranhão como requisito para obtenção do título de Mestre em Ciências Agrárias.

Orientador: Prof. Dr. Francisco Charles dos Santos Silva

Coorientador: Prof. Dr. Alan Mario Zuffo

LETICIA CARRAZANA MARTINEZ

Dissertação apresentada ao Programa de Pós-Graduação em Ciências Agrárias da Universidade Estadual do Maranhão, como requisito para obtenção do título de Mestre em Ciências Agrárias.

Aprovada em: 12 / 02/ 2026

BANCA EXAMINADORA



Documento assinado digitalmente

FRANCISCO CHARLES DOS SANTOS SILVA

Data: 11/03/2026 22:15:44-0300

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

Prof. Dr. Francisco Charles dos Santos Silva – Orientador



Documento assinado digitalmente

ALAN MARIO ZUFFO

Data: 12/03/2026 17:15:21-0300

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

Prof. Dr. Alan Mario Zuffo -Coorientador



Documento assinado digitalmente

CLEMILTON ALVES DA SILVA

Data: 12/03/2026 15:34:42-0300

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

Prof. Dr. Clemilton Alves da Silva

Balsas-MA 202

M385c

Martinez, Letícia Carrazana

Classificação de cultivares de soja com imagens de sementes e folhas usando inteligência artificial e otimização do número mínimo dessas estruturas a ser fotografadas. Letícia Carrazana Martinez /. – Balsas, 2026.

89 f.

Dissertação (Mestrado em Ciências Agrárias) Universidade Estadual do Maranhão – UEMA / Balsas, 2026.

Orientador: Prof. Doutor Francisco Charles dos Santos Silva

Coorientador: Prof. Doutor Alan Mario Zuffo

1. Machine Learning. 2. Variáveis Morfológica. 3. Variáveis Espectrais. I Título.

CDU: 633.34

Dedicatória

A DEUS !

AGRADECIMENTOS

A meu grande Deus, sem ele nada seria possível, a meu orientador o professor Francisco pelos ensinamentos e pela orientação que permitiu esse trabalho se concretizar, a meus amigos Teresinha e Edimilson pelo seu apoio e acolhida durante as aulas em São Luis, a minha mãe pelo carinho mesmo na distância, a UEMA campus-Balsas e aos professores da instituição e do programa pelos ensinamentos e suporte na pesquisa, e a instituição FAPEMA pelo apoio financeiro.

Lista de Figuras

Capítulo II

- Figura 1.** Diagrama das etapas e procedimentos de análise de dados para classificação de cultivares de soja usando processamento de imagens e aprendizado de máquina.....46
- Figura 2.** Resumo dos 10 clusters obtidos pela aplicação do agrupamento k-means as características morfológicas e espectrais, obtidas de imagens RGB de sementes e folhas de dez cultivares de soja 48
- Figura 3.** Matriz de confusão das características morfológicas e espectrais, obtidas de imagens RGB de sementes e folhas de dez cultivares de soja..... 49
- Figura 4.** Resumo da Análise das duas Componentes Principais (ACP) das características morfológicas e espectrais, obtidas de imagens RGB de sementes e folhas de dez cultivares de soja (A). Porcento de autovalores de todas as componentes principais da análises das características morfológicas e espectrais, obtidas de imagens RGB de sementes e folhas de dez cultivares de soja (B)..... 50
- Figura 5.** Porcentagem de classificações corretas de 10 cultivares de soja a partir de 5 algoritmos de aprendizado de máquina aplicado para caracteres morfológicos das folhas, caracteres espectrais das folhas e caracteres morfológicos e espectrais das folhas combinado. NB: Naive Bayes, ROF: Rotation Forest, RAF: Random Forest, IB1: Instance Based, MP: Multilayer Perceptron52
- Figura 6.** Porcentagem de classificações corretas de 10 cultivares de soja a partir de 5 algoritmos de aprendizado de máquina aplicado para caracteres morfológicos das sementes, caracteres espectrais das sementes e caracteres morfológicos e espectrais das sementes combinado. NB: Naive Bayes, ROF: Rotation Forest, RAF: Random Forest, IB1: Instance-Based, MP: Multilayer Perceptron.....53
- Figura 7.** Porcentagem de classificações corretas de 10 cultivares de soja a partir de 5 algoritmos de aprendizado de máquina aplicado para caracteres morfológicos das F/S, caracteres espectrais das F/S e caracteres morfológicos e espectrais das F/S combinado. NB: Naive Bayes, ROF: Rotation Forest, RAF: Random Forest, IB1: Instance-Based, MP: Multilayer Perceptron..... 54

Capítulo III

Figura 1. Números de folhas para obtenção do 95% de precisão dos caracteres morfológicos das folhas resultante das análises de repetibilidade: Anova: Análise de Variância; CP-COV: Covariâncias Fenotípicas; CP-COR: Matriz de Correlações; A. Estructural: Análise Estructural; Area: Área; Perim: Perímetro; Cir: Circularidade; AR: Aspect Ratio; Raund: Redondeza; Solidity: Solidez 75

Figura 2. Números de folhas para obtenção do 95% de precisão para caracteres espectrais das folhas resultante das análises de repetibilidade: Anova: Análise de Variância; CP-COV: Covariâncias Fenotípicas; CP-COR: Matriz de Correlações; A.Estructuta: Análise Estructural; Red: Vermelho; Green: Verde; Blue: Azul; H: Valor de Matriz HSB; L: Luminosidade no modelo Lab; S: Saturação; a: Coordenada cromática Lab; B: Banda do azul; b: Coordenada cromática do amarelo Lab..... 77

Figura 3. Números de sementes para obtenção do 95% de precisão dos caracteres morfológicos das sementes resultante das análises de repetibilidade: Anova: Análise de Variância; CP-COV: Covariâncias Fenotípicas; CP-COR: Matriz de Correlações; A. Estructural: Análise Estructural; Area: Área; Perim: Perímetro; Cir: Circularidade, AR: Aspect Ratio; Raund: redondeza, Solidity: Solidez 79

Figura 4. Números de sementes para obtenção do 95% de Acurácia dos caracteres espectrais das sementes resultante das análises de repetibilidade: Anova: Análise de Variância; CP-COV: Covariâncias Fenotípicas; CP-COR: Matriz de Correlações; A. Estructural: Análise Estructural; H:Valor de Matriz HSB; L: Luminosidade no modelo Lab; S: Saturação; a: Cordenada cromática Lab; B: Banda do azul; b: Coordenada cromática do amarelo Lab..... 81

Lista de Tabelas

Capítulo II

Tabela 1. Modelos de aprendizado de máquina utilizados para a classificação dos cultivares de soja..... 44

Tabela 2. Resumo da análise de variância das características morfológicas e espectrais, obtidas de imagens RGB de sementes e folhas de dez cultivares de soja 47

Capítulo III

Tabela 1. Resumo da análise de variância das características morfológicas e espectrais, obtidas de imagens RGB de sementes e folhas de dez cultivares de soja 74

Lista das siglas e abreviações

| | |
|-------------------------------------------|---------------------------------------------|
| A. Estrutural: Análise Estrutural | Green: Verde |
| a: Cordenada cromática R/G Lab | H: Valor de Matriz HSB |
| ACP : Análise de Componentes Principais . | IB1: Instance-Based Learning |
| AD: Análise de Discriminante. | L: Luminosidade no modelo Lab |
| ANOVA: Análise de Variância | ML: Machine Learning |
| AR: Aspect Ratio | MP: Multilayer Perceptron |
| Área: Área | NB: Naive Bayes |
| B: Banda do azul | Perim: Perímetro |
| b: Coordenada cromática do amarelo | QM: Quadrado Médio |
| Lab | RAF: Random Forest |
| Blue: Azul | Raund: Redondeza |
| Circ: Circularidade | Red: Vermelho |
| CP-COR: Matriz de Correlações | RF: Floresta aleatória |
| CP-COV:Covariâncias | RNA: Redes Neurais Artificiais |
| Fenotípicas | ROF: Rotation forest |
| CV:Coeficiente de Variação | S: Saturação |
| FB: é a classificação de falso negativo | Solidity: Solidez |
| FP: é a classificação de falso positivo | TP:é a classificação de verdadeiro positivo |

SUMÁRIO

| | |
|--------------------------------------------------|----|
| Capítulo I..... | 14 |
| 1. INTRODUÇÃO..... | 15 |
| 2. REFERENCIAL TEÓRICO | 17 |
| 2.1 Origem da soja | 17 |
| 2.2 Taxonomia, botânica e morfologia | 18 |
| 2.3 Exigências edafoclimáticas..... | 20 |
| 2.4 Processamento de imagens..... | 21 |
| 2.5 Inteligência artificial na Agricultura | 22 |
| 2.6 Produção de sementes de soja | 25 |
| 2.7 Repetibilidade..... | 26 |
| REFERÊNCIAS | 27 |
| Capítulo II..... | 34 |
| 1. INTRODUÇÃO..... | 37 |
| 1.1 Instalação do ensaio em campo..... | 38 |
| 1.2 Coleta das informações das folhas | 39 |
| 1.3 Coleta das informações das sementes | 39 |
| 1.4 Análises das imagens..... | 40 |
| 1.5 Processamento dos dados..... | 40 |
| 2. RESULTADOS E DISCUSSÃO | 43 |
| 4.CONCLUSÃO | 52 |
| REFERÊNCIAS | 52 |
| Capítulo III..... | 58 |
| 1.INTRODUÇÃO | 61 |
| 2. MATERIAL E MÉTODOS | 62 |
| 2.1 Instalação do ensaio em campo..... | 62 |
| 2.2 Coleta das informações das folhas..... | 63 |
| 2.3 Coleta das informações das sementes | 63 |
| 2.4 Análises das Imagens | 64 |
| 2.5 Processamento dos dados..... | 64 |
| 3. RESULTADOS E DISCUSÃO | 67 |
| 4. CONCLUSÃO | 76 |
| REFERÊNCIAS | 76 |
| CONSIDERAÇÕES FINAIS | 80 |
| Anexos..... | 81 |

Resumo

A soja (*Glycine max* (L.) Merrill) destaca-se como uma das culturas agrícolas mais cultivadas no planeta. Atualmente, o processo de identificação de sementes é realizado por inspeção visual, que é muito impreciso e oneroso, uma solução é o desenvolvimento de métodos baseados em análise de imagens e aplicação de algoritmos de Machine Learning, o qual exige a prévia determinação do número ideal de sementes ou folhas de uma mesma cultivar a serem coletadas dados morfológicos e espectrais. Dessa forma, o estudo teve como objetivos: (I) identificar algoritmos de Machine Learning para a identificação de cultivares de soja, utilizando imagens de sementes e folhas capturadas por câmeras de smartphones, em combinação com técnicas de inteligência artificial, (II) otimizar o número mínimo de folhas e sementes de soja que devem ser fotografados para prever o valor real de essas cultivares para variáveis morfológicas e espectrais obtidas via análise de imagens. Para isso foram coletadas folhas e sementes de dez cultivares de soja, depois extraídas as informações morfológicas e espectrais a partir das imagens, e com esses dados foi aplicada as análises estatísticas (ANOVA, ACP, AD, K-means) e algoritmos de Machine Learning para classificar as cultivares de soja, aos mesmos dados se aplicou análises de repetibilidade (ANOVA, CP-COV, CP-COR, A. Estructural) para determinar o número ideal de folhas e sementes. Os resultados obtidos demonstram que as variáveis espectrais possuem maior poder classificatório na identificação de cultivares de soja do que as variáveis morfológicas independentemente da estrutura avaliada (folhas ou sementes). Entre os algoritmos testados, o (RAF) destacou-se como o mais eficaz para a classificação das cultivares de soja, com porcentagens corretas superiores a 93% sendo a maior para todos os algoritmos aplicados. Para análises de repetibilidade, devem ser fotografados no mínimo 124 folhas para as variáveis morfológicas das folhas, e 102 folhas para as variáveis espectrais, por outro lado devem ser fotografados no mínimo 344 sementes para variáveis morfológicas das sementes e 183 sementes para as variáveis espectrais.

Palavras - chaves: Machine Learning; Variáveis morfológicas, Variáveis espectrais.

Abstract

Soybean (*Glycine max* (L.) Merrill) stands out as one of the most widely cultivated agricultural crops on the planet. Currently, the seed identification process is carried out by visual inspection, which is very imprecise and costly. One solution is the development of methods based on image analysis and the application of Machine Learning algorithms, which requires the prior determination of the ideal number of seeds or leaves of the same cultivar from which morphological and spectral data will be collected. Therefore, this study aimed to: (I) identify Machine Learning algorithms for the identification of soybean cultivars, using images of seeds and leaves captured by smartphone cameras, in combination with artificial intelligence techniques; (II) optimize the minimum number of soybean leaves and seeds that should be photographed to predict the real value of these cultivars for morphological and spectral variables obtained via image analysis. To this end, leaves and seeds from ten soybean cultivars were collected, then morphological and spectral information was extracted from the images, and statistical analyses (ANOVA, PCA, DA, K-means) and machine learning algorithms were applied to these data to classify the soybean cultivars. Repeatability analyses (ANOVA, CP-COV, CP-COR, Structural Analysis) were applied to the same data to determine the ideal number of leaves and seeds. The results obtained demonstrate that spectral variables have greater classifying power in identifying soybean cultivars than morphological variables, regardless of the structure evaluated (leaves or seeds). Among the algorithms tested, the (RAF) stood out as the most effective for classifying soybean cultivars, with correct percentages exceeding 93%, the highest among all applied algorithms. For repeatability analyses, at least 124 leaves should be photographed for leaf morphological variables, and 102 leaves for spectral variables; conversely, at least 344 seeds should be photographed for seed morphological variables and 183 seeds for spectral variables.

Keywords: Machine Learning; Morphological variables; Spectral variables.

Capítulo I

Introdução e Referencial Teórico

1. INTRODUÇÃO

A soja [*Glycine max* (L.) Merrill] é uma das principais culturas agrícolas no mundo devido ao seu alto potencial econômico, valor proteico, diversos usos na alimentação humana e animal, papel como matéria-prima para fins industriais e bioenergéticos e excelente adaptabilidade a diferentes regiões (Seixas et al., 2020). Como commodity, gera ganhos significativos na balança comercial e contribui diretamente para o fortalecimento da economia global (Montoya et al., 2019).

No entanto, o desempenho produtivo das lavouras de soja pode ser limitado devido a diversos fatores, sendo um deles a baixa qualidade sanitária, fisiológica e/ou genética das sementes utilizadas para semeadura (Ekramirad et al., 2024). A mistura de sementes de soja de cultivares distintas causa problemas substanciais para os agricultores o que eleva a complexidades do mercado de sementes (Liu et al., 2016).

O crescente interesse por estratégias rápidas e não destrutivas para identificação de cultivares tem promovido métodos baseados em análise de imagens. Dentre os métodos disponíveis, a análises de imagens multiespectrais tem se mostrado uma alternativa promissora para avaliar a pureza varietal (Boaelt et al., 2018). Estudos em culturas já relataram a eficácia da aplicação de imagem na identificação de cultivares como em pimentas (Kurtulmus et al., 2016) e arroz (Huang; Chien, 2017; Chaugule; Mali, 2016).

Sensores imageadores de modo geral geram um grande volume de dados, que muitas vezes podem não apresentar relação linear entre si, o que torna técnicas estatísticas paramétricas ineficientes para sua análise. Uma alternativa, seria a aplicação de algoritmos de Machine Learning (ML), que permite o desenvolvimento de algoritmos para serem aplicados em grandes conjuntos de dados e com informações complexas, como dados de imagens (Santana et al., 2023, Marconi et al., 2022).

No entanto, a capacidade dessas técnicas avançadas de Machine Learning aplicadas em dados de imagem de sementes ou de folhas de soja, para distinção de cultivares ainda é um tema pouco elucidado na cultura.

O desenvolvimento de técnicas de identificação de cultivares de soja através de imagens, como de sementes e folhas, exige a prévia determinação do número ideal de órgãos de uma mesma cultivar a serem coletadas dados espectrais. Essa etapa define o número mínimo de essas estruturas da planta que devem ser fotografados para assegurar que o resultado da avaliação reflita o valor real da cultivar com confiabilidade, ao mesmo tempo em que se otimiza o tempo de análise.

Para esse fim, existem diversas metodologias estatísticas, como, por exemplo, a estimativa do coeficiente de repetibilidade. Trata-se de um parâmetro amplamente utilizado para quantificar se as medições repetidas de uma característica são as mesmas, expressando a variação total em termos da contribuição do genótipo e do ambiente (Ferreira et al., 2010). Assim, valores elevados da estimativa de repetibilidade da característica indicam que é possível prever o valor real do indivíduo com um pequeno número de medições (Cruz et al., 2012).

Esses métodos têm sido aplicados em diferentes culturas segundo a literatura (Diel et al., 2020; Ferreira et al., 2020; Pedrozo et al., 2011). Na soja, a estimativa do coeficiente de repetibilidade foi aplicado na avaliação de descritores morfológicos (Matsuo et al., 2012), na avaliação do teor de óleo e proteína (Jiang, 2020), e nas associações entre métodos de análise de adaptabilidade, estabilidade e produtividade (Woyann et al., 2018).

No entanto, a utilização de análises de repetibilidade favorece a melhor definição da quantidade de plantas a serem medidas, para se obter o valor real do indivíduo, com confiabilidade e otimização do tempo na avaliação. E como até o momento, não foram identificados estudos que visassem a determinação do número mínimo necessário de estruturas de plantas que devem ser fotografadas para estudos com imagens em soja, diante disto as hipóteses do trabalho foram: (I): será que o processamento de imagens de folhas e sementes aplicando inteligência artificial é eficiente na classificação de cultivares de soja?, (II): qual será o número mínimo de folhas e sementes a ser fotografadas para as cultivares de soja a ser classificadas?.

Para dar resposta a isso, os objetivos do trabalho foram: (I) : identificar algoritmos de Machine Learning para a identificação de cultivares de soja, utilizando imagens de sementes e folhas capturadas por câmeras de smartphones, em combinação com técnicas de inteligência artificial, (II): otimizar o número mínimo de folhas e sementes de soja que devem ser fotografados para prever o valor real de essas cultivares para diferentes variáveis morfológicas e espectrais obtidas via análise de imagens.

2. REFERENCIAL TEÓRICO

2.1 Origem da soja

A soja (*Glycine max* (L.)) tem centro de origem no nordeste da China (entre 45 e 50° N). A introdução da soja no Ocidente ocorreu em latitudes similares ao centro de origem, iniciando pela Europa em 1712, tendo sido descrita em 1737 por Linné (*Linnaeus*) nas *Cliffortianus Hortus* (Gazzoni, 2018). O primeiro relato sobre cultivo de soja nos Estados Unidos é de 1765. Embora lentamente, a soja se expandiu naquele país, exigindo investimento em pesquisas que propiciaram um desenvolvimento tecnológico que redundou em cultivares de soja produtivos, resistentes a doenças, em um sistema produtivo adaptado às diferentes condições de produção norte-americanas.

Em 1882, Gustavo D'Utra efetuou primeiro cultivo de soja na Bahia. Empreitada que fracassou posto que o material genético, desenvolvido para climas frios ou temperados, não se adaptou às condições da Bahia. Em 1891, cultivares de soja foram testados no Instituto Agrônomo de Campinas (IAC-SP). A efetiva trajetória de sucesso da produção comercial de soja somente tem início no Rio Grande do Sul, no período de 1920 a 1940. Antes disso, em 1901, o professor Guilherme Minssen, da então Escola Superior de Agronomia e Veterinária, atualmente vinculada à Universidade Federal do Rio Grande do Sul (UFRGS), iniciara pesquisas com a leguminosa naquele estado (Gazzoni, 2018). A produção da soja no Brasil em larga escala teve início na década 1950, expandindo-se do Sul do país para o Norte, atravessando os biomas da Mata Atlântica, do Cerrado e, finalmente, atingiu o bioma Amazônico (Lima et al., 2019; Piras et al., 2021; Sauer, 2018).

Na atualidade, a maior parte da produção do país é oriunda do bioma do Cerrado (Rausch et al., 2019). A soja se expandiu no Estado Maranhense, ganhando destaque. Isso se deu pela combinação de preços, subsídios governamentais, parceria com a Empresa Brasileira de pesquisa Agropecuária (EMBRAPA) para a realização de estudos com a sojicultura, e a Companhia Vale do Rio Doce (CVRD) para o uso da Estrada Ferroviária de Carajás (EFC) no escapamento da produção por meio do porto de Itaqui (Rodrigues; Terra, 2018).

Até o final da década de 1980, as áreas cultivadas com soja no Maranhão eram bastante reduzidas. A partir de 1990, observou-se uma expressiva evolução do cultivo de terras com essa cultura. De fato, até o final de 1980 o IBGE (Instituto Brasileiro de Geografia e Estatística) registrou a existência de apenas 80 hectares de soja no Maranhão. A partir da década de 1990, observou-se uma verdadeira explosão no cultivo dessa commodity, inicialmente nos cerrados do sul do estado e depois no Baixo Parnaíba, no leste maranhense, de tal forma que em 1995 as

áreas ocupadas com soja atingiram 87,7 mil hectares no Maranhão; 178,7 mil hectares em 2000; e 400 mil hectares em 2010 (IBGE, 2020).

2.2 Taxonomia, botânica e morfologia

A espécie vegetal *Glycine max* (L.) Merrill, é uma planta leguminosa pertencente ao reino *Plantae*, divisão *Magnoliophyta*, classe *Magnoliopsida*, ordem *Fabales*, família *Leguminosae*, subfamília *Papilionoideae*, gênero *Glycine*, espécie *max* (Oliveira et al., 2017). A família *Fabaceae* é uma das maiores famílias botânicas composta por 730 gêneros e 19.325 espécies (Mello et al., 2015). O gênero *Glycine* se divide em dois subgêneros, a *G. max* ($2n=40$ cromossomos), que é a espécie cultivada, e a *G. soja* ($2n=40$ cromossomos), que é a espécie selvagem (Oliveira et al., 2017).

A soja é caracterizada como uma planta herbácea, isto é, planta de porte pequeno e de caule flexível. A cultura é descrita como anual, ou seja, é uma espécie que conclui seu ciclo produtivo em menos de 1 ano. As sementes da espécie apresentam germinação do tipo epígea, logo, os cotilédones e a gema apical se elevam acima do solo devido ao alongamento do hipocótilo da plântula (Matsuo et al., 2015).

Em relação ao sistema radicular da soja, ele é caracterizado como difuso e predominantemente axial fasciculado, composto pela raiz principal e suas raízes secundárias (Matsuo et al., 2015). Ainda, podem ser observados nódulos, consequentes da relação simbiótica com as bactérias do gênero *Bradyrhizobium*, responsáveis por realizar o processo denominado fixação biológica de nitrogênio (Matsuo et al., 2015; Sedimaya et al., 2015).

O sistema caulinar da soja é descrito como herbáceo, ereto, isto é, apresenta crescimento perpendicular ao solo, pubescente, ou seja, com penugem, e, por fim, ramificado (Matsuo et al., 2017). A coloração de seus tricomas é acinzentada ou em tons variados de marrom (Matsuo et al., 2015; Sedimaya et al., 2015). Vale ressaltar que na maioria dos genótipos da espécie o caule apresenta crescimento ortótropo, ou seja, apresenta crescimento na vertical em relação ao estímulo da luz, entretanto, as plantas podem atingir até 3,0 m se forem cultivadas sob-baixa luminosidade e fotoperíodo longo, o que proporciona caules delgados e volúveis que necessitam de tutoramento (Matsuo et al., 2017).

As plantas de soja apresentam ao longo do seu desenvolvimento quatro tipos de folhas, as cotiledonares, responsáveis pelo suprimento de reserva energética ao desenvolvimento da plântula, de forma elíptica oval; as folhas unifolioladas, que ocorrem de forma e desenvolvem-se no primeiro nó logo acima do nó cotiledonar, podem apresentar forma auriculada, truncada

ou lanceolada, com base estreita ou larga, as folhas trifolioladas com filotaxia alterna compostas por três folíolos, variando no formato de lanceolada estreita, lanceolada, triangular, oval pontiaguda e oval arredondada e desenvolvem-se nos nós subsequentes à folha unifoliolada, as prófios presentes em pares nas bases das ramificações caulinares, com pulvínulo e pecíolos ausentes (Matsuo et al., 2015; Oliveira et al., 2017).

A soja é uma espécie autógama, isto é, apresenta taxa de autofecundação superior a 95%, a qual ocorre de maneira natural, através de mecanismos físicos e biológicos (Borém et al., 2021; Matsuo et al., 2015). Entretanto, foram verificadas taxas de 1,27% de hibridação natural, por abelhas e tripes (Borém et al., 2009).

A soja apresenta um mecanismo para impedir que ocorra a fecundação cruzada, denominado cleistogâmica, ou seja, este faz com que ocorra a polinização da flor antes da abertura do botão floral (Borém et al., 2009; Borém et al., 2021). Outra característica reprodutiva fundamental para as hibridações artificiais da soja é o processo denominado protogenia, isto é, o pistilo amadurece antes que as anteras (Borém et al., 2009).

As flores da soja são completas e ocorrem em racemos terminais ou axilares. Cada inflorescência contém de 2 a 35 flores, variando de 6 a 7 mm de comprimento (Matsuo et al., 2015). A planta é hermafrodita com flores monoclinas, ou seja, apresenta tanto as estruturas femininas como as masculinas na mesma flor. A coloração das suas pétalas pode ser branca ou tons variados de roxo, conforme o genótipo e, vale ressaltar, que devido ao efeito pleiotrópico do gene que condiciona a cor de flor em relação a cor do hipocótilo, quando a flor é branca, o hipocótilo é verde, e quando é roxa, o hipocótilo, também, é roxo (Verneti; Junior, 2009).

O fruto da soja é descrito botanicamente como legume, também comumente chamado de vagem, variando desde o formato arredondado, achatado, reto até o curvado. Em relação à sua dimensão, esta apresenta de 2 a 7 cm de comprimento e 1 a 2 cm de largura, podendo conter de 1 a 5 grãos, mas geralmente varia de 2 a 3 grãos (Matsuo et al., 2015). O hilo possui variação na coloração, podendo ser cinza clara, cinza escura, marrom clara, marrom média e marrom escura, preto, preto imperfeito, enquanto as vagens cinza, amarelo-palha, marrom ou preta (Oliveira et al., 2017).

As sementes de soja são descritas quanto à forma como esférica, esférica-achatada, alongada e alongada achatada, enquanto que a coloração do tegumento pode variar entre amarelo, amarelo esverdeado, verde, marrom claro, marrom médio, marrom escuro até o preto (Matsuo et al., 2015).

2.3 Exigências edafoclimáticas

A cultura da soja depende de condições edafoclimáticas favoráveis para o desenvolvimento no campo. Tais condições abrangem as exigências hídricas, fotoperíodo e temperatura. A atuação direta da água no metabolismo da planta é fundamental para manutenção das funções vitais de transporte de substâncias, controle estomático, termorregulação, dentre outras. Logo na fase de germinação da soja é necessário que o nível de água no solo se mantenha entre 50% e 85%, valores inferiores ou superiores a essa margem afetam o desenvolvimento da planta (Wani et al., 2012). No período de floração e enchimento de grãos, o consumo de água aumenta atingindo valores de 7-9 mm/dia; o consumo total varia, em geral, de 300 mm/ciclo a 800 mm/ciclo para bons rendimentos produtivos, dependendo da região, tipo de solo e cultivares (Farias et al., 2007; Wani et al., 2012). Nos casos de déficit hídrico durante a floração, o valor de perda estimado para a produtividade pode atingir até 52%, quando o período de desenvolvimento atingido pela seca abrange o enchimento dos grãos, os efeitos se tornam mais severos (Domuta et al., 2018).

O cultivo da soja demonstra maior desenvolvimento e produtividade em regiões com temperaturas entre 20 °C e 30 °C, com melhores valores obtidos a 30 °C, para o solo, a temperatura ideal pode variar, também, entre 20 °C e 30 °C, com melhor desempenho a 25 °C, temperaturas baixas (10 °C) reduzem o crescimento e, em alguns casos, altas temperaturas (por exemplo, 40 °C) estimulam o crescimento, em todos os casos, a floração e a produção das vagens são comprometidas (Wani et al., 2012; Embrapa, 2013).

Fotoperíodo é conceituado basicamente pelo número de horas de luz ocorrido durante o dia, ou seja, intervalo de tempo (em horas) do amanhecer ao anoitecer. É um aspecto importante na definição dos cultivares nas diferentes regiões de produção. As escolhas das cultivares de acordo com o fotoperíodo é um indicador para definir a duração do período vegetativo, para sempre incidir com o período de maior radiação solar, fator este que possui alta relação com a produtividade de grãos (Mundstock, 2005).

A soja é uma planta de dia curto no qual o seu fotoperíodo estabelece relação direta com o florescimento. À medida que a latitude se altera o comprimento do dia também é alterado gerando problemas de adaptação do cultivar e conseqüente atraso do florescimento. Cultivares com período juvenil longo têm mais facilidade para se adaptar, por dispor de permanência maior no estágio vegetativo que as cultivares convencionais (Farias et al., 2007).

A influência do fotoperíodo sobre o desenvolvimento da soja é variável entre as cultivares, ou seja, cada cultivar possui um número de horas de luz, chamado de fotoperíodo

crítico, acima do qual, o florescimento é atrasado. Por isso a soja é considerada uma planta de dias curtos. Plantas de dias curtos, são induzidas ao florescimento quando a duração do dia é menor do que o seu fotoperíodo crítico. Quando a duração do dia é maior do que o fotoperíodo crítico, as plantas não recebem o sinal para florescer e continuam no período vegetativo. Portanto o manejo do fotoperíodo e temperatura, através da escolha da cultivar e da data de semeadura é preponderante para obter sucesso na produção de soja (Rodrigues et al., 2001).

Na cultura da soja, a sensibilidade fotoperiódica varia com o genótipo, e o grau de resposta ao estímulo fotoperiódico é o principal determinante da área de adaptação das diferentes cultivares. Genótipos que possuem um maior período juvenil (período inicial do desenvolvimento, em que a planta não é sensível ao fotoperíodo) são mais adaptáveis ao cultivo em diferentes latitudes e épocas de plantio (Farias et al., 2007).

2.4 Processamento de imagens

De acordo com Gonzales; Woods (2002), o processamento de imagens digitais, ou simplesmente processamento de imagens pode ser definido como o processo de modificação de imagens digitais. Estas modificações podem ter por objetivo a interpretação humana ou ainda a percepção automática de máquinas. A definição de imagem digital é baseada no conceito de imagem monocromática ou simplesmente imagem.

Os modelos de cor podem ser orientados ao hardware ou orientados a aplicação. Dentre os modelos orientados ao hardware mais comumente utilizados na prática está o RGB (Red, Green, Blue). Este modelo é utilizado em monitores coloridos e também por uma ampla classe de câmeras de vídeo (Gonzales; Woods, 2002). Imagens no modelo RGB consistem em três planos de imagens independentes (camadas). Quando alimentadas num monitor RGB, estas camadas combinam-se sobre a tela para produzir uma imagem de cores compostas.

O processamento de imagens é normalmente dividido em: aquisição da imagem, pré-processamento, segmentação, extração de características, reconhecimento e interpretação da imagem (Márquez; Vieira, 1999).

1-Aquisição da imagem: A primeira etapa é a aquisição das imagens. Para tanto, são necessários alguns sensores e um digitalizador. Os sensores convertem a informação óptica em sinais elétricos e o digitalizador transformará a imagem analógica em imagem digital. Dentre os aspectos de projeto envolvidos nesta etapa, pode-se mencionar: a escolha do tipo do sensor,

os conjuntos de lentes a utilizar, as condições de iluminação da cena, a resolução espacial e o número de níveis de cinza da imagem. Esta etapa produz uma imagem digitalizada como saída.

2- Pré-processamento: A imagem resultante do passo anterior pode apresentar diversas imperfeições, tais como: presença de pixels ruidosos além de contraste ou brilho inadequado.

A função da etapa de pré-processamento é aprimorar a qualidade da imagem para as etapas subsequentes. A imagem resultante desta etapa é uma imagem digitalizada de melhor qualidade que a original.

3-Segmentação: A tarefa básica da segmentação é dividir uma imagem digital em suas unidades significativas, ou seja, nos objetos de interesse que a compõem. Esta tarefa, apesar de simples de descrever, é das mais difíceis de implementar.

4-Extração das características: Esta etapa procura extrair características das imagens resultantes da segmentação através de descritores que permitam caracterizar estes objetos. Estes descritores devem ser representados por uma estrutura de dados adequada ao algoritmo de reconhecimento. É importante observar que nesta etapa a entrada ainda é uma imagem, mas a saída é um conjunto de dados correspondentes àquela imagem.

5-Reconhecimento e interpretação: Nesta última etapa, denominamos reconhecimento, o processo de atribuição de um rótulo a um objeto baseado em suas características, traduzidas por seus descritores. A tarefa de interpretação, por outro lado, consiste em atribuir significado a um conjunto de objetos reconhecidos.

2.5 Inteligência artificial na Agricultura

A IA surgiu na década de 1950, com uso de conceitos matemáticos e teoria da computação, disseminando-se com o avanço do poder de processamento de informatização e uso de big data. O desenvolvimento da IA inclui marcos como sistemas especialistas, redes neurais e aprendizado profundo, permitindo aplicações complexas em reconhecimento de padrões, processamento de linguagem natural e automação avançada (Mendoza et al., 2023).

As ferramentas de IA vêm sendo aplicadas à gestão com o objetivo de otimizar decisões, reduzir custos e expandir a produtividade. Por meio do uso de sensores, por exemplo, dados são coletados e analisados em nuvem, permitindo ações automatizadas e preditivas nas áreas de produção, como clima, análise e manejo de solo. Isso fortalece a eficácia e eficiência operacional e auxilia no alcance aos requisitos relativos à sustentabilidade nas propriedades rurais (Sebrae, 2021).

A digitalização altera a lógica da produção e reprodução do capital no setor do agronegócio, pois a integração de sensores, com uso de técnicas e ferramentas de IA e automação transformam tarefas rotineiras em fluxos de informação contínua que sustentam decisões gerenciais em tempo real (Massruhá et al., 2020). Na Agricultura 4.0, as tecnologias deixam de ser mero suporte técnico e passam a orientar estratégia, produtividade e sustentabilidade em toda a cadeia (Bolfé, 2020; Massruhá et al., 2020).

Contudo, mesmo com o avanço de máquinas modernas e sistemas digitais, muitos agricultores não utilizam os recursos disponíveis por falta de conhecimento, o que expõe a importância da formação técnica nesse novo cenário. Os demais pilares complementam essa estrutura: a produção, voltada à eficiência operacional; a gestão, relacionada à análise e uso dos dados na tomada de decisões; e a sustentabilidade, que visa aumentar a produtividade com responsabilidade ambiental e minimização de custos operacionais. Essa transformação exige reconfigurar processos, cultura organizacional e modelos de negócios inteligentes. A conectividade em nuvem, análise de grandes volumes de dados e plataformas colaborativas criam propriedades rurais orientadas por inteligência operacional (Massruhá et al., 2020). Conseqüentemente, gestores precisam dominar técnicas de inovação, gestão de dados e indicadores de desempenho para alinhar rentabilidade e responsabilidade socioambiental e empresarial.

Modelos preditivos sustentados por imagens de satélite, drones e sensores permitem otimizar irrigação, adubação e controle fitossanitário, mesmo em pequenas propriedades rurais, desde que disponham de conectividade a redes de comunicação (Munz et al., 2024). Assim, a transformação digital no campo impõe uma agenda de gestão estratégica da tecnologia. A literatura aponta que conectividade, interoperabilidade e governança de dados são condicionantes para que a Agricultura 4.0 evolua rumo a modelos de negócio baseados em serviços e plataformas, reduzindo riscos gerados a partir de fatores operacionais (OECD, 2022).

O mercado global de IA na agricultura está em acelerado processo de expansão. No ano de 2023, os negócios relacionados ao ramo foram estimados em US\$ 1,7 bilhão, com expectativa de crescimento de US\$ 4,7 bilhões até 2028, o que corresponde a uma taxa de crescimento anual composta de 23,1% (Markets, 2023).

Outra importante técnica utilizada no âmbito de aplicação da IA é o Machine Learning (ML) ou “Aprendizado de Máquina”. Esta técnica diz respeito à capacidade das máquinas reconhecerem padrões a partir do uso de bancos de dados disponíveis. Por exemplo, existe a possibilidade de desenvolver um modelo que antecipa a informação quanto a qualidade do café examinando dados como a altura do cultivo, volume de chuva, temperatura, propriedades do

solo (tipo, fertilidade e umidade), exposição à luz solar, a coloração dos grãos, entre outros indicadores. Apesar de parecer uma tecnologia avançada, esse processo pode ser tão elementar quanto a aplicação de uma simples técnica estatística, como a regressão linear (Queiroz et al., 2022). Existe diversos algoritmos de Machine Learning, entre eles destaca-se:

1- O algoritmo “Naive Bayes” é um classificador probabilístico muito utilizado em machine learning. Baseado no “Teorema de Bayes”, ele assume que todas as características (features) são independentes entre si, simplificando significativamente os cálculos (Raschka; Mirjalili, 2017).

2- O algoritmo Instance-Based Learning (IB1) é um algoritmo de aprendizado baseado em instâncias (instance-based learning), pertencente à família dos lazy learners. Ele funciona basicamente como uma versão simples do k-Nearest Neighbors (k-NN), com $k = 1$ e distância euclidiana como métrico padrão. Foi desenvolvido por David W. Aha, Dennis Kibler e Marc K. Albert em 1991 (David et al. 1991).

3- O algoritmo Rotation Forest (ROF) chamado Floresta de Rotação foi desenvolvido por Juan J Rodríguez, Ludmila I, Kuncheva e Carlos J 2006 , na qual cada árvore de decisão é treinada aplicando primeiro a análise de componentes principais (ACP) em um subconjunto aleatório dos recursos de entrada (Rodriguez et al. 2006).

4- O algoritmo Random Forest, ou Floresta Aleatória, é um método de aprendizagem de conjunto para classificação, regressão e outras tarefas que funcionam criando uma infinidade de árvores de decisão durante o treinamento (Ho Tk, 1998). Floresta aleatórias corrigem o hábito das árvores de decisão de se sobreajustarem ao seu conjunto de treinamento , sendo o primeiro algoritmo para florestas de decisão aleatórias foi criado em 1995 por Tin Kam Ho (Tin Kan, 2016).

5- O algoritmo Multilayer Perceptron foi desenvolvido em 1958 por Frank Rosenblatt quem foi um psicólogo americano , propôs o modelo constituído por uma camada de entrada, uma camada oculta com pesos aleatórios que não aprendiam e uma camada de saída com conexões aprendíveis (Rosenblatt, 1958). Consiste em neurônios totalmente conectados com funções de ativação não lineares , organizados em camadas, notáveis por serem capazes de distinguir dados que não são linearmente separáveis (Cybenko, 1989).

Destaca-se a diferença entre IA e ML, a IA engloba sistemas capazes de imitar processos cognitivos humanos, enquanto o ML, subcampo da própria IA, concentra-se em algoritmos que aprendem de forma autônoma a partir da análise de grandes volumes de dados (Gyamfi et al., 2024). Nesse sentido, o ML é parte integrante da IA que confere às máquinas mecanismo de aprendizado independente a partir de um histórico de experiências acumuladas, sem necessitar

de programação explícita. Inclui três categorias: supervisionado, não supervisionado e por reforço (Ahmad; Nabi, 2021). Desta maneira, ML é um método que utiliza dados de treinamento para criar modelos matemáticos capazes de fazer previsões e tomar decisões (Bal; Kayaalp, 2021).

2.6 Produção de sementes de soja

São muitas as decisões de manejo que interferem na produtividade de uma lavoura tais como: manejo do solo e de irrigação, controle de pragas e doenças, adubação, controle de plantas daninhas, o posicionamento de cultivares e a utilização de sementes de qualidade. Entre esses, a utilização de sementes de qualidade pode-se dizer que é o mais importante, pois nenhuma outra prática de manejo citadas acima pode melhorar a produtividade além do limite imposto pela semente (Zimmer, 2017).

Produzir sementes de soja de elevada qualidade em regiões tropicais e subtropicais é um grande desafio, pois regiões em que há predominância destes climas, necessitam de técnicas específicas para se produzir sementes. Quando o produtor de sementes não leva em consideração as boas práticas citadas anteriormente, o resultado pode ser o decréscimo na qualidade das sementes produzidas, e que quando utilizadas resultará em prejuízo, pois a produtividade será comprometida (França-Neto et al., 2007). A baixa disponibilidade nutricional, as mudanças climáticas bruscas e os danos ocasionados por insetos ou microrganismos são as principais causas da perda de qualidade da semente no campo. A elevação da umidade da semente por um longo período ou períodos intermitentes após a maturação fisiológica e antes da colheita, também ocasionam a deterioração da semente (Henning, 2025).

No Sul do Brasil, onde o clima predominante é o subtropical, são estabelecidas datas de semeaduras diferentes para a produção de grão e para a produção de semente. A produção de grão tem como objetivo alcançar a máxima produção por área, já a produção de semente tem como objetivo alcançar a máxima qualidade produzida. A data de semeadura para a produção de semente deve ser escolhida de forma que a maturação da semente ocorra em uma época predominante de temperaturas amenas e menores índices pluviométricos (França-Neto et al., 2007).

Estresses ambientais, como seca ou veranicos, elevadas temperaturas ou ocorrências de geada intensa, podem levar a morte prematura da planta ou à maturação forçada. Eventos esses que levam a redução de produtividade, produção de sementes esverdeadas, doenças de raiz e de

folhas e intenso ataque de pragas. Quando há ocorrência desses eventos, a qualidade de uma lavoura para semente é totalmente comprometida (França-Neto et al., 2016).

2.7 Repetibilidade

De acordo com Lush (1937), citado por Resende (2002), a repetibilidade é conceituada como correlação fenotípica intraclasse e, portanto, no caso de indivíduos, refere-se às correlações fenotípicas entre medições repetidas no mesmo indivíduo. O coeficiente de repetibilidade mede, portanto, a capacidade dos organismos em repetir a expressão do caráter, ao longo de vários períodos de tempo ou no espaço, no decorrer de suas vidas (Resende, 2002). Ao selecionar um genótipo, espera-se que suas características ou o seu desempenho sejam mantidos por toda sua vida, e tal expectativa pode ser confirmada pela determinação do coeficiente de repetibilidade da característica desejada (Cruz ; Regazzi, 2001).

Conforme Cruz e Regazzi (2001), a repetibilidade representa o limite superior da herdabilidade, tanto no sentido amplo como no restrito. Os valores de repetibilidade são sempre mais elevados do que os da herdabilidade, pois a repetibilidade inclui, além dos efeitos aditivos, os efeitos não aditivos e também são acrescidas algumas diferenças relacionadas ao ambiente permanente, cuja variância geralmente permanece confundida com a variância genotípica.

O coeficiente de repetibilidade (r), varia de 0 a 1, sendo esta a repetibilidade máxima, que se manifesta quando o caráter ocorre com muita constância (Soares, 2005). A repetibilidade admite determinar o número de medidas necessárias para o cálculo do coeficiente de determinação (R^2) para predição do valor real dos indivíduos, sendo necessário pequeno número de medidas repetidas se a repetibilidade é alta, e grande número de avaliações quando o coeficiente de repetibilidade é baixo. Permitindo-se avaliar o ganho de precisão com a realização de várias medidas, pois o aumento do número de avaliações reduzirá o valor da variância ambiental e a redução da variância fenotípica representará ganho em precisão. Entretanto, o aumento do número de medições proporciona pequenos acréscimos de precisão quando a repetibilidade for alta, mas se baixa, as medidas múltiplas podem conduzir a um ganho de precisão, apesar de que são rapidamente reduzidos com o aumento de medições (Falconer, 1981; Soares, 2005).

A repetibilidade pode ser estimada por diferentes métodos, que são: o método de correlação intraclasse da análise de variância (ANOVA), utilizando modelo estatístico com um ou dois fatores de variação, sendo que ao utilizar dois fatores é possível remover efeitos do ambiente temporário, que certamente ficam confundidos com a variação dentro do genótipo,

levando a uma subestimação do coeficiente de repetibilidade. O método dos componentes principais é citado como uma das formas mais eficientes para o cálculo do coeficiente de repetibilidade, principalmente em espécies que apresentam comportamento cíclico ou alternância de produção, variável entre os genótipos, efeito não eliminado quando da realização de análise de variância, levando a uma subestimação da repetibilidade.

O método da análise estrutural, semelhante ao dos componentes principais, com diferenças conceituais, considera as correlações entre genótipos, em cada par de avaliação, na constituição da matriz paramétrica, para definição do estimador de repetibilidade (Cruz ; Regazzi, 2001). As estimativas do coeficiente de repetibilidade, do número de medições necessárias para uma predição adequada e do coeficiente de determinação (R^2) para o número de medições realizadas também podem ser obtidas por meio do programa Genes, para todos os métodos descriptos (Cruz, 2006).

REFERÊNCIAS

AHMAD, L.; NABI, F. **Agriculture 5.0**. Ed.1. Boca Raton: CRC Press, 2021. Ebook. ISBN 978-1-003-12543-3.

ALMEIDA, L. A; KIIHL, R. A. **Melhoramento da soja no Brasil - desafios e perspectivas**. In: Soja: Tecnologia da Produção. Gil. M. S. Câmara (ed.). Piracicaba, SP, USP-ESALQ, p. 40-54, 1998.

BAL, F.; KAYAALP, F. Review of machine learning and deep learning models in agriculture. **International Advanced Researches and Engineering Journal**, v. 5, n. 2, p. 309-323, 2021. DOI: [10.35860/iarej.848458](https://doi.org/10.35860/iarej.848458).

BERNS, R. S. **Billmeyer and Saltzman's principles of color technology**. John Wiley & Sons, 2019.

BOELT. B.; SHRESTHAS, S.; SALIMI, Z.; JORGENSEN, J. R.; NICOLAISEN, M.; CARSTENSE, J.M. Multispectral imaging - a new tool in seed quality assessment? **Seed Science Research**, v. 28, n. 3, p. 222-228, 2018. DOI: <https://doi.org/10.1017/S0960258518000235>.

BOLFE, É. L.; MASSRUHÁ, S. M. A transformação digital e a sustentabilidade agrícola. **Agroanalysis**, Rio de Janeiro, v. 40, n. 3, p. 32-34, 2020.

BORÉM, A. **Hibridação Artificial de Plantas**. 2º Edição, Editora UFV, p.625, 2009.

- BORÉM, A.; MIRANDA, G.V.; FRISTSCHÉ-NETO, R. **Melhoramento de Plantas**. 8. ed. Viçosa, MG: Editora UFV, 2021.
- CHAUGULE, A. A; MALI, S. N. Identification of rice varieties based on new angular characteristics of the seeds. **Comput. Electron. Agric**, v. 123, p. 415–422, 2016.
- CRUZ, C. D.; REGAZZI, A. J. **Modelos biométricos aplicados ao melhoramento genético**. Viçosa, UFV. P. 390, 2001.
- CRUZ, C. D. **Programa genes: Biometria**. Viçosa: UFV, p. 382, 2006.
- CRUZ, C. D.; REGAZZI, A. J.; CARNEIRO, P. C. S **Modelos biometricos aplicados ao melhoramento genetico**. Viçosa, MG: UFV, v. 1, 2012.
- CYBENCO, G. Aproximação por superposições de uma função sigmoideal Matemática de Controle. **Sinais e Sistemas**, v. 2, p. 303–314, 1989.
- DAVID, W.; KLIBER, D.; ALBERT, K. M. **Instance-based learning algorithms. Machine Learning**, Department of Information and Computer Science, University of California, Irvine, CA 92717, p. 37, 1991.
- DIEL, M. I.; LÚCIO AD.; OLIVOTO, T.; PINHEIRO, M.V.M.; KRYSCZUN, D.K.; SARI, B.G.; Schmidt, D. Repeatability coefficients and number of measurements for evaluating traits in strawberry. **Acta Scientiarum. Agronomy**, v. 42, p. 1-9, 2020.
- DOMUTA, C.; PEREŞ, A.; BREJEA, R.; BORZA, I.; BEI, M.; NANDOR, K.; GÎTEA, M.; JUDE, E. Pedological drought influence on some parameters. **Romanian Agricultural Research**, n. 35, 2018.
- EMBRAPA. Tecnologias de Produção de Soja: Região Central do Brasil 2014. **Sistemas de Produção**, n. 16. Londrina: Embrapa Soja, 2013.
- KRAMIRAD, N.; DOYLE, L.; LOEB, J.; SANTRA, D.; ADEDEJI, A. A. Hyperspectral imaging and machine learning as a nondestructive method for proso millet seed detection and classification. **Foods**, v. 13, n. 9, p. 1330, 2024.
- FALCONER, D. S. **Introdução à genética quantitativa**. Viçosa: UFV, p. 279, 1981.
- FARIAS, J. R.; NEPOMUCENO, A. L.; NEUMAIER, N. Ecofisiologia da Soja. Londrina: Embrapa Soja. (Circular Técnica 48, **Embrapa Soja**), 2007.

FERREIRA, R. D. P., DE VASCONCELOS, E. S., CRUZ, C. D., JÚNIOR, W. B., RASSINI, J. B., DE FREITAS, A. R., VIELA, D.; MOREIRA, A. Determinação do coeficiente de repetibilidade e estabilização genotípica das características agronômicas avaliadas em genótipos de alfafa no ano de estabelecimento. **Revista Ceres**, v. 57, n. 5, p. 642-647, 2010.

FERREIRA, F. M.; DE CARVALHO ROCHA, J. R.; ALVES, R. S.; ELIZEU, A. M.; BENEDITES, F. R. G.; DE RESENDE, M. D.V. Estimates of repeatability coefficients and optimum number of measures for genetic selection of *Cynodon spp.* **Euphytica**, v. 216, n. 5, 2020.

FRANÇA-NETO, J. B.; KRZYZANOWSKI, F.C; HENNING, A. A; PÁDUA, G. P; LORINI, I; HENNING, F. A. Tecnologia da produção de semente de soja de alta qualidade – Série Sementes. Londrina: **Embrapa Soja**, n. 40, p. 12, 2007.

FRANÇA-NETO, J. B.; KRZYZANOWSKI, F.C; HENNING, A. A; PÁDUA, G. P; LORINI, I; HENNING, F. A. Tecnologia da produção de semente de soja de alta qualidade – Série Sementes. Londrina: **Embrapa Soja**, n. 40, p. 12, 2016.

GAZZONI, D. L. A soja no Brasil é movida por inovações tecnológicas. **Ciência e Cultura**, v. 70, n. 3, p. 16-18, 2018.

GONZALES, R. C.; WOODS, R. E. **Digital Image Processing**. 2. ed. Saddle River: Prentice Hall, p.793, 2002.

GYAMFI, E. K.; ELSAYED, Z.; KROPCZYNSKI, J.; YAKUBU, M. A.; ELSAYED, N. **Agricultural 4.0 leveraging on technological solutions: study for smart farming sector**. University of Cincinnati, 2024.

HARTWIG, E. E; KIIH, R. A. L. **Identification and utilization of a delayed flowering character in soybean for short-day conditions**. Field Crops Res. 2: p.145- 151, 1979.

HENNING, A. A. **Patologia e tratamento de sementes: Noções gerais**. 2.ed. Londrina: Embrapa Soja, n. 264, p. 52, 2025.

HINSON, K. **The use of long juvenile trait in cultivar development**. In. Conferência Mundial de Investigación en Soja, 4. B. Aires, Argentina. Actas. A.J. Pascale (ed.). p.983-987,1989.

- HO, T. K. Random Subspace Method for Decision Forest Construction. **IEEE Transactions on Pattern Analysis and Machine Intelligence** v. 20, p. 832–844, 1998.
- HUAN, K.Y; CHEIN, M.C. Um novo método para identificar variedades de sementes de arroz. **Sensors**, v. 17, p. 809, 2017.
- INSTITUTO BRASILEIRO DE GEOGRAFIA E ESTATÍSTICA – IBGE (2020). Produção Agrícola Municipal (PAM).Disponível:<https://sidra.ibge.gov.br/pesquisa/pam/tabelas>.
- JIANG, G. L. Comparison and application of non-destructive NIR evaluations of seed protein and oil content in soybean breeding. **Agronomy**, v. 10, n. 1, p. 77, 2020.
- KIIHL, R. A; GARCIA. A. **The use of the long juvenile trait in breeding soybean cultivars**. In: Conferencia Mundial de Investigacion en Soja, 4. Buenos Aires, Argentina. Actas. A.J. Pascale (ed.). p. 994-1000, 1989.
- KURTULMUS, F.; ALIBAS, I.; KAVDIR, I. Classification of pepper seeds using machine vision based on neural network.**Int. J. Agric. Biol. Eng.** v. 9, p. 51–62, 2016.
- LIMA, M.; SILVA JUNIOR, C. A.; RAUSCH, L.; GIBBS, H. K.; JOHANN, J. A. Demystifying sustainable soy in Brazil. **Land Use Policy**, v. 82, p. 349-352, 2019.
- LIU, Y.; XIE, H.; CHEN, Y.; TAN, K.; WANG, L.; XIE, W. Mutual neighborhood information and its application in hyperspectral band selection for classification. **Chemom. Intell. Lab. Syst**, v. 157, p. 140–151, 2016.
- MARCONI, S.; WEINSTEIN, B. G.; ZOU, S.; ZARE, A.; STERWARTS, P.; HARMON, I. Hyperspectral classification of tree species at the continental scale in the United States National Network of Ecological Observatories **Remote Sens. Environ**, v. 282, p. 113264, 2022.
- MARQUES, F. O.; VIEIRA, N. H. **Processamento Digital de Imagens**. Rio de Janeiro: Brasport, p.331, 1999.
- MARKETS AND MARKETS. **Artificial Intelligence in Agriculture Market Size & Growth**. São Francisco, 22 out, 2024.
- MASSRUHÁ, S. M.; LEITE, A. M.; OLIVEIRA, M. S. MEIRA, A.C.; JUNIOR, L. A; BOLFE, L.E.**Agricultura digital: pesquisa, desenvolvimento e inovação nas cadeias produtivas**. Campinas: Embrapa Informática Agropecuária, cap. 1, 2020.

- MATSUO, É.; SEDIMAYA, T.; CRUZ, C.D.; OLIVEIRA, R.C.T. Hibridação. In: SEDIYAMA, T. (Ed.) **Melhoramento genético da soja**. Londrina: Mecenaz, p. 57-72, 2015.
- MELLO, I. S.; DUARTE, D. G; NETO, G.G. Sinopse de Fabaceae Caesalpinioideae para a flora de Mato Grosso, Brasil. **Biodiversidade**, v. 2, n. 14, p. 43-49, 2015.
- MENDONÇA, M.; BOVOLENTA, M.G.; ROSA, B.O.; SOUZA, L.B.; GONGORA, V.L.; GARCIA, E.I. **Inteligência artificial: fundamentos, conceitos, aplicações e tendências**. Ciência, tecnologia e inovação: experiências, desafios e perspectivas 3. Cornélio Procópio: UTFPR, cap. 4, 2023.
- MONTOYA, M.A.; BERTUSSI, L.A; LOPES, R.L.; FINAMORE, E.B. Uma nota sobre consumo energético. emissões. Renda e emprego na cadeia de soja no Brasil. **Revista Brasileira de Economia**, v.73, p. 345-369, 2019.
- MUNDSTOCK, C. M.; THOMAS, A. L. **Soja: fatores que afetam o crescimento e o rendimento de grãos**. Porto Alegre: Universidade Federal do Rio Grande do Sul, 2005.
- MUNZ, J.; MAURMANN, I.; SHUELE, H.; DOLUSCHITZ, R. Digital transformation at what cost? **Smart Agricultural Technology**, v. 9, p. 100585, 2024.
- FARIAS, J. R. B.; NEPOMUCENO, A. L. ; NEUMAIER, N. **Ecofisiologia da soja**. Londrina: EMBRAPA. Circular Técnica 48, p. 9, 2007.
- OECD. **The digital transformation of agriculture and food systems**. Paris: Organisation for Economic Co-operation and Development, 2022.
- OLIVEIRA, M. F. *et al.* **Melhoramento da soja**. Centro de diversidade e recursos genéticos Viçosa, MG: Editora UFV. cap. 3. p. 44-61, 2017.
- PEDROZO, C. Â.; BARBOSA, H.P.; SILVA, F.L.; RESENDE, M.D.V.; PETERNELLI, L.A. Repeatability of full-sib sugarcane families across harvests and the efficiency of early selection. **Euphytica**, v. 182, n. 3, p. 423-430, 2011.
- PIRAS, S.; WESZ, V.J.; GHINOI, S. Soy Expansion, Environment, and Human Development: An Analysis across Brazilian Municipalities. **Sustainability**, v. 13, p. 7246, 2021.

QUEIROZ, M. M; WAMBA, F. S. A Framework Based on Blockchain, Artificial Intelligence, and Big Data Analytics to Leverage Supply Chain Resilience considering the COVID-19, *ScienceDirect*, v.55, p. 2396-2401, 2022.

RASCHIKA, S.; MIRJALILI, V. Python machine learning: Machine learning and deep learning with python. **Scikit-Learn, and Tensor Flow**. 2nd Edition, p.17, 2017.

RAUSCH, L.L.; GIBBS, K.H.; SCHELLY, I.; BRANDÃO, J.A; DOUGLAS, C. M.; FILHO, C.A. Soy expansion in Brazil's Cerrado. **Conservation Letters**, v.12, p.12671, 2019.

RESENDE, M. D. V. **Genética Biométrica e Estatística no Melhoramento de plantas perenes**. Brasília: Embrapa Informação Tecnológica, p. 975. ed. 1, 2002.

RODRIGUEZ, C. A compact supermassive binary black hole system. **The Astrophysical Journal**, v. 646, n. 1, p. 49, 2006.

RODRIGUES, V. V.; TERRA, A. Conflitos socioespaciais entre os agricultores tradicionais e sojicultores no município de Balsas Maranhão. **Brazilian Journal of Development**, v. 4, n. 4, p. 1196-1216, 2018.

RODRIGUES, O.; DIDONET, D. A; LHAMBY, B. J. C; BERTAGNOLLI, F.P.; LUZ, S. J. Resposta quantitativa do florescimento da soja à temperatura e ao fotoperíodo. **Pesquisa Agropecuária Brasileira**, v. 36, n. 3, p. 431-437, 2001.

ROSENBLATT, F. The Perceptron: A Probabilistic Model for Information Storage and Organization in the Brain.**Psychological Review**, v. 65, p. 386–408, 1958.

SANTANA, D. C.; TEODORO, R. L.; BAILO, R. F.; DOS SANTOS, G. R.; CORADI, C.P.; BIDUSKI, B. Classification of soybean genotypes for industrial traits using UAV multispectral imagery and machine learning. **Remote Sensing Applications: Society and Environment**, v. 29, p. 100919, 2023.

SAUER, S. Soy expansion into the agricultural frontiers of the Brazilian Amazon: The agribusiness economy and its social and environmental conflicts. **Land Use Policy**, [s.l.],v.79, p. 326–338, 2018.

SEBRAE. A inteligência artificial (IA) já chegou ao agronegócio. Brasília, nov. 2021.

SEDIYAMA, T. FELIPE, S.; BOREM, A. **Soja do Plantio à Colheita**. 1 ed. Viçosa: Editora UFV, p. 312, 2015.

SEIXAS, C. D. S.; NEUMAIER, N.; BALBINOT JUNIOR, A. A.; KRZYZANOWSKI, F. C.; LEITE, R. M. V. B. de C. **Tecnologias de produção de soja. Sistemas de Produção / Embrapa Soja**, 2020.

TIN, K. **Decisão Aleatórias (PDF)**. **Anais da 3ª Conferência Internacional sobre Análise e Reconhecimento de Documentos, Montreal, QC**, p. 278–282, 2016.

VERNETTI, F. J.; JUNIOR, F. J. V. **Genética da soja: caracteres quantitativos e diversidade genética**. Brasília: Embrapa Informação Tecnológica, p. 221, 2009.

WANG, X.; LI, Z.; ZHENG, D.; WANG, W. Nondestructive identification of millet varieties using hyperspectral imaging technology. **J.Appl. Spectrosc**, v. 87, p. 54–61, 2020.

WOYANN, L. G.; MILIOLI, A.S.; BOZI, A.H.; Dalló, S.C.; MATEI, G.; STORCK, L.; BENIN, G. Repeatability of associations between analytical methods of adaptability, stability, and productivity in soybean. **Pesq. agropec. bras., Brasília**, v. 53, n. 1, p. 63-73, 2018.

ZIMMER, G. **Avaliação técnica e econômica do uso de sementes de soja no Rio Grande do Sul**. 57 f. Tese (Doutorado) - Curso de Tecnologia de Produção de Sementes, Universidade Federal de Pelotas, Pelotas, 2017.

Capítulo II

Classificação de cultivares de soja com imagens de sementes e folhas usando Aprendizado de Máquina

Resumo

A soja é uma das culturas de maior importância econômica a nível mundial, a contaminação cruzada na identificação de cultivares de soja é um problema a ser resolvido, por isso o objetivo desse estudo foi identificar algoritmos de Machine Learning para a identificação de cultivares de soja, utilizando imagens de sementes e folhas capturadas por câmeras de smartphones, em combinação com técnicas de inteligência artificial. Para isso foram coletadas folhas e sementes de dez cultivares de soja e posteriormente extraídas as informações dos caracteres morfológicos e espectrais e aplicado as estatísticas (ANOVA, ACP, AD, K-means, algoritmos de Machine Learning). Demonstrou-se que as variáveis espectrais possuem maior poder classificatório na identificação de cultivares de soja, e o (RAF) destacou-se como o mais eficaz para a classificação das cultivares de soja, com porcentagens corretas superiores a 93% sendo a maior para todos os algoritmos aplicados.

Palavras-chaves: *Glycine max* (L.); Machine Learning; Classificação.

Abstract

Soybean is one of the most economically important crops worldwide. Cross-contamination in the identification of soybean cultivars is a problem to be solved. Therefore, the objective of this study was to identify Machine Learning algorithms for the identification of soybean cultivars, using images of seeds and leaves captured by smartphone cameras, in combination with artificial intelligence techniques. To this end, leaves and seeds of ten soybean cultivars were collected, and information on morphological and spectral characteristics was subsequently extracted and statistical analyses were applied (ANOVA, PCA, DA, K-means, Machine Learning algorithms). It was demonstrated that spectral variables have greater classification power in the identification of soybean cultivars, and the (RAF) stood out as the most effective for the classification of soybean cultivars, with correct percentages exceeding 93%, the highest for all applied algorithms.

Keywords: Glycine max (L.); Machine Learning; Classification.

1. INTRODUÇÃO

A soja (*Glycine max* (L.) Merrill) destaca-se como uma das culturas agrícolas mais cultivadas no planeta, exercendo papel fundamental na economia e na segurança alimentar. Na atualidade a mistura de sementes de soja causa problemas substanciais para os agricultores (Liu et al. 2016), sendo o processo de identificação de sementes realizado por inspeção visual, que é muito impreciso e oneroso, ou por meio de testes de DNA (Wang et al. 2020, Zhu et al. 2019). Portanto, é necessário desenvolver métodos eficazes e rápidos para a identificação de cultivares de soja, com aplicação na detecção de contaminação de cultivares, que pode ocorrer em vários pontos da cadeia de produção de sementes da leguminosa.

A identificação de diferentes cultivares a partir da análise de imagens de sementes foi relatado com sucesso em *Avena sativa* L. (Aveia) por Fu et al. 2023 e em *Panicum miliaceum* L. (Cereal) por Ekramirad et al. 2024, ambos estudos utilizando informações hiperespectrais na faixa do infravermelho próximo, mas nenhum avaliou a extração de formas geométricas das sementes pelas imagens.

Uma forma eficaz para a coleta de informações de sementes ou folhas de soja por meio de imagens é a utilização de sensores proximais, como câmeras RGB (vermelho, verde e azul), presentes nos smartphones. Esses sensores exibem um enorme potencial para serem utilizadas como sensores proximais de sementes, visto que, com os avanços tecnológicos recentes, passaram a ser capazes de capturar imagens digitais de elevada qualidade o que permite adquirir sem esforço os valores de intensidade vermelho, verde e azul de sementes pertencentes ao modelo de cores RGB (Perry et al. 2021).

Devido as vantagens óbvias do uso de smartphones, como conveniência e simplicidade de uso e rapidez, as câmeras RGB desses equipamentos têm sido consideradas para diversas análises agrícolas (Perry et al. 2021, Tobiszewaki e Vakh, 2023). Entretanto, ainda são escassas a aplicação desses sensores na identificação de genótipos de espécies vegetais.

Dispositivos imageadores geram um grande volume de dados, o que torna técnicas estatísticas paramétricas ineficientes para sua análise. Uma alternativa, seria a aplicação de algoritmos de Machine Learning (ML), que permite o desenvolvimento de algoritmos para serem aplicados em grandes conjuntos de dados e com informações complexas, como dados de imagens (Marconi et al. 2022, Santana et al. 2023).

Dentre os algoritmos de ML utilizados para processar dados obtidos por sensores imageadores destacam-se as redes neurais artificiais (RNA) e a floresta aleatória (RF) (Santana et al. 2023), os quais são capazes de resolver problemas complexos (Goyal, 2014), como a classificação de cultivares. Outra vantagem, é a possibilidade da utilização dos modelos criados em instrumentos e/ou software.

No entanto, a capacidade dessas técnicas avançadas de Machine Learning aplicadas em dados de imagem de sementes ou de folhas de soja, para distinção de cultivares ainda não foi muito bem explorada. Dessa forma, o estudo teve como objetivo identificar algoritmos de Machine Learning para a classificação de cultivares de soja, utilizando imagens de sementes e folhas capturadas por câmeras de smartphones, em combinação com técnicas de inteligência artificial.

MATERIAL E MÉTODOS

1.1 Instalação do ensaio em campo

Para realização do estudo instalou-se um experimento em campo com dez cultivares de soja em janeiro de 2025, na área experimental da Accert Pesquisa Agronômica, em Balsas (MA), localizada a latitude 7°53'23.37" S e longitude 46°16'03.98" W, com altitude média de 283 metros. A região possui clima tropical, com inverno seco e verão chuvoso (Aw), seguindo a classificação climática de Köppen (Maranhão, 2022). O solo da área é classificado como Latossolo amarelo com textura arenosa (Santos et al. 2012).

No ensaio foram conduzidas dez cultivares de soja (T1 -CZ48B18; T2 -DOMINIO; T3- B44; T4- B74I2X; T5- B66; T6- CRUZADA CE; T7- B412X; T8 -CZ48B44; T9 -ITIQUIRA; T10 -82I78),

adotando-se o delineamento experimental de blocos casualizados, com 3 repetições. Os tratos culturais foram realizados conforme a recomendação de (Sedimaya et al. 2015).

1.2 Coleta das informações das folhas

Quando as plantas atingiram o estágio de floração plena (R2), foram coletados três trifólios por parcela para obtenção de imagens. Sendo as coletas realizadas no terceiro nó superior da haste principal com folha verdadeira. Após a coleta, os trifólios foram acondicionados individualmente em recipientes térmicos mantidos a 13 °C e, em seguida, transportados ao laboratório para a captura das imagens.

As imagens RGB foram obtidas no laboratório utilizando um smartphone Samsung A05, equipado com câmera de alta resolução (8165 × 6124 pixels). Para a captura, os folíolos foram separados e posicionados sobre uma superfície plana de cor azul escura, a fim de garantir contraste adequado. A amostra foi disposta em uma estrutura de suporte que mantinha a câmera a uma distância fixa de 20 cm e em um ângulo perpendicular (90°) em relação ao folíolo. Essa estrutura também incluía uma lâmpada com iluminação uniforme minimizando sombras e variações de luz, além de uma régua graduada para calibração da escala das imagens e extração precisa das características morfológicas. As imagens geradas apresentaram três bandas espectrais distintas (vermelho, verde e azul), correspondentes ao modelo RGB.

1.3 Coleta das informações das sementes

No estágio R8 (maturação plena), todas as plantas das parcelas foram colhidas, sendo a debulha foi realizada quando as sementes atingiram umidade de 15%, condição ideal para padronização das amostras. Em seguida, foram selecionadas 20 sementes de cada parcela para a obtenção das imagens RGB.

Para obtenção das imagens, cada amostra de 20 sementes, que representava uma repetição de uma cultivar, foi posicionado sobre uma superfície plana de cor arroxeadada, sendo a estrutura e a

câmera utilizada de obtenção das imagens de sementes idênticos aos utilizados para obtenção das imagens das folhas. As imagens RGB das sementes também foram registradas com resolução de 8165 × 6124 pixels

1.4 Análises das imagens

Após a aquisição, as imagens foram inicialmente pré-processadas para destacar a diferenciação dos pixels das folhas ou sementes dos pixels de fundo, usando o processo separação de bandas de correção gama (Rueden et al. 2017). Após o pré-processamento, as imagens foram segmentadas aplicando o algoritmo threshold desenvolvido por Huang e Wang, (1995). No processo de segmentação das imagens, os pixels das folhas ou sementes receberam o valor um (1) e os pixels de fundo preto receberam o valor zero (0)

A partir da calibração dos valores de pixel para milímetros, utilizando a régua presente nas imagens, e com base nas imagens previamente segmentadas, foram extraídas para cada semente e para cada folíolo seis variáveis morfológicas (área, perímetro, circularidade, solidez, AR e raund) e nove variáveis espectrais (reflectância das faixas do vermelho, verde e azul; valores de matiz, saturação e valor do espaço de cores HSB (Persson, 2005) e valores de luminosidade L^* , coordenadas cromáticas a^* e b^* do espaço de cores Lab (Berns, 2019). Todas as etapas do processamento e extração dos dados foram realizadas com auxílio do software ImageJ® (Rueden et al. 2017).

A área corresponde à superfície total ocupada pela semente ou folíolo na imagem, enquanto o perímetro é o comprimento da linha que a contorna. A relação entre essas duas medidas dá origem à circularidade, um parâmetro adimensional que avalia a quão próxima a forma da semente está de um círculo perfeito, por sua vez a solidez (do inglês, solidity), quantifica a integridade do contorno da forma, ou quanto enrugada é estrutura. (Masseto e Faleiros, 2017).

1.5 Processamento dos dados

ANOVA

Os dados dos caracteres morfológicos e espectrais das variáveis das sementes e folhas foram submetidos à análise de variância (ANOVA), a fim de verificar a existência de diferenças significativas dentro do ensaio. A ANOVA avalia se as diferenças entre os cultivares são significativas. Se $p < 0,05$, conclui-se que existe diferença real entre pelo menos uma cultivar, foi usado o software R versão 4.1.0. Essa abordagem estatística e gráfica permitiu uma análise detalhada das discrepâncias entre as amostras de um mesmo tratamento, melhorando a percepção dos dados de modo a facilitar a interpretação dos resultados.

Análises Multivariadas

Para averiguar a viabilidade da identificação das cultivares, as variáveis morfológicas e espectrais obtidas foram submetidas a um conjunto de análises multivariadas, incluindo Análise de Componentes Principais (ACP), Análise de Discriminante (AD) e a técnica de agrupamento não supervisionado K-means.

Embora fundamentadas em princípios matemáticos distintos, todas essas técnicas têm como objetivo comum promover o agrupamento de indivíduos semelhantes, a partir da análise simultânea de múltiplas variáveis. A eficácia dessas abordagens em discriminar corretamente as cultivares foram posteriormente comparadas ao desempenho de diferentes algoritmos de aprendizado de máquina (Machine Learning – ML). Todas essas análises foram realizadas com auxílio do software R versão 4.1.0.

Machine Learning – ML

Posteriormente os dados também foram submetidos à análise de Machine Learning pelos algoritmos listados na Tabela 1, para os quais 70% dos dados são utilizados para calibração e 30% para validação dos modelos. Os algoritmos foram selecionados por serem os mais utilizados na agricultura para diversas tarefas, como relatado por De Queiroz Otone et al. 2024.

Tabela 1. Modelos de aprendizado de máquina utilizados para a classificação dos cultivares de soja.

| Sigla | Modelo de Machine Learning | Referência |
|-------|----------------------------|--------------------------------|
| NB | Naive Bayes | RASCHKA; MIRJALILI, 2017 |
| IB1 | Instance-Based Learning | DAVID <i>et al.</i> , 1991 |
| ROF | Rotation forest | RODRIGUEZ <i>et al.</i> , 2006 |
| RAF | Random Forest | TIN KAM, 1995; BREIMAN, 2001 |
| MP | Multilayer Perceptron | ROSENBLATT, 1958 |

As configurações padrão do software Weka versão 3.8.6 foram utilizadas para definir os parâmetros de todos os modelos. Para avaliar o desempenho dos modelos foi utilizada como métrica a porcentagem de classificações corretas (CC, %), refletindo a acurácia na identificação das cultivares, dada pela equação 1

$$CC = \frac{TP}{TP + FN + FP} \times 100$$

Em que:

TP: é a classificação de verdadeiro positivo

FB: é a classificação de falso negativo

FP: é a classificação de falso positivo

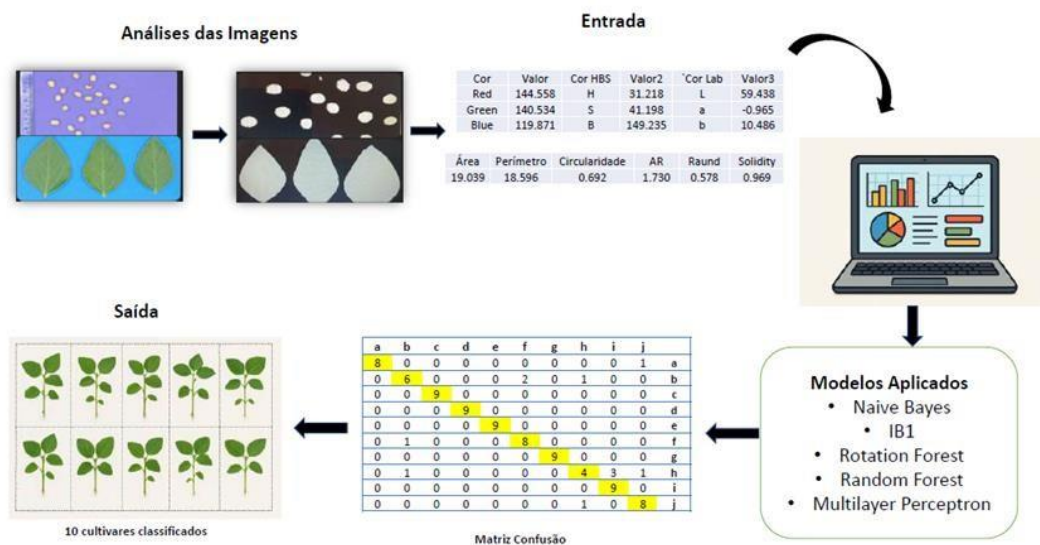


Figura 1. Diagrama das etapas e procedimentos de análise de dados para classificação de cultivares de soja usando processamento de imagens e aprendizado de máquina.

2. RESULTADOS E DISCUSSÃO

A análise de variância revelou diferenças significativas para todas as características avaliadas nas dez cultivares de soja (Tabela 1). Quanto aos coeficientes de variação (CV%), todos apresentaram valores dentro dos limites aceitáveis estabelecidos pela literatura (Bispo et al. 2025, Duc et al. 2023, Kassout et al. 2022, Martin et al. 2020, Oliveira et al. 2025), o que demonstra boa precisão experimental e, por consequência, reforça a confiabilidade dos dados obtidos.

Esses resultados demonstram a robustez do uso de imagens digitais na caracterização fenotípica de genótipos de soja e também evidenciam o potencial de identificação de variáveis capazes de diferenciar com precisão, cultivares de soja por meio de imagens digitais.

Tabela 2. Resumo da análise de variância das características morfológicas e espectrais, obtidas de imagens RGB de sementes e folhas de dez cultivares de soja.

| | | ----- Variáveis das Sementes ----- | | | | | | | | | |
|--------|--|------------------------------------|-------|---------|------|---------|----|---------|----|---------|----|
| | | Area | Perim | Circ | Ar | Raund | | | | | |
| QM | | 0,0026 | ** | 0,0416 | ** | 0,0019 | ** | 0,0387 | ** | 0,0094 | ** |
| CV (%) | | 4,58 | | 2,28 | | 1,47 | | 3,4 | | 3,25 | |
| | | Solidity | Blue | B | b | Green | | | | | |
| QM | | 0,0001 | ** | 210,042 | ** | 145,742 | ** | 94,2274 | ** | 57,2021 | ** |
| CV (%) | | 0,31 | | 1,8 | | 0,41 | | 4,99 | | 0,77 | |
| | | S | a | Red | H | L | | | | | |
| QM | | 391,82 | ** | 5,3998 | ** | 189,327 | ** | 2,8383 | ** | 5,8491 | ** |
| CV (%) | | 4,17 | | 20,66 | | 0,46 | | 2,29 | | 0,81 | |
| | | ----- Variáveis das Folhas ----- | | | | | | | | | |
| | | Area | Perim | Circ | Ar | Raund | | | | | |
| QM | | 187,385 | ** | 37,2372 | ** | 0,0064 | ** | 0,0958 | ** | 0,0099 | ** |
| CV (%) | | 21,01 | | 10,79 | | 5,56 | | 7,79 | | 7,87 | |
| | | Solidity | Red | Green | Blue | H | | | | | |
| QM | | 0,0002 | * | 1099,1 | * | 364,859 | ** | 317,491 | ** | 2009,69 | ** |
| CV (%) | | 1,02 | | 21,83 | | 4,54 | | 8,87 | | 16,61 | |
| | | L | S | a | B | b | | | | | |
| QM | | 10,2478 | ** | 5130,71 | ** | 780,062 | ** | 224,574 | ** | 130,482 | ** |
| CV (%) | | 3,2 | | 29,58 | | 35,82 | | 4,84 | | 23,41 | |

QM: Quadrado Médio; CV: Coeficiente de Variação; Área: Área; Perim: Perímetro; Circ: Circularidade; AR: Aspect Ratio; Raund: Redondeza; Solidity: Solidez; Red: Vermelho; Green: Verde; Blue: Azul B: Banda do azul; b: Coordenada cromática do amarelo Lab; S: Saturação; a: Cordenada cromática R/G Lab; H: Valor de Matriz HSB; L: Luminosidade no modelo Lab; ** e * significativos a 1 e 5% de probabilidade; respectivamente; pelo teste F, ns: não-significativo; pelo teste F.

Afim de avaliar a possibilidade de intensificação das cultivares de soja estudadas com as variáveis obtidas de imagens RGB das folhas e sementes, inicialmente procedeu-se aplicação de duas técnicas de classificação não supervisionados, K-means e Análise Discriminante.

Conforme descrito na seção de materiais e métodos, cada cultivar foi representada por informações provenientes de dez indivíduos. Nesse contexto, a eficácia das técnicas de agrupamento está diretamente relacionada à sua capacidade de identificar, com precisão, todos os indivíduos pertencentes a cada variedade.

Com base nas variáveis obtidas, a análise K-means não foi eficaz na identificação de nenhuma cultivar (Figura 2). Visto que os grupos formados apresentaram sobreposição de cultivares, sendo que nenhum grupo agrupou apenas um único cultivar.

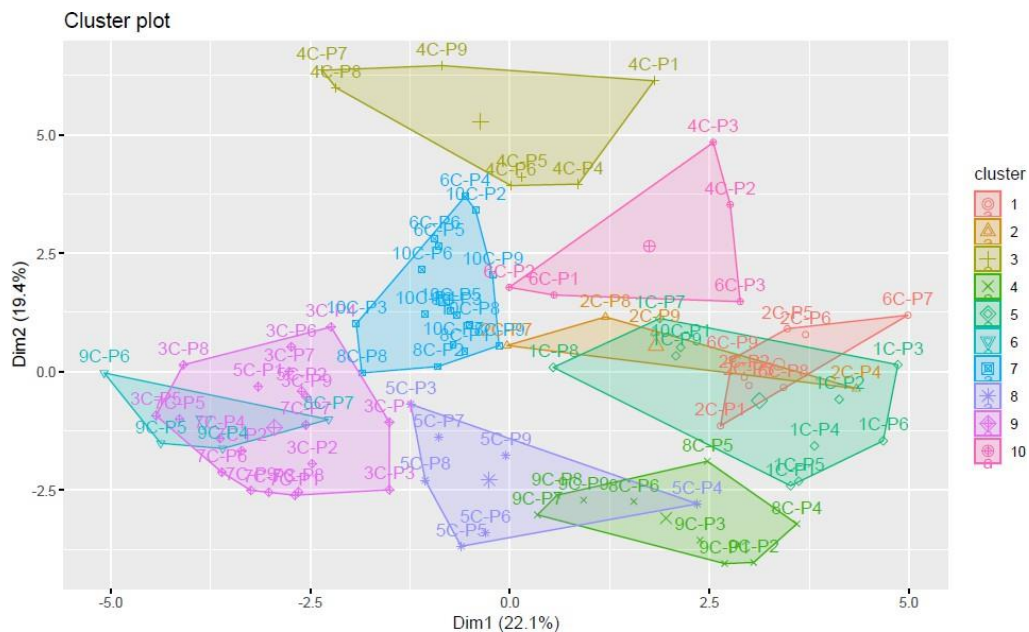


Figura 2. Resumo dos 10 clusters obtidos pela aplicação do agrupamento k-means às características morfológicas e espectrais, obtidas de imagens RGB de sementes e folhas de dez cultivares de soja.

A aplicação do algoritmo K-means já demonstrou eficácia na diferenciação de genótipos em diversas espécies agrícolas como em arroz a partir de variáveis de cor extraídas de imagens digitais (Jumi et al. 2021) em milho (Conceição et al. 2021) e batata (Islam et al. 2025). Esses resultados indicam limitações da técnica na identificação de cultivares pertencentes a espécies com base genética

estreita, como a soja (Priolli et al. 2002), em que as características fenotípicas entre diferentes cultivares apresentam elevada similaridade.

Por sua vez, a Análise Discriminante resultou em 44% de classificações corretas das plantas nas diferentes cultivares de soja, conforme ilustrado na matriz de confusão (Figura.3). Esse resultado indica que, embora a classificação tenha apresentado desempenho satisfatório para algumas cultivares, como C1 (70%), C4 (60%) e C5 (60%), a alocação incorreta de indivíduos, especialmente nas cultivares C6 e C8, comprometeu a eficiência global da metodologia, ficando essa distante de um valor ideal de 100% de eficácia.

| | C1 | C2 | C3 | C4 | C5 | C6 | C7 | C8 | C9 | C10 | | % de Acertos |
|-----|------|----|----|----|----|----|----|----|----|-----|-----|--------------|
| C1 | 7 | 0 | 2 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | C1 | 70 |
| C2 | 0 | 5 | 0 | 0 | 0 | 2 | 0 | 0 | 0 | 2 | C2 | 50 |
| C3 | 0 | 0 | 5 | 0 | 0 | 0 | 0 | 0 | 0 | 4 | C3 | 50 |
| C4 | 0 | 0 | 0 | 6 | 0 | 2 | 0 | 0 | 0 | 1 | C4 | 60 |
| C5 | 0 | 0 | 0 | 0 | 6 | 0 | 0 | 0 | 0 | 3 | C5 | 60 |
| C6 | 0 | 0 | 0 | 0 | 0 | 2 | 0 | 0 | 0 | 7 | C6 | 20 |
| C7 | 0 | 0 | 1 | 0 | 0 | 4 | 4 | 0 | 0 | 0 | C7 | 40 |
| C8 | 0 | 0 | 0 | 0 | 0 | 2 | 0 | 2 | 0 | 5 | C8 | 20 |
| C9 | 0 | 0 | 0 | 0 | 0 | 2 | 0 | 2 | 4 | 1 | C9 | 40 |
| C10 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 5 | 0 | 3 | C10 | 30 |
| | % CC | | | | | | | | | | | 44 |

Figura 3. Matriz de confusão das características morfológicas e espectrais, obtidas de imagens RGB de sementes e folhas de dez cultivares de soja.

Apesar dos avanços recentes no uso de técnicas de análise multivariada aplicadas à identificação de cultivares agrícolas, observa-se uma escassez de estudos que explorem especificamente a análise discriminante com base em variáveis espectrais e morfológicas obtidas de imagens RGB para a diferenciação de cultivares de soja. A maior parte das pesquisas disponíveis concentra-se no emprego da análise discriminante em informações de sensores hiperespectrais, como os realizados por Crusiol et al. 2021 e Jeong et al. 2024, cujas características diferem substancialmente das imagens RGB.

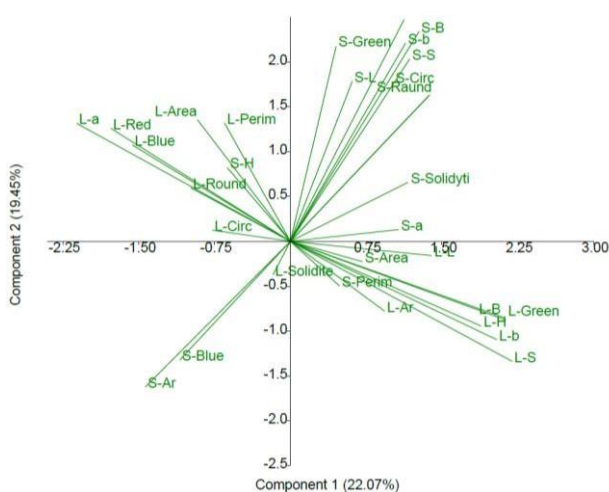
Entretanto, em outras culturas, utilizando informações de imagens RGB a eficácia da técnica de classificação se mostrou superior. Em milho (*Zea mays*), por exemplo, Chen et al. 2010 obtiveram

100% de precisão na identificação de cultivares. Enquanto Banerje et al. 2020, relataram 96,3% de acerto na identificação de diferentes cultivos, conseguindo identificar diferentes espécies vegetais como trigo (*Triticum spp.*), mostarda (*Brassica spp.*), repolho (*Brassica oleracea*), feijão-guandu (*Cajanus cajan*), berinjela (*Solanum applanarum*) e grão-de-bico (*Cicer arietinum*).

Esses resultados sugerem que, em soja, a maior similaridade entre cultivares pode dificultar a classificação correta das variedades, exigindo a combinação de diferentes fontes de dados ou técnicas de classificação mais complexas para melhorar a identificação dos materiais. Além disso dados coletados de imagens multiespectrais frequentemente apresentam estrutura não linear (Ghamisi et al. 2017, Plaza et al. 2009), dificultando a aplicação de métodos estatísticos tradicionais como K-means e Análise Discriminante.

Com intuito de analisar a estrutura das variáveis morfológicas e espectrais que foram obtidas pelas imagens RGB, as mesmas foram submetidas à Análise de Componentes Principais (PCA), [Fig. 4 (A)]. É possível averiguar que os dois primeiros componentes principais (CP1 e CP2) explicaram conjuntamente apenas 41,52% da variação total dos dados [Fig.4 (A)], sendo necessário unir o efeito das sete primeiras componentes para acumular um mínimo de 81,77% da variância total [Fig.4 (B)].

(A)



(B)

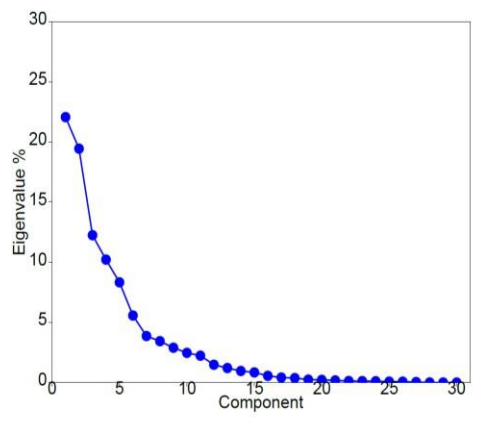


Figura 4. Resumo da Análise das duas Componentes Principais (PCA) das características morfológicas e espectrais, obtidas de imagens RGB de sementes e folhas de dez cultivares de soja (A). Porcento de autovalores de todas as componentes principais da análises das características morfológicas e espectrais, obtidas de imagens RGB de sementes e folhas de dez cultivares de soja (B).

Para aplicação e interpretação da ACP, se usa como critério, que os componentes principais escolhidos expliquem pelo menos 80% da variância acumulada dos dados (Peres-Neto et al. 2005).

Com os primeiros componentes não capturaram uma quantidade significativamente maior de variância, igual ou superior a 80%, isso sugere que não existe uma direção linear predominante para resumir a estrutura dos dados. Além disso, como apresentado na Figura 4(B), que mostra a variância explicada por cada componente principal, não apresentou uma queda acentuada nos primeiros componentes, significa que a ACP não encontrou uma estrutura linear clara para resumir os dados em poucas dimensões (Jolliffe e Cadima, 2021). Comprovando a existência de estrutura não linear dos dados obtidos pela análise das imagens RGB de sementes e folhas de soja.

Assim, para superar tais limitações, torna-se necessário aplicação de métodos mais robustos, como algoritmos de Machine Learning, capazes de explorar relações não lineares e capturar padrões complexos nos dados, ampliando o potencial de identificação das cultivares de soja a partir de variáveis de imagens RGB (Hastie, Tibshirani e Friedman, 2009).

Afim de avaliar a capacidade de identificação dos cultivares de soja a partir de informações espectrais e morfológicas extraídas de imagens digitais de folhas e sementes, utilizando diferentes algoritmos de aprendizado de máquina, os modelos de classificação foram inicialmente testados exclusivamente com as variáveis provenientes das folhas.

A identificação correta de variedades de soja pela análise de informações foliares, em campos de produção de sementes permitiria que o controle da pureza genética fosse feito ainda na fase vegetativa. Essa abordagem traria vantagens agronômicas e econômicas significativas, como a possibilidade de intervenções imediatas para a eliminação de plantas fora do padrão varietal, evitando o descarte de áreas inteiras ou a multiplicação de sementes contaminadas.

Quando utilizados apenas os caracteres morfológicos das folhas, as taxas de classificação correta variaram entre 22,2% (Random Forest) e 31,1% (Naive Byes), indicando baixo poder discriminatório desse conjunto de variáveis de forma isolada (Figura 5). Com a utilização dos caracteres espectrais das folhas, os percentuais de acerto apresentaram melhora, chegando a 36,6% com Multilayer Perceptron e Instance Based (Figura 5).

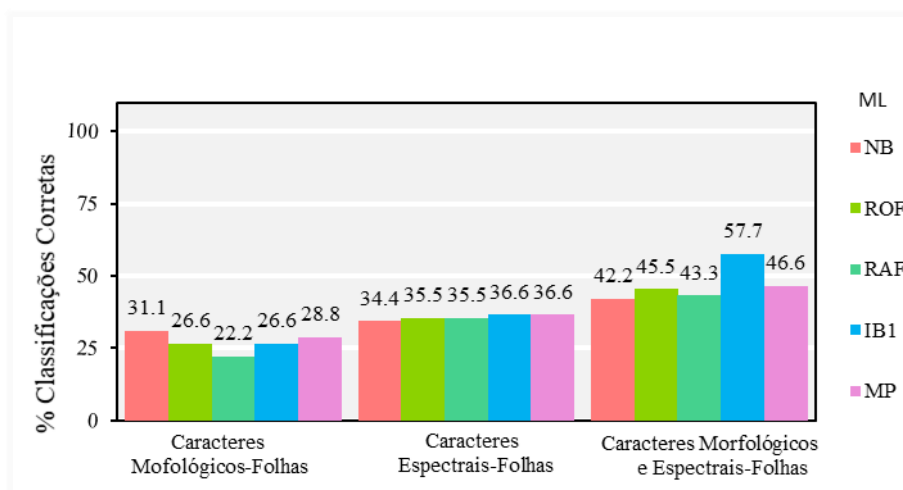


Figura 5. Porcentagem de classificações corretas de 10 cultivares de soja a partir de 5 algoritmos de aprendizado de máquina aplicado para caracteres morfológicos das folhas, caracteres espectrais das folhas e caracteres morfológicos e espectrais das folhas juntos. NB: Naive Bayes; ROF: Rotation Forest; RAF: Random Forest; IB1: Instance Based; MP: Multilayer Perceptron.

Esse resultado demonstra que as informações espectrais são mais eficazes para a diferenciação entre cultivares quando comparadas às variáveis morfológicas das folhas. Enquanto variáveis morfológicas apresentam elevada variação dentro de uma mesma cultivar, como demonstrados pelos valores de CV% na Tabela 1, reduzindo sua capacidade discriminatória, as variáveis espectrais, com exceção da variável “a”, fornecem maior sensibilidade a variações sutis de pigmentação foliar e padrões de refletância, aspectos cruciais para diferenciação varietal.

Em relação às variáveis extraídas das imagens das sementes, conforme ilustrado na Figura 6 e na Tabela 2, observa-se que essas apresentam maior estabilidade fenotípica e, conseqüentemente, maior capacidade discriminatória quando comparadas às variáveis obtidas a partir das imagens das folhas. Essa diferença é evidenciada pela menor magnitude dos valores de coeficiente de variação (CV%) observados para uma mesma variável quando esta é obtida nas sementes (Tabela 2).

A escolha do órgão vegetal para a fenotipagem digital é crítica, pois diferentes órgãos exibem níveis distintos de plasticidade fenotípica em resposta a fatores ambientais (Buchaillet et al. 2019, Rao et al. 2019). As folhas, como principal interface da planta com o ambiente, são altamente plásticas, alterando sua morfologia e cor em resposta à disponibilidade de água, luz e nutrientes (He et al. 2025, Wang et al. 2022, Wang et al. 2024). Em contrapartida, as características das sementes são predominantemente controladas por fatores genéticos (Silva et al. 2016, Sulistyo et al. 2021, Yirga et al. 2022). Por esta razão, as variáveis das sementes de soja podem ser consideradas mais eficazes para a discriminação de cultivares, pois demonstram alta estabilidade mesmo quando as plantas-mãe são submetidas a condições de cultivo.

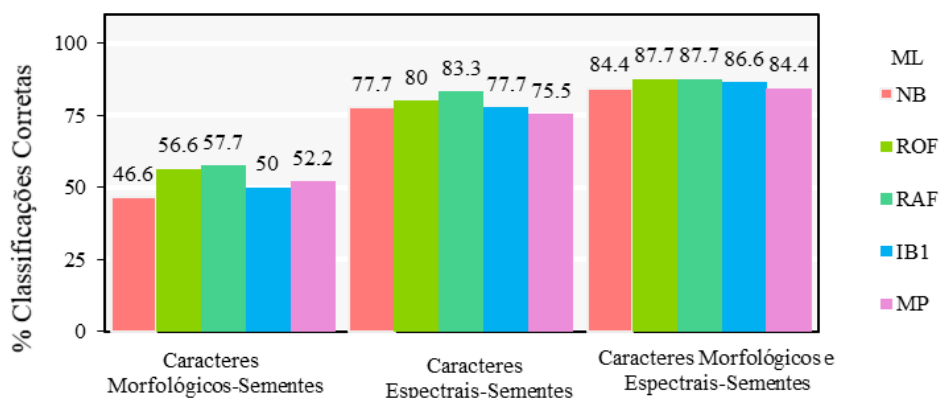


Figura 6. Porcentagem de classificações corretas de 10 cultivares de soja a partir de 5 algoritmos de aprendizado de máquina aplicado para caracteres morfológicos das sementes, caracteres espectrais das sementes e caracteres morfológicos e espectrais das sementes simultaneamente. NB: Naive Bayes; ROF: Rotation Forest; RAF: Random Forest; IB1: Instance-Based; MP: Multilayer Perceptron.

Com aplicação apenas dos caracteres morfológicos das sementes, os modelos de classificação alcançaram uma taxa de acerto de no máximo 57,7% utilizando o algoritmo Random Forest. Por sua vez, o uso apenas dos caracteres espectrais das sementes elevou a precisão dos modelos, alcançando 83,3% de classificação correta com o mesmo algoritmo. Esses resultados demonstram que, assim como observado nas folhas, as informações espectrais extraídas das sementes de soja apresentam maior eficiência discriminatória em relação às variáveis morfológicas.

Segundo Yang et al. (2020), variáveis espectrais demonstram desempenho superior às morfológicas na discriminação de cultivares por refletirem diferenças químicas e estruturais internas

das sementes, relacionadas à composição e aos pigmentos, que não são detectadas por características externas. As variáveis morfológicas, por sua vez, embora mostrem claramente que as sementes de diferentes cultivares são distintas (Tabela 2), a grande variação dentro das cultivares reduz a diferença entre elas.

Embora nenhum dos algoritmos de aprendizado de máquina tenha alcançado 100% de acerto na identificação das cultivares, seja utilizando variáveis obtidas das folhas ou das sementes, observou-se que, em ambas as condições, a combinação de atributos morfológicos e espectrais proporcionou o melhor desempenho geral dos classificadores. As maiores porcentagem de classificações corretas foram obtidos com o algoritmo Instance-Based Learning (IB1), atingindo 57,7% com dados das folhas, e 87,7% com os algoritmos Rotation Forest e Random Forest, quando utilizadas as variáveis extraídas das sementes.

Considerando que os resultados das classificações anteriores não alcançaram 100% de precisão na identificação das variedades, os algoritmos foram analisados utilizando simultaneamente as variáveis extraídas das folhas e das sementes de soja (Figura. 7).

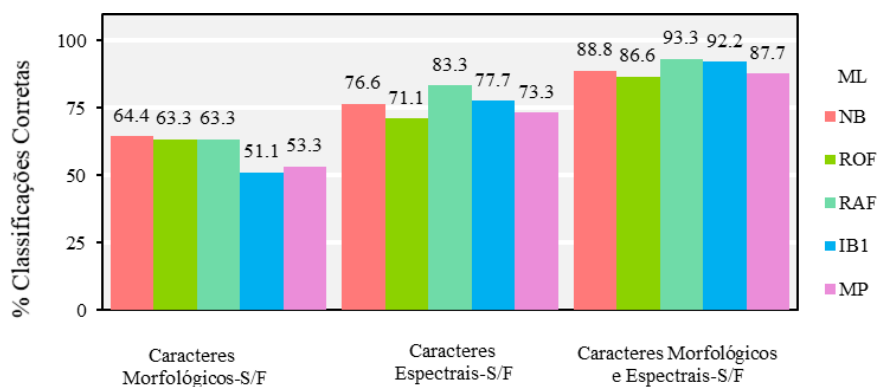


Figura 7. Porcentagem de classificações corretas de 10 cultivares de soja a partir de 5 algoritmos de aprendizado de máquina aplicado para caracteres morfológicos das F/S, caracteres espectrais das F/S e caracteres morfológicos e espectrais das F/S juntos. NB: Naive Bayes; ROF: Rotation Forest; RAF: Random Forest; IB1: Instance-Based; MP: Multilayer Perceptron.

Os resultados demonstram o incremento no desempenho de todas as metodologias de aprendizado de máquina, possivelmente em decorrência do aumento do número de variáveis e consequentemente do aumento de parâmetros do modelo de cada algoritmo (Figura 7). Embora

nenhum tenha alcançado precisão de máxima na identificação das cultivares de soja, o modelo Random Forest (RAF) obteve o melhor desempenho geral (93,3%) (Figura 7), e se destacou em quase todas as condições que foi submetido nesse estudo.

O destaque do Random Forest está alinhado com evidências de estudos anteriores que demonstraram sua superioridade na identificação de cultivares, como relatado por (Bao et al. 2021, Chen et al. 2025 e Kong et al. 2013). Além disso, nos últimos anos, tem sido amplamente utilizado na classificação de imagens (Chowdhury, 2024) e em diversos estudos agrícolas (Amankulova et al. 2023, Anbananthen et al. 2021, Khanal et al. 2018, Venkatesan et al. 2022).

O método Random Forest (RF) aplicado seguiu a suposição e a equação desenvolvida por Breiman, 2001, em que cada árvore no RF é construída usando um subconjunto das variáveis preditoras, e o número de árvores e variáveis preditoras, usados para formar cada árvore, pode variar dependendo do conjunto de dados.

Essa estrutura torna o modelo ideal para encontrar padrões não lineares (Db, 2021), e resolver o problema de colinearidade variável (Jabed e Murad, 2024). Além de proporcionar a capacidade de gerenciar conjuntos de dados de alta dimensão (Jabed e Murad, 2024), como proveniente de imagens digitais.

A obtenção porcentagem de classificação corretas superiores a 93% pelo algoritmo Random Forest, com a combinação de diferentes tipos de variáveis, indica o elevado potencial dessa abordagem para a identificação precisa de cultivares de soja a partir de imagens RGB de folhas e sementes. Cabe destacar que nenhuma das abordagens tradicionais de classificação supervisionada, K-means e Análise Discriminante, alcançou sequer metade dessa capacidade.

Tal desempenho reforça a viabilidade do uso dessa técnica em áreas de produção e certificação de sementes, reduzindo tempo e custos em comparação aos métodos convencionais na identificação de misturas varietais e no monitoramento da pureza genética da soja.

No entanto, estudos adicionais são necessários para validar a aplicação prática dessa técnica na classificação de cultivares de soja. Entre as etapas futuras, tem-se a inclusão de um número maior

de cultivares para o refinamento dos modelos, testes com sensores RGB de diferentes resoluções e a avaliação do poder discriminatório de variáveis morfológicas e espectrais obtidas com o infravermelho. Com esses avanços, no futuro, será possível utilizar câmeras de smartphones como ferramenta acessível e eficiente para o monitoramento da pureza genética em lotes de sementes de soja.

4.CONCLUSÃO

Entre os algoritmos testados, o Random Forest (RAF) destacou-se como o mais eficaz na classificação dos dez cultivares de soja avaliados.

REFERÊNCIAS

Amankulova K, Farmonov N, Muksi L (2023) Sunflower crop yield prediction by advanced statistical modeling using satellite-derived vegetation indices and crop phenology. **Geocarto international** **38**: e2197509.

Anbananthen KSM, Subbiah S, Chelliah D, Sivakumar P, Somasundaram V, Velshankar KH, Khan MKAA (2021) An intelligent decision support system for crop yield prediction using hybrid machine learning algorithms. **F1000Research** **10**: 1143.

Banerje K, Dutta S, Das B, Roy D, Sen S, Bhabani P, Chatterjee A (2025) Crop type discrimination through low cost proximal RGB imaging and multivariate analysis. **Arabian Journal of Geosciences** **18**:31.

Bao F, Bambil D (2021) Applicability of computer vision in seed identification: deep learning, random forest, and support vector machine classification algorithms. **Acta Botanica Brasilica** **35**:17-21.

Berns RS (2019). **Billmeyer and Saltzman's principles of color technology**. John Wiley e Sons.

Bispo R B, Fukuji KKN, Fukuji AYS, Sudré CP, Cassaro S, Amaral Júnior AT, Rodrigues R (2025) Physiological and colorimetric phenotyping in obtaining genetic gains by indirect selection in Capsicum. **Brazilian Journal of Biology** **85**: e 290261.

- Breiman L (2001) **Random forests**. Machine learning. Springer Nature Link, 5-32p.
- Buchaillet ML, Gracia-Romero A, Vergara-Diaz O, Zaman-Allah MA, Tarekegne A, Cairns JE, Prasanna BM, Araus JL, Kefauver SC (2019) Evaluating maize genotype performance under low nitrogen conditions using RGB UAV phenotyping techniques. **Sensors 19**: e1815.
- Chen M, Chang Z, Jin C, Cheng G, Wang S, Ni Y (2025) Classification and Recognition of Soybean Quality Based on Hyperspectral Imaging and Random Forest Methods. **Sensors 25**: 1539.
- Chen X, Xun Y, Li W, Zhang J (2010) Combinando análise discriminante e redes neurais para identificação de variedades de milho. **Computadores e eletrônica na agricultura 71**: 48-53.
- Chowdhury MS (2024) Comparison of accuracy and reliability of random forest, support vector machine, artificial neural network and maximum likelihood method in land use/cover classification of urban setting. **Environmental Challenges14**: e100800.
- Concepcion R, Alejandrino J, Palconit MG, Lamboloto I, Dadios E, Duarte B e Lauguico S (2021) Identificação de variedades de milho filipinas usando rede neural convolucional com caracterização de fenótipos morfológicos do grão. **In 2021 IEEE 9th Region 10 Humanitarian Technology Conference (R10-HTC). IEEE**. <https://doi.org/10.1109/r10-htc53172.2021.9641688>
- Yousefi M, Aziz AS, Azrad S (2021) Classification of oil palm female inflorescences anthesis stages using machine learning approaches. **Information Processing in Agriculture 8**: 537-549.
- Duc NT, Ramlal A, Rajendran A, Raju D, Lal SK, Kumar S (2023) Image-based phenotyping of seed architectural traits and prediction of seed weight using machine learning models in soybean. **Frontiers in Plant Science 14**: e1206357.
- Ekramirad N, Doyle L, Loeb J, Santra D, Adedeji AA. (2024) Hyperspectral imaging and machine learning as a nondestructive method for proso millet seed detection and classification. **Foods 13**: e1330.
- Fu X, Bai M, Xu Y, Wang T, Hui Z, Hu X. (2023) Cultivar's identification of oat (*Avena sativa* L.) seed via multispectral imaging analysis. **Frontiers in Plant Science 14**: e1113535.

Ghamisi P, Yokoya N, Li J, Liao W, Liu S, Plaza J (2017) Advances in Hyperspectral Image and Signal Processing: A Comprehensive Overview of the State of the Art, in **IEEE Geoscience and Remote Sensing Magazine** 5: 37-78.

Goyal S (2014) Artificial Neural Networks in fruits: A comprehensive review. **International Journal of Image, Graphics and Signal Processing** 6: 53.

Crusiol GT, Nanni MR, Furlanetto RH, Sibaldelli RNR, Cezar E, Sun L, Foloni JSS, Mertz-Henning LM, Nepomuceno AL, Neumaier N (2021) Using leaf-based hyperspectral reflectance for genotype classification within a soybean germplasm collection assessed under different levels of water availability. **International Journal of RemoteSensing** 42: 8165–8184.

Hastie, T, Tibhhirani R, Friedman J (2009) **The Elements of Statistical Learning: Data Mining, Inference, and Prediction**. Springer.

He D, Li H, Zhou P, Guo J, Yuan J, Wang J, Liu Y, Zhang Z, Lei Y (2025) Leaf Plasticity Responses of Four Urban Garden Plants to Low-Light Environments Under Viaducts. **Forests**16: e651.

Yang L, Zhang Z, Hu X (2020) Cultivar discrimination of single alfalfa (*Medicago sativa* L.) seed via multispectral imaging combined with multivariate analysis. **Sensors** 20; e 6575.

Huang LK, Wang M J (1995) Image thresholding by minimizing the measures of fuzziness. **Pattern Recognition** 28: 41–51.

Islam S, Rahman T, Islam M, Momim A (2025) Leaf-Based Varietal Categorization of Sweetpotato (*Ipomoea batatas*), a Potentially Healthful Vegetable, Using Image Processing and K-Means Clustering. **Journal of Agriculture, Food, Environment and Animal Sciences** 6: 31-49.

Jabed MA e Murad MA (2024) Crop yield prediction in agriculture: A comprehensive review of machine learning and deep learning approaches, with insights for future research and sustainability. **Heliyon** 24: e40836.

Jolliffe IT e Cadima J (2021) Principal Component Analysis: A Review and Recent Developments. **Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences** 379: e2192.

Jumi J, Zaenuddin A, Mulyono T (2021) Model for identification of rice type using combination of shape and color features. In: **IOP conference series: Materials science and engineering**. IOP Publishing, 012038p.

Kassout J, Hmimsa Y, Fatehi SE, Kadaoui K, Houssni M, Chakkour S (2022) Image analysis of Moroccan carob seeds (*Ceratonia siliqua* L.) revealed substantial intraspecific variations depending on climate and geographic origin. **Ecological Processes** **11**: e34.

Khanal S, Fulton J, Klopfenstein A, Douridas N, Shearer S (2018) Integration of high resolution remotely sensed data and machine learning techniques for spatial prediction of soil properties and corn yield. **Computers and electronics in agriculture** **153**:213-225.

Kong W, Zhang C, Liu F, Nie P, He Y (2013) Rice seed cultivar identification using near-infrared hyperspectral imaging and multivariate data analysis. **Sensors** **13**: 8916-8927.

Liu Y, Xie H, Chen Y, Tan K, Wang L e Xie W (2016) Mutual neighborhood information and its application in hyperspectral band selection for classification. **Chemom. Intell. Lab. Syst** **157**:140–151.

MARANHÃO - GOVERNO DO ESTADO DO MARANHÃO. Gerência de Planejamento e Desenvolvimento Econômico - GEPLAN. Atlas do Maranhão. São Luís: Universidade Estadual do Maranhão, p.39, 2022.

Marconi S, Weinstein BG, Zou S, Bohlman S, Singh A (2022) Hyperspectral classification of tree species at the continental scale in the United States National Network of Ecological Observatories. **Remote Sens. Environ** **282**: e113264.

Martin TN, Fipke GM, Winck JEM, Márchese JA (2020) ImageJ software as an alternative method for estimating leaf area in oats. **Acta Agronómica** **69**:162-169.

Masetto TE, Faleiros RS (2017) Análise de imagens digitais para caracterização de formas e tamanhos de sementes. **Engenharia Agrícola, Jaboticabal** **37**:82-91.

Oliveira CA, Vedana NG, Mendonça WA, Gonçalves JVF, Matos DHS e Furlanetto RH (2025) High-Throughput Identification and Prediction of Early Stress Markers in Soybean Under

Progressive Water Regimes via Hyperspectral Spectroscopy and Machine Learning. **Remote Sensing** **17**: e 3409.

De Queiroz Otone JD, Theodoro GdF, Santana DC, Teodoro LPR, de Oliveira JT e De Oliveira IC (2024) Hyperspectral response of the soybean crop as a function of target spot (*Corynespora cassiicola*) using machine learning to classify severity levels. **Agri Engineering** **6**: 330-343.

Peres-Neto PR, Jackson DA, Somers KM (2005) “How many principal components? Stopping rules for determining the number of non-trivial axes revisited.” **Computational Statistics & Data Analysis** **49**: 974-997.

Pery T, Hitesh V, Prasad D e Asim B (2021) Multi-algorithm comparison to predict soil organic matter and soil moisture content from cell phone images. **Geoderma** **385**: e114863.

Persson M (2005) Estimating surface soil moisture from soil color using image analysis. **Vadose Zone Journal** **4**: 1119-1122.

Plaza A, Benediktsson JA, Boardman JW, Brazile J e Bruzzone L (2009) Recent advances in techniques for hyperspectral image processing. **Remote Sensing of Environment** **113**:110–122.

Priolli1 RH, Mendes-Junior CT, Arantes NE e Contel EPB (2002) Characterization of Brazilian soybean cultivars using microsatellite markers. **Genetics and Molecular Biology** **25**: 185-193.

Rao DE, Chaitanya, KV (2019) Morphological and physiological responses of seven different soybean (*Glycine max* M) cultivars to drought stress. **Journal of Crop Science and Biotechnology** **22**: 355-362.

Rodriguez JJ, Kuncheva LI, Alonso CJ (2006) Rotation Forest: A new classifier ensemble method. **IEEE Transactions on Pattern Analysis and Machine Intelligence** **28**: 1619–1630.

Rueden CT, Schindelin J, Hiner MC, DeZonia BE, Walter AE e Arena ET (2017) ImageJ for the next generation of scientific image data. **BMC Bioinformatics** **18**: e 529.

Santana DC, Teixeira Filho MCM, da Silva MR, Chagas PHMd, de Oliveira JLG e Baio FHR (2023) Classification of soybean genotypes for industrial traits using UAV multispectral imagery and machine learning. **Remote Sensing Applications: Society and Environment** **29**: e100919.

Jeong S W, Lyu J, Jeong HW, Baek J, Moon JK e Lee C (2024) Spectral unmixing of hyperspectral images for phenotyping soybean seed traits. **Plant Cell Reports** **43**: e164.

Silva FC, Sedimaya T, Silva FA, Bezerra GRA, Rosa PD, Ferreira VL, Cruz DC (2016) Identification of new descriptors for differentiation of soybean genotypes by Gower algorithm. **African Journal of Agricultural Research** **11**: p. 961-966

Silva F e Azevedo C (2016) The Assistat Software Version 7.7 and its use in the analysis of experimental data. **Revista Brasileira de Biomassa** **2**: 36-41.

Tobiszewski M e Vakh C (2023) Analytical applications of smartphones for agricultural soil analysis. **Analytical and Bioanalytical Chemistry** **415**:3703-3715.

Venkatesan S, Lim J e Cho Y (2022) A Crop Growth Prediction Model Using Energy Data Based on Machine Learning in Smart Farms. **Computational Intelligence and Neuroscience** **1**: e 2648695.

Wang X, Li Z, Zheng D, Wang W (2020) Nondestructive identification of millet varieties using hyperspectral imaging technology. **Journal of Applied Spectroscopy** **87**: 54-61.

Wang S, Zhou H, He Z, Ma D, Sun W, Xu X, Tian Q (2024) Effects of drought stress on leaf functional traits and biomass characteristics of *Atriplex canescens*. **Plants****13**: e 2006.

Wang MZ, Li HL, Liu KC, Dong CB, Yu HF (2022) Adaptive plasticity in response to light and nutrient availability in the clonal plant *Duchesnea indica*. **Journal of Plant Ecology** **15**:795-807.

Sulistyo A, Purwantoro, Mejaya J M, Nugrahaeni N (2021) Determination of genetic parameters of seed characteristics in edible soybean. **Legume Research-An International Journal** **44**: 522-526.

Yirga M, Sileshi1 Y, Tesfaye A, Hailemariam M (2022) Genetic variability and association of traits in soybean (*Glycine max* (L.) genotypes in Ethiopia. **Ethiopian Journal of Crop Science** **9**: 49-74.

Zhu S, Chao M, Zhang J, Xu X, Song P, Zhang J, Huang Z. (2019) Identification of soybean seed varieties based on hyperspectral imaging technology. **Sensors** **19**: 5225.

Capítulo III

**Otimização do número mínimo de folhas e sementes para
classificar cultivares de soja usando imagens**

Resumo

A identificação de cultivares de soja através de imagens, como de sementes e folhas, exige a prévia determinação do número ideal de órgãos a ser avaliados. Para isso o estudo visou otimizar o número mínimo de folhas e sementes de soja que devem ser fotografados para prever o valor real de essas cultivares para diferentes variáveis morfológicas e espectrais obtidas via análise de imagens. Foram coletadas folhas e sementes de dez cultivares de soja, extraídas as informações morfológicas e espectrais a partir das imagens e aplicada a análises de repetibilidade (ANOVA, CP-COV, CP-COR, A.Estructural). Os resultados confirmam que devem ser fotografados no mínimo 124 folhas para as variáveis morfológicas das folhas, e 102 folhas para as variáveis espectrais, por outro lado devem ser fotografados no mínimo 344 sementes para variáveis morfológicas das sementes e 183 sementes para as variáveis espectrais.

Palavras Chaves: *Glycine max* ; Repetibilidade; Morfológicas; Espectrais.

Abstract

Identifying soybean cultivars through images, such as seeds and leaves, requires the prior determination of the ideal number of organs to be evaluated. Therefore, this study aimed to optimize the minimum number of soybean leaves and seeds that should be photographed to predict the true value of these cultivars for different morphological and spectral variables obtained via image analysis. Leaves and seeds from ten soybean cultivars were collected, morphological and spectral information was extracted from the images, and repeatability analyses were applied (ANOVA, CP-COV, CP-COR, Structural Analysis). The results confirm that a minimum of 124 leaves should be photographed for leaf morphological variables and 102 leaves for spectral variables; conversely, a minimum of 344 seeds should be photographed for seed morphological variables and 183 seeds for spectral variables.

Keywords: Glycine max; Repeatability; Morphological; Spectral.

1. INTRODUÇÃO

Uma das espécies mais cultivadas no mundo, com grande importância socioeconômica é a soja [*Glycine max* (L.) Merrill], visto que gera empregos, receita e é uma das mais importantes fontes de proteína e óleo vegetal do mundo (Hartman et al. 2015; Sedimaya et al. 2015).

Na safra 2024/25, a soja se destacou como uma das culturas agrícolas mais produzidas no mundo, com um total de 421 milhões de toneladas (Embrapa, 2025). No entanto, o desempenho produtivo das lavouras de soja pode ser limitado devido a diversos fatores, sendo um deles a baixa qualidade sanitária, fisiológica e/ou genética das sementes utilizadas para semeadura. A mistura de sementes de soja de cultivares distintas causa problemas substanciais para os agricultores o que eleva a complexidades do mercado de sementes (Liu et al. 2016).

O crescente interesse por estratégias rápidas e não destrutivas para identificação de cultivares tem promovido métodos baseados em análise de imagens. Dentre os métodos disponíveis, a análise de imagens multiespectrais tem se mostrado uma alternativa promissora para avaliar a pureza varietal (Boelt et al. 2018). Estudos em diversas culturas já relatam a eficácia da aplicação de imagem na identificação de cultivares como em pimentas (Kurtulmus et al. 2016) e arroz (Huang e Chien 2017, Chaugule e Mali 2016).

A identificação de cultivares de soja através de imagens, como de sementes e folhas, exige a prévia determinação do número ideal de órgãos de uma mesma cultivar a serem coletados os dados. Essa etapa define o número mínimo de essas estruturas da planta que devem ser fotografados para assegurar que o resultado da avaliação reflita o valor real da cultivar com confiabilidade, ao mesmo tempo em que se otimiza o tempo de análise.

Para esse fim, existem diversas metodologias estatísticas, como, por exemplo, a estimativa do coeficiente de repetibilidade. Trata-se de um parâmetro amplamente utilizado para quantificar se as medições repetidas de uma característica são as mesmas, expressando a variação total em termos da contribuição do genótipo e do ambiente (Feirreira et al. 2010). Assim, valores elevados da estimativa

de repetibilidade da característica indicam que é possível prever o valor real do indivíduo com um pequeno número de medições (Cruz et al. 2012).

O coeficiente de repetibilidade pode ser estimado por meio de diferentes métodos, como, por exemplo, o método de análise de variância, o método de componentes principais e o método de análise estrutural (Guilhen et al. 2019, Rodrigues et al. 2020). Esses métodos têm sido aplicados nas mais diversas culturas (Diel et al. 2020, Ferreira et al. 2020, Pedrozo et al. 2011). Na soja, esses métodos foram utilizados na avaliação de descritores morfológicos (Matsuo et al. 2012) e na avaliação do teor de óleo e proteína (Jiang 2020), associações entre métodos de análise de adaptabilidade, estabilidade e produtividade (Woyann et al. 2018), e outros.

A utilização de medidas de repetibilidade favorece a melhor definição da quantidade de plantas a serem medidas, para se obter o valor real do indivíduo, com confiabilidade e otimização do tempo de avaliação. Como até o momento, não foram identificados estudos que visassem a determinação do número mínimo necessário de estruturas de plantas que devem ser fotografadas para estudos com imagens em soja.

Este estudo visa determinar o número mínimo de folhas e sementes de soja que devem ser fotografados para prever o valor real de cultivares de soja para diferentes variáveis morfológicas e espectrais obtidas via análise de imagens.

2. MATERIAL E MÉTODOS

2.1 Instalação do ensaio em campo

Para realização do estudo inicialmente instalou-se um experimento em campo com dez cultivares de soja na área experimental da Accert Pesquisa Agronômica, na cidade de Balsas (MA), localizada a latitude 7°53'23.37" S e longitude 46°16'03.98" W, com altitude média de 283 metros. A região possui clima tropical, com inverno seco e verão chuvoso (Aw), seguindo a classificação climática de Köppen (Maranhão 2022).

No ensaio foram conduzidas dez cultivares de soja (T1 -CZ48B18; T2 -DOMINIO; T3- B44; T4- B74I2X; T5- B66; T6- CRUZADA CE; T7- B412X; T8 -CZ48B44; T9 -ITIQUIRA; T10 -82I78),

adotando-se o delineamento experimental de blocos casualizados, com 3 repetições. Os tratos culturais foram realizados conforme a recomendação de (Sedimaya et al. 2015).

2.2 Coleta das informações das folhas

Quando as plantas atingiram o estágio de floração plena (R2), foram coletados três trifólios por parcela para obtenção de imagens. Sendo as coletas realizadas no terceiro nó superior da haste principal com folha verdadeira. Após a coleta, os trifólios foram acondicionados individualmente em recipientes térmicos mantidos a 13 °C e, em seguida, transportados ao laboratório para a captura das imagens.

As imagens RGB foram obtidas utilizando um smartphone Samsung A05, equipado com câmera de alta resolução (8165 × 6124 pixels). Para a captura, os folíolos foram separados e posicionados sobre uma superfície plana de cor azul escura, a fim de garantir contraste adequado. A amostra foi disposta em uma estrutura de suporte que mantinha a câmera a uma distância fixa de 20 cm e em um ângulo perpendicular (90°) em relação ao folíolo. Essa estrutura também incluía iluminação uniforme, minimizando sombras e variações de luz, além de uma régua graduada para calibração da escala das imagens e extração precisa das características morfológicas. As imagens geradas apresentaram três bandas espectrais distintas (vermelho, verde e azul), correspondentes ao modelo RGB.

2.3 Coleta das informações das sementes

No estágio R8 (maturação plena), todas as plantas das parcelas foram colhidas, sendo a debulha foi realizada quando as sementes atingiram umidade de 15%, condição ideal para padronização das amostras. Em seguida, foram selecionadas 20 sementes de cada parcela para a obtenção das imagens RGB.

Para obtenção das imagens, cada amostra de 20 sementes, que representava uma repetição de uma cultivar, foi posicionado sobre uma superfície plana de cor arroxeadada, sendo a estrutura e a câmera utilizada de obtenção das imagens de sementes idênticos aos utilizados para obtenção das

imagens das folhas. As imagens RGB das sementes também foram registradas com resolução de 8165 × 6124 pixels.

2.4 Análises das Imagens

Após a aquisição, as imagens foram inicialmente pré-processadas para destacar a diferenciação dos pixels das folhas ou sementes dos pixels de fundo, usando o processo separação de bandas de correção gama (Rueden et al. 2017). Após o pré-processamento, as imagens foram segmentadas aplicando o algoritmo threshold desenvolvido por HUANG e WANG, (1995). No processo de segmentação das imagens, os pixels das folhas ou sementes receberam o valor um (1) e os pixels de fundo preto receberam o valor zero (0).

A partir da calibração dos valores de pixel para milímetros, utilizando a régua presente nas imagens, e com base nas imagens previamente segmentadas, foram extraídas para cada semente e para cada folíolo seis variáveis morfológicas (área, perímetro, circularidade, solidez, AR e raund) e nove variáveis espectrais (reflectância das faixas do vermelho, verde e azul; valores de matiz, saturação e valor do espaço de cores HSB (Persson 2005) e valores de luminosidade L^* , coordenadas cromáticas a^* e b^* do espaço de cores Lab (Berns 2019). Todas as etapas do processamento e extração dos dados foram realizadas com auxílio do software ImageJ® (Rueden et al. 2017).

A área corresponde à superfície total ocupada pela semente ou folíolo na imagem, enquanto o perímetro é o comprimento da linha que a contorna. A relação entre essas duas medidas dá origem à circularidade, um parâmetro adimensional que avalia a quão próxima a forma da semente está de um círculo perfeito, por sua vez a solidez (do inglês, solidity), quantifica a integridade do contorno da forma, ou quanto enrugada é estrutura (Masetto e Faleiros 2017).

2.5 Processamento dos dados

ANOVA

Os dados dos caracteres morfológicos e espectrais das variáveis das sementes e folhas foram submetidos à análise de variância (ANOVA), a fim de verificar a existência de diferenças significativas das variáveis dentro do ensaio.

Análises de repetibilidade

Assim, o coeficiente de repetibilidade foi estimado por diferentes procedimentos estatísticos, segundo Cruz et al. 2012: análise de variância, componentes principais com base nas matrizes de variâncias e covariâncias fenotípicas; componentes principais com base nas matrizes de correlação e análise estrutural, conforme descrito a seguir.

Método de Análise de Variância (ANOVA) com um fator de variação

Ao avaliar p cultivares e \bar{y} medições repetidas, o coeficiente de repetibilidade pode ser estimado por meio da correlação intraclasses obtida a partir da análise de variância. Este modelo permite remover os efeitos do ambiente temporário que podem ser confundidos com a variação dentro dos genótipos, contribuindo para uma subestimação do coeficiente de repetibilidade (Cruz et al. 2012, Rodrigues et al. 2020).

Modelo com um fator de variação na equação 1

$$Y_{ij} = \mu + g_i + \epsilon_{ij} \quad (1)$$

Em que:

Y_{ij} = observação referente à i -ésima cultivar na j -ésima ambiente, na k -ésima repetição

μ = média geral

g_i = efeito aleatório da i -ésima cultivar sob a influência do ambiente permanente ($i = 1, 2, \dots, p$)

ϵ_{ij} = erro experimental estabelecido pelos efeitos temporários do ambiente na j -ésima medição da i -ésima cultivar.

O coeficiente de repetibilidade é dado de acordo com a equação 2:

$$r = \frac{\delta^2g}{\delta^2 + \delta^2g} \quad (2)$$

Em que:

δ^2 = estimador de variância

δ^2g = estimador de variância genotípica

Métodos de componentes principais

Este método foi proposto por Abeywardena 1972 e amplamente utilizado por diversos autores (Guilhen et al. 2019, Marçal et al. 2016). As estimativas dos coeficientes de repetibilidade foram obtidas com base na matriz de correlações entre cultivares em cada par de medições, por meio da qual são determinados os autovalores e os autovetores normalizados. O autovetor cujos elementos apresentam o mesmo sinal e magnitudes próximas é aquele que expressa a tendência das cultivares em manter suas posições relativas nos diversos intervalos de tempo. A proporção do autovalor, associada ao autovetor, e o estimador do coeficiente de repetibilidade são dados conforme mostrado na equação 3:

$$r = \frac{\lambda_k}{\sum_j \lambda_j} \quad (3)$$

Em que:

λ_k = autovalor associado ao autovetor, cujos elementos têm o mesmo sinal e magnitude semelhante

O coeficiente de repetibilidade pode ser estimado alternativamente, através da técnica de componentes principais aplicada à matriz de variâncias e covariâncias fenotípicas. Nesse caso, o estimador do coeficiente de repetibilidade é dado pela equação 4:

$$r = \rho \frac{\lambda_1 - \delta^2y}{\delta^2y(n-1)} \quad (4)$$

Em que:

λ_1 = autovalor de G associado ao autovetor cujos elementos têm o mesmo sinal e magnitude semelhante.

Método de análise estrutural

Este método foi proposto por Mansour et al. 1981. Diversos estudos utilizaram este método para estimar o coeficiente de repetibilidade (Rodrigues et al. 2020) e apresenta algumas diferenças conceituais em relação ao método dos componentes principais. Neste método, R é considerada a matriz paramétrica de correlações entre os genótipos em cada par de avaliação e \hat{y} é o seu estimador. Um estimador do coeficiente de repetibilidade baseado na análise estrutural é dado pela equação 5:

$$r = \frac{\alpha' R \alpha - 1}{n-1} \quad (5)$$

Em que:

α' = é o autovetor com elementos paramétricos associado ao maior vetor R.

Da mesma forma, o coeficiente de repetibilidade pode ser obtido usando o autovetor paramétrico (\hat{y}) e a matriz de covariância (G), e é dado de acordo com a equação 6:

$$r = \frac{\alpha' R \alpha - \delta^2 y}{\delta^2 y (n-1)} \quad (6)$$

Após a estimativa do coeficiente de repetibilidade (r), estimou-se o número de medições (n_0) necessárias para prever o valor real de estruturas a serem fotografadas a um nível de 95% de precisão. O valor da determinação genotípica (R^2) foi obtido de acordo com a equação 7:

$$n = \frac{R^2(1-R)}{(1-R^2)r} \quad (7)$$

As estimativas do coeficiente de repetibilidade, do número de medições necessárias e do coeficiente de determinação foram obtidas por meio do programa Genes (Cruz 2016).

3. RESULTADOS E DISCUSSÃO

A análise de variância revelou diferenças significativas para todas as características avaliadas nas dez cultivares de soja (Tabela 1).

Tabela 1. Resumo da análise de variância das características morfológicas e espectrais, obtidas de imagens RGB de sementes e folhas de dez cultivares de soja.

| | | ----- Variáveis das Sementes ----- | | | | |
|--------|--|------------------------------------|------------|------------|------------|------------|
| | | Área | Perim | Circ | AR | Raund |
| QM | | 0,0026 ** | 0,0416 ** | 0,0019 ** | 0,0387 ** | 0,0094 ** |
| CV (%) | | 4,58 | 2,28 | 1,47 | 3,4 | 3,25 |
| | | Solidity | Blue | B | b | Green |
| QM | | 0,0001 ** | 210,042 ** | 145,742 ** | 94,2274 ** | 57,2021 ** |
| CV (%) | | 0,31 | 1,8 | 0,41 | 4,99 | 0,77 |
| | | S | a | Red | H | L |
| QM | | 391,82 ** | 5,3998 ** | 189,327 ** | 2,8383 ** | 5,8491 ** |
| CV (%) | | 4,17 | 20,66 | 0,46 | 2,29 | 0,81 |
| | | ----- Variáveis das Folhas ----- | | | | |
| | | Area | Perim | Circ | AR | Raund |
| QM | | 187,385 ** | 37,2372 ** | 0,0064 ** | 0,0958 ** | 0,0099 ** |
| CV (%) | | 21,01 | 10,79 | 5,56 | 7,79 | 7,87 |
| | | Solidity | Red | Green | Blue | H |
| QM | | 0,0002 * | 1099,1 * | 364,859 ** | 317,491 ** | 2009,69 ** |
| CV (%) | | 1,02 | 21,83 | 4,54 | 8,87 | 16,61 |
| | | L | S | a | B | b |
| QM | | 10,2478 ** | 5130,71 ** | 780,062 ** | 224,574 ** | 130,482 ** |
| CV (%) | | 3,2 | 29,58 | 35,82 | 4,84 | 23,41 |

QM: Quadrado Médio; CV: Coeficiente de Variação; Área: Área; Perim: Perímetro; Circ: Circularidade; AR: Aspect Ratio; Raund: Redondeza; Solidity: Solidez; Red: Vermelho; Green: Verde; Blue: Azul B: Banda do azul; b: Coordenada cromática do amarelo Lab; S: Saturação; a: Cordenada cromática R/G Lab; H: Valor de Matriz HSB; L: Luminosidade no modelo Lab; ** e * significativos a 1 e 5% de probabilidade; respectivamente; pelo teste F, ns: não-significativo; pelo teste F.

Quanto aos caracteres morfológicos das folhas (Figura.1), a variável circularidade exigiu maior número de folhas a ser fotografadas, variando de 31 a 53, dependendo do método, o maior valor foi observado para a ANOVA (53 folhas), indicando maior sensibilidade dessa variável à variação morfológica. Isso sugere que a circularidade é mais influenciada por variações sutis no formato da folha, demandando maior amostragem para estabilizar as estimativas.

Por sua vez para a variável AR (Aspect Ratio), o número de folhas variou entre 15 e 41, o menor valor foi observado para CP-COV com 15 folhas, enquanto a análise estrutural exigiu mais amostras. Esses resultados indicam que a razão comprimento/largura apresenta estabilidade intermediária, sendo moderadamente eficiente para discriminação de cultivares de soja. A variável Round necessitou entre 21 e 49 imagens de folhas, mostrando comportamento semelhante ao da circularidade. A necessidade de maior número de folhas reforça que atributos relacionados à forma

geométrica tendem a ser mais sensíveis a pequenas variações morfológicas e ambientais. A solidez foi a variável que demandou o maior número de folhas, especialmente na ANOVA (124 folhas) e na análise estrutural (84 folhas). Isso indica alta variabilidade intra-amostral, tornando a utilização dessa variável mais onerosa para distinção de cultivares de soja.

Por outro lado, a variável área apresentou valores muito semelhantes entre os métodos de análise de repetibilidade, variando de 18 a 20 folhas, essa característica torna a área foliar um descritor robusto e confiável para estudos de distinção de cultivares. Por sua vez, para determinar o valor real do perímetro são necessárias 22 imagens de folhas levemente superior ao da área da folha.

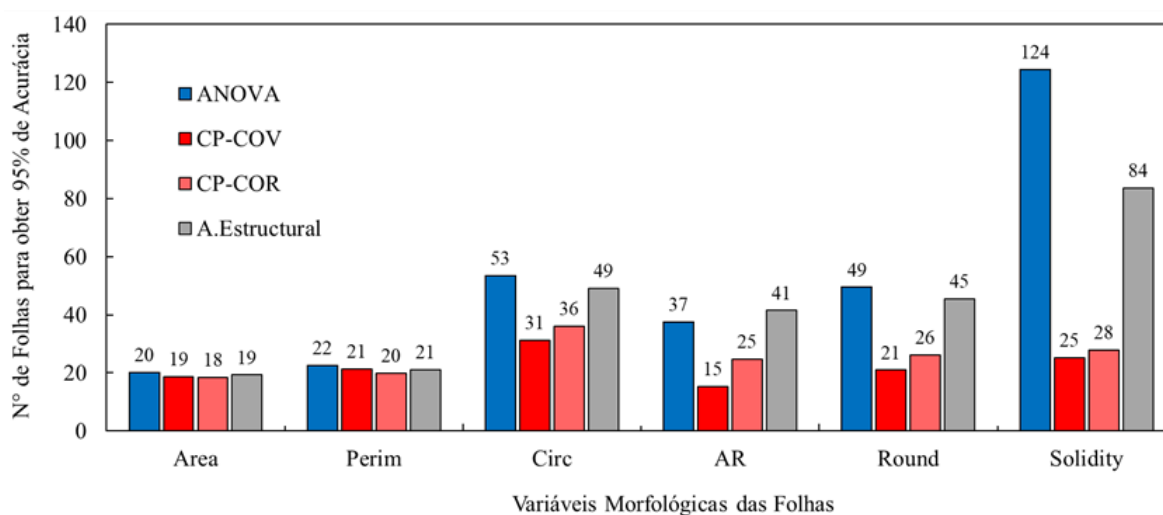


Figura 1. Números de folhas para obtenção do 95% de precisão dos caracteres morfológicos das folhas resultante das análises de repetibilidade: Anova: Análise de Variância; CP-COV: Covariâncias Fenotípicas; CP-COR: Matriz de Correlações; A. Estructural: Análise Estrutural; Area: Área; Perim: Perímetro; Cir: Circularidade; AR: Aspect Ratio; Raund: Redondeza; Solidity: Solidez.

Esse desempenho pode ser atribuído ao fato de que ambas serem variáveis relacionadas ao crescimento foliar, refletindo diretamente processos fisiológicos e genéticos associados à expansão e divisão celular, os quais tendem a ser mais conservados dentro de um genótipo e mais contrastantes entre genótipos (Nicotra et al. 2011, Perez-Harguindeguy et al. 2013).

Semelhante a este trabalho aonde as variáveis área e perímetro apresentaram o menor número de folhas para os métodos CP-COV e CP-COR, no estudo de Matsuo et al. 2012 as variáveis

comprimentos do hipocótilo, do epicótilo e do internódio requereram menor número de repetições quando se tratava dos métodos CP-COV e CP-COR, isso ressalta que a escolha da abordagem estatística impacta diretamente no número de repetição ou número de folhas necessários para alcançar acurácia de 95%.

Assim, a consistência do desempenho de área e perímetro das folhas, entre os diferentes caracteres avaliados, evidencia a robustez dessas variáveis na classificação de cultivares de soja de forma rápida e barata.

Quanto aos caracteres espectrais das folhas (Figura.2), a variável Red apresentou números elevados, variando de 9 folhas (CP-COV) até 87 folhas (A. Estrutural) a serem fotografados. Esse resultado indica alta variabilidade espectral, especialmente quando analisada por métodos mais sensíveis à estrutura da variância. A forte influência da clorofila e de variações fisiológicas na reflectância do vermelho aumenta a variabilidade biológica e reduz a estabilidade estatística das estimativas exigindo, portanto, maior tamanho amostral, sobretudo em métodos baseados em estrutura de covariância. Para a variável Green, o número de folhas variou entre 19 e 29, valores intermediários em relação às demais bandas. Isso indica uma estabilidade moderada, com menor variabilidade do que o vermelho, porém ainda sensível a diferenças fisiológicas e estruturais entre folhas.

Por sua vez a variável Blue apresentou exigência de 39 a 69 folhas, indicando alta sensibilidade a variações espectrais e variações morfológicas. Essa faixa espectral costuma ser mais afetada por espalhamento da luz e interferência atmosférica, o que justifica a maior necessidade amostral observada (Gaida et al. 2020).

A variável H (Matriz HSB) apresentou os menores valores absolutos entre todas as variáveis espectrais, variando de 13 a 15 folhas, independentemente do método. Isso indica alta estabilidade, tornando-a uma das variáveis mais eficientes para discriminação espectral com menor esforço amostral. A variável L(Luminosidade) apresentou comportamento contrastante entre métodos, variando de 28 a 102 folhas.

Entretanto, a variável S (Saturação) apresentou valores intermediários (27 a 29 folhas), com baixa variação entre métodos. Esse comportamento sugere boa estabilidade espectral e menor influência de ruídos ambientais quando comparada às bandas RGB puras. A variável a (Coordenada cromática R/G Lab) apresentou baixa exigência amostral (19 a 21 folhas), indicando elevada consistência e potencial para discriminação espectral de folhas, especialmente relacionada ao teor de clorofila (Gitelson et al. 2003). Para variável b (Coordenada cromática do amarelo Lab), o número de folhas variou entre 37 e 44, indicando comportamento intermediário.

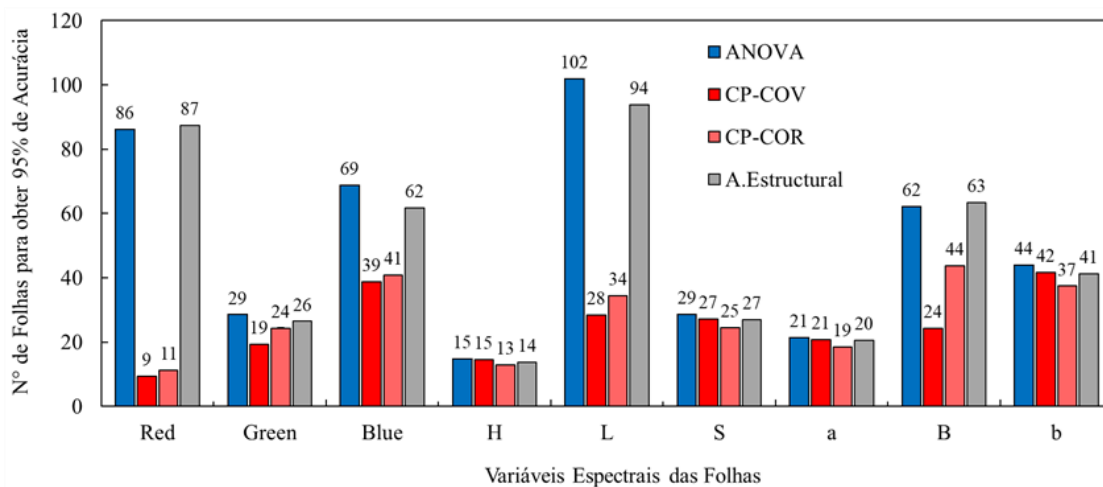


Figura 2. Números de folhas para obtenção do 95% de precisão para caracteres espectrais das folhas resultante das análises de repetibilidade: Anova: Análise de Variância; CP-COV: Covariâncias Fenotípicas; CP-COR: Matriz de Correlações; A.Estructuta: Análise Estrutural; Red: Vermelho; Green: Verde; Blue: Azul; H: Valor de Matriz HSB; L: Luminosidade no modelo Lab; S: Saturação; a: Coordenada cromática Lab; B: Banda do azul; b: Coordenada cromática do amarelo Lab.

O estudo de Li et al. 2023 demonstrou que a variação na reflectância espectral foliar é fortemente influenciada por diferenças genéticas ou ambientais, especialmente na região visível do espectro. Isso indica que algumas bandas espectrais (Red e Blue) podem ter alta variabilidade dentro da mesma planta, o que reforça os achados deste trabalho de que essas variáveis exigem mais folhas/amostras para alta acurácia e variáveis espectrais (H, S, a, b) podem efetivamente reduzir essa variabilidade e melhorar repetibilidade precisando de menor número de folhas.

O estudo de Nansen et al. 2024 analisou repetibilidade radiométrica em espectros foliares de soja sob diferentes condições de luz e estresses, observou que a repetibilidade espectral nas bandas entre 380–700 nm (visível) pode ser menor (< 80 %) sob certas condições, implicando alta variabilidade nesses comprimentos de onda. Em contraste, as bandas no intervalo 730–1015 nm (próximo ao infravermelho) apresentaram repetibilidade próxima de 100 %, implicando baixa variabilidade nesses comprimentos de onda. Esse padrão explica por que a variável (Red) da banda visível exige mais número de folhas para obter alta acurácia (menor repetibilidade), entretanto as variáveis (H, S, a, b) que são bandas transformadas muitas vezes têm maior repetibilidade, alinhando-se com o menor número de folhas dessas variáveis para atingir o 95% de precisão neste estudo.

Quanto aos caracteres morfológicos das sementes (Figura 3.), as variáveis área e perímetro apresentaram valores intermediários de amostragem, com necessidade entre 120 e 147 sementes nos métodos baseados em ANOVA e análise estrutural. Esses resultados indicam que, embora sejam descritores amplamente utilizados, essas variáveis são sensíveis a pequenas variações no contorno da semente, o que aumenta a variabilidade dentro da amostra de uma mesma planta. Já a variável circularidade apresentou os menores valores de amostragem entre as variáveis morfológicas, especialmente nos métodos CP-COV e CP-COR (24–28 sementes). Isso indica alta repetibilidade, provavelmente porque a circularidade é uma métrica adimensional que normaliza variações de tamanho, tornando-se menos sensível a ruídos experimentais.

Por sua vez, a variável AR apresentou comportamento intermediário, com exigência de aproximadamente 54–61 sementes nos métodos baseados em covariância e correlação, e valores mais altos quando avaliada por ANOVA. A variável Raund mostrou necessidade moderada de sementes (54–59), semelhante a AR, indicando que métricas baseadas em proporção geométrica apresentam maior estabilidade do que medidas absolutas. A variável Solidity apresentou o maior contraste entre métodos, tendo valores extremamente elevados pelo método ANOVA (até 344 sementes) e valores bem menores nos métodos CP-COV e CP-COR (26–32 sementes). Esse comportamento indica que a solidez é altamente sensível a pequenas irregularidades no contorno, o que amplifica a variabilidade

quando analisada apenas por ANOVA. Métodos multivariados ou baseados em correlação conseguem capturar melhor a estrutura do dado e reduzir o número necessário de amostras.

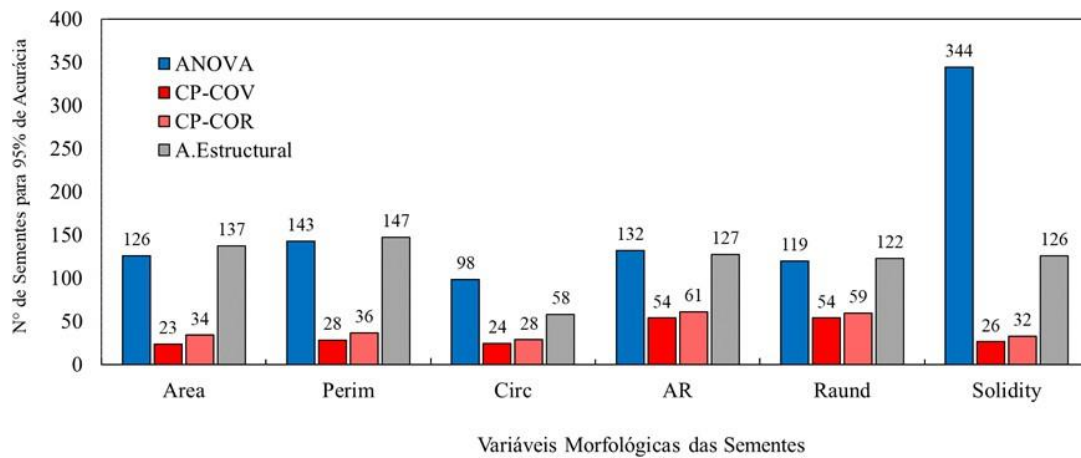


Figura 3. Números de sementes para obtenção do 95% de precisão dos caracteres morfológicos das sementes resultante das análises de repetibilidade: Anova: Análise de Variância; CP-COV: Covariâncias Fenotípicas; CP-COR: Matriz de Correlações; A. Estructural: Análise Estrutural; Area: Área; Perim: Perímetro; Cir: Circularidade, AR: Aspect Ratio; Raund: Redondeza, Solidity: Solidez.

Observou-se que as variáveis relacionadas à forma das sementes (Raund, AR, Circularidade e Solidity), demandaram maior número de unidades amostrais quando comparadas às variáveis dimensionais (área e perímetro). Esse comportamento indica maior variabilidade intrínseca desses descritores, refletindo menor repetibilidade experimental.

Esse padrão está de acordo com os resultados reportados por Silva et al. 2023, que, ao avaliarem sementes de soja por meio de análise de imagens digitais, verificaram que descritores de forma apresentaram maior sensibilidade a variações morfológicas e, conseqüentemente, menor estabilidade estatística. De forma semelhante, Abud et al. 2022 observaram que variáveis dimensionais apresentaram maior variância residual em sementes de feijão, exigindo maior número de observações para atingir níveis adequados de precisão.

Entre as variáveis morfológicas das sementes, a circularidade destacou-se por apresentar o menor número de sementes fotografadas para atingir 95% de precisão nos métodos aplicados. Esse desempenho está associado ao fato de a circularidade representar a regularidade geométrica global da

semente, característica fortemente relacionada à identidade genotípica e menos suscetível às variações ambientais que afetam atributos de tamanho, como área e perímetro assim relatado por Duc et al. 2023.

Os resultados obtidos neste estudo indicam que o número mínimo de sementes necessário para atingir 95% de precisão variou de 23 a 54 sementes, dependendo da variável morfológica avaliada, quando utilizado o método CP-COV, evidenciando elevada eficiência amostral. Esses valores são inferiores ou semelhantes relatados na literatura para outras espécies vegetais. Em um estudo com sementes de mamoeiro, Favero et al. 2025 observaram que o número de sementes necessário para estimar com precisão características dimensionais variou entre 39 e 113 unidades, dependendo da variável analisada.

De forma semelhante, Cargelluti et al. 2018 relataram que, em diferentes culturas, o tamanho amostral exigido para caracteres morfológicos pode variar amplamente, alcançando valores superiores a 100 unidades quando a variabilidade é elevada. Esses autores destacam que características mais estáveis tendem a demandar menor número de amostras, enquanto atributos mais heterogêneos exigem amostragens maiores. Assim, os resultados do presente estudo indicam que as variáveis morfológicas analisadas apresentam baixa a moderada variabilidade, permitindo estimativas confiáveis com um número reduzido de sementes, em consonância com o observado em estudos anteriores e reforçando a eficiência do método CP-COV para a determinação do tamanho amostral adequado.

Quanto aos caracteres espectrais das sementes (Figura. 4), os números de sementes obtidos por variável para atingir 95% de precisão variou, observa-se que a variável Red apresentou a menor exigência amostral, variando de 18 a 24 sementes entre os métodos, indicando elevada estabilidade espectral. Para Green, a exigência foi significativamente maior, com valores entre 37 e 163 sementes, evidenciando maior variabilidade. A variável Blue demandou entre 39 e 123 sementes, enquanto H (matiz) apresentou uma das maiores exigências, variando de 38 a 183 sementes. A variável L (Luminosidade) requereu entre 40 e 153 sementes, refletindo sensibilidade a variações de brilho. Para

S (saturação), os valores oscilaram entre 35 e 72 sementes, caracterizando exigência intermediária. Já a variável (a) variou de 22 a 164 sementes, indicando comportamento heterogêneo entre os métodos, enquanto B apresentou baixa exigência amostral, entre 17 e 25 sementes. Por fim, a variável b exigiu de 34 a 52 sementes, mostrando variabilidade moderada em comparação às demais variáveis espectrais.

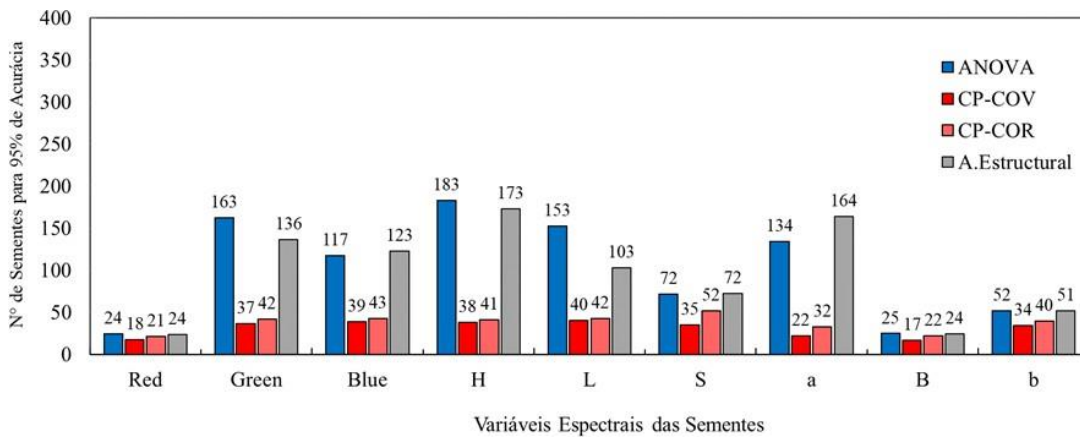


Figura 4. Números de sementes para obtenção do 95% de Acurácia dos caracteres espectrais das sementes resultante das análises de repetibilidade: Anova: Análise de Variância; CP-COV: Covariâncias Fenotípicas; CP-COR: Matriz de Correlações; A.Estructural: Análise Estrutural; H: Valor de Matriz HSB; L: Luminosidade no modelo Lab; S: Saturação; a: Coordenada cromática Lab; B: Banda do azul; b: Coordenada cromática do amarelo Lab.

Para obter o valor real de uma cultivar de soja para cada variável morfológica das folhas deve ser fotografadas o número mínimo de 20 folhas para área, 22 folhas para o perímetro, 53 folhas para a circularidade, 41 folhas para o Aspect Ratio (AR), 49 folhas para a redondeza (Raund) e 124 folhas para a Solidity; por outro lado em variáveis espectrais, deve ser fotografada 87 folhas para a variável Red, 29 folhas para variável Green, 69 folhas para variável Blue, 15 folhas para a Matriz HSB (H), 102 folhas para a Luminosidade (L), 27 folhas para a Saturação (S), 21 folhas para a coordenada cromática R/G Lab (a), 63 folhas para Banda do azul (B) e 44 folhas para a coordenada cromática do amarelo Lab (b).

Entretanto, para as variáveis morfológicas das sementes deve ser fotografadas o número mínimo de 137 sementes para a área, 147 sementes para o perímetro, 98 sementes para a circularidade,

122 sementes para a redondeza (Raund), 344 sementes para a Solidity, por outro lado para as variáveis espectrais das sementes deve ser fotografadas o número mínimo de 24 sementes para a variável Red, 163 sementes para Green, 123 sementes para Blue, 183 sementes para a Matriz HSB (H), 153 sementes para a Luminosidade (L), 72 sementes para a Saturação (S), 164 sementes para a coordenada cromática R/G Lab (a), 24 sementes para a Banda do azul (B) e 52 sementes para a coordenada cromática do amarelo Lab (b).

Assim como no estudo de DUARTE *et al.*, (2022), métodos multivariados (CP-COV, CO-COR) tenderam a reduzir substancialmente o número de unidades amostrais necessárias em comparação com a ANOVA, exatamente como se observou nas variáveis espectrais deste estudo. O método anova proporcionou os menores valores dos coeficientes de repetibilidade, entre os métodos avaliados, independentemente da variável analisada. Os métodos multivariados de componentes principais, baseados na CP-COV e na CP-COR, obtiveram os melhores resultados, com os maiores coeficientes de repetibilidade.

4. CONCLUSÃO

Os resultados confirmam que devem ser fotografados no mínimo 124 folhas para as variáveis morfológicas das folhas, e 102 folhas para as variáveis espectrais, por outro lado devem ser fotografados no mínimo 344 sementes para variáveis morfológicas das sementes e 183 sementes para as variáveis espectrais.

REFERÊNCIAS

- Abeywardena V (1972) An application of principal componente analysis in genetics. **Journal of Genetics** 61: 27-51.
- Berns RS (2019). **Billmeyer and Saltzman's principles of color technology**. John Wiley e Sons.
- Abud HF, Mesquita CM, Sarmiento EC, Melo RS, Lima KP e Silva AF (2022) Image analysis of the seeds and seedlings of *Vigna radiata* L. **Rev. Ciênc. Agron** 53: e20207303.

Boelt B, Shrestha S, Salimi Z, Jørgensen JR, Nicolaisen M, Carstensen JM (2018) Multispectral imaging - a new tool in seed quality assessment? **Seed Science Research** 28: 222-228.

Chaugule AA, Mali SN (2016) Identification of rice varieties based on new angular characteristics of the seeds. **Comput. Electron. Agric** 123: 415–422.

Diel MI LúcioAD, Olivoto T, Pinheiro MVM, Krysczun DK, Sari BG e Schmidt D (2020) Repeatability coefficients and number of measurements for evaluating traits in strawberry. **Acta Scientiarum. Agronomy** 42: 1-9.

Cargnelutti-Filho A, Toebe M, Mendonça-Alves B, Burin C e Wartha CA (2018) Sample size to estimate the mean of traits in jack bean. **Brazilian Journal of Agricultural Sciences** 13:1-7.

Cruz CD, Regazzi AJ e Carneiro PCS (2012) **Modelos Biométricos Aplicados ao Melhoramento Genético**. 4th Edition, Editora UFV, Viçosa, 514 p.

Cruz CD (2016) Genes Software – extended and integrated with the R, Matlab and Selegen. **Acta Scientiarum** 38:547-552.

Favero M, Arpini BS, Costa NS, Moitinho ACR, Souza LF, Moreira SO (2025) **Tamanho amostral ótimo para variáveis morfoagronômicas em semente de mamão da cultivar Rubi Incaper** 511.XXIII Encontro Latino Americano de Iniciação Científica, XIX Encontro Latino Americano de Pós-Graduação e IX Encontro de Iniciação à Docência - Universidade do Vale do Paraíba.

Ferreira FM, Rocha JRS, Alves RS, Elizeu AM, Benites RFG, Resende MDV (2020) Estimates of repeatability coefficients and optimum number of measures for genetic selection of *Cynodon spp.* **Euphytica** 216:70.

Ferreira RP, Vasconcelos ES, Cruz CD, Júnior WB, Rassini JB e Freitas AR (2010) Determinação do coeficiente de repetibilidade e estabilização genotípica das características agronômicas avaliadas em genótipos de alfafa no ano de estabelecimento. **Revista Ceres** 57: 642-647.

Gaida W, Breunig FM, Galvão LS e Ponzoni FJ (2020). Correção Atmosférica em Sensoriamento Remoto: Uma Revisão. **Revista Brasileira De Geografia Física** 13:229–248.

Gitelson_ AA, Gritz Y, Merzlyak MN (2003) Relationships between leaf chlorophyll content and spectral reflectance and algorithms for non-destructive chlorophyll assessment in higher plant leaves.

Journal of Plant Physiology **160**:271-282.

Guilhen JH, Noia LR, Bernardes CO, Oliveira WB, Marçal TB e Ferreira MS (2019) Repeatability analysis of guava fruit and leaf characteristics. **Bioscience Journal** **35**: 2.

Hartman GL, Rupe JC, Sikora EJ, Domier LL, Davis JÁ, Steffey KL (2015) **Compendium of soybean diseases**. 5 ed. Saint Paul: APS Press, 56p.

Jiang GL (2020) Comparison and application of non-destructive NIR evaluations of seed protein and oil content in soybean breeding. **Agronomy****10**: 77.

Kurtulmus F, Ilknur A, Ismail K (2016) Classificação de sementes de pimenta usando visão computacional baseada em rede neural. **Int. J. Agric. Biol. Eng** **9**: 51–62.

Li C Czyż EA, Halitschke R, Baldwin IT, Schaepman ME e Schuman MC (2023) Evaluating potential of leaf reflectance spectra to monitor plant genetic variation. **Plant Methods** **19**:108.

Liu Y, Xie H, Chen Y, Tan K, Wang L e Xie W (2016) Mutual neighborhood information and its application in hyperspectral band selection for classification. **Chemom. Intell. Lab. Syst** **157**:140–151

Matsuo É, Sedyamail T, Cruz CD e Oliveira RCT (2012) Análise da repetibilidade em alguns descritores morfológicos para soja. **Ciência Rural** **42**:189-196.

MARANHÃO - GOVERNO DO ESTADO DO MARANHÃO. Gerência de Planejamento e Desenvolvimento Econômico - GEPLAN. Atlas do Maranhão. São Luís: Universidade Estadual do Maranhão, p.39, 2022.

Marçal TDS, Guilhen HS, Oliveira WB, Ferreira MS, Resende MD e Ferreira A (2016) Repeatability of biometric characteristics of juçara palm fruit. **Embrapa Florestas** **32**:890-898.

Masetto TE, Faleiros RS (2017) Análise de imagens digitais para caracterização de formas e tamanhos de sementes. **Engenharia Agrícola, Jaboticabal** **37**:82-91.

Mansour H, Nordheim EV e Rutledge JJ (1981) Estimadores de repetibilidade. **Genética Teórica e Aplicada**, **60**:151-156.

Nansen C, Savi P, Mantri A (2024) Methods to optimize optical sensing of biotic plant stress – combined effects of hyperspectral imaging at night and spatial binning. **Plant Methods** **20**:163.

Nicotra AB, Leigh A, Boyce CK, Jones CS, Niklas KJ, Royer DL, Tsukaya H. (2011) The evolution and functional significance of leaf shape in angiosperms. **Functional Plant Biology** **38**: 535–552.

Pedrozo C, Barbosa HP, Silva FL, Resende MD e Peternell LA (2011) Repeatability of full-sib sugarcane families across harvests and the efficiency of early selection. **Euphytica** **182**: 423-430.

Persson M (2005) Estimating surface soil moisture from soil color using image analysis. **Vadose Zone Journal** **4**:1119-1122.

Perez- Harguindeguy N, Díaz S, Garnier E, Lavorel S, Poorter H e Jaureguiberry P (2013) New handbook for standardised measurement of plant functional traits worldwide. **Australian Journal of Botany** **61**:167–234.

Rodrigues EV, Dhaer RF, Gravina GH, Viana AP, Araújo MB e Oliveira MS (2020) Repeatability estimates and minimum number of evaluations for selection of elephant-grass genotypes for herbage production. **Bioscience Journal** **36**:1.

Rueden CT, Schindelin J, Hiner MC, DeZonia BE, Walter AE e Arena ET (2017) ImageJ for the next generation of scientific image data. **BMC Bioinformatics** **18**: e 529.

Sedimaya T, Filipe S e Borém A (2015) Soja **do Plantio à Colheita**. 1 ed. Viçosa: Editora UFV, 312p.

Silva FF, Gomes-Junior FG, Rego CH, Marassi AG e Tannús A (2023) Advances in imaging technologies for soybean seed analysis. **Journal of Seed Science** **45**: e202345022.

Woyann LG, Milioli AS, Bozi AH, SC, Matei G, Storck L e Benin G (2018) Repeatability of associations between analytical methods of adaptability, stability, and productivity in soybean. **Pesq. agropec. bras., Brasília** **53**:63-73.

CONSIDERAÇÕES FINAIS

O estudo mostrou que a aplicação de algoritmos de Machine Learning na classificação de cultivares de soja mediante processamento de imagens RGB de folhas e sementes, pode ser uma solução viável rápida e com baixo custo para a identificação de cultivares de soja, destacando-se entre os algoritmos aplicados que o Random Forest com resultados de 93% de porcentagem de classificações corretas é uma alternativa promissora viável a ser aplicada na agricultura.

Por outro lado, neste estudo foi possível determinar o número real e mínimo de folhas e sementes de uma mesma cultivar de soja a ser fotografadas sempre que fosse necessário fazer um estudo de classificação de soja usando variáveis morfológicas e espectrais, como as avaliadas neste trabalho o qual viabiliza a coleta de dados para os futuros estudos de identificação de cultivares de soja na agricultura.

Anexos

Instruções para Autores

Processo de Submissão

CBAB – MELHORAMENTO DE CULTURAS E BIOTECNOLOGIA APLICADA

Desde 2001

Os autores de artigos na revista CBAB beneficiam-se dos seguintes itens:

- CBAB é uma revista de acesso aberto diamante.
- Todos os artigos são publicados gratuitamente.
- Submissão e revisão de manuscritos através da plataforma ScholarOne
- Publicação rápida: tempo médio de 91 dias, em 2023.
- Artigos em PDF disponíveis gratuitamente online.
- Avaliação por renomados revisores nacionais e internacionais da Web of Science.
- Aplicação do software iThenticate contra plágio
- Todos os artigos são publicados com DOI, localização eletrônica e código QR.
- Ampla visibilidade, com publicação de artigos exclusivamente em inglês desde 2001.

Processo de Submissão

1. Acesse a plataforma de submissão:

<https://mc04.manuscriptcentral.com/cbab-scielo>

2. Criar uma conta

3. Criar um ID ORCID

4. Conecte-se

5. Iniciar submissão:

- a. Selecione o tipo de manuscrito, insira o título e o resumo.
- b. Faça o upload dos arquivos
- c. Insira as palavras-chave do seu manuscrito.
- d. Insira as informações do seu coautor.
- e. Insira as informações do avaliador
- f. Digite ou cole sua carta de apresentação.
- g. Analise as informações e envie o manuscrito de acordo com as instruções abaixo.

Taxa de publicação

Não há taxas de submissão nem custos de processamento de artigos (APCs) para autores que submetem trabalhos à CBAB.

Responsabilidades Éticas

Os autores devem evitar apresentar resultados de pesquisa imprecisos, pois isso pode prejudicar a credibilidade da revista e a competência do autor. O manuscrito não deve ser submetido a múltiplas revistas para consideração simultânea. Os resultados não devem ser fabricados, falsificados ou usados de forma inadequada. Os autores são responsáveis pelas opiniões e ideias expressas, que podem não refletir a visão do Conselho Editorial. No entanto, o Conselho Editorial reserva-se o direito de sugerir ou solicitar quaisquer modificações.

Plágio

A revista adota o software iThenticate para identificação de plágio. O manuscrito submetido deve ser original e não ter sido publicado anteriormente em nenhum outro lugar, seja parcial ou integralmente, a menos que o novo manuscrito seja uma continuação do anterior. Solicitamos transparência quanto à reutilização de material para evitar preocupações com reciclagem de texto (autoplágio). Nenhum dado, texto ou teoria de outras fontes deve ser apresentado como se fosse de autoria do autor (plágio). Outros trabalhos devem ser devidamente creditados, incluindo material que seja copiado, resumido e/ou parafraseado.

Aceitação de pré-publicações

Aceitamos submissões de preprints na CBAB. Um preprint refere-se um manuscrito preparado para submissão e depositado em servidores de preprints de boa reputação, seja antes ou simultaneamente à submissão um periódico. Aceitamos apenas artigos previamente depositados nesses servidores, cada um dos quais passa por uma avaliação rigorosa de reconhecimento e qualidade pela CBAB. Após aceitação e publicação do artigo na CBAB, os autores devem atualizar o status de registro no servidor de preprints, fornecendo a referência completa da publicação no periódico.

Disponibilidade de dados e informações complementares

Para garantir a qualidade e o reconhecimento dos dados gerados, o periódico CBAB recomenda e segue os princípios FAIR (Localizáveis, Acessíveis, Interoperáveis e Reutilizáveis). Os autores são fortemente encorajados incluir em seu artigo uma declaração opcional sobre a disponibilidade dos dados. Os dados de pesquisa abrangem qualquer informação observada, coletada, gerada ou criada para validar resultados de pesquisa originais. Isso inclui dados brutos e processados, código, software, algoritmos, protocolos e métodos. Os autores são responsáveis por garantir que os dados confidenciais não sejam compartilhados, a fim de manter os padrões éticos, legais e de privacidade. O método preferencial para o compartilhamento de dados de pesquisa é por

meio de um repositório de dados, que permite que autores/colaboradores arquivem e compartilhem dados de pesquisa originais e processados, organizados, com ou sem material suplementar.

Diversos repositórios de dados estão disponíveis para o depósito de dados de pesquisa, acessíveis em <https://repositoryfinder.datacite.org/>. Entre os repositórios de dados recomendados pelo CBAB está o repositório SciELO Data <https://data.scielo.org/dataverse/brcbab>, um sistema aberto e isento de taxas que permite aos colaboradores depositar dados de pesquisa associados a um manuscrito específico. Os autores que utilizam o SciELO Data devem apresentar os dados somente após a aceitação do manuscrito.

Essas Declarações de Disponibilidade de Dados devem incluir detalhes sobre a localização onde os dados que sustentam as conclusões apresentadas no artigo estão acessíveis, como hiperlinks para conjuntos de dados publicamente arquivados, analisados ou gerados durante o estudo. Além disso, podem especificar se os dados estão disponíveis mediante solicitação aos autores. Um modelo a ser seguido é o seguinte:

Os conjuntos de dados gerados e/ou analisados durante a presente pesquisa estão acessíveis no repositório [Nome], disponível em [Link para os conjuntos de dados].

Defendemos a adesão às práticas de Ciência Aberta para garantir segurança e transparência. Para cumprir as políticas da SciELO, os autores devem preencher o formulário de Conformidade com a Ciência Aberta.

Caso o manuscrito contenha arquivos suplementares, os autores devem removê-los do documento principal e especificar nos Agradecimentos que esses arquivos estão acessíveis por meio do autor correspondente, seguindo o modelo:

Os conjuntos de dados gerados e/ou analisados durante a presente pesquisa estão disponíveis mediante solicitação razoável ao autor correspondente.

Autoria

Recomenda-se enfaticamente aos autores que verifiquem se o grupo de autores e a ordem dos autores estão corretos no momento da submissão. Geralmente, não é permitido adicionar ou remover autores durante as etapas de revisão. No entanto, em alguns casos, isso pode ser permitido. Não é possível alterar a autoria de um manuscrito após a aceitação.

Todos os autores cujos nomes constam na submissão devem ter:

1. contribuiu significativamente para a concepção ou o projeto do trabalho ou para a aquisição, análise ou interpretação dos dados.
2. redigiu o trabalho ou o revisou cuidadosamente quanto ao conteúdo intelectual importante.

3. aprovou a versão final
4. Concordou em ser responsável por todos os aspectos do trabalho, garantindo que as questões relacionadas à precisão ou integridade sejam analisadas e resolvidas.

Recomenda-se aos autores que incluam uma declaração especificando a contribuição de cada autor para a pesquisa e preparação do manuscrito. Exemplo de declaração:

Todos os autores contribuíram para a concepção e o planejamento do estudo. [Nome completo], [Nome completo] e [Nome completo] prepararam o material, coletaram e analisaram os dados. A versão inicial foi escrita por [Nome completo], e todos os autores forneceram feedback sobre as versões subsequentes. Todos os autores leram e aprovaram a versão final do manuscrito.”

Autor correspondente

Um autor correspondente é designado para agir em nome de todos os coautores e garantir que as questões relacionadas à precisão ou integridade do artigo sejam devidamente abordadas. Este autor é responsável por garantir que todos os autores aprovaram o manuscrito antes do envio, incluindo seus nomes e ordem; coordenar toda a comunicação entre a revista e os coautores, antes e depois da publicação; e assegurar que as declarações e os termos de transparência de todos os autores sejam incluídos.

Instruções para Manuscritos

A CBAB publica exclusivamente em inglês, portanto, os autores devem submeter seus artigos em inglês. É obrigatório que o manuscrito, após aprovação, seja revisado linguisticamente, o que deve ser feito exclusivamente pelo revisor oficial da revista tradutores, sendo a responsabilidade pela organização deste serviço do autor.

Exclua todas as informações de autoria e correspondência do arquivo do manuscrito. Como a revista adota uma política de revisão anônima (revisão por pares duplo-cega), os autores não devem revelar suas identidades no manuscrito. O autor será solicitado a inserir essas informações em um formulário separado durante o processo de submissão, antes de enviar o arquivo do manuscrito.

Tipos de Manuscritos

Artigo

O manuscrito completo deve seguir a seguinte sequência: TÍTULO, RESUMO, PALAVRAS-CHAVE, INTRODUÇÃO, MATERIAL E MÉTODOS, RESULTADOS E DISCUSSÃO, AGRADECIMENTOS, REFERÊNCIAS, TABELAS e FIGURAS. A INTRODUÇÃO deve incluir uma breve revisão da literatura sobre o tema e os objetivos do estudo. A seção MATERIAL E MÉTODOS deve permitir que outros pesquisadores

reproduzam o experimento. A seção RESULTADOS E DISCUSSÃO deve ser apresentada em conjunto para facilitar a leitura. Os AGRADECIMENTOS devem ser concisos e limitados a colaboradores relevantes e agências financiadoras.

Lançamento de cultivares

Novas cultivares merecem atenção especial devido ao seu papel fundamental no melhoramento de plantas e, conseqüentemente, na agricultura. Uma contribuição para esta seção deve incluir um resumo, palavras-chave, uma introdução, menção aos métodos de melhoramento aplicados, características de desempenho (com apresentação de testes comparativos de médias), produção de sementes básicas e conter, no mínimo, referências, tabelas e figuras.

Observação

As notas são destinadas informar sobre novos estudos ou observações, não sendo, portanto, necessárias ferramentas analíticas. Elas podem abordar um tema de interesse geral, descrever brevemente um estudo original, relatar uma pesquisa participativa, expressar observações de especial interesse nas áreas de pesquisa, ensino e ciências aplicadas, ou comentar o lançamento de um novo software em uma área relacionada ao melhoramento de plantas.

Lançamento de software/dispositivo

Novos softwares e dispositivos são importantes para o aprimoramento atual e, portanto, para a agricultura. Esta nova seção deve conter um resumo, palavras-chave, introdução, métodos, histórico utilizado para o desenvolvimento do manuscrito, características de desempenho, aplicações, formas de acesso e, no mínimo, referências, tabelas e figuras.

Artigo de revisão

Autores de referência em determinados temas serão convidados pela Comissão Editorial a escrever uma Revisão, que deverá abordar especificamente assuntos instigantes que merecem uma análise mais aprofundada do seu estágio de desenvolvimento.

Programa de Melhoramento de Plantas

Programas de melhoramento genético de destaque em termos de inovação, eficiência, impacto e/ou continuidade podem ser exemplificados no CBAB.

Texto do Manuscrito

O manuscrito deve ser submetido em **formato Word**. A fonte deve ser Times New Roman, tamanho 12, espaçamento duplo, com margens de 20 mm e numeração consecutiva no canto superior direito.

O texto não deve exceder 18 páginas para artigos, revisões e programas de

melhoramento de plantas, e 12 páginas para cultivares, notas e lançamentos de software/dispositivos. Isso inclui tabelas e figuras separadas (uma em cada página) ao final do texto. Todas as equações, modelos e símbolos devem ser escritos usando a ferramenta de equações.

O TÍTULO deve ser claro, conciso e expressar a essência do artigo. Não deve ultrapassar 15 palavras e deve estar em negrito, alinhado à esquerda, com a inicial maiúscula. O RESUMO não deve conter mais de 150 palavras para artigos, revisões, notas e programas de melhoramento de plantas, e 50 para cultivares e lançamentos de software/dispositivos. É permitido um máximo de cinco PALAVRAS-CHAVE, diferentes do título.

As coordenadas devem seguir o exemplo: lat 34° 36' 40.20" S, long 56° 26' 33.15" W e alt 14 m acima do nível do mar.

Denote e inter-relacione unidades como potências positivas ou negativas, não com barras, por exemplo:

kg ha⁻¹, não kg/ha.

Tabelas

Cada tabela deve estar em uma página separada após as Referências. As tabelas devem ser criadas com a função “Tabelas” do MS Word. As notas de rodapé devem estar em Times New Roman, tamanho 10. O título da tabela deve aparecer imediatamente acima e seguir o modelo: **Tabela 1.** [Inserir título aqui]

Figuras

Cada figura deve estar em uma página separada após as Referências e Tabelas. A figura deve ser colorida e de boa qualidade. Dê preferência a arquivos de imagem no formato TIFF, com resolução mínima de 300 dpi. Caso não seja possível adicioná-la em cores, certifique-se de que as informações principais permaneçam visíveis. As notas de rodapé devem estar em Times New Roman, tamanho 10. O título da figura deve aparecer abaixo dela e seguir o modelo: **Figura 1.** [Inserir título aqui].

Referências

Tenha cuidado com as REFERÊNCIAS. Nunca cite dissertações, teses ou qualquer outra literatura não publicada. As citações mencionadas no texto devem ser feitas pelo sobrenome do autor, seguido do ano e em ordem cronológica (por exemplo, Liu 2021, Chaves et al. 2020, Pereira e Amaral Júnior 2001). As referências listadas no item REFERÊNCIAS devem estar em ordem alfabética, conforme os exemplos a seguir:

Artigos

Pereira MG e Amaral Júnior AT (2001) Estimativa de componentes genéticos em

milho pipoca com base no delineamento aninhado. **Melhoramento de Culturas e Biotecnologia Aplicada 1**: 3-10.

Chaves MS, Silva GBP, Caierão E, Federizzi LC e Martinelli JA (2020) Um século de melhoramento de trigo no Brasil: origem e herança do locus Lr34 em variedades de trigo lançadas de 1922 a 2016. **Crop Breeding and Applied Biotechnology 20**: e27952027.

Obs.: Não utilize abreviações para o nome do periódico.

Livro

Hallauer AR, Carena MJ e Miranda Filho JB (2010) **Genética quantitativa no melhoramento do milho**. Springer, Nova York, 664p.

Capítulo do livro

Morais PPP e Borem A (2017) Cultivares GM. Em Silva FL, Borem A, Sedyama T e Ludke WH (eds) **Melhoramento de soja**. Springer, Nova York, p. 174-189.

Congresso

Frey KJ (1992) Perspectivas de melhoramento de plantas para a década de 1990. Em Stalker HT e Murphy JP (eds.) **Anais do simpósio sobre melhoramento de plantas na década de 1990**. CAB, Wallingford, p. 1-13.

Internet

CONAB – Companhia Nacional de Abastecimento (2020) Acompanhamento da safra brasileira de café no Brasil. Disponível em <website>. Acesso em 18 de fevereiro de 2019.

Boletim informativo

Casela CR e Ferreira AS (2003) A cercosporiose na cultura do milho. Embrapa Milho e Sorgo, Sete Lagoas, 17h. (**Boletim Informativo, 24**).

Informações complementares

Caso o manuscrito contenha arquivos suplementares, os autores podem removê-los do documento principal e informar nos agradecimentos que esses arquivos estão disponíveis com o autor correspondente.

Submissão e Avaliação

O processo de revisão editorial só se inicia se o manuscrito seguir as normas estabelecidas (instruções para autores); caso contrário, será devolvido para avaliação de adequação. O Conselho Editorial decidirá sobre a sua aprovação com base na importância, originalidade e clareza do manuscrito, bem como na sua relevância para a revista. Manuscritos submetidos que não sejam adequados ao escopo da revista serão rejeitados. Manuscritos com problemas relativos à estrutura da composição textual serão devolvidos e, se corrigidos, poderão ser reenviados.

Após aprovação inicial pelo Conselho Editorial, os manuscritos passarão para a fase subsequente, na qual os revisores serão selecionados de um grupo de pesquisadores com reconhecida experiência na área específica e, preferencialmente, afiliados a instituições distintas daquelas de origem do manuscrito.

Antes de iniciar análise de um manuscrito, o Conselho Editorial solicita aos Revisores que considerem a possibilidade de situações ou relações que possam comprometer a imparcialidade do processo de avaliação, tais como colaboração regular com possíveis autores em atividades de pesquisa, relação orientador-aluno, interesse comercial na publicação e relações familiares. Nesses casos, o Revisor deve comunicar imediatamente a impossibilidade de realizar a revisão.

Processo de revisão por pares

A revisão por pares refere-se à obtenção de aconselhamento técnico-científico sobre manuscritos por parte de especialistas na área, denominados Consultores *ad hoc*. Estes revisarão os manuscritos sem conhecer a identidade dos autores. Cada Consultor *ad hoc* sugerirá uma das seguintes opções: aceitar a submissão, solicitar revisões (menores ou maiores) ou rejeitar a submissão. O Consultor pode sugerir alterações, adicionar excertos e seções, apontar informações ambíguas, redundantes e incoerentes, indicar o uso incorreto de terminologia e solicitar explicações mais detalhadas. O autor correspondente receberá todas as opiniões sem conhecer a identidade dos Consultores. Caso os autores discordem das sugestões, deverão apresentar uma justificativa e enviá-la ao Consultor.

Editor/Revisor com a nova versão.

Após aprovação

Após aprovação, todas as etapas seguintes são realizadas por e-mail. Os autores devem entrar em contato com um dos tradutores da CBAB para a revisão linguística e pagar pelo serviço. Assim que receberem o manuscrito revisado, o autor correspondente deverá enviá-lo para o e-mail da revista (cbabjournal@gmail.com). Caso seja necessária alguma modificação, a revista o notificará.

Após a revisão e as correções serem concluídas, será enviada uma prova de leitura. A prova serve para verificar erros de formatação, bem como a integridade e a precisão do texto, das tabelas e das figuras. O Editor-Chefe deve aprovar quaisquer alterações de conteúdo, como novos resultados, valores corrigidos, título ou autoria. Portanto, recomenda-se que os autores leiam atentamente todo o manuscrito, fazendo as correções propostas e destacando possíveis erros ou modificações a serem feitas pelo Editor.

Declaração de direitos autorais

Os autores não serão reembolsados pela publicação de artigos no CBAB e cederão seus direitos autorais ao periódico. No entanto, os autores mantêm a responsabilidade exclusiva pelo conteúdo do artigo publicado. Os editores reservam-se o direito de fazer ajustes textuais durante o processo de edição do manuscrito.