



UFC

UNIVERSIDADE FEDERAL DO CEARÁ

CENTRO DE CIÊNCIAS AGRÁRIAS

DEPARTAMENTO DE FITOTECNIA

PROGRAMA DE PÓS-GRADUAÇÃO EM AGRONOMIA/FITOTECNIA

LUCIMARA RIBEIRO VENIAL

**VISÃO COMPUTACIONAL ASSOCIADA A DEEP LEARNING NA AVALIAÇÃO DA
QUALIDADE DE SEMENTES**

FORTALEZA

2024

LUCIMARA RIBEIRO VENIAL

VISÃO COMPUTACIONAL ASSOCIADA A DEEP LEARNING NA AVALIAÇÃO DA
QUALIDADE DE SEMENTES

Tese apresentada ao Programa de Pós-Graduação em Agronomia/Fitotecnia do Departamento de Fitotecnia da Universidade Federal do Ceará, como parte dos requisitos para obtenção do título de Doutorado em Agronomia/Fitotecnia. Área de concentração: Sementes

Orientador (a): Alek Sandro Dutra

FORTALEZA

2024

Dados Internacionais de Catalogação na Publicação
Universidade Federal do Ceará
Sistema de Bibliotecas

Gerada automaticamente pelo módulo Catalog, mediante os dados fornecidos pelo(a) autor(a)

- V572v Venial, Lucimara Ribeiro.
Visão computacional associada a deep learning na avaliação da qualidade de sementes / Lucimara Ribeiro Venial. – 2024.
143 f. : il. color.
- Tese (doutorado) – Universidade Federal do Ceará, Centro de Ciências Agrárias, Programa de Pós-Graduação em Agronomia (Fitotecnia), Fortaleza, 2024.
Orientação: Prof. Dr. Alek Sandro Dutra.
1. Classificação de Sementes. 2. Análise de Imagens. 3. Reconhecimento de Padrões. 4. Redes Neurais Convolucionais. I. Título.

CDD 630

LUCIMARA RIBEIRO VENIAL

VISÃO COMPUTACIONAL ASSOCIADA A DEEP LEARNING NA AVALIAÇÃO DA
QUALIDADE DE SEMENTES

Tese apresentada ao Programa de Pós-Graduação em Agronomia/Fitotecnia do Departamento de Fitotecnia da Universidade Federal do Ceará, como parte dos requisitos para obtenção do título de Doutorado em Agronomia/Fitotecnia. Área de concentração: Sementes.

Aprovada em 29/08/2024

BANCA EXAMINADORA

Prof. Dr. Alek Sandro Dutra (Orientador)
Universidade Federal do Ceará (UFC)

Prof. Dr. Marcelo de Almeida Guimarães
Universidade Federal do Ceará (UFC)

Dra. Caris dos Santos Viana
Universidade Federal de Viçosa (UFV)

Dra. Patrícia Alvarez Cabanez
Universidade Federal do Espírito Santo (UFES)

Dra. Cíntia Machado de Oliveira Moulin Carias
Instituto Federal do Espírito Santo (IFES)

À minha mãe Solange, por ter sido um anjo em vida. Uma mulher banhada de benevolência e compaixão, mostrou-me somente suas cores belas e me cobriu com seu manto. De ti só recebi amor, amor e mais nada. Não houve um só dia que não pensei na sua ausência. De certo, hoje, lhe faço eterna, pois cada linha desta tese é uma prova do seu amor por mim;

Ao meu pai Djalma, que sempre me disse que eu poderia ser quem eu quisesse, me incentivou a ser dona de mim. Me deu asas, me ensinou a voar, apoia meu voo;

E, a todas as mulheres.

DEDICO

AGRADECIMENTOS

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

A Deus e Nossa Senhora que sempre guiaram meu caminho nessa jornada.

A minha mãe, que nunca aprendeu a ler, e meu pai, que chegou até a quarta série. Agradeço profundamente por todo o esforço e trabalho duro, enfrentando o sol com uma enxada, para que eu pudesse ter oportunidades e chegar até aqui!

Aos meus irmãos queridos Andréia, Júnior e Adriano pelo sentido de ser e estar em família, pelo apoio emocional, por acreditar 100% em mim.

Ao Tio Zé e Tio Tião (*in memoriam*), por terem me amado como filha e a eles, eu amado como pais.

Meus amados sobrinhos: Gabriel, Luiz Henrique, Sofia, Cristal, Miguel, Clara e Alice por serem a alegria da minha vida e me ensinarem o amor mais puro desse mundo.

As minhas cunhadas Carolina e Leidiane, por fazer da nossa família a de vocês. A Keren que sempre me ouviu e me deu apoio.

Meu orientador Alek, por todos esses anos, por ter aguentado meu choro, meu riso, por permitir que o LAS fosse meu lar por tantos anos. Por ter realizado o sonho da minha vida que era estudar em outro país. Por ter sido uma figura paterna em muitos momentos.

Ao Centro Conservazione Biodiversità na Itália que me recebeu, ao professor Gianluigi Bacchetta, em especial ao doutorando Marco Sarigu e todos os amigos do laboratório.

A Raylson (*in memoriam*), minha dupla dinâmica no doutorado, obrigada por todas as conversas até altas horas no LAS que só a gente entendia, pelas saídas no Vinil, os almoços no RU, os sábados no laboratório. Você sempre será o Pink do meu Cérebro.

Minha querida Paloma, o reconhecimento que foi nossa amizade desde o primeiro dia de LAS, acadêmica e pessoal, tanto amor que não cabe no Lattes, me acolheu, me protegeu e me amou.

Meu trio: Dona Regina, Priscila e Ana Kelly, pelo acolhimento no LAS, pela amizade, por ouvir minhas lamúrias todo dia na hora do café, as saídas em Fortaleza, os risos no fim da tarde e as fofocas na hora do almoço. A Luci, por ser extraordinária.

A Caris, não somente a amizade, mas por me acolher na sua família e me fazer sentir parte dela.

A Alana, por absolutamente tudo. Foi a minha primeira IC, realização de um sonho acadêmico e que virou amiga/irmã, jamais teria esse trabalho sem o seu esforço junto comigo, sem os sábados no LAS, os dias de semana até muito tarde, essa vitória é nossa.

A Iza, por ter lido meu recado no Instagram e vim trabalhar comigo, ser minha duplinha, acreditar no meu trabalho, pela pesquisadora impecável, por ser uma grande amiga, pelo apoio incondicional nos últimos tempos.

A Carlos, por todos os anos de amizade, por ser um porto seguro.

A Itálo, pela amizade, por sempre me ajudar mesmo quando não pode. Meus estagiários: Lívia, Hellen, Rian e Giselle.

Meus queridos amigos da pós: Diene, Domingos e Mayume, pela amizade.

Guilherme e Ana, pelo tempo que moramos juntos no apartamento 304 e fomos uma família. Em especial a Ana por tudo que fez e faz por mim, por toda a ajuda na estatística deste trabalho, pela atenção e cuidado, por me amar.

A Lucas e a Matheus pela amizade, por me ajudar nesse trabalho.

Ao Professor Paulo Cortes pela parceria. Ao seu aluno Bruno, que sempre me ajudou quando precisei. E Luan Ícaro e a todos da turma de DPI.

Ao Rodrigo, pelo envio das sementes, tetrazólio usado neste trabalho e amizade.

Minha amada Maria, por ser amiga desde o dia que coloquei os pés em Fortaleza, por acreditar em mim, ouvir o mesmo assunto como se fosse a primeira vez.

A João por toda a amizade quando estávamos na Itália, pela paciência, carinho e cuidado.

“Caro amico Nicola, grazie per il supporto in Italia, la compagnia e per aiutarmi sempre quando ne avevo bisogno.”

Aos queridos amigos do LAS-UFES 2015: Patrícia, Manoel, Nohora e Eloá, jamais chegaria aqui sem vocês.

Amigos: Caio, Rafael e Eloá por tudo que sempre fizeram por mim.

A Eloá, por todos os anos de amizade desde a graduação, por ser a luz da minha vida.

A Paula, pela amizade, companheirismo e carinho sempre.

A Karina e Sara, pelo apoio e amizade, por tudo que fizeram por mim.

A minha atual casa IFES-Campus Montanha, pelo sentimento de pertencimento, aos amigos aqui criados. Em especial a Diretora Cláudia pelo acolhimento no campus e por acreditar no meu trabalho.

A Cíntia por ser tão incrível. Euzilene, pela amizade e por tudo. Caio, pela força nessa reta final. Carol pelo cuidado. Ronaldo por absolutamente tudo. Ana por me ajudar quando mais precisei. Guilherme, por ser incrível comigo sempre.

Os meus alunos: Débora, Maria Eduarda, Clara, Lázaro e Anily.

Aos membros da Banca pelo aceite no convite.

Minha amada Coordenação de Aperfeiçoamento de Pessoal de Nível Superior- Brasil (CAPES) – Código de Financiamento 001. pela concessão da bolsa.

Ao programa Erasmus pela oportunidade do sanduíche.

A UFC, e todos que passaram pelo LAS.

Meus sinceros agradecimentos!

“Eu posso não ter ido para onde eu pretendia ir, mas eu acho que acabei terminando onde eu pretendia estar.” (Douglas Adams)

RESUMO

A semente é o insumo indispensável ao agronegócio, carregando consigo toda a pesquisa e o desenvolvimento do melhoramento genético. Nesse contexto, a utilização de tecnologias é um diferencial, ganhando cada vez mais espaço nesse setor importante da economia. A visão computacional (VC) associada a sistemas de detecção baseados em imagem é de grande importância na determinação da qualidade agroalimentar aliada à confiabilidade, precisão e rapidez, removendo a instabilidade e inconsistência através de intervenção humana. Técnicas de deep learning (aprendizado profundo) associadas ao processamento de imagens digitais podem oferecer inúmeros benefícios, incluindo o desenvolvimento de modelos computacionais para tomada de decisões, possibilitando capacitar grandes e pequenos agricultores, podendo otimizar a tomada de decisão em diferentes etapas da produção. Dessa forma, o objetivo do trabalho é utilizar a VC associada a deep learning na avaliação da qualidade de sementes, visando eliminar a subjetividade e otimizar o processo de análise. O trabalho foi desenvolvido no Laboratório de Análise de Sementes do Departamento de Fitotecnia da Universidade Federal do Ceará (UFC), em colaboração com o Departamento de Computação da UFC e a Universidade de Cagliari, na Itália. Dez lotes de sementes de feijão-caupi (*Vigna unguiculata* (L.) Walp.) foram utilizados. A pesquisa se concentrou na criação de um sistema de visão artificial para análise e classificação das sementes após a realização do teste de tetrazólio. O software tinha como objetivo permitir a análise das imagens, gerar resultados baseados em cores de vermelho carmim claro, vermelho carmim escuro e branco leitoso e classificar automaticamente as sementes de acordo com a literatura vigente quanto ao vigor e a viabilidade. Além disso, foi avaliado as imagens das sementes não tratadas com uso do plugin Seeds Analysis, pertencente ao software ImageJ, que foi validado em solo italiano para discriminar as características morfo-colorimétricas. Foram realizadas a avaliação do vigor das sementes, por meio dos testes tradicionais para caracterização, comparação e validação com as técnicas automatizadas. Todas as metodologias abordadas na tese demonstraram alta eficiência para a análise automatizada, eliminando a subjetividade e aumentando a precisão e a eficiência, reconhecendo padrões escondidos, correlacionar dados, agrupar, classificar e disponibilizar esse conhecimento possibilitando sua atualização de acordo com as inovações no setor sementeiro. O desenvolvimento de um sistema de visão artificial, demonstrou ser uma solução eficaz para superar as limitações dos métodos tradicionais. Além disso, o uso de ferramentas como o plugin Seeds Analysis validaram o potencial dessas tecnologias em identificar e quantificar padrões morfo-colorimétricos em

sementes, contribuindo significativamente para o avanço do setor sementeiro. Assim, este trabalho não apenas destaca a viabilidade e a relevância do uso dessas inovações tecnológicas, mas também abre caminho para futuras atualizações e melhorias no campo da análise de sementes.

Palavras-chave: classificação de sementes; análise de imagens; reconhecimento de padrões; redes neurais convolucionais.

ABSTRACT

The seed is an indispensable input for agribusiness, carrying with it all the research and development of genetic improvement. In this context, the use of technologies is a differentiator, gaining increasing importance in this vital sector of the economy. Computer vision (CV) associated with image-based detection systems is of great importance in determining agri-food quality, providing reliability, precision, and speed, and eliminating instability and inconsistency through human intervention. Deep learning techniques associated with digital image processing can offer numerous benefits, including the development of computational models for decision-making, enabling both large and small farmers to optimize decision-making at various stages of production. Therefore, the objective of this work is to use CV associated with deep learning in seed quality assessment, aiming to eliminate subjectivity and optimize the analysis process. The study was conducted at the Seed Analysis Laboratory of the Department of Crop Science at the Federal University of Ceará (UFC), in collaboration with the UFC's Department of Computer Science and the University of Cagliari in Italy. Ten batches of cowpea seeds (*Vigna unguiculata* (L.) Walp.) were used. The research focused on creating an artificial vision system for analyzing and classifying seeds after performing the tetrazolium test. The software was designed to allow image analysis, generate results based on light crimson red, dark crimson red, and milky white colors, and automatically classify the seeds according to current literature on vigor and viability. In addition, images of untreated seeds were analyzed using the Seeds Analysis plugin, part of the ImageJ software, which was validated in Italy to discriminate morphocolorimetric characteristics. The seed vigor was evaluated using traditional tests for characterization, comparison, and validation with automated techniques. All methodologies addressed in the thesis demonstrated high efficiency for automated analysis, eliminating subjectivity and increasing precision and efficiency by recognizing hidden patterns, correlating data, grouping, classifying, and providing this knowledge for updating according to innovations in the seed sector. The development of an artificial vision system proved to be an effective solution to overcome the limitations of traditional methods. Furthermore, the use of tools such as the Seeds Analysis plugin validated the potential of these technologies in identifying and quantifying morphocolorimetric patterns in seeds, significantly contributing to advances in the seed sector. Thus, this work not only highlights the feasibility and relevance of using these technological innovations but also paves the way for future updates and improvements in the field of seed analysis.

Keywords: seed classification; image analysis; pattern recognition; convolutional neural networks.

LISTA DE FIGURAS

Figura 1 –	Processo de redução do sal de tetrazólio (2,3,5-trifenil cloreto de tetrazólio) para trifenilformazan.....	30
Figura 2 –	Classes de viabilidade e vigor de sementes de feijão.....	31
Figura 3 –	Tipos de arquiteturas de RNAs.....	39
Figura 4 –	Interface do plugin e metodologia para usá-la no software ImageJ.....	39
Figura 5 –	Parâmetros disponíveis no plugin Seeds Analysis.....	39
Figura 6 –	Disposição das sementes para a captura e a variação do número de sementes.....	48
Figura 7 –	Imagem da captura das sementes: a) Parte externa da semente; b) Parte interna da semente.....	49
Figura 8 –	Análise Individual por semente.....	50
Figura 9 –	Classificação final das sementes.....	50
Figura 10 –	Planilha exportada do excel com os dados finais da repetição do lote avaliado gerado pelo Tetrased.....	51
Figura 11 –	Diagrama de sequência.....	52
Figura 12 –	Diagrama de atividade.....	52
Figura 13 –	Diagrama de interface.....	53
Figura 14 –	Diagrama do pré-processamento.....	53
Figura 15 –	Diagrama de classificação.....	53
Figura 16 –	Imagens do Scanner.....	55
Figura 17 –	Imagem original (esquerda) e imagem resultante após a remoção de fundo (direita).....	57

Figura 18 – Extração do Vermelho Carmim Claro: Cor majoritária Carmim escuro (esquerda) e cor majoritária Carmim claro (direita).....	57
Figura 19 – Extração do Vermelho Carmim escuro: Cor majoritária Carmim escuro (esquerda) e cor majoritária Carmim claro (direita).....	58
Figura 20 – Segmentação do branco leitoso.....	59
Figura 21 – Padrões pré-definidos pelo algoritmo MarchingSquares.....	60
Figura 22 – Extração dos buracos da semente.....	60
Figura 23 – 1º Teste - Input de arquivos em outros formatos.....	64
Figura 24 – 3º Teste - Clique no botão prosseguir, sem arquivos enviados.....	65
Figura 25 – 5º Teste - Análise estática de código do Front-End usando ferramenta SonarQube.....	66
Figura 26 – Tela inicial para upload das imagens.....	67
Figura 27 – Tela com os arquivos prontos para serem enviados.....	67
Figura 28 – Processamento das imagens.....	68
Figura 29 – Resultados gerados e dispostos para download.....	68
Figura 30 – Interface inicial do TetraSeed.....	69
Figura 31 – Tela inicial para o upload das imagens.....	70
Figura 32 – Tela com os arquivos prontos para serem enviados.....	71
Figura 33 – Classes de viabilidade e vigor de sementes de feijão-caupi.....	79
Figura 34 – Aquisição das imagens do teste de tetrazólio.....	80
Figura 35 – Resumo do fluxo de operação do TETRASEED.....	82
Figura 36 – Esquema da separação automatizada das sementes seguidas da segmentação da imagem.....	84

Figura 37 – Fases da segmentação e processamento das imagens nos canais RGB...	85
Figura 38 – Gráfico Radar entre as variáveis IVE, G, PC, MSRA, MSPA, CRA, CPA e VIGOR com as classes obtidas pelo teste de tetrazólio para os dez lotes.....	91
Figura 39 – Gráfico para Acurácia e Perda em função das épocas de treinamento...	94
Figura 40 – Matriz de confusão para as classificações preditas pela rede em função das classificações visuais.....	97
Figura 41 – Classes de viabilidade e vigor de sementes de feijão-caupi.....	98
Figura 42 – Balanço do número de imagens por classe.....	99
Figura 43 – Distribuição de frequências das classes.....	102
Figura 44 – Gráfico Radar entre as variáveis Vigor, Viabilidade e Classes do TETRASEED.....	104
Figura 45 – Lotes de feijão utilizados nos experimentos.....	109
Figura 46 – Biometria de semente por paquímetro digital.....	111
Figura 47 – Aquisição das imagens digitais.....	112
Figura 48 – Coeficientes de correlação de Spearman estimados entre os parâmetros índice de velocidade de emergência (IVE), massa seca de parte aérea (MSPA), massa seca de raiz (MSRA), comprimento de parte aérea (CPA), comprimento de raiz (CRA), peso de mil sementes (PMS), condutividade elétrica (CE), germinação (G), primeira contagem (PC) e parâmetros obtidos na análise de imagem.....	126
Figura 49 – Análise de componentes principais (PCA) (A e B) dos dez lotes; A) PCA para os parâmetros físicos e B) PCA para os parâmetros de cor RGB; PC1 = componente principal 1; PC2 = componente principal.....	128

Figura 50 – (A) Representação gráfica da análise discriminante (B) histograma dos resíduos padronizados, (C) gráfico de dispersão dos resíduos padronizados testados com o teste de Levene (F) e (D) gráfico de probabilidade normal (P-P) testado com o teste de Kolmogorov-Smirnov (K-S)..... 131

LISTA DE TABELAS

Tabela 1 –	Características morfológicas obtidas no plugin Seeds Analysis no ImageJ.....	41
Tabela 2 –	Recursos de cores dos espaços de cores RGB e HSV obtidas pelo plugin Seeds Analysis no ImageJ.....	43
Tabela 3 –	Recursos de textura da imagem em tons de cinza obtidas pelo plugin Seeds Analysis no ImageJ.....	44
Tabela 4 –	Resultados de primeira contagem - PC; plântulas anormais- PA; comprimento de parte aérea - CPA; comprimento de raiz - CRA; matéria seca de parte aérea - MSPA; matéria seca de raiz - MSRA; índice de velocidade de emergência - IVE obtidos em 10 lotes de feijão-caupi.....	88
Tabela 5 –	Métricas de desempenho para medir o desempenho do modelo CNN (Inception-ResNet v2).....	100
Tabela 6 –	Parâmetros de características morfológicas avaliados com o plugin Seeds Analysis.....	115
Tabela 7 –	Resultados de primeira contagem - PC; plântulas anormais- PA; comprimento de parte aérea - CPA; comprimento de raiz - CRA; matéria seca de parte aérea - MSPA; matéria seca de raiz - MSRA; condutividade elétrica - CE; peso de mil sementes - PMS; índice de velocidade de emergência - IVE obtidos em 10 lotes de feijão.....	116
Tabela 8 –	Resultados de comprimento - C; largura- L obtidos em 10 lotes de sementes de feijão.....	110
Tabela 9 –	Médias dos parâmetros de características morfológicas para sementes avaliados com o plugin Seeds Analysis nos 10 lotes de feijão.....	121

Tabela 10 – Resultados de média de vermelho - MeanRed; média de verde - MeanGreen; média de azul - MeanBlue e média de RGB - MeanRGB avaliados com o plugin Seeds Analysis nos 10 lotes de sementes de feijão.....	124
Tabela 11 – Ranking das variáveis morfocolorimétricas discriminantes mais importantes selecionadas e utilizadas pela LDA.....	130
Tabela 12 – Porcentagem de classificação correta e validação cruzada para os 10 lotes.....	132
Tabela 13 – Porcentagem de classificação correta e validação cruzada para os 10 lotes.....	133

SUMÁRIO

1	INTRODUÇÃO GERAL	24
2	OBJETIVOS	26
2.1	Objetivo geral	26
2.2	Objetivos específicos	26
3	CAPÍTULO 1 – VISÃO COMPUTACIONAL ASSOCIADA A DEEP LEARNING NA AVALIAÇÃO DA QUALIDADE DE SEMENTES - REVISÃO DE LITERATURA	27
3.1	Qualidade da semente	27
3.2	Testes tradicionais	27
3.3	Teste de tetrazólio	28
3.4	Histórico do teste de tetrazólio	28
3.5	Princípios do teste	29
3.6	Interpretação dos resultados	30
3.7	Teste de tetrazólio no feijão	32
3.8	Visão computacional	33
3.9	Redes neurais	35
3.10	Software Image J	38
4	CAPÍTULO 2 – CONSTRUÇÃO DO SOFTWARE TETRASEED	45
4.1	Introdução	45
4.2	Criação do software TETRASEED	46
4.2.2	<i>Definições de uso</i>	46
4.2.2.1	<i>Requisitos básicos</i>	46
4.2.2.2	<i>Restrições do sistema</i>	47
4.2.3	<i>Requisitos do sistema</i>	49
4.2.3.1	<i>Requisitos funcionais</i>	49

4.2.3.2	<i>Requisitos não funcionais</i>	51
4.2.4	<i>Estrutura do software</i>	52
4.2.4.1	<i>Diagrama de sequência</i>	52
4.2.4.2	<i>Diagrama de atividades</i>	52
4.2.4.3	<i>Interface</i>	53
4.2.4.4	<i>Pré-processamento</i>	53
4.2.4.5	<i>Classificação</i>	53
4.2.5	<i>Ambiente de desenvolvimento</i>	53
4.2.6	<i>Implementação</i>	54
4.2.6.1	<i>Interface</i>	55
4.2.6.2	<i>Aquisição</i>	55
4.2.7	<i>Pré-processamento</i>	56
4.2.7.1	<i>Alinhamento da malha</i>	56
4.2.8	<i>Corte da malha</i>	56
4.2.8.1	<i>Com grid</i>	56
4.2.8.2	<i>Sem grid</i>	56
4.2.9	<i>Processamento - TETRASEED</i>	57
4.2.9.1	<i>Remoção do fundo</i>	57
4.2.9.2	<i>Extração da cor vermelha</i>	57
4.2.9.3	<i>Extração do branco leitoso</i>	59
4.2.9.4	<i>Identificação de buracos na semente</i>	59
4.2.10	<i>Classificação</i>	61
4.2.10.1	<i>Classificação das sementes</i>	61
4.2.11	<i>Descrição do algoritmo CNN</i>	61
4.2.11.1	<i>Hiperparâmetros</i>	61

4.2.12	Detecção e classificação do embrião	63
4.2.13	Testes	64
4.2.13.1	<i>1º Teste - Input de arquivos em outros formatos</i>	64
4.2.13.2	<i>2º Teste - Redimensionamento de janela</i>	64
4.2.13.3	<i>3º Teste - Clique no botão prosseguir, sem arquivos</i>	65
4.2.13.4	<i>4º Teste - Alinhamento de itens</i>	65
4.2.13.5	<i>5º Teste - Análise estática de código do Front-End usando ferramenta SonarQube</i>	65
4.2.14	TETRASEED	67
5	CAPÍTULO 3 - TETRASEED: SISTEMA DE VISÃO ARTIFICIAL PARA ANÁLISE E CLASSIFICAÇÃO DE SEMENTES DE FEIJÃO CAUPI (<i>Vigna unguiculata</i> (L.) Walp.)	72
5.1	Introdução	74
5.2	Materiais e métodos	76
5.2.1	<i>Etapa I: Análise de qualidade dos lotes de sementes com testes tradicionais</i>	76
5.2.1.1	<i>Teste de germinação e primeira contagem</i>	76
5.2.1.2	<i>Comprimento de plântulas</i>	76
5.2.1.3	<i>Massa seca de plântulas</i>	77
5.2.1.4	<i>Teste de emergência</i>	77
5.2.2	<i>Teste de tetrazólio</i>	77
5.2.3	<i>Aquisição das imagens</i>	79
5.2.3.1	<i>Visão geral do TETRASEED</i>	80
5.2.3.2	<i>Processamento</i>	84
5.2.3.3	<i>Classificação</i>	85
5.3	Resultados e discussão	88
5.3.1	<i>Testes tradicionais</i>	88

5.3.2	<i>Teste de tetrazólio e aquisição de imagens</i>	90
5.3.3	TETRASEED	93
5.4	Conclusão	104
6	CAPÍTULO 4 - CARACTERIZAÇÃO MOFORCOLORÍMETRICA DE SEMENTES FEIJÃO CAUPI (<i>Vigna unguiculata (L.) Walp.</i>) POR MEIO DO USO DE ANÁLISE DE IMAGEM	105
6.1	Introdução	107
6.2	Materiais e métodos	108
6.2.1	<i>Etapa I: Análise de qualidade dos lotes de sementes com testes tradicionais</i>	109
6.2.1.1	<i>Teste de germinação e primeira contagem</i>	109
6.2.1.2	<i>Comprimento de plântulas</i>	109
6.2.1.3	<i>Massa seca de plântulas</i>	109
6.2.1.4	<i>Teste de emergência</i>	110
6.2.1.5	<i>Peso de mil sementes</i>	110
6.2.2	<i>Biometria de sementes manual</i>	111
6.2.3	<i>Etapa II: morfo-colorimetria por análise de imagens</i>	111
6.2.3.1	<i>Aquisição de imagem</i>	111
6.2.3.2	<i>Plugin para extração de características das imagens</i>	113
6.3	Resultados e discussão	115
6.3.1	<i>Etapa I: Análise de qualidade dos lotes de sementes com testes tradicionais</i>	115
6.3.2	<i>Etapa II: morfo-colorimetria por análise de imagens</i>	120
6.4	Conclusão	133
	CONSIDERAÇÕES FINAIS;	135
	REFERÊNCIAS	136

1 INTRODUÇÃO GERAL

A semente é o insumo fundamental para o sucesso do agronegócio, servindo como o ponto de partida para a produção agrícola e carregando consigo os avanços do melhoramento genético. A qualidade das sementes desempenha um papel crucial no desenvolvimento das culturas e, por conseguinte, na produtividade agrícola, visto que ela é essencial para o início de um novo ciclo de cultivo (Marcos Filho, 2016). Portanto, sementes de alta qualidade é uma das principais condições necessárias para se alcançar uma população com alto padrão em campo, visando com isso obter o máximo do potencial produtivo de uma cultura (Javorski *et al.*, 2018). Dessa forma, os agricultores cada vez mais buscam sementes de qualidade, visando um pleno estabelecimento em campo, desse modo as empresas devem oferecer sementes de alta qualidade para assegurar sua posição competitiva nos mercados agrícolas (Grasso *et al.*, 2018).

Diferentes testes de qualidade são realizados para avaliar o potencial das sementes em campo objetivando aferir sobre o vigor e ranquear lotes (Marcos Filho *et al.*, 2015). De acordo com McDonald (1975) os testes são divididos entre físicos, fisiológicos, bioquímicos, e de resistência. Os métodos de avaliação são importantes para a avaliação da qualidade da semente, mas são realizados de modo manual e demorado, dado que os processos demandam muito tempo e precisão para que não haja erros de análise e de interpretação nos dados obtidos (Abud *et al.*, 2018; Mondo e Cicero, 2005).

Se os resultados estiverem equivocados, a avaliação da qualidade das sementes pode ser alterada, prejudicando diretamente o produtor (Mahajan *et al.*, 2018). Se a qualidade exigida pelo mercado não for atendida esse lote pode não alcançar os parâmetros adequados de qualidade para o mercado, resultando em prejuízos significativos (Grasso *et al.*, 2018). Uma alternativa possível, devido ao avanço da tecnologia, é executar as avaliações com o método de análise de imagem (Hoffmaster *et al.*, 2003). Nesse cenário, a integração de tecnologias avançadas, como a VC e o deep learning, tem se mostrado um diferencial significativo para a indústria agrícola (Basso *et al.*, 2016).

A VC é um campo dentro da inteligência artificial que visa permitir que os sistemas de computadores interpretem e compreendam informações visuais da mesma forma que os seres

humanos, envolvendo o desenvolvimento de algoritmos e sistemas que podem processar, analisar e entender imagens. Estudos de Zhao *et al.*, (2022) destacam que a aplicação de técnicas avançadas como a VC, é importante nos testes de sementes, pois essas técnicas permitem a classificação automática de sementes, aumentando a precisão e eficiência dos testes de pureza e germinação. Portanto, a incorporação de técnicas de VC vem possibilitando novas formas de análise e interpretação de dados visuais em uma ampla gama de aplicações, incluindo a avaliação da qualidade das sementes.

O acelerado crescimento no campo da VC tem, principalmente com o avanço das técnicas de deep learning (aprendizado profundo), grandes avanços para resolver problemas complexos de processamento de imagens (Pratap *et al.*, 2024). O uso de técnicas de deep learning na VC para a avaliação de sementes permite a automação do processo de inspeção, melhorando significativamente a precisão e a velocidade da análise. O deep learning é uma subárea dentro do aprendizado de máquinas que utiliza redes neurais profundas (RNAs), com um potencial para análise de imagens digitais (Wang *et al.*, 2021).

Dentre as arquiteturas utilizadas, as redes neurais convolucionais (CNNs) tem uma grande capacidade de aprender representações hierárquicas dos dados visuais, permitindo a identificação e a classificação precisa de padrões complexos nas imagens de sementes (Lecun *et al.*, 2015). Essa capacidade é particularmente útil para desenvolver modelos computacionais que suportam a tomada de decisões informadas na seleção e avaliação das sementes.

Sistemas de automação da avaliação de sementes por meio de VC e deep learning, aumenta a eficiência e reduz a dependência de processos manuais, podendo processar grandes volumes de imagens de sementes rapidamente, identificando e classificando cada semente com alta precisão, o que é essencial para operações em larga escala (Mnih *et al.*, 2015). Portanto, a combinação dessas técnicas oferece uma abordagem inovadora e eficiente para a avaliação da qualidade das sementes, melhorando a precisão e a eficiência do processo de avaliação, contribuindo para a inovação e a sustentabilidade na agricultura moderna. Dessa forma, a tese foi separada em capítulos, referencial teórico, descrição do software, artigo 1 e artigo 2.

2 OBJETIVOS

2.1 Objetivo Geral

Utilizar a visão computacional associada a deep learning na avaliação da qualidade de sementes, visando eliminar a subjetividade e otimizar o processo de análise.

2.2 Objetivos específicos

- Conduzir testes de laboratório para avaliar o vigor das sementes de forma tradicional e criar uma base de comparação com os resultados obtidos pelo software;
- Desenvolver um banco de dados de imagens robusto;
- Validar a eficácia dos modelos de deep learning através da comparação com os métodos tradicionais;
- Implementar algoritmos de deep learning para análise de sementes;
- Desenvolver e treinar redes neurais convolucionais (CNNs) para detectar e classificar características visuais das sementes;
- Utilizar os algoritmos para identificar padrões escondidos nos dados de imagem e correlacioná-los com a qualidade das sementes;
- Agrupar e classificar as sementes com base nos padrões identificados, fornecendo *insights* valiosos para a tomada de decisão;
- Integrar o sistema de visão computacional em um software de avaliação;
- Implementar um sistema de aprendizado contínuo para que o software possa ser atualizado com novas informações e inovações no setor sementeiro.

3 CAPÍTULO 1 – VISÃO COMPUTACIONAL ASSOCIADA A DEEP LEARNING NA AVALIAÇÃO DA QUALIDADE DE SEMENTES - REVISÃO DE LITERATURA

3.1 Qualidade da semente

A semente é ponto inicial mais crucial nos sistemas agrícolas, visto que ela é essencial para o início de um novo ciclo de cultivo (Marcos Filho, 2016). Segundo Vizcayno *et al.*, (2014), a semente se difere do grão por apresentar todas as estruturas viáveis para a geração de uma nova planta e por possuir qualidade suficiente para seu estabelecimento em campo.

A qualidade de uma semente compreende um conjunto de características, atributos, índices ou componentes que determinam seu estado para semeadura, bem como seu desempenho em condições de campo Brasil, (2009). A alta qualidade de sementes é importantíssima principalmente para uma germinação uniforme que é necessária para garantir o ideal estande das plantas. Os principais critérios para definir a qualidade da semente são pureza, germinação, umidade e ausência de doenças e insetos (Kapadia; Sasidharan; Patil, 2017).

Testes de vigor são utilizados para avaliar o vigor das sementes (Zucareli *et al.*, 2011). A definição de vigor é complexa e explorada por muitos autores, de acordo com Krzyzanowski e França-Neto (2001), ele pode ser definido por uma junção de atributos que irão garantir a semente potencial realizar uma boa germinação, emergir e gerar plântulas normais sob diferentes condições edafoclimáticas. A avaliação do vigor possibilita diferenciar e detectar possíveis diferenças na qualidade fisiológica entre lotes de lotes de semente, porém por ser realizado em condições controladas podem não representar o potencial do lote em condições de campo (Marcos Filho, 2009).

3.2 Testes Tradicionais

Os testes são utilizados para avaliar a qualidade dos lotes e ranquear os lotes. O teste tradicional mais utilizado é o teste de germinação no qual é possível observar com o crescimento inicial a semente e o desenvolvimento de estruturas importantes presentes no embrião, tornando possível uma previsão para propensão de desenvolvimento em uma condição ideal de campo (BRASIL, 2009). Existem outros testes utilizados para avaliar o

vigor e a qualidade de sementes, permitindo uma melhor seleção dos lotes que serão levados para a comercialização (Ullmann *et al.*, 2015). McDonald (1975) propõe uma divisão entre físicos, fisiológicos, bioquímicos, e de resistência.

Dessa forma, os testes mais utilizados para avaliação de vigor são teor de água, germinação, índice de velocidade de germinação, primeira contagem, porcentagem de emergência, índice de velocidade de emergência, condutividade elétrica, envelhecimento acelerado, tetrazólio, e condutividade elétrica, e são nomeadas com testes tradicionais. Esses testes são reconhecidos e recomendados pelas ISTA (International Seed Testing Association), AOSA (Association of Official Seed Analysts), e ABRATES. (Associação Brasileira de Tecnologia de Sementes) (Rodrigues *et al.*, 2020), as principais associações e instituições de análise de sementes nacionais e internacionais.

3.3 Teste de tetrazólio

Entre os métodos de avaliação de qualidade de sementes, destaca-se o teste de tetrazólio. Este teste é rápido, preciso e de baixo custo, permitindo a identificação de fatores como deterioração e danos mecânicos, que podem interferir na qualidade das sementes. Além disso, ele fornece resultados sobre a viabilidade e o vigor das sementes. (Fogaça *et al.*, 2006; Ribeiro *et al.*, 2010; Rodrigues *et al.*, 2015).

3.4 Histórico do Teste de tetrazólio

Historicamente os primeiros testes aconteceram por Turina utilizando bio-corantes para o estudo da viabilidade de sementes, utilizando sais de telúrio e de selênio nas células das sementes e por Neljubow em 1925 com a utilização de índigo carmine (França Neto; Krzyzanowski, *et al.*, 1998). Na década de 30, o pesquisador Hasegawat aperfeiçoou a aplicação de sais de telúrio e de selênio para a coloração de embriões de sementes, em seguida o método do selênio aprimorado por F.E. Eidmann, após conversa com Hasegawat (França Neto *et al.*, 1988).

Houve o aperfeiçoamento da técnica de selênio por Hasegawa e Eidmann na década de 1940, Georg Lakon, trabalhando em Hohenheim, Alemanha, desenvolveu um novo método topográfico utilizando selênio para determinar a viabilidade das sementes. No entanto, ao descobrir a toxicidade desse elemento, Lakon buscou corantes similares que não fossem

tóxicos. Após testar diversos compostos de tetrazólio em várias sementes, ele concluiu que o 2,3,5-trifenil cloreto de tetrazólio era o mais apropriado (França Neto; Krzyzanowski, 2018).

Na década de 1960, Delouche, Still, Raspet e Leinhard, da Universidade Estadual do Mississippi, publicaram o primeiro manual sobre o teste de tetrazólio. Em 1970, o teste foi aceito pela AOSA (Associação Oficial dos Analistas de Sementes) com a publicação do Manual do Teste de Tetrazólio. Em 1983, a AOSA lançou o Manual de Testes de Vigor, que incluía a metodologia do teste para sementes de soja, algodão, milho e trigo. (França Neto, Krzyzanowski, 2018).

Para a determinação do vigor das sementes, métodos de coloração foram realizados na década de 1920 na Rússia por Neljubow, que formulou cinco classes de padrões de coloração com base na intensidade da coloração dos tecidos das sementes. Já na década de 1930, Eidmann propôs três classes de avaliação: sementes plenamente germináveis, sementes germináveis fracas e sementes mortas, dependendo da intensidade e extensão da coloração. (França Neto *et al.*, 1988).

Em 1950, Lakon propôs um método para classificar as sementes em alto e baixo vigor com base na localização e extensão da coloração, assim como na aparência e textura dos tecidos das sementes. Em 1956, Moore e Smith aprimoraram esses princípios, classificando as sementes em três classes: “A” (vigorosas), “B” (viáveis não vigorosas) e “C” (não viáveis). (França Neto; Krzyzanowski, 2018; França Neto *et al.*, 1988).

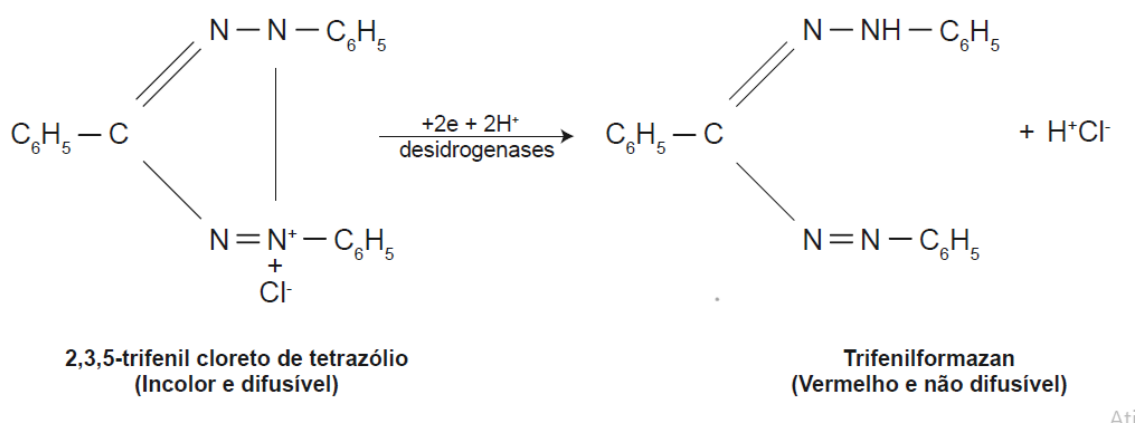
A nível nacional, o teste de tetrazólio foi introduzido por pesquisadores que receberam treinamento na Universidade Estadual do Mississippi, através do Convênio AGIPLAN/Ministério da Agricultura (França Neto *et al.*, 1988). Em 1999, a ABRATES (Associação Brasileira de Tecnologia de Sementes) publicou o manual “Vigor de Sementes: Conceitos e Testes”, detalhando os procedimentos para o teste de tetrazólio objetivando a determinação do vigor em sementes (França-Neto; Krzyzanowski e Da Costa, 1998).

3.5 Princípios do teste

O princípio do teste de tetrazólio baseia-se na atividade das enzimas desidrogenase, responsáveis pelas reações respiratórias das mitocôndrias durante a glicólise e o ciclo de Krebs (ciclo do ácido cítrico). Dessa forma, o teste determina indiretamente a atividade

respiratória das sementes (França-Neto; Krzyzanowski e Da Costa,1998). As sementes quando são imersas na solução de tetrazólio (incolor) a enzima malato desidrogenase reduz o sal tetrazólio (2,3,5–trifenil cloreto de tetrazólio ou TCT) nos tecidos vivos onde ocorre uma reação de redução resultando em um composto vermelho, estável e não-difusível, o trifenilformazan (França Neto; Krzyzanowski, 2018).

Figura 1 – Processo de redução do sal de tetrazólio (2,3,5-trifenil cloreto de tetrazólio) para trifenilformazan.



Fonte: França Neto; Krzyzanowski (2018).

Dessa forma, a redução do TCT indica, através da coloração, a atividade respiratória do tecido e a viabilidade celular. A coloração vermelho carmim clara indica tecido vigoroso, enquanto um vermelho mais intenso sugere deterioração devido à maior difusão do TCT. Por outro lado, os tecidos não viáveis não reagem e, conseqüentemente, não apresentam coloração (AOSA, 1983; França-Neto; Krzyzanowski e Da Costa,1998).

3.6 Interpretação dos resultados

Cada espécie possui uma metodologia específica para a coloração e avaliação da viabilidade e vigor das sementes, baseada na intensidade da coloração, localização e extensão dos danos. É necessária uma análise interna e externa das sementes (França Neto, 1999). Deve-se dar maior atenção ao eixo hipocótilo-radícula e à área de ligação dos cotilédones ao eixo embrionário, pois danos nessas regiões são mais prejudiciais do que em outras áreas menos vulneráveis (França-Neto; Krzyzanowski e Da Costa,1998).

Para cada espécie, são definidos critérios específicos para a formação de classes, e o número de classes pode variar conforme estabelecido na literatura. Para sementes de algodão são adotadas 8 classes, para amendoim são 3 classes e para feijão são 7 classes (França Neto, 1999).

Além das sementes de grandes culturas como feijão, soja, milho, algodão, entre outras o teste de tetrazólio se mostrou eficiente para avaliar a viabilidade de diversas espécies como *Cucurbita pepo* (Barros *et al.*, 2005), *Peltophorum dubium* (Sprengel) Taubert (Carvalho *et al.*, 2005), *Gleditschia amorphoides* Taub. (Fogaça *et al.*, 2006), *Panicum maximum* Jacq (Dias; Alves, 2008), *Jatropha curcas* L. (Pinto *et al.*, 2009), *Clitorea ternatea* L. (Deminicis *et al.*, 2009), *Sebastiania commersoniana* (Baill.) Smith & Downs (Santos; De Paula, 2009), *Sesamum indicum* L. (Jesus *et al.*, 2015), *Vigna unguiculata* (L.) Walp. (Rodrigues *et al.*, 2015).

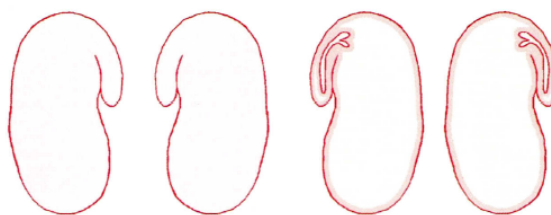
3.7 Teste de tetrazólio no feijão

As sementes são seccionadas longitudinalmente ao longo do eixo embrionário e seus tegumentos são removidos para que possa fazer a interpretação dos dados. Em seguida, as sementes são analisadas individualmente, tanto internamente quanto externamente, quanto aos danos (mecânicos, por insetos e por umidade), coloração e consistência dos tecidos.

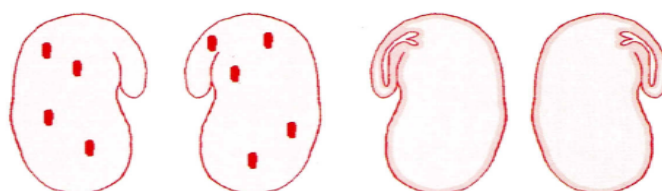
Para a caracterização dos níveis de viabilidade no feijão, utilizam-se 7 classes. As classes de 1 a 3 correspondem a sementes viáveis e vigorosas, as classes 4 e 5 a sementes viáveis, porém não vigorosas, e as classes 6 e 7 a sementes não viáveis.

Figura 2 – Classes de viabilidade e vigor de sementes de feijão.

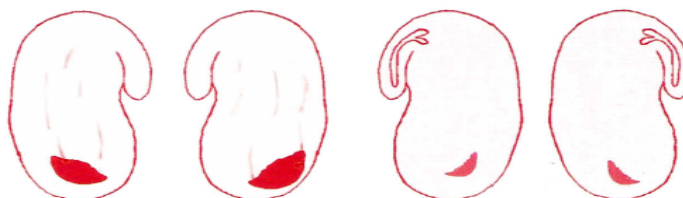
Classe 1- Mais alto vigor



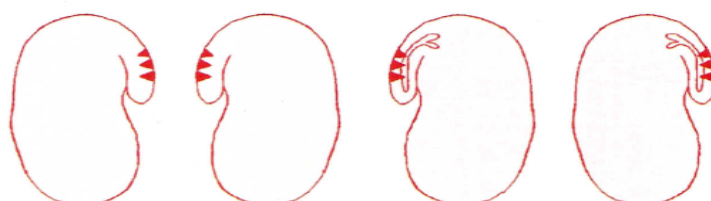
Classe 2 - Alto vigor



Classe 3 - Médio vigor



Classe 4 - Baixo vigor



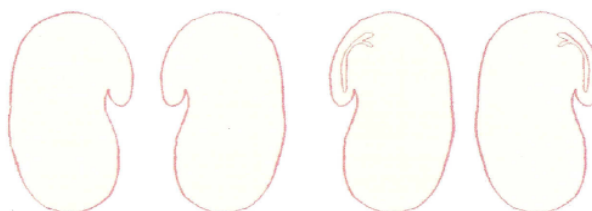
Classe 5 - Muito baixo vigor



Classe 6 - Não viável



Classe 7 - Morta



Fonte: França-Neto; Krzyzanowski e Da Costa (1998)

Os resultados da interpretação irão variar de acordo com a classificação das sementes e sua percentagem final entre classes. A viabilidade é expressa pelo somatório das sementes germináveis das classes de 1 a 5, e o potencial de vigor expresso pelo somatório de 1 a 3 (França Neto, 1999).

3.8 Visão computacional

A VC é um campo da inteligência artificial (IA) que se concentra em como os computadores podem ser treinados para interpretar e compreender o mundo visual da mesma

forma que os seres humanos. Envolve o desenvolvimento de algoritmos e técnicas que permitem que máquinas analisem imagens e vídeos, extraindo informações úteis a partir desses dados visuais. E para isso a VC integra a IA ao processamento digital de imagens (Li; Huang, 2024). A IA surgiu com o objetivo de criar máquinas capazes de pensar, aprender e tomar decisões para resolver problemas, simulando a inteligência humana (Lobo, 2018).

Uma imagem digital é definida como uma representação bidimensional composta por pixels (do inglês, "picture elements"), que são as menores unidades de uma imagem. Esses pixels formam uma matriz binária composta por 0s e 1s, que são essenciais para a codificação da imagem (Jain, 1989; Gonzalez *et al.*, 2004).

Cada imagem contém numerosos pixels, e a resolução da imagem aumenta com o número de pixels. A informação de cor em cada pixel permite a criação de imagens coloridas, possibilitando a representação detalhada e precisa (Gunturk; Li, 2018; Szeliski, 2022). A combinação dessas informações de pixel é fundamental para o processamento e análise de imagens digitais em sistemas de VC (Szeliski,2022)

Diversos sistemas de cores são utilizados na VC, sendo o sistema RGB (Red, Green, Blue) um dos mais comuns. Este sistema possibilita a formação de imagens coloridas com até 16.777.216 combinações diferentes de tons de Queiroz, J. E. R., e Gomes, H. M. (2006). Cada pixel, como mencionado anteriormente, representa uma variação de cor que pode ser descrita em uma escala de 0 a 1, onde valores próximos de zero indicam baixa intensidade de cor e valores próximos de um indicam alta intensidade de cor (Gonzalez, 2009). A combinação das cores RGB em diferentes proporções permite a criação de uma ampla gama de cores e tons.

Na análise de imagens, a inteligência artificial realiza suas funções ao interpretar os pixels e suas informações de cor, destacando a importância de utilizar um sistema padrão de cores e garantir imagens de alta resolução para obter resultados precisos e eficientes Wang *et al.*, (2021). Realizada com base na obtenção de informações a respeito de objetos que estão presentes em uma imagem digital, considerando características do objeto em estudo como a cor, a textura e outras informações (Cícero *et al.*, 1998).

De acordo com Gonzalez, Woods e Eddins (2004), o processamento de imagens envolve cinco etapas principais: obtenção das imagens, pré-processamento, segmentação, reconhecimento e interpretação. Na primeira etapa, é necessário um equipamento capaz de

capturar e digitalizar a imagem. No pré-processamento, a imagem digitalizada é aprimorada para melhorar a eficácia nas etapas subsequentes, incluindo a remoção de ruídos e o realce dos contrastes. Durante a segmentação, o objeto de interesse é separado do fundo da imagem, criando uma camada binária. Na etapa de reconhecimento, o objeto é rotulado com base nas informações fornecidas por seu descritor. Finalmente, na interpretação, é atribuído um significado ao conjunto de objetos identificados.

Com o avanço da tecnologia e sua crescente aplicação no setor agrícola, a VC emergiu como uma solução para superar as limitações existentes. Esse método oferece resultados mais rápidos, precisos e não destrutivos na avaliação da qualidade das sementes, utilizando a análise de imagens para classificar espécies de plantas vasculares com base na morfologia e nas características colorimétricas das sementes (Vale *et al.*, 2020; Zhao, Haque e Wang, 2022).

A análise é realizada de forma semi-automatizada ou totalmente automatizada, empregando algoritmos e softwares especializados em análise de imagens. Esse processo reduz o tempo necessário para as análises, acelera a avaliação, minimiza distorções causadas por luz natural e microscópios, diminui as taxas de erro e aumenta a eficiência, resultando em análises mais rápidas e precisas (Marcos Filho, 2015; Sharma *et al.*, 2020).

3.9 Redes neurais

As RNAs são um campo dentro da IA, são algoritmos computacionais inspirados no funcionamento do cérebro humano, são confeccionadas por unidades de processamento simples chamadas neurônios artificiais, que estão interconectados em uma estrutura em camadas. São frequentemente utilizadas para resolver problemas complexos, nos quais o comportamento das variáveis não é precisamente conhecido (Fleck *et al.*, 2016). Sua propriedade mais importante está relacionada com sua capacidade de aprender por meio de exemplos, generalizar e melhorar seu desempenho, criando um modelo não linear, o que torna sua aplicação na análise espacial altamente eficiente (Spörl, Castro e Luchiari, 2011).

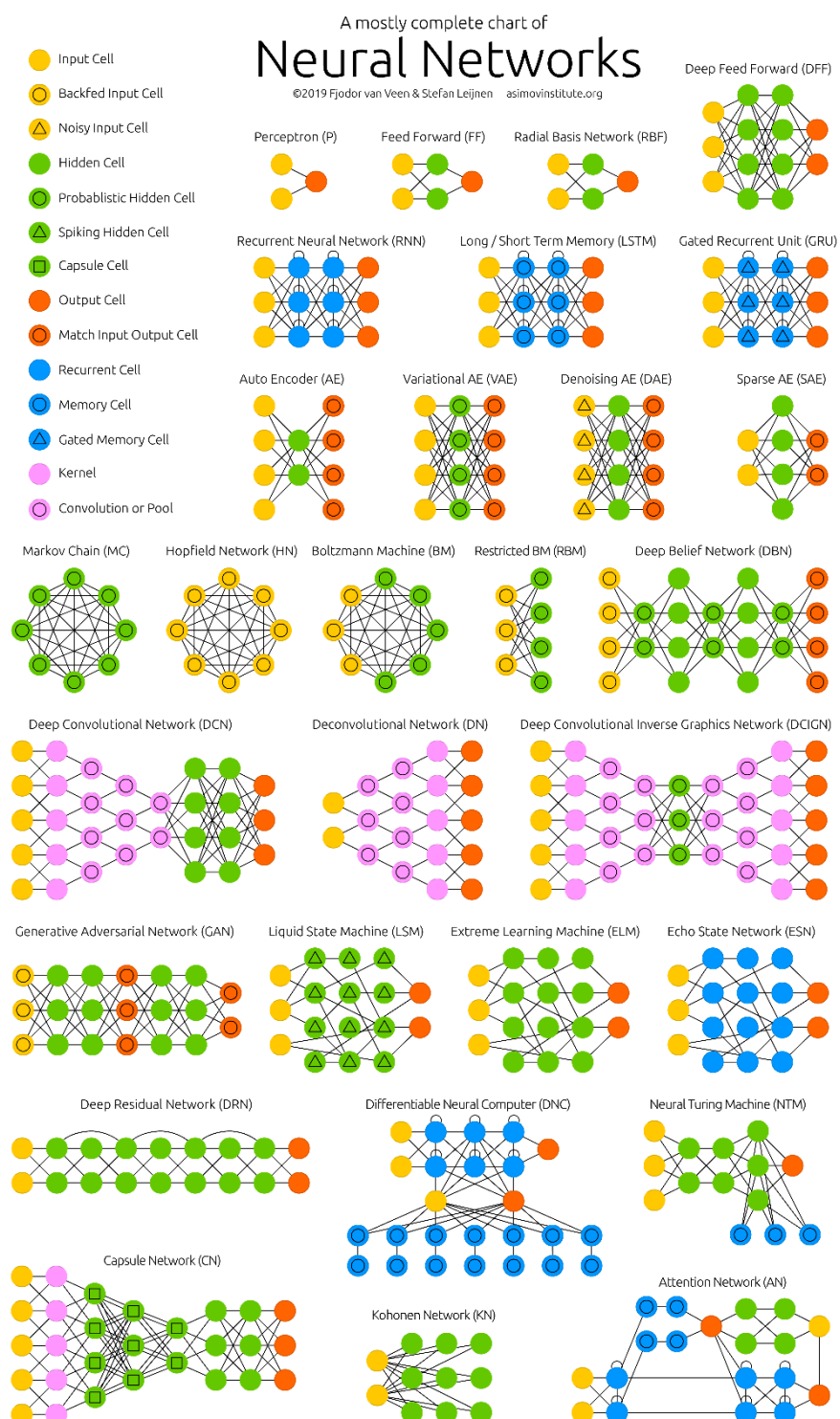
O uso de RNAs foi introduzida na década de 1940, quando Mcculloch e Pitts (1943) propuseram o primeiro modelo de neurônio artificial (Mcculloch; Pitts 1943), a partir desses anos as RNAs têm sido amplamente estudadas e aplicadas em diversas áreas, como reconhecimento de padrões, previsão de séries temporais, e otimização de processos

complexos. Essas redes aprendem a partir de exemplos, ajustando seus pesos sinápticos para minimizar o erro nas previsões, tornando-se extremamente eficazes em problemas onde os dados são ruidosos ou não lineares.

Uma RNA padrão é composta por três tipos principais de camadas: camada de entrada, camadas ocultas e camada de saída. A camada de entrada recebe os dados brutos. Cada neurônio nesta camada representa uma característica dos dados, como cada pixel de uma imagem ou cada palavra de um texto (Lecun *et al.*, 1998). Na próxima etapa os dados são transferidos para as camadas ocultas, onde acontece o processamento por neurônios que aplicam funções de ativação não lineares para extrair características mais complexas e abstratas (Goodfellow, 2016).

Existem diversos tipos de arquiteturas de RNAs (Figura 3), que estarão ligadas ao objetivo e a função da rede, ou seja, os dados no qual se deseja avaliar irão determinar escolha da arquitetura assim como o sucesso na análise de dados.

Figura 3 – Tipos de arquiteturas de RNAs.



Fonte: Van Veen, 2017.

Entre as diversas arquiteturas de RNAs, as redes neurais convolucionais (Convolutional Neural Networks, CNNs) se destacam por sua eficácia em tarefas de reconhecimento de padrões em imagens e vídeos. São projetadas especificamente para processar imagens, utilizando camadas convolucionais para aplicar filtros sobre as entradas e extrair características espaciais e temporais de forma hierárquica (Rezaee *et al.*, 2018). Cada filtro na camada convolucional detecta características específicas, como bordas e texturas, em regiões locais da imagem, enquanto camadas posteriores combinam essas características para formar representações mais complexas e abstratas (Stahl; Weimann, 2022). Essa função torna possível que as CNNs realizem tarefas de reconhecimento de padrões e classificação com alta precisão e eficiência, tornando-as particularmente eficazes em aplicações de VC e processamento de imagens (Liu; Abd-Elrahman, 2018).

3.10 Software Image J

Lançado em 1997 o software ImageJ é um programa de domínio público projetado para o processamento de imagens. Possui uma arquitetura aberta, ele permite que os usuários adicionem plugins personalizadas utilizando o editor integrado ao software (Schneider; Rasband e Eliceiri, 2012). Dessa forma, o software permite a criação e implementação de macros e plugins.

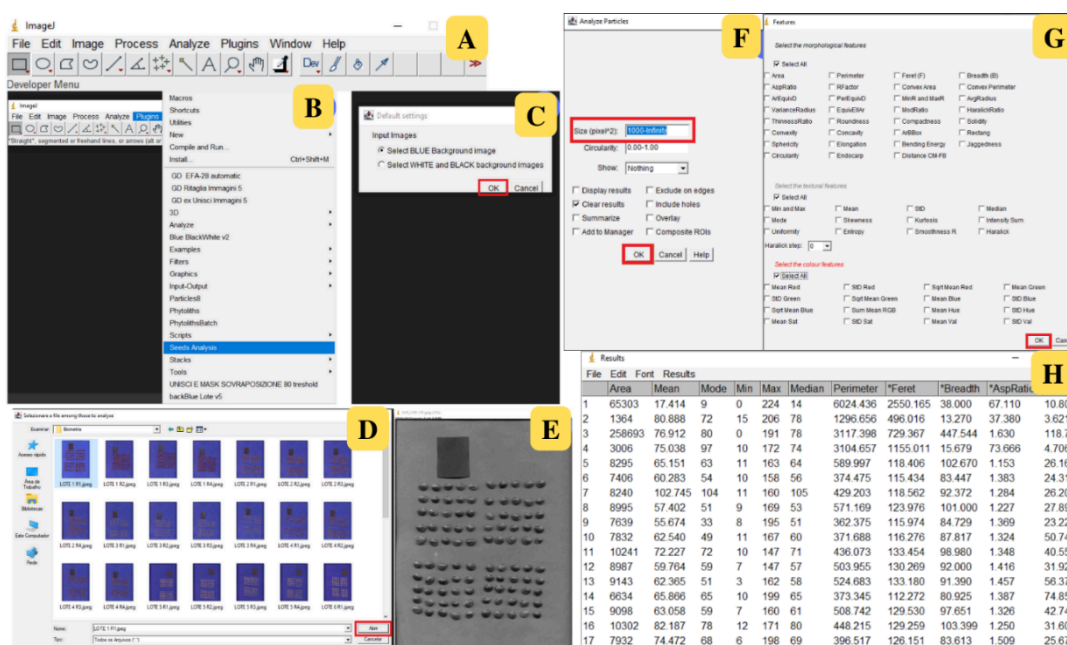
A macro é caracterizada por ser uma sequência automatizada de comandos ou ações que podem ser gravadas e reproduzidas para automatizar tarefas repetitivas em um software. Elas são utilizadas amplamente para simplificar processos complexos, permitindo que os usuários executem uma série de comandos com um único clique ou atalho. Segundo Hunt e Thomas (1999), macros são ferramentas poderosas que ajudam a aumentar a produtividade ao automatizar tarefas frequentes e repetitivas.

O plugin por sua vez é um módulo de software adicional que se integra a um programa principal para expandir suas funcionalidades, tornando o software mais especializado e ajustado de acordo com as necessidades específicas dos usuários. De acordo com Fowler (2003), plugins oferecem uma maneira modular de adicionar novas características a um software sem alterar seu núcleo, proporcionando flexibilidade e extensibilidade. Portanto, são importantes para solucionar problemas específicos de processamento e análise de imagens (Rasband, 2012).

Um exemplo de plugin é a Seeds Analysis, criada por pesquisadores da Universidade de Cagliari, na Itália, foi desenvolvida com o intuito de avaliar diversas sementes em uma única imagem digital. O plugin realiza análises de múltiplas imagens contidas em uma pasta (Loddo *et al.*, 2022). Ele se baseia em imagens com o fundo de coloração padrão azul, branco e preto, e na remoção do mesmo para fazer o destaque das sementes em relação ao fundo da imagem.

Após fazer o recorte do fundo, a imagem é pré-processada para fazer a seleção das partes desejadas, sendo possível também fazer a seleção das sementes de modo individual. Após a segmentação da imagem, é necessário fazer a seleção de certos parâmetros, como o de tamanho mínimo e máximo da área, que é medido em pixels, a seleção dessas opções chaves são necessários, pois isso fará com que facilite a análise das regiões de interesse. Por fim, após essas etapas, o usuário poderá então fazer a escolha de todos os parâmetros de seu interesse (Figura 4).

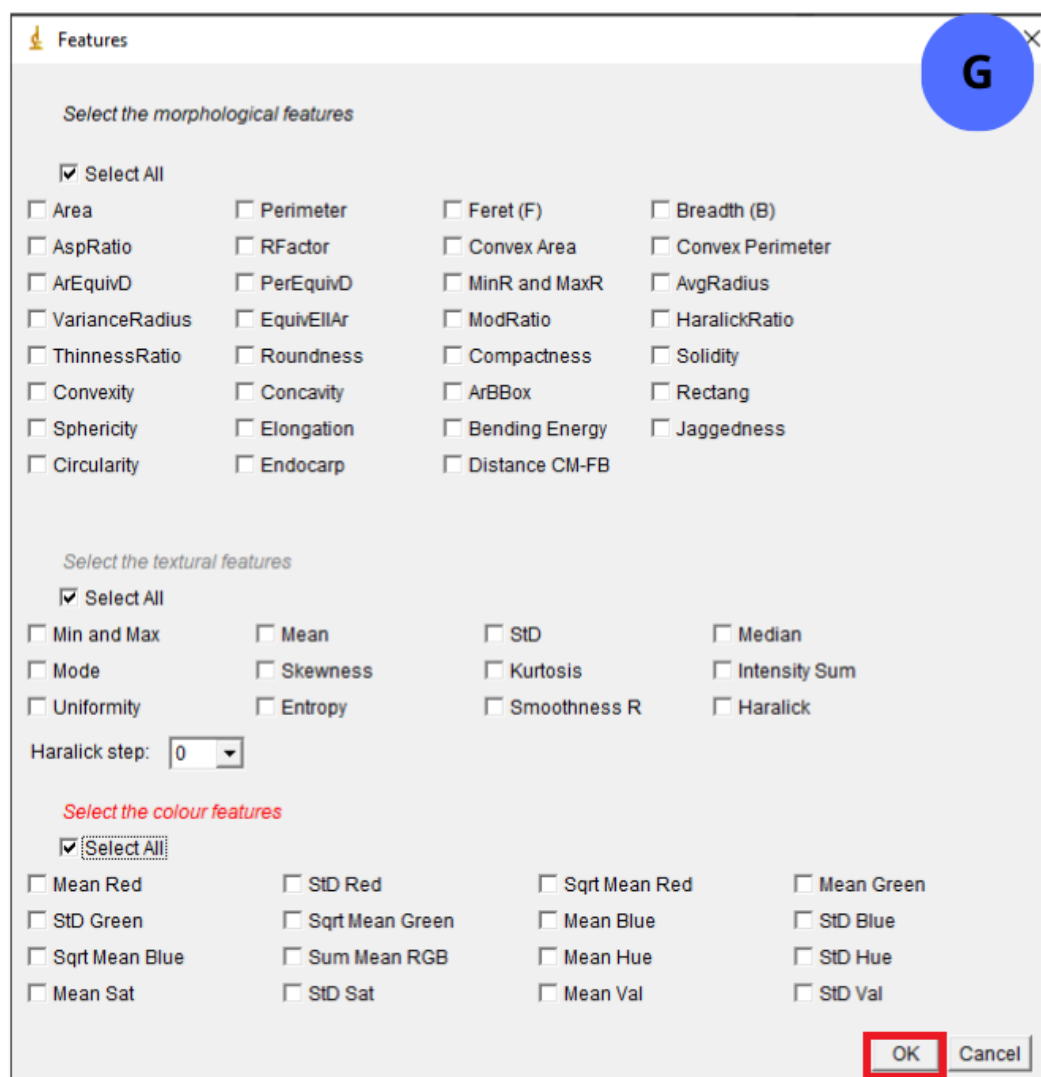
Figura 4 – Interface do plugin e metodologia para usá-la no software ImageJ.



Fonte: Loddo *et al.*, 2022. Legenda: A – Interface do software; B – Selecionando plugin Seeds Analysis; C – Escolhendo fundo de coloração azul; D – Selecionando o arquivo; E – Aplicação da máscara cinza; F – Definindo “1000 infinity” e prosseguindo; G – Escolhendo os parâmetros a serem analisados; H – Arquivo dos resultados para ser exportado para planilha.

São dos 64 parâmetros (Figura 5), sendo 32 de recursos morfológicos como comprimento, largura, relação L/C, área, perímetro e arredondamento (Tabela 1), 16 de cor como R, G, B, H, S, V e Cinza (Tabela 2), e outras 16 de textura calculados para cada canal de cor do RGB e cinza (Loddo *et al.*, 2022), que podem ser visualizados na Tabela 3.

Figura 5 – Parâmetros disponíveis no plugin Seeds Analysis.



Fonte: Loddo *et al.*, 2022.

Tabela 1 – Características morfológicas obtidas no plugin Seeds Analysis no ImageJ.

Variáveis para Características Morfológicas	
Recursos	Descrição
Área	Área de semente (em pixels)
Perimeter	Comprimento do contorno da semente
Feret	Maior diâmetro rastreável com dois pontos do contorno da semente como extremidades, chamado Comprimento
Breadth	Comprimento do maior eixo rastreável perpendicular ao Feret, também chamado de Largura
AspRatio	Feret/largura, também chamado de razão de excentricidade ou retangularidade
RFactor	Fator de forma, definido como $CovexArea/(Ferate \times \pi)$
ConvexArea	Área do polígono convexo traçado entre os pontos externos da região
ArEquivD	Diâmetro do círculo com área equivalente da região, definido $\sqrt{4/\pi \times \acute{A}rea}$
PerEquivD	Diâmetro do círculo com o mesmo perímetro da região, $\acute{A}rea/\pi$
AvgR	Comprimento médio dos raios calculado a partir do centro de massa
VarianceR	Variância dos raios
EquivEllAr	Área da elipse tendo Feret e Largura como eixos
ModRatio	Taxa de modificação, definida como $(2 \times MinR)/Feret$
HaralickRatio	Razão entre a média e o desvio padrão dos raios
ThinnessR	Razão de magreza, também chamada de forma, dada por $Perimeter^2/\acute{A}rea$
Roundness	Medida de arredondamento, definida como $4 \times \acute{A}rea/(\pi \times Feret^2)$

Compactness	Medida de compacidade, expressa por $\sqrt{4/\pi \times \acute{A}rea/Feret}$
Solidity	Medida de solidez, definida como $\acute{A}rea/ConvexArea$
Convexity	Medida de convexidade, definida como $ConvexPerimeter/Perimeter$
Concavity	Medida de concavidade, definida como $ConvexArea - \acute{A}rea$
ArBBox	$\acute{A}rea$ da caixa delimitadora que contém a região
Sphericity	Também chamada de razão de raio, expressa por $M \text{ in}R/M \text{ ax}R$
Elongation	Inverso da circularidade, definido como $Perim^2/(4 \times \pi \times \acute{A}rea)$
Bending Energy	Definido como a soma da curvatura ao quadrado ao longo de todo o contorno
Jaggedness	Medida que representa se uma semente é "serrilhada", definida como $(2 \times \sqrt{4/\pi \times \acute{A}rea})/Perímetro$
Circularity	Também chamado de fator de forma, obtido por $2 \times \pi \times \acute{A}rea/Perímetro^2$
Endocarp	Número de pixels que formam o endocarpo da semente

Fonte: Adaptado de Loddo *et al.*, 2022.

Tabela 2 – Recursos de cores dos espaços de cores RGB e HSV obtidas pelo plugin Seeds Analysis no ImageJ.

Variáveis de cores dos espaços de cores RGB e HSV	
Recursos	Descrição
Mean Red	Média dos valores do canal vermelho
Sqrt Mean Red	Raiz quadrada do valor médio do canal vermelho
StD Red	Desvio padrão dos valores do canal vermelho
Mean Green	Média dos valores do canal Verde
Sqrt Mean Green	Raiz quadrada do valor médio do canal Verde
StD Green	Desvio padrão dos valores do canal verde
Mean Blue	Média dos valores do canal Blue
Sqrt Mean Blue	Raiz quadrada do valor médio do canal azul
StD Blue	Desvio padrão dos valores do canal Azul
Mean RGB	Raiz quadrada do valor médio do canal Azul
Mean H	Tom médio do canal Hue
StD H	Desvio padrão dos valores do canal de Hue
Mean S	Tom médio do canal de saturação
StD S	Desvio padrão dos valores do canal de saturação

Fonte: Adaptado de Loddo *et al.*, 2022

Tabela 3 – Recursos de textura da imagem em tons de cinza obtidas pelo plugin Seeds Analysis no ImageJ.

Recursos de textura da imagem em tons de cinza	
Recursos	Descrição
Mean	Valor médio de cinza na região
Mode	Modo dos valores de cinza
Min	Valor de cinza mínimo na região
Max	Valor de cinza máximo na região
Median	Mediana dos valores de cinza
StD	Desvio padrão de intensidade como medida de contraste
Skewness	Medida da simetria do histograma de nível de cinza em torno do valor médio
Kurtosis	Medida do "tailedness" do histograma de nível de cinza
Intensity sum	Soma dos valores de cinza da região
Uniformity	Máximo quando todos os níveis de cinza no histograma são iguais
Entropy	Medida da variabilidade da distribuição do nível de cinza
Smoothness R	Medida de suavidade

Fonte: Adaptado de Loddo *et al.*, 2022.

Segundo Abràmoff, Magalhães e Ram (2004), o software ImageJ também permite a aquisição de imagens diretamente de dispositivos como scanners, máquinas fotográficas, entre outros equipamentos similares. Além disso, o ImageJ é capaz de realizar conversões entre espaços de cores, como de RGB para matiz, saturação e intensidade, além de oferecer ferramentas para plotagem bidimensional e tridimensional.

4 CAPÍTULO 2 – CONSTRUÇÃO DO SOFTWARE TETRASEED

4.1 Introdução

A problemática que motivou o desenvolvimento deste software foi a necessidade de classificar sementes de feijão-caupi com base em parâmetros específicos, como a porcentagem de cores presentes nas sementes e o estado do embrião. Para resolver essa questão, o software foi projetado para processar imagens escaneadas das sementes, tanto da parte interna quanto da externa.

O sistema realiza a classificação das sementes e gera um arquivo de saída no formato .csv contendo os resultados. Além disso, as informações processadas são exibidas em uma interface gráfica, permitindo ao usuário analisar facilmente os dados gerados, tornando o processo de classificação mais eficiente e preciso.

A criação de um software é um processo complexo e multifacetado que envolve várias etapas, desde o planejamento inicial até a implementação e manutenção. Todo o processo requer uma compreensão profunda das necessidades do usuário, bem como das especificações técnicas e funcionais do sistema a ser desenvolvido (Al-rousan; Abualese, 2019). O desenvolvimento inclui fases como análise de requisitos, design, codificação, teste e manutenção.

O processo inicial acontece na fase de análise de requisitos que é crucial para garantir que o software atenda às necessidades dos usuários finais. Isso envolve a coleta e a documentação das necessidades funcionais e não funcionais, que servirão de base para o design do sistema assim como a coleta de informações sobre o que os usuários esperam do software. A falha em capturar corretamente os requisitos pode resultar em um software que não atende às expectativas, o que é uma das principais causas de falhas em projetos de software (Pressman, 2005).

Finalizado a fase de definição dos requisitos, o design do software começa. A arquitetura do sistema é esboçada nessa fase, incluindo a estrutura de dados, a arquitetura do sistema e a interface do usuário. É essencial um design eficaz para garantir que o software seja escalável, manutenível e fácil de usar. O design de software deve considerar tanto os

aspectos técnicos quanto os aspectos de experiência do usuário, equilibrando a funcionalidade com a usabilidade (Fowler, 2018).

A próxima fase é a de codificação, na qual transforma o design em código executável. Esta fase é caracterizada por exigir uma forte compreensão das linguagens de programação e das práticas de codificação eficiente. De acordo com Otero (2012), a qualidade do código é fundamental para a robustez do software e pode impactar significativamente a facilidade de manutenção futura.

Uma vez que o código é escrito, ele deve ser rigorosamente testado para garantir que funciona conforme o esperado e que é livre de erros. Existem vários tipos de testes, incluindo testes unitários, testes de integração e testes de aceitação, cada um desempenhando um papel crucial na garantia da qualidade do software (Beizer, 2003).

Por fim, temos a manutenção do software é uma fase contínua que envolve a correção de erros, a adição de novas funcionalidades e a adaptação do software às mudanças tecnológicas ou de mercado. Pressman (2005) destaca que a manutenção pode representar até 60% dos custos totais de um projeto de software, o que enfatiza a importância de um design e codificação robustos desde o início.

Nesse contexto, o objetivo deste capítulo é trazer a documentação do TETRASEED, suas configurações, como foi construído, como funciona e o detalhamento do processo de classificação.

4.2 Criação do software TETRASEED

4.2.2 Definições de uso

4.2.2.1 Requisitos básicos

A máquina que executar o software deve ter, no mínimo, 4gb de memória RAM, pois durante o pré-processamento, processamento e classificação, vários processos serão executados em paralelo, no intuito de não tornar estas etapas prolongadas.

4.2.2.2 Restrições do sistema

Tamanho máximo do arquivo a ser enviado: 10Mb

Extensão da imagem: JPEG, JPG e PNG

Navegadores: Google Chrome v103.

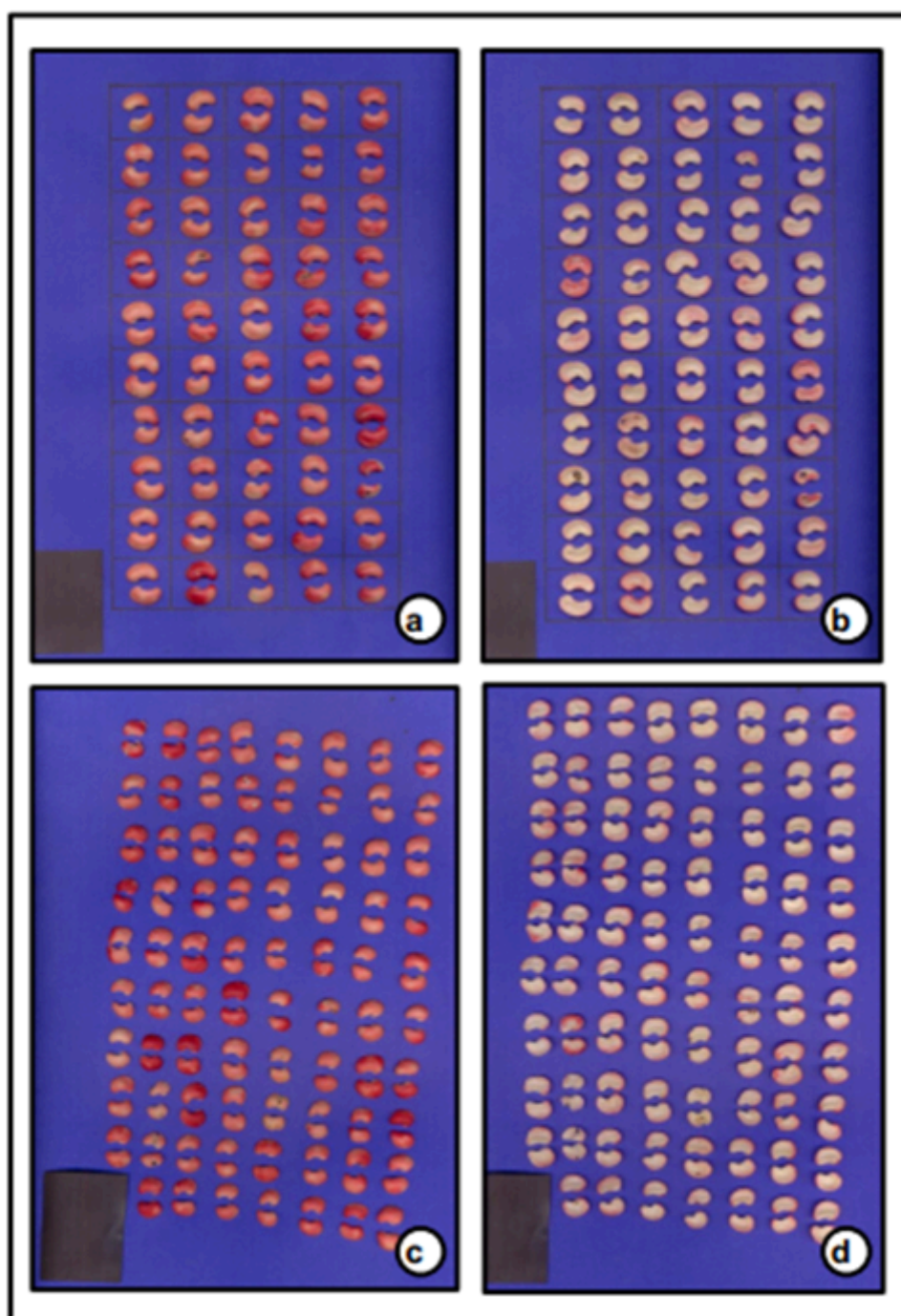
Sistema operacional: Windows, Linux e MacOS.

Aquisição das imagens: O espaçamento entre sementes distintas deve ser maior que o espaçamento entre as partes da semente dividida. Caso a restrição não seja atendida, o sistema não consegue separar as sementes.

Fundo da imagem: A cor de fundo da imagem pode ser azul.

Número máximo de sementes por imagem: varia em relação ao tamanho da semente embebida após a realização do teste de Tetrazólio.

Figura 6 – Disposição das sementes para a captura e a variação do número de sementes.



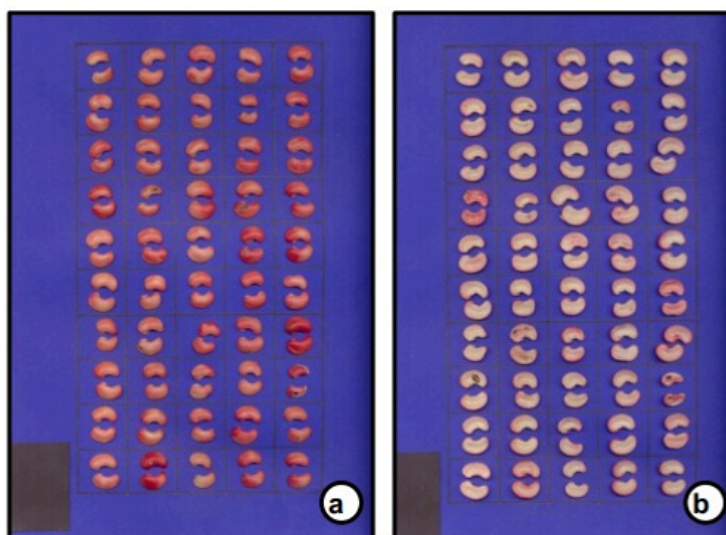
Fonte: Elaborado pela autora. Legenda: a) parte externa das sementes com gride; b) parte interna das sementes com gride; c) parte externa das sementes sem gride e d) parte interna das sementes sem gride.

4.2.3 Requisitos do sistema

4.2.3.1 Requisitos funcionais

Deve ser criada uma interface capaz de receber, por upload, duas imagens em cada análise. Essas imagens são um conjunto de sementes de um lote dispostas em uma matriz, o 1º representa a parte externa e o 2º a parte interna das sementes (Figura 2).

Figura 7 – Imagem da captura das sementes: a) Parte externa da semente; b) Parte interna da semente.



Fonte: Elaborado pela autora.

O sistema deve recortar essas imagens e criar um conjunto de 100 novas imagens, sendo 50 da parte exterior da semente e 50 da parte interior. O sistema deve fornecer, após o processamento da imagem, a porcentagem das cores vermelho carmim claro, vermelho carmim escuro, branco leitoso e cores divergentes (Figura 4). Estes resultados precisam ser distinguidos para a semente em geral e para o embrião.

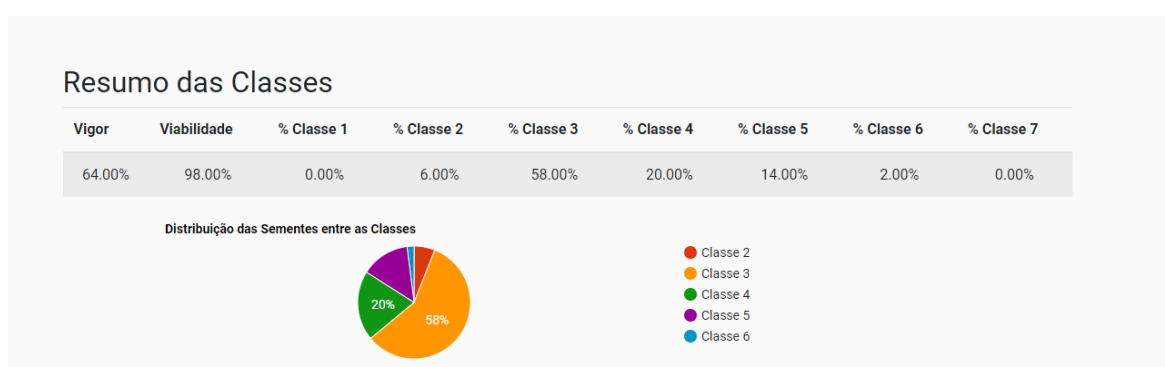
Figura 8 – Análise Individual por semente.



Fonte: Elaborado pela autora.

O sistema deve detectar a quebra somente da parte externa da semente. O sistema deve detectar danos no embrião. Após a análise da semente como um todo, o embrião deve ser localizado e analisado separadamente. Classificar as sementes com base nas classes dispostas na pesquisa base do trabalho (Figura 4).

Figura 9 – Classificação final das sementes.



Fonte: Elaborado pela autora.

A classificação será feita para cada semente (Figura 3) e para o lote de sementes (Figura 4). Por fim, o software deve gerar um relatório no Excel contendo todos os resultados, classificações e/ou falhas de cada semente e do lote (Figura 5).

Figura 10 – Planilha exportada do excel com os dados finais da repetição do lote avaliado gerado pelo Tetrased.

A	B	C	D	E
Semente	Lado	Porcentagem de Branco	Porcentagem de Branco Leitoso	Porcentagem de Vermelho
1	interno	37.13%	6.95%	4.01%
1	externo	9.03%	2.86%	38.79%
2	interno	38.62%	18.36%	3.99%
2	externo	6.19%	1.58%	39.67%
3	interno	35.35%	20.21%	10.41%
3	externo	5.29%	2.29%	38.71%
4	interno	35.29%	15.95%	3.30%
4	externo	3.81%	1.22%	40.52%
5	interno	37.65%	17.73%	5.85%
5	externo	9.98%	7.37%	41.13%
6	interno	40.67%	24.96%	7.79%
6	externo	10.10%	2.74%	39.98%
7	interno	9.54%	2.33%	36.23%
7	externo	2.81%	0.87%	33.51%
8	interno	31.01%	18.22%	6.23%
8	externo	1.51%	0.96%	39.52%
9	interno	38.75%	19.99%	9.03%
9	externo	7.23%	4.39%	38.71%
10	interno	32.92%	20.89%	4.19%
10	externo	5.23%	5.46%	32.14%
11	interno	32.24%	20.65%	23.76%
11	externo	0.34%	0.03%	24.39%
12	interno	36.01%	20.87%	10.21%
12	externo	5.26%	4.00%	37.49%
13	interno	35.57%	24.67%	10.38%

Fonte: Elaborado pela autora.

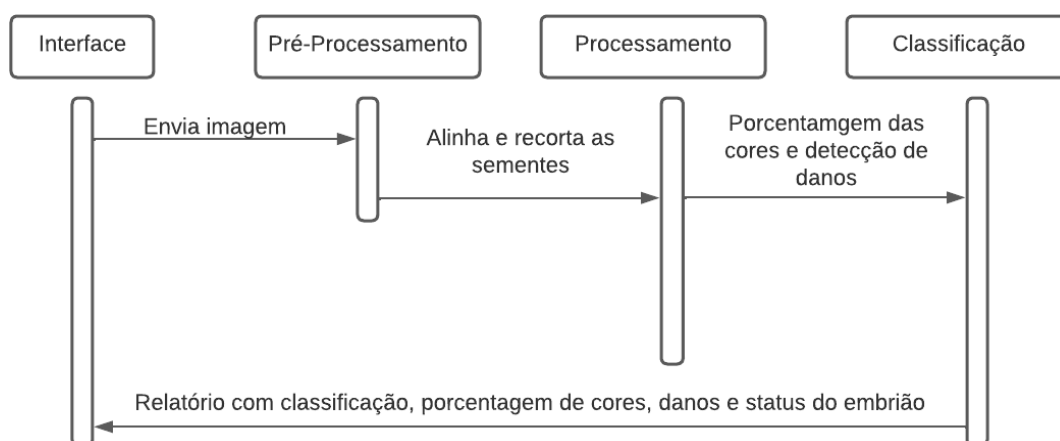
4.2.3.2 Requisitos não funcionais

- Os resultados devem ser organizados em um padrão específico, considerando uma ordem que começa na parte inferior, partindo da esquerda para a direita, e subindo as linhas de sementes até o topo
- As imagens precisam ser numeradas. A numeração das sementes começa a partir do canto inferior esquerdo, indicado pelo bloco preto, e cresce da esquerda para direita, depois subindo a linha.
- A interface precisa ser de fácil utilização.
- O sistema deve ser capaz de ser executado no computador pessoal do usuário, para evitar a contratação e um provável desuso da ferramenta.

4.2.4 Estrutura do software

4.2.4.1 Diagrama de sequência

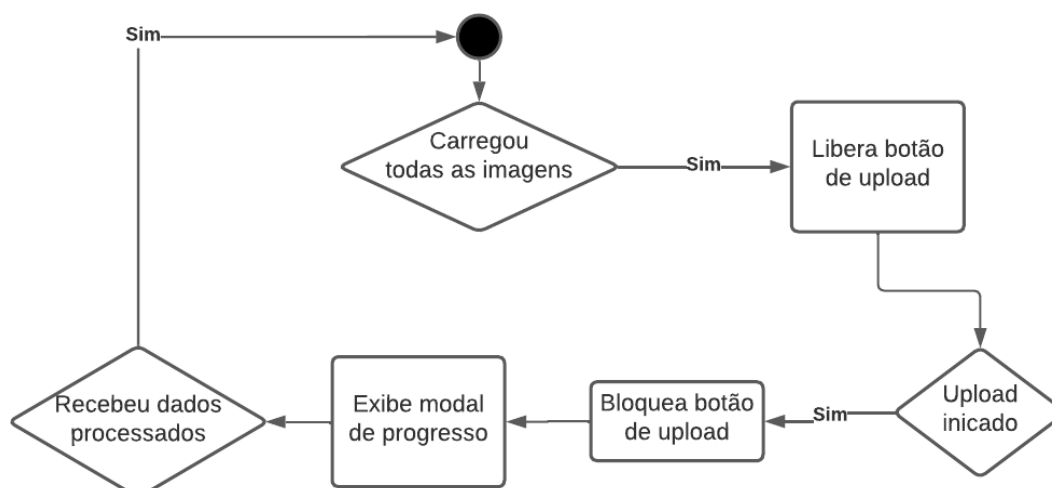
Figura 11 – Diagrama de sequência.



Fonte: Elaborado pelo grupo de pesquisa .

4.2.4.2 Diagrama de atividades

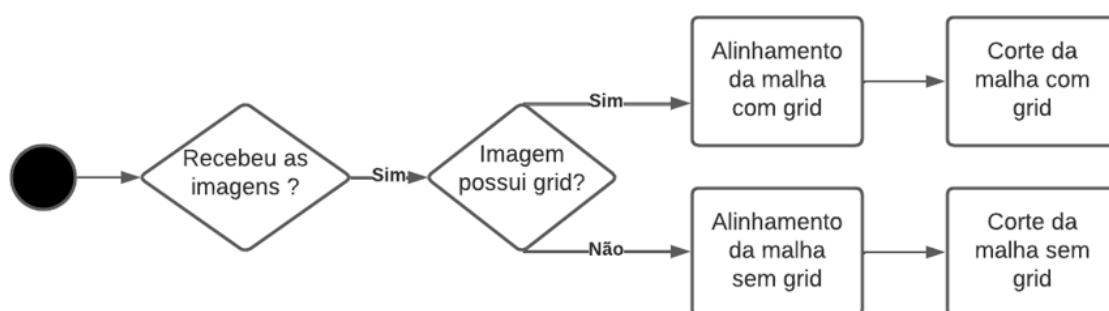
Figura 12 – Diagrama de atividade.



Fonte: Elaborado pelo grupo de pesquisa.

4.2.4.3 Interface

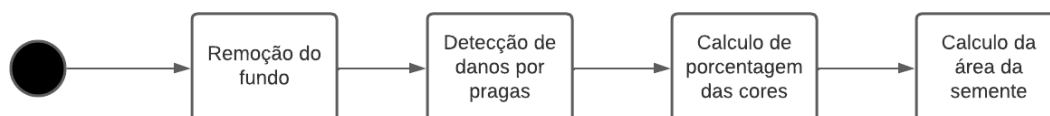
Figura 13 – Diagrama de interface.



Fonte: Elaborado pelo grupo de pesquisa.

4.2.4.4 Pré-Processamento

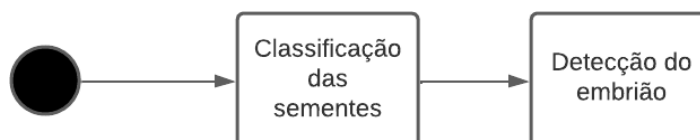
Figura 14 – Diagrama do pré-processamento.



Fonte: Elaborado pelo grupo de pesquisa.

4.2.4.5 Classificação

Figura 15 – Diagrama de classificação.



Fonte: Elaborado pelo grupo de pesquisa.

4.2.5 Ambiente de desenvolvimento

Abaixo estão descritas as linguagens, API e ferramentas utilizadas no desenvolvimento

do projeto e suas respectivas versões.

Linguagens:

- Html - Front End;
- CSS - Front End;
- JavaScript - Front End;
- Python - Back End.

APIs:

- Flask - API em python que faz a integração entre o Front End e o Back End.

Ferramentas

- Jupyter Notebook - Para o desenvolvimento dos algoritmos de processamento das imagens;
- Lucid app - Criação dos diagramas;
- Google documentos - Criação da documentação;
- Canva - Criação do manual do usuário.

4.2.6 Implementação

4.2.6.1 Interface

O FrontEnd da interface foi construído com HTML, CSS e JavaScript para facilitar na construção do software e manutenções futuras.

Com o intuito de não dificultar a comunicação entre a interface de interação com o usuário e os demais módulos do sistema como a aquisição, o processamento, a classificação e o teste, a equipe de interface optou por usar uma API de python para compor o BackEnd. A API adotada foi o Flask, por ser de fácil implementação e contar com uma documentação

detalhada.

A interface do software conta somente com a tela inicial que é modificada para exibir o status do processo, desde quando as imagens são carregadas, durante o processamento e após a conclusão do processamento, quando os resultados estão disponíveis para download.

Para evitar envios de forma incorreta, a interface só aceita o upload de duas imagens por vez, sendo o botão de upload liberado somente quando as duas imagens são selecionadas, e bloqueado quando o upload e processamento está acontecendo.

4.2.6.2 Aquisição

Para a aquisição das imagens, foi usado um scanner da marca HP, modelo HP Scanjet 2004, um scanner de mesa, pois era o modelo disposto no almoxarifado da UFC e tem uma resolução de 1200 dpi, gerando imagens nítidas. O programa utilizado é o próprio Fax e scanner do windows. O scanner teve que ser acoplado a um revestimento, no modelo de uma maleta, para que a imagem tenha o mínimo de ruídos.

Figura 16 – Imagens do Scanner.



Fonte: Elaborado pela autora.

O scanner conta com uma iluminação interna para que a imagem seja capturada com maior nitidez. O fundo das imagens geralmente é da cor azul, por contrastar com a cor das sementes, neste caso vermelha, e assim facilitar a remoção do fundo e identificação das

sementes.

Esta não é a única forma para realizar a aquisição das imagens, tendo em vista outros equipamentos como celular ou câmera fotográfica que podem também serem usados. Contudo o scanner é o mais indicado pois garante uma posição e resolução adequada para capturar as imagens das sementes.

4.2.7 Pré-Processamento

4.2.7.1 Alinhamento da malha

Após a segmentação com o algoritmo thresholding para que a borda da semente seja de fácil detecção, são definidos dois pontos na imagem, um no canto superior esquerdo e outro no canto inferior esquerdo. Estes pontos percorrem a imagem da esquerda para a direita até encontrarem a borda da malha.

A diferença desses pontos é usada para calcular o ângulo de rotação da malha, usando a distância entre os pontos como a hipotenusa do triângulo.

4.2.8 Corte da malha

4.2.8.1 Com grid

São definidos 4 pontos na imagem, no lado direito, no lado esquerdo, em cima e em baixo. Esses pontos percorrem a imagem até acharem as bordas da malha com as sementes.

É usado um tamanho padrão de 5x10 para o corte da malha e salvo a imagem de cada semente em um vetor. Neste modo, há um limite de 50 sementes.

4.2.8.2 Sem grid

Neste caso, é usado o histograma para detectar a posição das sementes na imagem e os espaços entre elas é usado para realizar o corte da imagem.

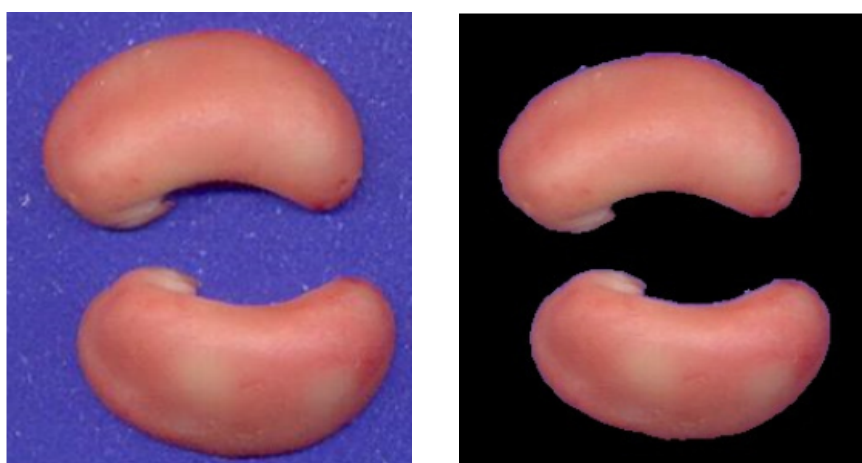
Sem o grid, não há limite de colunas ou linhas, somente a necessidade de que o espaço entre sementes diferentes seja maior que o espaço entre as metades da mesma semente e que seja mantido um alinhamento das colunas e linhas.

4.2.9 Processamento – *TETRASEED*

4.2.9.1 Remoção do Fundo

A primeira etapa do processamento consiste na extração dos pixels da semente. Para isso, a imagem é convertida para o formato de representação RGB. Como o fundo da imagem possui uma coloração azul, esse formato de representação mostra-se adequado por possuir uma componente azul. A partir disso, foi definido um intervalo nos valores R, G e B deste canal. Utilizando esse intervalo, é realizada uma limiarização, processo que consiste em zerar todos os pixels que não pertencem ao intervalo especificado. Após a limiarização, é criada uma máscara de pixels utilizando apenas os valores do canal vermelho resultante. A máscara consiste de uma matriz 0 e 1 's que indicam se determinado pixel deve ser zerado ou não. Por fim, a máscara é aplicada na imagem original.

Figura 17 – Imagem original (esquerda) e imagem resultante após a remoção de fundo (direita).



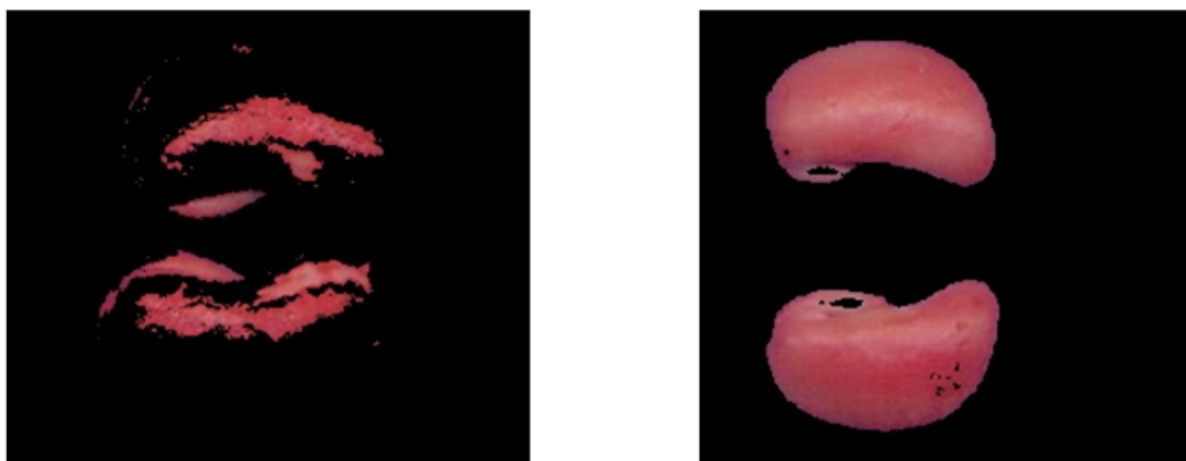
Fonte: Elaborado pelo grupo de pesquisa.

4.2.9.2 Extração da Cor Vermelha

Para extração da cor vermelha foi utilizado o sistema de representação HSV, dado que, após uma série de testes utilizando outros sistemas de representação, foi constatado que a informação desejada para este problema estava codificada no canal H, que representa a matiz da cor. Após análises de média e desvio padrão nos valores dos canais da imagem utilizando exemplos selecionados de vermelho carmim claro e vermelho carmim escuro, foi possível

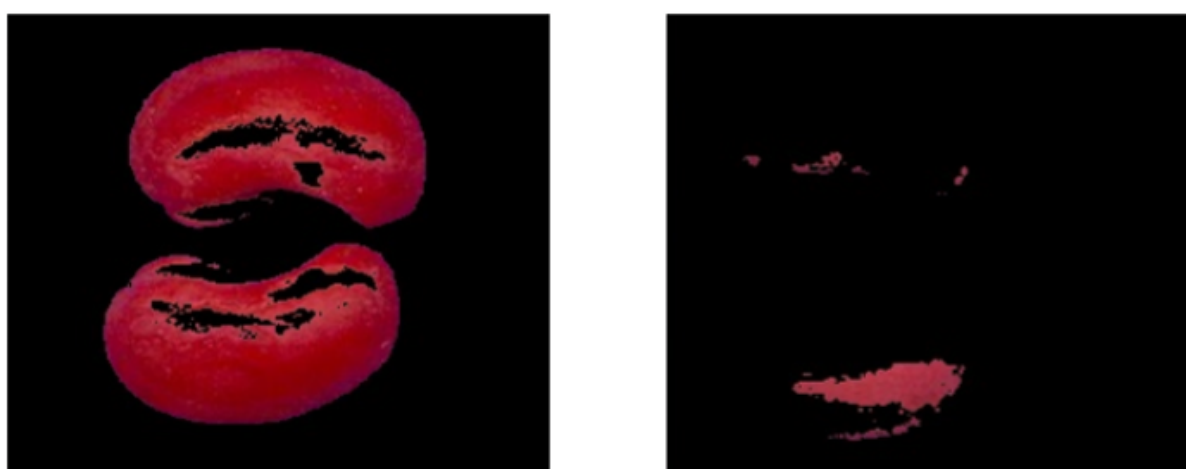
definir intervalos para os canais do formato HSV que extraíam os pixels que representavam cada um dos vermelhos. Após isso, os pixels extraídos são computados e é calculada a sua área em relação à área total da semente.

Figura 18 – Extração do Vermelho Carmim Claro: Cor majoritária Carmim escuro (esquerda) e cor majoritária Carmim claro (direita).



Fonte: Elaborado pelo grupo de pesquisa.

Figura 19 – Extração do Vermelho Carmim escuro: Cor majoritária Carmim escuro (esquerda) e cor majoritária Carmim claro (direita).

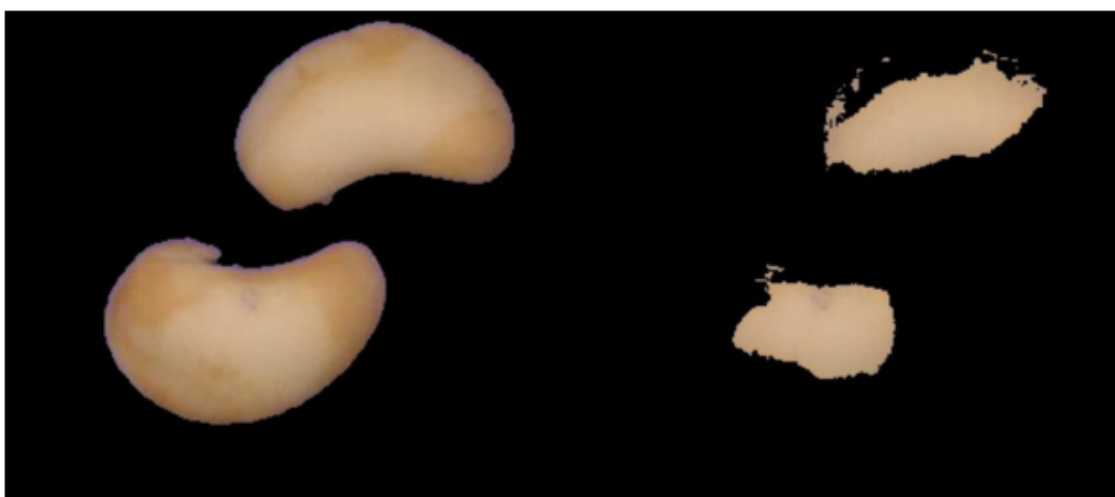


Fonte: Elaborado pelo grupo de pesquisa.

4.2.9.3 *Extração do Branco Leitoso*

O processo de extração do branco leitoso foi semelhante ao do vermelho carmim. Contudo, a informação necessária não estava codificada apenas em um canal, como no caso do vermelho carmim, fato que dificultou a extração. Após uma série de testes, o sistema de representação RGB mostrou-se mais adequado para a extração. Assim como para o vermelho carmim, foram realizadas análises a partir da média e desvio padrão de amostras selecionadas de sementes branco leitoso e foram gerados intervalos para cada um dos canais do modelo de representação. Assim, é realizada uma limiarização utilizando esses intervalos e é calculada a área de pixels extraída relativo à área total da semente.

Figura 20 – Segmentação do branco leitoso.

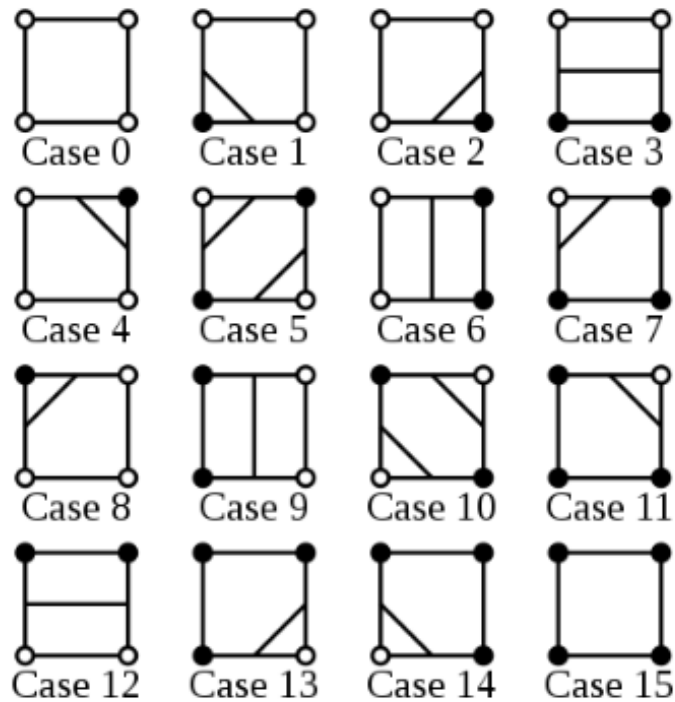


Fonte: Elaborado pelo grupo de pesquisa.

4.2.9.4 *Identificação de buracos na semente*

Para realizar a identificação dos buracos na semente, foi utilizado o algoritmo *MarchingSquares*, que é um algoritmo utilizado para gerar contornos em imagens. De acordo com um limiar fornecido, inicialmente o algoritmo calcula uma matriz binária com mesma dimensão da imagem de acordo com o limiar, onde os pixels que apresentam valor maior que esse limiar recebem 1 e os que apresentam valor menor recebem 0. Após isso, o algoritmo percorre a matriz binária analisando os valores em grupos de 4 e, de acordo com 16 padrões predefinidos, desenha linhas na imagem que formam os seus contornos.

Figura 21 – Padrões pré-definidos pelo algoritmo MarchingSquares.



Fonte: Elaborado pelo grupo de pesquisa.

Utilizando este algoritmo, temos então todos os contornos presentes na imagem da semente. A partir disso, é considerado que o maior polígono é respectivo à semente e é calculado quais polígonos estão inscritos à esse polígono maior. Dessa forma, temos a contagem da quantidade de buracos presentes na semente.

Figura 22 – Extração dos buracos da semente.



Fonte: Elaborado pelo grupo de pesquisa.

4.2.10 Classificação

4.2.10.1 Classificação das sementes

A classificação das sementes é baseada em classes existentes na literatura e usa a porcentagem de vermelho carmim, claro e escuro, que a semente possui. Para classificar as sementes, o software conta com 7 classes de sementes. A classificação inicial foi feita manualmente usando a análise de histograma das imagens.

A partir dessa análise, houve a tentativa de criar modelos de aprendizagem para a rede neural. Alguns algoritmos como KNN, Naive Bayes, Adaboost e SVM foram usados, porém não tiveram uma acurácia superior a 50%.

O algoritmo com maior acurácia foi a rede neural convolucional (CNN), chegando a 88% para a classificação das imagens.

4.2.11 Descrição do algoritmo de CNN

A CNN é um algoritmo de aprendizado profundo que analisa uma imagem e pode definir pesos e vieses a características a partir dos aspectos detectados nos objetos contidos na imagem. Um algoritmo CNN tende a usar menos processamento que os demais algoritmos de análise por conta da sua estrutura, que usa uma variação de perceptrons multi-camada.

A rede Inception-ResNet v2 foi utilizada, ela é uma arquitetura para rede neural profunda, que foi desenvolvida para tarefas de VC, como a classificação de imagens. Para treinamento da rede, foi utilizada para cada um dos lotes uma das 4 imagens obtidas pelo scanner. Os valores dos pixels foram normalizados para a faixa [0, 1] para garantir a uniformidade dos dados e melhorar a convergência durante o treinamento. Técnicas de aumento de dados (Data Augmentation), como rotação, tradução, espelhamento e zoom, foram aplicadas para aumentar a diversidade do conjunto de treinamento e evitar o overfitting.

4.2.11.1 Hiperparâmetros

- **Taxa de Aprendizado (Learning Rate):** Iniciada em 0,001 com decaimento adaptativo.

- **Tamanho do Lote (Batch Size):** 32 imagens por lote.
- **Épocas:** O treinamento foi realizado por 50 épocas.

Os dados foram divididos em conjuntos de treinamento (70%), validação (20%) e teste (10%). Métricas de desempenho utilizadas para avaliar o desempenho do modelo foram acurácia, precisão, recall e F1-score. Foram utilizadas as métricas Acurácia (Equação 1), precisão (Equação 2), Recall (Equação 3) e F-Measure (Equação 4) para avaliar a eficiência da rede, calculadas a partir das Equações (1, 2, 3 e 4).

A acurácia é a proporção de predições corretas (tanto verdadeiros positivos quanto verdadeiros negativos) sobre o total de instâncias avaliadas.

$$\text{Equação 1: Acurácia} = \frac{TP+TN}{TP+TN+FP+FN}$$

- TP (Verdadeiro Positivo): Casos verdadeiros positivos.
- TN (Verdadeiro Negativo): Casos verdadeiros negativos.
- FP (Falso Positivo): Casos falsos positivos.
- FN (Falso Negativo): Casos falsos negativos.

A precisão (ou valor preditivo positivo) é a proporção de predições positivas corretas sobre o total de predições positivas feitas pelo modelo.

$$\text{Equação 2: Precisão} = \frac{TP}{TP+FP}$$

Alta precisão indica que a maioria das instâncias classificadas como positivas são realmente positivas.

O recall (ou sensibilidade) é a proporção de predições positivas corretas sobre o total de instâncias positivas reais.

$$\text{Equação 3: Recall} = \frac{TP}{TP+FN}$$

Alto recall indica que a maioria das instâncias positivas reais são corretamente identificadas pelo modelo.

O F1-score é a média harmônica entre a precisão e o recall, proporcionando um balanço entre os dois. É especialmente útil quando há um desequilíbrio entre as classes.

$$\text{Equação 4: } F1 - Score = 2 \cdot \frac{\text{Precisão} \cdot \text{Recall}}{\text{Precisão} + \text{Recall}}$$

Um F1-score alto indica que o modelo tem tanto alta precisão quanto alto recall.

Técnicas estatísticas de validação cruzada foram aplicadas para avaliar a capacidade de generalização de um modelo de aprendizado de máquina. Ao invés de simplesmente dividir os dados em um conjunto de treinamento e um conjunto de teste, a validação cruzada permite uma avaliação mais robusta e confiável do desempenho do modelo para garantir a generalização do modelo e evitar o overfitting. O modelo de CNN Inception-ResNet v2 alcançou uma acurácia de 86% no conjunto de teste, superando outras arquiteturas comparativas.

Após o treinamento da rede, foi construída uma matriz de confusão, tendo-se as classificações preditas pela rede em função das classificações visuais:

(i) Dividimos os dados em 70% para treinamento, 20% para validação e 10% para conjunto de teste;

(ii) Ajustamos as CNNs no conjunto de treinamento, usando o conjunto de validação para evitar overfitting;

(iii) Usamos validação cruzada estratificada de 10 vezes no conjunto de treinamento e validação combinados, para treinar os cinco algoritmos de classificação;

(iv) Finalmente avaliamos a classificação realizada no conjunto de teste.

4.2.12 Detecção e classificação do embrião

Para a detecção do embrião, o algoritmo analisa a área do branco leitoso da parte interna da semente.

O CNN também foi utilizado na detecção do embrião, chegando a 98% de acurácia para a detecção da posição do embrião.

4.2.13 Testes

4.2.13.1 1º Teste - Input de arquivos em outros formatos

Foi usada a plataforma SonarQube para verificação da existência de bugs e análise da qualidade da interface.

Para a verificação do Input de arquivos foram testados os formatos pdf, png, mp4, zip, jpg e jpeg, sendo aceitos somente os arquivos de formato jpg e jpeg. Quando o usuário tenta enviar algum arquivo fora do tipo aceito, é notificado com uma mensagem de alerta.

Figura 23 – 1º Teste - Input de arquivos em outros formatos.



Fonte: Elaborado pelo grupo de pesquisa.

4.2.13.2 2º Teste: Redimensionamento de janela

Para garantir que o usuário tenha a melhor visualização da tela, foram realizados testes quanto ao redimensionamento de janela, buscando verificar a alocação dos elementos da página conforme o redimensionamento da janela. Na versão 2.0, o software já aceita diferentes formatos de tela, tanto na orientação retrato quanto na paisagem, com diferentes resoluções.

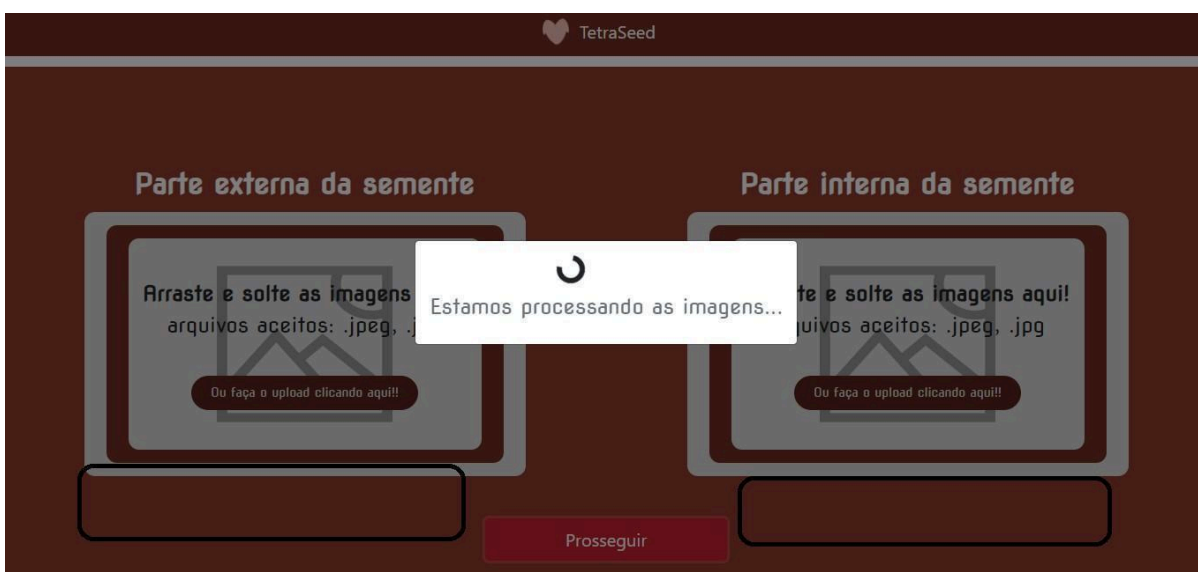
4.2.13.3 3º Teste - Clique no botão prosseguir, sem arquivos enviados

O botão “prosseguir” foi testado em relação a sua funcionalidade quando não tem arquivos a serem enviados. Por padrão, o botão prosseguir permanece com a propriedade desativada enquanto não é enviado um arquivo de formato correto.

O teste consistiu em desativar manualmente essa propriedade no código da página, usando o inspecionar elementos do Google Chrome, e tentar clicar no botão Prosseguir, sem nenhum arquivo ter sido enviado.

Foi constatado que após o clique no botão Prosseguir, a página apresentou o modal de carregamento como se o arquivo tivesse sido submetido. Essa é uma funcionalidade que precisa ser melhorada, pois o ideal é que apareça uma mensagem de alerta indicando que nenhum arquivo foi selecionado para ser enviado.

Figura 24 – 3º Teste - Clique no botão prosseguir, sem arquivos enviados.



Fonte: Elaborado pelo grupo de pesquisa.

4.2.13.4 4º Teste - Alinhamento de itens

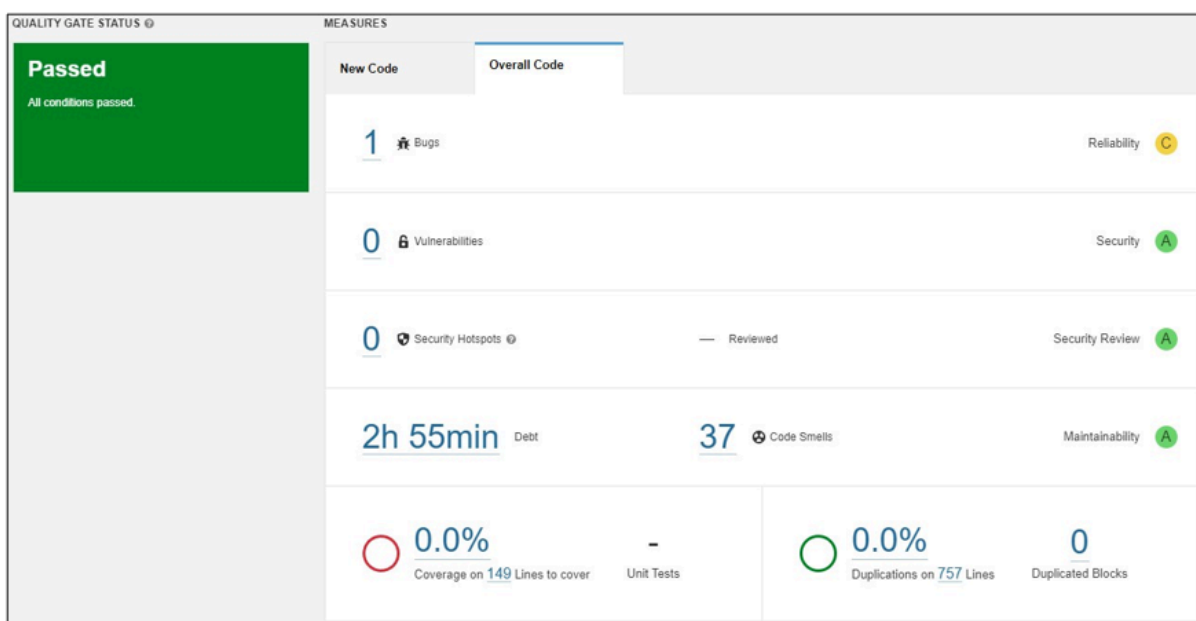
Foi testado o alinhamento dos componentes da página para diferentes resoluções, sendo corrigido o alinhamento na versão 2.0, que posiciona os itens de melhor forma de acordo com a tela.

4.2.13.5 5º Teste - Análise estática de código do Front-End usando ferramenta SonarQube

Para a realização deste teste foi utilizada a SonarQube que é uma plataforma de código aberto desenvolvida pela SonarSource para inspeção contínua da qualidade do código para realizar revisões automáticas com análise estática de código para detectar bugs, code smells em 29 linguagens de programação.

A ferramenta usada separa os problemas identificados em quatro grandes categorias, chamadas de Bugs, Vulnerabilities, Code Smells e Security Hotspots. Cada uma delas é relacionada a um tipo de falha, com diferentes níveis de severidade em questão da segurança da aplicação e de problemas de execução. Ao executar a ferramenta no projeto do Front-End, ela mostrou o seguinte resultado:

Figura 25 – 5º Teste - Análise estática de código do Front-End usando ferramenta SonarQube.



Fonte: Elaborado pelo grupo de pesquisa.

O Bug em questão é apenas relacionado a um objeto instanciado que não foi utilizado, enquanto os Code Smells se referem a más práticas de programação que, embora não afetem a execução da aplicação, podem tornar o código confuso ou de difícil leitura e entendimento.

Pela sua própria definição, os Code Smells podem ser ignorados na análise do

funcionamento do sistema. Já sobre o Bug, ele pode não estar relacionado a qualquer função crítica da aplicação, considerando que o objeto instanciado não é utilizado, ou ele pode ser um falso-positivo encontrado pela ferramenta. Ambas as situações não afetam a capacidade da aplicação em realizar as tarefas necessárias.

Diante do exposto, embora tenha sido encontrado alguns problemas, nenhum deles interfere de forma significativa na execução da tarefa realizada pela aplicação. Dessa forma, é possível concluir que em relação a falhas de programação, a aplicação não apresenta problemas que venham a interferir com sua tarefa.

4.2.14 TETRASEED

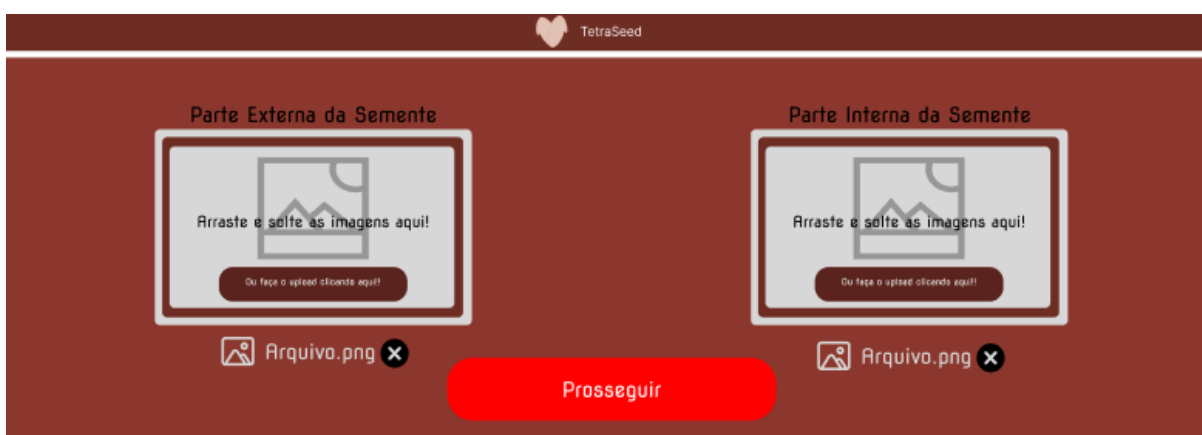
A seguir estão dispostas a interface versão inicial do sistema e suas aplicações.

Figura 26 – Tela inicial para upload das imagens.



Fonte: Elaborado pelo grupo de pesquisa.

Figura 27 – Tela com os arquivos prontos para serem enviados.



Fonte: Elaborado pelo grupo de pesquisa.

Figura 28 – Processamento das imagens.



Fonte: Elaborado pelo grupo de pesquisa.

Figura 29 – Resultados gerados e dispostos para download.



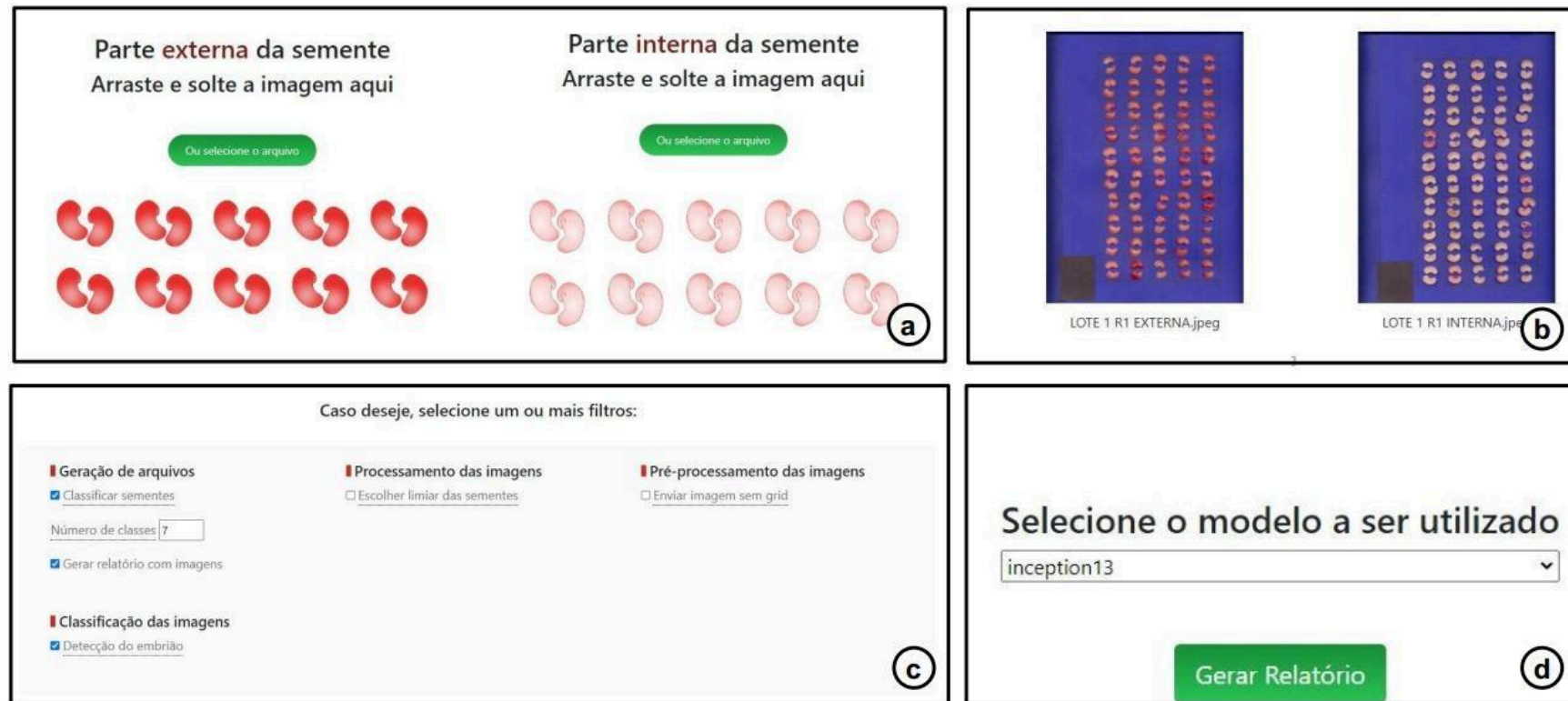
Fonte: Elaborado pelo grupo de pesquisa.

Figura 30 – Interface inicial do TetraSeed.



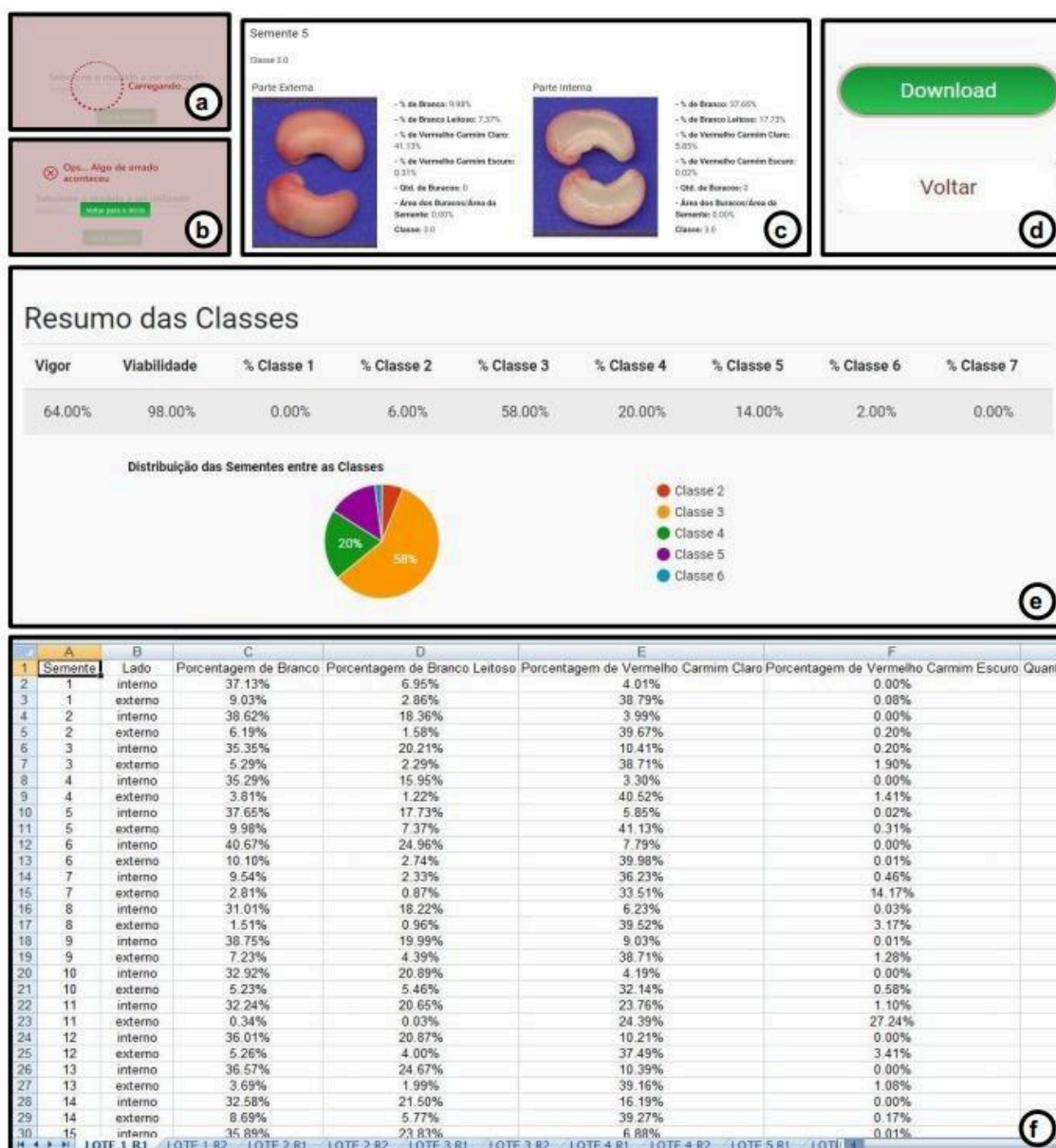
Fonte: Elaborado pela autora. Legenda: a) Página inicial do software; b) Área para login e senha; c) Menu inicial e d) Descrição do software.

Figura 31 – Tela inicial para o upload das imagens.



Fonte: Elaborado pela autora. Legenda: a) interface para fazer o upload das imagens; b) Imagens selecionadas para análise; c) interface para escolha de classes e entre outras ferramentas e d) local para selecionar o modelo de rede neural ou fazer upload de outro.

Figura 32 – Tela com os arquivos prontos para serem enviados.



Fonte: Elaborado pela autora. Legenda: a) Sistema realizando a análise dos dados; b) Sistema configurando erro; c) Avaliação individual de cada semente; d) Download dos dados e) Resumo dos resultados e classificação com base na literatura e f) Excel exportado no software.

5 CAPÍTULO 3 - TETRASEED: SISTEMA DE VISÃO ARTIFICIAL PARA ANÁLISE E CLASSIFICAÇÃO DE SEMENTES DE FEIJÃO CAUPI (*Vigna unguiculata* (L.) Walp.)

Resumo - O teste de tetrazólio é uma técnica eficiente e de baixo custo para avaliar a qualidade fisiológica das sementes, oferecendo informações rápidas sobre viabilidade e vigor ao identificar deterioração e danos. No entanto, a interpretação manual dos resultados pode ser subjetiva e trabalhosa, especialmente com grandes volumes de amostras. Para superar essas limitações, a visão computacional (VC) tem sido aplicada para automatizar e padronizar a análise de imagens, oferecendo maior precisão e eficiência. Portanto, o objetivo deste trabalho foi criar um sistema de VC para a avaliação nas imagens de sementes submetidas ao teste de tetrazólio e classificá-las de maneira automatizada visando diminuir a subjetividade e otimizar o processo de classificação. Desenvolvido em Python TETRASEED é um sistema de visão artificial para análise e classificação de sementes de feijão caupi após a realização do teste de tetrazólio. Visa analisar as imagens das sementes e gerar os resultados em porcentagens de cor de vermelho carmim claro, vermelho carmim escuro, branco leitoso e classificar de forma automatizada as sementes com base na literatura atual para o teste de tetrazólio de feijão caupi. A criação do sistema de VC começou com a introdução dos dados qualitativos do teste de tetrazólio, validados pela correlação com os resultados dos testes de vigor. Foi utilizado redes neurais convolucionais (CNN), para análise e classificação. O TETRASEED apresentou uma acurácia de 88%, na avaliação das imagens de sementes submetidas ao teste de tetrazólio indicando um desempenho sólido na classificação automatizada das sementes de feijão-caupi a partir de dados qualitativos do teste de tetrazólio, demonstrando uma boa capacidade de aprendizagem e generalização, ajustes adicionais são necessários para garantir uma performance mais equilibrada em todas as classes, aumentando a confiabilidade do sistema para aplicações práticas na avaliação de sementes.

Palavras-chave: classificação de sementes; análise de imagens; reconhecimento de padrões; redes neurais convolucionais (CNN).

5 CHAPTER 3 - TETRASEED: ARTIFICIAL VISION SYSTEM FOR ANALYSIS AND CLASSIFICATION OF COWPEA BEAN SEEDS (*Vigna unguiculata* (L.) Walp.)

Abstract - The tetrazolium test is an efficient and low-cost technique for assessing the physiological quality of seeds, providing quick information on viability and vigor by identifying deterioration and damage. However, manual interpretation of results can be subjective and laborious, especially with large volumes of samples. To overcome these limitations, computer vision (CV) has been applied to automate and standardize image analysis, offering greater accuracy and efficiency. Therefore, the aim of this work was to develop a CV system for evaluating images of seeds subjected to the tetrazolium test and to classify them automatically, aiming to reduce subjectivity and optimize the classification process. Developed in Python, TETRASEED is an artificial vision system for the analysis and classification of cowpea seeds after the tetrazolium test. It aims to analyze seed images and generate results in percentages of light carmine red, dark carmine red, and milky white colors, and to automatically classify the seeds based on the current literature for the cowpea tetrazolium test. The creation of the computer vision system began with the introduction of qualitative data from the tetrazolium test, validated by correlation with vigor test results. Convolutional Neural Networks (CNNs) were used for analysis and classification. TETRASEED achieved an accuracy of 88% in evaluating the images of seeds subjected to the tetrazolium test, indicating a solid performance in the automated classification of cowpea seeds from qualitative tetrazolium test data. It demonstrated good learning and generalization capabilities, although further adjustments are needed to ensure a more balanced performance across all classes, increasing the system's reliability for practical seed evaluation applications.

Key words: seed classification; image analysis; pattern recognition; convolutional neural networks (CNN).

5.1 Introdução

O feijão-caupi (*Vigna unguiculata* (L.) Walp.), também conhecido como feijão-de-corda, é uma leguminosa de grande importância econômica e nutricional, especialmente em regiões tropicais e subtropicais (Silva, 2019). Possui alta adaptabilidade a condições adversas, como solos pobres e baixos índices pluviométricos, características que o tornam uma cultura de escolha em regiões semiáridas (de Abreu Araújo *et al.*, 2019). Além disso, ele possui um alto valor nutricional, sendo uma fonte significativa de proteínas, fibras, vitaminas e minerais, contribuindo para a dieta de milhões de pessoas em todo o mundo (Freire Filho *et al.*, 2011; Freire Filho *et al.*, 2005; Freire Filho *et al.*, 1988).

Para garantir a alta produtividade e a qualidade das colheitas, a utilização de sementes viáveis e vigorosas é essencial (Marcos Filho, 2020). Nesse contexto, o teste de tetrazólio se destaca como uma técnica eficaz para a avaliação da qualidade fisiológica das sementes, oferecendo informações rápidas sobre a viabilidade e o vigor das sementes, fatores críticos para o sucesso no campo (França Neto; Krzyzanowski, 2018). O teste de tetrazólio é um teste rápido, preciso e de baixo custo que permite a identificação de fatores como deterioração, danos mecânicos entre outros fatores que podem interferir na qualidade das sementes, permitindo obter resultados sobre a viabilidade e vigor das mesmas ((Fogaça *et al.*, 2006; Ribeiro *et al.*, 2010; Rodrigues *et al.*, 2015). Esse teste, baseia-se de forma indiretamente na atividade respiratória das sementes (França Neto; Krzyzanowski e Da Costa, 1998). As sementes ao serem imersas na solução de tetrazólio (incolor), devido a atividade da enzima malato desidrogenase, reduz o sal tetrazólio (2,3,5-trifenil cloreto de tetrazólio ou TCT) nos tecidos vivos formando em um composto vermelho, estável e não- difusível, o trifenilformazan (França Neto; Krzyzanowski, 2018).

Apesar de sua eficácia, a interpretação dos resultados do teste de tetrazólio em sementes é um processo que pode ser subjetivo e depende da experiência do analista. A análise manual das sementes, além de ser trabalhosa, está sujeita a variações nos resultados, especialmente quando se trata de grandes volumes de amostras (Marcos Filho, 2015). A padronização e a automação da análise são desafios que têm levado

pesquisadores a explorar novas tecnologias, como a VC, para melhorar a eficiência e a precisão do teste (Patrício; Rieder, 2018).

Nos últimos anos, a VC é utilizada como uma ferramenta poderosa para a análise automatizada de imagens em diversas áreas, incluindo a agricultura. Essa tecnologia envolve a aplicação de algoritmos de processamento de imagens e técnicas de inteligência artificial para extrair informações relevantes de imagens digitais, permitindo uma análise mais rápida e precisa (Kamilaris; Prenafeta-Boldú, 2018). A integração da VC com o teste de tetrazólio oferece um grande potencial para a automatização do processo de avaliação de sementes, reduzindo a subjetividade e o tempo de avaliação (Howarth; Stanwood, 1993; Qiao *et al.*, 2023).

Uma aplicação prática da VC no teste de tetrazólio envolve a identificação e quantificação automática das áreas coloridas das sementes, que indicam os tecidos viáveis. Esse processo utiliza técnicas como segmentação de imagem baseada em cor, filtragem de regiões de interesse e aprendizado de máquina para classificar as áreas conforme padrões pré definidos (Gonzalez; Woods, 2010).

Com o uso de redes neurais convolucionais (CNN), é possível distinguir automaticamente entre áreas viáveis e não viáveis nas sementes, proporcionando uma avaliação quantitativa precisa da qualidade das sementes (Ghaffari, 2024). Essa abordagem não só acelera o processo, como também aumenta a precisão da análise, a automação pode melhorar efetivamente a eficiência de avaliações (Fernandes *et al.*, 2023)

Além disso, a VC facilita a integração de dados, permitindo a criação de bancos de dados que podem ser usados para treinar modelos de aprendizado de máquina, melhorando continuamente a precisão da análise (Lecun *et al.*, 2015). Isso é especialmente relevante para a indústria de sementes, onde a uniformidade e a qualidade são fundamentais para a competitividade no mercado (Tavares *et al.*, 2016).

Portanto, a associação de VC no teste de tetrazólio para sementes de feijão-caupi representa uma inovação significativa que pode revolucionar a avaliação da qualidade das sementes. Automatizando e padronizando a análise, essa tecnologia não apenas

melhora a eficiência e a precisão do teste, mas também oferece uma ferramenta valiosa para decisões em programas de melhoramento e produção de sementes (Chandra *et al.*, 2020). Nesse contexto, o objetivo deste trabalho foi criar um sistema de VC para a avaliação nas imagens de sementes submetidas ao teste de tetrazólio e classificá-las de maneira automatizada visando diminuir a subjetividade e otimizar o processo de classificação.

5.2 Material e métodos

O trabalho foi desenvolvido no Laboratório de Análise de Sementes, do Departamento de Fitotecnia, do Centro de Ciências Agrárias da Universidade Federal do Ceará-UFC em parceria com o Departamento de Computação. Foram utilizados dez lotes de sementes de feijão-caupi.

5.2.1 Etapa I: Análise de qualidade dos lotes de sementes com testes tradicionais.

5.2.1.1 Teste de germinação e primeira contagem

No teste de germinação, para cada lote, foram utilizadas quatro repetições de cinquenta sementes. A semeadura foi feita no papel de (Germitest®) umedecido com água destilada, na proporção de 2,5 vezes a massa do substrato. Após isso, as sementes foram acondicionadas em câmara de germinação tipo Biochemical Oxygen Demand (BOD) regulada à temperatura de 25 °C, com fotoperíodo de 12h, sendo feita a primeira contagem no 5º dia, considerando apenas as plântulas normais, finalizando no 8º dia. Os dados foram obtidos por meio da média das quatro repetições, transformando o resultado em porcentagem.

5.2.1.2 Comprimento de plântulas

Foi realizado em conjunto com o teste de germinação, onde após oito dias da semeadura, ocorreu a medição do comprimento de dez plântulas por repetição com o auxílio de régua graduada em milímetros. O comprimento médio foi obtido somando-se as medidas de cada repetição e dividido pelo número de plântulas avaliadas.

5.2.1.3 Massa seca de plântulas

As plântulas foram colocadas em sacos de papel e levadas para a estufa com circulação de ar forçada à temperatura de 65°C até atingir peso constante. O material seco será pesado em balança com precisão de 0,001g e o resultado dividido pelo número de plântulas normais.

5.2.1.4 Teste de emergência

No teste de emergência de plântulas, foram utilizadas quatro repetições de 50 sementes por lote. A semeadura ocorreu de forma manual em canteiros de 10 x 1m, com 10 cm entre repetições, contendo substrato terra/areia na proporção 1:1. As sementes foram dispostas de maneira equidistante em sulcos com 3cm de profundidade no canteiro. Diariamente foram feitas observações e irrigações sempre que necessário, visando o fornecimento de água para a germinação das sementes e emergência das plântulas e a partir da emergência da primeira plântula, iniciou-se a contagem. O teste foi avaliado aos dez dias após a semeadura, sendo computada a porcentagem de plântulas. Ao final da contagem os dados foram transformados em porcentagem conforme Nakagawa (1999).

Foi avaliado o Índice de Velocidade de Emergência (IVE), que é capaz de estimar o número de dias médio necessários para ocorrer a emergência e o número de plantas emergidas por dia. Os índices foram calculados seguindo a metodologia proposta por Maguire (1962), em que: $IVE = (G1/N1) + (G2/N2) + \dots + (GN/N)$, no qual, G = número de plântulas normais computadas nas contagens; N = número de dias da semeadura do primeiro ao último dia de contagem.

5.2.2 Teste de Tetrazólio

Dez lotes de sementes de feijão caupi foram submetidos ao teste de tetrazólio e ao final foram realizadas a captura das imagens.

Foram utilizadas 4 repetições de 50 sementes por lote, que foram pré-condicionadas em copos descartáveis com 50mL de água destilada em B.O.D no período de hidratação de 16 horas a temperatura de 25°C, dessa forma com a pré

hidratação acontece a ativação da respiração e amolecimento dos tecidos com finalidade de facilitar os cortes e a absorção do sal de tetrazólio.

Após hidratação as sementes foram submersas em solução de tetrazólio na concentração de 0,050% por um período de três horas em B.O.D na temperatura de 42°C no escuro. Esse processo é realizado no escuro, devido o sal de tetrazólio ser sensível a luz, podendo ser reduzido, ocasionando uma coloração indevida das sementes (BRASIL, 2009).

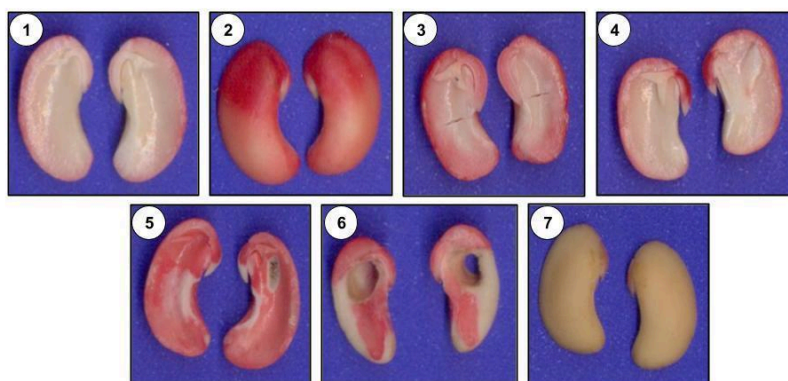
Finalizado o período de coloração as sementes são retiradas da B.O.D., lavadas em água corrente e mantidas em água até o final da avaliação, para que a turgidez seja mantida (Brasil, 2009). Para a interpretação dos resultados de forma manual as sementes foram seccionadas longitudinalmente ao longo do eixo embrionário e os tegumentos removidos, na sequência analisadas individualmente interna e externamente quanto os danos (mecânico, por insetos e pela umidade), a coloração e a consistência dos tecidos.

No momento da avaliação, as características de intensidade de cor são de suma importância, permitindo a diferenciação para a classificação das sementes em relação ao vigor e viabilidade, sementes com coloração vermelho carmim apresentam tecidos vivos e vigorosos, sementes de coloração vermelho carmim forte demonstram tecidos em deterioração e sementes com cor branco leitoso apresentam tecidos mortos.

Para a caracterização dos níveis de viabilidade no feijão, usa-se 7 classes onde as classes de 1 a 3 correspondem às sementes viáveis e vigorosas, as classes 4 e 5 correspondem às sementes viáveis, porém não vigorosas e as classes 6 e 7 correspondem às sementes não viáveis.

Os resultados da interpretação da análise manual das sementes (Figura 1) foram expressos de forma binária na qual 0 é para as classes de 1-3 e 1 para as classes de 4-7, essa forma de classificação foi necessária para a utilização dos dados iniciais na classificação automatizada.

Figura 33 – Classes de viabilidade e vigor de sementes de feijão-caupi.



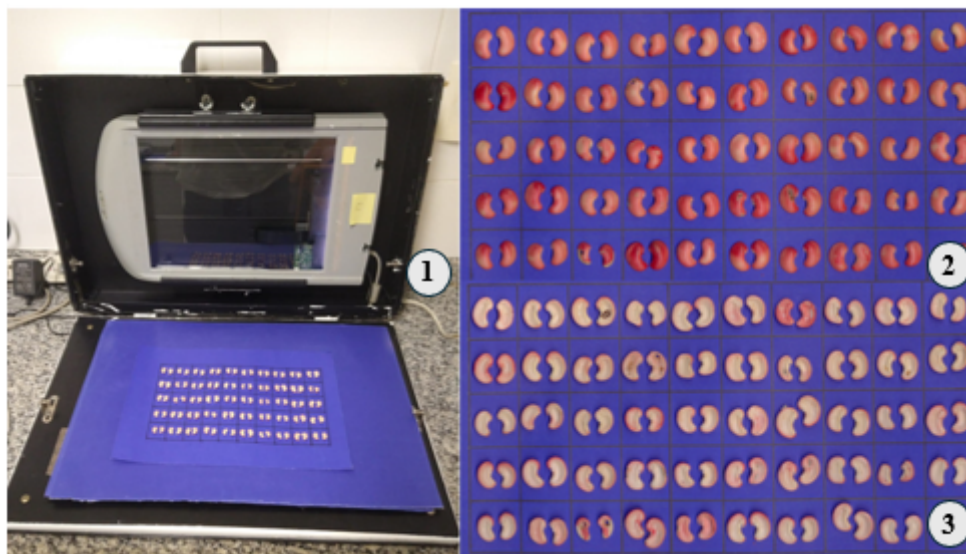
Fonte: Elaborado pela autora . Legenda Classes de viabilidade e vigor de sementes de feijão. (1) Alto vigor; (2) Médio vigor; (3) Baixo vigor; (4) Baixo vigor (5) Muito baixo vigor; (6) Não viável e (7) Morta.

O delineamento experimental utilizado foi o inteiramente casualizado (DIC). Os resultados dos testes de germinação e vigor, bem como o teste de tetrazólio foram submetidos à análise de variância e as médias submetidas ao teste Tukey em até 5% de probabilidade. Também foi realizada uma correlação de Pearson entre as variáveis estudadas pelos testes de vigor e o tetrazólio. As análises estatísticas foram realizadas por meio do software R 4.3.1 (R Core Team, 2020).

5.2.3 Aquisição das imagens

As sementes após realização do teste de tetrazólio foram distribuídas de forma equidistantes em folhas de papel de coloração azul, de maneira que haja alto contraste com a coloração após pigmentação com o sal de tetrazólio nas sementes, para facilitar a identificação de suas partes e realização de análise comparativa e foram utilizadas 2 repetições para cada lote, essas repetições primeiro receberam a classificação manual para depois ser realizada a captura. As imagens foram capturadas por meio de um escâner modelo HP Scanjet 2004, adaptado de maneira invertida dentro de uma caixa de alumínio, digitalizadas em 300 dpi, salvas em formato JPG e armazenadas para posterior análise (Figura 2).

Figura 34 – Aquisição das imagens do teste de tetrazólio.



Fonte: Elaborado pela autora. Legenda: 1- Scanner para a captura das imagens; 2- Sementes de feijão-caupi após o teste de tetrazólio face axial e 3- Sementes de feijão-caupi após o teste de tetrazólio face abaxial.

5.2.3.1 Visão geral do TETRASEED

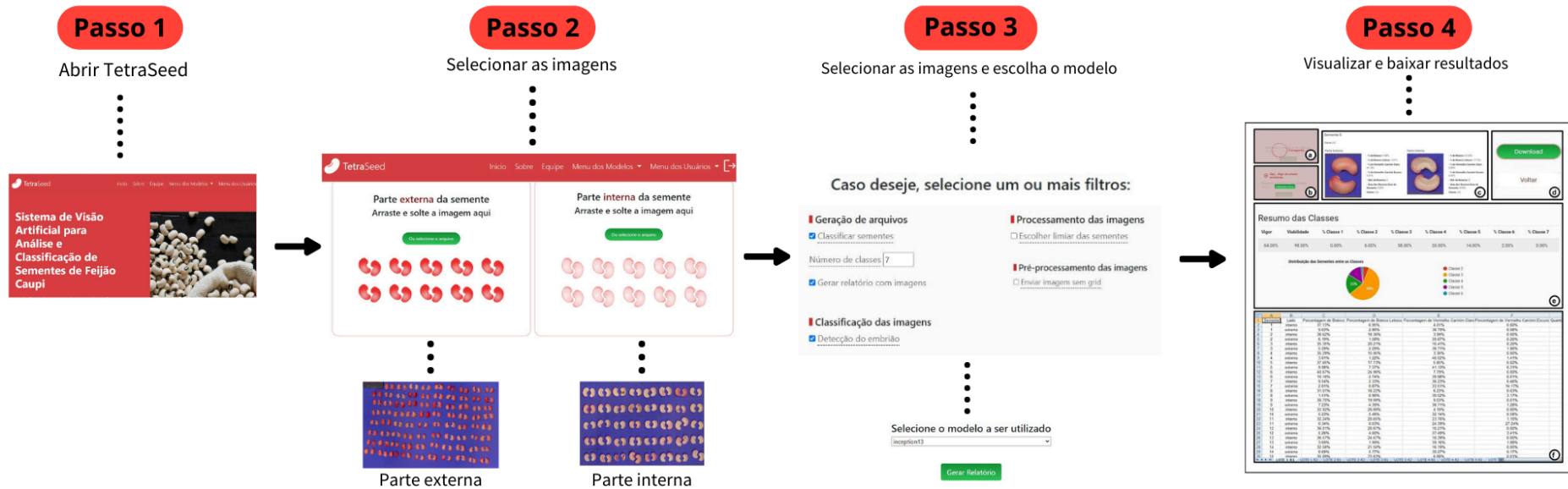
O TETRASEED é um sistema de visão artificial para análise e classificação de sementes de feijão caupi após a realização do teste de tetrazólio (Figura 3). Visa analisar as imagens das sementes e gerar os resultados em porcentagens de cor de vermelho carmim claro, vermelho carmim escuro, branco leitoso e classificar de forma automatizada as sementes com base na literatura atual para o teste de tetrazólio de feijão caupi (França Neto, 1999).

Desenvolvido em Python com Flask, utiliza Python como a linguagem de programação para desenvolver a aplicação e Flask como o framework que gerencia o lado servidor da aplicação. Com o intuito de não dificultar a comunicação entre a interface de interação com o usuário e os demais módulos do sistema como a aquisição, o processamento, a classificação e o teste. O FrontEnd da interface foi construído com HTML, CSS e JavaScript para facilitar na construção do software e manutenções futuras.

A interface do software conta somente com a tela inicial que é modificada para

exibir o status do processo, desde quando as imagens são carregadas, durante o processamento e após a conclusão do processamento, quando os resultados estão disponíveis para download.

Figura 35 – Resumo do fluxo de operação do TETRASEED.



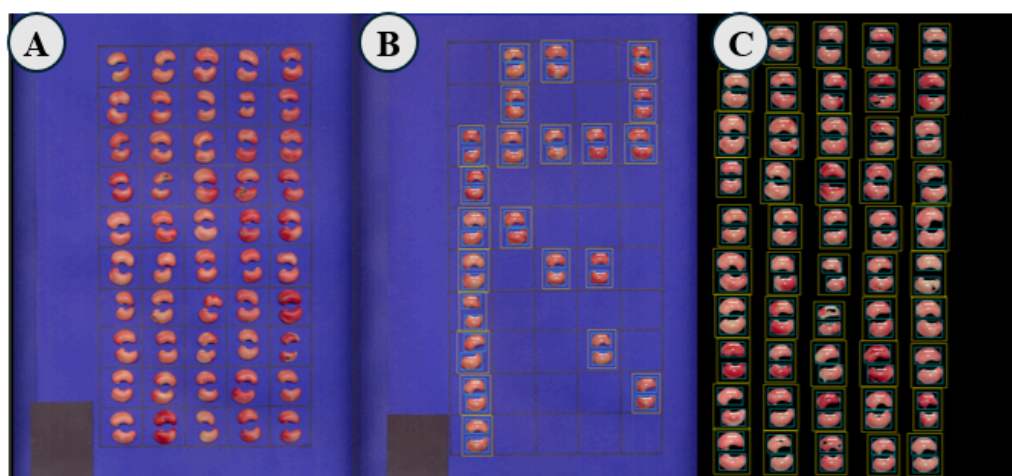
Fonte: Elaborado pela autora.

Para executar o software, o usuário deve acessar a página inicial, inserir o e-mail previamente cadastrado e a senha na área de login. Após o login, será redirecionado para o menu inicial, onde encontrará a descrição do software, informações sobre a equipe idealizadora do projeto, e funções de trabalho intuitivas. Não é necessário instalar nenhuma adaptação adicional em computadores ou notebooks para o funcionamento da ferramenta.

Após abrir o software (Passo 1), o usuário será encaminhado para o menu (Passo 2), onde a interface permite o upload das imagens a serem analisadas pelo software. O usuário pode fazer upload de duas imagens por vez: uma para avaliação externa e outra para avaliação interna da semente. Para garantir que o envio seja feito corretamente, a interface aceita apenas o upload de duas imagens simultaneamente; o botão de upload é liberado somente quando ambas as imagens são selecionadas e bloqueado durante o processo de upload e análise. As imagens selecionadas passam para a próxima etapa (Passo 3), onde o usuário pode escolher o número de classes para a avaliação, optar pela geração de relatórios com os dados das sementes, e verificar preliminarmente a melhor limiarização para o lote analisado, com base na colorimetria na escala de vermelho. O pré-processamento é então aplicado automaticamente à imagem.

Além disso, a interface oferece a opção de envio da imagem sem a grade de divisão das sementes, disponível apenas para imagens em que as sementes não estejam tocando umas nas outras. A separação é feita de forma automatizada, mas para que essa função funcione corretamente, deve haver uma separação clara entre cotilédones e sementes na imagem enviada (Figura 4). Essa etapa é finalizada com a opção de selecionar uma rede neural ou fazer o upload de outra.

Figura 36 – Esquema da separação automatizada das sementes seguidas da segmentação da imagem.



Fonte: Elaborado pela autora. Legenda: A- Sementes dispostas para a aquisição da imagem; B- Separação automatizada das sementes e C- Segmentação.

Ao selecionar a opção para iniciar a análise, o Passo 4 é ativado. Nesta etapa, o software realiza uma análise de alta capacidade de todas as sementes contidas no diretório de entrada. As sementes são classificadas individualmente, gerando dados percentuais sobre a classificação geral da imagem analisada, incluindo um gráfico e a porcentagem final de vigor e viabilidade das sementes. Nesta etapa, um documento Excel é gerado automaticamente, podendo ser baixado pelo usuário para posterior análise.

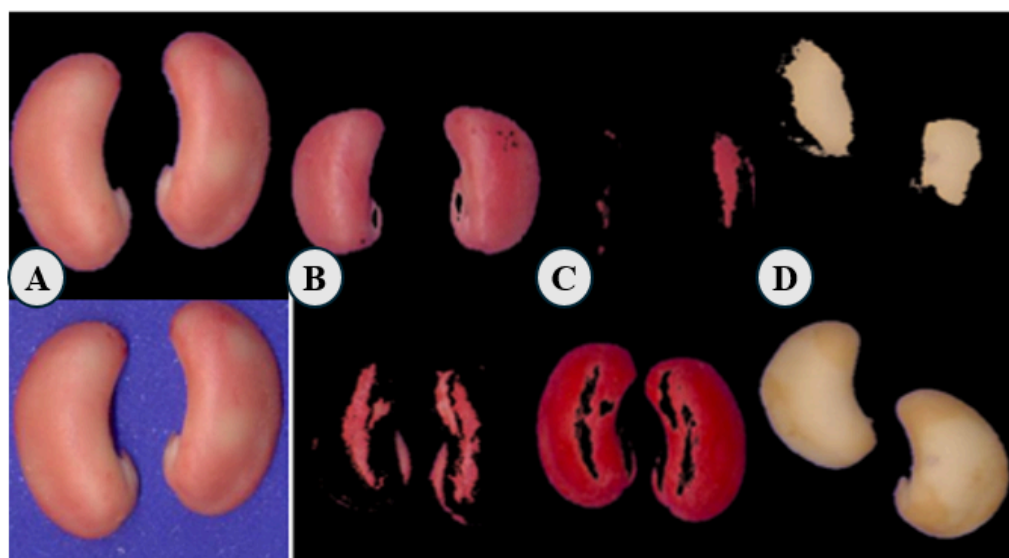
A maioria dos principais passos do TETRASEED são interligados, ou seja, ao final de cada processo, o usuário pode ser redirecionado para a opção anterior e pode escolher outro caminho, sem precisar começar novamente.

5.2.3.2 Processamento

Na primeira etapa do processamento temos a remoção do fundo, que consiste na extração dos pixels da semente Figura 5A. Para isso, a imagem é convertida para o formato de representação RGB. Para a extração das cores vermelhas das sementes foi utilizado o sistema de representação de cores HSV (Hue (Matiz), Saturation (Saturação) e Value (Valor ou Brilho)), que foi constatado que a informação desejada para este problema estava codificada no canal H, que representa a matiz da cor (Figura 5 B-C). O processo de extração do branco leitoso foi semelhante ao do vermelho carmim. Contudo, a informação necessária não estava

codificada apenas em um canal, como no caso do vermelho carmim, mas sim no sistema de representação RGB (Figura 5D).

Figura 37 – Fases da segmentação e processamento das imagens nos canais RGB.



Fonte: Elaborado pela autora. Legenda: A- Remoção do fundo; B- Extração do vermelho carmim claro; C- Extração do vermelho carmim escuro e D- Extração do branco leitoso.

5.2.3.3 Classificação

Para a classificação das sementes a rede Inception-ResNet v2 foi utilizada, ela é uma arquitetura para rede neural profunda, que foi desenvolvida para tarefas de VC, como a classificação de imagens. Para treinamento da rede, foi utilizada para cada um dos lotes uma das 4 imagens obtidas pelo scanner. Os valores dos pixels foram normalizados para a faixa [0, 1] para garantir a uniformidade dos dados e melhorar a convergência durante o treinamento. Técnicas de aumento de dados (Data Augmentation), como rotação, tradução, espelhamento e zoom, foram aplicadas para aumentar a diversidade do conjunto de treinamento e evitar o overfitting.

Os dados manuais foram inseridos na camada inicial da CNN e nos pesos das camadas intermediárias, visando treinar a rede por meio de dados de entrada oriundos da classificação manual. Foram realizados por 50 épocas para o treinamento:

(i) Os dados foram divididos em conjuntos de treinamento (70%), validação (20%) e teste (10%);

(ii) Ajustamos as CNNs no conjunto de treinamento, usando o conjunto de validação para evitar overfitting;

(iii) Usamos validação cruzada estratificada de 10 vezes no conjunto de treinamento e validação combinados, para treinar os cinco algoritmos de classificação;

(iv) Finalmente avaliamos a classificação realizada no conjunto de teste.

Métricas de desempenho utilizadas para avaliar o desempenho do modelo foram acurácia, precisão, recall e F1-score. Foram utilizadas as métricas Acurácia (Equação 1), precisão (Equação 2), Recall (Equação 3) e F-Measure (Equação 4) para avaliar a eficiência da rede, calculadas a partir das Equações (1, 2, 3 e 4).

A acurácia é a proporção de predições corretas (tanto verdadeiros positivos quanto verdadeiros negativos) sobre o total de instâncias avaliadas.

$$\text{Equação 1: Acurácia} = \frac{TP+TN}{TP+TN+FP+FN}$$

- TP (Verdadeiro Positivo): Casos verdadeiros positivos.
- TN (Verdadeiro Negativo): Casos verdadeiros negativos.
- FP (Falso Positivo): Casos falsos positivos.
- FN (Falso Negativo): Casos falsos negativos.

A precisão (ou valor preditivo positivo) é a proporção de predições positivas corretas sobre o total de predições positivas feitas pelo modelo.

$$\text{Equação 2: Precisão} = \frac{TP}{TP+FP}$$

Alta precisão indica que a maioria das instâncias classificadas como positivas são realmente positivas.

O recall (ou sensibilidade) é a proporção de predições positivas corretas sobre o total de instâncias positivas reais.

Equação 3:
$$Recall = \frac{TP}{TP+FN}$$

Alto recall indica que a maioria das instâncias positivas reais são corretamente identificadas pelo modelo.

O F1-score é a média harmônica entre a precisão e o recall, proporcionando um balanço entre os dois. É especialmente útil quando há um desequilíbrio entre as classes.

Equação 4:
$$F1 - Score = 2x \frac{Precisão \times Recall}{Precisão + Recall}$$

Um F1-score alto indica que o modelo tem tanto alta precisão quanto alto recall.

Técnicas estatísticas de validação cruzada foram aplicadas para avaliar a capacidade de generalização de um modelo de aprendizado de máquina. Ao invés de simplesmente dividir os dados em um conjunto de treinamento e um conjunto de teste, a validação cruzada permite uma avaliação mais robusta e confiável do desempenho do modelo para garantir a generalização do modelo e evitar o overfitting.

5.3 Resultados e Discussão

5.3.1 Testes tradicionais

A avaliação inicial do vigor (Tabela 4) caracterizou os lotes estudados além de permitir correlacionar com os dados do teste de tetrazólio, validando as informações obtidas.

Tabela 4 – Resultados de primeira contagem - PC; plântulas anormais- PA; comprimento de parte aérea - CPA; comprimento de raiz - CRA; matéria seca de parte aérea - MSPA; matéria seca de raiz - MSRA; índice de velocidade de emergência - IVE obtidos em 10 lotes de feijão-caupi.

Lote	PC (%)	G (%)	PA (%)	CPA (cm)	CRA (cm)	MSPA (g)	MSR A (g)	IVE
1	74,00b	74,00b	55,00ab	9,25cd	6,62c	0,697bc	0,075 b	1,55ns
2	95,50a	95,50a	82,00a	8,56d	8,62bc	0,522cd	0,087 b	1,77ns
3	89,50ab	89,50ab	8,5d	18,95a	13,40ab	0,392d	0,085 b	1,57ns
4	99,00a	99,00a	55,50ab	10,75bcd	13,02ab	0,960b	0,117 ab	1,77ns
5	92,00a	92,00a	23,00cd	18,46a	14,34ab	0,710bc	0,122 ab	1,55ns
6	89,00ab	89,00ab	40,50bc	12,99b	11,19abc	1,267a	0,090 b	1,67ns
7	90,00a	90,00a	46,50bc	12,04bc	11,66abc	0,962b	0,115 ab	1,60ns
8	89,00ab	89,00ab	48,50bc	13,54b	16,12a	0,672cd	0,167 a	2,59ns
9	94,50a	94,50a	28,50bcd	14,17b	15,84a	0,640cd	0,152 a	1,65ns
10	88,00ab	88,00ab	33,00bcd	18,34a	13,90ab	0,562cd	0,112 ab	1,91ns
CV(%)	7,18	7,18	28,79	10,46	20,16	15,79	21,00	15,32

Médias seguidas da mesma letra minúscula na coluna não diferem pelo teste Tukey a 5% de probabilidade.

Fonte: Elaborado pela autora.

A porcentagem de G% de todos os lotes avaliados foi superior ao valor mínimo de 70% estabelecido para a comercialização de sementes dessa espécie no Brasil (Brasil, 2009),

o que indica uma alta viabilidade dos lotes. No entanto, observou-se diferença na germinação entre os lotes. Os lotes 4, 2, 9, 5 e 7 apresentaram as maiores médias para as variáveis de comprimento de plântula (PC) e G, com valores de 99,00%, 95,50%, 94,50%, 92,00% e 90,00%, respectivamente. Por outro lado, o lote 1 apresentou a menor média de PC, com 74,00%. O teste mais comumente utilizado para avaliar o vigor é o teste de germinação..

Na variável porcentagem PA, o lote 2 apresentou a maior média, com 82,00%. Em contrapartida, o lote 3 teve a menor média para essa variável, com 8,5%. A baixa taxa de plântulas anormais sugere que o lote 3 possui melhores condições para o desenvolvimento no campo.

No teste de CPA, os lotes 3, 5 e 10 se destacaram com as maiores médias, apresentando 18,95 cm, 18,46 cm e 18,34 cm, respectivamente, o que indica um bom desenvolvimento da parte aérea das plântulas. Em contraste, o lote 2 apresentou a menor média, com apenas 8,56 cm.

No teste de CRA, os lotes 8 e 9 se destacaram, com médias de 16,12 cm e 15,84 cm de comprimento de raiz, respectivamente. Por outro lado, o lote 1, que teve a menor média com 6,62 cm, pode enfrentar dificuldades no estabelecimento inicial da planta, o que pode comprometer o rendimento final.

Em relação à MSPA, o lote 6 obteve a maior média, com 1,267 g, enquanto o lote 3 apresentou a menor média, com 0,392 g. Para a Massa Seca da Raiz (MSRA), os lotes 8 e 9 se destacaram, com médias de 0,167 g e 0,152 g, respectivamente. Os lotes 1, 3, 2 e 6 apresentaram as menores médias para essa variável, com valores de 0,075 g, 0,085 g, 0,087 g e 0,090 g, respectivamente.

Por fim, na variável IVE, não foram observadas diferenças estatísticas entre os lotes.

A caracterização dos lotes de sementes por meio dos testes de vigor é fundamental para identificar aqueles que, apesar de apresentarem alta germinação, podem ter desempenho inferior em condições de campo ou em outros testes de vigor (Gough, 2020). Essa avaliação é especialmente importante na agricultura, onde as sementes enfrentam diversas condições edafoclimáticas, tornando a uniformidade de emergência crucial para o sucesso da lavoura (Savage; Bassel, 2016). Lotes de sementes com alto vigor tendem a produzir plântulas vigorosas e uniformes, o que é essencial para uma competição eficiente por recursos e,

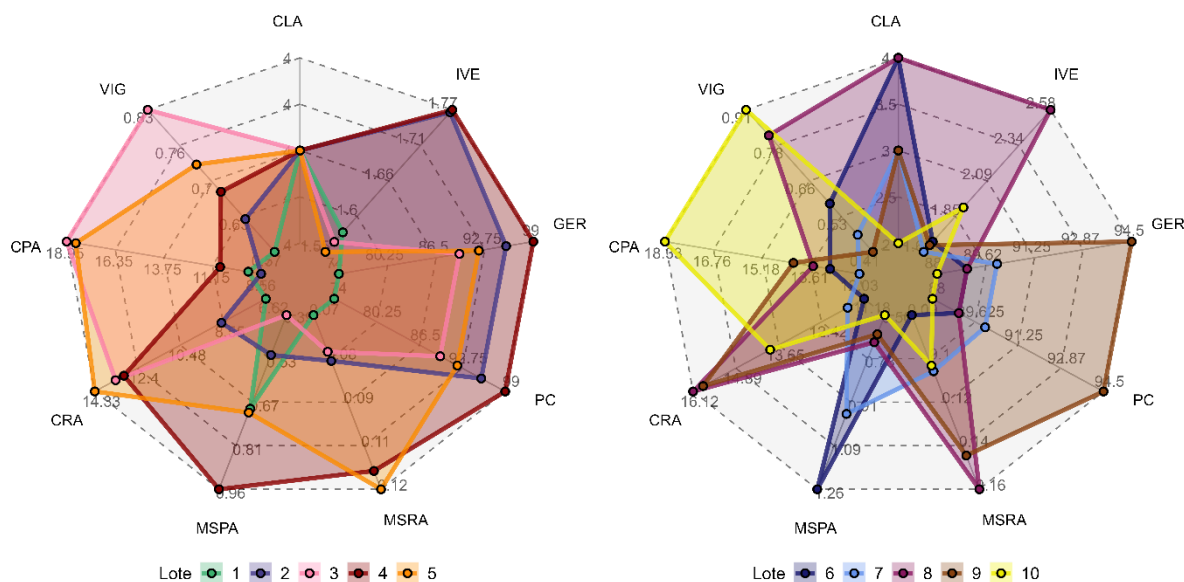
consequentemente, para uma maior produtividade (Marcos Filho, 2015). A avaliação inicial do vigor foi importante para correlacionar com os dados do teste de tetrazólio e validar as informações obtidas.

5.3.2 Teste de Tetrazólio e aquisição de imagens

A metodologia utilizada no teste de tetrazólio demonstrou-se eficiente para a sua correta aplicação, permitindo classificar as sementes em sete categorias com base na coloração gerada pelo sal, além de fornecer informações sobre o vigor dos lotes. A eficácia do teste de tetrazólio está intimamente relacionada à padronização da metodologia, com a adaptação das condições ideais para o pré-condicionamento, preparo e coloração específicas de cada espécie (Cervi; Mendonça, 2009). A execução adequada da metodologia é crucial, pois a coloração obtida reflete diretamente o estado fisiológico das sementes, possibilitando a diferenciação entre sementes vigorosas e aquelas viáveis, mas com menor capacidade de emergência ou desenvolvimento inicial.

Os resultados do teste de tetrazólio proporcionaram não apenas dados quantitativos, mas, sobretudo, informações qualitativas, permitindo a classificação dos lotes e a avaliação do vigor e viabilidade de cada um. Além disso, foi possível correlacionar os dados com os resultados de outros testes realizados nos dez lotes (Figura 6), evidenciando a confiabilidade do teste. A correlação entre diferentes testes de vigor e métodos de avaliação é fundamental para garantir a consistência e robustez das avaliações da qualidade das sementes (Perry, 1981). Em sistemas de visão computacional, a validação dos dados de entrada por meio da correlação com testes de vigor é essencial para assegurar que os algoritmos funcionem corretamente e que as classificações automáticas estejam alinhadas com as avaliações tradicionais (Tian *et al.*, 2020).

Figura 38 – Gráfico Radar entre as variáveis IVE, G, PC, MSRA, MSPA, CRA, CPA e VIGOR com as classes obtidas pelo teste de tetrazólio para os dez lotes.



Fonte: Elaborado pela autora.

A análise de dados apresentada na Figura 6 ressalta a relevância das classes 1, 2 e 3 para a qualidade final dos lotes de sementes, sugerindo que estas classes estão relacionadas a lotes com maior vigor, em contraste com as classes 4, 5, 6 e 7. Esses resultados indicam que lotes classificados nas três primeiras categorias possuem maior chance de estabelecer-se bem em campo, confirmando a correlação entre a classificação obtida pelos testes de vigor e o desempenho das sementes em condições reais de cultivo. Na prática, pode-se afirmar que sementes de alta qualidade (classes 1, 2 e 3) apresentam maior potencial para gerar plântulas vigorosas, com melhor estabelecimento em campo, sendo menos suscetíveis a estresses bióticos e abióticos, o que proporciona maiores ganhos ao agricultor (Finch-Savage; Bassel, 2016).

Outro ponto relevante mostrado na Figura 6 é a correlação das classes do teste de tetrazólio com os demais testes (IVE, G, PC, MSRA, MSPA, CRA, CPA e Vigor) dos 10 lotes analisados. Os resultados do teste de tetrazólio confirmam as observações dos outros testes de vigor, evidenciando a consistência e a confiabilidade desse método na avaliação da qualidade das sementes. Essa concordância entre diferentes metodologias reforça a eficiência do teste de tetrazólio como uma ferramenta complementar valiosa para a análise da viabilidade e vigor das sementes.

Conforme destacado por Silva *et al.*, (2013), a utilização de concentrações mais baixas de tetrazólio pode ser uma estratégia eficaz para otimizar recursos em laboratórios de análise de sementes. Isso é particularmente relevante em contextos em que a redução de custos e o aumento da eficiência são fundamentais. A adoção de metodologias mais econômicas permite que mais amostras sejam analisadas sem comprometer a precisão dos resultados, o que é essencial para a tomada de decisões informadas na seleção de lotes de sementes.

Estudos de viabilidade e vigor dos mostram que o teste é um indicador confiável da qualidade fisiológica das sementes, especialmente quando utilizado em conjunto com outros testes de vigor e germinação (Fogaça *et al.*, 2006; Ribeiro *et al.*, 2010 e Rodrigues *et al.*, 2015. Fornecendo informações rápidas e precisas sobre a viabilidade das sementes, fato esse que o torna uma ferramenta indispensável nos laboratórios de análise de sementes, especialmente em situações que é necessário avaliar muitas amostras em um curto período.

A etapa seguinte é a aquisição de imagens que desempenham um papel vital na construção de sistemas de VC. A aquisição das imagens foi eficiente para a construção de um banco de dados robustos, influenciando de maneira positiva na etapa de processamento de imagens, além de ser uma aquisição reprodutível e de fácil acessibilidade, o que contribuem para a eficácia de modelos de aprendizado de máquina, assegurando que a análise seja consistente e reprodutível.

O processo de aquisição é fundamental para a análise de imagem, pois determina a precisão e a confiabilidade dos resultados subsequentes. As imagens com alta qualidade capturam detalhes essenciais, os ruídos são minimizados e garantem uma segmentação precisa, reduzindo a necessidade de processamento adicional e diminuindo a chance de erros (Dodge; Karam, 2016). Imagens bem adquiridas também aceleram o processamento e tornam a análise mais eficiente, sendo, portanto, uma etapa crucial para o sucesso em aplicações de VC e outras áreas relacionadas (Alginahi, 2010).

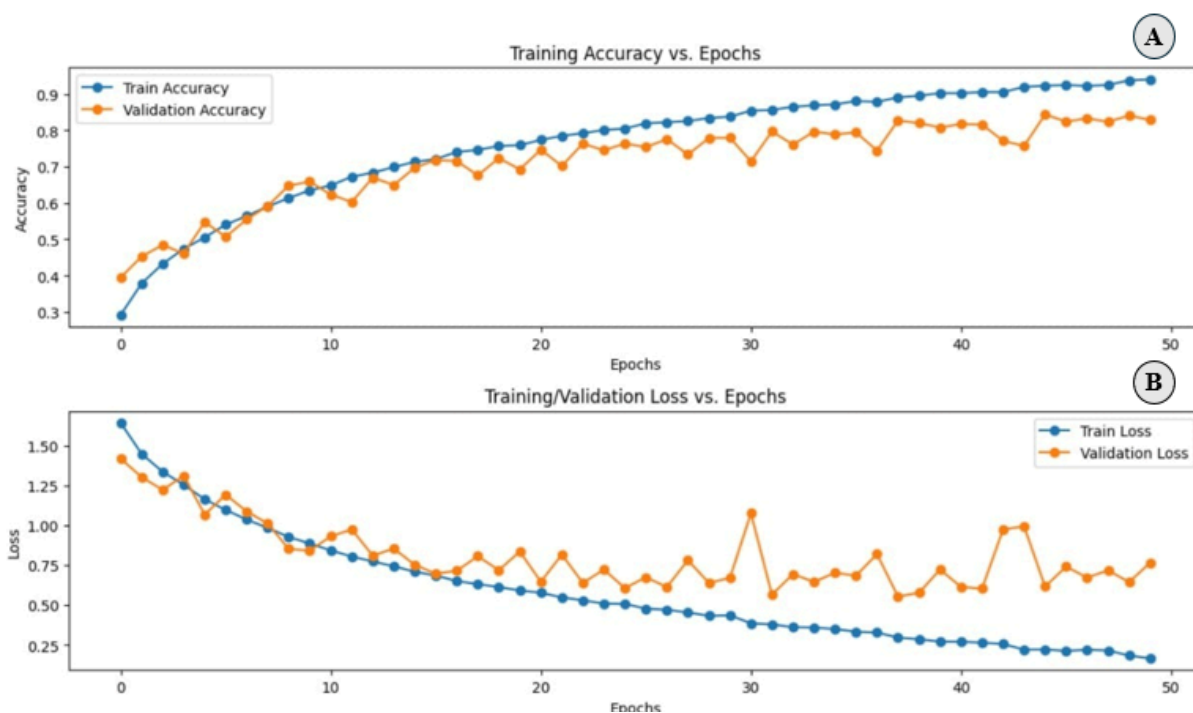
O processamento de imagens acontece em cinco etapas: a obtenção das imagens, o pré-processamento, a segmentação, o reconhecimento e a interpretação (Gonzales; Woods, 2000), portanto o ponto inicial para uma análise se imagem adequada acontece na aquisição.

5.3.3 TETRASEED

A criação do sistema de VC começou com a introdução dos dados qualitativos do teste de tetrazólio, validados pela correlação com os resultados dos testes de vigor. Os dados qualitativos de entrada do teste de tetrazólio deram início ao treinamento da rede. Os dados foram processados em várias camadas da rede, extraídas características importantes automaticamente. Os pesos iniciais foram calculados de maneira aleatória, após passar pela rede a saída é gerada e comparada com as classificações reais. A diferença entre a saída prevista e a real (o erro) é então propagada de volta pela rede, ajustando os pesos através de um algoritmo de otimização. Esse ciclo de ajuste e validação se repete ao longo de várias épocas, onde cada época corresponde a uma passagem completa por todo o conjunto de dados. Durante as épocas, o desempenho da rede foi monitorado (Figura7), com o objetivo de minimizar a perda e melhorar a capacidade da CNN de generalizar para novos dados.

Durante o treinamento da CNN, foi construído um gráfico para acurácia e perda em função das épocas de treinamento (Figura 7), para monitorar o desempenho do modelo e para garantir que ele esteja aprendendo de forma eficaz e generalizando bem para novos dados. Além de auxiliar na tomada de decisões sobre ajustes no processo de treinamento, como a modificação da taxa de aprendizado ou a implementação de regularização.

Figura 39 – Gráfico para Acurácia e Perda em função das épocas de treinamento.



Fonte: Elaborado pela autora. Legenda: A- Gráfico para acurácia em função das épocas de treinamento e B- Gráfico para acurácia em função das épocas de treinamento.

O eixo “x” é determinado pelo número de épocas, cada época é uma rodada de treinamento da CNN, começado na época zero até a 50, no eixo “y” temos o aumento da acurácia na figura 7A e o aumento da perda na Figura 7B.

Na figura 7A temos a acurácia em função da época de treinamento, como era esperado, a acurácia vai aumentando com o número de épocas de treinamento, seja para os dados de treinamento ou validação cruzada. Esse comportamento é esperado e está alinhado com o conceito fundamental de aprendizado supervisionado, onde o modelo ajusta seus parâmetros para minimizar a função de perda e melhorar a performance na tarefa de classificação.

Com o aumento do número de épocas a CNN tem mais oportunidades de ajustar seus pesos de acordo com os dados de treinamento, dessa forma o modelo consegue durante o processo aprender sobre padrões e características complexas nos dados, resultando em um aumento na acurácia, que é um indicativo de que o modelo está sendo eficaz para capturar

padrões do conjunto de dados. Importante salientar, que a acurácia quando segue um padrão contínuo dentro do gráfico pode sinalizar que o modelo está começando a se ajustar demais (overfitting) aos dados de treinamento, especialmente se a acurácia dos dados de validação não acompanhar o mesmo ritmo.

O overfitting ocorre quando um modelo de aprendizado de máquina profundo como as CNNs, que operam com dados complexos, o modelo se ajusta excessivamente aos dados de treinamento, capturando não apenas os padrões reais, mas também o ruído e as variações aleatórias presentes nos dados, resultando em uma performance excelente nos dados de treinamento, mas uma capacidade reduzida de generalização para novos dados (Goodfellow, 2016). A validação cruzada é uma estratégia eficaz para mitigar esse problema.

A validação cruzada, uma técnica que avalia a capacidade de generalização do modelo, muitas vezes mostra uma melhoria mais gradual na acurácia em comparação ao treinamento, sugerindo que o modelo não está apenas memorizando, mas aprendendo a generalizar a partir dos dados (Berrar, 2019.).

Na Figura 7A conseguimos comprovar um comportamento esperado que é a acurácia em dados de treinamento aumentando mais rapidamente nas primeiras épocas e, eventualmente, se estabilizando. Se a acurácia de validação começar a se estabilizar ou até a diminuir, pode ser um sinal de que o modelo está começando a sofrer de overfitting, o que sugere a necessidade de técnicas de regularização ou ajustes no treinamento, fato esse que não aconteceu para a CNN escolhida, demonstrando a eficiência dela na análise dos dados (Hawkins, 2004).

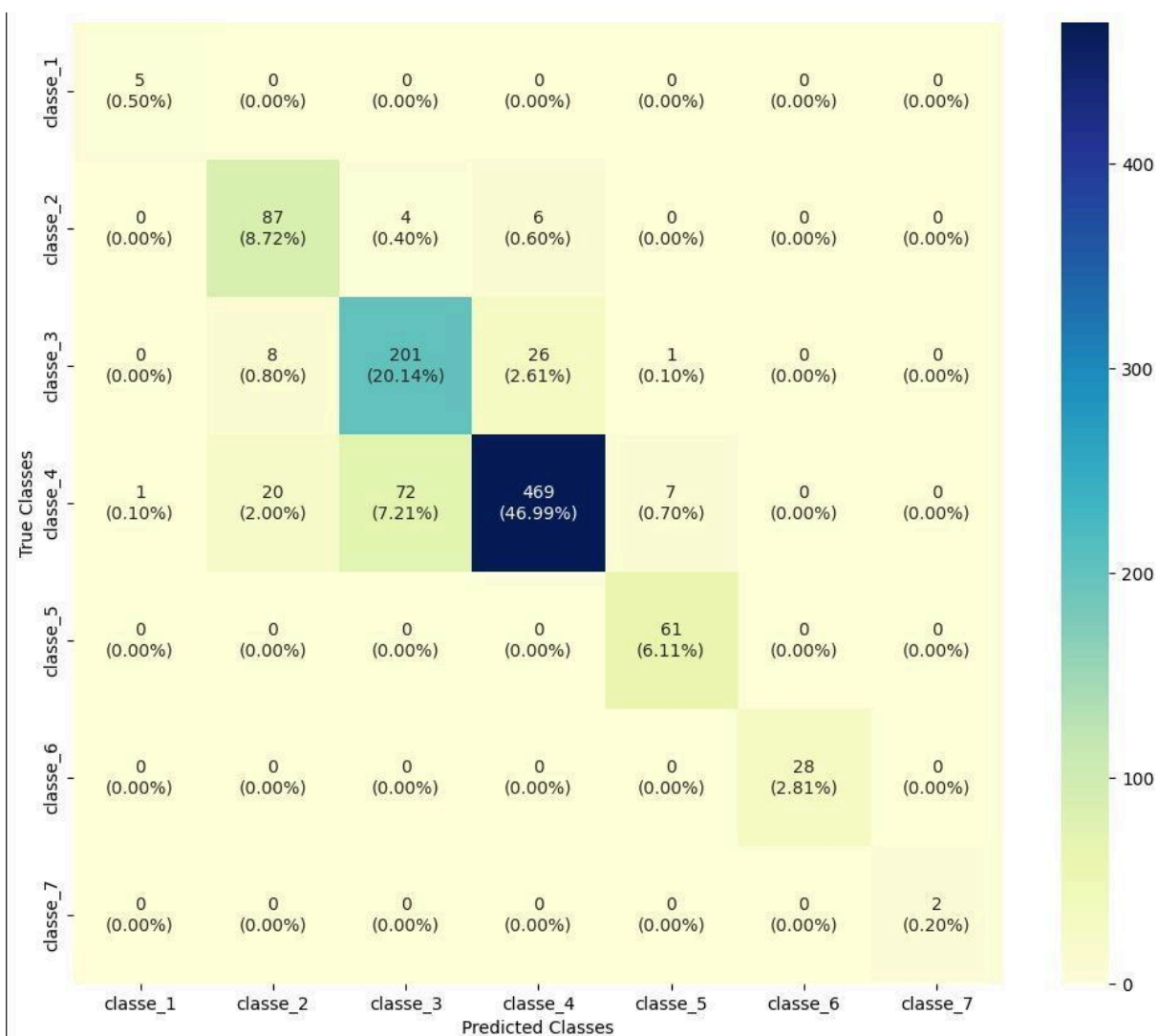
Na Figura 7B, o gráfico representa a função de Loss (Função de perda), que mede o quanto as previsões do modelo diferem dos valores reais ou esperados. Neste estudo, utilizou-se a função Cross-entropy, uma métrica amplamente adotada em tarefas de classificação por sua eficácia em quantificar a diferença entre a distribuição de probabilidade prevista e a distribuição verdadeira (Jadon, 2020). Essa métrica é especialmente útil para otimizar modelos em problemas de classificação multiclasse, onde a precisão das previsões é crucial para o desempenho global (Jadon, 2020). A função também incentiva o modelo a concentrar-se nas classes corretas, aplicando penalidades maiores às previsões incorretas, o

que tem se mostrado eficaz em aprimorar a capacidade do modelo de aprender a partir dos dados fornecidos (Grandini *et al.*, 2020; Valverde-Albacete; Peláez-Moreno, 2010).

O esperado é que a função de Loss diminua com o aumento das épocas, pois o objetivo do treinamento é minimizar a perda, ajustando os pesos da rede para que as previsões se tornem cada vez mais precisas. Esta formulação incentiva o modelo a gerar probabilidades mais altas para a classe correta, pois minimizar a perda durante o treinamento fará com que a distribuição de probabilidade prevista pelo modelo convergindo para a distribuição verdadeira (Janocha; Czarnecki, 2017).

Após o treinamento da rede, foi construída uma matriz de confusão (figura 8), tendo-se as classificações preditas pela rede em função das classificações visuais.

Figura 40 – Matriz confusão para as classificações preditas pela rede em função das classificações visuais.



Fonte: Elaborado pela autora.

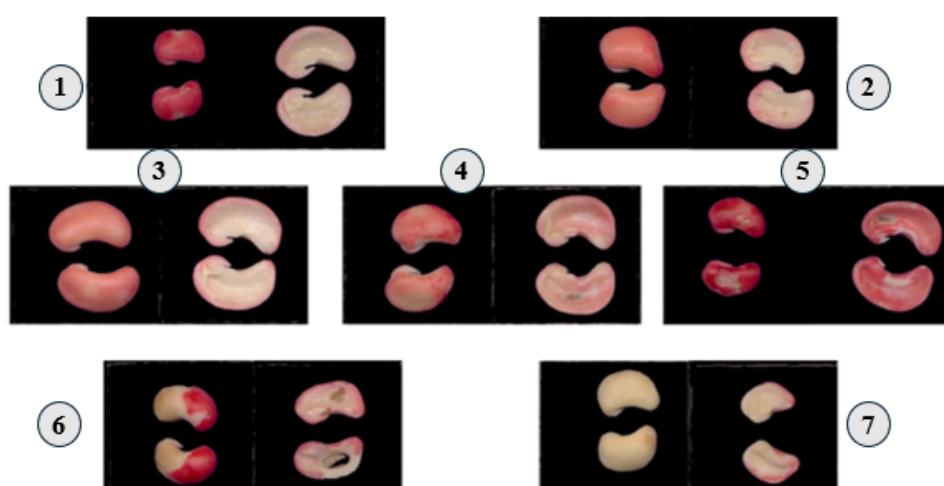
A matriz confusão é uma forma de avaliação utilizada principalmente em problemas de classificação para comparar as previsões feitas por um modelo com as verdadeiras classificações dos dados. Os resultados são apresentados em uma tabela que mostra a quantidade de observações classificadas corretamente e incorretamente para cada classe.

Na Figura 8, observamos um ponto crítico na comparação entre as classificações preditas pela rede e as classificações visuais. Este ponto de interseção ocorre entre a classe 3, que é avaliada como vigorosa, e a classe 4, que, conforme a literatura, é classificada apenas como viável, as diferenças entre as classes 3 e 4 refletem variações no vigor das sementes.

Sementes classificadas como classe 3 são consideradas vigorosas, indicando não apenas viabilidade, mas também um potencial de germinação e desenvolvimento. Em contraste, sementes da classe 4 são viáveis, o que significa que têm a capacidade de germinar, mas não apresentam o mesmo nível de vigor ou desempenho potencial no campo que as da classe 3 (França Neto; Krzyzanowski; Costa, 1998).

Para apresentar os resultados a viabilidade é expressa pelo somatório das sementes germináveis das classes de 1 a 5, e o potencial de vigor expresso pelo somatório de 1 a 3 (França Neto, 1999). Outra periclitância seria a similaridade visual entre a classe 3 e 4 (Figura 9), o que influencia na avaliação do teste e nos resultados.

Figura 41 – Classes de viabilidade e vigor de sementes de feijão-caupi.



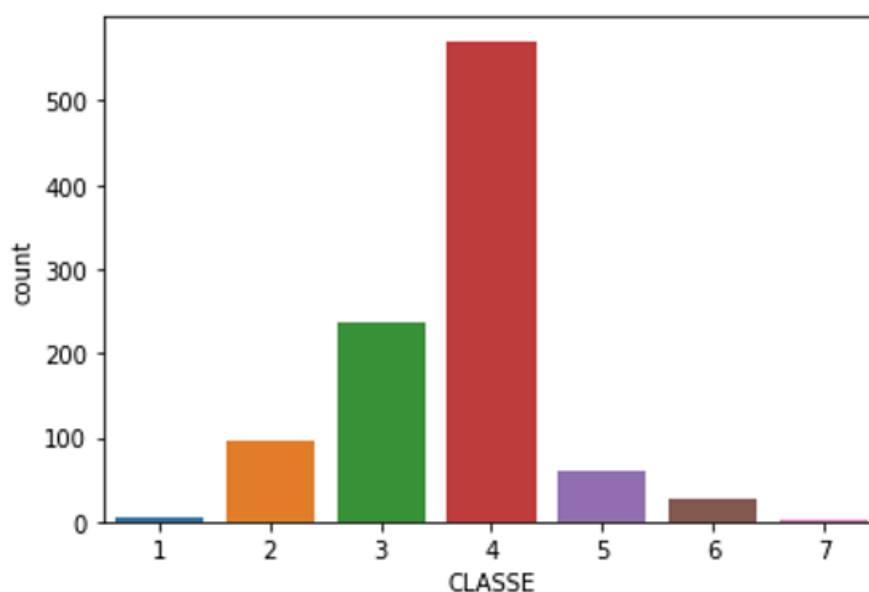
Fonte: Elaborado pela autora. Legenda Classes de viabilidade e vigor de sementes de feijão. (1) Alto vigor; (2) Médio vigor; (3) Baixo vigor; (4) Baixo vigor; (5) Muito baixo vigor; (6) Não viável e (7) Morta.

Portanto, quando a classificação ocorre de maneira errada há uma variação final na porcentagem de vigor do lote. A classificação manual de sementes é suscetível a erros humanos e estes erros podem resultar em uma avaliação incorreta do vigor das sementes, o que compromete a qualidade final dos lotes. A interpretação dos resultados, é uma limitação do teste, esse processo pode apresentar dificuldades de padronização e depende significativamente da habilidade do analista em determinar se uma semente é vigorosa, o que se torna uma das principais fontes de variação nos resultados (McDonald; Phaneendranath, 1978; França Neto; Krzyzanowski, 2019). A automação, por outro lado, oferece uma

abordagem mais padronizada e objetiva, utilizando algoritmos e técnicas de VC para analisar as sementes de forma mais precisa e reprodutível.

Durante a construção do banco de imagens, ocorreu um desequilíbrio no número de imagens por classe, conforme ilustrado na Figura 9. Isso pode ser explicado pela pouca variabilidade entre os lotes e que seria necessário para ter essa maior variação entre os dados.

Figura 42 – Balanço do número de imagens por classe.



Fonte: Elaborado pela autora.

Esse problema foi mitigado com a aplicação da técnica de “Data Augmentation” utilizada em aprendizado de máquina, especialmente em VC, para aumentar o volume e a diversidade de dados de treinamento sem a necessidade de coletar novas amostras. Essa técnica atua aplicando transformações como rotação, espelhamentos, ajustes de brilho e contraste, entre outras, às imagens originais (Saleem, *et al.*, 2024). As variações geradas aumentam a robustez e consequentemente a capacidade de generalização do modelo.

Dessa forma, foi possível a construção da matriz de confusão, além de permitir calcular métricas como acurácia, precisão, recall e F1-score (Tabela 5). Facilita a análise do desempenho do modelo, ajuda a identificar padrões de erro e áreas que necessitam de ajustes, como a tendência do modelo a confundir certas classes, contribuindo para a melhoria contínua da eficácia de modelos de classificação.

As métricas de desempenho são utilizadas para avaliar a eficácia do modelo e para quantificar a precisão do modelo, a capacidade de generalização e a identificação de áreas que precisam de melhorias

Tabela 5 – Métricas de desempenho para medir o desempenho do modelo CNN (Inception-ResNet v2).

Classes	Precision	recall	F1-score
1	1,00	1,00	1,00
2	0,72	0,55	0,62
3	0,72	0,58	0,64
4	0,80	0,87	0,83
5	0,76	0,90	0,83
6	0,71	0,86	0,77
7	1,00	0,50	0,67
Acurácia	-	-	0,88
Média avg	0,81	0,75	0,88
Média ponderada avg	0,77	0,77	0,88

Fonte: Elaborada pela autora.

A tabela 5 mostra as métricas de desempenho para o modelo CNN (Inception-ResNet v2), para as 7 classes.

A métrica de precisão mede a proporção de previsões positivas corretas para uma classe específica (Dinga *et al.*, 2019). No caso da classe 1, a precisão é 1,00, indicando que todas as previsões feitas para esta classe estão corretas. Por outro lado, para a classe 2, a precisão é 0,72, significando que 72% das previsões para esta classe são corretas, enquanto a classe 3 apresenta uma precisão similar de 0,72. As classes 4, 5 e 6 mostram boas precisões, variando de 0,76 a 0,80, o que sugere um desempenho eficaz nessas classes. Já a classe 7 possui uma precisão perfeita de 1,00, embora seu recall seja significativamente baixo, indicando que muitas instâncias da classe não estão sendo identificadas corretamente.

O recall avalia a capacidade do modelo de identificar todas as instâncias de uma classe. Para a classe 1, o recall é 1,00, indicando que todas as instâncias dessa classe foram corretamente identificadas dessa maneira o modelo conseguiu identificar todas as instâncias dessa classe, indicando uma performance excelente nessa categoria (Saito; Rehmsmeier, 2015).

Em contraste, a classe 2 tem um recall de 0,55, sugerindo que o modelo está perdendo 45% das instâncias reais dessa classe. A classe 3, com um recall de 0,58, também apresenta dificuldades similares. As classes 4, 5 e 6 têm recalls altos, variando de 0,80 a 0,90, indicando boa identificação das instâncias dessas classes. A classe 7, com um recall de 0,50, indica uma identificação insuficiente. Estudos anteriores têm mostrado que baixos valores de recall podem estar associados a problemas de desequilíbrio de classe ou à dificuldade do modelo em generalizar para todas as instâncias de uma classe (Chicco; Jurman, 2020).

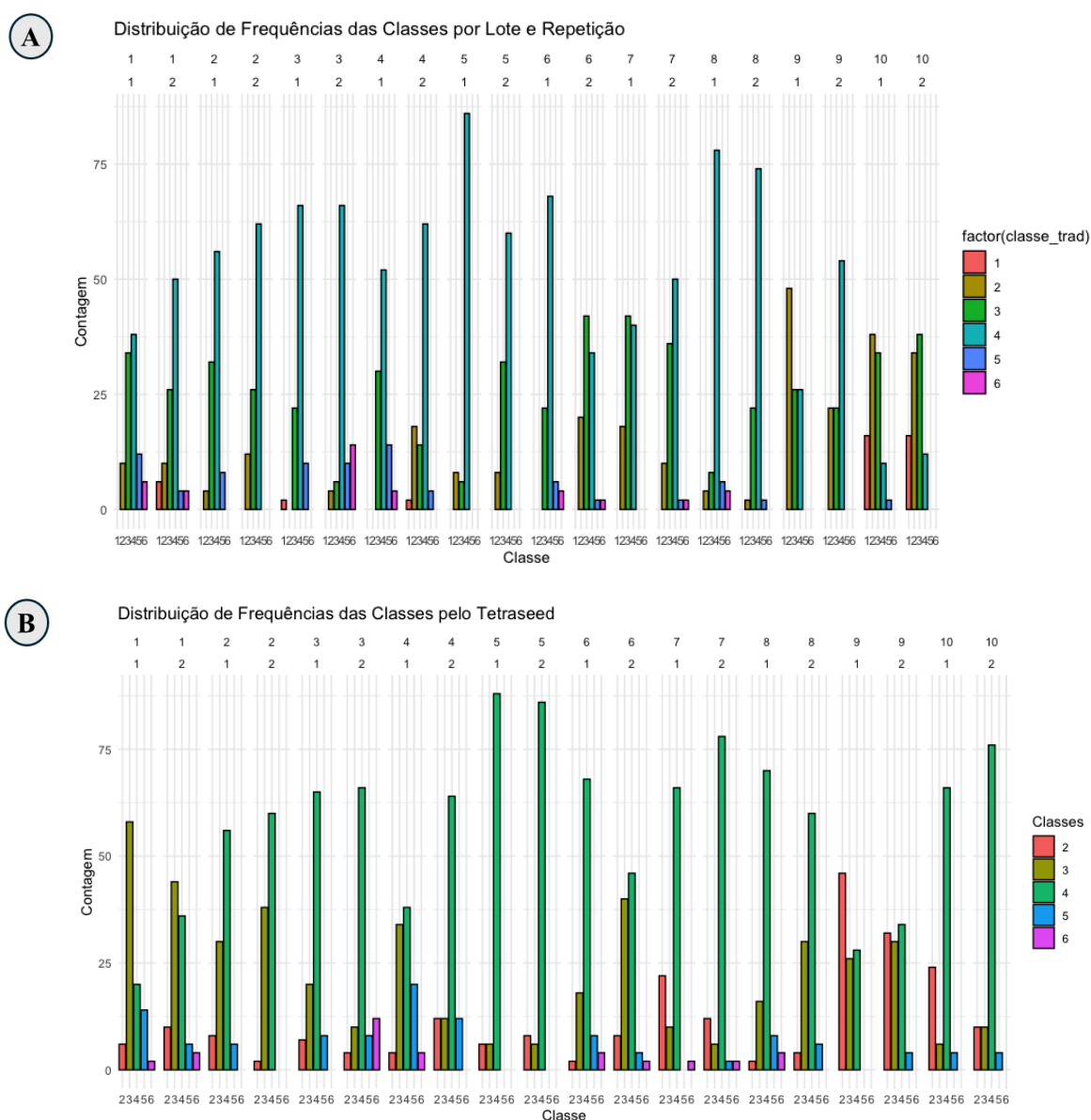
O F1-score combina precisão e recall em uma única métrica, proporcionando um equilíbrio entre essas duas dimensões. A classe 1 tem um F1-score de 1,00, demonstrando um desempenho excelente. As classes 4, 5 e 6 também mostram F1-scores elevados, variando de 0,77 a 0,83, o que sugere um bom equilíbrio entre precisão e recall. No entanto, a classe 2 tem um F1-score mais baixo de 0,62, e a classe 3 um pouco melhor, com 0,64. A classe 7 tem um F1-score de 0,67, refletindo uma performance intermediária. Esse resultado é explicado devido à distribuição desequilibrada dos dados entre as classes (Altuntaş; Cömert; Kocamaz, 2019).

A acurácia geral do modelo é 0,88, indicando que 88% das previsões feitas pelo modelo estão corretas. As médias calculadas incluem a média simples e a média ponderada para precisão, recall e F1-score. A média simples fornece uma visão geral do desempenho médio do modelo, enquanto a média ponderada ajusta essas métricas de acordo com a proporção de instâncias de cada classe, oferecendo uma visão mais equilibrada, especialmente quando as classes têm tamanhos diferentes. As métricas são essenciais para compreender o desempenho do modelo, principalmente para a classificação com muitas classes, onde pode ter variação em termos de complexidade e dificuldade de identificação. A aplicação de técnicas como validação cruzada e ajuste de hiperparâmetros pode contribuir significativamente para melhorar essas métricas e, conseqüentemente, aprimorar o desempenho geral do modelo (Kohavi, 1995; He; Garcia, 2009).

O modelo de CNN Inception-ResNet v2 alcançou uma acurácia de 88%, mas os resultados indicam que ajustes podem ser necessários para melhorar a identificação de instâncias em classes específicas.

Após o treinamento houve upload das imagens para análise pelo TETRASEED e coleta dos dados gerados pelo sistema para comparação entre os métodos (Figura 11).

Figura 43 – Distribuição de frequências das classes.



Fonte: Elaborado pela autora. Legenda: A- Para a avaliação manual e B- Para a avaliação do TETRASEED.

As sementes foram inicialmente classificadas manualmente com base na literatura, e apesar do rigor do estudo, é possível que algumas sementes tenham sido rotuladas

incorretamente no conjunto de dados. Embora o modelo de CNN tenha demonstrado alta acurácia e excelentes resultados, essas possíveis rotulações incorretas podem ter impactado negativamente o processo de aprendizado. Como parte de projetos futuros, planeja-se aumentar o banco de dados e re-treinar a CNN, visando melhorar a acurácia e a capacidade de generalização do modelo.

Ao comparar as frequências entre os métodos de classificação manual e automatizada, observam-se diferenças que eram esperadas. Na classificação manual, os dados são tratados qualitativamente, o que em estatística refere-se a informações categóricas que descrevem atributos ou características não numéricas, como gênero, cor ou tipo, e podem ser classificadas como nominais ou ordinais (Agresti, 2012). Em contraste, na classificação automatizada, os dados foram gerados quantitativamente através de algoritmos de segmentação, que isolaram as sementes do fundo da imagem e mediram diretamente as propriedades das regiões segmentadas em pixels (Gonzalez, 2009). Esse processo transforma características visuais subjetivas em dados mensuráveis e objetivos (Lecun *et al.*, 2015).

Dados qualitativos e quantitativos desempenham papéis complementares na pesquisa estatística (Shi *et al.*, 2023). Enquanto dados qualitativos capturam a essência de características não mensuráveis, dados quantitativos fornecem uma base sólida para análises estatísticas rigorosas e generalizações (Yang *et al.*, 2021; Shi *et al.*, 2023). A compreensão e a distinção entre esses dois tipos de dados são essenciais para a correta aplicação de técnicas estatísticas em pesquisas, impactando diretamente na interpretação dos resultados.

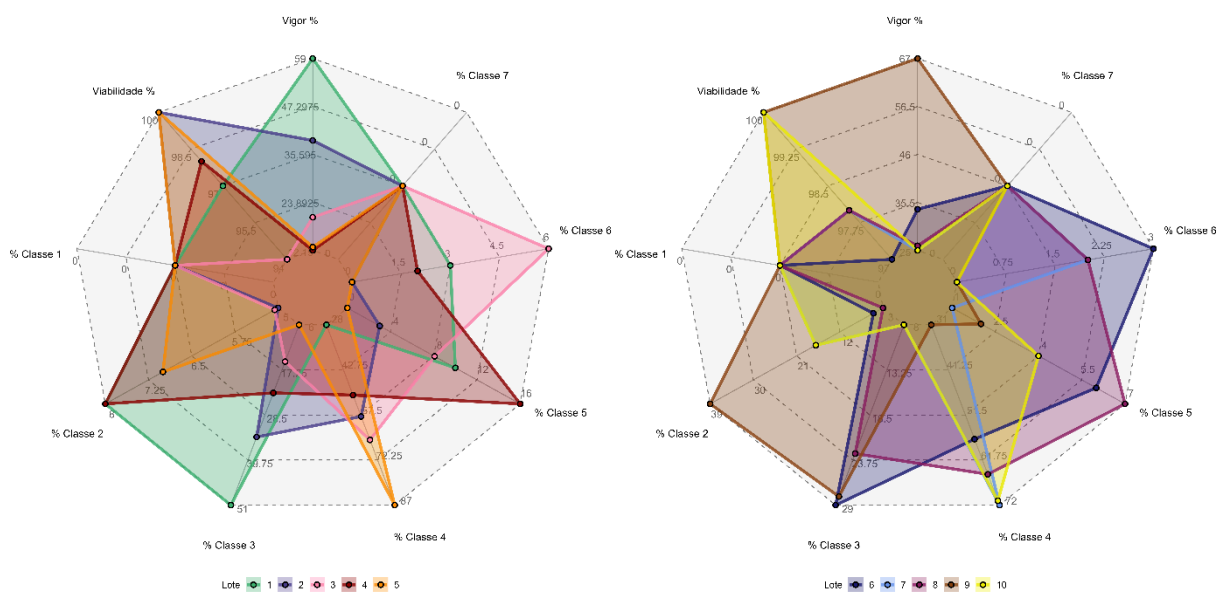
A análise de dados apresentada na Figura 12, traz os resultados no Gráfico radar das variáveis Vigor, Viabilidade e Classes obtidas por meio da análise no TETRASEED.

O TETRASEED analisou e classificou as imagens de sementes submetidas ao teste de tetrazólio com acurácia de 88%, esse resultado é particularmente relevante considerando que a acurácia é um dos principais indicadores de desempenho em sistemas de visão computacional aplicados na agricultura (Lecun *et al.*, 2015). Portanto, o sistema demonstra grande importância e potencial para uso futuro, e embora ainda necessite de aprimoramentos, os dados atuais são concretos e promissores para a área de sementes.

O processo de geração de dados totalmente automatizado, desde a segmentação da imagem até a classificação final, elimina a subjetividade associada à análise manual, um fator

que tradicionalmente pode introduzir variabilidade nos resultados (Liu *et al.*, 2023.). A automatização não apenas melhora a consistência dos dados, mas também otimiza o tempo de análise, o que é crucial em operações de larga escala (Guo *et al.*, 2016).

Figura 44 – Gráfico Radar entre as variáveis Vigor, Viabilidade e Classes do TETRASEED.



Fonte: Elaborado pela autora.

No entanto, apesar dos resultados promissores, o TETRASEED ainda requer melhorias. A necessidade de ajustes adicionais é comum em sistemas de aprendizado de máquina, especialmente em suas fases iniciais de desenvolvimento (Chen *et al.*, 2020.). Com aprimoramentos contínuos, a expectativa é de que o TETRASEED possa alcançar ainda maior acurácia e se tornar uma ferramenta indispensável na avaliação da qualidade das sementes.

5.4 Conclusão

O sistema de VC TETRASEED apresentou uma acurácia de 88%, na avaliação das imagens de sementes submetidas ao teste de tetrazólio indicando um desempenho sólido na classificação automatizada das sementes de feijão-caupi a partir de dados qualitativos do teste de tetrazólio, demonstrando uma boa capacidade de aprendizagem e generalização, diminuindo a subjetividade da classificação manual. Ajustes adicionais são necessários para garantir uma performance mais equilibrada em todas as classes, aumentando a confiabilidade do sistema para aplicações práticas na avaliação de sementes.

6 CAPÍTULO 4 - CARACTERIZAÇÃO MORFORCOLORÍMETRICA DE SEMENTES FEIJÃO CAUPI (*Vigna unguiculata (L.) Walp.*) POR MEIO DO USO DE ANÁLISE DE IMAGEM

Resumo - A caracterização morfo-colorimétrica de sementes analisa a forma (morfologia) e cor (colorimetria) das sementes, utilizando ferramentas como análise de imagem para avaliar sua qualidade e vigor com precisão, possibilitando diferenciar lotes de sementes. O presente trabalho teve como objetivo realizar a caracterização morfo-colorimétrica de lotes de sementes de feijão-caupi (*Vigna unguiculata (L.) Walp.*) por meio da análise de imagem digital, com o uso do plugin Seeds Analysis do software Image J, visando uma análise mais precisa, confiável e com otimização de tempo. O trabalho foi desenvolvido no Laboratório de Análise de Sementes, do Departamento de Fitotecnia, do Centro de Ciências Agrárias da Universidade Federal do Ceará-UFC em parceria com a Universidade de Cagliari, na Itália. Foram utilizados dez lotes de sementes de feijão-caupi para a avaliação do vigor, foram conduzidos testes de germinação, primeira contagem de germinação, emergência de plântulas, comprimento das plântulas, peso de massa seca, peso de 1000 sementes e biometria das sementes com comprimento, largura e relação C/L. A aquisição das imagens foi feita utilizando um scanner de bancada (HP Scanjet 2004[®]) adaptado de maneira invertida. A coleta dos dados das imagens foi feita através do plugin pertencente ao software ImageJ. Foram avaliados parâmetros relacionados a características morfo-colorimétricas disponíveis no plugin, realizado uma correlação entre teste de vigor e imagem para validar a técnica, além de realizar uma análise discriminante linear (LDA) para separação e ranqueamento dos lotes. Os resultados indicam que a análise de imagem digital é uma alternativa viável e vantajosa para a diferenciação da qualidade fisiológica de sementes, proporcionando uma avaliação precisa do vigor dos lotes. O uso da análise de imagem digital para a caracterização biométrica de sementes de feijão-caupi demonstrou ser uma ferramenta eficiente e precisa.

Palavras-chave: ImageJ; qualidade fisiológica; biometria digital

6 CHAPTER 4 - MOPHORCOLORIMETRIC CHARACTERIZATION OF COWPEA SEEDS (*Vigna unguiculata* L.) USING IMAGE ANALYSIS

Abstract - Seed biometrics is a widely used variable in differentiating the physiological quality of various species. However, traditional biometric methods require time and human resources and are susceptible to measurement errors. This study aimed to characterize the biometrics of cowpea seed lots using digital image analysis, validating the use of the Seeds Analysis plugin to discriminate morphocolorimetric characteristics. The work was conducted in the Seed Analysis Laboratory of the Department of Crop Science at the Federal University of Ceará (UFC), in collaboration with the University of Cagliari in Italy. Ten cowpea seed lots were used. To assess vigor, tests were conducted for germination, first germination count, seedling emergence, seedling length, dry mass weight, 1000-seed weight, and seed biometrics, including length, width, and length/width ratio. Images were acquired using an inverted desktop scanner (HP Scanjet 2004®). Data collection from the images was done through the plugin belonging to the ImageJ software. Parameters related to morphometric characteristics available in the plugin were evaluated, a correlation between vigor tests and image analysis was conducted to validate the technique, and a Linear Discriminant Analysis (LDA) was performed for the separation and ranking of the lots. The results indicate that digital image analysis is a viable and advantageous alternative for differentiating the physiological quality of seeds, providing an accurate evaluation of the vigor of the lots. The use of digital image analysis for the biometric characterization of cowpea seeds proved to be an efficient and precise tool.

Keywords: ImageJ; physiological quality; digital biometrics.

6.1 Introdução

O feijão-caupi é uma leguminosa de grande importância econômica e nutricional, especialmente em regiões tropicais e subtropicais. No Brasil, é amplamente cultivado nas regiões Norte e Nordeste devido à sua capacidade de adaptação às condições adversas, como altas temperaturas e períodos de estiagem, características que são comuns nessas áreas (dos Santos *et al.*, 2021). Além de seu valor econômico, o feijão-caupi possui um elevado teor proteico, tornando-se uma fonte significativa de proteínas vegetais na dieta das populações locais (dos Santos *et al.*, 2021). A cultura é também resistente e possui baixo custo de cultivo, o que a torna uma opção viável para pequenos produtores e agricultores familiares, reforçando seu papel crucial na segurança alimentar e na sustentabilidade agrícola (Câmara *et al.*, 2018). No entanto, apesar de sua importância, o uso de sementes de baixa qualidade ainda é frequente na implantação dessa cultura.

A qualidade das sementes é um fator determinante para o sucesso da agricultura, sendo a base para o estabelecimento de culturas saudáveis e produtivas (Marcos Filho, 2015). A utilização de sementes vigorosas e de alta qualidade é essencial para garantir um bom desempenho em campo (dos Santos *et al.*, 2021). Entre as diversas técnicas para avaliar essa qualidade, a caracterização morfo-colorimétrica, que analisa simultaneamente a forma (morfologia) e a cor (colorimetria) das sementes, tem se destacado como uma abordagem inovadora e eficiente.

Tradicionalmente, a avaliação da qualidade das sementes era realizada por meio de testes manuais, como o teste de germinação e o de vigor (Xia *et al.*, 2019). Embora esses métodos sejam amplamente aceitos, eles apresentam limitações, como a subjetividade na interpretação dos resultados, o tempo necessário para execução e a suscetibilidade a erros humanos (Neto *et al.*, 2020). Além disso, esses testes são frequentemente destrutivos, o que os torna menos adequados para a avaliação de grandes lotes de sementes (Ma; Tsuchikawa; Inagaki, 2020). Nesse sentido, há uma demanda crescente por métodos que sejam mais rápidos, precisos e não invasivos.

A análise de imagem digital surge como uma solução promissora para superar essas limitações (Brandani, 2017). Essa tecnologia permite a captura e o processamento de imagens de sementes, possibilitando a extração de informações morfométricas e colorimétricas de forma automatizada e objetiva (Bekkering; Huang; Tian, 2020). No caso das sementes,

características como comprimento, largura, proporção e tonalidade de cor podem ser analisadas com alta precisão, fornecendo dados valiosos para a diferenciação entre lotes de diferentes qualidades fisiológicas (Altuntaş, 2019).

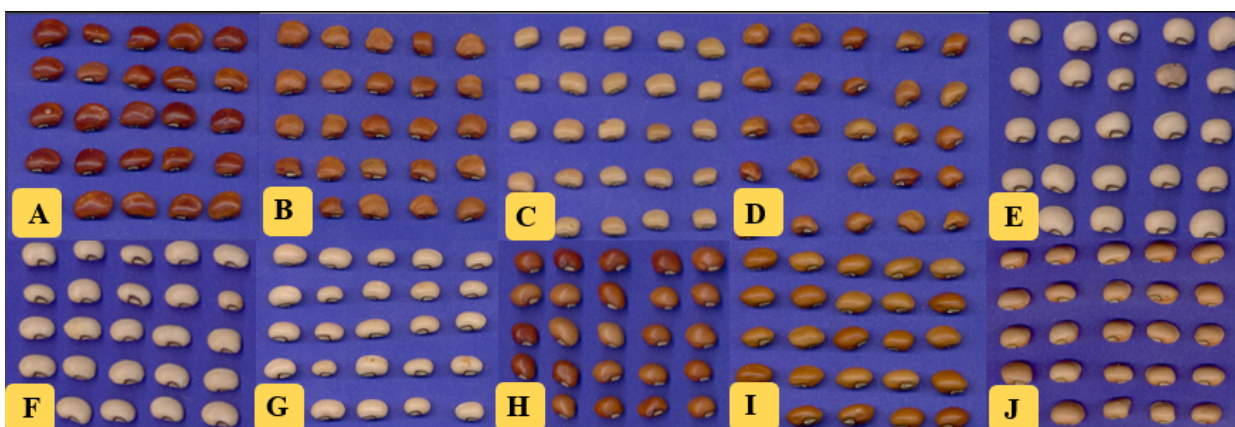
Ferramentas de software, como o ImageJ, desempenham um papel fundamental nesse processo, facilitando a análise automatizada de imagens. Com a adição de plugins especializados, como o Seeds Analysis, é possível adaptar o software para as necessidades específicas da análise de sementes, tornando-o uma ferramenta eficiente para pesquisadores e profissionais do setor agrícola (Loddo *et al.*, 2022). O uso dessa tecnologia em conjunto com testes de vigor tradicionais pode gerar uma validação robusta dos resultados, corroborando sua eficácia na diferenciação de lotes de sementes (Tang; Cai, 2013).

Diante disso, o presente trabalho teve como objetivo realizar a caracterização morfo-colorimétrica de lotes de sementes de feijão-caupi por meio da análise de imagem digital, com o uso do plugin Seeds Analysis do software Image J, visando uma análise mais precisa, confiável e com otimização de tempo.

6.2 Materiais e métodos

O trabalho foi desenvolvido no Laboratório de Análise de Sementes, do Departamento de Fitotecnia, do Centro de Ciências Agrárias da Universidade Federal do Ceará-UFC em parceria com a Universidade de Cagliari, na Itália. Para a realização dos testes foram utilizados dez lotes de sementes de feijão caupi (Figura 1). Para a avaliação da qualidade dos lotes, foram conduzidos testes de peso de 1000 sementes (PMS) germinação (G%), primeira contagem (PC), comprimento de parte aérea e raiz (CPA e CPR), peso de massa seca de parte aérea e raiz (MSPA e MSR), índice de velocidade de emergência (IVE) e condutividade elétrica (CE) seguindo os padrões das Regras para Análises de Sementes (RAS) (BRASIL, 2009). Os experimentos foram divididos em duas partes, Etapa I e II, sendo na Etapa I, foram avaliados a qualidade das sementes por meio de testes tradicionais, e na Etapa II, morfo-colorimetria por análise de imagem.

Figura 45 – Lotes de feijão utilizados nos experimentos.



Fonte: Elaborado pela autora. Legenda: A) Lote 1; B) Lote 2; C) Lote 3; D) Lote 4; E) Lote 5; F) Lote 6; G) Lote 7; H) Lote 8; I) Lote 9 e J) Lote 10.

6.2.1 Etapa I: Análise de qualidade dos lotes de sementes com testes tradicionais.

6.2.1.1 Teste de germinação e primeira contagem

No teste de germinação, para cada lote, foram utilizadas quatro repetições de cinquenta sementes. A semeadura foi feita no papel de (Germitest®) umedecido com água destilada, na proporção de 2,5 vezes a massa do substrato. Após isso, as sementes foram acondicionadas em câmara de germinação tipo Biochemical Oxygen Demand (BOD) regulada à temperatura de 25 °C, com fotoperíodo de 12h, sendo feita a primeira contagem no 5º dia, considerando apenas as plântulas normais, finalizando no 8º dia. Os dados foram obtidos por meio da média das quatro repetições, transformando o resultado em porcentagem.

6.2.1.2 Comprimento de plântulas

Foi realizado em conjunto com o teste de germinação, onde após oito dias da semeadura, ocorreu a medição do comprimento de dez plântulas normais por repetição com o auxílio de régua graduada em milímetros. O comprimento médio foi obtido somando-se as medidas de cada repetição e dividido pelo número de plântulas avaliadas.

6.2.1.3 Massa seca de plântulas

As plântulas foram colocadas em sacos de papel e levadas para a estufa com circulação de ar forçada à temperatura de 65°C até atingir peso constante. O material seco

será pesado em balança com precisão de 0,001g e o resultado dividido pelo número de plântulas normais.

6.2.1.4 Teste de emergência

No teste de emergência de plântulas, foram utilizadas quatro repetições de 50 sementes por lote. A semeadura ocorreu de forma manual em canteiros de 10 x 1m, com 10 cm entre repetições, contendo substrato terra/areia na proporção 1:1. As sementes foram dispostas de maneira equidistante em sulcos com 3cm de profundidade no canteiro. Diariamente foram feitas observações e irrigações sempre que necessário, visando o fornecimento de água para a germinação das sementes e emergência das plântulas e a partir da emergência da primeira plântula, iniciou-se a contagem. O teste foi avaliado aos dez dias após a semeadura, sendo computada a porcentagem de plântulas. Ao final da contagem os dados foram transformados em porcentagem conforme Nakagawa (1999).

Foi avaliado o Índice de Velocidade de Emergência (IVE), que são capazes de estimar o número de dias médio necessários para ocorrer a emergência e o número de plantas emergidas por dia. Os índices foram calculados seguindo a metodologia proposta por Maguire (1962), em que: $IVE = (G1/N1) + (G2/N2) + \dots + (GN/N)$, no qual, G = número de plântulas normais computadas nas contagens; N = número de dias da semeadura do primeiro ao último dia de contagem.

6.2.1.5 Peso de mil sementes

Para a realização do peso de mil sementes, foram utilizadas oito repetições de 100 sementes de cada lote, separadas manualmente. Após isso, é feita a pesagem do material em gramas, em balança de precisão com duas casas decimais. Para a obtenção dos resultados foram feitos cálculos de variância, desvio padrão e coeficiente de variação dos valores obtidos nas pesagens, seguindo as fórmulas:

$$Variância = \frac{n(\sum x^2) - (\sum x)^2}{n(n-1)}$$

Onde:

x: peso de cada repetição;

n: número de repetições;

Σ : somatório.

$$\text{Desvio padrão } (S) = \sqrt{\text{variância}}$$

$$\text{Coeficiente de variação } (CV) = \frac{S}{\bar{X}} \times 100$$

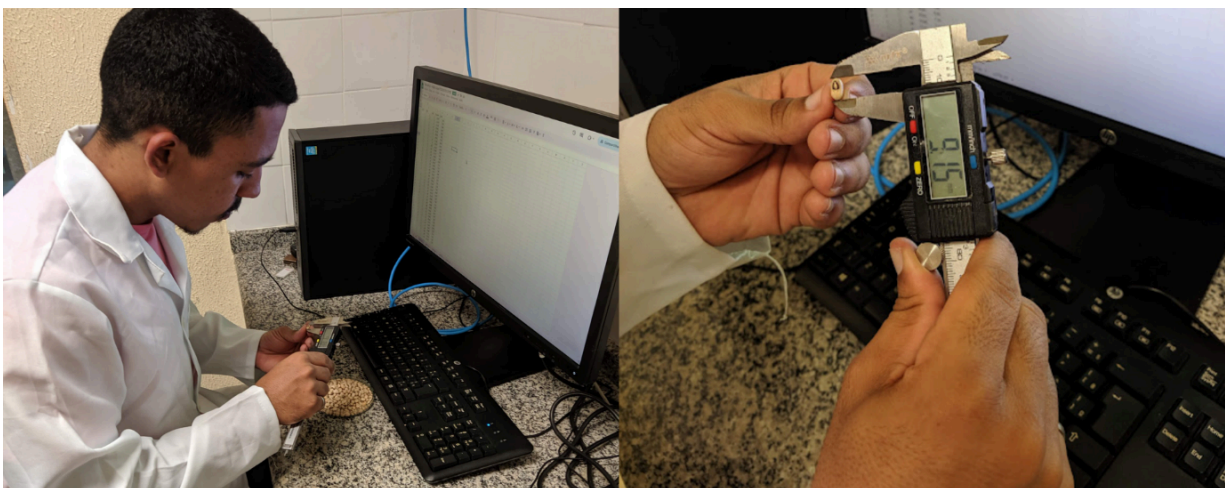
Onde:

X: peso médio de 100 sementes.

6.2.2 Biometria de sementes manual

Para a análise biométrica das sementes, foram utilizadas quatro repetições e 100 sementes por unidade amostral por lote. Os dados foram obtidos por meio de um paquímetro digital com resolução de 0,01mm sendo avaliado comprimento, largura das sementes, em seguida obteve-se a relação comprimento e largura (C/L).

Figura 46 –Biometria de semente por paquímetro digital.



Fonte: Elaborado pela autora.

6.2.3 Etapa II: mofo-colorimetria por análise de imagem.

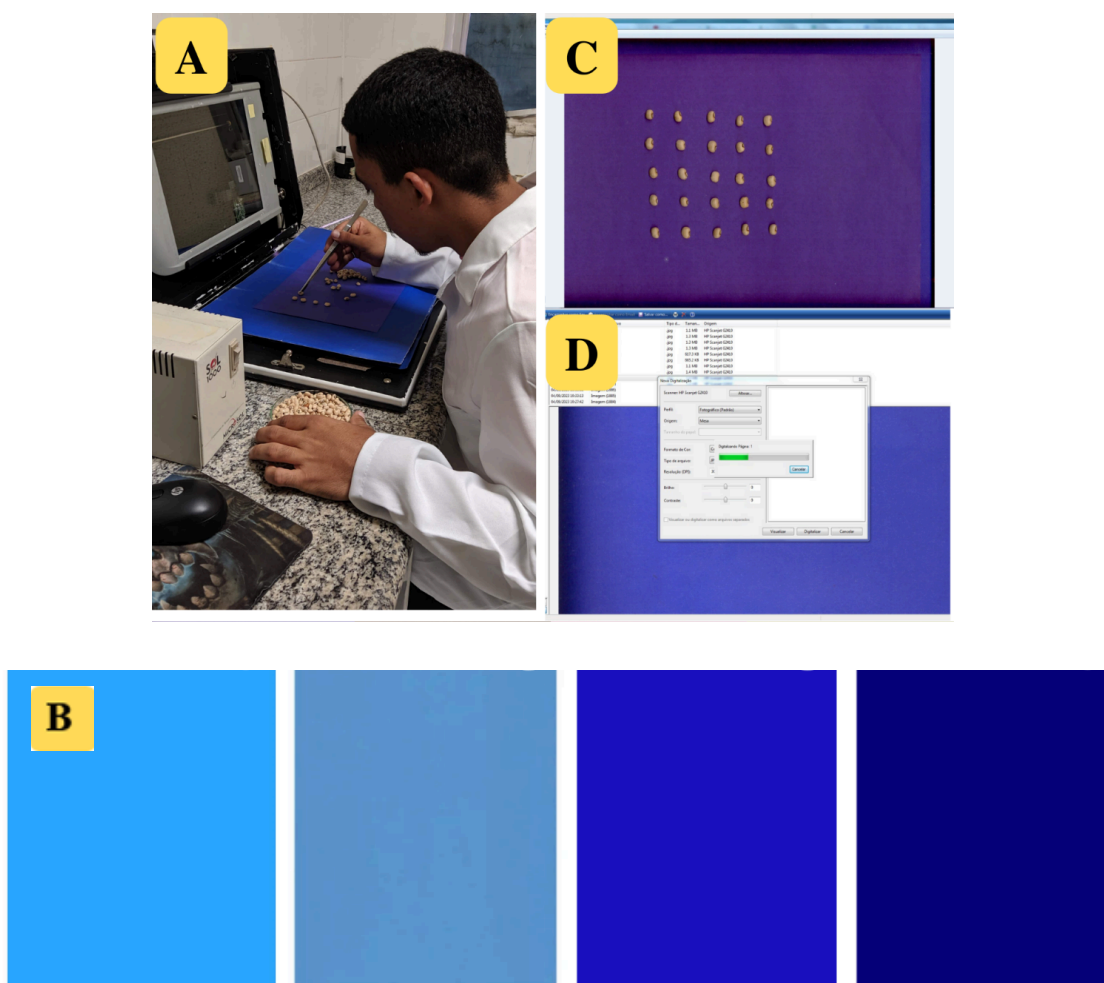
6.2.3.1 Aquisição de imagem

Para a captura de imagem foram selecionadas aleatoriamente quatro repetições de 100 sementes para cada lote, totalizando 4000 sementes. A captura de imagem ocorreu com o auxílio de um scanner de bancada (HP Scanjet 2004®) adaptado de maneira invertida (Figura

3A). As sementes foram dispostas em folhas de sulfite de diferentes tonalidades de azul (Figura 3B). As diferentes colorações foram utilizadas para encontrar a melhor cor que se adequa às sementes, possibilitando alto contraste entre o fundo da imagem e as sementes, facilitando posteriormente a segmentação da imagem e extração dos dados.

As sementes foram dispostas no papel de maneira equidistante, não ocorrendo sobreposição umas com as outras e o scanner foi utilizado com a tampa fechada, proporcionando assim uma iluminação adequada e uniforme, evitando quaisquer interferências do ambiente (Figura 4C). Após as capturas, as imagens foram salvas em formato JPEG em formato A4 com dimensões (210x297mm), essas imagens foram encaminhadas para o software ImageJ para a realização da coleta de dados (Figura 4D).

Figura 47 – Aquisição das imagens digitais.



Fonte: Elaborado pela autora. Legenda: A - Scanner de bancada invertido; B - Captura da imagem com diferentes colorações de fundo; C – Disposição das sementes; D - Processamento das imagens em JPEG.

6.2.3.2 Plugin para extração de características das imagens

O plugin utilizado foi o Seeds Analysis, que é um plugin capaz de analisar múltiplas sementes em uma única imagem de forma digital. O plugin se baseia em imagens com o fundo de coloração padrão ou preto, e na remoção do mesmo para fazer o destaque das sementes em relação ao fundo da imagem.

Após fazer o recorte do fundo, a imagem é pré-processada para fazer a seleção das partes desejadas, sendo possível também fazer a seleção das sementes de modo individual. Após a segmentação da imagem, é necessário fazer a seleção de certos parâmetros, como o de tamanho mínimo e máximo da área, que é medido em pixels, a seleção dessas opções chaves são necessários, pois isso fará com que facilite a análise das regiões de interesse. Por fim, após essas etapas, o usuário poderá então fazer a escolha de todos os parâmetros de seu interesse.

Foram avaliados 16 parâmetros dos 64 disponíveis no plugin Seeds Analysis (Tabela 6).

Tabela 6 – Parâmetros de características morfológicas avaliados com o plugin Seeds Analysis.

Parâmetro	Descrição
Parâmetros Biométricos	
Área	Área de semente (em pixels)
Perimeter	Comprimento do contorno da semente
Feret	Maior diâmetro rastreável com dois pontos do contorno da semente como extremidades, chamado Comprimento
Breadth	Comprimento do maior eixo rastreável perpendicular ao Feret, também chamado de Largura
AspRatio	Feret/largura, também chamado de razão de excentricidade ou retangularidade
RFactor	Fator de forma, definido como $\text{CovenxArea}/(\text{Feret} \times \pi)$
Roundness	Medida de arredondamento, definida como $4 \times \text{Área}/(\pi \times \text{Feret}^2)$
Solidity	Medida de solidez, definida como $\text{Área}/\text{ConvexArea}$
Convexity	Medida de convexidade, também chamada de rugosidade, definida como $\text{ConvexPerimeter}/\text{Perimeter}$

Sphericity	Também chamada de razão de raio, expressa por MinR/MaxR
Circularity	Também chamado de fator de forma, obtido por $2 \times \pi \times \text{Área}/\text{Perímetro}^2$
Endocarp	Número de pixels que formam o endocarpo da semente

Parâmetros Colorimétricos

MeanRed	Média dos valores do canal vermelho
MeanGreen	Média dos valores do canal verde
MeanBlue	Média dos valores do canal azul
MeanRGB	Média dos valores do canal RGB

Fonte: Adaptado de Loddo *et al.*, (2022).

O delineamento experimental utilizado foi o inteiramente casualizado (DIC). Os resultados dos testes de germinação e vigor foram submetidos à análise de variância e as médias submetidas ao teste Tukey a 5% de probabilidade. Os resultados também foram submetidos aos testes de correlação de Spearman entre as variáveis estudadas pelo método tradicional e pela análise de imagem. As análises estatísticas foram realizadas por meio do software R 4.3.1 (R Core Team, 2020).

As características morfo-colorimétricas foram utilizadas para a construção de um banco de dados.

Para a técnica multivariada exploratória a análise de Componentes Principais (PCA), foi realizado o processamento com a matriz de covariância das variáveis originais, obtendo-se dela os autovalores que construíram os autovetores. Por meio dessas combinações lineares há a denominação de componentes principais.

O poder discriminatório de cada variável num componente foi medido pela fórmula:

$$r_{xj}(cp_h) = \frac{a_{jh}\sqrt{\lambda_h}}{s_j}$$

Nessa fórmula, s_j = desvio padrão da variável j , a_{jh} = coeficiente da variável j no h -ésimo componente principal, e λ_h = h -ésima raiz característica (autovalor) da matriz de covariância (HAIR *et al.*, 2005).

Os dados foram submetidos a classificação pela multivariada Análise Discriminante Linear (LDA) segundo Sarigu, *et al.*, (2019) e implementado no Software IBM SPSS (Statistical Package for Social Science) versão 16.0 (SPSS Inc. for Windows, Chicago, Illinois, EUA) (2006).

6.3 Resultados e Discussão

6.3.1 Etapa I: Análise de qualidade dos lotes de sementes com testes tradicionais.

De acordo com os resultados expostos na Tabela 7 foi possível definir estatisticamente a diferença de vigor entre os lotes de feijão avaliados. Os testes de vigor têm sido utilizados com objetivo de identificar diferenças entre lotes de sementes, destacando aqueles que possuem maior potencial fisiológico, com maiores chances de sucesso em campo (Marcos Filho *et al.*, 2015). Segundo os testes de vigor são utilizados para diferenciar o potencial fisiológico entre os lotes, classificando-os em lotes de alto e baixo vigor (Zucareli *et al.*, 2011).

Tabela 7 – Resultados de primeira contagem - PC; plântulas anormais- PA; comprimento de parte aérea - CPA; comprimento de raiz - CRA; matéria seca de parte aérea - MSPA; matéria seca de raiz - MSRA; condutividade elétrica - CE; peso de mil sementes - PMS; índice de velocidade de emergência - IVE obtidos em 10 lotes de feijão.

Lote	PC (%)	G (%)	PA (%)	CPA (cm)	CRA (cm)	MSPA (g)	MSRA (g)	CE ($\mu\text{S cm}^{-1} \text{g}^{-1}$)	PMS	IVE
1	74,00b	74,00b	55,00ab	9,25cd	6,62c	0,697bc	0,075b	1646,50b	235,11ab	1,55ns
2	95,50a	95,50a	82,00a	8,56d	8,62bc	0,522cd	0,087b	1493,75b	169,80e	1,77ns
3	89,50ab	89,50ab	8,5d	18,95a	13,40ab	0,392d	0,085b	849,97d	114,81f	1,57ns
4	99,00a	99,00a	55,50ab	10,75bcd	13,02ab	0,960b	0,117ab	1903,25a	191,33de	1,77ns
5	92,00a	92,00a	23,00cd	18,46a	14,34ab	0,710bc	0,122ab	2,29e	205,14cd	1,55ns
6	89,00ab	89,00ab	40,50bc	12,99b	11,19abc	1,267a	0,090b	2,60e	250,74a	1,67ns
7	90,00a	90,00a	46,50bc	12,04bc	11,66abc	0,962b	0,115ab	2,20e	221,55bc	1,60ns
8	89,00ab	89,00ab	48,50bc	13,54b	16,12a	0,672cd	0,167a	1537,75b	187,07de	2,59ns
9	94,50a	94,50a	28,50bcd	14,17b	15,84a	0,640cd	0,152a	781,35d	189,06de	1,65ns
10	88,00ab	88,00ab	33,00bcd	18,34a	13,90ab	0,562cd	0,112ab	1283,00c	183,70de	1,91ns
CV(%)	7,18	7,18	28,79	10,46	20,16	15,79	21,00	8,40	5,76	51,32

Médias seguidas pela mesma letra não diferem entre si pelo teste de Tukey a 5% de probabilidade. Fonte: Elaborada pela autora.

As maiores médias para as variáveis de PC e G foram observadas nos lotes 4, 2, 9, 5 e 7, com valores de 99,00%, 95,50%, 94,50%, 92,00% e 90,00%, respectivamente. Em contrapartida, o lote 1 apresentou a menor média de PC, com 74,00%. O teste de germinação é amplamente utilizado para avaliar o vigor das sementes, permitindo a observação do crescimento inicial e o desenvolvimento de estruturas importantes do embrião, as quais indicam a capacidade da semente formar uma planta normal em condições ideais de campo (Brasil, 2009). No entanto, o teste de germinação requer condições controladas e favoráveis, o que pode não refletir o desempenho das sementes em condições reais de campo (Silva *et al.*, 2014). Por isso, é fundamental a realização de outros testes de qualidade das sementes, possibilitando uma seleção mais criteriosa dos lotes destinados à comercialização (Ullmann *et al.*, 2015). De maneira semelhante, o teste de PC é realizado em condições favoráveis, mas é considerado um indicador de vigor, uma vez que a avaliação da quantidade de plântulas normais na primeira contagem permite a determinação do vigor relativo entre os lotes de sementes (Silveira *et al.*, 2002).

Na variável PA, que se refere a porcentagem de plântulas anormais, o lote 2 apresentou a maior média dentre as observadas, com 82,00%, a taxa alta de plântulas anormais pode estar associada a fatores como baixa viabilidade das sementes e por condições de armazenamento (Marcos Filho, 2015). O lote 3 obteve a menor média para esta variável, com 8,5%, baixa taxa de plântulas anormais mostra a predição do lote de ter melhores condições sem campo. Portanto, a fase de plântula pode ser considerada um dos estágios mais vulneráveis do ciclo da planta, estando suscetíveis a estresses bióticos e abióticos (Bray; West, 2015).

Para o teste de CPA, foi possível distinguir os lotes 3, 5 e 10 com as maiores médias observadas, com 18,95cm, 18,46cm e 18,34cm, respectivamente, indicando um bom desenvolvimento da parte aérea das plântulas. O CPA é uma variável indicativa de vigor de plântulas como discutido por Finch-Savage e Bassel (2016), que associam o vigor inicial ao desempenho futuro da planta em campo. Para esta variável, o lote 2 obteve a menor média, com 8,56cm esse resultado pode ser explicado possivelmente pela baixa qualidade do lote ou reflete uma possível deficiência nas condições iniciais de crescimento intrínseca das sementes.

Já para o teste de CRA foi observado que os lotes que se destacaram com as maiores médias foram os lotes 8 e 9, com 16,12 cm e 15,84 cm de comprimento de raiz, respectivamente, um sistema radicular bem desenvolvido é fundamental para a absorção de água e nutrientes e está diretamente relacionado ao vigor da semente, (Carvalho; Nakagawa, 2012). O lote 1, com a menor média (6,62 cm), pode enfrentar dificuldades no estabelecimento inicial da planta, o que pode comprometer o rendimento final.

Para avaliação de MSPA, o lote 6 obteve a maior média com 1,267g indicando uma maior acumulação de biomassa na parte aérea, um sinal de vigor e potencial produtivo (Maguire, 1962). Em contraste, o lote 3 obteve a menor média com 0,392g sugere uma menor eficiência na conversão de recursos em biomassa.

Dentre os valores médios de MSRA, os que se destacaram foram os dos lotes 8 e 9, com 0,167g e 0,152g, respectivamente, a massa seca da raiz é um indicador importante do vigor, sendo crucial para o crescimento inicial. Os lotes 1, 3, 2 e 6 se destacaram com as menores médias para esta variável, com 0,075g, 0,085g, 0,087g e 0,090g, respectivamente. Segundo Carvalho e Nakagawa (2012), sementes com menor massa seca da raiz podem enfrentar dificuldades no estabelecimento inicial, resultando em plântulas mais fracas e menos competitivas.

Nos resultados do teste de CE o lote 4 se apresentou com o maior resultado dentre os demais lotes, com 1903,25 $\mu\text{S cm}^{-1} \text{ g}^{-1}$ de acúmulo de lixiviado. A CE é uma medida da quantidade de eletrólitos lixiviados pelas sementes depois do processo de embebição, sendo um indicativo da integridade das membranas celulares das sementes. Sementes que apresentam alta condutividade elétrica indicam maior grau de lixiviação de solutos, sugerindo danos nas membranas celulares e menor vigor. Dessa forma, a integridade das membranas celulares é fundamental para a viabilidade das sementes, e danos a essas membranas resultam em maior lixiviação de eletrólitos, comprometendo a capacidade das sementes de reter nutrientes essenciais (Rajjou *et al.*, 2012). O aposto encontramos em sementes com baixa condutividade elétrica indicam uma menor lixiviação de solutos, sugerindo que as membranas celulares estão em melhor condição como os lotes 7, 5 e 6, apresentaram os menores valores, com 2,20, 2,29 e 2,60 $\mu\text{S cm}^{-1} \text{ g}^{-1}$, respectivamente. Portanto, sementes com membranas celulares intactas mantêm melhor a integridade estrutural, permitindo uma retenção eficiente de nutrientes e água,

que são essenciais para um desenvolvimento saudável.

Para o PMS, o lote 6 apresentou a maior média, com 250,74 g, enquanto o lote 3 teve a menor média, com 114,81 g. Esse resultado pode ser justificado pela possibilidade de o processo de deterioração ter ocorrido de maneira menos intensa nos lotes menores em comparação aos maiores, uma vez que a deterioração das sementes armazenadas é inevitável (Bragantini, 2005). Diversos fatores podem influenciar esses resultados, como a diferença de idade entre os lotes e as condições de armazenamento inadequadas, especialmente em relação à temperatura e umidade (Santos *et al.*, 2005). Além disso, a qualidade fisiológica das sementes de feijão pode variar conforme o genótipo (Silva *et al.*, 2008).

Para a variável IVE, não houve diferença estatística entre os lotes.

Na tabela 8, que diz respeito às variáveis biométricas avaliadas manualmente para sementes, o lote 1 se destacou com as maiores médias de comprimento e largura de sementes, com 10,81 mm e 7,31mm, respectivamente, sementes com maiores dimensões, geralmente apresentam uma maior reserva de nutrientes, o que contribui para uma melhor germinação e desenvolvimento inicial.

Tabela 8 – Resultados de comprimento - C; largura- L obtidos em 10 lotes de sementes de feijão.

Lote	C (mm)	L (mm)
1	10,81a	7,31a
2	8,94bc	6,97ab
3	6,74d	5,37c
4	8,51c	6,81ab
5	8,53c	6,34b
6	9,85ab	6,82ab
7	8,90bc	6,53ab
8	8,64c	6,95ab
9	8,09c	6,46b
10	8,38c	6,67ab
CV(%)	5,51	5,00

Médias seguidas da mesma letra minúscula na coluna não diferem pelo teste Tukey a 5% de

probabilidade. Fonte: Elaborado pela autora.

O lote 3 apresentou as menores médias para as duas variáveis, com 6,74 mm para comprimento e 5,37 mm para largura, sementes menores podem ter menor vigor e enfrentar mais desafios durante a germinação e o desenvolvimento inicial (Bewley; Black, 1994). Segundo Silva (2019), a partir da análise biométrica é possível elucidar sobre características, vantagens e desvantagens de uma variedade sobre a outra, se tornando uma ferramenta estratégica do ponto de vista de escolha da semente a ser cultivada. Venial *et al.*, (2017) em seu trabalho avaliando sementes de *Theobroma cacao* L. demonstram a importância das variáveis biométricas e como elas podem ser usadas para programas de melhoramento.

Dessa forma, podemos aferir que por meio das características biométricas o lote 1 apresentou as melhores médias e o lote 3 as piores médias.

6.3.2 Etapa II: Morfo-colorimetria por análise de imagem digital

Com a utilização do plugin Seeds Analysis foi possível realizar a segmentação das sementes do fundo, permitindo a extração e análise dos dados mensurados em “pixels” (Tabela 9). Dentre as variáveis utilizadas para analisar a sementes, RFactor e Solidity não foram significativas pelo teste F a 1% de probabilidade

Tabela 9 – Médias dos parâmetros de características morfológicas para sementes avaliados com o plugin Seeds Analysis nos 10 lotes de feijão.

Lote	Área	Perimeter	Feret	Breadth	AspRatio	RFactor	Roudness	Solidity	Convexity	Sphericity	Circularity	Endocarp
	----- Pixels -----											
1	8290,12a	412,86a	122,38ab	90,24a	1,36bc	33,81ns	702,85abc	761,37ns	825,25b	737,75ab	612,50b	7725,96ab
2	7050,87ab	350,60bc	107,83c	87,77ab	1,23c	34,16ns	767,87a	716,62ns	885,5a	814,00a	722,00a	6700,27abc
3	4994,50c	290,93d	92,27d	71,36d	1,29bc	35,51ns	746,5ab	658,50ns	898,75a	773,62ab	740,75a	4703,56d
4	6335,5bc	336,99bc d	105,42cd	81,02abcd	1,30bc	35,81ns	724,0abc	675,62ns	875,37ab	768,62ab	701,00ab	5998,51cd
5	6086,87bc	310,46cd	103,89cd	74,14cd	1,40ab	30,24ns	718,00abc	728,00ns	927,62a	714,75bc	791,75a	5776,41cd
6	8468,25a	379,18ab	127,85a	83,25abc	1,53a	31,08ns	663,00c	763,12ns	910,12a	654,25c	743,00a	8089,06a
7	6356,12bc	332,02cd	102,52cd	79,68bcd	1,29bc	48,32ns	766,62a	629,12ns	884,5a	777,25ab	724,62a	6024,10cd
8	6684,25b	333,60bc d	108,77bc	80,07bcd	1,36bc	29,47ns	720,00abc	736,25ns	904,25a	738,75ab	753,75a	6350,64bc
9	5764,75bc	319,69cd	103,98cd	73,67cd	1,42ab	33,80ns	680,25bc	696,00ns	886,00a	710,25bc	711,12ab	5445,05cd
10	6541,875b	337,92bc	108,40c	78,80bcd	1,37bc	29,67ns	710,00abc	726,87ns	895,12a	729,00bc	737,62a	6203,95c
CV(%)	9,06	5,69	5,28	5,21	4,73	28,05	4,55	14,45	2,67	4,39	5,88	9,35

Médias seguidas da mesma letra minúscula na coluna não diferem pelo teste Tukey a 5% de probabilidade. Fonte: Elaborado pelo autora.

Nas variáveis area e perimeter os lotes 1 e 6 se destacaram com as maiores médias em comparação com as outras médias (8290,12 e 412,86 pixels) e (8468,25 e 379,18 pixels). Da mesma forma que acontece com a biometria manual, sementes com maiores area e perimeter geralmente tem mais reservas nutricionais favorecendo a germinação. As variáveis área e perimeter são métricas fundamentais para caracterizar as dimensões das sementes. Estudos com variedades de milho demonstraram a utilização dessas variáveis para poder distinguir fenotípicamente as variedades (Zhao; Wu; Zhang, 2011).

O lote 3 apresentou a menor média para essas variáveis, com 4994,50 e 290,93 pixels. Os lotes 1 e 6 por apresentarem as maiores médias têm uma probabilidade de estabelecimento inicial em campo maior, quando comparado com o lote 3 que tem as menores médias. Portanto as sementes que possuem maior quantidade de reserva, aumentam as chances de um melhor estabelecimento da plântula (Haig; Westoby, 1991).

A variável feret é uma medida de dimensão longitudinal, indica que sementes com uma maior dimensão longitudinal, pode ser benéfico para o plantio e a germinação, como o lote 6 obteve a maior média, com 127,85 pixels, já o lote 3 que apresentou a menor média com 92,27 pixels, poder ter um desempenho inferior. Segundo Tang e Cai (2013), essa variável é crucial para identificação de diferentes tipos de sementes e para prever o comportamento de germinação.

Na variável breadth, que diz respeito sobre a largura do objeto de estudo, o lote 1 se destacou com a maior média, 90,24 pixels, e o lote 3 ficou com a menor média dentre as outras, 71,36 pixels. A medida de largura da semente é um indicador da capacidade de armazenamento de nutrientes da semente. Segundo Xu *et al.*, (2023) em seu trabalho, foi observado que sementes com maiores tamanho apresentaram melhores porcentagens de germinação e consequentemente de vigor, devido a maior concentração de nutrientes nas sementes. Este fator é crucial para a seleção e classificação de lotes de sementes. Por exemplo, o lote 1, com uma maior largura média de sementes (90,24 pixels), provavelmente apresentará um desempenho superior em campo comparado ao lote 3, que possui médias menores.

Para a variável aspratio, o lote 6 se destacou dentre as outras médias com uma média de 1,53 pixels, já o lote 2 se destacou com a menor média, 1,23 pixels. Para a variável

sphericity, o lote 2 obteve a maior média, com 814,00 pixels. O lote 6 apresentou a menor média, 654,25 pixels. Essas medidas estão relacionadas à proporção entre o comprimento e a largura das sementes, sugerindo que sementes mais alongadas podem apresentar melhores características de germinação e crescimento em comparação com aquelas que possuem proporções menores. Dentro da variável roundness que é a medida de arredondamento, avaliando o formato da semente, assim, quanto menor o valor, a semente será mais circular, encontramos, portanto, as maiores médias com o lote 2 com 767,87 e a menor média para o lote 6 com 663,0.

A convexity mede a regularidade da forma das sementes, temos lotes 5, 6, 8, 3, 10, 9, 2 e 7 se destacaram com as maiores médias, 927,62, 910,12, 904,25, 898,75, 895,12 886,00, 885,20 e 884,50 pixels respectivamente, sementes com alta convexidade são indicativas de uma forma mais uniforme e consistente, o que pode facilitar o manejo e a germinação McDonald, (1994). O lote 1 obteve a menor média dentre as demais, com 825,25 pixels.

Em circularity, que é uma medida de quão próxima a forma da semente está de um círculo perfeito, portanto, sementes com alta circularidade tendem a ser mais uniformes e fáceis de manusear. As melhores médias foram dos lotes 5, 8, 6, 3, 10, 7 e 2 apresentaram 791,75, 753,75, 743,00, 740,75, 737,62, 724,62 e 722,00 pixels, respectivamente. O lote 1 obteve a menor média dentre as demais, com 612,50 pixels.

Na variável endocarp que está relacionada ao endocarpo que envolve o embrião da semente o lote 6 se destacou com a maior média de 8089,06 pixels dentre as demais. Já o lote 3 apresentou a menor média, com 4703,56 pixels. O endocarpo atua como uma camada protetora que pode influenciar a germinação ao controlar a absorção de água e a proteção das sementes durante o desenvolvimento inicial Baskin e Baskin, (2022). Essa variável só é possível aferir por meio do uso de análise imagens, a biometria manual não permite extrair esse dado, tenho como vantagens quando se comparada com a análise manual, a agilidade do processo de análise, identificação automática de características específicas e a classificação automática de família ou gêneros (Lind *et al.*, 2012).

Os dados biométricos são fundamentais para determinar a qualidade e o desempenho das sementes. Os lotes 1 e 6, que se destacam em várias dessas métricas, provavelmente possuem melhores características para o cultivo, enquanto o lote 3 pode enfrentar desafios

maiores devido às suas menores médias em várias dessas variáveis.

Na tabela 10, que se refere às variáveis colorimétricas extraídas do plugin, é possível observar que para a variável mean read, o lote 7 apresentou o maior resultado, com 199,43 pixels, já o lote 1 apresentou a menor média, com 126,41 pixels.

Tabela 10 – Resultados de média de vermelho - MeanRed; média de verde - MeanGreen; média de azul - MeanBlue e média de RGB - MeanRGB avaliados com o plugin Seeds Analysis nos 10 lotes de sementes de feijão.

Lote	MeanRed	MeanGreen	MeanBlue	MeanRGB
----- Pixels -----				
1	126,41g	45,55f	39,87d	70,61
2	151,10e	82,49d	55,79c	96,46d
3	176,67c	117,93bc	92,47b	129,02bc
4	136,46fg	75,08de	48,43cd	86,66de
5	186,25bc	126,58ab	110,89a	141,24ab
6	191,62ab	134,60a	115,55a	147,26a
7	199,43a	133,60ab	114,01a	149,01a
8	136,21fg	63,72e	40,75d	80,23ef
9	146,99ef	81,22d	48,14cd	92,12de
10	163,91d	108,31c	82,20b	118,14c
CV(%)	2,90	6,73	7,43	4,59

Médias seguidas da mesma letra minúscula na coluna não diferem pelo teste Tukey a 5% de probabilidade.

Fonte: Elaborado pela autora.

Para a variável mean green, o lote 6 apresentou a melhor média, com 134,60 pixels. O lote 1 se destacou com a menor média dentre as demais, com 45,55 pixels. Na variável mean blue os lotes 6,7 e 5 se destacaram com as maiores médias, 115,55, 114,01 e 110,89 pixels, respectivamente. Os lotes 8 e 1 apresentaram as menores médias, com 40,75 e 39,87 pixels, respectivamente. Para a variável mean rgb, os lotes 7 e 6 obtiveram as melhores médias, com 149,01 e 147,26 pixels, respectivamente. O lote 1 obteve a menor média, com 70,61 pixels. As variáveis biométricas (tamanho, forma, cor e textura) são consideradas características discriminatórias importantes para a identificação automática de sementes utilizadas para identificação de diferentes tipos de sementes. Tang e Cai, (2013). Podendo ser utilizadas para programas de melhoramento por meio da realização de fenotipagem, uma vez que ela se refere à medição e análise das características observáveis de um organismo, que são influenciadas pela interação entre o genótipo (conjunto de genes) e o ambiente (Omari *et al.*,

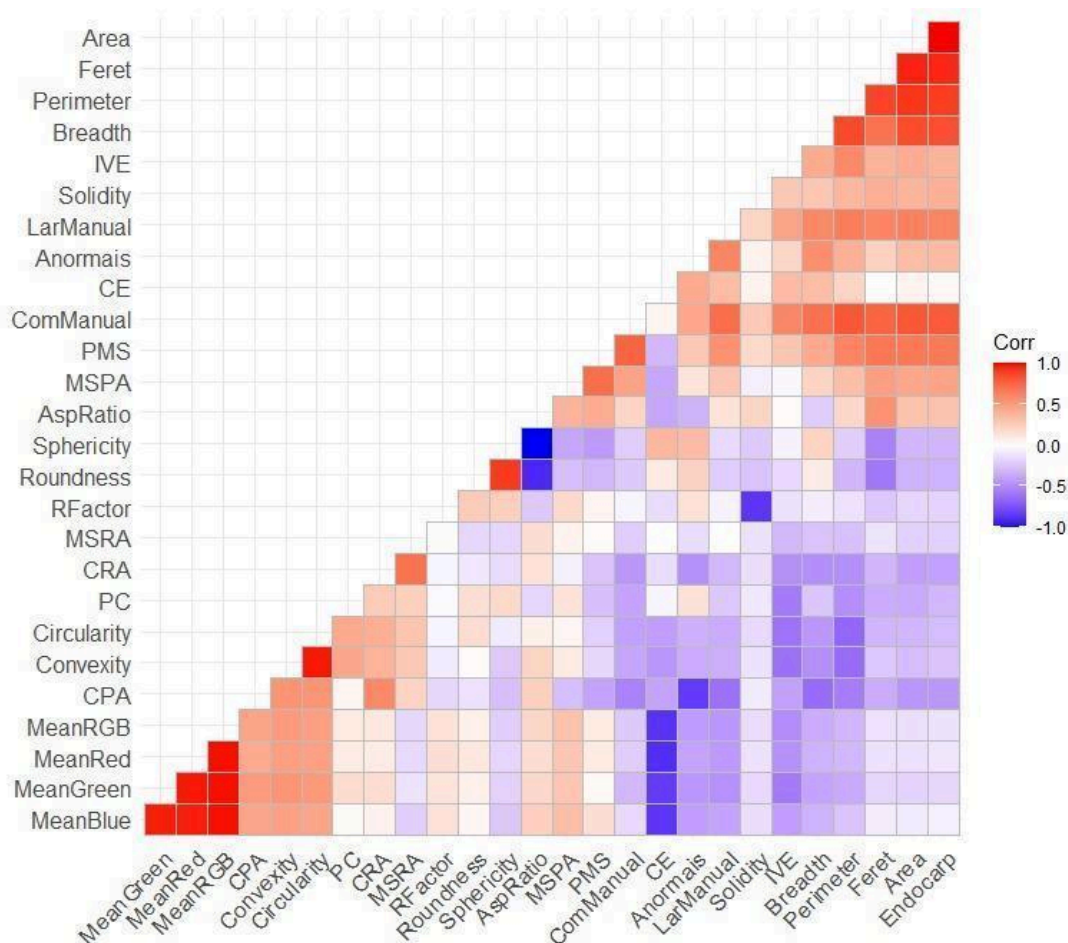
2020).

Estes resultados sugerem que as características baseadas na cor, particularmente o verde médio, o azul médio e o RGB médio, podem ser eficazes na discriminação entre diferentes tipos de sementes. Sistemas automatizados que utilizam o RGB podem classificar as sementes com base em suas características visuais, aumentando a eficiência e a precisão na seleção de semente (Nehoshtan *et al.*, 2021).

Os dados demonstrados neste artigo comprovam a eficácia da biometria por meio de análise de imagem. Na Figura 23, apresentamos a correlação de Spearman realizada entre os dados obtidos nos testes de vigor, biometria manual e análise de imagem pelo software ImageJ.

O coeficiente de correlação de Spearman, varia entre -1 e +1 e não dependem das unidades de medida das variáveis, o que facilita a sua interpretação, de modo que se avalia a intensidade e a direção da relação monótona sejam lineares ou não, não apresentando sensibilidade a assimetria na distribuição nem à presença de outliers (valores atípicos), não exigindo que os dados provenham de duas populações com distribuições normais (Sousa, 2019).

Figura 48 – Coeficientes de correlação de Spearman estimados entre os parâmetros índice de velocidade de emergência (IVE), massa seca de parte aérea (MSPA), massa seca de raiz (MSRA), comprimento de parte aérea (CPA), comprimento de raiz (CRA), peso de mil sementes (PMS), condutividade elétrica (CE), germinação (G), primeira contagem (PC) e parâmetros obtidos na análise de imagem.

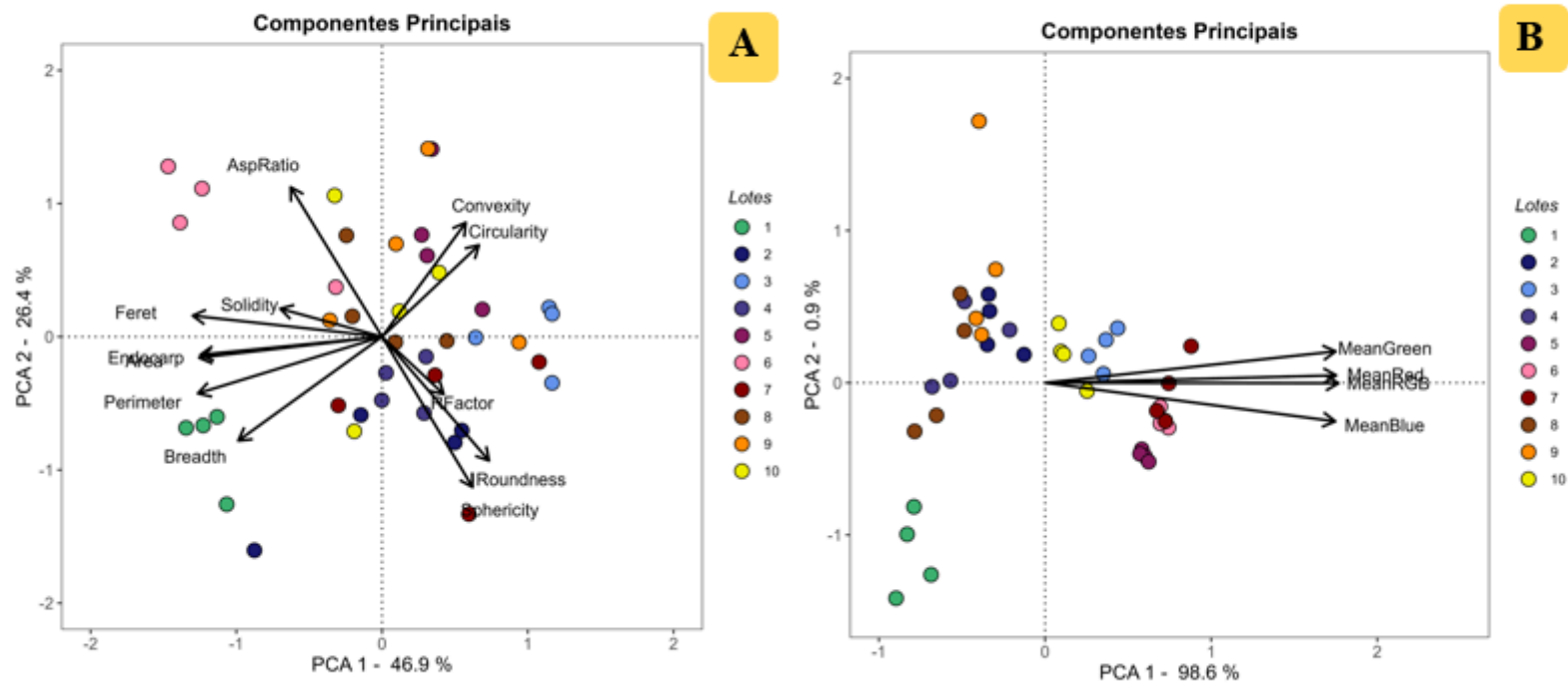


Fonte: Elaborado pela autora.

Podemos observar uma correlação positiva entre a variável G e PC com as variáveis area, perimeter, feret e breadth. Portanto, fica possível correlacionar os dados de vigor com os dados obtidos por meio da análise de imagem, demonstrando o potencial para uma futura substituição dos testes tradicionais para os realizados de forma automatizada. Com base nesses resultados, os lotes que apresentam uma alta percentagem de germinação também demonstram elevados valores nas variáveis físicas obtidas com o uso do plugin. Isso permite uma extração de dados mais robusta, proporcionando informações adicionais que vão além dos dados obtidos pelos testes tradicionais.

Com a finalidade de resumir os dados experimentais da análise de imagem, foi usada a análise PCA (Figura 5). Com base nos dados gerados pela análise de imagem foi possível realizar uma estatística multivariada de componentes principais que são importantes para compreender os dados gerados pela plug-in. As análises mostraram que os dois primeiros componentes principais permitiram explicar 73,3% da variância contida nas variáveis (Figura 5A) e 99,5 (Figura 5B) (Jolliffe; Cadima, 2016).

Figura 49 – Análise de componentes principais (PCA) (A e B) dos dez lotes; A) PCA para os parâmetros físicos e B) PCA para os parâmetros de cor RGB; PC1 = componente principal 1; PC2 = componente principal.



Fonte: Elaborado pela autora.

O componente principal 1 (CP1) e o componente principal 2 (CP2) contribuíram com 46,9% e 26,4%, respectivamente, na Figura 5A, e 98,6% e 0,9%, respectivamente, na Figura 5B. Um percentual elevado de variância explicada indica que a PCA consegue capturar de forma eficaz a estrutura subjacente dos dados, facilitando a interpretação dos resultados e a identificação de padrões relevantes.

Na Figura 5A as variáveis que apresentaram maior poder discriminatório em CP1 foram: Roundness, rfactor e sphericity essas variáveis estão relacionadas com a características de circularidade da semente, importante para manter um padrão biométrico entre os tratamentos estudados e na Figura 5.B. Esse resultado indica que estas variáveis estão fortemente correlacionadas com a CP1 e entre si, sugerindo que essas características são cruciais para a variação observada nos dados das sementes. Segundo Faulvel *et al.*, (2008) a análise de PCA pode ser usada como forma de extrair características importantes dos dados e identificar variáveis que têm maior impacto na variabilidade observada.

Variáveis com o mesmo sinal atuam de forma direta, ou seja, quando o valor de uma aumenta, o valor da outra também aumenta, e vice-versa. Por outro lado, variáveis com sinais opostos atuam de forma inversa, significando que quando o valor de uma aumenta, o valor da outra diminui. Isso é importante para entender como as diferentes características das sementes estão inter-relacionadas e como elas podem influenciar o desempenho geral. Estudos com PCA de Girgel, (2021) realizados com *Phaseolus vulgaris* indicam que as características morfoagronômicas são cruciais para o melhoramento genético dos genótipos de feijão, proporcionando variabilidade genética essencial para programas de melhoramento. E as variáveis avaliadas nesse trabalho podem ser utilizadas para estudos futuros de divergência genética assim como para fenotipagem.

A análise de PCA é uma técnica essencial para resumir e compreender dados experimentais complexos, como os dados de imagem de sementes. Ela permite a redução da dimensionalidade, facilita a visualização de padrões e ajuda na identificação das variáveis mais importantes.

Com dados gerados pelo software ImageJ foi realizada a LDA, que é uma técnica estatística usada para encontrar uma combinação linear de características que melhor separa duas ou mais classes de objetos ou eventos. O software ImageJ apresenta a vantagem de ser

gratuito e oferecer um vasto conjunto de ferramentas para análise de imagens (Schneider; Rasband; Eliceiri, 2012).

A LDA é amplamente utilizada em reconhecimento de padrões, como reconhecimento de voz e imagem, onde é importante classificar dados em diferentes categorias (Xanthopoulos *et al.*, 2013).

A utilização de LDA permitiu a avaliação da contribuição das variáveis gerada pelo plugin na classificação e separação dos lotes, o uso do algoritmo possibilitou compreender as características mais relevante que as outras no processo de discriminação dos lotes (Tabela 11), foi possível observar que os parâmetros selecionados estavam mais relacionados às características do tamanho da semente e as colorimétricas.

Tabela 11 – Ranking das variáveis morfocolorimétricas discriminantes mais importantes selecionadas e utilizadas pela LDA.

Features	Tolerance	F remove	Wilcky's lambda
Area	,006	161,516	,000
Mean Blue	,008	117,795	,000
Mean Green	,004	112,242	,000

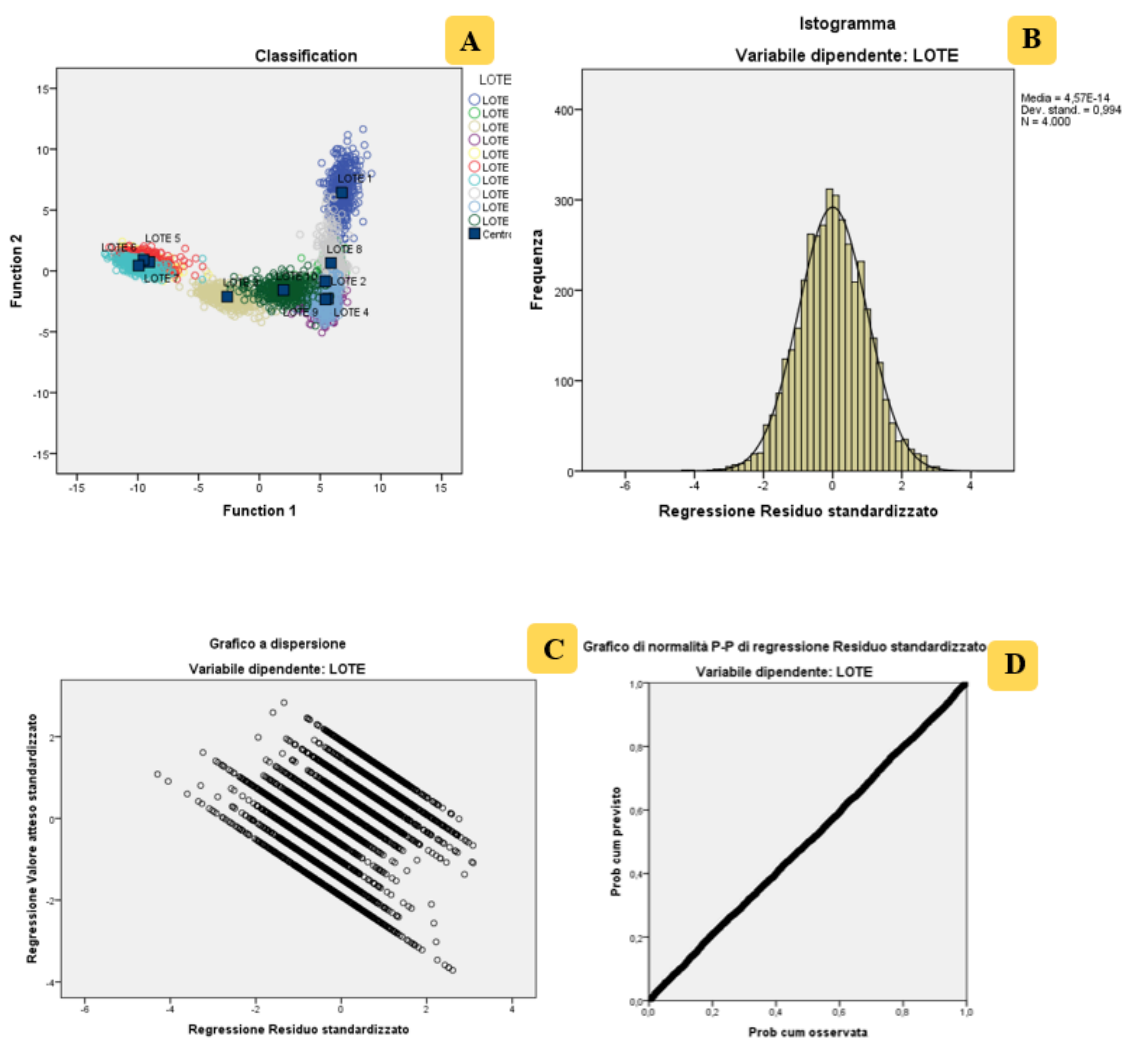
Fonte: Elaborado pela autora.

Os resultados obtidos pela LDA demonstraram que os parâmetros colorimétricos se mostraram mais discriminantes do que descritivos de forma e tamanho da semente. Essa maior discriminação pode ser atribuída aos fatores de variação entre os lotes, onde os parâmetros colorimétricos apresentaram maiores discrepâncias entre si. A literatura existente sobre a análise de sementes frequentemente destaca a importância dos atributos colorimétricos na distinção de diferentes genótipos, devido à sua maior variabilidade inter-lotes e ao seu papel crucial na identificação e classificação de variedades de sementes Tang e Cai (2013).

Os gráficos da Figura 6 possibilitaram entender graficamente a distribuição normal dos dados utilizados para comparar os lotes. Foi conduzida também avaliação da homogeneidade da variância das variáveis. A Figura 6 (B–C) mostra, respectivamente, a frequência e a

dispersão dos resíduos padronizados, enquanto o Gráfico de Probabilidade Normal (PP) apresenta a comparação entre a probabilidade cumulativa esperada e a observada (Fig. 6 (D)). O teste de normalidade de Kolmogorov–Smirnov (K–S) também foi realizado para verificar a distribuição normal dos dados, fornecendo valores de significância menores que 0,05.

Figura 50 – (A) Representação gráfica da análise discriminante (B) histograma dos resíduos padronizados, (C) gráfico de dispersão dos resíduos padronizados testados com o teste de Levene (F) e (D) gráfico de probabilidade normal (P-P) testado com o teste de Kolmogorov-Smirnov (K-S).



Fonte: Elaborada pela autora.

O gráfico de classificação da Figura 6-A nos mostra uma separação oriunda do

agrupamento dos lotes com liminaridades, gerando 3 classes. O lote 1 como é visível se mantêm distantes dos demais lotes, assim como o observado nos testes de vigor, onde ele apresenta em sua grande maioria as piores médias. Os lotes 5, 6 e 7 formam outra classe de agrupamento, o que pode responder o seu comportamento discrepante no teste de vigor de CE, e a similaridade entre os lotes com relação às médias dos demais testes de vigor. Os lotes restantes (2,3,4,8,9 e 10) formam o outro grupo de classificação.

Esses resultados são significativos, pois permitem a classificação dos lotes em alto, médio e baixo vigor utilizando o algoritmo de LDA. Este tipo de análise é suportado por diversos estudos que destacam a eficácia na discriminação de classes com base em múltiplos parâmetros. Wang *et al.*, (2021) demonstraram que a LDA é uma ferramenta poderosa para a classificação de sementes com base em características fenotípicas e genéticas, proporcionando uma base sólida para a seleção de lotes de alta qualidade.

A LDA conseguiu atingir o percentual de classificação correta de 91,02% (Tabela 12). Esses resultados são importantes para demonstrar a robustez da técnica.

Tabela 12 – Porcentagens de classificação correta e validação cruzada para os 10 lotes.

		Resultado da Classificação											
LOTE		Grupo de membros previstos											
		LOTE 1	LOTE 2	LOTE 3	LOTE 4	LOTE 5	LOTE 6	LOTE 7	LOTE 8	LOTE 9	LOTE 10	Total	
O r i g i n a l	Resu mo	LOTE 1	379	5	0	0	0	0	14	2	0	400	
		LOTE 2	0	381	3	6	0	0	5	3	2	400	
		LOTE 3	0	0	398	0	1	0	1	0	0	400	
		LOTE 4	0	9	0	333	0	0	0	7	48	3	400
		LOTE 5	0	0	0	0	360	32	8	0	0	0	400
	Resu mo	LOTE 6	0	0	0	0	59	330	11	0	0	0	400
		LOTE 7	0	0	2	0	5	5	388	0	0	0	400
		LOTE 8	6	0	0	9	0	0	0	365	8	12	400
		LOTE 9	0	9	0	47	0	0	0	3	341	0	400
		LOTE 10	0	0	4	2	0	0	0	0	5	389	400
O r i g i n a l	%	LOTE 1	94,8	1,3	0,0	0,0	0,0	0,0	0,0	3,5	,5	0,0	100,0
		LOTE 2	0,0	95,3	,8	1,5	0,0	0,0	0,0	1,3	,8	,5	100,0
		LOTE 3	0,0	0,0	99,5	0,0	,3	0,0	,3	0,0	0,0	0,0	100,0
		LOTE 4	0,0	2,3	0,0	83,3	0,0	0,0	0,0	1,8	12,0	,8	100,0
		LOTE 5	0,0	0,0	0,0	0,0	90,0	8,0	2,0	0,0	0,0	0,0	100,0
		LOTE 6	0,0	0,0	0,0	0,0	14,8	82,5	2,8	0,0	0,0	0,0	100,0
		LOTE 7	0,0	0,0	,5	0,0	1,3	1,3	97,0	0,0	0,0	0,0	100,0
		LOTE 8	1,5	0,0	0,0	2,3	0,0	0,0	0,0	91,3	2,0	3,0	100,0
		LOTE 9	0,0	2,3	0,0	11,8	0,0	0,0	0,0	,8	85,3	0,0	100,0
		LOTE 10	0,0	0,0	1,0	,5	0,0	0,0	0,0	0,0	1,3	97,3	100,0

Fonte: Elaborado pela autora. Legenda: A- para. 91,6% dos casos originais agrupados foram classificados corretamente; B- A validação cruzada é realizada apenas para esses casos na análise. Para validação cruzada, cada caso é classificado pelas funções derivadas de todos os outros casos; C- 91,2% dos casos agrupados com validação cruzada foram classificados corretamente.

Tabela 13 – Porcentagens de classificação da validação cruzada para os 10 lotes.

LOTE	Grupo de membros previstos										Total	
	LOT											
	LOTE 1	LOTE 2	LOTE 3	E 4	LOTE 5	LOTE 6	LOTE 7	LOTE 8	LOTE 9	LOTE 10		
Validação Cruzada	LOTE 1	376	6	0	0	0	0	0	15	3	0	400
	LOTE 2	0	380	3	6	0	0	0	6	3	2	400
	LOTE 3	0	0	398	0	1	0	1	0	0	0	400
	LOTE 4	0	9	0	331	0	0	0	7	50	3	400
	LOTE 5	0	0	0	0	360	32	8	0	0	0	400
	LOTE 6	0	0	0	0	59	329	12	0	0	0	400
	LOTE 7	0	0	2	0	6	6	386	0	0	0	400
	LOTE 8	6	0	0	9	0	0	0	364	8	13	400
	LOTE 9	0	9	0	49	0	0	0	3	338	1	400
	LOTE 10	0	0	4	2	0	0	0	1	5	388	400
%	LOTE 1	94,0	1,5	0,0	0,0	0,0	0,0	0,0	3,8	,8	0,0	100,0
	LOTE 2	0,0	95,0	,8	1,5	0,0	0,0	0,0	1,5	,8	,5	100,0
	LOTE 3	0,0	0,0	99,5	0,0	,3	0,0	,3	0,0	0,0	0,0	100,0
	LOTE 4	0,0	2,3	0,0	82,8	0,0	0,0	0,0	1,8	12,5	,8	100,0
	LOTE 5	0,0	0,0	0,0	0,0	90,0	8,0	2,0	0,0	0,0	0,0	100,0
	LOTE 6	0,0	0,0	0,0	0,0	14,8	82,3	3,0	0,0	0,0	0,0	100,0
	LOTE 7	0,0	0,0	,5	0,0	1,5	1,5	96,5	0,0	0,0	0,0	100,0
	LOTE 8	1,5	0,0	0,0	2,3	0,0	0,0	0,0	91,0	2,0	3,3	100,0
	LOTE 9	0,0	2,3	0,0	12,3	0,0	0,0	0,0	,8	84,5	,3	100,0
	LOTE 10	0,0	0,0	1,0	,5	0,0	0,0	0,0	,3	1,3	97,0	100,0

Fonte: Elaborado pela autora. Legenda: A- para. 91,6% dos casos originais agrupados foram classificados corretamente; B- A validação cruzada é realizada apenas para esses casos na análise. Para validação cruzada, cada caso é classificado pelas funções derivadas de todos os outros casos; C- 91,2% dos casos agrupados com validação cruzada foram classificados corretamente.

6.4 Conclusão

Evidenciou-se o potencial da análise de imagem digital como ferramenta para caracterizar biometricamente e colorimetricamente os lotes de sementes de feijão-caupi, demonstrando correlações significativas entre dados de vigor, germinação e características morfo-colorimétricas extraídas pelo plugin Seeds Analysis no ImageJ. Lotes com melhores

médias de germinação e vigor também se destacaram nas variáveis extraídas das imagens, especialmente nos índices de comprimento, largura, área, perímetro e tonalidade RGB, os quais se mostraram bons indicadores do potencial inicial de desenvolvimento das sementes. Além disso, a Análise de Componentes Principais (PCA) permitiu agrupar as características avaliadas, explicando grande parte da variância e possibilitando uma eficiente discriminação entre os lotes. Da mesma forma, a Análise Discriminante Linear (LDA) destacou-se na separação dos lotes, reforçando a precisão da metodologia digital na distinção de sementes com alto potencial de desempenho. Esses resultados sugerem que a análise de imagem digital pode ser uma alternativa precisa e viável para otimizar o processo de seleção de sementes de alta qualidade.

CONSIDERAÇÕES FINAIS

A pesquisa desenvolvida evidencia o enorme potencial das tecnologias de VC e deep learning na otimização do processo de análise de sementes, especificamente no contexto do feijão-caupi. A implementação de um sistema de visão artificial para a análise do teste de tetrazólio, em particular, mostrou-se uma solução prática e poderosa para o setor sementeiro, permitindo a classificação automatizada e a identificação de padrões complexos que seriam difíceis de detectar por métodos convencionais. Através da integração de ferramentas avançadas como o plugin Seeds Analysis do software ImageJ e a construção de um banco de dados robusto de imagens, foi possível não apenas validar a eficácia dessas metodologias, mas também demonstrar sua capacidade de superar as limitações dos métodos tradicionais. O uso dessas tecnologias eliminou a subjetividade inerente às avaliações manuais, aumentando a precisão e a eficiência na classificação e análise de sementes.

Além de contribuir para a melhoria das práticas agrícolas, esse trabalho estabelece uma base sólida para futuras pesquisas e inovações no campo da análise de sementes, demonstrando que a aplicação de tecnologias de aprendizado profundo e VC pode capacitar agricultores de diferentes escalas a tomarem decisões mais informadas e precisas. A integração dessas inovações tecnológicas no setor sementeiro representa não apenas um avanço significativo em termos de qualidade e eficiência, mas também um passo crucial na modernização do agronegócio, promovendo uma agricultura mais sustentável e competitiva.

REFERÊNCIAS

- ABRAMOFF, Michael D.; MAGALHÃES, Paulo J.; RAM, Sunanda J. Image processing with ImageJ. **Biophotonics international**, v. 11, n. 7, p. 36-42, 2004.
- ABUD, Haynna Fernandes *et al.* Caracterização morfológica dos frutos de pimenta malagueta e biquinho. **Revista Brasileira de Agropecuária Sustentável (RBAS)**, vol. 8, n. 2, p. 29-39, junho, 2018.
- AGRESTI, Alan. **Categorical data analysis**. John Wiley & Sons, 2012.
- AL-ROUSAN, Thamer; ABUALESE, Hasan. Simplifying the structural complexity of software systems. **Cybernetics and Information Technologies**, v. 19, n. 3, p. 57-73, 2019.
- ALGINAHI, Yasser *et al.* Preprocessing techniques in character recognition. **Character recognition**, v. 1, p. 1-19, 2010.
- ALTUNTAŞ, Yahya; CÖMERT, Zafer; KOCAMAZ, Adnan Fatih. Identification of haploid and diploid maize seeds using convolutional neural networks and a transfer learning approach. **Computers and Electronics in Agriculture**, v. 163, p. 104874, 2019.
- ANTUNES Neto, A. *et al.* Determinação de sementes de soja esverdeadas por meio de análise de imagens. Sete Lagoas, MG: Embrapa Milho e Sorgo, 22 p. (**Embrapa Milho e Sorgo**. Boletim de Pesquisa e Desenvolvimento, 212). 2020.
- BARROS, Daniela Inácio *et al.* Uso do teste de tetrazólio para avaliação da qualidade fisiológica de sementes de abobrinha. **Revista Brasileira de Sementes**, v. 27, p. 165-171, 2005.
- BASKIN, Jerry M.; BASKIN, Carol C. Seed (true seed plus endocarp) dormancy in Anacardiaceae in relation to infrafamilial taxonomy and endocarp anatomy. **Seed Science Research**, v. 32, n. 4, p. 187-199, 2022.
- BASSO, B., Cammarano, D., & Carfagna, E.. “A review of crop yield prediction using machine learning techniques.” **Computers and Electronics in Agriculture**, 127, 185-194, 2016.
- BEIZER, Boris. **Software testing techniques**. dreamtech Press, 2003.
- BEKKERING, Cody S.; HUANG, Jin; TIAN, Li. Image-based, organ-level plant phenotyping for wheat improvement. **Agronomy**, v. 10, n. 9, p. 1287, 2020.
- BERRAR, Daniel *et al.* Cross-validation. 2019.
- BEWLEY, J. D., & BLACK, M.. **Seeds: physiology of development and germination**, 1994.
- BRASIL. Ministério da Agricultura, Pecuária e Abastecimento. **Regras para análise de sementes**. Brasília: Mapa/ACS, p. 399, 2009.

CARVALHO, Maria Laene Moreira de et al. Teste de raios X na avaliação da qualidade de sementes de abóbora. **Revista Brasileira de Sementes**, v. 31, p. 221-227, 2009.

CERVI, Fernanda; MENDONÇA, Elisabeth Aparecida Furtado. Adequação do teste de tetrazólio para sementes de algodoeiro. **Revista Brasileira de Sementes**, v. 31, p. 177-186, 2009.

CHANDRA, Akshay L. *et al.* Computer vision with deep learning for plant phenotyping in agriculture: A survey. **arXiv preprint arXiv:2006.11391**, 2020.

CHEN, Junde *et al.* Using deep transfer learning for image-based plant disease identification. **Computers and Electronics in Agriculture**, v. 173, p. 105393, 2020.

CHICCO, Davide; JURMAN, Giuseppe. The advantages of the Matthews correlation coefficient (MCC) over F1 score and accuracy in binary classification evaluation. **BMC genomics**, v. 21, p. 1-13, 2020.

DA SILVA, Fernando Henrique Alves et al. Atributos físicos e fisiológicos de sementes salvas de feijão-caupi utilizadas no semiárido brasileiro. **Revista Caatinga**, v. 32, n. 1, p. 113-120, 2019.

DE ABREU ARAÚJO, Beatriz *et al.* Veranicos na produtividade de feijão-caupi (*Vigna unguiculata*) no município de Crateús, Ceará. **Revista verde de agroecologia e desenvolvimento sustentável**, v. 14, n. 2, p. 312-316, 2019.

DE OLIVEIRA, Luciana Magda; DE CARVALHO, Maria Laene Moreira; DAVIDE, Antonio Claudio. Teste de tetrazólio para avaliação da qualidade de sementes de *Peltophorum dubium* (Sprengel) Taubert-Leguminosae Caesalpinioideae. **Cerne**, v. 11, n. 2, p. 159-166, 2005.

DE QUEIROZ, José Eustáquio Rangel; GOMES, Herman Martins. Introdução ao processamento digital de imagens. **Rita**, v. 13, n. 2, p. 11-42, 2006.

DEMINICIS, Bruno Borges; VIEIRA, Henrique Duarte; SILVA, Roberto Ferreira da. Teste de tetrazólio para avaliação da qualidade de sementes de *Clitoria ternatea* L. **Revista Brasileira de Sementes**, v. 31, p. 54-62, 2009.

DIAS, Maria Cristina Leme de Lima; ALVES, Sérgio José. Avaliação da viabilidade de sementes de *Panicum maximum* Jacq pelo teste de tetrazólio. **Revista Brasileira de Sementes**, v. 30, p. 152-158, 2008.

DODGE, Samuel; KARAM, Lina. Understanding how image quality affects deep neural networks. In: **2016 eighth international conference on quality of multimedia experience (QoMEX)**. IEEE, 2016. p. 1-6.

FAUVEL, M. *et al.*, Spectral and spatial classification of hyperspectral data using SVMs and morphological profiles. **IEEE Transactions on Geoscience and Remote Sensing**, 46(11), 2008.

FERNANDES, Ana Clara Gonçalves *et al.* Convolutional neural networks in the qualitative improvement of sweet potato roots. **Scientific Reports**, v. 13, n. 1, p. 8429, 2023.

FINCH-SAVAGE, William E.; BASSEL, George W. Seed vigour and crop establishment: extending performance beyond adaptation. **Journal of experimental botany**, v. 67, n. 3, p. 567-591, 2016.

FLECK, Leandro *et al.* Redes neurais artificiais: Princípios básicos. **Revista Eletrônica Científica Inovação e Tecnologia**, v. 1, n. 13, p. 47-57, 2016.

FOGAÇA, Cristiane Alves *et al.* Aplicação do teste de tetrazólio em sementes de *Gleditschia amorphoides* Taub. Caesalpinaceae. **Revista Brasileira de Sementes**, v. 28, p. 101-107, 2006.

FOWLER, Martin. **Patterns of enterprise application architecture**. Addison-Wesley, 2012.

FOWLER, Martin. **Refactoring: improving the design of existing code**. Addison-Wesley Professional, 2018.

FRANÇA-NETO, *et al.* Metodologia do teste de tetrazolio em semente de soja. **EMBRAPA-CNPSO. Documentos**, v. 32, 1988.

FRANÇA-NETO, J.B; KRZYZANOWSKI, Francisco Carlos; DA COSTA, Nilton Pereira. **O teste de tetrazólio em sementes de soja**. Embrapa Soja, 1998.

FRANÇA NETO, José B.; Kryzanowski, F. C. Teste de tetrazólio para determinação do vigor de sementes. **Vigor de sementes: conceitos e testes. Londrina: ABRATES**, p. 1-7, 1999.

FRANÇA NETO, J. D. B., Krzyzanowski, F. C. Metodologia do teste de tetrazólio em sementes de soja. 1ª ed. Londrina : **Embrapa Soja**, 53p. 2018.

FRANÇA-NETO, José de Barros; KRZYZANOWSKI, Francisco Carlos. Tetrazolium: an important test for physiological seed quality evaluation. **Journal of Seed Science**, v. 41, n. 3, p. 359-366, 2019.

FREIRE FILHO, F.R.; Origem, evolução e domesticação do caupi. 1º ed. Teresina, Piauí, Brasil: **EMBRAPA/UEPAE**, 21 p., 1988.

FREIRE FILHO, Francisco Rodrigues. **Feijão-caupi: avanços tecnológicos**. Embrapa Informação Tecnológica; Teresina, PI: Embrapa Meio-Norte, 2005.

FREIRE FILHO, Francisco Rodrigues *et al.* Feijão-caupi no Brasil: produção, melhoramento genético, avanços e desafios. 2011.

GHAFFARI, Akram. Precision seed certification through machine learning. **Technology in Agronomy**, n. tia-0024-0013, p. 1-12, 2024.

GIRGEL, U. Principle component analysis (PCA) of bean genotypes (*Phaseolus vulgaris* L.) concerning agronomic, morphological and biochemical characteristics. **Applied Ecology and Environmental Research**, v. 19, n. 3, p. 1999-2011, 2021.

GONZALEZ, Rafael C.; WOODS, Richard E. **Processamento de imagens digitais**. Editora Blucher, 2000.

GONZALEZ, R. C., Woods, R. E., & Eddins, S. L.. Digital image processing using MATLAB. **Pearson Prentice Hall**, 2004.

GONZALEZ, Rafael C. **Digital image processing**. Pearson education india, 2009.

GONZALEZ, Rafael C.; WOODS, Richard E. Processamento digital de imagem. **Pearson**, ISBN-10: 8576054019, v. 10, p. 11-27, 2010.

GOODFELLOW, Ian. Deep learning. **MIT press**, 2016.

GOUGH, Robert E. **Seed quality: basic mechanisms and agricultural implications**. CRC Press, 2020.

GULZAR, Yonis *et al.* A convolution neural network-based seed classification system. **Symmetry**, v. 12, n. 12, p. 2018, 2020.

GUNTURK, Bahadir; LI, Xin. **Image restoration**. CRC Press, 2018.

GUO, Yanming et al. Deep learning for visual understanding: A review. **Neurocomputing**, v. 187, p. 27-48, 2016.

GRASSO, Rosaria et al. Non-destructive evaluation of watermelon seeds germination by using Delayed Luminescence. **Journal of Photochemistry and Photobiology B: Biology**, v. 187, p. 126-130, 2018.

HAIG, D.; Westoby, M. Seed size, pollination casts and angiosperm success. **Evolutionary Ecology**, v.5, p.231-247, 1991.

HAWKINS, Douglas M. The problem of overfitting. **Journal of chemical information and computer sciences**, v. 44, n. 1, p. 1-12, 2004.

HE, Haibo; GARCIA, Edwardo A. Learning from imbalanced data. **IEEE Transactions on knowledge and data engineering**, v. 21, n. 9, p. 1263-1284, 2009.

HOFFMASTER, A. L. *et al.* An automated system for vigor testing three-day-old soybean seedlings. **Seed Science and Technology**, v. 31, n. 3, p. 701-713, 2003.

HOWARTH, M. S.; STANWOOD, P. C. Tetrazolium staining viability seed test using color image processing. **Transactions of the ASAE**, v. 36, n. 6, p. 1937-1940, 1993.

HUNT, A., e Thomas, H. The Pragmatic Programmer: From Journeyman to Master. **Addison-Wesley**. 1999.

ISTA (International Seed Testing Association). International rules for seed testing. Bassersdorf, Suíça: **ISTA 2018**.

ISTA. Regras internacionais para testes de sementes. Associação Internacional de Testes de Sementes, **Bassersdorf**, 2019.

- JAIN, Anil K. **Fundamentals of digital image processing**. Prentice-Hall, Inc., 1989.
- JADON, Shruti. A survey of loss functions for semantic segmentation. In: **2020 IEEE conference on computational intelligence in bioinformatics and computational biology (CIBCB)**. IEEE, p. 1-7, 2020.
- JANOCHA, Katarzyna; CZARNECKI, Wojciech Marian. On loss functions for deep neural networks in classification. **arXiv preprint arXiv:1702.05659**, 2017.
- JAVORSKI, Maicon *et al.* Image analysis to evaluate the physiological potential and morphology of pearl millet seeds. **Journal of Seed Science**, v. 40, n. 2, p. 127-134, 2018.
- JESUS, Lorena L. *et al.* Teste de tetrazólio para sementes de *Sesamum indicum*. **Revista de Ciências Agrárias**, v. 38, n. 3, p. 422-428, 2015.
- JOLLIFFE, I. T., & Cadima, J.. Principal component analysis: A review and recent developments. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 374, 2016.
- KAMILARIS, Andreas; PRENAFETA-BOLDÚ, Francesc X. Deep learning in agriculture: A survey. **Computers and electronics in agriculture**, v. 147, p. 70-90, 2018.
- KUO, Tzu-Yi *et al.* Identifying Rice Grains of Various Cultivars Using Machine Vision. In: **2015 ASABE Annual International Meeting**. American Society of Agricultural and Biological Engineers, . p. 1, 2015.
- KURTULMUŞ, F.; İlknur Alibaş, I.; Kavdir, I. Classification of pepper seeds using machine vision based on neural network, **Int J Agric & Biol Eng**, v.9, n.1 p.51- 54, 2016.
- LECUN, Yann *et al.* Gradient-based learning applied to document recognition. **Proceedings of the IEEE**, v. 86, n. 11, p. 2278-2324, 1998.
- LECUN, Yann; BENGIO, Yoshua; HINTON, Geoffrey. Deep learning. **nature**, v. 521, n. 7553, p. 436-444, 2015.
- LIU, Fei *et al.* Digital techniques and trends for seed phenotyping using optical sensors. **Journal of Advanced Research**, 2023.
- LEITE, Cam *et al.* Validação do sistema de análise de imagens Vigor-S para a determinação de fitotoxicidades em plântulas de soja. 2018.
- LI, Chen; HUANG, Zengyi. Using Artificial Intelligence to Refine the Implementation Trajectory of Digital Image Processing Technology. **Frontiers in Computing and Intelligent Systems**, v. 8, n. 1, p. 112-115, 2024.
- LIND, R.: Open source software for image processing and analysis: picture this with imagej. IN: HARLAND, L., FORSTER, M. (eds.) *Open Source Software in Life Science Research*. Woodhead Publishing Series in Biomedicine, pp. 131–149. **Woodhead Publishing, Sawston**, 2012.

- LIU, Tao; ABD-ELRAHMAN, Amr. Deep convolutional neural network training enrichment using multi-view object-based analysis of Unmanned Aerial systems imagery for wetlands classification. **ISPRS Journal of Photogrammetry and Remote Sensing**, v. 139, p. 154-170, 2018.
- LOBO, Luiz Carlos. Inteligência artificial, o futuro da medicina ea educação médica. **Revista Brasileira de Educação Médica**, v. 42, n. 3, p. 3-8, 2018.
- LODDO, Andrea *et al.* An effective and friendly tool for seed image analysis. **The Visual Computer**, v. 39, n. 1, p. 335-352, 2023.
- MA, Te; TSUCHIKAWA, Satoru; INAGAKI, Tetsuya. Rapid and non-destructive seed viability prediction using near-infrared hyperspectral imaging coupled with a deep learning approach. **Computers and Electronics in Agriculture**, v. 177, p. 105683, 2020.
- MAGUIRE, J. D. Speed of germination aid in selection and evaluation for seedling emergence and vigor. *Crop Science*, Madison, v. 2, n. 2, p.176-77, 1962.
- MAHAJAN, Shveta; MITTAL, Sudesh Kumar; DAS, Amitava. Machine vision based alternative testing approach for physical purity, viability and vigour testing of soybean seeds (Glycine max). **Journal of food science and technology**, v. 55, n. 10, p. 3949-3959, 2018.
- MARCOS FILHO, Júlio. Teste de envelhecimento acelerado. **Vigor de sementes: conceitos e testes**, 1999.
- MARCOS FILHO, Julio; KIKUTI, Ana Lúcia Pereira; LIMA, Liana Baptista de. Métodos para avaliação do vigor de sementes de soja, incluindo a análise computadorizada de imagens. **Revista Brasileira de sementes**, v. 31, p. 102-112, 2009.
- MARCOS FILHO, Julio. Seed vigor testing: an overview of the past, present and future perspective. **Scientia agricola**, v. 72, n. 4, p. 363-374, 2015.
- MARCOS FILHO, Júlio. Fisiologia de sementes de plantas cultivadas. 2015.
- MARCOS FILHO, Julio. Seed vigor testing: an overview of the past, present and future perspective. **Scientia agricola**, v. 72, p. 363-374, 2015.
- MARCOS FILHO, J. Seed Physiology of Cultivated Plants. **Abrates**. 2016.
- MCCULLOCH, Warren S.; PITTS, Walter. A logical calculus of the ideas immanent in nervous activity. **The bulletin of mathematical biophysics**, v. 5, p. 115-133, 1943.
- McDONALD JR, M.B. A review and evaluation of seed vigor test. **Proceedings of the International Seed Testing Association**. Lansing, v.65, n. 1, p.109-139, 1975.
- MCDONALD, M. B. A review and evaluation of seed vigor tests. **Proceedings of Association of Official Seed Analysts** 65, p. 109-139, 1975.

MCDONALD JR, Miller B.; PHANEENDRANATH, Bangalore R. A modified accelerated aging seed vigor test for soybeans. **Journal of Seed Technology**, p. 27-37, 1978.

MCDONALD, M. B.. **Seed** germination and seedling establishment. Physiology and determination of crop yield, 37-60,1994.

MNIH, Volodymyr et al. Human-level control through deep reinforcement learning. **nature**, v. 518, n. 7540, p. 529-533, 2015.

MONDO, Vitor Henrique Vaz; CICERO, Silvio Moure. Análise de imagens na avaliação da qualidade de sementes de milho localizadas em diferentes posições na espiga. **Revista Brasileira de Sementes**, v. 27, p. 9-18, 2005.

NAKAGAWA, J. Testes de vigor baseados no desempenho das plântulas. In: Krzyzanowski, F.C.; Vieira, R.D.; França Neto, J.B. (Ed.). Vigor de sementes: conceitos e testes. Londrina: **ABRATES**. p.2.1-2.24, 1999

NEHOSHTAN, Yuval *et al.* Robust seed germination prediction using deep learning and RGB image data. **Scientific reports**, v. 11, n. 1, p. 22030, 2021.

OMARI, Mohammad Kamran et al. Digital image-based plant phenotyping: a review. **Korean Journal of Agricultural Science**, v. 47, n. 1, p. 119-130, 2020.

OTERO, Carlos. **Software engineering design: theory and practice**. CRC Press, 2012.

PATRÍCIO, Diego Inácio; RIEDER, Rafael. Computer vision and artificial intelligence in precision agriculture for grain crops: A systematic review. **Computers and electronics in agriculture**, v. 153, p. 69-81, 2018.

PERRY, D.A. A vigor test for seeds of barley (*Hordeum vulgare*) based on measurement of plumule growth. **Seed Science and Technology**, vol. 5, p. 709 – 719, 1977.

PRESSMAN, Roger S. **Software engineering: a practitioner's approach**. Palgrave macmillan, 2005.

PINTO, Tais Leite Ferreira et al. Avaliação da viabilidade de sementes de pinhão manso pelos testes de tetrazólio e de raios X. **Revista Brasileira de Sementes**, v. 31, p. 195-201, 2009.

PRATAP, Harshita *et al.* SoyNet: Deep Learning Approaches for Automated Soybean Seed Quality Assessment. In: **2024 2nd International Conference on Advancement in Computation & Computer Technologies (InCACCT)**. IEEE, 2024. p. 613-617.

QIAO, Juxiang *et al.* Vigour testing for the rice seed with computer vision-based techniques. **Frontiers in Plant Science**, v. 14, p. 1194701, 2023.

R CORE TEAM. R: A Language and Environment for Statistical Computing. **R Foundation for Statistical Computing**, Vienna, Austria. <<https://www.R-project.org/>>.2023.

RAJJOU, Loïc *et al.* Seed germination and vigor. **Annual review of plant biology**, v. 63, n.

1, p. 507-533, 2012.

RANDINI, Margherita; BAGLI, Enrico; VISANI, Giorgio. Metrics for multi-class classification: **an overview**. **arXiv preprint arXiv:2008.05756**, 2020.

RASBAND, Wayne. Image processing and analysis in java [online available]. **http://rsbweb.nih.gov/ij**. v. 9, 2012.

REGO, Carlos Henrique Queiroz *et al.* Assessing the vigor of cowpea seeds using the Vigor-S software. **Journal of Seed Science**, v. 43, p. e202143018, 2021.

REZAEI, Mohammad *et al.* Deep convolutional neural network for complex wetland classification using optical remote sensing imagery. **IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing**, v. 11, n. 9, p. 3030-3039, 2018.

RIBEIRO, Leonardo Monteiro *et al.* Critérios para o teste de tetrazólio na estimativa do potencial germinativo em macaúba. **Pesquisa Agropecuária Brasileira**, v. 45, p. 361-368, 2010.

RODRIGUES, Ana Paula Medeiros dos Santos *et al.* Teste de tetrazólio para avaliação da qualidade fisiológica de sementes de *Vigna unguiculata* (L.) Walp. **Revista Ciência Agrônômica**, v. 46, p. 638-644, 2015.

RODRIGUES, Mayara; GOMES-JUNIOR, Francisco G.; MARCOS-FILHO, Julio. Vigor-S: system for automated analysis of soybean seed vigor. **Journal of Seed Science**, v. 42, p. e202042039, 2020

dos SANTOS, S. R. G., & de Paula, R. C.. Testes de vigor para avaliação da qualidade fisiológica de sementes de *Sebastiania ommersoniana* (Baill.) Smith and Downs. **Scientia Forestalis (Brazil)**, v. 37, n. 81, 2009.

SAITO, Takaya; REHMSMEIER, Marc. The precision-recall plot is more informative than the ROC plot when evaluating binary classifiers on imbalanced datasets. **PloS one**, v. 10, n. 3, p. e0118432, 2015.

SALEEM, Nasla *et al.* Class-specific data augmentation for plant stress classification. **The Plant Phenome Journal**, v. 7, n. 1, p. e20112, 2024.

SARIGU, Marco *et al.* Taxonomic discrimination of the *Paeonia mascula* group in the Tyrrhenian Islands by seed image analysis. **Systematics and biodiversity**, v. 17, n. 8, p. 801-810, 2019.

SARIGU, Marco *et al.* Phenotypic identification of plum varieties (*Prunus domestica* L.) by endocarps morpho-colorimetric and textural descriptors. **Computers and Electronics in Agriculture**, v. 136, p. 25-30, 2017.

SCHNEIDER, Caroline A.; RASBAND, Wayne S.; ELICEIRI, Kevin W. NIH Image to ImageJ: 25 years of image analysis. **Nature methods**, v. 9, n. 7, p. 671-675, 2012.

- SHARMA, Abhinav et al. Machine learning applications for precision agriculture: A comprehensive review. **IEEE Access**, v. 9, p. 4843-4873, 2020.
- SHI, Tingting *et al.*, Recent advances in plant disease severity assessment using convolutional neural networks. **Scientific Reports**, v. 13, n. 1, p. 2336, 2023.
- SILVA, Felipe *et al.* **Soja: do plantio à colheita**. Oficina de Textos, 2022.
- SILVA, Rosemeire Carvalho da et al. Adaptação do teste de tetrazólio para avaliação da viabilidade e do vigor de sementes de girassol. **Pesquisa Agropecuária Brasileira**, v. 48, p. 105-113, 2013.
- SILVEIRA, Maria Angelica Moreira; VILLELA, Francisco Amaral; TILLMANN, Maria Ângela André. Comparação de métodos para avaliação da qualidade fisiológica em sementes de calêndula. **Revista Brasileira de Sementes**, v. 24, p. 24-30, 2002.
- SOUSA, Á.. Coeficiente de correlação de Pearson e coeficiente de correlação de Spearman: o que medem e em que situações devem ser utilizados? **Correio dos Açores**, 19-19, 2019.
- SPÖRL, Christiane; CASTRO, Emiliano; LUCHIARI, Aílton. Aplicação de redes neurais artificiais na construção de modelos de fragilidade ambiental. **Revista do Departamento de Geografia**, v. 21, p. 113-135, 2011.
- SPSS. SPSS base 16.0. **Application guide**. New Jersey: Prentice Hall, 2006.
- STÅHL, Niclas; WEIMANN, Lisa. Identifying wetland areas in historical maps using deep convolutional neural networks. **Ecological Informatics**, v. 68, p. 101557, 2022.
- SZELISKI, Richard. **Computer vision: algorithms and applications**. Springer Nature, 2022.
- TANG, Minnan; CAI, Cheng. Weed seeds identification based on structure elements' descriptor. In: **2013 Asia-Pacific Signal and Information Processing Association Annual Summit and Conference**. IEEE, 2013. p. 1-4.
- TAVARES, Lizandro Ciciliano *et al.* Estratégias de marketing na área de sementes. **Arquivos do Instituto Biológico**, v. 83, p. e0882014, 2016.
- TEAM, R. Core. R language and environment for statistical computing, R Foundation for Statistical. **Computing**, 2020.
- ULLMANN, Renan *et al.* Qualidade fisiológica das sementes de sorgo sacarino submetidas à secagem em diferentes condições de ar. **Revista Brasileira de Engenharia Agrícola e Ambiental**, v. 19, n. 1, p. 64-69, 2015.
- VALE, A. M. P. G. *et al.* A new automatic approach to seed image analysis: From acquisition to segmentation. **arXiv preprint arXiv:2012.06414**, 2020.
- VALVERDE-ALBACETE, Francisco J.; PELÁEZ-MORENO, Carmen. Two information-theoretic tools to assess the performance of multi-class classifiers. **Pattern Recognition Letters**, v. 31, n. 12, p. 1665-1671, 2010.

- VAN VEEN, Fjodor. Neural network zoo prequel: Cells and layers. **Retried from <https://www.asimovinstitute.org/author/fjodorvanveen>**, 2017.
- VENIAL, Lucimara Ribeiro *et al.* Biometria e armazenamento de sementes de genótipos de cacauero. **Pesquisa Florestal Brasileira**, v. 37, n. 89, p. 39-46, 2017.
- VIZCAYNO, J. F., Hugo, W., & Sanz Alvarez, J.. Variedades de sementes apropriadas para pequenos agricultores: práticas fundamentais para implementadores de RRC. 2014.
- VN, Kapadia; SASIDHARAN, N.; PATIL, Kalyanrao. Seed Image Analysis and Its Application in Seed Science Research. **Seed**, v. 7, n. 2, 2017.
- WANG, Shaohui *et al.* Practical implementation of artificial intelligence-based deep learning and cloud computing on the application of traditional medicine and western medicine in the diagnosis and treatment of rheumatoid arthritis. **Frontiers in pharmacology**, v. 12, p. 765435, 2021.
- WANG, Xuemeng *et al.* Non-destructive identification of naturally aged alfalfa seeds via multispectral imaging analysis. **Sensors**, v. 21, n. 17, p. 5804, 2021.
- WATERWORTH, Wanda M.; BRAY, Clifford M.; WEST, Christopher E. The importance of safeguarding genome integrity in germination and seed longevity. **Journal of Experimental Botany**, v. 66, n. 12, p. 3549-3558, 2015.
- XANTHOPOULOS, Petros *et al.* Linear discriminant analysis. **Robust data mining**, p. 27-33, 2013.
- XU, Shan *et al.* Identification of a Seed Vigor–Related QTL cluster associated with weed competitive ability in direct–seeded rice (*Oryza Sativa* L.). **Rice**, v. 16, n. 1, p. 45, 2023.
- YANG, Guofeng *et al.* Convolutional rebalancing network for the classification of large imbalanced rice pest and disease datasets in the field. **Frontiers in Plant Science**, v. 12, p. 671134, 2021.
- ZHAO, Min; WU, Wen Fu; ZHANG, Ya Qiu. The corn seed image segmentation and measurement of the geometrical features based on image analysis. **Applied Mechanics and Materials**, v. 66, p. 1100-1105, 2011.
- ZHAO, Liang; HAQUE, S. M.; WANG, Ruoqing. Automated seed identification with computer vision: challenges and opportunities. **Seed Science and Technology**, v. 50, n. 2, p. 75-102, 2022.
- ZUCARELI, Claudemir *et al.* Teste de deterioraç o controlada na avalia o do vigor de sementes de milho. **Revista Brasileira de Sementes**, v. 33, p. 732-742, 2011.