

PROGRAMA DE PÓS-GRADUAÇÃO EM PRODUÇÃO VEGETAL

Sandra Eulália Santos Faria

**FENOTIPAGEM DE ALTA EFICIÊNCIA PELA ANÁLISE COMPUTACIONAL
DE IMAGENS NO MELHORAMENTO GENÉTICO DO TOMATEIRO**

Montes Claros

2024

Sandra Eulália Santos Faria

**FENOTIPAGEM DE ALTA EFICIÊNCIA PELA ANÁLISE COMPUTACIONAL
DE IMAGENS NO MELHORAMENTO GENÉTICO DO TOMATEIRO**

Tese apresentada ao Programa de Pós-Graduação em Produção Vegetal da Universidade Federal de Minas Gerais, como requisito parcial para a obtenção do título de Doutora em Produção Vegetal.

Orientador: Alcinei Místico Azevedo

Montes Claros-MG

Março de 2024

F224f Faria, Sandra Eulália Santos.

2024

Fenotipagem de alta eficiência pela análise computacional de imagens no melhoramento genético do tomateiro [manuscrito] / Sandra Eulália Santos Faria.

Montes Claros, 2024.

113f. .il.

Tese (doutorado) - Área de concentração em Produção Vegetal. Universidade Federal de Minas Gerais / Instituto de Ciências Agrárias.

Orientador: Alcinei Místico de Azevedo

Banca examinadora: Silvia Nietsche, Derly José Henrique da Silva, Tiago Olivoto, Ana Clara Gonçalves Fernandes.

Inclui referências: f. 37-44; 69-72; 90-92; 110-112

1. Inteligência Computacional -- Teses. 2. Plantas -- Melhoramento genético - Teses. 3. Análise de imagem -- Teses. 4. Genética vegetal -- Teses. 5. Tomate -- Teses. I. Azevedo, Alcinei Místico. II. Universidade Federal de Minas Gerais. Instituto de Ciências Agrárias. III. Título.

CDU: 631.52

ELABORADA PELA BIBLIOTECA UNIVERSITÁRIA DO ICA/UFGM

Nádia Cristina Oliveira Pires / CRB-6/2781

ATA DE DEFESA DE TESE

Aos 15 dias do mês de março de 2024 às 8:10 horas, sob a Presidência do Professor Alcinei Místico Azevedo, D. Sc. (Orientador - UFMG/ICA) e com a participação dos Professores Sílvia Nietsche, D. Sc. (UFMG/ICA), Derly José Henrique da Silva, D. Sc. (UFV), Tiago Olivoto, D. Sc. (UFSC), e de Ana Clara Gonçalves Fernandes, D. Sc., reuniu-se, por videoconferência, a Banca de tese de **Sandra Eulália Santos Faria**, aluna do Curso de Doutorado em Produção Vegetal. O resultado da defesa de tese intitulada: **FENOTIPAGEM DE ALTA EFICIÊNCIA PELA ANÁLISE COMPUTACIONAL DE IMAGENS NO MELHORAMENTO GENÉTICO DO TOMATEIRO**", sendo a aluna considerada aprovada. E, para constar, eu, Professor Alcinei Místico Azevedo, Presidente da Banca, lavrei a presente ata que depois de lida e aprovada, será assinada por mim e pelos demais membros da Banca examinadora.

OBS.: A aluna somente receberá o título após cumprir as exigências do **ARTIGO 65** do regulamento do Curso de Doutorado em Produção Vegetal, conforme apresentado a seguir:

Art. 65 Para dar andamento ao processo de efetivação do grau obtido, o candidato deverá, após a aprovação de sua Dissertação ou Tese e da realização das modificações propostas pela banca examinadora, se houver, encaminhar à secretaria do Colegiado do Programa, com a anuência do orientador, 1 (um) exemplar impresso e 1 (um) exemplar eletrônico da dissertação ou tese, no prazo de 60 (sessenta) dias.

Montes Claros, 15 de março de 2024.

Documento assinado digitalmente
gov.br SILVIA NIETSCHÉ
Data: 18/03/2024 09:50:55-0300
Verifique em <https://validar.iti.gov.br>

Sílvia Nietsche
Membro

Documento assinado digitalmente
gov.br DERLY JOSE HENRIQUES DA SILVA
Data: 18/03/2024 17:00:45-0300
Verifique em <https://validar.iti.gov.br>

Derly José Henrique da Silva
Membro

Documento assinado digitalmente
 TIAGO OLIVOTO
Data: 18/03/2024 08:30:07-0300
CPF: ***.689.590-**
Verifique as assinaturas em <https://v.ufsc.br>

Tiago Olivoto
Membro


Ana Clara Gonçalves Fernandes
Membro

Documento assinado digitalmente
gov.br ALCINEI MISTICO AZEVEDO
Data: 18/03/2024 09:57:16-0300
Verifique em <https://validar.iti.gov.br>

Alcinei Místico Azevedo
Orientador

Dedico

À minha mãe, Maria Bernadete, pelo amor e apoio incondicional, pela dedicação e por ter sido minha maior incentivadora em todos os meus projetos de vida. Também ao meu marido Adriano, pelo companheirismo, amor e por sempre acreditar em mim.

AGRADECIMENTOS

À minha família que sempre me incentivou e apoiou para que eu conseguisse alcançar meus ideais e foi sempre meu porto seguro. Em especial, a meu sobrinho Luís Felipe e à minha mãe Maria Bernadete.

Ao meu marido Adriano e a meus filhos: Joao Pedro, Maria Paula, Alice e Beatriz, por estarem ao meu lado nessa caminhada, enchendo-me de amor e motivação e por sempre entenderem minha ausência. Um agradecimento especial às minhas filhas Alice e Bia por terem sido meu braço direito e maior suporte, durante a condução do experimento, quando, em época de pandemia, estávamos todos isolados.

Ao Professor Alcinei pela orientação e pela disponibilidade, sempre muito solícito quando eu precisava. Pela dedicação a este trabalho, pela confiança, paciência e ensinamentos compartilhados que me possibilitaram chegar ao final desta etapa.

Aos 'irmãos' de orientador, Janete, Ana Luíza, Jaílson, Luan e Ana Clara, pelo companheirismo, pelo apoio na pesquisa e pela amizade que tornou esta etapa mais fácil de ser cumprida.

Aos colegas do Grupo de Estudo em Experimentação Agrícola (GEEA), em especial, Nayany, Deltimara, Valentina, Elias, Varlen e Phelipe, pela amizade, companheirismo, auxílio na condução das pesquisas, sem os quais tudo teria sido muito mais difícil.

Às amigas da pós, Caroline, Karla, Geís, Verônica e Érika, pelo carinho e amizade com que me ajudaram, apoiaram e incentivaram nesses anos.

Aos professores do Programa de Pós-Graduação pela grande contribuição e ensinamentos compartilhados nas disciplinas cursadas e na condução do experimento. Sou-lhes muito grata.

À banca pela disponibilidade e contribuições.

À Universidade Federal de Minas Gerais, em especial, ao Instituto de Ciências Agrárias pelo Programa de Pós-Graduação em Produção Vegetal, pela oportunidade de realização do meu doutorado.

À Coordenação de Aperfeiçoamento de Pessoal de Nível Superior- Brasil (CAPES), ao Conselho Nacional de Desenvolvimento Científico (CNPQ) e à Fundação de Amparo à Pesquisa de Minas Gerais (FAPEMIG) pelo apoio a este estudo.

Meu muito obrigada!

“O melhoramento de plantas exige que os procedimentos sejam cada vez mais rápidos, com menores custos, e cada vez mais eficientes. Atualmente, o melhoramento é arte, ciência e negócio”.

(Borém; Miranda; Fritsche-Neto)

FENOTIPAGEM DE ALTA EFICIÊNCIA PELA ANÁLISE COMPUTACIONAL DE IMAGENS NO MELHORAMENTO GENÉTICO DO TOMATEIRO

RESUMO

O tomate (*Solanum lycopersicum* L.) destaca-se por sua importância no cenário agrícola e econômico, pois é a segunda hortaliça mais produzida mundialmente. Apesar do avanço biotecnológico, no melhoramento genético do tomateiro, a integração de inovações, como visão computacional e análise de imagens, mostra-se promissora para acelerar o desenvolvimento de novas cultivares. A fenotipagem tornou-se mais precisa e eficiente, ao longo da última década, graças à essa associação, possibilitando avaliações em larga escala, maior quantidade de descritores morfoagronômicos e redução de tempo, recursos humanos e financeiros. O uso de análises computacionais pode ainda facilitar a obtenção dos parâmetros genéticos, bem como facilitar as avaliações. Logo objetivou-se estabelecer metodologias, para a avaliação quantitativa e qualitativa de caracteres morfoagronômicos no tomateiro, por meio da análise computacional de imagens. Foram realizados dois experimentos: o primeiro envolveu linhagens comerciais que foram cruzadas em dialelos balanceados e conduzidas até a maturação dos frutos. Nessa etapa, foram adquiridas imagens, em diferentes fases do ciclo da cultura e realizada a coleta das sementes; o segundo experimento, F1 do primeiro, foi constituído por 10 híbridos e cinco progenitores, avaliados da forma tradicional e por meio da análise de imagens. Para a aquisição de imagens, foram obtidos vídeos no campo antes da colheita. Além disso, imagens dos frutos foram adquiridas em miniestúdios por câmeras digital e pré-processadas no software R. Foram avaliadas, em nível de genótipo, características de produção e qualidade dos frutos como formato, grupo, cor e defeitos. Foi observada correlação significativa entre os caracteres avaliados de forma tradicional e computacionalmente, mostrando que o emprego da análise de imagens, aliada à visão computacional e aprendizado profundo, é uma ferramenta eficaz na fenotipagem do tomateiro. Ademais, houve consistência na estimativa de parâmetros genéticos em comparação com a fenotipagem tradicional. Essa abordagem eficiente tem grande potencial aos programas de melhoramento genético do tomateiro, pois simplifica a tomada de decisões e automatiza a fenotipagem, reduzindo gastos de tempo, mão de obra e recursos financeiros.

Palavras-Chave: Inteligência Computacional. Melhoramento genético de hortaliças. Análise de imagens. Fenotipagem de alto rendimento. *Solanum lycopersicum* L.

HIGH EFFICIENCY PHENOTYPING BY COMPUTATIONAL IMAGE ANALYSIS IN GENETIC IMPROVEMENT OF TOMATO PLANT

ABSTRACT

Tomato (*Solanum lycopersicum L.*) stands out for its importance in the agricultural and economic scenario, as it is the second most produced vegetable worldwide. Despite biotechnological advances in tomato plant genetic improvement, the integration of innovations, such as computer vision and image analysis, shows promise in accelerating the development of new cultivars. Phenotyping has become more accurate and efficient over the last decade, thanks to this association, enabling large-scale evaluations, a greater number of morphoagronomic descriptors and a reduction in time, human and financial resources. The use of computational analyzes can still facilitate the obtaining of genetic parameters, as well as facilitate evaluations. Therefore, the objective was to establish methodologies for the quantitative and qualitative evaluation of morphoagronomic characters in tomato plants, through computational image analysis. They were carried out two experiments: the first involved commercial lines that were crossed in balanced diallels and conducted until the fruits matured. At this stage, images were acquired at different phases of the crop cycle and seeds were collected; the second experiment, F1 of the first, was constituted of 10 hybrids and five parents, evaluated in the traditional way and through image analysis. For image acquisition, they were obtained videos in the field before harvesting. Furthermore, images of the fruits were acquired in mini-studios using digital cameras and pre-processed in the R software. They were evaluated on the genotype level, characteristics of fruit production and quality such as shape, group, color and defects. It was observed a significant correlation between the characters evaluated in the traditional and computational ways, showing that the use of image analysis, combined with computer vision and deep learning, is an effective tool in tomato plant phenotyping. Furthermore, there was consistency in the estimation of genetic parameters compared to traditional phenotyping. This efficient approach has great potential for tomato plant genetic improvement programs, as it simplifies decision-making and automates phenotyping, reducing time, labor and financial resources.

Keywords: Computational Intelligence. Genetic improvement of vegetables. Image analysis. High-throughput phenotyping. *Solanum lycopersicum L.*

LISTA DE ILUSTRAÇÕES

FIGURAS

REVISÃO BIBLIOGRÁFICA

Figura 1- Diagrama representativo das etapas de um sistema de processamento de imagens.....	28
Figura 2-Tipos básicos de tarefas realizadas na área de visão computacional. A) Classificação da imagem como um todo dentro de um conjunto de classes. B) Detecção de objetos entre diferentes classes. Cada objeto é individualmente localizado de maneira aproximada por uma região retangular da imagem. C) Segmentação de todos os objetos de uma classe, feito pixel a pixel para a imagem inteira. D) Segmentação dos objetos de maneira individual, de forma que a segmentação separada as diversas ocorrências de objetos.....	30
Figura 3-Esquema de: A- modelo de um neurônio artificial- perceptrons; B- Modelo de uma rede Neural Artificial perceptrons.....	32
Figura 4-Ilustração da convolução entre um filtro 3x3 e o volume de entrada.....	34
Figura 5-Modelo de uma CNN e suas fases de processamento ilustrando o processo de classificação de cultivares de guaraná.....	34

CAPÍTULO I

Figura 1- Condução das plantas em casa de vegetação.....	50
Figura 2 - A) Pólen retirado das flores do genitor masculino; (B) Flor emasculada; (C) Sucesso no pegamento e crescimento dos frutos 15 dias após a polinização.....	51
Figura 3 - Área do experimento com a cultura do tomateiro.....	52
Figura 4 - Obtenção de vídeos da linha de plantio antes da colheita (A) e Orthomosaio de uma parcela criado a partir de imagens decompostas do vídeo, utilizando o programa Agisoft- Metashape (B).....	53
Figura 5 - Exemplo de imagens do banco sintético com imagens com a extensão png sobre o background (A) e sua respectiva máscara (B) obtida pela biblioteca cocosynth.....	54
Figura 6 - Estimativa da função de perda (loss) no decorrer de 100 épocas de treinamento de redes	58
Figura 7 - Imagens demonstrando a classificação realizada pelo algoritmo Mask-RCNN na detecção dos frutos imaturos e maduros.....	58
Figura 8 - Análise de regressão múltipla para a estimação da produtividade do tomateiro a partir da produtividade real.....	59
Figura 9 - Correlação fenotípica entre a produtividade de frutos de tomateiro e seis variáveis estimadas por meio de redes neurais convolucionais (Mask-RCNN). * significativo pelo teste t ao nível de 5%. .61	
Figura 10 - Correlação dos ranks de parâmetros genéticos capacidade geral de combinação – (CGC), capacidade específica de combinação (CEC) e heterose (HET) entre a produtividade de frutos de tomateiro e seis variáveis estimadas por redes neurais convolucionais.....	62

CAPÍTULO II

Figura 1 - Classificação do fruto do tomateiro quanto ao: A- formato, B- Grupo, C- cor e D- Defeitos leves e E- Defeitos graves.	78
Figura 2 - Mini estúdio para aquisição das imagens com iluminação artificial com lâmpada fluorescente (A), posicionamento dos frutos dentro do estúdio (B), Imagem adquirida no estúdio (C) e imagem dos frutos individualizados pelo software R (D).	79
Figura 3 - Exemplificação do agrupamento de imagens em suas respectivas classificações por análise visual, utilizadas para criar as pastas de treinamento das CNNs. A - Grupo, B - Cor, C - Defeito e D - Formato.	80

CAPÍTULO III

Figura 1 - Mini estúdio para aquisição das imagens com iluminação artificial com lâmpada fluorescente (A), posicionamento dos frutos dentro do estúdio (B) e Imagem adquirida no estúdio (C). Para a automação da classificação por inteligência artificial.....	99
Figura 2 - Fluxograma que mostra as etapas, desde a preparação das imagens até a determinação da área de cada fruto em termos percentuais de pixels realizado pelo software R, pacote ExplImage.....	100
Figura 3 - Análise de regressão para a estimação do peso do fruto a partir da área em pixels ocupada pelo fruto na imagem (A) na amostra de treinamento e avaliação da qualidade da predição para a amostra de validação (B).	101
Figura 4 - Erro quadrático médio obtido considerando o peso de frutos observado e estimado na amostra de validação para cada um dos tratamentos e estimativa do teste Hartley.....	102
Figura 5 - Correlação de Spearman e ordenamento de cinco genitores para a capacidade geral de combinação para o peso observado e estimado de frutos de tomateiro	103
Figura 6 - Correlação de Spearman e ordenamento das 15 combinações híbridas para a capacidade específica de combinação (A) e heterose (B) para o peso observado e estimado de frutos de tomateiro.	114

TABELAS

CAPÍTULO I

Tabela 1- Dialelo balanceado envolvendo os 5 progenitores, suas respectivas autofecundações e 10 híbridos.....	49
Tabela 2- Avaliadores da qualidade da classificação de frutos maduros e imaturos por redes neurais convolucionais (Mask-RCNN).	59
Tabela 3 - Estimativas dos quadrados médios, coeficiente de variação residual (CV), coeficiente de variação genético (CVg) e herdabilidade no estudo de um experimento em dialelo com a cultura do tomateiro.....	60

CAPÍTULO II

Tabela 1- Número de imagens originais de frutos do tomateiro utilizadas para a fenotipagem do Formato, Grupo, Cor e Defeitos e número de imagens destinadas ao ajuste das RCNNs (Treinamento) e avaliação da qualidade do ajuste (Teste).	81
Tabela 2 - Número de épocas, tempo de treinamento e classificações realizadas por diferentes arquiteturas de RCNNs para frutos do tomateiro quanto ao formato, grupos, cor e defeitos.	83
Tabela 3 - Classificação das arquiteturas quanto aos acertos e erros na identificação de frutos.	84
Tabela 4 - Avaliadores da qualidade do ajuste de RCNNs com diferentes arquiteturas, na classificação de frutos do tomateiro, quanto ao formato, grupos, cor e defeitos.	86

CAPÍTULO III

Tabela 1- Dialelo balanceado envolvendo os 5 progenitores, suas respectivas autofecundações e 10 híbridos.....	97
Tabela 2 - Parâmetros estimados na avaliação de um experimento em dialelo considerando o peso de frutos observado e estimado a partir da análise computacional de imagens.	103

SUMÁRIO

1 INTRODUÇÃO GERAL	15
2 OBJETIVOS	17
2.1 OBJETIVO GERAL	17
2.2 OBJETIVOS ESPECÍFICOS	17
3 REVISÃO DE LITERATURA	17
3.1-A CULTURA DO TOMATEIRO	17
3.2 MELHORAMENTO GENÉTICO DO TOMATEIRO	19
3.3 USO DE ANÁLISE DIALÉLICA NO MELHORAMENTO GENÉTICO.....	21
3.4 FENOTIPAGEM DE ALTA EFICIÊNCIA.....	23
3.5 ANÁLISE COMPUTACIONAL DE IMAGENS	26
3.5.1- <i>Imagens</i>	26
3.5.2- <i>Visão Computacional e a Imagem</i>	29
3.5.3 <i>Redes Neurais Artificiais</i>	30
3.5.4 <i>Redes Neurais Convolucionais</i>	33
3.6 -REFERÊNCIAS.....	37
4 ARTIGOS.....	45
4.1 CAPÍTULO I: FENOTIPAGEM DE ALTO RENDIMENTO NA AVALIAÇÃO DA PRODUTIVIDADE DE UM DIALELO COM TOMATEIRO	45
RESUMO	45
ABSTRACT	46
4.1.1 <i>INTRODUÇÃO</i>	47
4.1.2 <i>MATERIAL E MÉTODOS</i>	48
4.1.3 <i>RESULTADOS</i>	57
4.1.4 <i>DISCUSSÃO</i>	62
4.1.5 <i>CONCLUSÃO</i>	68
4.1.6 <i>REFERENCIAS BIBLIOGRÁFICAS</i>	69
4.2 CAPÍTULO II: FENOTIPAGEM AVANÇADA NA CLASSIFICAÇÃO DE FRUTOS DO TOMATEIRO POR INTELIGÊNCIA ARTIFICIAL	73
RESUMO	73
ABSTRACT	74
4.2.1 <i>INTRODUÇÃO</i>	75
4.2.2 <i>MATERIAL E MÉTODOS</i>	76
4.2.3 <i>RESULTADOS</i>	82
4.2.4 <i>DISCUSSÃO</i>	87
4.2.5 <i>CONCLUSÃO</i>	89
4.2.6 <i>REFERÊNCIAS</i>	90
4.3 CAPÍTULO III: ANÁLISE DE IMAGENS NA ESTIMAÇÃO DE MASSA FRESCA DE TOMATES ...	93
RESUMO	93
ABSTRACT	94
4.3.1 <i>INTRODUÇÃO</i>	95
4.3.2 <i>MATERIAL E MÉTODOS</i>	96
4.3.3 <i>RESULTADOS</i>	101
4.3.4 <i>DISCUSSÃO</i>	105
4.3.5 <i>CONCLUSÃO</i>	109
4.3.6 <i>REFERÊNCIAS</i>	110
5 CONSIDERAÇÕES FINAIS	113

1 INTRODUÇÃO GERAL

O tomateiro (*Solanum lycopersicum* L.), hortaliça- fruto, pertence à família das Solanaceae, espécie diploide ($2n=2x=24$) e autógama. Nativa da América do Sul, o tomateiro destaca-se pela sua importância no cenário agrícola e econômico (NICK; SILVA, 2016). As Solanáceas têm uma relevância significativa no cenário agrícola mundial, com mais de 3000 variedades cultivadas, e o tomateiro é a segunda hortaliça mais produzida (FAOSTAT, 2023). É reconhecido como um alimento funcional por seus elevados teores de vitaminas A e C, além de ser rico em licopeno (SIERRA-OROZCO *et al.*, 2021). Da domesticação até os dias atuais, o tomate foi continuamente sujeito à seleção, para a melhoria do formato, cor, tamanho, sabor das frutas, *tolerância* das plantas a estresses abióticos e bióticos (NICK e SILVA, 2016; RAZIFARD *et al.*, 2020). Sua diversidade genética e relevância comercial o tornam um modelo a estudos de desenvolvimento de frutos, com aplicações em outras espécies hortícolas. O elevado consumo impulsionou a necessidade de desenvolver variedades novas, de alto rendimento e melhor qualidade nutricional (ANWAR *et al.*, 2019, SIERRA-OROZCO *et al.*, 2021).

Atualmente, a maioria dos programas de melhoramento genético para o tomateiro concentram-se em aprimorar a qualidade dos frutos, aumentar a produção e desenvolver resistência a doenças (RAZIFARD *et al.*, 2020). No melhoramento genético vegetal, a seleção criteriosa dos genitores é crucial para o sucesso do programa. O cruzamento desses genitores visa gerar híbridos e, posteriormente, populações segregantes promissoras, facilitando o trabalho do melhorista no alcance de progresso genético (BOREM *et al.*, 2017).

Diversas metodologias eficazes podem ser empregadas na seleção de cultivares e linhagens para programas de hibridação. Entre as metodologias disponíveis, os cruzamentos dialélicos são amplamente empregados, com os objetivos de adquirir informações sobre os parâmetros genéticos das linhagens parentais e para compreender os efeitos genéticos relacionados à determinação dos caracteres (ONOFRI *et al.*, 2021, CRUZ *et al.*, 2012).

A análise dialélica é amplamente usada para identificar combinações híbridas heteróticas, visando a híbridos com características comerciais, qualidade nutricional superior, resistência a doenças e melhor conservação pós-colheita (MATOS *et al.*,

2021). No entanto, ao avaliar esses experimentos e outros do melhoramento vegetal, enfrentam-se desafios relacionados à fenotipagem, exigindo avanços em escala, precisão, reprodutibilidade e exploração de novas características, superados em parte pelo uso crescente da análise de imagens na última década (CHANDRA *et al.*, 2020; WALTER *et al.*, 2015).

A análise computacional de imagens para fenotipagem oferece benefícios, sendo rápida, precisa, econômica e não invasiva, com boa resolução espacial e temporal. Essa abordagem permite explorar grandes áreas, realizar muitas repetições e garantir a segurança do operador, facilitando estudos comparativos e seleção fenotípica em larga escala (GEBREMEDHIN *et al.*, 2019).

A integração de inovações, como o uso de imagens em conjunto com a inteligência computacional, destaca-se como uma vertente promissora no processo de fenotipagem de plantas (SAMBASIVAM e OPIYO, 2021; CHANDRA *et al.*, 2020). A análise computacional de imagens envolve o uso de algoritmos e técnicas computacionais para extrair informações e realizar análises a partir de imagens. O uso de tecnologias avançadas, como o Aprendizado de máquina (Machine learning) e o Aprendizado profundo (Deep learning), para processar imagens, está se tornando cada vez mais confiável, ao mesmo tempo que permanece acessível em termos de custo. Essas técnicas oferecem alta precisão e velocidade para as avaliações (LU *et al.*, 2021).

A fenotipagem por imagem torna-se importantes ao melhoramento do tomateiro, pois permite a geração de fenótipos, por meio da associação de imagens com programas de computador. Isso viabiliza a construção de bancos de dados com as imagens, os quais podem ser fenotipados e consultados, ao longo do tempo, conforme a necessidade do pesquisador. Diante da urgência de aprimorar a obtenção e análise de dados fenotípicos no tomateiro, é importante explorar o potencial das imagens tanto para as características de produção quanto para as qualitativas, com o objetivo de simplificar a fenotipagem da cultura. Assim, é possível supor que a fenotipagem por imagem tem o potencial de acelerar o processo de melhoramento do tomateiro, simplificando e agilizando a obtenção de dados fenotípicos importantes para o avanço da pesquisa e desenvolvimento nesta cultura.

Nesse contexto, o objetivo deste estudo foi estabelecer metodologias para a fenotipagem de características morfoagronômicas dos frutos do tomateiro, tanto para

caracteres qualitativos quanto quantitativos, por meio da análise computacional de imagens e testar a eficiência dessas metodologias em um cruzamento em dialelo.

2 OBJETIVOS

2.1 Objetivo Geral

Estabelecer metodologias, para a avaliação quantitativa e qualitativa de caracteres morfoagronômicos no tomateiro, por meio da visão computacional com análise de imagens e testar a eficiência dessas metodologias em um dialelo.

2.2 Objetivos Específicos

- Estabelecer uma metodologia para a obtenção de imagens do tomateiro com o auxílio de câmeras fotográficas.
- Estabelecer uma metodologia, para a automação da análise computacional de imagens, a fim de mensurar quantitativa e qualitativamente caracteres morfoagronômicos de interesse no melhoramento genético.
- Comparar os parâmetros genéticos estimados, por meio de dados mensurados de forma tradicional e pela análise computacional de imagens, no estudo de um dialelo balanceado sem recíprocos.

3 REVISÃO DE LITERATURA

3.1-A cultura do tomateiro

A planta do tomateiro (*Solanum lycopersicum* L.), pertencente à família Solanaceae, é uma hortaliça- fruto originária das regiões andinas da América do Sul, incluindo partes do Equador, Peru e Chile. Seu cultivo como cultura começou no século XVIII e, em pouco tempo, foi extensivamente cultivada globalmente e adaptada em diversas variedades. É provável que a domesticação do tomateiro ocorreu no México, onde foi cultivado pelos Astecas (ANDRADE *et al.*, 2009). Alguns botânicos

também atribuem a origem do cultivo e consumo do tomate como alimento à civilização Inca no antigo Peru, em virtude da persistência de uma grande variedade de tomates selvagens e espécies domesticadas exclusivas daquela região. A diversidade ambiental nessas regiões contribuiu para uma ampla diversidade genética entre as espécies relacionadas ao tomateiro. No entanto, apesar da grande diversidade morfológica na espécie cultivada, sua base genética é estreita, em razão do processo de domesticação que ocorreu na América Central, fora de seu centro de origem (BAI; LINDHOUT, 2007; CHETELAT *et al.*, 2009).

Em 2022, a produção global de tomates atingiu 186,1 milhões de toneladas em uma área de 4,91 milhões de hectares. É cultivado em mais de 100 países e a China lidera como o maior produtor, seguida pelos Estados Unidos, Índia, Turquia e Egito. O Brasil ocupa o décimo lugar, com uma produção de 3,9 milhões de toneladas, em 54,5 mil hectares e uma produtividade média de 70 toneladas por hectare, conforme dados da FAO (2023).

O tomateiro é uma planta perene que pode atingir até dois metros de altura, tem um caule cilíndrico e tenro, sendo necessário, com frequência, o uso de estacas (tutores) para suporte. Essa hortaliça exibe crescimento simpodial, característico de plantas cuja estrutura se desenvolve a partir de várias gemas. As plantas podem ter hábito de crescimento indeterminado, semideterminado ou determinado. Cultivares com hábito indeterminado são preferenciais, para a produção de frutos consumidos *in natura*, apresentando crescimento contínuo e vigoroso, com predominância da gema apical. Por outro lado, cultivares com crescimento determinado são mais comuns na produção industrial de tomate, caracterizadas por crescimento vegetativo menos robusto, com a haste principal terminando em inflorescências (FILGUEIRA, 2008; NICK; SILVA, 2018).

As partes verdes do tomateiro são cobertas por pelos que emanam um cheiro forte e característico. Esse odor contribuiu para a crença, que persistiu por muitos anos na Europa, de que o tomate poderia ser uma planta venenosa, desencorajando seu consumo humano (PADOVANI, 1989).

Planta autógama, conferida por suas flores hermafroditas, com $2n = 2x = 24$ cromossomos. Apresenta inflorescência em forma de racimo. Suas flores regulares e hipóginas possuem cinco ou mais pétalas dispostas helicoidalmente, com igual número de estames e ovário bi ou plurilocular (ALVARENGA, 2013). A disposição

desses elementos na flor favorece a autopolinização e dificulta a polinização cruzada (PADOVANI, 1989).

Os frutos do tomateiro, do tipo baga, são carnosos e atingem a maturação entre 30 e 40 dias, após a fecundação do óvulo. A colheita no campo varia de 1(um) a três meses (FILGUEIRA, 2008). As sementes, pequenas e reniformes têm coloração marrom- clara e podem apresentar minúsculos pelos. Cada fruto é composto por um número variável de 50 a 200 sementes (ALVARENGA, 2013). Geralmente consumidos crus em saladas, como vegetais cozidos e em diversos pratos, os frutos de tomate têm uma ampla aplicação na alimentação. Além do consumo direto, uma parte significativa da colheita global é destinada ao processamento, resultando em produtos como tomate enlatado, suco de tomate, ketchup, purê, pasta, tomate seco ou polpa desidratada (LIU *et al.*, 2015). Esses frutos são fontes essenciais de vitaminas e minerais para uma dieta humana equilibrada (GURBUZ ÇOLAK *et al.*, 2020).

O cultivo do tomateiro exige um alto nível tecnológico e intensa utilização de mão de obra, o que dá a essa cultura elevada importância econômica e social (DOSSA *et al.* 2017). O grande consumo da cultura criou a necessidade de produzir variedades novas, de alto rendimento e com melhor qualidade nutricional (ANWAR *et al.*, 2019). Diante das crescentes exigências do mercado consumidor por frutos visualmente atraentes, com maior qualidade nutricional e menos agrotóxicos, torna-se importante aprimorar todos os aspectos do sistema de cultivo (BERTIN; GÉNARD, 2018).

3.2 Melhoramento Genético do Tomateiro

O início do melhoramento genético do tomate no Brasil ocorreu simultaneamente ao desenvolvimento do melhoramento de hortaliças em 1938. O cultivar Santa Cruz de tomate de mesa foi pioneiro no país, originado de um cruzamento natural entre os cultivares Rei Humberto e Redondo Japonês. Os agricultores, ao observarem plantas com características distintas em suas plantações, tiveram o cuidado de selecioná-las e propagá-las. Rapidamente, esse novo cultivar se destacou como o tomate mais amplamente aceito no mercado brasileiro (MELO *et al.*, 2016).

Desde a domesticação ao melhoramento genético, o tomateiro enfrenta significativos desafios, especialmente no aprimoramento de características de

mercado. Isso inclui o aumento do rendimento, baseado no tamanho e número de frutos por planta e na melhoria da qualidade dos frutos diante de estresses abióticos e bióticos. A doçura do tomate, crucial para a sua qualidade (SHARMAI *et al.*, 2018), representa um desafio, pois seu aumento, por vezes, compromete o rendimento. O enfoque inclui também o aumento do conteúdo de componentes nutricionais e funcionais, a extensão da vida pós-colheita, considerando sua natureza climatérica, a abordagem de estresses abióticos, como seca, calor, frio e salinidade, que podem impactar seriamente a produtividade (LIU *et al.* 2020). Como o tomateiro é uma hortaliça- fruta, a tolerância ao estresse é vital, não apenas nas fases vegetativas e de mudas, mas também durante o crescimento reprodutivo, além de lidar com danos causados por pragas (NAGANIME *et al.*, 2023).

Foi demonstrado que uma grande variação genética está presente e é explorável, a partir de espécies selvagens de *Solanum* e, ainda, que tomates silvestres abrigam uma grande riqueza de genes de resistência, alguns dos quais foram identificados e usados, em programas de melhoramento, mas a maior parte ainda é inexplorada por uma combinação de complexidade genética, limitações tecnológicas, prioridades de melhoramento, barreiras regulatórias e recursos limitados (NICK e SILVA, 2016).

A biotecnologia tem sido aplicada com sucesso na cultura do tomateiro. Diversos estudos de mapeamento genético, para características agrônômicas importantes, foram conduzidos, nos últimos anos, resultando no desenvolvimento de marcadores moleculares, para a seleção assistida de genes-chave relacionados à resistência e tolerância, transgênicos capazes de acumular antocianinas nos frutos, entre outros (LIU *et al.* 2020). Além disso, tomateiro foi uma das primeiras culturas a ter seu genoma completamente sequenciado (TOMATO GENOME CONSORTIUM, 2012), destacando-se como um exemplo significativo de avanço biotecnológico (NAGANIME *et al.*, 2023).

A expressão fenotípica de qualquer gene no tomateiro pode ser influenciada pela interação com outros genes e pelo ambiente circundante. Embora muitos genes que regem características de herança simples tenham sido empregados no processo de melhoramento, aproveitando eficientemente a biodiversidade disponível, existem numerosas características poligênicas ou quantitativas, altamente suscetíveis à influência ambiental e de difícil mensuração. Entre essas características, destacam-

se a resistência a pragas e doenças, produtividade, rendimento industrial e teor de sólidos solúveis (brix) nos frutos (SILVA, 2014).

A grande maioria das lavouras, para consumo in natura e para processamento no Brasil, tem sido cultivada com híbridos F1 que têm sido preferidos pelas suas elevadas produtividades e por um conjunto de qualidades nutricionais e fabris (frutos firmes, secos, de boa coloração e elevada viscosidade). Algumas características se expressam de maneira comercialmente vantajosa somente quando em condição heterozigota. Produzir híbridos de tomate é a maneira mais eficaz de aumentar a produtividade e aprimorar a qualidade e adaptabilidade das culturas (MARÍN-MONTES *et al.*, 2022).

Embora, nos últimos 50 anos, a produção global de tomate tenha apresentado um crescimento contínuo, os 10 principais países, com exceção da China e da Índia, que lideram a produção, observa-se uma estagnação ou mesmo uma redução (FAOST, 2023). Essa tendência ressalta a necessidade de se adotar a fenotipagem de alto rendimento, em vários países produtores, para o desenvolvimento de genótipos mais produtivos. Diante desse panorama, é evidente que a produção de tomate exigirá uma abordagem mais focada à obtenção de maior qualidade, aumento da produtividade em áreas limitadas, incorporação de características e ampliação do prazo de validade. O rápido avanço de técnicas genômicas gerou, como resultado, elevado volume de informações moleculares de alta qualidade, que tem se mostrado de grande utilidade ao melhoramento vegetal (CROSSA *et al.*, 2017).

3.3 Uso de Análise Dialélica no Melhoramento Genético

A eficiência de programas de melhoramento genético depende da escolha eficaz de genitores para cruzamentos, visando à produção de híbridos e populações segregantes promissoras (BOREM *et al.*, 2017). Entre as metodologias disponíveis, os cruzamentos dialélicos são amplamente utilizados, proporcionando avanços significativos na seleção de genitores e no entendimento da ação gênica. No entanto a restrição mais significativa desses cruzamentos ocorre, quando há um grande número de genitores, tornando a análise trabalhosa e, por vezes, inviável (VEIGA *et al.*, 2000).

Os cruzamentos dialélicos envolvem os possíveis cruzamentos dentro de um grupo de genótipos, podendo resultar em um número considerável de combinações híbridas a serem avaliadas. Assim, um conjunto de $p(p-1)/2$ híbridos, resultante do cruzamento entre p progenitores (linhagens, variedades, clones, entre outros), pode incluir, além dos respectivos pais, os híbridos recíprocos e/ou outras gerações relacionadas, tais como F_2 's, retrocruzamentos, etc (RAMALHO *et al.*, 1993).

Existem diversas classificações de cruzamento dialélicos, como balanceados ou desbalanceados, completos, parciais e circulantes. As metodologias de análise, como as propostas por Griffing (1956), Gardner; Eberhart (1966), e Hayman (1954), oferecem ferramentas para estimar e interpretar parâmetros genéticos (CRUZ E VENCOVSKY, 1989; CRUZ *et al.*, 2012). Parâmetros genéticos, como capacidade geral de combinação, capacidade específica de combinação, variância genotípica, coeficiente de variação genético, herdabilidade e índice de variação, são importantes para a escolha de métodos de melhoramento mais adequados à cultura, pois as análises biométricas fornecem essas estimativas. Essas informações são essenciais para fazer inferências sobre a predição de ganhos com a seleção (CRUZ *et al.*, 2012).

A metodologia de Griffing (1956) é amplamente utilizada, parametrizando os efeitos de populações em capacidade geral de combinação (CGC) e capacidade específica de combinação (CEC), nas quais são estimados os efeitos e a soma dos quadrados dos efeitos. A CGC reflete o desempenho médio de um pai em cruzamentos, enquanto a CEC está relacionada aos desvios específicos de cruzamentos. A CEC é crucial, para identificar combinações híbridas superiores ou inferiores ao esperado, considerando o desempenho médio das cultivares envolvidas e ainda a obtenção de informações sobre a heterose (CRUZ *et al.*, 2012).

As análises dialélicas presumem a homogeneidade e independência dos erros relativos às médias, exceto em casos de dialelos desbalanceados. A aplicação dessas análises tem sido referenciada no melhoramento genético do tomateiro, especialmente na avaliação de características quantitativas, como qualidade nutricional, resistência a doenças e conservação pós-colheita como mostra os trabalhos de SALEEM *et al.* (2013), KAUSHIK; DHALIWAL (2018), LUPP *et al.* (2023).

3.4 Fenotipagem de alta eficiência

No contexto do melhoramento vegetal, a fenotipagem, uma área em constante expansão, desempenha um papel crucial na compreensão das características associadas às culturas, caracterizando e quantificando os aspectos físicos e fisiológicos típicos de uma planta para uma avaliação quantitativa e qualitativa de suas propriedades e comportamento em várias condições ambientais (CHANDRA *et al.*, 2020). Esse processo fenotípico, que se mantém como pilar central do melhoramento de plantas, é fundamental, apesar dos avanços genéticos e da aplicação de tecnologias moleculares na pesquisa das culturas.

A expressão da produtividade e de características agronomicamente relevantes desempenha um papel significativo no processo de melhoramento. Essa expressão pode ser influenciada por condições edafoclimáticas, tipo de solo, práticas e tecnologias agrícolas, variedade, recursos hídricos, entre outros, sendo assim essencial para seleções e definição de produtos comerciais à avaliação do fenótipo (GODWIN *et al.*; 2019). Esses diversos fatores ambientais destacam a importância contínua da fenotipagem, no avanço e sucesso do melhoramento de plantas, assim como sua adequação às novas tecnologias para o melhoramento mais preciso e eficaz.

O termo "tecnologias para o melhoramento" abrange não apenas a modificação genética, por meio de técnicas de biologia molecular, mas também uma variedade de outras abordagens provenientes de disciplinas como ciência da computação, biologia vegetal, estatística, automação, robótica e inteligência artificial, todas contribuindo para o avanço no melhoramento de plantas (GODWIN *et al.*, 2019). Esse conceito, longe de ser novo, remonta ao início do século passado. A fenotipagem tradicional de plantas, utilizada na agricultura há mais de cem anos, consiste basicamente, na observação, influenciando o desempenho do fenótipo no genótipo. A endogamia e o vigor híbrido eram explorados como parte da genética teórica que impulsionou a ciência agrícola. À medida que o século avançava, os princípios genéticos teóricos se traduziam em avanços concretos, no melhoramento de plantas, passando pela compreensão molecular de genes e técnicas para manipulá-los. Na década de 1990, marcadores moleculares foram empregados, na seleção de linhagens superiores, e o melhoramento assistido por genômica surgiu para aumentar a precisão e eficiência

(VARSHNEY *et al.* 2005). Hoje, em virtude da convergência de diversas disciplinas, o papel do melhorista de plantas evoluiu de um tomador de decisões, baseado em fenótipos, para um líder em equipes multidisciplinares, evidenciando um notável avanço na entrega da ciência no mundo moderno (GODWIN *et al.*, 2019).

Independentemente da abordagem de fenotipagem escolhida pelos melhoristas, a eficiência no manuseio dos dados é um fator crucial para aumentar as chances de sucesso de um programa de melhoramento (REINOLDS *et al.*, 2020). A fenotipagem de alto rendimento ou fenômica é uma alternativa promissora, por sua alta velocidade, precisão e baixo custo na obtenção de informações fenotípicas (TARDIEU *et al.*, 2017; LI *et al.*, 2021).

A habilidade das Redes Neurais Artificiais (RNAs), em modelar sistemas complexos, confere a essa técnica um vasto potencial de aplicação na fenotipagem de germoplasma. Ela oferece a vantagem de automatizar processos que, se realizados manualmente, demandariam um tempo considerável. Assim, a combinação de imagens com RNAs se mostra extremamente funcional nesse contexto (MOURA *et al.*, 2015).

O uso de técnicas computacionais, como visão computacional e processamento de imagens, eleva o nível científico na obtenção e interpretação de dados fenotípicos das plantas. Além disso, a fenômica permite a obtenção de dados fenotípicos, ao longo do ciclo da cultura, de maneira que pode não ser destrutiva, avaliando todas as plantas do experimento (FRITSCH NETO; BORÉM, 2016). Tecnologias baseadas em imagens medem, cada vez com maior frequência, atributos bioquímicos, fisiológicos e morfológicos, que nunca foram mensurados antes ou que eram avaliados apenas em situações específicas, tais como atividade fotossintética, atividade estomática, composição química foliar e arquitetura de folhas e raízes (CHANDRA *et al.*, 2020).

A pesquisa em fenotipagem de plantas teve um rápido crescimento em razão da acessibilidade de dispositivos de imagem digital econômicos, como câmeras RGB, multiespectrais e hiperespectrais, que simplificaram a coleta de grandes volumes de dados. Essa abundância de dados, combinada com o uso de aprendizado de máquina, impulsionou o desenvolvimento de algoritmos e a criação de ferramentas fenotípicas eficientes, incluindo a contagem de frutos, diagnóstico de doenças e estimativa de rendimento. Dada a diversidade de conhecimentos necessários, frequentemente,

essa área é colaborativa entre melhoristas e cientistas de dados (CHANDRA, *et al.*, 2020).

A associação de imagens com programas de computador, para a avaliação de fenótipos, é alternativa promissora ao futuro mapeamento genético detalhado de diversas espécies, permitindo consultas, ao longo dos anos, conforme a conveniência do pesquisador. Essa abordagem contrasta com as medidas manuais, que, muitas vezes, requerem execução em um intervalo de tempo pré-definido, enquanto as plantas estão vivas (HERRMANN JUNIOR *et al.*, 2019). Mas, apesar dos avanços e da necessidade urgente da fenotipagem de alto rendimento, a fenotipagem tradicional, que é realizada manual e visualmente, ainda é frequente. Isso é atribuído ao fato de que, à medida que o ambiente de crescimento das plantas se torna mais desafiador, em termos de adaptação, os modelos fisiológicos e genéticos precisam incorporar novas características para compreender o sistema como um todo (GODWIN *et al.*; 2019). A fenotipagem de alta eficiência acelera programas de melhoramento genético de plantas e animais e melhora a compreensão das respostas das culturas às condições ambientais. No entanto desafios, como padronização e automação dos métodos, custo, interpretação de grandes volumes de dados, integração multidisciplinar e validação são questões a serem abordadas para sua maior proeminência na pesquisa (CHANDRA *et al.*, 2021)

No melhoramento do tomateiro, algumas tecnologias de fenotipagem de alto rendimento têm sido utilizadas com o uso de imagens. Esses trabalhos abrangem características, como: previsão da massa fresca da parte aérea, o número de frutos de tomate e a massa produtiva na colheita, com base em uma série de variáveis derivadas das imagens RGB capturadas por drones (JOHANSEN *et al.* 2020); investigação na análise de parâmetros morfológicos que caracterizam a emergência de raízes primárias na avaliação do vigor das sementes (OLIVEIRA *et al.*, 2021); classificação de doenças em folhas de tomate utilizando diferentes arquiteturas de Redes Neurais de Convolucionais(CNNs) (GEHLOT *et al.*, 2020); identificação de diversas partes da planta, como nós, frutos e flores, em um conjunto de dados desafiador, adquirido durante um experimento de estresse conduzido em vários genótipos de tomate, utilizando Yolo-V5 que é uma arquitetura de redes neurais convolucionais (CARDELLICCHIO *et al.*, 2023.); classificação de frutos quanto à maturação CNNs (PHAN *et al.* 2023), entre muitos outros.

Essas abordagens permitem estabelecer conexões entre fenótipos complexos de plantas e sua expressão genética, enfrentando desafios atuais e emergentes na agricultura, relacionados à segurança alimentar, considerando a crescente escassez de recursos naturais e as projeções de variabilidade e mudanças climáticas (DAS CHOUDHURY *et al.*, 2019, HERRMANN JUNIOR *et al.*, 2019). Dessa forma, a transição para métodos automatizados destaca a relevância da análise de imagens para fenotipagem.

3.5 Análise computacional de Imagens

3.5.1- Imagens

As análises visuais constituem uma prática essencial na caracterização fenotípica de plantas, remontando aos primórdios da agricultura (BARBOSA *et al.*, 2016). Essa abordagem empírica envolve a observação visual de características morfológicas e agronômicas das plantas, permitindo aos agricultores e pesquisadores avaliar o desenvolvimento, saúde e outras propriedades relevantes.

Com o avanço tecnológico, especialmente a introdução de câmeras digitais e scanners, as análises visuais evoluíram para incorporar métodos mais precisos e eficientes. A captura de imagens tornou-se uma ferramenta importante, na análise de plantas, proporcionando uma visão detalhada que vai além da observação a olho nu. Essa integração de tecnologia permitiu uma abordagem mais aprofundada, na compreensão e avaliação das características das plantas, contribuindo para o avanço da agricultura e da pesquisa agronômica (CHANDRA *et al.*, 2020).

O processo inicial nas avaliações fenotípicas de alta eficiência é a captura de imagens, uma etapa de extrema importância que demanda cuidado durante todo o processamento. Qualquer ruído ou imperfeição na imagem pode comprometer as etapas subsequentes do processo (PERSECHINO *et al.*, 2015).

No mercado, há uma variedade de câmeras digitais disponíveis. As câmeras comuns, também conhecidas como câmeras RGB, captam luz no espectro visível para os humanos (400-700 nm) com dimensões espaciais x e y. É importante destacar que dispositivos, como celulares e webcams, pertencem a esse grupo (GUERRERO *et al.*, 2012). Embora essas câmeras possuam a capacidade de captar comprimentos de

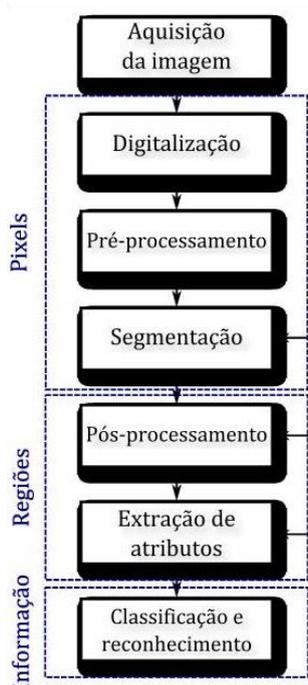
onda fora do espectro visível, elas geralmente vêm equipadas de fábrica com filtros que limitam essa capacidade (SAKAMOTO *et al.*, 2011). Câmeras especiais, por outro lado, possibilitam a captura de imagens geradas em comprimentos de onda fora do espectro visível, desde o ultravioleta (200-400 nm) até o térmico (4000-50000 nm), com a maioria das emissões ocorrendo entre 8000-14000 nm (JONES; VAUGHAN, 2010).

A unidade básica das imagens digitais é o pixel, o menor ponto que compõe uma imagem digital colorida. Milhares de pixels formam a imagem completa. O número de pixels depende do número de unidades fotossensíveis no sensor da câmera (GUERRERO *et al.*, 2012). Uma imagem pode ser visualizada como uma matriz de pontos (ou pixels) com m linhas por n colunas. Cada elemento da matriz representa a magnitude total de energia eletromagnética refletida ou emitida pela área observada pelo sensor. Desse modo, é possível aplicar diversas técnicas de processamento digital de imagens às matrizes numéricas (MATHER, 1997).

O processamento de imagens passa por etapas importantes. As fases de pré-processamento e segmentação atuam nos pixels da imagem. A transformação de dados (pixels) em informação ocorre nas etapas de pós-processamento e extração de atributos. Essas etapas, que podem envolver realimentação, são interdependentes com a segmentação (FIGURA 1). A etapa mais crítica desse processamento é a segmentação, pois é o resultado dessa fase que será processado e avaliado por etapas posteriores (PERSECHINO *et al.*, 2015).

O processamento de imagem envolve a aplicação de operações matemáticas nos dados, visando melhorar os aspectos espectrais e espaciais conforme as necessidades de uma determinada aplicação. Existem duas principais categorias de técnicas de processamento: realce e classificação. Essas técnicas são aplicadas, após o pré-processamento, com o propósito de corrigir eventuais erros encontrados nos dados processados (MENESES; ALMEIDA, 2012).

Figura 1- Diagrama representativo das etapas de um sistema de processamento de imagens.



Fonte: PERSECHINO *et al.*, 2015.

Muitos estudos no campo do melhoramento vegetal fundamentam-se em imagens convencionais adquiridas, por meio de câmeras digitais, scanners ou celulares. No entanto, para investigações mais complexas, têm-se utilizado câmeras térmicas ou multiespectrais, capazes de captar informações adicionais que os dispositivos convencionais não conseguem (TATTARIS, 2016),

Tecnologias baseadas em imagens podem ser usadas, para avaliar o que nunca foi mensurado antes ou somente, em situações específicas, estão começando a ser medidas, com maior frequência, como atributos bioquímicos, fisiológicos e morfológicos, tais como atividade fotossintética, atividade estomática, composição química foliar e arquitetura de folhas e raízes (CLARK *et al.*, 2011).

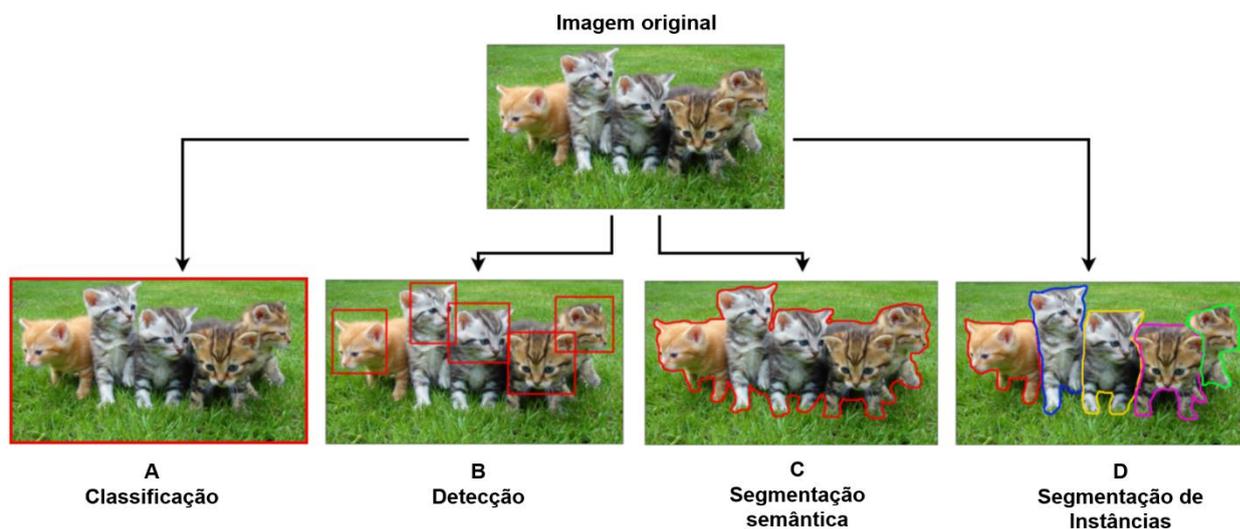
O uso de imagens digitais, para fins acadêmicos, já foi limitado a um grupo restrito de especialistas, principalmente pelo alto custo dos equipamentos para aquisição e processamento dessas imagens (BURGER; BURGER, 2008). Contudo a redução desses custos, a acessibilidade de computadores com capacidade de processamento elevada e a disponibilidade de pacotes de análise possibilitaram o crescimento das áreas de processamento digital de imagens, visão computacional e fenômica no campo do melhoramento vegetal (SABROL, 2015).

3.5.2- Visão Computacional e a Imagem

A visão computacional é um campo da ciência e tecnologia dedicado ao desenvolvimento de teorias e tecnologias para a construção de sistemas artificiais capazes de extrair informações de imagens ou dados multidimensionais (PRINCE *et al* 2012). Em essência, trata-se da capacidade das máquinas de "enxergar" e interpretar dados visuais. Para efetuar decisões sobre objetos reais, é frequentemente necessário criar alguma descrição ou modelo desses objetos a partir das imagens. Dessa forma, muitos pesquisadores definem o objetivo da visão computacional como a construção de descrições de cenas obtidas por meio de imagens (STOCKMAN; SHAPIRO, 2001).

A visão computacional abrange todo o processo, desde a aquisição da imagem, seu pré-processamento e tratamento, até a tentativa de aprender relações entre as imagens, extrair informações e apresentar um resultado. A classificação relaciona a imagem dentro de um conjunto de possíveis classes conhecidas (Figura 2 A). Além da classificação, podemos fazer a detecção e segmentação semântica e de instâncias de objetos dentro delas (YU *et al.*, 2020) (Figura 2- B, C e D). A segmentação semântica classifica individualmente os pixels de uma imagem entre o conjunto de possíveis classes conhecidas; já a segmentação de instâncias gera máscaras ou contornos para cada objeto em uma imagem, acompanhados por rótulos de classe e índices de confiança. Essa abordagem é valiosa, quando se necessita conhecer tanto a localização quanto a forma precisa dos objetos em uma imagem (CARVALHO, 2022)

Figura 2-Tipos básicos de tarefas realizadas na área de visão computacional. A) Classificação da imagem como um todo dentro de um conjunto de classes. B) Detecção de objetos entre diferentes classes. Cada objeto é individualmente localizado de maneira aproximada por uma região retangular da imagem. C) Segmentação de todos os objetos de uma classe, feito pixel a pixel para a imagem inteira. D) Segmentação dos objetos de maneira individual, de forma que a segmentação separe as diversas ocorrências de objetos



Fonte: https://repositorio.ufsc.br/bitstream/handle/123456789/197793/TCC_Roberto_Philippi.pdf?sequence=1;isAllowed=y

3.5.3 Redes Neurais Artificiais

Avanços no aprendizado profundo (*deep learning*), um subcampo do aprendizado de máquinas (*machine learning*), levou a um aumento no desenvolvimento de métodos de fenotipagem por imagem de plantas. No campo de machine learning, as máquinas têm a capacidade de aprender. Isso implica que a solução, para um determinado problema, resultará do processo de aprendizado, em vez de depender de um código pré-estabelecido detalhado por um programador. Em vez de organizar os dados para serem executados por meio de equações predefinidas, o *deep learning* configura parâmetros básicos sobre os dados e treina o computador para aprender sozinho pelo reconhecimento de padrões em várias camadas de processamento. Portanto são particularmente conhecidas por sua eficácia, no manuseio baseado na visão, tarefas como classificação de imagens, detecção de objetos, segmentação semântica e entendimento da cena (Li *et al.*, 2021). Coincidentemente, muitas dessas tarefas são a base das várias outras tarefas de

fenotipagem de plantas, como detecção de doenças, detecção de frutos e estimativa de rendimento (CHANDRA, 2020).

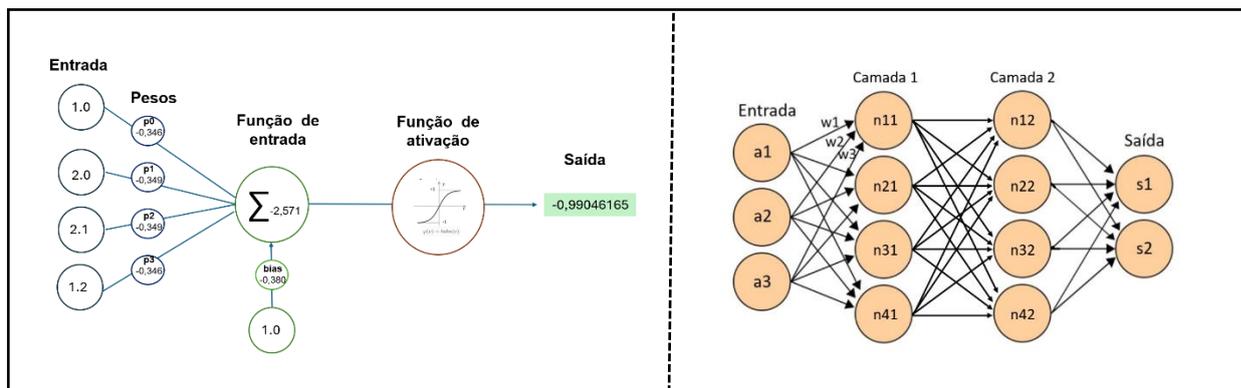
A abordagem Deep learning, como o nome sugere, envolve uma aprendizagem profunda, por meio de múltiplas camadas de processamento, utilizando Redes Neurais Artificiais para a resolução de problemas complexos como o reconhecimento de imagens (OLIVEIRA; CÂMARA, 2019).

As Redes Neurais Artificiais (RNAs) têm sua inspiração nos modelos iniciais de processamento sensorial do cérebro. O cérebro humano é composto por uma rede altamente interconectada de neurônios, que se comunicam, por meio do envio de pulsos elétricos, por meio de sua fiação neural, constituída por axônios, sinapses e dendritos (DE SOUZA *et al.*, 2022). No contexto computacional, é possível criar uma rede neural artificial, simulando um modelo de rede de neurônios em um computador. Ao aplicar algoritmos, que replicam os processos de neurônios reais, podemos fazer com que a rede "aprenda" a resolver diversos tipos de problemas (HAYKIN, 2009).

Um neurônio modelo (FIGURA 3A), denominado unidade limite, desempenha um papel fundamental nesse processo. Essa unidade recebe informações de várias outras unidades ou fontes externas, pondera cada entrada e as soma (KROGH, 2008).

A rede neural é formada por camadas, que representam a organização dos neurônios. A profundidade da rede é determinada pelo agrupamento dessas camadas. A forma como as camadas são organizadas influencia o comportamento da rede, resultando em diferentes arquiteturas (OLIVEIRA; CÂMARA, 2019). As entradas da rede são introduzidas na primeira camada, denominada camada de entrada a qual distribui as informações para a(s) camada(s) escondida(s) da rede. A última camada é a de saída, em que a solução para o problema é obtida. Pode haver uma ou mais camadas intermediárias (ou escondidas) entre a camada de entrada e a de saída. Além disso, os neurônios de uma camada estão conectados apenas aos neurônios da imediatamente posterior, sem realimentação (comunicação unidirecional) ou conexões entre neurônios da mesma camada, ou seja, todos os neurônios da rede estão conectados a todos os neurônios em camadas adjacentes (HAYKIN, 2009), ilustrados na Figura 3B.

Figura 3-Esquema de: A- modelo de um neurônio artificial- perceptron; B- Modelo de uma rede Neural Artificial perceptrons.



Fonte: Adaptado de: <https://www.wisertecnologia.com.br/blog/post/redes-neurais-artificiais> e <https://didatica.tech/introducao-a-redes-neurais-e-deep-learning/>

O pré-processamento da imagem é uma parte importante que pode impactar a acurácia máxima do modelo durante o treinamento. O modelo requer que as imagens sejam decodificadas e redimensionadas para se adequarem. Modificar constantemente as imagens, antes de inseri-las no modelo, é benéfico, pois cada imagem específica se torna um pouco diferente a cada época de treinamento. O pré-processamento de imagens é uma prática que envolve tanto arte quanto ciência, e um estágio bem projetado pode significativamente aprimorar as capacidades de reconhecimento do modelo. Por outro lado, um estágio de pré-processamento muito simples pode impor um limite artificial na precisão alcançada pelo modelo durante o treinamento (OLIVEIRA; CÂMARA, 2019.).

As redes neurais artificiais são uma alternativa promissora em vários setores por sua capacidade de realizar ajustes eficazes. Em contraste com a regressão, essas redes têm a vantagem de incorporar tanto variáveis quantitativas quanto qualitativas categóricas em suas modelagens (HAYKIN, 2001). A capacidade de aprendizado e adaptação das RNAs permitem lidar com dados imprecisos e situações não completamente definidas no início do treinamento. Essas características tornam esses modelos cada vez mais atraentes, para aplicações em técnicas de processamento de imagem, reconhecimento de padrões, problemas de classificação, controle de processos, entre outros (SOUZA *et al.*, 2020). As RNAs representam uma ferramenta tecnológica importante na fenotipagem de alta precisão, destacando-se pela eficácia na modelagem de problemas complexos (AZEVEDO *et al.*, 2015).

Basicamente, o treinamento de uma RNA envolve a resolução de um problema de minimização não linear sem restrições. Nesse processo, os pesos sinápticos da rede são iterativamente ajustados, para minimizar o erro quadrático médio entre a resposta desejada, com base nos dados de entrada e a saída obtida no neurônio de saída. Do ponto de vista estatístico, o treinamento implica a estimativa dos parâmetros do modelo considerando um conjunto de dados específico (DE OLIVEIRA *et al.*, 2010). No contexto da simulação de aprendizado por computador, realizam-se ajustes incrementais, nos pesos e no limite, cada vez que um novo exemplo é apresentado, visando aprimorar a precisão da classificação. Os algoritmos empregados, assim como os parâmetros e coeficientes utilizados, durante o treinamento, representam outro elemento que exerce impacto considerável tanto no tempo de treinamento quanto na eficiência da RNA (BINOTI *et al.*, 2014).

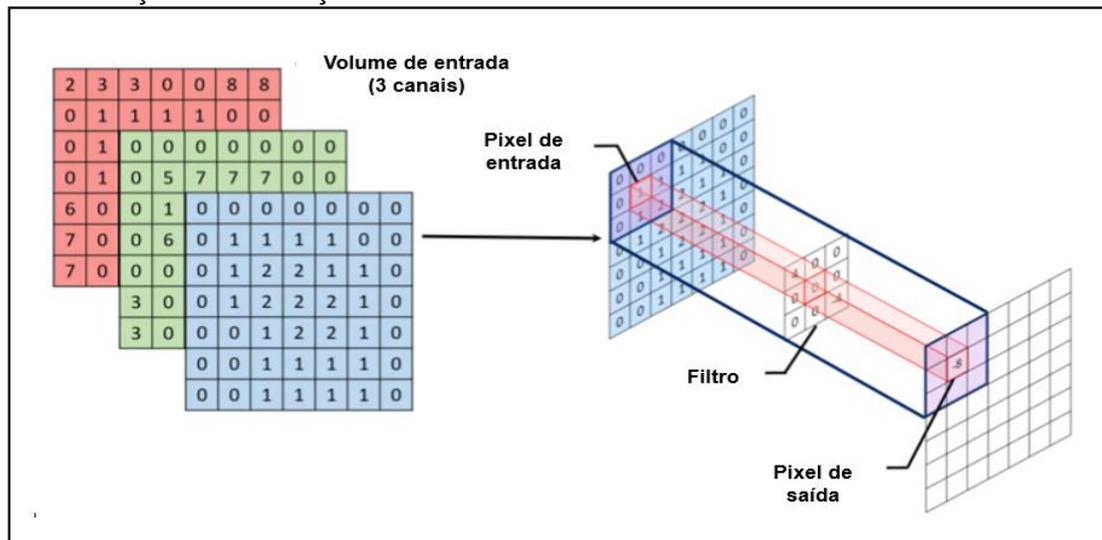
3.5.4 Redes Neurais Convolucionais

Redes Neurais Convolucionais (CNNs) são uma categoria de Redes Neurais multicamadas treináveis para aprender recursos essenciais à classificação. Utilizam a convolução em pelo menos uma de suas camadas. A aplicação principal é no processamento de informações visuais. São fáceis de treinar quando existe grande quantidade de amostra (ARAUJO *et al.*, 2017). Elas se destacam em aplicações de visão computacional, como reconhecimento facial, detecção de objetos, visão em robótica e veículos autônomos (CHANDRA *et al.*, 2020).

A CNN é composta basicamente de três camadas neurais: camadas convolucionais, camadas de *pooling* e camadas totalmente conectadas. Essas camadas são estruturadas em arquiteturas hierárquicas (GUO *et al.*, 2016), reconhecem imagens como objetos tridimensionais, em contraste com estruturas planas que são medidas apenas por largura e altura. Essa característica está vinculada à codificação RGB (Red-Green-Blue) das imagens, que combina essas três cores para criar o espectro de cores percebido pelos seres humanos. Uma CNN interpreta imagens como esses três estratos distintos de cores empilhados verticalmente (FIGURA 4). Assim, recebe uma imagem como uma caixa retangular, em que a largura e altura são determinadas pelo número de pixels, ao longo dessas dimensões, e a profundidade é composta por três camadas, uma para cada

componente RGB. Essas camadas de profundidade são denominadas canais (Li *et al.*, 2021).

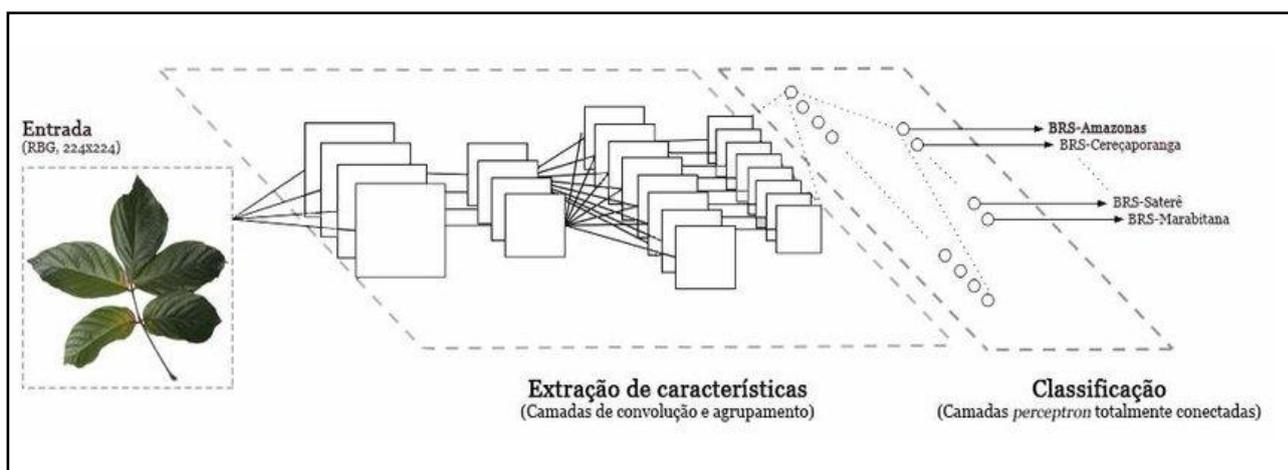
Figura 4-Ilustração da convolução entre um filtro 3x3 e o volume de entrada.



Fonte: ARAÚJO *et al.* (2017).

Diferente das RNAs clássicas, as camadas convolucionais das CNNs (FIGURA 5) têm a função de aplicar diversos filtros aos dados de entrada, formando uma hierarquia de características que resultam em descritores precisos da imagem (KRIZHEVSKY *et al.*, 2012).

Figura 5-Modelo de uma CNN e suas fases de processamento ilustrando o processo de classificação de cultivares de guaraná



Fonte: SOUSA *et al.*, 2017.

As CNNs podem ser divididas em várias arquiteturas, que auxiliam na melhoria do desempenho de redes neurais profundas (SZEGEDY *et al*, 2016). Essas arquiteturas variam de acordo com o tamanho, precisão, quantidade de parâmetros, profundidade e tempo por etapa de inferência. Dessa forma, podem-se destacar as arquiteturas, como: VGG, Inception, Inception-v3, ResNet, InceptionResNetV2, descritas abaixo:

- VGG - O algoritmo de agrupamento de geometria visual (VGG) foi desenvolvido, em seis arquiteturas distintas, variando até 19 camadas de convolução. Em contraste com os algoritmos anteriores de aprendizado profundo, o VGG utiliza filtros de tamanho 2x2 e 3x3. A nomenclatura do VGG reflete o número de camadas incorporadas em sua estrutura. Por exemplo, se forem utilizadas 16 camadas, o algoritmo é referido como VGG16; se a arquitetura incluir 19 camadas, é denominado VGG19. A rede VGG processa imagens RGB de 224 x 224 pixels, submetendo-as a camadas consecutivas de convolução com filtros, seguidas por uma camada de max-pooling (SIMONYAN; ZISSERMAN, 2014). A VGG apresenta algumas desvantagens, como alto consumo de tempo computacional, uso excessivo de memória e uma estrutura que contém um grande número de parâmetros (HULSTAERT, 2018).

- A ResNet, ou Redes Neurais Convolucionais Residuais, é uma arquitetura clássica amplamente utilizada em tarefas de visão computacional. Essa rede permite treinar modelos extremamente profundos, ultrapassando 150 camadas, mantendo uma complexidade inferior a redes menos profundas. A utilização de blocos residuais ajuda a evitar problemas de saturação ou rápida degradação de desempenho à medida que a rede se aprofunda (SOUZA *et al.*, 2020).

- Inception - O principal propósito do módulo Inception é funcionar como um extrator de características em múltiplos níveis, realizando operações de convolução com filtros de tamanhos 1x1, 3x3 e 5x5 dentro do mesmo módulo da rede. Cada bloco Inception é seguido por uma camada de expansão de filtro (convolução 1 x 1 sem ativação) que é usada, para aumentar a dimensionalidade do banco de filtros, antes da adição para corresponder à profundidade da entrada (Souza *et al* 2020).

- O Inception V3 - Apresenta-se como uma variação da Inception. É um modelo composto por elementos básicos, tanto simétricos quanto assimétricos. Esses elementos incluem convoluções, pool médio, pool máximo, concatenações, dropouts e camadas totalmente conectadas. A normalização em lote é extensivamente aplicada

às entradas de ativação em todo o modelo. Desenvolvida para resolver tarefas de classificação de imagens, a arquitetura InceptionV3 foi treinada no conjunto de dados ImageNet. Uma característica notável é a redução dos parâmetros estimados pela rede, resultando em um desempenho computacional superior em comparação com as redes VGG durante o treinamento (SOUZA, 2020; SZEGEDY *et al*, 2016)

- InceptionResNet-V2 - A fusão das arquiteturas Inception e ResNet resultou no Inception-ResNet-v2, uma poderosa rede neural convolucional. Treinada em mais de um milhão de imagens do banco de dados ImageNet, essa rede de 164 camadas demonstra desempenho notável no reconhecimento de imagens, é capaz de classificar imagens em 1.000 categorias distintas, abrangendo desde objetos cotidianos até uma variedade de animais. Com uma entrada de imagem de 299 por 299, a saída da rede é uma lista de probabilidades estimadas para cada classe, evidenciando sua habilidade em aprender representações ricas de características para diversos tipos de imagens (SZEGEDY *et al*, 2017).

As CNNs se destacam por sua facilidade de treinamento, especialmente quando há uma grande quantidade de amostras rotuladas representando diversas classes-alvo. Suas vantagens incluem a habilidade de extrair características relevantes, por meio do aprendizado de transformações (kernels) e a capacidade de depender de um número menor de parâmetros de ajuste, em comparação com redes totalmente conectadas (ARAUJO *et al.*, 2017). Além disso, a capacidade das CNNs de misturar múltiplos mapas de características, simultaneamente, possibilita a extração de características cada vez mais complexas (VARGAS *et al.* 2016). Dessa forma, as CNNs demonstram a capacidade de criar filtros extremamente complexos, aproveitando ao máximo as informações disponíveis nos dados de treinamento.

3.6 -Referências

- ANWAR, R.F.T.; MATTOO, A. K. Tomatoes: A model crop of solanaceous plants. In: Oxford Research Encyclopedia of Environmental Science. **Oxford University Press**. p. 1-50. 2019. DOI <https://doi.org/10.1093/acrefore/9780199389414.013.223>
- ALVARENGA, M. A. R. **Tomate: Produção em campo, casa de vegetação e hidroponia**. 3a ed. Lavras-MG: Editora Un. Lavras: UFLA, 2022, 517P.
- ARAÚJO, F. H.; CARNEIRO, A. C.; SILVA, R. R.; MEDEIROS, F. N.; USHIZIMA, D. M. Redes neurais convolucionais com tensorflow: Teoria e prática. **SOCIEDADE BRASILEIRA DE COMPUTAÇÃO. III Escola Regional de Informática do Piauí. Livro Anais-Artigos e Minicursos**, v. 1, p. 382-406, 2017.
- AZEVEDO, A. M.; ANDRADE JÚNIOR, V. C.; PEDROSA, C. E.; OLIVEIRA, C. M. DE; DORNAS, M. F.; CRUZ, C. D.; VALADARES, N. R. Application of artificial neural networks in indirect selection: a case study on the breeding of lettuce. **Bragantia**, v.74, n.4, p.387-393, 2015. DOI <https://doi.org/10.1590/1678-4499.0088>
- BAI, Y.; LINDHOUT, P. Domestication and breeding of tomatoes: what have we gained and what can we gain in the future?. *Annals of Botany*, v.100, n.5, p.1085–1094, 2007. DOI <https://doi.org/10.1093/aob/mcm150>
- BARBOSA, J. Z.; CONSALTER, R.; PAULETTI, V.; MOTTA, A. C. Uso de imagens digitais obtidas com câmeras para analisar plantas. **Revista de Ciências Agrárias**, v.39, n.1, p.15-24, 2016. DOI <https://doi.org/10.19084/RCA15006>
- BERTIN, N.; GÉNARD, M. Tomato quality as influenced by preharvest factors. **Scientia Horticulturae**, v. 233, p. 264–276. 2018. DOI <https://doi.org/10.1016/j.scienta.2018.01.056>
- BINOTI, D. H. B.; DA SILVA, M. L. M. B.; LEITE, H. G. Configuração de redes neurais artificiais para estimação do volume de árvores. **Revista Ciência da Madeira**, v.5, n. 1, p. 58-67, 2014. DOI [10.12953/2177-6830.v05n01a06](https://doi.org/10.12953/2177-6830.v05n01a06)
- BORÉM, A.; MIRANDA, G.V.; FRITSCHÉ-NETO, R. **Melhoramento de plantas**. 7 ed. Viçosa: UFV, 2017. 543 p.
- BURGER W.; BURGE M.J. **Digital Imaging Processing: An algorithmic introduction using Java**. Verlag New York: Springer. 2008. 564 p. Disponível em: https://imagingbook.files.wordpress.com/2013/06/burgerburgeen20071104_ijreferencie_letter.pdf
- CARDELLICCHIO, A.; SOLIMANI, F.; DIMAURO, G.; PETROZZA, A.; SUMMERER, S.; CELLINI, F.; RENÓ, V. Detection of tomato plant phenotyping traits using YOLOv5-based single stage detectors. **Computers and Electronics in Agriculture**, v. 207, p. 107757, 2023. DOI <https://doi.org/10.1016/j.compag.2023.107757>

CRUZ, C. D.; VENCOVSKY, R. Comparação de alguns métodos de análise dialélica. **Revista Brasileira de Genética**, v. 12, n. 2, p. 425-438, 1989.

CRUZ, C. D.; REGAZZI, A. J.; CARNEIRO, P. C. **Métodos biométricos aplicados ao melhoramento genético**. 4. ed. Viçosa MG: Imprensa Universitária UFV, 2012. 514p.

CHANDRA, A. L.; DESAI, S. V.; GUO, W.; BALASUBRAMANIAN, V. N. Computer vision with deep learning for plant phenotyping in agriculture: A survey. **arXiv preprint arXiv:2006.11391**, v. 1, 2020. DOI <https://doi.org/10.34048/ACC.2020.1.F1>

CHETELAT, R. T.; PERTUZÉ, R. A.; FAUNDEZ, L.; GRAHAM, E. B.; JONES, C. M. Distribution, ecology and reproductive biology of wild tomatoes and related nightshades from the Atacama Desert region of northern Chile. **Euphytica**, v. 167, n. 1, p. 77-93, 2009. DOI [10.1007/s10681-008-9863-6](https://doi.org/10.1007/s10681-008-9863-6)

CLARK R. T.; MAC CURDY, R. B.; JUNG, J. K.; SHAFF, J. E.; MC COUCH, S. R.; ANESHANSLEY, D. J.; KOCHIAN, L. V. Three-dimensional root phenotyping with a novel imaging and software platform. **Plant physiology**, v. 156, n. 2, p. 455-465, 2011. DOI <https://doi.org/10.1104/pp.110.169102>

CROSSA, J.; PÉREZ-RODRIGUEZ, P.; CUEVAS, J.; MONTESINOS-LÓPEZ, O.; JARQUÍN, D.; DE LOS CAMPOS, G.; VARSHNEY, R. K. Genomic selection in plant breeding: methods, models, and perspectives. **Trends in plant science**, v. 22, n. 11, p. 961-975, 2017. DOI <https://doi.org/10.1016/j.tplants.2017.08.011>

DAS CHOUDHURY, S.; SAMAL, A.; AWADA, T. Leveraging image analysis for high-throughput plant phenotyping. **Frontiers in plant science**, v. 10, p. 508, 2019. Doi: <https://doi.org/10.3389/fpls.2019.00508>

DE OLIVEIRA, A. C. S.; DE SOUZA, A. A.; LACERDA, W. S.; GONÇALVES, L. Application of artificial neural networks in the forecast of alcohol production. **Ciência e Agrotecnologia**, v. 34, n. 2, p. 279-284, 2010. DOI <http://dx.doi.org/10.1590/S1413-70542010000200002>

DOSSA, D.; FUCHS, F. Tomate: análise técnico-econômica e os principais indicadores da produção nos mercados mundial, brasileiro e paranaense. **Boletim Técnico**, v. 3, 2017.

FAOSTAT. 2023. Food and Agriculture Organization of the United Nations. In: Acesso em: 10 jan. 2024. <https://www.fao.org/faostat/en/#compare>

FILGUEIRA, F. A.R. **Novo Manual de Olericultura**: agrotecnologia moderna na produção e comercialização de hortaliças. 3 ed. Viçosa MG: Editora UFV. 2008. 421p.

FRITSCH NETO, R; BORÉM, A. **Fenômica: como a fenotipagem de próxima geração está revolucionando o melhoramento de plantas**. Viçosa, MG: Editora UFV. 2016. 216p.

GEHLOT, M.; SAINI, M. L. Analysis of different CNN architectures for tomato leaf disease classification. **2020 5th IEEE international conference on recent advances and innovations in engineering (ICRAIE)**. IEEE, p. 1-6, 2020. DOI [10.1109/ICRAIE51050.2020.9358279](https://doi.org/10.1109/ICRAIE51050.2020.9358279).

GODWIN, I. D.; RUTKOSKI, J.; VARSHNEY, R. K.; HICKEY, L. T. Technological perspectives for plant breeding. **Theoretical and Applied Genetics**, v. 132, n. 3, p. 555-557, 2019. DOI <https://doi.org/10.1007/s00122-019-03321-4>

GURBUZ ÇOLAK, N.; EKEN, N.T.; ÜLGER, M., FRARY, A.; DOGANLAR, S. Mapping of quantitative trait loci for antioxidant molecules in tomato fruit: Carotenoids, vitamins C and E, glutathione and phenolic acids. **Plant Science**, vol. 292, p. 110393, 2020. DOI <https://doi.org/10.1016/j.plantsci.2019.110393>

GUERRERO, N. R.; QUINTERO, M. A. O.; NARANJO, J. C. P. Determinación del área foliar en fotografías tomadas con una cámara web, un teléfono celular o una cámara semiprofesional. **Revista Facultad Nacional de Agronomía-Medellín**, v. 65, n. 1, p. 6399-6405, 2012.

GUO, K.; SUI, L.; QIU, J.; YAO, S.; HAN, S.; WANG, Y.; YANG, H. Angel-eye: A complete design flow for mapping cnn onto customized hardware. In: **IEEE Computer Society Annual Symposium on VLSI (ISVLSI) 2016**. IEEE, p. 24-29, 2016. DOI <https://doi.org/10.1109/ISVLSI.2016.129>

HAYKIN, S. Tradução: ENGEL, P. M. **Redes neurais: princípios e prática**. 2 ed. Porto Alegre: Bookman, 2001. 913 p.

HAYKIN, S. **Neural networks and learning machines, 3/E**. Pearson Education India, 2009. 944 p.

HERRMANN JUNIOR, P. D. P.; CRESTANA, S.; RIBEIRO JUNIOR, W. Q.; DE SOUSA, C. A. F.; SANTOS, T.; LANNA, A. **Cenário atual dos novos métodos de fenotipagem de plantas e urgência nas ações de imersão do Brasil na era da bioeconomia**. Siagro. 2019. p. 289-294. Disponível em: <https://www.embrapa.br/busca-de-publicacoes/-/publicacao/1114694/cenario-atual-dos-novos-metodos-de-fenotipagem-de-plantas-e-urgencia-nas-acoes-de-imersao-do-brasil-na-era-da-bioeconomia>

HULSTAERT, L. **Going deep into image classification**. Toward Data Science. 2018 Disponível em: <https://towardsdatascience.com/an-overview-of-imageclassification-networks-3fb4ff6fa61b>. Acesso em: 30 de abril de 2020

JOHANSEN, K.; MORTON, M. J.; MALBETEAU, Y.; ARAGON, B.; AI-MASHHARAWI, S.; ZILIANI, M. G.; MC CABE, M. F. Predicting biomass and yield in a tomato phenotyping experiment using UAV imagery and random forest. **Frontiers in Artificial Intelligence**, v. 3, p. 28, 2020. DOI <https://doi.org/10.3389/frai.2020.00028>

JONES, H. G.; VAUGHAN, R. A. Remote sensing of vegetation: principles, techniques, and applications. **Journal of Vegetation Science**. p. 1-3, 2010. DOI [10.1111/j.1654-1103.2011.01319](https://doi.org/10.1111/j.1654-1103.2011.01319)

KROGH, A. What are artificial neural networks?. **Nature biotechnology**, v. 26, n. 2, p. 195-197, 2008. DOI <https://doi.org/10.1038/nbt1386>

KRIZHEVSKY, A.; SUTSKEVER, I.; HINTON, G. E. Imagenet classification with deep Convolutional neural networks. **Communications of the ACM**, v. 60, n.6, p. 84-90, 2012. DOI <https://doi.org/10.1145/3065386>

LI, D; QUAN, C; SONG, Z; LI, X; YU, G; LI, C; MUHAMMAD, A. High-throughput plant phenotyping platform (HT3P) as a novel tool for estimating agronomic traits from the lab to the field. **Frontiers in Bioengineering and Biotechnology**, v. 8, p. 1533, 2021. DOI <https://doi.org/10.3389/fbioe.2020.623705>

LI, Z.; LIU, F.; YANG, W.; PENG, S.; ZHOU, J. A survey of convolutional neural networks: analysis, applications, and prospects. **IEEE transactions on neural networks and learning systems**, v. 33, n. 12, p. 6999-7019, 2021. DOI <https://doi.org/10.1109/tnnls.2021.3084827>

LIU, R.; HOW-KIT, A.; STAMMITTI, L.; TEYSSIER, E.; ROLIN, D.; MORTAIN-BERTRAND, A.; HALLE S.; LIU, M.; KONG, J.; WU, C.; DEGRAEVE-GUIBAULT, C.; CHAPMAN N. H.; MAUCOURT, M.; HODGMAN, T. C.; TOST, J.; BOUZAYEN, M.; HONG, Y.; SEYMOUR, G. B.; GIOVANNONI, J. J.; GALLUSCI, P. A. A DEMETER-like DNA demethylase governs tomato fruit ripening. **Proceedings of the National Academy of Sciences**, v. 112, n. 34, p. 10804-10809, 2015. DOI <https://doi.org/10.1073/pnas.1503362112>

LIU, Z.; JIANG, J.; REN, A.; XU, X.; ZHANG, H.; ZHAO, T.; JIANG, X.; SUN, W.; LI, J.; YANG, H. Heterosis and combining ability analysis of fruit yield, early maturity, and quality in tomato. **Agronomy**, v. 11, n. 4, p. 807, 2021. DOI <https://doi.org/10.3390/agronomy11040807>

LU, T.; HAN, B.; CHEN, L.; YU, F.; XUE, C. A generic intelligent tomato classification system for practical applications using DenseNet-201 with transfer learning. **Scientific Reports**, v. 11, n. 1, p. 15824, 2021. DOI <https://doi.org/10.1038/s41598-021-95218-w>

LUPP, R. M.; MARQUES, D. N.; LIMA NOGUEIRA, M.; CARVALHO, M. E. A.; AZEVEDO, R. A.; PIOTTO, F. A. Cadmium tolerance in tomato: determination of organ-specific contribution by diallel analysis using reciprocal grafts. **Environmental Science and Pollution Research**, p. 1-13, 2023. DOI <https://doi.org/10.1007/s11356-023-31230-z>

KAUSHIK, P.; DHALIWAL, M. S. Diallel analysis for morphological and biochemical traits in tomato cultivated under the influence of tomato leaf curl virus. **Agronomy**, v. 8, n. 8, p. 153, 2018. DOI <https://doi.org/10.3390/agronomy8080153>

MARIN-MONTES, I.M.; RODRÍGUEZ-PÉREZ, J.E.; ROBLEDO-PAZ, A.; DE LA CRUZ, E.T; PENA-LOMELÍ, A.; SAHAGÚN-CASTELLANOS, J. Haploid Induction in Tomato (*Solanum lycopersicum* L.) via Gynogenesis. **Plants**, v. 11, p. 1595. 2022. DOI <https://doi.org/10.3390/plants11121595>

MATHER, G. The use of image blur as a depth cue. **Perception**, v. 26, n. 9, p. 1147-1158, 1997. DOI <https://doi.org/10.1068/p261147>

MELO, P. C. T.; MELO, A. M. T.; ARAGÃO, F. A. S. Melhoramento de hortaliças no Brasil. In: NICK, C.; BORÉM, A. **Melhoramento de Hortaliças**. 1 ed. Viçosa: Editora UFV, 2016. p. 9-60.

MENESES, P. R.; ALMEIDA, T. Introdução ao processamento de imagens de sensoriamento remoto. Brasília. **Universidade de Brasília UNB**, 2012. Disponível em: https://www.researchgate.net/publication/332292728_INTRODUCAO_AO_PROCCESSAMENTO_DE_IMAGENS_DE_SENSORIAMENTO_REMOTO

MOCHIDA, K.; KODA, S.; INOUE, K.; HIRAYAMA, T.; TANAKA, S.; NISHII, R.; MELGANI, F. Computer vision-based phenotyping for improvement of plant productivity: a machine learning perspective. **GigaScience**, v. 8, n. 1, p.153, 2019. DOI <https://doi.org/10.1093/gigascience/giy153>.

MOURA, M. C. C. L.; AZEVEDO, A. M.; SILVA, D. J. H.; CRUZ, C. D. Potencialidades das redes neurais artificiais na avaliação de recursos genéticos em bancos de germoplasma. **Revista RG News**, Brasília, DF. Sociedade Brasileira de Recursos Genéticos, p. 14-19, 2015. Disponível em: https://www.recursosgeneticos.org/Recursos/Arquivos/4._Potencialidades_das_rede_s_neurais_artificiais_na_avaliao_de_recursos_gen_ticos_em_bancos_de_germoplasma.pdf

NAGAMINE, A; TAKAYAMA, M.; EZURA, H. Genetic improvement of tomato using gene editing technologies. **The Journal of Horticultural Science and Biotechnology**, v. 98, n. 1, p. 1-9, 2023. Doi: <https://doi.org/10.1080/14620316.2022.2091489>

NICK, C.; DA SILVA, D. J. H. Melhoramento de Tomate. In: . In: NICK, C.; BORÉM, A. **Melhoramento de Hortaliças**. 1 ed. Viçosa: Editora UFV, 2016. p. 396-464.

NUNES, M. L.; DA COSTA, J. C.; DE SOUZA, D. G. **Entendendo o funcionamento do cérebro ao longo da vida**. 1 ed. Editora da PUCRS, 2022. 212 p.

OLIVEIRA, G. R. F. D.; SALLES, F. K. L. D.; BATISTA, T. B.; SILVA, M. S. D.; CICERO, S. M.; GOMES-JUNIOR, F. G. Morphological parameters of image processing to characterize primary root emergence in evaluation of tomato seed vigor. **Journal of Seed Science**, v. 43, 2021. DOI <https://doi.org/10.1590/2317-1545v43245215>

OLIVEIRA, P. F.; CÂMARA, C. E. Análise de desempenho de um algoritmo desenvolvido para solução de deep learning utilizando redes neurais convolucionais para Análise de contraste de imagens. **Revista de Ubiquidade**, v. 2, n. 1, p. 84-105, 2019. Disponível em: <https://revistas.anchieta.br/index.php/RevistaUbiquidade/article/view/1010/893>

ONOFRI, A.; TERZAROLI, N.; RUSSI, L. Linear models for diallel crosses: a review with R functions. **Theoretical and Applied Genetics**, v. 134, p. 585-601, 2021. DOI <https://doi.org/10.1007/s00122-020-03716-8>

PADOVANI, M. I. **Tomate: o fruto do amor que conquistou o mundo**. 2a ed. São Paulo: Icone, 1989. 156 p.

PERSECHINO, A.; DE ALBUQUERQUE, M. P. Image processing: fundamental concepts. **Monografias-CBPF**, v. 1, n. 4, 2015.

DOI <http://dx.doi.org/10.7437/MO2447-1119/2015.04.001>

PHAN, Q. H; NGUYEN, V. T.; LIEN, C. H.; DUONG, T. P.; HOU, M. T. K.; LE, N. B. Classification of Tomato Fruit Using Yolov5 and Convolutional Neural Network Models. **Plants**, v. 12, n. 4, p. 790, 2023. DOI <https://doi.org/10.3390/plants12040790>

PRINCE, S. J. D. **Computer vision: models, learning, and inference**. Cambridge University Press, 2012. 600p

RAZIFARD, H.; RAMOS, A.; DELLA VALLE, A.L; BODARY, C.; GOETZ, E., MANSER, E.J.; LI, X.; ZHANG, L.; VISA, S.; TIEMAN, D.; VAN DER KNAAP, E.; CAICEDO, A.L.; Genomic evidence for complex domestication history of the cultivated tomato in Latin America. **Molecular biology and evolution**, v. 37, n. 4, p. 1118-1132, 2020.

DOI <https://doi.org/10.1093/molbev/msz297>

RAMALHO, M. A. P.; SANTOS, J. D.; ZIMMERMANN, M. D. O. **Genética quantitativa em plantas autógamas: aplicações ao melhoramento do feijoeiro**. Goiânia: UFG, 1993. 271 p.

REYNOLDS, M.; CHAPMAN, S.; CRESPO-HERRERA, L.; MOLERO, G.; MONDAL, S.; PEQUENO, D. N.; PINTO, F.; PINERA-CHAVEZ, F. J.; POLAND, J.; RIVERA-AMADO, C.; PIERRE, C. S.; SUKUMARAN, S. Breeder friendly phenotyping. **Plant Science**, v. 295, p. 110396, 2020. DOI <https://doi.org/10.1016/j.plantsci.2019.110396>

SAMBASIVAM, G.; OPIYO, G.D. A predictive application of machine learning in agriculture: cassava disease detection and classification with imbalanced dataset using networks convolutional neural networks. **Journal of Egyptian Informatics**, v.22, n.1, p.27-34, 2021. DOI <https://doi.org/10.1016/j.eij.2020.02.007>

SABROL, H. Recent Studies of Image and Soft Computing Techniques for Plant Disease Recognition and Classification. International. **Journal of Computer Applications**, v. 126, n. 1, p. 44–55, 2015.

DOI <http://dx.doi.org/10.5120/ijca2015905982>

SAKAMOTO, T.; SHIBAYAMA, M.; KIMURA, A.; TAKADA, E. Assessment of digital camera-derived vegetation indices in quantitative monitoring of seasonal rice growth. **ISPRS Journal of Photogrammetry and Remote Sensing**, v. 66, n.6, p. 872-882, 2011. DOI <https://doi.org/10.1016/j.isprsjprs.2011.08.005>

SALEEM, M. Y.; ASGHAR, M.; IGBAL, Q.; RAHMAN, A.; AKRAM, M. Diallel analysis of yield and some yield components in tomato (*Solanum lycopersicum* L.). **Pak. J. Bot**, v. 45, n. 4, p. 1247-1250, 2013.

SHAMMAI, A.; PETREIKOV, M.; YESELSON, Y.; FAIGENBOIM, A.; MOY-KOMEMI, M.; COHEN, S.; COHEN, D.; BESAULOV, E.; EFRATI, A.; HOUMINER, N.; BAR, M.; AST, T.; SCHULDINER, M.; KLEMENS, A. W.; NEUHAUS, E.; BAXTER, C. J.; RICKETT, D.; BONNET, J.; WHITE, R.; GIOVANNONI, J. J.; LEVIN, I.; SCHAFFER, A. Natural genetic variation for expression of a sweet transporter among wild species of *Solanum lycopersicum* (tomato) determines the hexose composition of ripening tomato fruit. **The Plant Journal**, v. 96, n. 2, p. 343-357, 2018. DOI <https://doi.org/10.1111/tpj.14035>

SIERRA-OROZCO, E.; SHEKASTEBAND, R.; ILLA-BERENGUER, E.; SNOUFFER, A.; VAN DER KANAAP, E.; LEE, T. G. Identification and characterization of GLOBE, a major gene controlling fruit shape and impacting fruit size and marketability in tomato. **Horticulture Research**, v. 8, 2021. DOI <https://doi.org/10.1038/s41438-021-00574-3>

SILVA DIAS, J. C. Guiding strategies for breeding vegetable cultivars. **Agricultural Sciences**, v. 5, n. 1, p. 9, 2014. DOI <http://dx.doi.org/10.4236/as.2014.51002>

SIMONYAN, K.; ZISSERMAN, A. Very deep convolutional networks for large-scale image recognition. **arXiv preprint arXiv:1409.1556**, 2014. DOI <https://doi.org/10.48550/arXiv.1409.1556>

SOUZA, V.; SILVA, L.; SANTOS, A.; ARAUJO, L. Análise comparativa de redes neurais convolucionais no reconhecimento de cenas. **Anais do Computer on the Beach**, v. 11, n. 1, p. 419-426, 2020.

SOUSA, A.D.L.; SALAME, M. F. A.; DO NASCIMENTO FILHO, F.J.; ATROCH, A.L. Redes neurais convolucionais aplicadas ao processo de classificação de cultivares de guaranazeiros. 2017. Disponível em: <https://www.alice.cnptia.embrapa.br/alice/bitstream/doc/1091093/1/AnaisENIAC2017.compressed.pdf>

STOCKMAN, G.; SHAPIRO, L. G. **Computer vision**. Prentice Hall PTR, 2001. 580 p.

SZEGEDY, C.; VANOUCKE, V.; IOFFE, S.; SHLENS, J.; WOJNA, Z. Rethinking the inception architecture for computer vision. **IEEE Conference on Computer Vision and Pattern Recognition (CVPR)**, p. 2818–2826, 2016. DOI <https://doi.org/10.1109/CVPR.2016.308>

SZEGEDY, C.; IOFFE, S.; VANHOUCHE, V.; ALEMI, A. Inception-v4, inception-resnet and the impact of residual connections on learning. In: **Proceedings of the AAAI conference on artificial intelligence**. v. 31, n. 1, 2017. DOI <https://doi.org/10.1609/aaai.v31i1.11231>

TARDIEU, F.; CABRERA-BOSQUET, L.; PRIDMORE T.; BENNETT, M. Plant phenomics, from sensors to knowledge. **Current Biology**, v. 27, n. 15, p. R770-R783, 2017. DOI <https://doi.org/10.1016/j.cub.2017.05.055>

TATTARIS, M.; REYNOLDS, M. P.; CHAPMAN, S. C. A direct comparison of remote sensing approaches for high-throughput phenotyping in plant breeding. **Frontiers in Plant Science**, v. 7, p. 1131, 2016. DOI <https://doi.org/10.3389/fpls.2016.01131>

TOMATO GENOME CONSORTIUM. "The tomato genome sequence provides insights into fleshy fruit evolution". **Nature**, v. 485, n. 7400, p. 635, 2012. DOI [10.1038/nature11119](https://doi.org/10.1038/nature11119)

VARGAS, A. C. G.; PAES, A.; VASCONCELOS, C. N. Um estudo sobre redes neurais convolucionais e sua aplicação em detecção de pedestres. In: Proceedings of the XXIX Conference on Graphics, **Patterns and Images**. p. 1–4, 2016.

VARSHNEY R.K; GRANER, A.; SORRELLS, M. E. Genomics-assisted breeding for crop improvement. **Trends in plant science**, v. 10, n. 12, p. 621-630, 2005. DOI <https://doi.org/10.1016/j.tplants.2005.10.004>

VEIGA, R. D.; FERREIRA, D. F.; RAMALHO, M. A. P. Eficiência dos dialelos circulantes na escolha de genitores. **Pesquisa Agropecuária Brasileira**, v. 35, n. 7, p. 1395–1406, 2000. DOI <https://doi.org/10.1590/S0100-204X2000000700014>

WALTER, A.; LIEBISCH, F.; HUND, A. Plant phenotyping: from bean weighing to image analysis. **Plant methods**, v. 11, n. 1, p. 1-11, 2015. DOI <https://doi.org/10.1186/s13007-015-0056-8>

YU, C.; FAN, X.; HU, Z.; XIA, X.; ZHAO, Y.; Li, R.; BAI, Y. Segmentation and measurement scheme for fish morphological features based on Mask R-CNN. **Information Processing in Agriculture**, v. 7, n. 4, p. 523-534, 2020. DOI <http://dx.doi.org/10.1016/j.inpa.2020.01.002>

4 ARTIGOS

4.1 CAPÍTULO I: FENOTIPAGEM DE ALTO RENDIMENTO NA AVALIAÇÃO DA PRODUTIVIDADE DE UM DIALELO COM TOMATEIRO

RESUMO

O tomate (*Solanum lycopersicum* L.) é uma das hortaliças mais importantes na economia mundial, conseqüentemente, apresenta-se como organismo- modelo para pesquisas de biotecnologia e programas de melhoramento vegetal. A fenotipagem por imagem é uma técnica que pode ser aplicada nesses programas e possibilita a análise do experimento de forma rápida, acurada, objetiva e sem destruição de amostras. Nesse sentido, objetivou-se com este estudo estabelecer metodologias, para a fenotipagem da produtividade no tomateiro, por meio da análise computacional de imagens pelo algoritmo Mask-RCNN e testar sua eficiência em um dialelo balanceado sem recíprocos com o tomateiro. Avaliaram-se as F1's na época da colheita dos frutos pela fenotipagem tradicional quanto à produtividade e, em paralelo, imagens foram capturadas para avaliação pela Mask-RCNN. Estimaram-se pela Mask-RCNN seis variáveis, sendo elas o número de frutos verdes, número de frutos maduros, número de frutos totais, porcentagem da área na imagem ocupada por frutos maduros (PFM), porcentagem da área ocupada por frutos verdes (PFV) e porcentagem de área ocupada por todos os frutos (PFT). A precisão para o reconhecimento foi de 85% para frutos maduros e 88% para frutos verdes. O modelo atingiu recall de 92% para frutos maduros e 84% para frutos verdes. Foi observada correlação significativa entre os caracteres avaliados de forma tradicional e computacionalmente, todos com valores positivos e próximos a 1. A correlação dos ranks entre a produtividade de frutos, obtida pela fenotipagem tradicional e as variáveis estimadas pela Mask-RCNN, foi alta, em especial, para o número de frutos maduros e totais. Concluiu-se pelos resultados que a análise de imagem é eficiente e pode ser utilizada na fenotipagem de alto rendimento no melhoramento genético do tomateiro.

Palavras-chave: Aprendizado de máquina. Fenômica. Redes Neurais Convolucionais. Fenotipagem por imagem. Melhoramento do Tomateiro.

HIGH-YIELD PHENOTYPING IN EVALUATING THE PRODUCTIVITY OF A DIALELL WITH TOMATO PLANT

ABSTRACT

Tomato (*Solanum lycopersicum L.*) is one of the most important vegetables in the world economy, consequently, it presents as a model organism for biotechnology research and plant breeding programs. The phenotyping by image is a technique that can be applied in these programs and allows the analysis of the experiment quickly, accurately, objectively and without destroying samples. In this sense, the objective of this study was to establish methodologies for phenotyping of the productivity in tomato plant, through computational analysis of images using the Mask-RCNN algorithm and to test its efficiency in a balanced diallel without reciprocals with the tomato plant. They were evaluated the F1's at the time of fruit harvest by the traditional phenotyping for productivity and, in parallel, images were captured for evaluation by Mask-RCNN. They were estimated six variables by Mask-RCNN, namely the number of green fruits, number of ripe fruits, number of total fruits, percentage of the area in the image occupied by ripe fruits (PFM), percentage of the area occupied by green fruits (PFV) and percentage of area occupied by all fruits (PFT). Accuracy for recognition was 85% for ripe fruits and 88% for green fruits. The model achieved recall of 92% for ripe fruits and 84% for green fruits. It was observed a significant correlation between the characters evaluated in traditional and computational way, all with positive values and close to 1. The ranks correlation between fruit productivity, obtained by traditional phenotyping and the variables estimated by Mask-RCNN, was high, especially for the number of ripe and total fruits. It was concluded from the results that image analysis is efficient and can be used in high-throughput phenotyping in tomato plant genetic improvement.

Keywords: Machine learning. Phenomics. Convolutional Neural Networks. Phenotyping by image. Tomato plant Improvement.

4.1.1 INTRODUÇÃO

O tomateiro (*Solanum lycopersicum* L.) é uma espécie diploide ($2n=2x=24$) e autógama. Nativa da América do Sul, destaca-se como hortaliça de importância econômica no mundo e com inúmeros benefícios à saúde humana. É considerado um alimento funcional por conter altos teores de vitaminas A e C, além de ser rico em licopeno. Da domesticação até os dias atuais, o tomate foi continuamente sujeito à seleção para a melhoria do formato, cor, tamanho, sabor das frutas, tolerância das plantas a estresses abióticos e bióticos (RAZIFARD *et al.*, 2020).

No melhoramento genético vegetal, a escolha dos genitores condiciona o sucesso do programa de melhoramento, que, ao serem cruzados, produzam híbridos e, posteriormente, populações segregantes promissoras, favorecendo o trabalho do melhorista na obtenção de progresso genético. O melhoramento de plantas é uma atividade cara e demorada, por isso, monitorar a eficiência dos genitores é de extrema importância, pois proporciona estimativas de ganho genético que permitem analisar criticamente a eficácia desses programas (STREK *et al.*, 2018). Existem muitas metodologias eficientes para a escolha de cultivares e linhagens a serem utilizadas nos programas de hibridação. Entre as metodologias disponíveis, os cruzamentos dialélicos são amplamente utilizados em diferentes espécies cultivadas (MAT *et al.*, 2020).

A análise dialélica tem sido amplamente utilizada na identificação de combinações híbridas heteróticas e explorada a possibilidade de se obterem híbridos com maior proporção de frutos com padrão comercial, melhor qualidade nutricional, resistentes a múltiplas doenças e com maior conservação pós-colheita (MATOS *et al.*, 2021). Porém um dos grandes desafios, para a avaliação desses experimentos, reside na fenotipagem dos indivíduos em estudo, que precisa avançar em escala, precisão, reprodutibilidade e exploração de novas características, em relação às técnicas tradicionais (CHANDRA *et al.*, 2020). Essa fenotipagem tem se tornado mais fácil e precisa, ao longo da última década, recorrendo-se ao uso da análise de imagens (WALTER *et al.*, 2015).

A fenotipagem por análise computacional de imagens é vantajosa, pois é rápida, não invasiva, precisa, de baixo custo, livre de resíduos, com boa resolução espacial e temporal, permite explorar áreas relativamente grandes, proporciona

grande número de repetições e segurança ao operador, viabilizando estudos comparativos e a seleção fenotípica em larga-escala (GEBREMEDHIN *et al.*, 2019). O número de estudos utilizando esta estratégia tem crescido rapidamente, sendo uma importante ferramenta para classificação de frutas na colheita e pós-colheita (ALKREEM; KARIM, 2021).

Para a obtenção de informações, a partir de imagens, pode-se recorrer ao uso de redes neurais (MOCHIDA *et al.*, 2019), permitindo a previsão de características que são difíceis de medir, oferecendo melhores resultados do que a seleção indireta tradicional (CHANDRA, *et al.*, 2020). As Redes Neurais se referem a um modelo matemático inspirado no cérebro humano que fornece um esquema não linear de computação de mineração de dados para modelar as complexas relações entre variáveis de entrada (preditores) e variáveis de saída (JAHAN *et al.*, 2020; PATTERSON; GIBSON, 2017). Consequentemente podem ter desempenho superior ao dos modelos estatísticos convencionais, em virtude de sua estrutura não linear, com a vantagem de serem não paramétricas, não necessitarem de informações detalhadas sobre os processos físicos do sistema a ser modelado e tolerarem perda de dados (AZEVEDO *et al.*, 2015).

Entre as técnicas de redes neurais para a análise de imagens, as Redes Neurais Convolucionais (CNNs) se destacam. Entre os algoritmos existentes, o Mask-RCNN é vantajoso, pois, além de possibilitar alta acurácia, permite a classificação e a segmentação dos objetos de interesse (SHARMA, 2019). Nesse sentido, objetivou-se verificar a eficiência da fenotipagem no tomateiro, por meio da análise computacional de imagens pelo algoritmo Mask-RCNN, na avaliação um experimento em dilelo.

4.1.2 MATERIAL E MÉTODOS

O estudo foi dividido em duas etapas, uma conduzida em casa de vegetação e a outra em campo, na fazenda experimental Professor Hamilton de Abreu Navarro (FEHAN) do Instituto de Ciências Agrárias (ICA)-UFMG, localizada no *Campus* Regional Montes Claros, (coordenadas: 16°40'58.16" S e 43°50'20.15" O).

4.1.2.1 Obtenção das sementes híbridas.

Como progenitores, foram utilizadas cinco linhagens de tomateiro comerciais, sendo eles: San Marzano e Santa Clara (da empresa Isla), Santa Cruz Kada gigante e Gaúcho Melhorado Nova Seleção (da empresa Feltrin) e Gaúcho (da empresa Top Seed). Como o objetivo nesta pesquisa está apenas no processo de fenotipagem, não pretendendo comparar a superioridade agrônômica entre estas linhagens, os denominamos aleatoriamente de A, B, C, D e E.

Por meio de um sistema de cruzamentos de dialelos balanceados de meia tabela, sem recíprocos, foi obtida uma população segregante por hibridação artificial constituindo 15 tratamentos (combinações híbridas e linhagens), a partir desses cinco progenitores, conforme a **Tabela 1**.

Tabela 1- Dialelo balanceado envolvendo os cinco progenitores, suas respectivas autofecundações e 10 híbridos.

PROGENITORES		Masculino				
		A	B	C	D	E
Feminino	A	AXA	AXB	AXC	AXD	AXE
	B		BXB	BXC	BXD	BXE
	C			CXC	CXD	CXE
	D				DXD	DXE
	E					EXE

Fonte: Da autora, 2024

Essa etapa foi conduzida em casa de vegetação. Para cada uma das cinco variedades, foram utilizadas quatro repetições de três plantas por parcela com uma planta por vaso, totalizando 60 plantas. Para genitor feminino, utilizaram-se 2/3 das plantas, o restante como genitor masculino.

As plantas foram tutoradas com bambu e conduzidas em duas hastas, eliminando-se as brotações das axilas das folhas (**Figura 1**). A adubação foi feita, de acordo com a análise de solo e as recomendações, para a cultura do tomateiro, utilizando-se como referência as recomendações de Furlani; Bataglia (2018). A irrigação foi feita diariamente com regador. O controle fitossanitário, sempre que necessário, foi realizado com produtos registrados para a cultura.

Figura 1- condução das plantas em casa de vegetação.



Fonte: Da autora, 2021.

A polinização artificial, para a obtenção das sementes híbridas, foi realizada pela manhã, no período das 6 às 10h. Os botões florais recém-abertos de cada um dos genitores masculinos foram coletados para a retirada do pólen (**Figura 2A**). O pólen obtido de cada genitor foi retirado e imediatamente levado até o estigma do genitor feminino.

As flores dos genitores femininos foram emasculadas antes da antese, quatro a cinco dias de desenvolvimento da flor. No processo de emasculação, foi utilizada uma pinça inox com ponta fina, para a abertura das pétalas, removendo os estames, deixando o estigma pronto para a polinização (**Figura 2B**) (SILVA; GIORDANO, 2000). Os estigmas dos botões emasculados receberam o pólen por contato direto.

Os botões emasculados e polinizados foram protegidos com envelopes de papel e identificados quanto ao cruzamento. Após um período de sete dias, os envelopes foram retirados para a averiguação do sucesso nas hibridações (**Figura 2C**).

Figura 2-(A) Polén retirado das flores do genitor masculino; (B) Flor emasculada; (C) Sucesso no pegamento e crescimento dos frutos 15 dias após a polinização.



Fonte: autora 2021.

Os frutos viáveis foram colhidos, depois de apresentarem completa maturação, 100 dias após o transplante das mudas e extraídas as sementes.

Para a obtenção das sementes, F1, os frutos foram cortados para retirar as sementes juntamente com a massa gelatinosa. Essa mistura foi mantida sob fermentação natural por imersão em água em laboratório durante 48 horas. Em seguida, fez-se a lavagem em água corrente, visando à separação apenas das sementes as quais foram tratadas com hipoclorito de sódio a 2,5 % p.p e colocadas para secar.

Após secagem completa, foram colocadas em sacos de papel, identificadas e armazenadas, em condições de 18°C e 40% de umidade relativa do ar, até a implantação do experimento II, de 30 dias.

Foram obtidas imagens das plantas, por meio de smartphones, durante todas as fases fenológicas da planta nessa etapa. O objetivo foi criar um banco de dados para treinar as redes neurais de modo a reconhecer o que são os frutos maduros e imaturos.

4.1.2.2 Instalação, condução e avaliação da F1 e genitores

Sementes das 10 combinações híbridas, junto com os cinco genitores, foram semeadas em bandejas de isopor de 128 células, contendo substrato comercial para hortaliças (Bioplant) e conduzidas até o transplante.

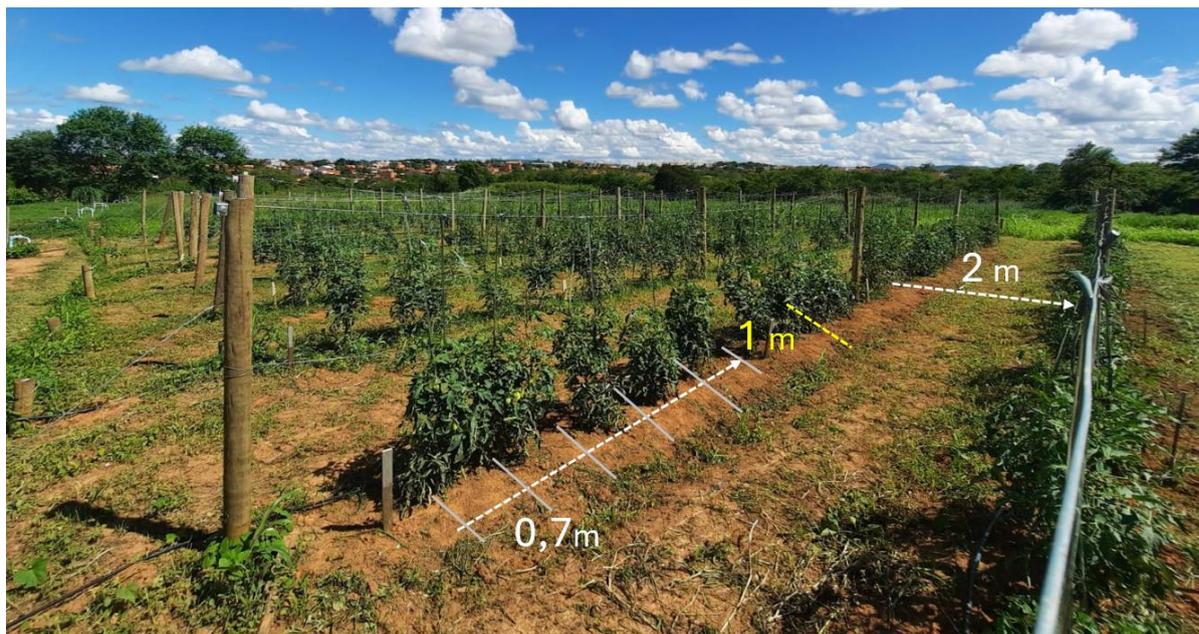
Após transplantadas para o campo, as plantas de tomateiro foram tutoradas e conduzidas em duas hastas. Cada haste foi manejada, para a formação de seis a

sete cachos de frutos e, como não houve desbastes de fruto na condução, o número de frutos variou, dessa forma, de acordo com as características dos genótipos.

Adubação foi feita, de acordo com a análise de solo e recomendações para a cultura do tomateiro, utilizando-se como referência Furlani; Bataglia (2018). A irrigação foi realizada diariamente por gotejamento. O controle fitossanitário, sempre que necessário, foi realizado com produtos registrados para a cultura.

O experimento foi conduzido, em delineamento em blocos casualizados (DBC) com quatro repetições, 15 tratamentos (dez híbridos e cinco linhagens) e cinco plantas por parcela, totalizando 300 plantas. Em 10 linhas de plantio (canteiros), com espaçamento entre plantas dentro das parcelas de 0,70m e 1m entre parcelas e espaçamento entre linhas de 2 metros (**Figura 3**), de forma a facilitar a aquisição das imagens.

Figura 3- Área do experimento com a cultura do tomateiro.



Fonte: Da autora, 2021.

4.1.2.3 Fenotipagem tradicional e por análise de imagens

As plantas do tomateiro foram avaliadas para a produtividade, durante doze colheitas, quando se iniciou a maturação dos frutos. Os frutos vermelhos eram colhidos duas vezes por semana, com um intervalo de três dias.

Antes da colheita era realizada a captura das imagens das parcelas por uma filmagem da linha de plantio (**Figura 4A**). Utilizou-se um smartphone Xiaomi modelo Poco M3, equipado com uma câmera de 48 megapixels capaz de capturar imagens

com resolução de até 8000x6000 pixels e vídeos em alta definição (Full HD) de até 1920x1080 pixels. Em razão do campo aberto, a luminosidade foi natural, variando conforme as condições climáticas do dia da colheita. A distância da câmera ao canteiro e ao solo foram balizadas com um fio de arame, de modo a não variar, ao longo da linha, sendo de 1,5 m dos canteiros e 1,1 m de altura. Um tecido azul foi utilizado, para cobrir o fundo das plantas evitando a interferência das plantas das linhas vizinhas de plantio (**Figura 4A**). Esses vídeos foram decompostos em imagens pelo software R e usadas para criar um ortomosaico frontal de cada parcela (**Figura 4B**), com o programa *Agisoft-Metashape*. Um ortomosaico é uma imagem composta criada pela sobreposição e combinação de várias fotografias. Essas imagens individuais são capturadas de diferentes ângulos e posições e depois são processadas por software especializado para corrigir distorções geométricas e criar uma imagem panorâmica. Utilizou-se essa ferramenta para obter-se uma imagem frontal de cada parcela no dia da colheita.

Figura 4-Obtenção de vídeos da linha de plantio antes da colheita (A) e Ortomosaico de uma parcela criado a partir de imagens decompostas do vídeo, utilizando o programa Agisoft- Metashape (B)



Fonte: autora 2022.

4.1.2.3_1 Análise Computacional das Imagens

Modelos de rede neural convolucional com o algoritmo Mask-RCNN foram treinados para identificar frutos verdes e maduros. Para o treinamento foi feita a criação de um banco sintético de imagens. Para isso, foram criadas imagens em formato *png* (com fundo transparente), contendo apenas os frutos, totalizando 100 frutos maduros e 100 de frutos imaturos. Na criação do banco sintético, essas imagens em formato *png* compuseram o *foreground* com duas subclasses (frutos maduros e imaturos). A edição para a obtenção destas imagens *png* foi feita com o auxílio do

software GIMP versão 2.10.22. Quanto ao *background* consideraram-se 200 imagens dos experimentos sem a presença dos frutos.

O banco sintético de imagens foi criado no formato COCO, com auxílio da biblioteca cocosynth (<https://github.com/akTwelve/cocosynth>), no ambiente do *Google colaboratory*, utilizando a linguagem de programação *python*. Para isso, foi estabelecido o número de 3000 imagens para treinamento e 1000 imagens para validação aleatoriamente. Dentro do algoritmo utilizado, estabeleceu-se, ao acaso, o número máximo de cinco objetos *png* por imagem. O posicionamento dessas imagens *png* sobre o *background* se deu de forma aleatória (**Figura 5**). Além disso, para a criação do banco sintético de imagens, foram permitidas alterações aleatórias na rotação, escala e brilho para a sobreposição das imagens *png* sobre o *background*. Por meio dessa metodologia, a máscara da imagem e arquivo com as definições das instâncias no formato json foram obtidas automaticamente. O tamanho das imagens de saída foi definido em 320x320 *pixels*. Tudo isso é útil para a obtenção de um banco com variabilidade de imagens suficiente para o treinamento mais eficiente.

Figura 5- Exemplo de imagens do banco sintético com imagens com a extensão *png* sobre o *background* (A) e sua respectiva máscara (B) obtida pela biblioteca cocosynth



Fonte: Da autora, 2022.

Para o treinamento das redes neurais convolucionais, utilizou-se o algoritmo Mask R-CNN (https://github.com/matterport/Mask_RCNN). Essa etapa também foi realizada no *Google colaboratory*, o ambiente de execução definido foi o GPU tornando o treinamento mais rápido. A rede utilizada, como extratora das

características, foi a ResNet101 e o modelo pré-treinado MS COCO. No uso do algoritmo Mask R-CNN, considerou-se o uso de cinco imagens por GPU, três classes (*background*, frutos maduros e frutos imaturos) e taxa de aprendizado de 0,001.

Foi utilizada a técnica do *Transfer learning*, a qual inicia o treinamento, a partir de pesos sinápticos já treinados no formato cocodata. Foram consideradas 100 épocas de treinamento, para o ajuste dos pesos sinápticos das últimas camadas da rede e, posteriormente, mais 100 épocas, para o ajuste dos pesos sinápticos de todas as camadas da rede.

Ao final do treinamento da rede, foi realizado o teste com as imagens. Nesse momento, estabeleceu-se a confiança mínima de 90% para a identificação e classificação dos frutos. Para avaliar a precisão na detecção dos objetos pela Mask R-CNN, construiu-se uma tabela de contingência, resultante da predição da classificação comparada com as classes reais. Para verificar a eficiência na classificação, foi estimado a Precisão, recall e F-measure.

$$Recall = \frac{VP}{VP + FN} \quad (eq1)$$

$$Precisão = \frac{VP}{VP + FP} \quad (eq2)$$

$$F - Measure = \frac{2 * Precisão * Recall}{Precisão + Recall} \quad (eq3)$$

Sendo:

- Verdadeiros Positivos (VP): Instâncias que foram corretamente previstas como pertencentes à classe positiva.
- Falsos Positivos (FP): Instâncias que foram erroneamente previstas como pertencentes à classe positiva quando, na verdade, pertencem à classe negativa.
- Verdadeiros Negativos (VN): Instâncias que foram corretamente previstas como pertencentes à classe negativa.
- Falsos Negativos (FN): Instâncias que foram erroneamente previstas como pertencentes à classe negativa quando, na verdade, pertencem à classe positiva.

Adicionalmente, foi estimada, por meio da análise de imagens, para cada parcela, a partir das imagens coletadas anteriormente à cada colheita, o número de frutos totais (NFT), número de frutos verdes (NFV) e número de frutos maduros (NFM). A partir da máscara obtida sobre cada fruto nas imagens, foram estimadas a porcentagem de pixels da imagem correspondentes aos frutos maduros (PPM), porcentagem de pixels da imagem correspondentes aos frutos verdes (PPV) e porcentagem de pixels da imagem correspondentes a todos os frutos (PPT).

4.1.2.3_2 Análise Estatística-Delineamento genético

Para a análise estatística, foi considerada a produtividade de frutos, obtida por meio de pesagem em t/ha. Além disso, foram consideradas as variáveis obtidas por meio da análise de imagem (NFT, NFV, NFM, PPM, PPV e PPT). Os dados dessas variáveis, para os 15 tratamentos (combinações híbridas e linhagens), em nível de média, foi submetida à análise de variância. A soma dos quadrados para tratamento foi decomposta em capacidade geral de combinação (CGC) e capacidade específica de combinação (CEC), com o auxílio do programa computacional GENES (CRUZ, 2013), considerando o Método 2 proposto por Griffing (1956), nos quais são incluídos $p(p+1)/2$ combinações, faltando os híbridos F1's recíprocos. Com o seguinte modelo matemático:

$$y_{ij} = m + g_i + g_j + s_{ij} + \bar{\varepsilon}_{ij}$$

Em que:

y_{ij} : valor médio da combinação híbrida ($i \neq j$) ou do progenitor ($i = j$)

m : média geral

g_i, g_j : efeito da capacidade geral de combinação do i -ésimo e do j -ésimo progenitor, respectivamente;

s_{ij} : efeito da capacidade específica de combinação para os cruzamentos entre os progenitores de ordem i e j ; e

$\bar{\varepsilon}_{ij}$: erro experimental médio

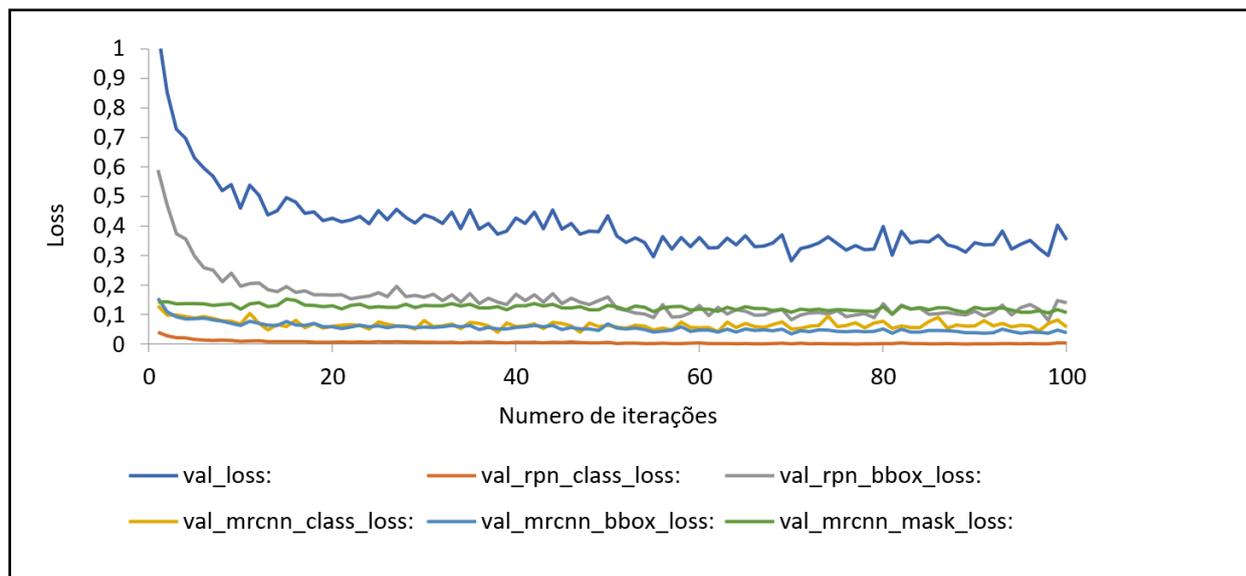
Nesse modelo é considerado que $s_{ij} = s_{ji}$

Para verificar a eficiência do uso da análise de imagens, o ranqueamento da CGC e CEC obtido, a partir das variáveis estimadas por essa abordagem (NFM, NFV, NFT, PPV, PPM e PPT), foi comparada com a produtividade. Para isso, foi utilizada a correlação de Spearman com a função *cor* do pacote *stats* do software R. As correlações foram apresentadas recorrendo a figuras obtidas pelo pacote *corplot*. Por fim, foi utilizada a regressão múltipla, para verificar a possibilidade de estimar a produtividade, por meio de todas as variáveis estimadas pelas redes simultaneamente (NFM, NFV, NFT, PPV, PPM e PPT). Para isso, considerou-se no modelo todos os efeitos simples e efeitos de interação. Posteriormente, a fim de manter apenas os efeitos mais relevantes, utilizou-se a metodologia *stepwise* pelas funções *lm* e *step* do pacote *stats*.

4.1. 3 RESULTADOS

Após o treinamento das redes, foram obtidas curvas da função de perda para os dados de treinamento e validação apresentados na **Figura 6**. Nesta figura, pode-se observar uma redução acentuada na função de perda (Loss) nas dez primeiras épocas. Sua estabilização, a partir da 60^a época, indica que a utilização de um número maior de iterações que a utilizada nesta pesquisa (100) é dispensável, visto que maior número de treinamento não ocasionaria maior ganho na classificação dos frutos. Em outras palavras, indica que o número de épocas utilizado foi suficiente.

Figura 6- Estimativa da função de perda (loss), no decorrer de 100 épocas de treinamento de redes neurais convolucionais (Mask-RCNN), na classificação de frutos maduros e imaturos do tomateiro.



Fonte: Da autora, 2024

A eficácia da rede Mask R-CNN, em classificar os frutos de tomates em verdes e maduros (**Figura 7**), foi avaliada utilizando-se valores de recall, precisão e F-measure (**Tabela 2**). O modelo atingiu recall de 92% para frutos maduros e 84% para frutos verdes. Os valores das métricas de desempenho tiveram valores bem próximos, sendo 85% de precisão para frutos maduros e 88% para frutos verdes. Essa métrica determina a eficiência de um modelo do aprendizado profundo, na classificação de dados, mostrando o total de acertos na classificação, ou seja, o quão bom o modelo é.

Figura 7- Imagens demonstrando a classificação realizada pelo algoritmo Mask-RCNN na detecção dos frutos imaturos e maduros



Fonte: Da autora, 2022.

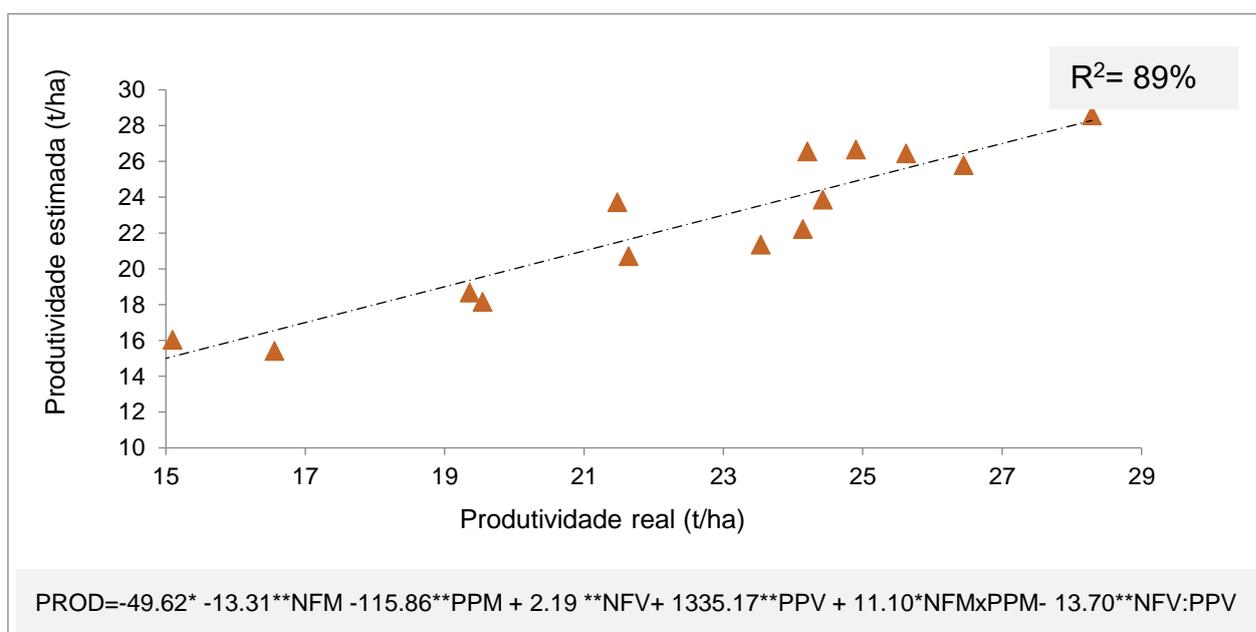
Tabela 2-Avaliadores da qualidade da classificação de frutos maduros e imaturos por redes neurais convolucionais (Mask-RCNN)

Classes	Recall	Precisão	F-measure
Frutos Maduros	0.92	0.85	0.88
Frutos Verdes	0.84	0.88	0.86

Fonte: da Autora 2022.

Ao utilizar o modelo de regressão múltipla, para estimar a produtividade, a partir das variáveis obtidas pelas redes neurais, verificou-se o coeficiente de determinação de 89% (**Figura 8**), que indica uma boa precisão na estimação da produtividade.

Figura 8- Análise de regressão múltipla, para a estimação da produtividade do tomateiro, a partir da produtividade real.



Fonte: Da autora, 2024

Foi verificada a presença de efeito significativo a 1 % de probabilidade pelo teste F, em todas as variáveis, para a fonte de variação tratamentos e para a capacidade específica de combinação (CEC). Já para a capacidade geral de combinação (CGC), não houve efeito significativo (**Tabela 3**).

O coeficiente de variação residual (CV) foi inferior a 20% para a produtividade de frutos. Porém, para as demais características estimadas pelas redes foi maior, chegando a atingir 27,66%. O coeficiente de variação genético (CVg) variou entre 17,52 a 24,17 (**Tabela 3**). O índice de variação (CVg/CVe) foi próximo a uma unidade

para a PPT, PPM, NFT e NFV. Valores do índice de variação maiores que uma unidade indica a predominância dos efeitos genéticos sobre os ambientais. O coeficiente de herdabilidade, em nível de média de progênes, foi alto para todos os caracteres, variando entre 72.25 e 80.57%, sendo o maior para PPT e o menor PPV.

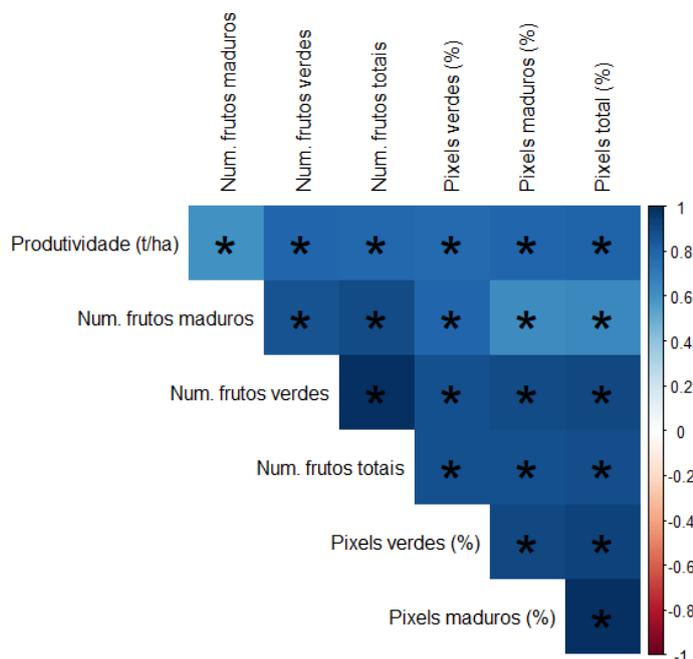
Tabela 3-. Estimativas dos quadrados médios, coeficiente de variação residual (CV), coeficiente de variação genético (CVg) e herdabilidade no estudo de um experimento em dialelo com a cultura do tomateiro

FV	Produtividade	Frutos Maduros	Frutos Verdes	Frutos Totais	Pixels Verdes	Pixels Maduros	Pixels Totais
Blocos	61.44	11.77	2152.51	2057.32	0.01	0.53	0.51
Tratamentos	78.15**	35.39**	1843.29**	2320.13**	0.01**	0.40**	0.51**
CGC	9.96 ^{ns}	28.72 ^{ns}	1287.52 ^{ns}	1569.33 ^{ns}	0.003 ^{ns}	0.31 ^{ns}	0.36 ^{ns}
CEC	105.43**	38.06**	2065.59**	2620.46**	0.009**	0.44**	0.60**
Resíduo	18.75	8.73	409.80	477.76	0.00	0.08	0.10
Média	21.99	10.68	92.33	103.01	0.17	1.23	1.40
CV(%)	19.69	27.66	21.93	21.22	26.11	23.46	22.42
CVg(%)	17.52	24.17	20.50	20.83	21.07	23.01	22.83
CVg/CVe	0.89	0.87	0.94	0.98	0.81	0.98	1.02
Herdabilidade (%)	76.00	75.34	77.77	79.41	72.25	79.37	80.57

Fonte: Da autora, 2024

A correlação fenotípica entre a produtividade de frutos estimada pela fenotipagem tradicional e as seis variáveis estimadas pela Mask-RCNN (NFV, NFM, NFT, PPM, PPV e PPT) apresentaram todos os valores significativos e próximos de 1,0 (Figura 9). Isso demonstra associação forte entre os caracteres avaliados de forma tradicional e por redes neurais. A correlação mais forte foi entre a produtividade, números de frutos verdes (NFV) e número de frutos totais (NFT).

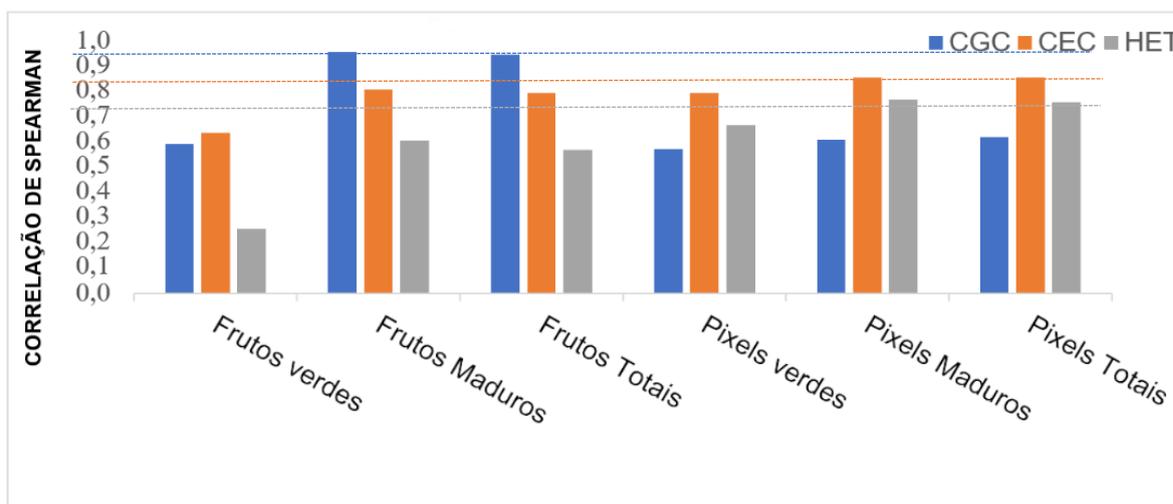
Figura 9-Correlação fenotípica entre a produtividade de frutos de tomateiro e seis variáveis estimadas por meio de redes neurais convolucionais (Mask-RCNN). * significativo pelo teste t ao nível de 5%



Fonte: Da autora, 2023.

A correlação de Spearman dos ranks entre a produtividade de frutos de tomateiro e seis variáveis estimadas pela Mask-RCNN mostrou um ordenamento similar entre os parâmetros considerando a fenotipagem tradicional e aquela obtida pela rede (**Figura 10**). As variáveis de menor desempenho foram o número de frutos verdes e a porcentagem da área da imagem ocupada por frutos verdes.

Figura 10-Correlação dos ranks de parâmetros genéticos capacidade geral de combinação – (CGC), capacidade específica de combinação (CEC) e heterose (HET) entre a produtividade de frutos de tomateiro e seis variáveis estimadas por redes neurais convolucionais (Mask-RCNN)



Fonte: Da autora, 2023.

Para NFM e NFT, de acordo com a CGC, o valor do coeficiente de correlação de Spearman foi próximo a 1, indicando concordância entre a produtividade estimada pela forma tradicional e pela Mask-RCNN. Existe uma relação forte, para CGC, entre a produtividade e os demais caracteres avaliados. A CEC apresentou valores altos na correlação dos ranks entre produtividade e as variáveis, NFM, NFT, PPM, PPV e PPT obtidos pela rede, não diferindo muito nos valores, sendo todos os valores acima de 0,8. Já para heterose, a correlação foi baixa para NFV, mas apresentou valor maior que 0,6 para as demais variáveis, obtendo valores próximos de 0,8 para as PPM e PPT.

4.1.4 DISCUSSÃO

A prática tradicional de fenotipagem de plantas, para a estimativa de produtividade e rendimento, utiliza características facilmente mensuráveis com base no peso, número e tamanho dos frutos. Esse é um processo muito demorado, caro e não é prático para grandes áreas. A estimativa por análise de imagens fornece uma solução viável nesse sentido. Os melhoristas de plantas têm buscado agregar modernas técnicas de fenotipagem ao processo de desenvolvimento de novas cultivares (CHANDRA *et al.*, 2020). Os métodos de estimativa de rendimento da cultura, baseados em visão computacional, podem ser divididos aproximadamente em

duas categorias: métodos baseados em região ou área e métodos baseados em contagem. (RAHNEMOONFAR; SHEPPARD, 2017; SCHILLACI. *et al.*, 2012; LI *et al.*, 2016; BARGOTI; UNDERWOOD, 2017; LIU *et al.*, 2019; HU *et al.*, 2019; CHU *et al.*, 2021).

Este trabalho relata resultados da detecção de tomates, em imagens feitas em campo, utilizando o algoritmo Mask-RCNN o qual detecta objetos e os pixels correspondentes a cada objeto, criando uma máscara para cada. Além disso, informações da imagem do tomate podem ser localizados facilmente analisando as características de forma e borda com as máscaras geradas a partir de Mask R-CNN (WAN *et al.*, 2018).

A tecnologia de reconhecimento de imagem está avançando em ritmo acelerado, e as CNNs permitiram aqui o treinamento de muitas amostras de forma rápida e eficiente. Modelos como a Mask-RCNN, como os que testamos neste estudo, indicam ser poderosos o suficiente para atingir uma precisão satisfatória na produtividade do tomateiro. O modelo proposto, Mask-RCNN, é robusto, para diferentes condições ambientais e cenas não estruturadas, como sobreposição, oclusão, iluminação inadequada e condições de sombreamento. Assim, pode ser aplicado, em situações de práticas reais de forma eficiente (HUANG *et al.*, 2020). CHU *et al.* (2021), ao comparar a eficiência na detecção de frutos entre as variantes das CNNs, como YOLOv3, DaSNet, Faster R-CNN, e Mask R-CNN apontaram o Mask-RCNN, como mais eficiente na detecção dessa tarefa.

A maioria das pesquisas anteriores sobre a detecção da maturidade do tomate foi feita, para aplicações de classificação, com o objetivo de distinção entre tomates vermelhos e verdes (WAN *et al.*, 2018; HUANG *et al.*, 2020; HSIEH *et al.*, 2021). Neste trabalho associamos essa classificação à quantificação da produtividade.

4.1.4.1 Avaliadores do ajuste

Os resultados de 100 imagens de teste, escolhidas aleatoriamente, mostraram que as taxas de precisão e recall foram de 88 % e 84%, respectivamente, para frutos verdes. Para os frutos maduros, a precisão foi de 85% e o recall igual a 92%.

O uso do Mask-RCNN, para detectar e classificar tomates em nossos dados, excede as métricas relatadas em trabalhos anteriores utilizando Deep Learning

(YAMAMOTO *et al.*, 2014; ZHOU *et al.*, 2017; RAHNEMOONFAR; SHEPPARD, 2017). O modelo treinado foi particularmente eficaz para a detecção de tomates verdes e maduros.

O F-measure apresentou valores de 88%, para frutos maduros e 86%, para frutos verdes, indicando a alta robustez do modelo treinado. Alcançar uma alta precisão na detecção de frutos é crucial para a estimativa de um bom rendimento. Neste estudo, as 600 imagens utilizadas, para a detecção de tomate, foram suficientes para explicar a grande variabilidade no conjunto de dados.

Falsos negativos foram observados, em maiores quantidades, nas fotos em que o número de frutos imaturos (verdes) foi maior. As prováveis causas dos erros de reconhecimento, em frutos imaturos serem maiores que a de frutos maduros, são o tamanho das amostras que também são maiores (THANAMMAL INDU *et al.*, 2022). Tem-se ainda que os modelos, baseados em detecção de objetos, para detecção de frutos, apresentam limitações, pois esses modelos não conseguem detectar frutos escondidos por folhagens ou outros frutos (KAMILARIS; PRENAFETA-BOLDU, 2018). Portanto o modelo não pode detectar todos os frutos, mas pode detectar a maioria deles (APOLO- APOLO *et al.*, 2020)

Outro fato importante que pode ter interferido, no valor do Recall e Precisão, para frutos verdes, é o problema da confusão com o fundo, pela semelhança dessa classe com o fundo. Um quarto fator seria que frutos maduros e verdes podem ter sido mal avaliados para o grau de maturação, pois, mesmo para o olho humano, a maturidade pode ser confundida, e os resultados da classificação do modelo são afetados por erros humanos, ao criar as máscaras para o conjunto de treino. Em um estudo com a cultura do tomateiro, realizado por Hsieh *et al.* (2021), foi destacado que os tomates considerados como alvos para a colheita, de acordo com a classificação da Mask-RCNN, foram os tomates que os especialistas identificaram como passíveis de serem colhidos e, portanto foram rotulados como tal. Os resultados do treinamento mencionados no trabalho indicam que tanto a Precisão quanto o Recall atingiram 98%, o que demonstra uma alta confiabilidade na distinção entre os tomates maduros determinados pela Mask R-CNN e aqueles avaliados como maduros pelos especialistas.

De forma a contornar esses fatores de detecção, Rahnemoonfar e Sheppard, (2017) desenvolveram uma nova arquitetura de CNN que conta objetos sem detectá-

los. Os resultados experimentais mostram uma precisão média de teste de 91% em imagens reais e 93% em imagens sintéticas.

Contudo os resultados desses avaliadores são bons e revelam que as Mask-RCNN se mostraram eficientes na classificação de tomates maduros e verdes, associados à produtividade, a partir de imagens, tendo em vista que os resultados dos parâmetros de acurácia apresentaram valores que indicam uma concordância entre os valores da produtividade e o número de frutos verdes e maduros detectados pela rede.

4.1.4.2 Análise de Regressão e Coeficiente de Determinação

De acordo com os coeficientes angulares e intercepto da equação da análise de regressão que foram significativos demonstrou-se que é possível, estatisticamente, estimar a produtividade, por meio das variáveis número estimado de frutos e porcentagem de pixels correspondentes aos frutos pelas redes neurais. Quando feita a regressão linear múltipla, bons resultados foram encontrados, com um $R^2 = 0,89$. Os resultados indicaram que o algoritmo tem alto desempenho, para fornecer a estimativa de produtividade da cultura, considerando os parâmetros avaliados pela rede. São ferramentas simples e eficazes para prever a produção de tomate (BELOUZ 2022). É importante acrescentar que, neste estudo, foram adquiridas imagens apenas de um lado da planta. Payne *et al.* (2013) observaram um aumento no R^2 , ao correlacionar as contagens de frutos nas imagens com a contagem total de frutos feita por humanos, em uma árvore, quando a somaram as contagens dos quatro lados da planta foi obtida, em comparação com as contagens das imagens de dois ou um lado.

4.1.4.3 Capacidade geral e Capacidade específica de Combinação, Herdabilidade e Heterose.

De acordo com os dados da ANOVA, o experimento realizado e avaliado, computacional e tradicionalmente, evidenciou a existência de diferença significativa, em nível de 1% de significância entre os genótipos, em relação a todos os caracteres avaliados, demonstrando a presença de variabilidade genética, ponto de partida para um programa de melhoramento. O conhecimento da natureza e da magnitude dos

efeitos genéticos que controlam determinado caráter apresentam grande importância na seleção e na predição do comportamento de indivíduos em populações segregantes e gerações híbridas (AHMAD LATIF *et al.*, 2021).

A avaliação da CGC permite a seleção dos melhores genitores. Já o efeito significativo da CEC indica a divergência genética entre genitores (TUNG *et al.*, 2018). O conhecimento sobre a significância de CGC e CEC é muito importante no melhoramento, pois orienta o melhorista sobre qual estratégia de melhoramento será mais efetiva para um determinado caractere (BADU-APRAKU *et al.*, 2015).

O maior coeficiente de variação (CV) encontrado para as variáveis estimadas pelas RCNN justifica-se pelo aumento do erro experimental, visto que, além de ser composto pelas variações ambientais no experimento, deve-se também ao erro de predição das redes.

Na teoria, os valores métricos, como os fenotípicos, resultam da interação entre fatores ambientais e genéticos em uma população. A variância fenotípica é uma medida observada e avaliada de acordo com a natureza de interesse. Parte dessa variância é atribuída à herdabilidade, que se manifesta devido às diferenças genéticas entre os indivíduos em uma população (AHMAD LATIF *et al.*, 2021).

A herdabilidade mede o grau de correspondência entre fenótipo e valor genético. As herdabilidades encontradas neste estudo variaram de 72,25% a 80,57%, já a relação CV_g/CV_e variou de 0,81 a 1,02. Alta variabilidade genética, altas estimativas de herdabilidade (h^2) e a relação CV_g/CV_e superior a um refletem uma situação favorável à seleção.

O ganho genético, dentro de um programa de melhoramento, pode ser acelerado de várias maneiras, como aumentar o tamanho dos experimentos de melhoramento para permitir maior intensidade de seleção, aumentar a precisão da seleção, garantir variação, acelerando os ciclos de reprodução e melhorando as ferramentas de apoio à decisão. Em todos esses componentes, a fenotipagem alto rendimento pode ser envolvida de forma direta ou indireta. A adoção desta fenotipagem em escala global poderá ser alcançada pelo melhorista, se for demonstrado como algo valioso em termos de ganhos genéticos (ARAUS *et al.*, 2018).

Observou-se boa correlação fenotípica entre a produtividades e os seis caracteres (GRÁFICO 2), também indicativos de produção, avaliados pela Mask-RCNN. Essas correlações indicam produtividade pelo uso de características das

plantas de fácil mensuração, pela análise computacional de imagens, como número de frutos maduros, verde e totais, os quais foram utilizados aqui.

A heterose é uma característica altamente relevante para o melhor aproveitamento de uma população. Em tomateiro se manifesta com o aumento do número de frutos por planta, peso médio do fruto e número de frutos por cacho, resultando em aumento de produtividade (SAH *et al.*, 2020). O desempenho *per se* dos híbridos F1, CxD e CxE excederam o valor da média de seus respectivos genitores na direção positiva para produtividade, NFV, NFT, PPM, PPV, PPT. A heterose implica um vigor híbrido superior ao parental, em características agronômicas, rendimento de grãos e outras características, em comparação aos seus progenitores. Isso ocorre pela ação gênica aditiva no controle genético das características (GABALLAH *et al.*, 2022). Para esses híbridos em relação a essas características, há também uma coincidência entre a seleção feita pela Mask-RCNN e pela fenotipagem tradicional.

Por meio da correlação de Spearman, observa-se uma relação forte, para CGC, entre a produtividade e o NFM e NFT, com valores próximos de 1 (FIGURA 5). Os ranks entre a produtividade de frutos de tomateiro e sem as variáveis estimadas pela Mask-RCNN mostraram um ordenamento similar entre a produtividade obtida pela fenotipagem tradicional e aquela obtida pela rede, para quase todas as variáveis estudadas, indicando concordância entre a produtividade estimada de forma tradicional e pela Mask-RCNN. O sinal positivo, para todos os caracteres avaliados, mostra que as duas avaliações variaram no mesmo sentido.

A CEC indica melhores duplas de genitores interessantes para serem utilizados para a obtenção de híbridos. O efeito da CEC é o desvio de híbrido, para melhor ou pior, em relação ao que seria esperado, com base na CGC dos pais, resultante dos efeitos de dominância, epistasia e vários tipos de interações gênicas (HAYMAN, 1954). Para a CEC, também, com um ordenamento similar entre as análises tradicionais e pela rede, indicando que se tende a selecionar os mesmos híbridos, considerando a produtividade real ou as análises pela rede. Demonstra a possibilidade em se deixar de fazer a fenotipagem tradicional e fazer apenas por imagem.

A Mask-RCNN detecta frutos e cria máscaras nos frutos, essa segmentação por pixel é mais precisa do que as caixas de detecção. Esse algoritmo é derivado do Faster RCNN e é atualmente o algoritmo de segmentação de instância mais

amplamente utilizado. Adicionando um novo ramo de máscara aos resultados de segmentação de saída, ele atende às necessidades na maioria das cenas (ZHAO *et al.*, 2023; CONG *et al.*, 2023).

A Convolutional Neural Network (CNN) demonstrou um notável êxito no domínio do reconhecimento de padrões, em imagens dos frutos do tomateiro, uma área também conhecida como "aprendizado profundo", por sua capacidade de interconectar camadas e extrair automaticamente parâmetros.

4.1.5 CONCLUSÃO

O modelo treinado foi eficiente, para a detecção de tomates com intensidade de luz variável nas imagens, sobreposição e oclusão. Houve similaridade nos parâmetros genéticos, considerando a análise tradicional e os resultados da análise computacional de imagens. Isso indica que a metodologia pode ser utilizada na fenotipagem de alto rendimento no melhoramento genético do tomateiro.

Este trabalho pode ser expandido, em outros estudos, para identificar produtividade por fruto com análise computacional de imagens em diferentes culturas. Além disso, reforça a capacidade de uso de drones e robôs na fenotipagem.

4.1.6 REFERENCIAS BIBLIOGRÁFICAS

AHMAD LATIF, N. H.; MOHD NAIN, F. N.; AHAMED, N. H. H. M.; ABDULLANH, R., ABDUL, M. F. R.; MOHAMAD, M.N.; MOHAMAD-FAUZI, N. S. Predicting Heritability of Oil Palm Breeding Using Phenotypic Traits and Machine Learning. **Sustainability**, v. 13, n. 22, p. 12613, 2021. DOI <https://doi.org/10.3390/su132212613>

ABD ALKREEM, M. H.; KARIM, A. A. Feature Selection Using Artificial Bee Colony for Fruit Classification. **Journal of Physics: Conference Series**. IOP Publishing, 2021. p. 012062. DOI <https://doi.org/10.1088/1742-6596/1818/1/0120>. 2021.

APOLO-APOLO, O. E.; MARTÍNEZ-GUANTER, J.; EGEA, G.; RAJA, P.; PEREZ-RUIZ, M. Deep learning techniques for estimation of the yield and size of citrus fruits using a UAV. **European Journal of Agronomy**, v. 115, p. 126030, 2020. DOI <https://doi.org/10.1016/j.eja.2020.126030>

ARAUS, J. L.; KEFAUVER, S. C.; ZAMAN-ALLAH, M.; OLSEN, M. S.; CAIRNS, J. E. Translating high-throughput phenotyping into genetic gain. **Trends in plant science**, v. 23, n. 5, p. 451-466, 2018. DOI <https://doi.org/10.1016/j.tplants.2018.02.001>

AZEVEDO, A. M.; ANDRADE JUNIOR, V. C. D.; PEDROSA, C. E.; OLIVEIRA, C. M. D.; DORNAS, M. F. S.; CRUZ, C. D.; VALADARES, N. R. Application of artificial neural networks in indirect selection: a case study on the breeding of lettuce. **Bragantia**, n. AHEAD, p. 00-00, 2015. DOI <https://doi.org/10.1590/1678-4499.0088>.

BADU-APRAKU, B.; FAKOREDE, M. A. B.; GEDIL, M.; TALABI, A. O.; ANNOR, B.; OYEKUNLE, M.; AKINWALE, R. O.; FASANMADE, T. Y.; AKAOGU, I. C.; ADEROUNMU, M. Heterotic responses among crosses of IITA and CIMMYT early white maize inbred lines under multiple stress environments. **Euphytica**, v. 206, p. 245-262, 2015. DOI <https://doi.org/10.1007/s10681-015-1506-0>

BARGOTI, S.; UNDERWOOD, J. P. Image segmentation for fruit detection and yield estimation in apple orchards. **Journal of Field Robotics**, v. 34, n. 6, p. 1039-1060, 2017. DOI <https://doi.org/10.1002/rob.21699>

BELOUZ, K.; NOURANI, A.; ZERENG, S.; BENCHEIKN, A. Prediction of greenhouse tomato yield using artificial neural networks combined with sensitivity analysis. **Scientia Horticulturae**, v. 293, p. 110666, 2022. DOI <https://doi.org/10.1016/j.scienta.2021.110666>

CHANDRA, A. L.; DESAI, S. V.; GUO, W.; BALASUBRAMANIAN, V. N. Computer vision with deep learning for plant phenotyping in agriculture: A survey. **arXiv preprint arXiv:2006.11391**, 2020. DOI <https://doi.org/10.34048/ACC.2020.1.F1>

CHU, P.; LI, Z., LAMMERS, K.; LU, R.; LIU, X. Deep learning-based apple detection using a suppression mask R-CNN. **Pattern Recognition Letters**, v. 147, p. 206-211, 2021. DOI <https://doi.org/10.1016/j.patrec.2021.04.022>

CONG, P.; LI, S.; ZHOU, J.; LV, K.; FENG, H. Research on instance segmentation algorithm of greenhouse sweet pepper detection based on improved mask RCNN. **Agronomy**, v. 13, n. 1, p. 196, 2023.

CRUZ, C. D. Genes: a software package for analysis in experimental statistics and quantitative genetics. **Acta Scientiarum. Agronomy**, v. 35, p. 271-276, 2013. DOI <https://doi.org/10.4025/actasciagron.v35i3.21251>

HAYMAN, B.I. The theory and analysis of diallel crosses. **Genetics**, Austin, v.39, p. 789-809, 1954. DOI <https://doi.org/10.1093/genetics/2F39.6.789>

HSIEH, K. W.; HUANG, B. Y.; HSIAO, K. Z.; TUAN, Y. H.; SHIH, F. P.; HSIEH, L. C.; CHEN, S.; YANG, I. C. Fruit maturity and location identification of beef tomato using R-CNN and binocular imaging technology. **Journal of Food Measurement and Characterization**, v. 15, n. 6, p. 5170-5180, 2021. DOI <https://doi.org/10.1007/s11694-021-01074-7>

HU, C.; LIU, X.; PAN, Z.; LI, P. Automatic detection of single ripe tomato on plant combining faster R-CNN and intuitionistic fuzzy set. **IEEE Access**, v. 7, p. 154683-154696, 2019. DOI <http://dx.doi.org/10.1109/ACCESS.2019.2949343>

HUANG, Y. P.; WANG, T. H.; BASANTA, H. Using fuzzy mask R-CNN model to automatically identify tomato ripeness. **IEEE Access**, v. 8, p. 207672-207682, 2020. DOI <http://dx.doi.org/10.1109/ACCESS.2020.3038184>

GABALLAH, M. M.; ATTIA, K. A.; GHONEIM, A. M.; KHAN, N.; EL-EZZ, A. F.; YANG, B.; XIAO, L.; IBRAHIM, E. I.; AI-DOSS, A. A. Assessment of genetic parameters and gene action associated with heterosis for enhancing yield characters in novel hybrid rice parental lines. **Plants**, v. 11, n. 3, p. 266, 2022. DOI <https://doi.org/10.3390/plants11030266>

GEBREMEDHIN, A.; BADENHORST, P. E.; WANG, J.; SPANGENBERG, G. C.; SMITH, K. F. Prospects for measurement of dry matter yield in forage breeding programs using sensor technologies. **Agronomy**, v. 9, n. 2, p. 65, 2019. DOI <https://doi.org/10.3390/agronomy9020065>.

JAHAN, M.; MAGHSOUDI, A.; ROKOUEI, M.; FARAJI-AROUGH, H. Prediction and optimization of slaughter weight in meat-type quails using artificial neural network modeling. **Poultry Science**, v. 99, n. 3, p. 1363-1368, 2020. DOI <https://doi.org/10.1016/j.psj.2019.10.072>.

KAMILARIS, A.; PRENAFETA-BOLDÚ, F. X. Deep learning in agriculture: A survey. **Computers and electronics in agriculture**, v. 147, p. 70-90, 2018. DOI <https://doi.org/10.1016/j.compag.2018.02.016>

LI, Y.; CAO, Z.; LU, H.; XIAO, Y. In-field cotton detection via region-based semantic image segmentation. **Computers and Electronics in Agriculture**, v. 127, p. 475-486, 2016. DOI <https://doi.org/10.1016/j.compag.2016.07.006>

LIU, X.; ZHAO, D.; WEIKUAN, J.; WEI, J.; YUEPING, S. A detection method for apple fruits based on color and shape features. **IEEE Access**, v. 7, p. 67923-67933, 2019. DOI <http://dx.doi.org/10.1109/ACCESS.2019.2918313>.

MAT, N. H. C.; BHUIYAN, M. A. R.; YAAKOB, Z.; WICKNESWARI, R. Full diallel crosses in superior parents of *Jatropha curcas* L. **Industrial crops and products**, v. 151, p. 112482, 2020. DOI <https://doi.org/10.1016/j.indcrop.2020.112482>

MATOS, R.; RESENDE, J. T. V. D.; ZEIST, A. R.; CORTE, L. E. D.; DA SILVA, P. R.; ZEFFA, D. M. Performance of the double-cross tomato hybrids from a partial diallel. **Ciência e Agrotecnologia**, v. 45, p. e027320, 2021. DOI <https://doi.org/10.1590/1413-7054202145027320>

MOCHIDA, K.; KODA, S.; INOUE, K.; HIRAYAMA, T.; TANAKA, S.; NISHII, R.; MELGANI, F. Computer vision-based phenotyping for improvement of plant productivity: a machine learning perspective. **GigaScience**, v. 8, n. 1, p. giy153, 2019. DOI <https://doi.org/10.1093/gigascience/giy153>.

PATTERSON, J.; GIBSON, A. **Deep learning: A practitioner's approach**. " O'Reilly Media, Inc.", 2017. 532 p.

PAYNE, A. B.; WALSH, K. B.; SUBEDI, P. P.; JARVIS, D. Estimation of mango crop yield using image analysis–segmentation method. **Computers and electronics in agriculture**, v. 91, p. 57-64, 2013. DOI <https://doi.org/10.1016/j.compag.2012.11.009>

RAHNEMOONFAR, M.; SHEPPARD, C. Real-time yield estimation based on deep learning. In: **Autonomous Air and Ground Sensing Systems for Agricultural Optimization and Phenotyping II**. SPIE, p. 59-65, 2017. DOI <http://dx.doi.org/10.1117/12.2263097>

RAZIFARD, H.; RAMOS, A.; DELLA VALLE, A.L; BODARY, C.; GOETZ, E., MANSER, E.J.; LI, X.; ZHANG, L.; VISA, S.; TIEMAN, D.; VAN DER KNAAP, E.; CAICEDO, A.L.; Genomic evidence for complex domestication history of the cultivated tomato in Latin America. **Molecular biology and evolution**, v. 37, n. 4, p. 1118-1132, 2020. DOI <https://doi.org/10.1093/molbev/msz297>

SAH, S. K.; SINGH, A. K.; SINGH, B. K.; BARMAN, K.; PAL, A. K.; KUMAR, A. Heterosis Studies for Growth and Yield Traits in Tomato (*Solanum lycopersicum* L.). **International Journal of Current Microbiology and Applied Sciences**, v. 9, n. 10, p. 2732-2738, 2020. DOI <https://doi.org/10.20546/ijcmas.2020.910.330>

SCHILLACI, G.; PENNISI, A.; FRANCO, F.; LONGO, D. Detecting tomato crops in greenhouses using a vision-based method. **Proceedings of International Conference on Safety, Health and Welfare in Agriculture and Agro**. Italy, p. 3-6, 2012.

SHARMA, P. Computer vision tutorial: Implementing mask R-CNN for image segmentation (with Python code). **Analytics Vidya**, 2019.

STRECK, E. A.; MAGALHAES, A. M.; AGUIAR, G. A., FACCHINELLO, P. K. H.; REIS, P. R., FRANCO, D. F.; NARDINO, M.; DE OLIVEIRA, A. C. Genetic progress in 45 years of irrigated rice breeding in Southern Brazil. **Crop Science**, v. 58, n. 3, p. 1094-1105, 2018.

DOI <https://doi-org.ez27.periodicos.capes.gov.br/10.2135/cropsci2017.06.0383>

THANAMMAL INDU, V.; SUJA PRIYADHARSINI, S. Crossover-based wind-driven optimized convolutional neural network model for tomato leaf disease classification. **Journal of Plant Diseases and Protection**, v. 129, n. 3, p. 559-578, 2022. DOI <https://doi.org/10.1007/s41348-021-00528-w>

TUNG, P. X.; VANDER ZAAG, P.; LI, C.; TANG, W. Combining ability for foliar resistance to late blight [*Phytophthora infestans* (Mont.) de Bary] of potato cultivars with different levels of resistance. **American Journal of Potato Research**, v. 95, p. 670-678, 2018. DOI <https://doi.org/10.1007/s12230-018-9673-x>

WAN, P.; TOUDESKI, A.; TAN, H.; EHSANI, R. A methodology for fresh tomato maturity detection using computer vision. **Computers and electronics in agriculture**, v. 146, p. 43-50, 2018. DOI <https://doi.org/10.1016/j.compag.2018.01.011>

WALTER, A; LIEBISCH, F; HUND, A. Plant phenotyping: from bean weighing to image analysis. **Plant methods**, v. 11, n. 1, p. 1-11, 2015. DOI <https://doi.org/10.1186/s13007-015-0056-8>

YAMAMOTO, K.; GUO, W.; YOSHIOKA, Y.; NINOMIYA, S. On plant detection of intact tomato fruits using image analysis and machine learning methods. **Sensors**, v. 14, n. 7, p. 12191-12206, 2014. DOI <https://doi.org/10.3390/s140712191>

ZHAO, Z.; HICKS, Y.; SUN, X.; LUO, C. Peach ripeness classification based on a new one-stage instance segmentation model. **Computers and Electronics in Agriculture**, v. 214, p. 108369, 2023. DOI <https://doi.org/10.1016/j.compag.2023.108369>

ZHOU, Y.; XU, T.; ZHENG, W.; DENG, H. Classification and recognition approaches of tomato main organs based on DCNN. **Transactions of the Chinese Society of Agricultural Engineering**, v. 33, n. 15, p. 219-226, 2017. DOI <http://dx.doi.org/10.11975/j.issn.1002-6819.2017.15.028>

4.2 CAPÍTULO II: FENOTIPAGEM AVANÇADA NA CLASSIFICAÇÃO DE FRUTOS DO TOMATEIRO POR INTELIGÊNCIA ARTIFICIAL

RESUMO

O tomate (*Solanum lycopersicum* L.) desempenha um papel crucial na agricultura global e serve como organismo- modelo em estudos genéticos. Com este trabalho, objetivou-se realizar a classificação de frutos do tomateiro, quanto ao formato, grupo, cor e defeitos, empregando Redes Neurais Convolucionais (CNNs). Avaliou-se o desempenho de cinco arquiteturas - VGG16, InceptionV3, ResNet50, EfficientNetB3 e InceptionResNetV2 – de forma a identificar e comparar a mais eficiente para essa classificação. A pesquisa foi conduzida considerando dez híbridos e suas cinco linhagens genitoras. O experimento foi conduzido em campo e a aquisição de imagens dos frutos maduros ocorreu em um mini- estúdio portátil. Utilizou-se o pacote ExlImage do software R, para individualização de fruto por imagem e auxílio, para a criação de um banco sintético, para treinamento das redes. As imagens foram agrupadas, conforme suas classificações quanto a formato, cor, grupos e defeitos. Mostrou-se com os resultados que a InceptionResNetV2 se destacou como a arquitetura mais eficiente, alcançando métricas, como precisão e recall superiores a 93%, para a maioria das variáveis analisadas, além de menores tempos de classificação. Este estudo não apenas avança na compreensão das aplicações de CNNs na agricultura e pesquisas, mas também fornece diretrizes valiosas para otimizar tarefas de classificação em diferentes tipos de frutos.

Palavras-chave: Análise de imagens. Deep learning. Melhoramento do tomateiro. Redes Neurais. *Solanum lycopersicum* L.

ADVANCED PHENOTYPING IN THE CLASSIFICATION OF TOMATO PLANT FRUITS BY ARTIFICIAL INTELLIGENCE

ABSTRACT

Tomato (*Solanum lycopersicum L.*) plays an important role in global agriculture and serves as a model organism in genetic studies. With this work, the objective was to classify tomato plant fruits, according to shape, group, color and defects, using Convolutional Neural Networks (CNNs). It was evaluated the performance of five architectures - VGG16, InceptionV3, ResNet50, EfficientNetB3 and InceptionResNetV2 - in order to identify and compare the most efficient one for this classification. The research was conducted considering ten hybrids and their five parental lines. The experiment was conducted in the field and images of ripe fruits occurred in a portable mini-studio. It was used the Expmage package of the R software for individualization of fruit by image and help, to create a synthetic database for training the networks. The images were grouped according to their classifications regarding format, color, groups and defects. It was showed with the results that InceptionResNetV2 stood out as the most efficient architecture, achieving metrics such as precision and recall above 93% for most of the variables analyzed, in addition to shorter classification times. This study not only advances the understanding of the applications of CNNs in agriculture and research, but also provides valuable guidelines for optimizing classification tasks on different types of fruits.

Keywords: Image analysis. Deep learning. Tomato plant improvement. Neural networks. *Solanum lycopersicum L.*

4.2.1 INTRODUÇÃO

O tomateiro (*Solanum lycopersicum* L.) é uma espécie nativa da América do Sul, sendo uma das hortaliças mais importantes, em termos de produção global, com uma produção aproximada de 186 milhões de toneladas em 2022 (FAOSTAT, 2023). Apresenta importância econômica, nutricional e, para as pesquisas biotecnológicas como organismo-modelo, uma vez que os resultados obtidos com o tomate também são aplicáveis a outras espécies hortícolas (HIWASA-TANASE, 2016). O grande consumo da cultura criou a necessidade de produzir variedades novas, de alto rendimento e com melhor qualidade nutricional (ANWAR *et al.*, 2019).

A classificação do fruto do tomateiro é crucial, em diversos contextos, incluindo a seleção de cultivares, padronização da produção, comercialização e pesquisa de melhoramento genético. Na comercialização, atributos como cor desempenham papel essencial. Esse fator, aliado ao tamanho, forma e aparência externa do fruto, influencia diretamente a escolha do consumidor (CAUSSE *et al.*, 2006). Na pesquisa, os melhoristas enfrentam o desafio de diversificar as variedades de tomateiro (NAGAMINE *et al.*, 2022).

Da domesticação até os dias atuais, o tomate passou por seleção para melhorar várias características, como formato, cor, tamanho e sabor (RAZIFARD *et al.*, 2020). Atualmente, os programas de melhoramento genético visam aprimorar a qualidade, aumentar a produção e desenvolver resistência a doenças. A diversidade genética no tomateiro cultivado é limitada, buscando-se novas fontes de variabilidade em bancos de germoplasma (BAI; LINDHOUT, 2007). Nos últimos 30 anos, novas cultivares e híbridos de tomateiro foram lançados, como o Santa Cruz e o Italiano, os quais apresentam características específicas (RAZIFARD *et al.*, 2020).

A classificação visual, para os programas de melhoramento do tomate, enfrenta desafios em função da diversidade das variedades, maturação desigual, danos externos e a subjetividade do avaliador. A utilização de recursos computacionais, como a fenotipagem por imagem, tem avançado, permitindo avaliações precisas e rápidas durante as fases pré e pós-colheita. Essas tecnologias oferecem vantagens, como baixo custo, rapidez, precisão e automação, reduzindo custos com mão de obra e eliminando a variabilidade devido à subjetividade humana na classificação (CHANDRA *et al.*, 2020).

A integração de inovações, como o uso de imagens em conjunto com a inteligência computacional, destaca-se como uma vertente promissora no processo de fenotipagem de plantas (SAMBASIVAM, OPIYO, 2021; CHANDRA *et al.*, 2020). Dentro do Deep learning, as Redes Neurais Convolucionais (CNNs) são o modelo mais amplamente reconhecido e empregado na resolução de tarefas relacionadas à visão computacional e à classificação de imagem. Esses algoritmos são capazes de identificar uma variedade de elementos visuais, incluindo rostos, indivíduos, sinais de rua, frutos, animais e diversos outros atributos presentes nos dados visuais (LI *et al.*, 2021).

Com isso, objetivou-se com esta pesquisa verificar a eficiência de uso das CNNs, para a automação da classificação de frutos em tomateiros, bem como comparar diferentes arquiteturas de CNN no desempenho dessa tarefa.

4.2.2 MATERIAL E MÉTODOS

4.2.2 .1 Instalação e avaliação do experimento

O experimento foi conduzido na fazenda experimental Professor Hamilton de Abreu Navarro (FEHAN) do Instituto de Ciências Agrárias (ICA)-UFMG, localizada no *Campus* Regional Montes Claros, (coordenadas: 16°40'58.16" S e 43°50'20.15"O). Foram avaliados cinco linhagens e suas 10 combinações híbridas F1s, originadas de um cruzamento em dialelos balanceados sem recíprocos de cinco linhagens comerciais, San Marzano e Santa Clara (empresa Isla), Santa Cruz Kada Gigante e Gaúcho Melhorado Nova Seleção (da empresa Feltrin) e Gaúcho (empresa Top Seed).

A semeadura ocorreu em bandejas de isopor de 128 células contendo substrato comercial para hortaliças (Bioplant) e conduzidas até o transplântio, 35 dias após a semeadura.

Após transplantadas para o campo, as plantas do tomateiro foram tutoradas, conduzidas em duas hastes. Cada haste foi manejada para a formação de seis a sete cachos de frutos. Não foram realizados desbastes de frutos na condução, o número de frutos variou, dessa forma, de acordo com as características dos genótipos. A adubação foi realizada de acordo com a análise de solo e as recomendações para a cultura do tomateiro (FURLANI; BATAGLIA, 2018). A irrigação, feita diariamente, por

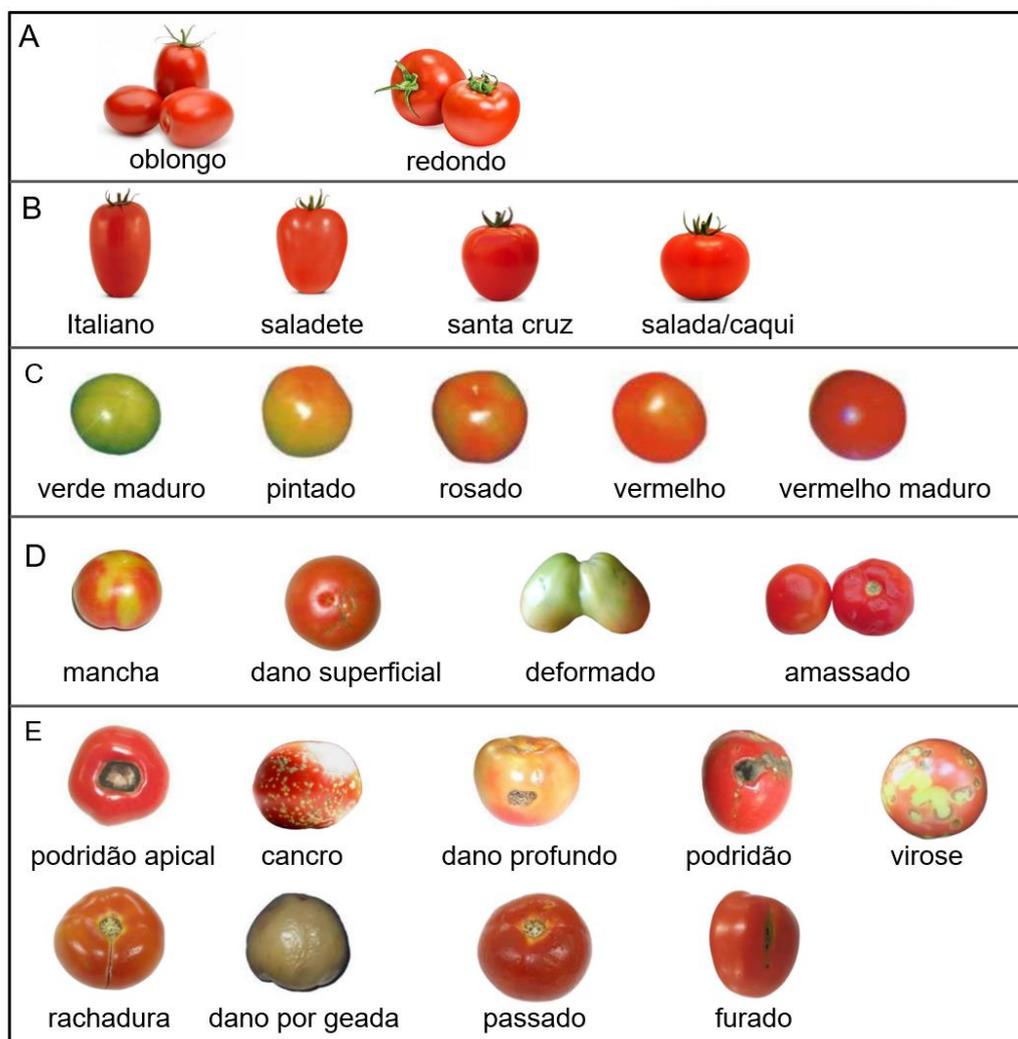
gotejamento. O controle fitossanitário, sempre que necessário, foi realizado com produtos registrados para a cultura.

O experimento foi conduzido, em delineamento em blocos casualizados (DBC) com quatro repetições, 15 tratamentos (dez híbridos e cinco linhagens) e cinco plantas por parcela, totalizando 300 plantas. Em 10 linhas de plantio (canteiros), com espaçamento entre plantas dentro das parcelas de 0,70m, 1m entre parcelas e entre linhas de 2 metros.

A colheita manual iniciou-se 90 dias após o transplântio. Os frutos do tomateiro foram avaliados durante doze colheitas. Os frutos, em estágio de maturação, na cor vermelho, eram colhidos duas vezes por semana.

Foram analisadas as variáveis quanto ao fruto do tomateiro: grupo (Italiano, Salada Caqui, Saladete e Santa Cruz), formato (Oblongo ou Redondo), cor (Pintado, Rosado, Verde Maduro, Vermelho e Vermelho Maduro) e defeitos (Graves, Leves e Sem defeitos). As avaliações foram realizadas, de acordo com os descritores e classificações da Portaria nº 553/95 do MAARA (BRASIL, 1995) e Anexo XVII da Portaria SARC nº 085/02 do MAPA (BRASIL, 2002), exemplificados na **Figura 1 (A, B, C, D e E)**.

Figura 1 -Classificação do fruto do tomateiro quanto ao: A- formato, B- Grupo, C- cor e D- Defeitos leves e E- Defeitos graves.



Fonte: Da autora, 2024, adaptado do BRASIL, 2002.

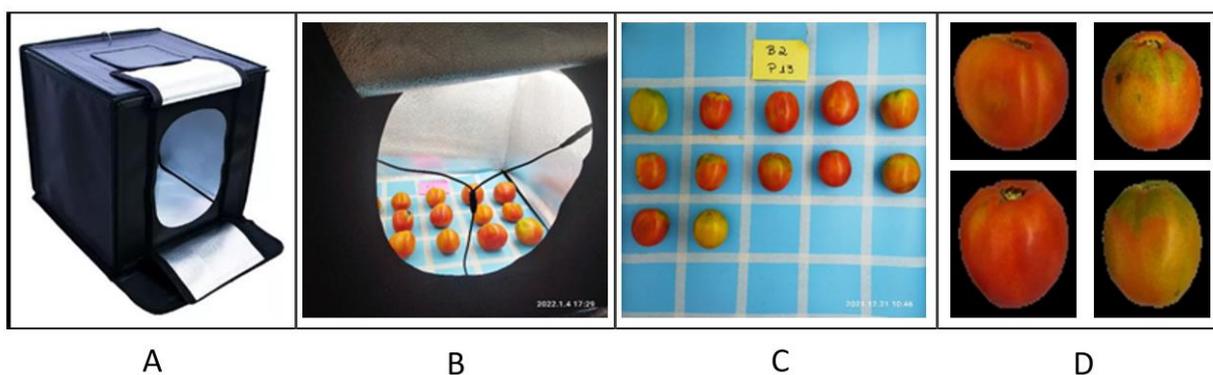
4.2.2 .2 Aquisição e processamento de imagens

A captura das imagens ocorreu em um mini-estúdio portátil com dimensões de 60 x 60 cm no fundo e altura de 60 cm (**Figura 2A**). Utilizou-se um smartphone Xiaomi modelo Poco M3, equipado com uma câmera de 48 megapixels capaz de capturar imagens com resolução de até 8000x6000 pixels e vídeos em alta definição (Full HD) de até 1920x1080 pixels. As imagens foram capturadas sob iluminação artificial, provida por lâmpadas fluorescentes. O smartphone foi acoplado a um suporte, na parte superior do mini- estúdio, com o objetivo de padronizar a obtenção das imagens, de forma que todas as imagens foram obtidas da mesma altura, 60 cm. Os frutos foram colocados sobre um fundo azul, quadriculado, de forma que cada fruto ficasse em um

quadrado (**Figura 2B**), até que 20 frutos fossem fotografados simultaneamente (**Figura 2C**).

O pacote ExplImage (AZEVEDO, 2024) do software R foi utilizado para individualização de um fruto por imagem (**Figura 2D**).

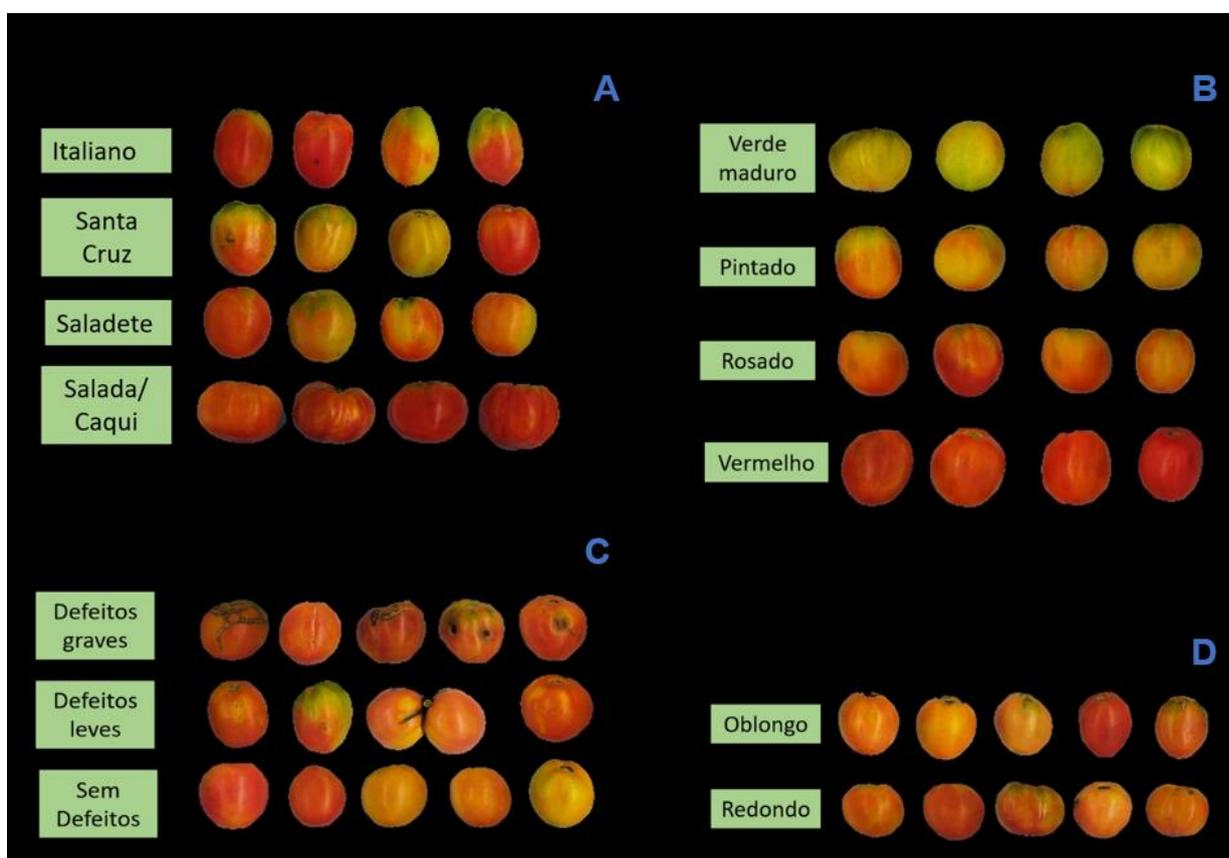
Figura 2 –Mini- estúdio para a aquisição das imagens com iluminação artificial com lâmpada fluorescente (A), posicionamento dos frutos dentro do estúdio (B), Imagem adquirida no estúdio (C) e imagem dos frutos individualizados pelo software R (D)



FONTE: Da autora, 2024.

As imagens foram agrupadas, de acordo com suas classificações, quanto ao formato (Italiano, Saladete, Salada e Santa Cruz), cor dos frutos (Vermelho, Vermelho maduro, Rosado, Pintado e Verde), grupos (Oblongo ou Redondo) e defeitos (Leves, Graves e Sem defeitos) por análise visual (**Figura 3**).

Figura 3- Exemplificação do agrupamento de imagens, em suas respectivas classificações por análise visual, utilizadas para criar as pastas de treinamento das CNNs. A - Grupo, B - Cor, C - Defeito e D - Formato.



Fonte: Da autora, 2024.

Após a obtenção das imagens de cada fruto maduro, realizou-se a classificação visual, obtendo-se os números apresentados na **Tabela 1** (Números originais). Com o desenvolvimento natural e sem restrição das plantas, o número de frutos para cada classificação se diferiu. Um número baixo de frutos foi encontrado para as classificações “Italiano” e “Salada Caqui” para a característica “grupo” e para as classificações “Verde maduro” e “Vermelho maduro” para a característica cor. Essas classificações foram desconsideradas neste trabalho, em razão desse baixo número de imagens, impossibilitando o treinamento das redes.

Para a composição do banco de imagens para treinamento e validação das redes, 1500 frutos foram considerados para cada classificação. A fim de ampliar o conjunto de dados no treinamento, cada uma dessas imagens foi replicada com três diferentes rotações (90°, 180° e 270°). Assim, para cada classificação, houve 4000 imagens no treinamento e 2000 na validação (1500 + 3x1500). As imagens destinadas ao banco de treinamento e validação foram selecionadas aleatoriamente. Para o

treinamento das redes, foi utilizada a linguagem python na plataforma do Google Colab. Utilizou-se a biblioteca Keras, considerando as arquiteturas VGG-16, Inception-v3, ResNet-50, InceptionResNetV2, EfficientNetB3. Considerou-se 100 o número máximo de iterações e early stopping com tolerância de 10 iterações e 0,001 de taxa de aprendizado.

Tabela 1-Número de imagens originais de frutos do tomateiro utilizadas para a fenotipagem do Formato, Grupo, Cor e Defeitos e número de imagens destinadas ao ajuste das RCNNs (Treinamento) e avaliação da qualidade do ajuste (Teste)

Característica	Classificação	Número originais	Fotos (originais + ampliadas)	Treinamento	Validação	Teste
Grupo	Italiano	479	-	-	-	-
	Salada Caqui	696	-	-	-	-
	Saladete	4667	1500 +1500x3	4000	2000	3167
	Santa Cruz	2736	1500 +1500x3	4000	2000	1236
Formato	Oblongo	3258	1500 +1500x3	4000	2000	1758
	Redondo	5416	1500 +1500x3	4000	2000	3916
Cor	Pintado	2085	1500 +1500x3	4000	2000	585
	Rosado	3128	1500 +1500x3	4000	2000	1628
	Vermelho	2102	1500 +1500x3	4000	2000	602
	Verde Maduro	594	-	-	-	-
	Vermelho Maduro	737	-	-	-	-
Defeito	Graves	2418	1500 +1500x3	4000	2000	918
	Leves	2625	1500 +1500x3	4000	2000	1125
	Sem defeitos	3766	1500 +1500x3	4000	2000	2266

*- Classificação não considerada no treinamento das redes pelo número pequeno de frutos.

Fonte: Da autora, 2024.

Neste trabalho, utilizaram-se redes neurais convolucionais para o treinamento e classificação das imagens. Cada classificador, então, determinou a probabilidade de uma instância pertencer a uma subclasse específica, selecionando a classe com a maior probabilidade como saída final.

Para avaliar os ajustes das redes convolucionais, foram construídas matrizes de confusão, tendo-se as classificações preditas pelas diferentes arquiteturas da rede em função das classificações visuais. Foram utilizadas as métricas Recall (Eq 1), Acurácia (Eq 2), Precisão (Eq 3), F-Measure (Eq 4) e Especificidade (Eq 5), para avaliar a eficiência da rede.

$$Recall = \frac{VP}{VP + FN} \quad (eq1)$$

$$Acurácia = \frac{VP + VN}{VP + VN + FP + FN} \quad (eq2)$$

$$Precisão = \frac{VP}{VP + FP} \quad (eq3)$$

$$F - Measure = \frac{2 * Precisão * Recall}{Precisão + Recall} \quad (eq4)$$

$$Especificidade = \frac{VN}{VN + FP} \quad (eq5)$$

Sendo:

- Verdadeiros Positivos (VP): Instâncias que foram corretamente previstas como pertencentes à classe positiva.
- Falsos Positivos (FP): Instâncias que foram erroneamente previstas como pertencentes à classe positiva quando, na verdade, pertencem à classe negativa.
- Verdadeiros Negativos (VN): Instâncias que foram corretamente previstas como pertencentes à classe negativa.
- Falsos Negativos (FN): Instâncias que foram erroneamente previstas como pertencentes à classe negativa quando, na verdade, pertencem à classe positiva

4.2.3 RESULTADOS

Cada arquitetura revelou uma quantidade distinta de épocas e tempo necessários para o treinamento (**Tabela 2**). Para o formato e grupo, a arquiteturas Inception-v3 apresentou os menores tempos de classificação, bem como menores números de épocas necessárias para o treinamento. Já para a classificação de cor e defeito, a arquitetura InceptionResNetV2 apresentou o menor tempo. Com exceção da variável grupo, a arquitetura EfficientNetB3 demandou o maior tempo e maior número de épocas, apresentando menor eficiência (**Tabela 2**).

Tabela 2: Número de épocas, tempo de treinamento e classificações realizadas por diferentes arquiteturas de CNNs para frutos do tomateiro quanto ao formato, grupos, cor e defeitos

Variáveis	Arquitetura	Épocas	Tempo (horas:min:seg)	Tempo médio/época (min)
Formato	VGG-16	80	2:01:11	1,51
	Inception-v3	14	1:17:36	5,54
	ResNet-50	88	1:33:16	1,05
	InceptionResNetV2	16	1:24:43	5,30
	EfficientNetB3	100	2:30:32	1,50
Grupos	VGG-16	100	2:04:14	1,25
	Inception-v3	16	0:18:05	1,13
	ResNet-50	78	1:32:43	1,19
	InceptionResNetV2	28	0:32:17	1,15
	EfficientNetB3	100	1:51:23	1,11
Cor	VGG-16	39	0:33:10	0,85
	Inception-v3	14	0:23:40	1,69
	ResNet-50	45	0:40:18	0,90
	InceptionResNetV2	17	0:22:59	1,35
	EfficientNetB3	42	0:50:31.	1,12
Defeito	VGG-16	45	0:42:27	0,94
	Inception-v3	17	0:22:25	1,32
	ResNet-50	63	1:22:48	1,31
	InceptionResNetV2	13	0:21:55	1,69
	EfficientNetB3	82	3:10:11	2,31

Fonte: Da autora, 2023

A arquitetura EfficientNetB3 teve maiores números de classificação verdadeiro positivo (VP) para grupos, mas não identificou nenhum verdadeiro negativo (VN) e obteve um alto valor de falsos positivos (FP) que são instâncias que foram erroneamente classificadas como positivas, mostrando uma ineficiência dessa arquitetura.

A arquitetura InceptionResNetV2 obteve a maior eficiência, maior acerto na classificação correta de instâncias positivas e negativas, altos valores de verdadeiro-positivo e verdadeiro-negativo para todas as variáveis avaliadas (**Tabela 3**). Os números de instâncias positivas (VP), casos em que o modelo acertou, ao reconhecer um padrão e as instâncias negativas (VN), o modelo acertou, ao rejeitar uma área que

não contém objeto de interesse, representam o sucesso da arquitetura na classificação.

Tabela 3- Classificação das arquiteturas quanto aos acertos e erros na identificação de frutos.

Descritores	Métricas	Arquiteturas das CNNs				
		VGG-16	Inception-v3	ResNet-50	InceptionResNet-V2	EfficientNet B3
Formato	VP	1033	1605	1378	1614	1528
	FN	487	3102	2731	414	0
	FP	698	126	353	117	3656
	VN	3169	554	925	3242	0
Grupos	VP	2267	428	2269	2687	2893
	FN	439	85	382	137	0
	FP	626	2465	624	206	1217
	VN	778	1132	835	1080	0
Cor (Pintado)	VP	396	521	463	500	458
	FN	163	38	96	59	101
	FP	312	164	232	0	159
	VN	1803	1951	1883	2115	1956
Cor (Rosado)	VP	820	1279	953	1528	1006
	FN	719	260	586	11	533
	FP	263	36	179	57	159
	VN	872	1099	956	1078	976
Cor (Vermelho)	VP	439	419	472	569	516
	FN	137	157	104	7	60
	FP	444	255	375	20	376
	VN	1654	1843	1723	2078	1722
Defeito (Grave)	VP	536	654	602	859	556
	FN	334	216	268	11	304
	FP	455	298	571	15	235
	VN	2757	2914	2641	3187	2977
Defeito (Leve)	VP	464	948	611	1057	686
	FN	608	124	461	15	386
	FP	650	250	890	26	851
	VN	2360	2760	2120	2974	2149
Defeito (Sem defeitos)	VP	1452	1883	1115	2115	1398
	FN	688	257	1025	15	742
	FP	525	49	293	0	346
	VN	1417	1893	1649	1942	1586

Fonte: Da autora, 2024

A eficácia das arquiteturas das redes neurais convolucionais, ao classificar os frutos do tomateiro quanto ao formato, grupos, cor e defeitos, foi avaliada, utilizando-se valores de precisão, recall, F-measure (F1), acurácia e especificidade (**Tabela 4**).

A arquitetura InceptionResNetV2 foi a que obteve maior sucesso com essas métricas. Maior precisão, melhor acurácia e especificidade para a maioria das variáveis analisadas, com valores acima de 84%. Essa arquitetura possibilitou precisão de 100%, para cor e para defeito e, para essa última variável, o valor de 100% ocorreu, para a maioria as métricas avaliadoras, ficando com 99% apenas para o recall.

A arquitetura EfficientNetB3 obteve um recall de 100% para formato e grupo, porém não teve bons valores para outras métricas, reforçando sua baixa eficiência na classificação. De forma geral, a arquitetura VGG-16 foi a que apresentou os piores resultados, com menores estimativas para os avaliadores da qualidade do ajuste.

Tabela 4-Avaliadores da qualidade do ajuste de RCNNS com diferentes arquiteturas, na classificação de frutos do tomateiro, quanto ao formato, grupos, cor e defeitos.

Descritor	Métricas de Avaliação	Arquiteturas de CNNs				
		VGG-16	Inception-v3	ResNet-50	InceptionResNet V2	EfficientNet B3
Formato	Precisão	0,60	0,93	0,80	0,93	0,29
	Recall	0,68	0,34	0,34	0,80	1,00
	F1	0,64	0,50	0,47	0,86	0,46
	Acurácia	0,78	0,40	0,43	0,90	0,29
	Espec.	0,82	0,81	0,72	0,97	0,00
Grupos	Precisão	0,78	0,15	0,78	0,93	0,70
	Recall	0,84	0,83	0,86	0,95	1,00
	F1	0,81	0,25	0,82	0,94	0,83
	Acurácia	0,74	0,38	0,76	0,92	0,70
	Espec.	0,55	0,31	0,57	0,84	0,00
Cor (Rosado)	Precisão	0,56	0,76	0,67	1,00	0,74
	Recall	0,71	0,93	0,83	0,89	0,82
	F1	0,63	0,84	0,74	0,94	0,78
	Acurácia	0,82	0,92	0,88	0,98	0,90
	Espec.	0,85	0,92	0,89	1,00	0,92
Cor (Pintado)	Precisão	0,76	0,97	0,84	0,96	0,86
	Recall	0,53	0,83	0,62	0,99	0,65
	F1	0,63	0,90	0,71	0,98	0,74
	Acurácia	0,63	0,89	0,71	0,97	0,74
	Espec.	0,77	0,97	0,84	0,95	0,86
Cor (Vermelho)	Precisão	0,50	0,62	0,56	0,97	0,58
	Recall	0,76	0,73	0,82	0,99	0,90
	F1	0,60	0,67	0,66	0,98	0,70
	Acurácia	0,78	0,85	0,82	0,99	0,84
	Espec.	0,79	0,88	0,82	0,99	0,82
Defeito (Grave)	Precisão	0,54	0,69	0,51	0,98	0,70
	Recall	0,62	0,75	0,69	0,99	0,65
	F1	0,58	0,72	0,59	0,99	0,67
	Acurácia	0,81	0,87	0,79	0,99	0,87
	Espec.	0,86	0,91	0,82	1,00	0,93
Defeito (Leve)	Precisão	0,42	0,79	0,41	0,98	0,45
	Recall	0,43	0,88	0,57	0,99	0,64
	F1	0,42	0,84	0,47	0,98	0,53
	Acurácia	0,69	0,91	0,67	0,99	0,70
	Espec.	0,78	0,92	0,70	0,99	0,72
Defeito (Sem defeito)	Precisão	0,73	0,97	0,79	1,00	0,80
	Recall	0,68	0,88	0,52	0,99	0,65
	F1	0,71	0,92	0,63	1,00	0,72
	Acurácia	0,70	0,93	0,68	1,00	0,73
	Espec.	0,73	0,97	0,85	1,00	0,82

Fonte: Da aurora, 2024

4.2.4 DISCUSSÃO

Os caracteres de aparência, como cor, textura, tamanho, forma e variedade de defeitos, constituem os atributos importantes de qualidade sensorial externa em frutos e vegetais. Os sistemas de visão computacional têm substituído amplamente a classificação visual e manual na avaliação da qualidade de produtos alimentícios e agrícolas (FRACAROLLI *et al.* 2021). Em comparação com os métodos tradicionais de visão computacional, as CNNs alcançam desempenhos superiores, apresentando tempos de inferência mais rápidos e taxas de detecção mais elevadas, como evidenciado por estudos anteriores (PAYMODE; MALODE, 2022). As redes aprendem, por meio de processos iterativos de ajustes nos pesos sinápticos, conhecidos como treinamento. O aprendizado efetivo ocorre, quando a rede neural alcança uma solução generalizada, para um problema específico (ZHANG *et al.*, 2020).

As diferentes arquiteturas testadas neste estudo mostraram que é possível detectar e classificar não apenas uma, mas várias classes diferentes do fruto do tomateiro com eficiência e rapidez. O treinamento das CNNs ocorre por meio de etapas chamadas de épocas ou ciclos. Cada uma dessas etapas representa o número de iterações do processo, durante as quais todos os dados de entrada de treinamento são aplicados à rede, com o objetivo de ajustá-la para reduzir o erro médio (GUIMARÃES *et al.*, 2008). No entanto, à medida que aumenta o número de iterações, durante o treinamento, as redes tendem a memorizar mais os dados, resultando no caráter não generalista do sistema, conhecido como "overfitting". Esse problema e desempenho surgem, à medida que os dados de treinamento diminuem. Seu desempenho melhora, à medida que aumenta o número de dados de treinamento (MAMAT *et al.*, 2023). Portanto é importante determinar um número ótimo de iterações, para os conjuntos de dados analisados, e isso pode ser realizado por meio da aplicação de uma estratégia conhecida como "early stopping" (FERNANDES *et al.*, 2023). Para este trabalho, de acordo com os resultados, o tamanho do conjunto de dados e o *early stopping* foram suficientes para ajustar todas as camadas das arquiteturas InceptionResNetV2 e de forma eficiente. Isso sugere que o modelo é robusto e generaliza bem para dados de teste.

Cada tipo de arquitetura, assim como cada variável utilizada, revelou diferentes números de épocas. Notavelmente, as arquiteturas Inception-v3 e InceptionResNetV2 demonstraram os menores números de épocas e, portanto os tempos mais curtos de classificação. Esses resultados apontam, para uma economia de tempo na obtenção de resultados (NI *et al.*, 2020), o que não necessariamente conduz a um bom ajuste. Por outro lado, as arquiteturas VGG-16, EfficientNetB3 e ResNet-50 demonstraram menor eficiência na classificação do conjunto de dados analisado, exigindo também o maior número de épocas e um tempo mais prolongado para a obtenção dos resultados.

O desempenho do modelo CNN para tarefas de classificação é geralmente avaliado, considerando vários critérios como Precisão, Recall, F1, Acurácia e Especificidade. Ao comparar a precisão das arquiteturas, a InceptionResNetV2 apresentou o melhor resultado. A escolha de uma arquitetura que permite a análise de forma mais eficiente qualifica características em grandes conjuntos de dados com pouca necessidade de mão de obra. Isso pode ajudar os melhoristas a avaliar experimentos de maneira mais eficaz, levando à identificação de potenciais de novas cultivares em menor espaço de tempo (HAQUE *et al.*, 2021).

No presente estudo, a arquitetura InceptionResNetV2 destaca-se, ao demonstrar maior precisão, melhor acurácia e especificidade, para a maioria das variáveis analisadas. A precisão, para essa rede, foi superior a 93% em todas as variáveis. A EfficientNetB3, embora tenha registrado maiores valores de verdadeiro-positivo para grupos, exibiu menor eficiência global e maior demanda computacional. Por outro lado, a VGG-16 apresentou menor eficácia, especialmente, ao detectar formato, cor e defeitos, sugerindo possíveis limitações nesses contextos.

Outro fator importante é que o desafio da identificação automática de defeitos e danos superficiais sempre representou uma dificuldade na classificação de produtos agrícolas (SUGAWARA *et al.*, 2018). Neste estudo, a arquitetura InceptionResNetV2 superou essa dificuldade, ao alcançar uma precisão e acurácia superiores a 98%, contribuindo assim para o sucesso na classificação dessa arquitetura. Essas altas precisões das arquiteturas, utilizadas na classificação dos frutos do tomateiro, podem levar a melhorar ainda mais o processo de tomada de decisão nas práticas agrícolas (VASCONEZ *et al.*, 2020). Estratégias voltadas para a aquisição de imagens e análise de dados, em ambientes agrícolas, têm o potencial de aprimorar as práticas

relacionadas ao melhoramento genético do tomateiro, incluindo contagem de frutos, estimativa de rendimento, detecção de agentes patogênicos e doenças, além da classificação da maturidade da cultura.

A classificação eficaz de frutos de tomateiro, por meio de análise computacional, destacando a superioridade da arquitetura InceptionResNetV2, evidencia a capacidade das redes neurais convolucionais em classificar os frutos com base em características específicas. A nitidez na coloração dos frutos exerce influência nas métricas, uma vez que regiões defeituosas podem apresentar padrões de cor distintos, aprimorando a precisão da classificação e fornecendo informações adicionais sobre a qualidade dos frutos (HAQUE *et al.*, 2020). Assim, a facilidade de distinção na matriz RGB da imagem, em relação aos objetos e suas partes, é um elemento crucial para o sucesso na detecção das classes avaliadas (JEONG *et al.*, 2020). Essas descobertas enfatizam a importância de selecionar uma arquitetura CNN apropriada, adaptada à tarefa específica de classificação. Os conhecimentos obtidos com este estudo não só avançam a nossa compreensão da classificação do tomate, mas também fornecem orientações valiosas, para otimizar as CNNs, em aplicações mais amplas de avaliação da qualidade de produtos agrícolas. Pesquisas futuras podem explorar estratégias de ajuste fino e abordagens de conjuntos para aumentar ainda mais a eficiência e a precisão dos modelos de classificação de tomate. Isso indica o potencial, para o desenvolvimento de aplicativos e dispositivos destinados à classificação de tomates, proporcionando apoio valioso aos melhoristas em diversas pesquisas.

4.2.5 CONCLUSÃO

Há diferenças entre as arquiteturas de Redes Neurais Convolucionais (CNNs), em relação ao tempo de treinamento, número de épocas e desempenho, nas diferentes variáveis de classificação.

A arquitetura InceptionResNetV2 destacou-se, ao classificar com precisão frutos de tomateiro, em relação ao formato, grupo, cor e defeitos, apresentando estimativas elevadas para os parâmetros de avaliação de qualidade.

A análise de imagens, combinada com aprendizado profundo, é uma alternativa valiosa, para aprimorar a qualidade dos frutos de tomateiro, reduzindo subjetividade

nas análises e diminuindo o tempo de fenotipagem da cultura, com uma fenotipagem mais precisa e eficiente de frutos de tomate. Isso pode prevenir com eficácia erros e aumentar a eficiência na utilização de dados relevantes para futuros trabalhos de melhoramento do tomate e de outras culturas.

4.2.6 REFERÊNCIAS

ANWAR, R.; FATIMA, T.; MATTOO, A. K. Tomatoes: A model crop of solanaceous plants. **Oxford Research Encyclopedia of Environmental Science**, p.1-50, 2019. DOI <https://doi.org/10.1093/acrefore/9780199389414.013.223>

AZEVEDO, Alcinei Místico. 2022. "ExplImage: Ferramenta para análise de imagens em experimentos. Disponível em: < <https://cran.r-project.org/web/packages/ExplImage/index.html>>.

BRASIL - Ministério da Agricultura, Pecuária e Abastecimento. Portaria SARC n° 085 de 06 de março de 2002. Propõe o Regulamento técnico de identidade e qualidade para classificação do tomate. Brasília, 06 mar. 2002.

CAUSSE, M.; DAMIDAUX, R.; ROUSSELLE, P. Traditional and enhanced breeding for quality traits in tomato. **Genetic improvement of Solanaceous crops**, v. 2, p. 153-192, 2006. DOI <https://doi.org/10.1201/b10744>

CHANDRA, A. L.; DESAI, S. V.; GUO, W.; BALASUBRAMANIAN, V. N. Computer vision with deep learning for plant phenotyping in agriculture: A survey. **arXiv preprint arXiv:2006.11391**, 2020. DOI <https://doi.org/10.34048/ACC.2020.1.F1>

CUI, J.; YANG, M.; SON, D.; PARK, S.; CHO, S. I. Estimation of tomato bruising by mechanical impact force using multivariate analysis. **HortScience**, v. 53, n. 9, p. 1352-1359, 2018. DOI <https://doi.org/10.21273/HORTSCI13066-18>

FAOSTAT- Food and Agriculture Organization of the United Nations. FAO. Disponível em: <https://www.fao.org/faostat/en/#data/QI>. Acesso em: 07 jan. 2024.

FRACAROLLI, J. A.; PAVARIN, F. F. A.; CASTRO, W.; BLASCO, J. Computer vision applied to food and agricultural products. **Revista Ciência Agronômica**, v. 51, 2021. DOI <https://doi.org/10.5935/1806-6690.2020008>

FERNANDES, A. C. G.; VALADARES, N. R, RODRIGUES, C. H. O.; ALVES, R. A; GUEDES, L. L. M.; ATHAYDE, A. L. M.; MISTICO, A. A. Convolutional neural networks in the qualitative improvement of sweet potato roots. **Scientific Reports**, v. 13, n. 1, p. 8429, 2023. DOI <https://doi.org/10.1038/s41598-023-34375-6>

GEBREMEDHIN, A.; BADENHORST, P. E.; WANG, J.; SPANGENBERG, G. C.; SMITH, K. F. Prospects for measurement of dry matter yield in forage breeding

programs using sensor technologies. **Agronomy**, v. 9, n. 2, p. 65, 2019. DOI <https://doi.org/10.3390/agronomy9020065>.

GUIMARÃES, A. M.; MATHIAS, I. M.; DIAS, A. H.; FERRARI, J. W.; JUNIOR, C. R. D. O. C. Módulo de validação cruzada para treinamento de redes neurais artificiais com algoritmos backpropagation e resilient propagation. **Publicatio UEPG: Ciências Exatas e da Terra, Agrárias e Engenharias**, v. 14, n. 01, p. 17-24, 2008. DOI <https://doi.org/10.5212/publicatio.v14i01.954>

JEONG, Y. S.; LEE, H. R.; BAEK, J. H.; KIM, K. H.; CHUNG, Y. S.; LEE, C. W. Deep learning-based rice seed segmentation for phynotyping. **Journal of the Korea Industrial Information Systems Research**, v.25, n.5, p.23-29, 2020. DOI <https://doi.org/10.9723/jksis.2020.25.5.023>

HAQUE, S.; LOBATON, E.; NELSON, N.; YENCHO, G. C.; PECOTA, K. V.; MIEROP, R.; KUDENOV, M. W.; BOYETTE, M.; WILLIAMS, C. M. Computer vision approach to characterize size and shape phenotypes of horticultural crops using high-throughput imagery. **Computers and Electronics in Agriculture**, v.182, 2021. DOI <https://doi.org/10.1016/j.compag.2021.106011>

HIWASA-TANASE, K. Fruit Ripening in Tomato and Its Modification by Molecular Breeding Techniques. **Functional Genomics and Biotechnology in Solanaceae and Cucurbitaceae Crops**, p. 155-174, 2016. DOI http://dx.doi.org/10.1007/978-3-662-48535-4_10

LI, Z.; LIU, F.; YANG, W.; PENG, S.; ZHOU, J. A survey of convolutional neural networks: analysis, applications, and prospects. **IEEE transactions on neural networks and learning systems**, 2021. DOI <https://doi.org/10.1109/tnnls.2021.3084827>

MAMAT, N.; OTHMAN, M. F.; ABDULGHAFOR, R; ALWAN, A. A.; GULZAR, Y. Enhancing image annotation technique of fruit classification using a deep learning approach. **Sustainability**, v. 15, n. 2, p. 901, 2023. DOI <https://doi.org/10.3390/su15020901>

NAGAMINE, A.; TAKAYAMA, M.; EZURA, H. Genetic improvement of tomato using gene editing technologies. **The Journal of Horticultural Science and Biotechnology**, v. 98, n. 1, p. 1-9, 2022. DOI <https://doi.org/10.1080/14620316.2022.2091489>

NI, X.; LI, C.; JIANG, H.; TAKEDA, F. Deep learning image segmentation and extraction of blueberry fruit traits associated with harvestability and yield. **Horticulture Research**, v7, n.1, p.1-14, 2020. DOI <https://doi.org/10.1038/s41438-020-0323-3>

PANTHEE, D. R.; CAO, C.; DEBENPORT, S. J.; RODRÍGUEZ, G. R.; LABATE, J. A.; ROBERTSON, L. D.; BREKSA III, A. P.; VAN DER KNAAP, E.; MCSPADDEN GARDENER, B. B. Magnitude of genotype× environment interactions affecting tomato fruit quality. **HortScience**, v. 47, n. 6, p. 721-726, 2012. DOI <https://doi.org/10.21273/HORTSCI.47.6.721>

PAPANDREOU, G.; CHEN, L. C.; MURPHY, K.; YUILLE, A. Weakly-and semi-supervised learning of a deep convolutional network for semantic image segmentation. **Proceedings of the IEEE international conference on computer vision**, p. 1742-1750, 2015. DOI <https://doi.org/10.48550/arXiv.1502.02734>

PAYMODE, A. S.; MALODE, V. B. Transfer learning for multi-crop leaf disease image classification using convolutional neural network VGG. **Artificial Intelligence in Agriculture**, v. 6, p. 23-33, 2022. DOI <https://doi.org/10.1016/j.aiaa.2021.12.002>

RAZIFARD, H.; RAMOS, A.; DELLA VALLE, A. L.; BODARY, C.; GOETZ, E.; MANSER, E. J.; LI, X.; ZHANG, L.; VISA, S.; TIEMAN, D.; VAN DER KNAAP, E.; CAICEDO, A. L. Genomic evidence for complex domestication history of the cultivated tomato in Latin America. **Molecular biology and evolution**, v. 37, n. 4, p. 1118-1132, 2020. DOI <https://doi.org/10.1093/molbev/msz297>

SAMBASIVAM, G.; OPIYO, G. D. A predictive application of machine learning in agriculture:cassava disease detection and classification with imbalanced dataset using networks convolutional neural networks. **Journal of Egyptian Informatics**, v.22, n.1, p.27-34, 2021. DOI <https://doi.org/10.1016/j.eij.2020.02.007>

SUGAWARA, T. Evaluation on Physiological Function and Development of Food Processing Technologies in Region Agricultural Products. **Journal of the japanese society for food science and technology-nippon shokuhin kagaku kogaku kaishi**, v. 65, n. 4, p. 163-169, 2018. DOI <http://dx.doi.org/10.3136/nskkk.65.163>

VASCONEZ, J. P.; DELPIANO, J.; VOUGIOUKAS, S.; CHEEIN, F. A. Comparison of convolutional neural networks in fruit detection and counting: A comprehensive evaluation. **Computers and Electronics in Agriculture**, v. 173, p. 105348, 2020. DOI <https://doi.org/10.1016/j.compag.2020.105348>

ZHANG, B.; HUANG, W.; Li, J.; ZHAO, C.; FAN, S.; WU, J.; LIU, C. Principles, developments and applications of computer vision for external quality inspection of fruits and vegetables: A review. **Food Research International**, v. 62, p. 326-343, 2014. DOI <http://dx.doi.org/10.1016/j.foodres.2014.03.012>

4.3 CAPÍTULO III: ANÁLISE DE IMAGENS NA ESTIMAÇÃO DE MASSA FRESCA DE TOMATES

RESUMO

Este estudo teve como objetivo desenvolver uma metodologia para fenotipagem dos frutos de tomateiro quanto à massa fresca, visando ao aprimoramento genético da cultura. Utilizando análise computacional de imagens, o desempenho dessa abordagem foi comparado com a metodologia tradicional de avaliação. Foram avaliadas 15 combinações híbridas, compreendendo 10 híbridos F1s de um cruzamento em dialelos balanceados e cinco linhagens (genitores). O experimento foi conduzido, em delineamento em blocos casualizados, com quatro repetições e cinco plantas por parcela, avaliando-se a massa dos frutos individualmente e em nível de genótipo. A aquisição de imagens foi realizada em um mini-estúdio portátil sob iluminação artificial. As análises foram conduzidas utilizando o pacote *ExpImage* do software R para a estimação da área dos frutos em pixels. Foi ajustado um modelo de regressão polinomial de primeiro grau, para a estimação da massa dos frutos, a partir da área em pixels, ocupada pelo fruto na imagem e outro, para validação do modelo entre a massa observada e a massa estimada. Os coeficientes de determinação (R^2) foram de 0,94 para a amostra de treinamento e de 0,91 para a validação. A correlação de Spearman entre os parâmetros genéticos estimados, para as massas dos frutos observados e massas preditas, apresentou coeficiente de correlação próximo e igual a 1,0. Os resultados indicaram que a análise de imagens realizou a predição precisa da massa dos frutos, mantendo o ranqueamento dos genótipos e, em consequência, preservando a similaridade entre os ganhos esperados com a seleção. Dessa forma, o uso de imagens revelou-se uma ferramenta eficaz nos programas de melhoramento genético dos frutos do tomateiro.

Palavras-chave: Fenotipagem. Análise computacional. *Solanum lycopersicum* L. Tomateiro. Fenômica.

CHAPTER III: IMAGE ANALYSIS IN ESTIMATION OF TOMATO FRESH MASS

ABSTRACT

This study aimed to develop a methodology for phenotyping tomato plant fruits (*Solanum lycopersicum* L.) in terms of fresh mass, aiming at genetic improvement of the crop. Using computational image analysis, the performance of this approach was compared with traditional evaluation methodology. They were evaluated 15 hybrid combinations, comprising 10 F1 hybrids from a cross in balanced diallels and five lines (parents). The experiment was conducted in a randomized block design, with four replications and five plants per plot, evaluating the mass of the fruits individually and at the genotype level. Image acquisition was performed in a portable mini-studio under artificial lighting. The analyzes were conducted using the ExplImage package of the R software to estimate the fruit area in pixels. It was adjusted a first-degree polynomial regression model to estimate the fruit mass, based on the area in pixels occupied by the fruit in the image and another model to validate the model between the observed mass and the estimated mass. The coefficients of determination (R^2) were 0.94 for the training sample and 0.91 for the validation. The Spearman correlation between the estimated genetic parameters, for the observed fruit masses and predicted masses, presented a correlation coefficient close to and equal to 1.0. The results indicated that the image analysis performed the accurate prediction of fruit mass, maintaining the ranking of the genotypes and, consequently, preserving the similarity between the expected gains with selection. Thus, the use of images proved to be an effective tool in genetic improvement programs for tomato plant fruits.

Keywords: Phenotyping. Computational analysis. *Solanum lycopersicum* L. Tomato plant. Phenomics.

4.3.1 INTRODUÇÃO

O tomate (*Solanum lycopersicum* L.), pertencente à família Solanaceae, destaca-se como um dos vegetais mais consumidos globalmente, com elevados níveis de vitaminas A e C e licopeno. Sua diversidade genética e importância comercial o tornam um modelo, para estudos de desenvolvimento de frutos, com resultados aplicáveis a outras espécies hortícolas. O elevado consumo da cultura impulsionou a necessidade de desenvolver variedades novas, de alto rendimento e melhor qualidade nutricional (SIERRA-OROZCO *et al.*, 2021).

Apesar do sequenciamento do genoma do tomate e da identificação de “loci” de características quantitativas, as técnicas tradicionais de melhoramento continuam predominantes, com a seleção de características dependendo da experiência do melhorista (ZHU *et al.*, 2018). A fenotipagem tradicional, embora seja prática, é demorada e suscetível a erros (COSTA *et al.*, 2018).

Nesse contexto, a necessidade de métodos quantitativos, para a avaliação de fenótipos de frutos de tomate, incluindo cor, tamanho e morfologia, é evidente (BHATTA *et al.*, 2021). O avanço do melhoramento genético integrado à análise computacional de imagens, baseada em dados bioinformáticos e agrônômicos, é essencial para impulsionar a eficiência do melhoramento de culturas (WALLACE *et al.*, 2018; WASHBURN *et al.*, 2019).

A análise computacional de imagens para fenotipagem apresenta vantagens, sendo rápida, não invasiva, precisa, econômica, sem resíduos, com boa resolução espacial e temporal. Ela permite explorar áreas extensas, oferece um grande número de repetições e a segurança do operador, possibilitando estudos comparativos e seleção fenotípica em larga escala (GEBREMEDHIN *et al.*, 2019).

Essas análises são baseadas em conceitos fundamentais, como a definição de uma imagem digital e a representação em escala de cinza ou RGB (sistema de cores aditivas formado pelos tons de Vermelho (Red), Verde (Green) e Azul (Blue)). A segmentação de imagem é importante, para selecionar e processar áreas específicas, sendo essencial para avançar na análise e extração de características das imagens (ZHU *et al.*, 2022).

A agricultura moderna depende fortemente de dados e informações, muitos dos quais são provenientes de imagens fotográficas. A tecnologia de processamento

digital de imagens é uma ferramenta valiosa para analisar e ampliar a análise de imagens agrícolas (BHARGAVA; BANSAL, 2021). À medida que a ciência da informação avança, técnicas como reconhecimento de padrões e processamento de imagens têm caminhado de maneira satisfatória e positiva no uso de novas tecnologias.

A incorporação de técnicas de fenotipagem de alto rendimento, em programas de melhoramento, pode aprimorar substancialmente os métodos de avaliação, para características agronômicas, proporcionando uma compreensão mais aprofundada de sua base genética, diversidade e das influências ambientais ao longo do ciclo de desenvolvimento da cultura (REYNOLDS *et al.*, 2020)

Nesse sentido, este trabalho teve como objetivo investigar a eficácia da análise de imagem, na estimação da massa fresca de frutos de tomates, visando aprimorar a precisão e a eficiência da avaliação desse caractere.

4.3.2 MATERIAL E MÉTODOS

4.3.2.1 Instalação e condução das plantas

Quinze genótipos de tomateiro, compostos por cinco linhagens: San Marzano e Santa Clara (da empresa Isla), Santa Cruz Kada gigante e Gaúcho Melhorado Nova Seleção (da empresa Feltrin) e Gaúcho (da empresa Top Seed) e 10 combinações híbridas F1s, resultantes de um cruzamento em dialelos balanceados sem recíprocos (**Tabela 1**), foram avaliadas para este estudo. Como o objetivo nesta pesquisa está apenas no processo de fenotipagem, não pretendendo comparar a superioridade agronômica entre essas linhagens, os denominamos aleatoriamente de A, B, C, D e E. Já os híbridos foram denominados como: AxB, AxC, AxD, AxE, BxC, BxD, BxE, CxD, CxE e DxE.

Tabela 1- Dialelo balanceado envolvendo os cinco progenitores, suas respectivas autofecundações e 10 híbridos.

PROGENITORES		Masculino				
		A	B	C	D	E
Feminino	A	AXA	AXB	AXC	AXD	AXE
	B		BXB	BXC	BXD	BXE
	C			CXC	CXD	CXE
	D				DXD	DXE
	E					EXE

Fonte: Da autora, 2024

A semeadura ocorreu em bandejas de isopor de 128 células contendo substrato comercial para hortaliças (Bioplant) e conduzidas até o transplântio. Após transplantadas para o campo, as plantas de tomateiro foram tutoradas, conduzidas em duas hastes. Cada haste foi manejada para formação de seis a sete cachos de frutos, não houve desbastes de fruto na condução, o número de frutos variou, dessa forma, de acordo com a performance dos genótipos. Adubação foi realizada de acordo com a análise de solo e recomendações para a cultura do tomateiro. As plantas foram irrigadas diariamente por gotejamento. O controle fitossanitário, sempre que necessário, foi realizado com produtos registrados para a cultura.

O experimento foi conduzido, em delineamento em blocos casualizados (DBC), com quatro repetições, 15 tratamentos (dez híbridos e cinco linhagens) e cinco plantas por parcela. Em 10 linhas de plantio (canteiros), com espaçamento entre plantas dentro parcelas de 0,70m e 1m entre parcelas e espaçamento entre linhas de 2 metros, totalizando 300 plantas.

A colheita manual foi iniciada 90 dias após o transplântio, estendendo-se por doze colheitas. Os frutos de tomateiro, em estágio de maturação vermelha, foram colhidos duas vezes por semana. Após a colheita, os frutos foram levados ao laboratório, onde foram submetidos à fenotipagem manual e à aquisição de imagens.

4.3.2.2 Avaliação

Duas avaliações foram realizadas: uma por meio de mensuração manual dos frutos (fenotipagem tradicional) e outra utilizando análise de imagens (fenotipagem por análise computacional).

4.3.2.2_1 Fenotipagem tradicional

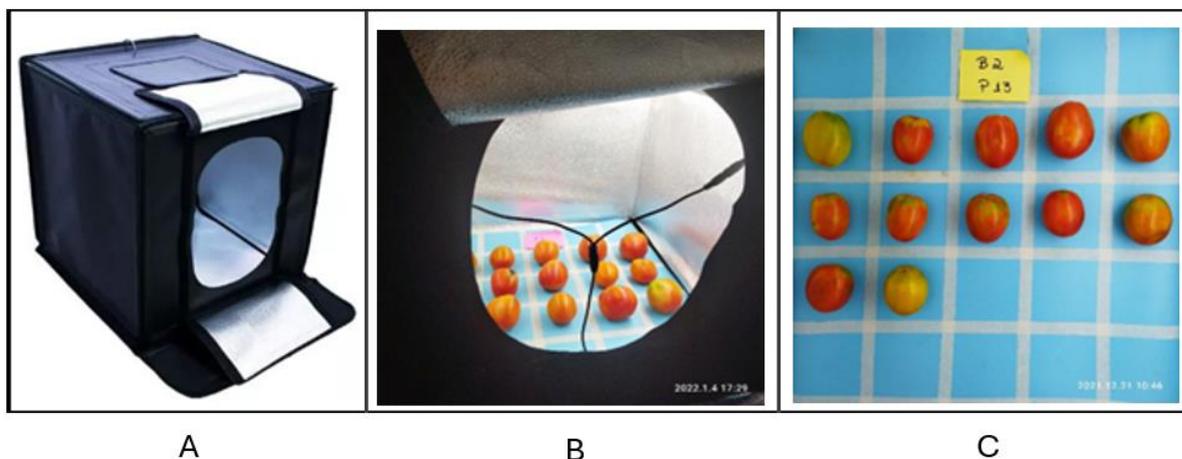
Todos os frutos colhidos dos 15 genótipos foram caracterizados individualmente quanto à massa fresca com o auxílio de uma balança de precisão de 0,001g.

4.3.2.2_2 Fenotipagem por análise de imagens

Um mini- estúdio portátil, medindo 60 x 60 x 60 cm, foi utilizado para adquirir as imagens (**Figura 1A**). Os frutos foram posicionados sobre um fundo azul quadriculado, de forma que cada fruto ficasse em um quadrado e até 20 frutos foram fotografados simultaneamente (**Figura 1B**). Utilizou-se um smartphone Xiaomi modelo Poco M3, equipado com uma câmera de 48 megapixels capaz de capturar imagens com resolução de até 8000x6000 pixels e vídeos em alta definição (Full HD) de até 1920x1080 pixels. As imagens foram capturadas sob iluminação artificial, provida por lâmpadas fluorescentes. O smartphone foi fixado em um suporte, localizado na parte superior do mini- estúdio, para padronizar a obtenção das imagens, garantindo que todas as imagens fossem capturadas a partir da mesma altura, 60cm.

O pacote ExpImage do software R (AZEVEDO, 2022) foi utilizado para individualização de um fruto por imagem.

Figura 1 -Mini-estúdio para a aquisição das imagens com iluminação artificial com lâmpada fluorescente (A), posicionamento dos frutos dentro do estúdio (B) e Imagem adquirida no estúdio (C). Para a automação da classificação por inteligência artificial.



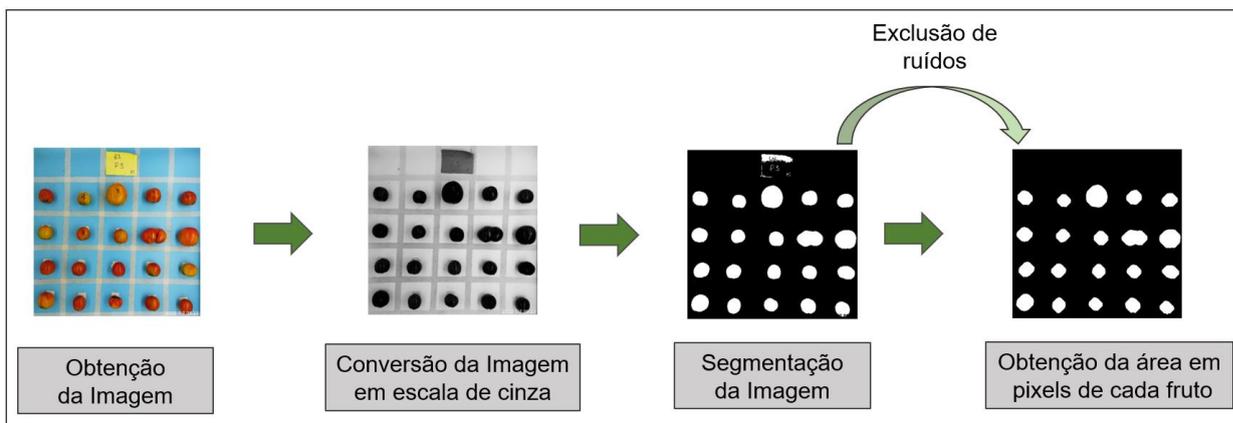
Fonte: Da autora, 2024.

No software R (R CORE TEAM, 2019), a análise das imagens foi conduzida utilizando o pacote *ExImage*. Essa rotina abrange a leitura da imagem, seu redimensionamento, conversão para escala de cinza, segmentação com base em um limiar e a medição da área em pixels para cada fruto (**Figura 2**).

As imagens originais foram redimensionadas para otimizar a análise computacional. As funções *read_image* e *resize_image* do pacote *ExImage* foram empregadas para redimensionar as imagens para uma dimensão de 480 x 480 pixels em largura e altura. Posteriormente, as imagens foram convertidas para escala de cinza, usando a função *gray_scale*, considerando a banda de azul, resultando em uma matriz com valores variando entre 0 e 1.

Para segmentar cada imagem e distinguir o fruto do fundo, um limiar (*threshold*) de 0,30 foi utilizado por meio da função *segmentation*. Pixels com valores acima desse limiar foram classificados como fundo (preto), atribuindo-se o valor 0, enquanto os demais pixels, correspondentes ao fruto, receberam o valor 1. A fim de excluir possíveis ruídos, foram realizadas seis erosões e, posteriormente, seis dilatações. Para isso, foram utilizadas as funções *erode_image* e *dilate_image* na matriz binária, respectivamente. Após a segmentação, a função *measure_image* foi empregada para obter a área em pixels para cada um dos frutos. A rotina utilizada para isso está disponível no link <https://github.com/AlcineiAzevedo/SandraPaper>.

Figura 2-Fluxograma que mostra as etapas, desde a preparação das imagens até a determinação da área de cada fruto em termos percentuais de pixels realizado pelo software R, pacote ExplImage



Fonte: Da autora, 2024.

Para a predição da massa fresca dos frutos, as amostras foram divididas de forma que todas as imagens referentes ao bloco 4 foram usadas para estimar o peso, a partir da área de pixels da imagem, por meio de regressão polinomial de primeiro grau. Já as demais imagens, pertencentes aos blocos 1, 2 e 3, foram utilizadas na validação do modelo e estimação de parâmetros genéticos.

Para verificar a qualidade do ajuste do modelo, estimou-se o coeficiente de determinação da regressão entre a massa fresca observada e predita. Além disso, foi estimado o erro quadrático médio, considerando os desvios das massas observadas e preditas para cada genótipo. A fim de verificar se há homogeneidade entre estes EQMs entre os genótipos foi feito o teste Hatley ($p \leq 0,05$).

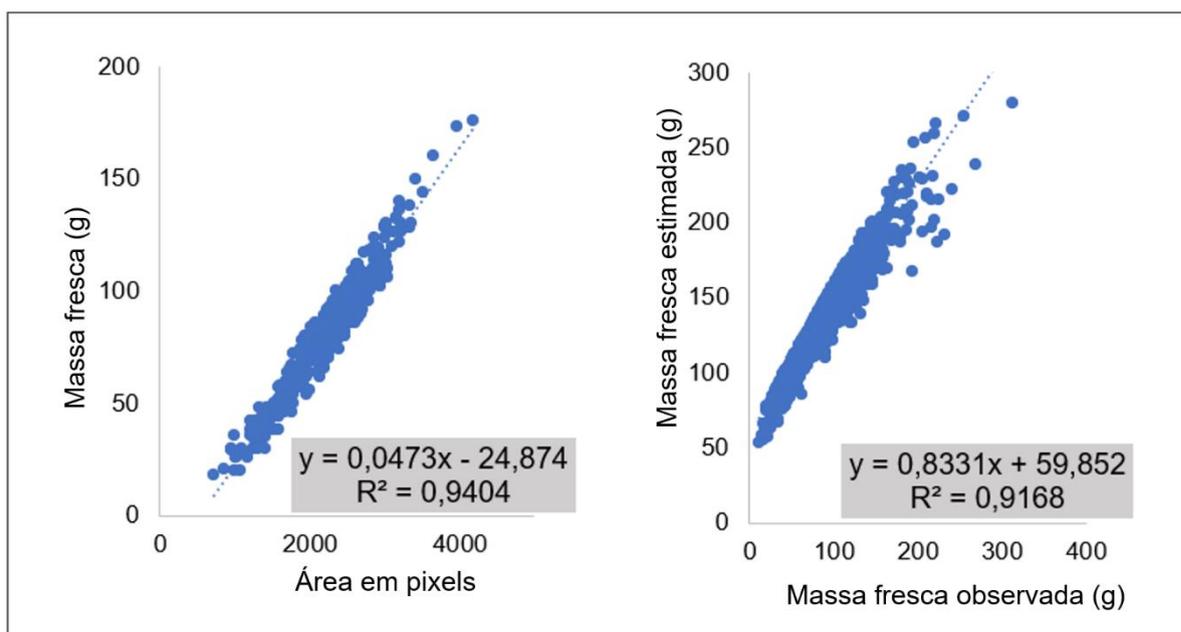
A análise genético-estatística da massa fresca observada e predita foi realizada utilizando o software GENES (CRUZ, 2013). Foram estimadas a capacidade geral de combinação (CGC), a capacidade específica de combinação (CEC), heterose e a herdabilidade.

Para verificar a eficiência do uso da análise de imagens, o ranqueamento da CGC, CEC e heterose obtidos a partir da massa fresca observada e estimada. Para isso, foi utilizada a correlação de Spearman com a função *cor* do pacote *stats* do software R.

4.3.3 RESULTADOS

Por meio do número de pixels, correspondente à área do fruto, ajustou-se um modelo, para a estimativa da massa fresca, para os 15 genótipos, por meio da análise de regressão. Obteve-se um alto coeficiente de determinação ($R^2=0,9404$) para o modelo, indicando a eficiência da metodologia desenvolvida, para prever a produtividade dos frutos do tomateiro com base em imagens **Figura 3A**). Ao analisar a precisão da metodologia, na amostra de validação, também foi observado um coeficiente de determinação elevado (0,9168) (**Figura 3B**).

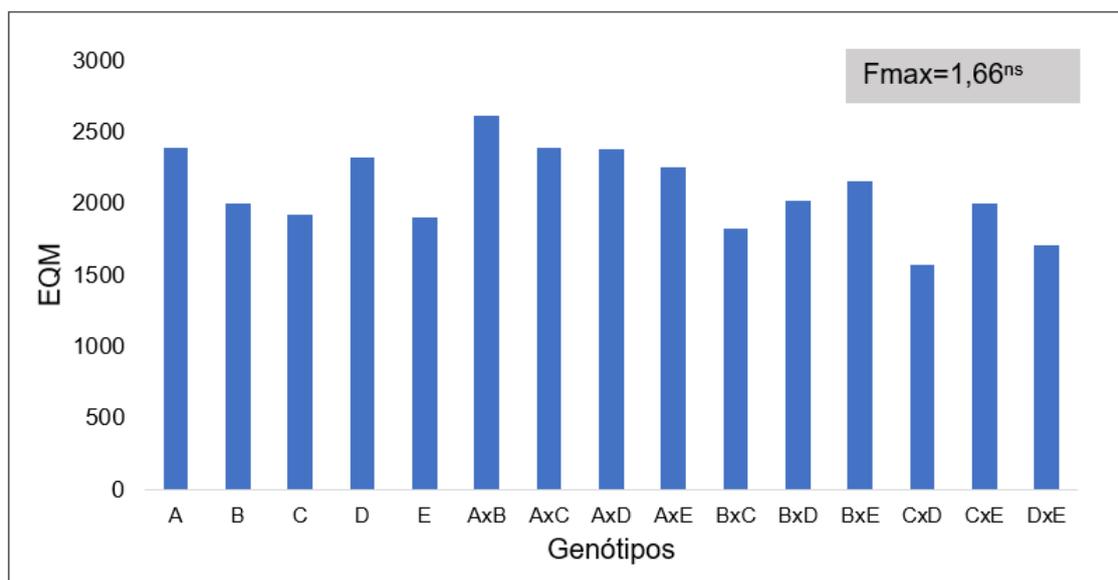
Figura 3- Análise de regressão, para a estimação da massa fresca do fruto, a partir da área em pixels, ocupada pelo fruto na imagem (A) na amostra de treinamento e avaliação da qualidade da predição para a amostra de validação (B).



Fonte: Da autora, 2022.

Na Figura 4, é apresentado o erro quadrático médio (EQM) para a estimativa da massa fresca. Valores próximos de EQM foram encontrados, para os 15 genótipos testados, sendo eles homogêneos de acordo com o teste de Hartley ($p>0,05$) (**Figura 4**).

Figura 4- Erro quadrático médio obtido, considerando o peso de frutos observado e estimado na amostra de validação, para cada um dos tratamentos e estimativa do teste Hartley



Fonte: Da autora, 2022.

Em relação aos parâmetros genéticos, as estimativas do quadrado médio (QM), para a capacidade geral de combinação (CGC), para os frutos do tomateiro, foram superiores aos valores de massa fresca observada (6375,13), em comparação com a massa fresca predita (4190,27), porém ambos foram significativos ao nível de 5% (**Tabela 2**). Esse comportamento também foi evidenciado, para a capacidade específica de combinação (CEC), cujo valor para a massa fresca observada foi de 1166,66, enquanto para a predita foi de 723,35.

Além disso, observou-se uma considerável variabilidade entre os genótipos, conforme indicado pelos altos coeficientes de variação genéticos (CVg), tanto para a massa fresca observada quanto para a predita. Também foram alcançados elevados e similares valores de herdabilidade (h^2a) para os genótipos estudados.

O coeficiente de variação residual (CV) foi de 13,13%, para a massa fresca observada e de 7,17% para a massa fresca predita dos frutos. Já a acurácia foi de 97%, para ambos os casos, indicando quão bem os valores preditos refletem os valores reais. (**Tabela 2**).

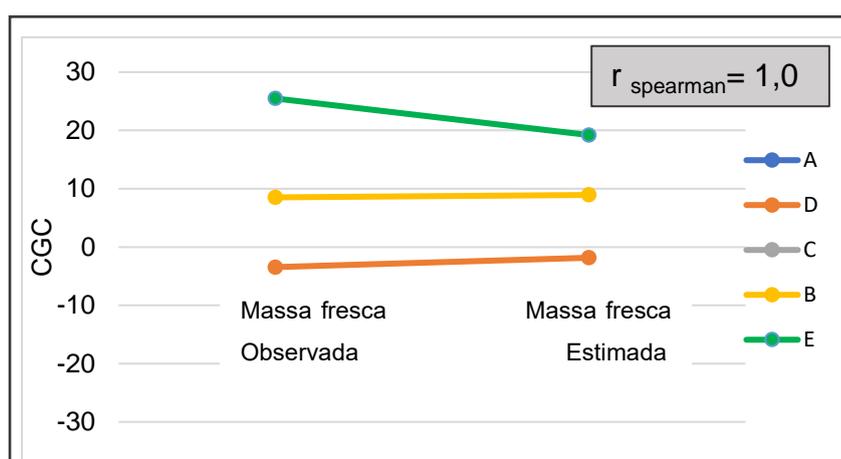
Tabela 2 - Parâmetros estimados, na avaliação de um experimento em dialelo, considerando a massa fresca de frutos observada e estimada, a partir da análise computacional de imagens

Parâmetros	Massa Fresca observada	Massa Fresca estimada
Capacidade geral de combinação (QM)	6375.13*	4190.27*
Capacidade específica de combinação (QM)	1166.66**	723.35**
Herdabilidade (%)	93.79	94.06
Correlação Intraclasse (US: parcela) - %	83.42	84.07
Coeficiente de Variação Residual (%)	13.13	7.17
Coeficiente de variação Genético(%)	29.45	16.47
Razão CVg/CVe	2.24	2.30
Acurácia	0.97	0.97

Fonte: Da autora, 2022.

A correlação de Spearman dos ranks entre a massa fresca observada dos frutos de tomateiro e a massa fresca estimada pela análise de imagem indicou uma semelhança no ordenamento dos genótipos para ambas as formas de fenotipagem (**Figura 5**). Essa correlação foi positiva e significativa ($r=1,00$) para a CGC, considerando as massas observadas e previstas, mostrando 100% de coincidência, no ordenamento dos genótipos. Ambas destacaram o progenitor e com a melhor CGC. Esse resultado evidencia a equivalência entre a técnica tradicional e a seleção, por meio da análise de imagens, para a seleção de genitores.

Figura 5-Correlação de Spearman e ordenamento de cinco genitores, para a capacidade geral de combinação, para a produtividade observada e estimada de frutos de tomateiro

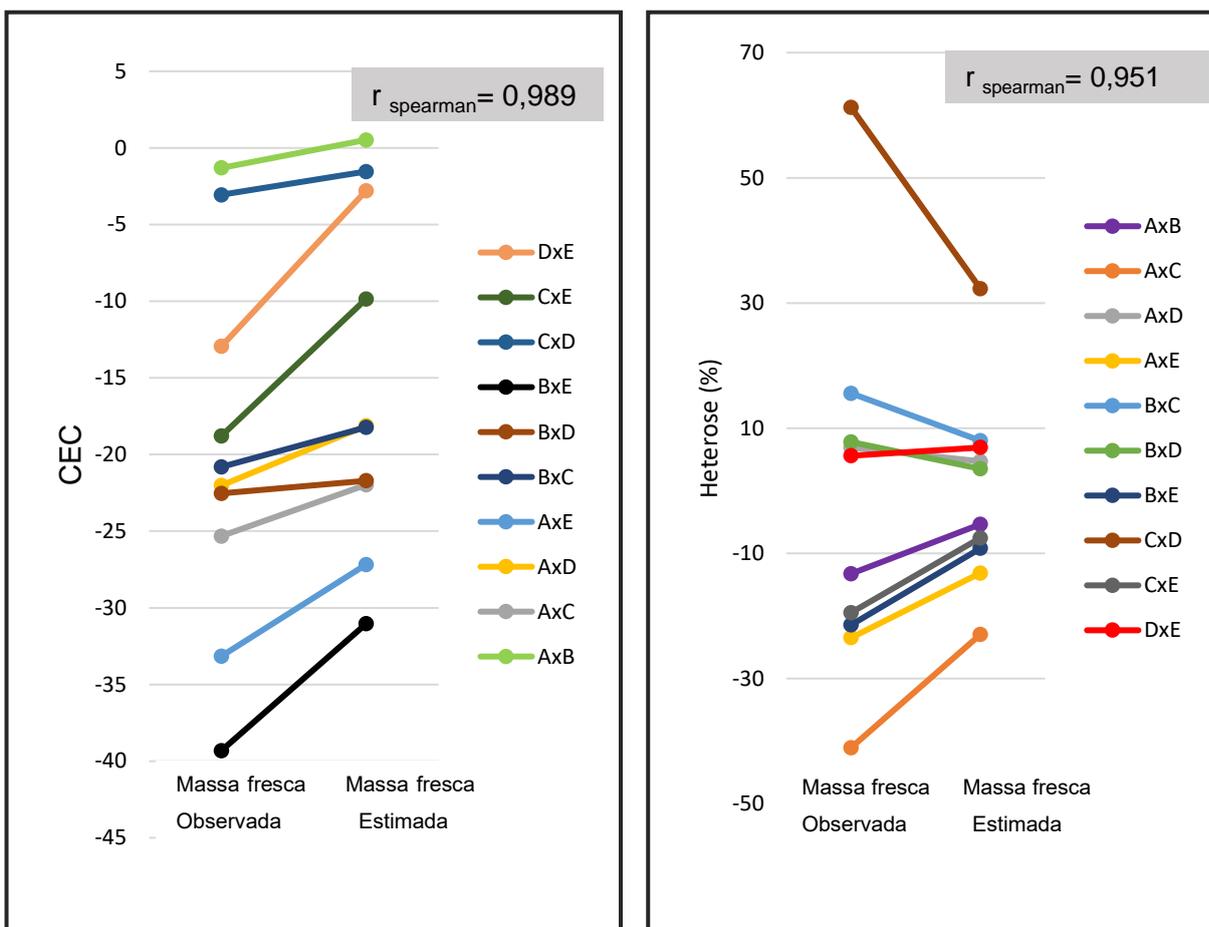


Fonte: Da autora, 2024.

A correlação de Spearman foi positiva e significativa para CEC ($r=0,989$) e para heterose ($r= 0,951$), considerando as massas frescas observadas e previstas (**Figura**

6). Observa-se uma boa similaridade entre o ranqueamento das melhores combinações híbridas selecionadas. Esse alto índice da correlação evidencia a equivalência entre a técnica tradicional e a seleção por meio da análise de imagens.

Figura 6- Correlação de Spearman e ordenamento das 10 combinações híbridas para a capacidade específica de combinação (A) e heterose (B) para a produtividade observada e estimada de frutos de tomateiro.



Fonte: Da autora, 2024.

4.3.4 DISCUSSÃO

Atualmente existe uma grande demanda por técnicas que aprimorem a área do melhoramento genético vegetal, visando aumentar a acurácia e precisão das avaliações experimentais, bem como automatizar as atividades dos programas de melhoramento (CHANDRA *et al.*, 2020). A aplicação dessas técnicas não só melhora as avaliações dos experimentos em campo e laboratório, mas também pode simplificar e reduzir os custos financeiros e de tempo associados aos programas de melhoramento. Uma ênfase particular é dada à detecção, mensuração e classificação de culturas, utilizando técnicas de aprendizagem profunda e processamento de imagens, o que se torna uma área de interesse relevante para os melhoristas (KHAN *et al.*, 2023). Ao adotar essa metodologia tecnológica, os programas de melhoramento fazem uso de métodos não destrutivos ou minimamente invasivos, incorporando conhecimentos e avanços de diversas áreas.

Os genótipos de tomateiro, submetidos a essa metodologia, revelaram que a análise de imagens proporcionou uma eficiente estimativa da massa fresca dos frutos. A capacidade preditiva baseia-se na premissa de que uma imagem é composta por pixels representando uma matriz de valores (PELT, SETHIAN, 2018). Conseqüentemente, a área ocupada por um fruto é quantificada por um número específico de pixels, que pode ser associado à sua massa. Sob essa perspectiva, é possível inferir que frutos de maiores dimensões apresentam massas superiores, concordando com as mensurações manuais e refletindo no elevado coeficiente de determinação obtido.

Quando se utilizou a análise de regressão, para comparar a massa fresca estimada com a área pixels ocupados pelo fruto, a metodologia demonstrou um desempenho notável. O coeficiente de determinação atingiu 0,9404, para a amostra de ajuste, indicando a viabilidade de estimar a massa do fruto, com base na área em pixels ocupada pela imagem, mostrando uma relação significativa entre essas variáveis. A habilidade de estimar com precisão a massa do fruto se mostra como uma ferramenta valiosa no processo de seleção de variedades. Para a amostra de validação, também, foi evidenciado um elevado coeficiente de determinação ($R^2=0,9168$), atestando a qualidade das predições da massa. A importância de validar o modelo em amostras independentes reforça a robustez e confiabilidade da metodologia.

Ao calcular-se o erro quadrático médio (EQM) entre as massas estimadas e observadas dos genótipos, conduziu-se o teste de Hartley para avaliar a homogeneidade das variâncias. Os resultados indicaram não significância estatística, sugerindo que os erros de predição foram similares, para os genótipos avaliados, isso fortalece a validade das inferências feitas com base no modelo de regressão.

Nas espécies autógamas, estuda-se o controle genético, usando populações segregantes, para estimar parâmetros como herdabilidade, número de genes, heterose, capacidade geral de combinação (CGC) e capacidade específica de combinação (CEC). Esses parâmetros fornecem informações sobre a herança, o tipo de genes envolvidos, a superioridade híbrida e as interações genéticas entre os pais (BALDISSERA *et al.*, 2013).

Os baixos coeficientes de variação residual (13,13 e 7,17) demonstraram boa precisão experimental e indicaram que as variações ambientais, incontroláveis, pouco afetaram os valores da massa dos frutos, tanto observada como predita.

A análise dialélica permitiu estimar os parâmetros genéticos, identificar as melhores combinações híbridas com base na avaliação das capacidades geral (CGC) e específica (CEC) de combinação. A análise de variância revelou a presença de variabilidade genética significativa, tanto aditiva quanto não aditiva, entre os genótipos avaliados, conforme evidenciado pelos quadrados médios da CGC e CEC. Para ambas as massas, observada e predita, as CGC foram superiores às CEC, indicando maior influência da ação gênica aditiva no controle do peso do fruto (CRUZ *et al.*, 2012). Obteve-se um resultado semelhante, para o carácter massa média dos frutos, quando estudado um dialelo em tomateiro (MATOS *et al.*, 2021), evidenciando, mais uma vez, a ação gênica aditiva na expressão fenotípica dessa característica.

Neste estudo, encontramos evidências da ação gênica aditiva pela CGC e uma herdabilidade alta para as duas estimações de massa. Assim, altas herdabilidades, conforme obtidas para os genótipos do tomateiro estudados, indicam que o fenótipo é mais influenciado por fatores genéticos que ambientais (MATOS *et al.*, 2021), tornando a análise mais confiável. A estimativa pela análise de imagem também obteve similaridade para esse parâmetro genético.

É conhecido que diversos elementos que contribuem positivamente na precisão das estimativas de herdabilidade estão sob o controle do melhorista, tais como a seleção da população, a escolha dos genitores e a determinação dos locais e número

de repetições nos ensaios de avaliação. Entretanto é importante destacar que existem fatores incontroláveis que podem influenciar a estimativa da herdabilidade, os quais estão relacionados ao genoma da planta (REYNOLDS *et al.*, 2020). O uso de análises computacionais pode aprimorar essa precisão e facilitar a obtenção dos parâmetros genéticos, bem como facilitar as avaliações (CHANDRA *et al.*, 2020).

O mérito da herdabilidade reside na sua capacidade de demonstrar de maneira quantitativa os efeitos genéticos expressos nos fenótipos, o que confere relevância ao valor genotípico, influenciando diretamente a próxima geração (FALCONER e MACKAY, 1996). Em um sentido restrito, a herdabilidade representa a proporção da variância fenotípica atribuída à variância genética aditiva (BALDISSERA *et al.*, 2013).

As estimativas dos efeitos da CGC revelaram a concentração de genes com efeitos aditivos. Assim, os parentais com estimativas elevadas e positivas de CGC impulsionam o aumento da expressão do traço, ao passo que aqueles com valores altos e negativos influenciam a redução de sua manifestação. Pela correlação de Serman, houve uma concordância de 100% no ordenamento dos melhores genótipos pela análise de imagem e a realizada de forma tradicional.

A CEC é interpretada como o desvio do híbrido em relação ao esperado com base CGC. Os valores dos efeitos da CEC foram na maioria negativos indicando a existência do desvio da dominância positiva e unidirecionais (CRUZ *et al.*, 2012). A adoção de cultivares híbridas tem sido utilizada para impulsionar a produção global de tomates (ROSA *et al.*, 2019). Na busca pela melhor combinação híbrida, é desejável encontrar aquela que apresenta a maior estimativa de valores absolutos de efeitos da CEC, resultante de um cruzamento em que pelo menos um dos genitores exiba uma CGC elevada. No entanto é importante destacar que, conforme observado por Cruz e Vencovsky (1989), dois genitores com CGC elevada nem sempre formarão a melhor combinação do dialelo. Neste estudo específico, os híbridos que exibiram os maiores valores absolutos dos efeitos da CEC (BxE e AxE) envolveram os genitores A e B, que apresentaram a maior CGC. As estimativas dos efeitos da CEC obtidas destacam a importância dos genes que manifestam efeitos não aditivos na determinação das características dos híbridos.

Além de analisar a CGC e a CEC, foi avaliada também a heterose. A magnitude do efeito da CEC no cruzamento está diretamente relacionada à intensidade da heterose do híbrido. A heterose ocorre, quando os híbridos exibem um desempenho

superior à média dos genitores, tanto nas gerações F1 quanto nas F2. Neste estudo, os progenitores B, C e D foram os que proporcionaram a maior heterose nos híbridos. No entanto, dado que os progenitores B e D apresentam os melhores desempenhos, em termos de efeito de CGC, seus híbridos são, sem dúvida, os mais promissores. Nesse ranqueamento dos melhores híbridos, tanto pela CEC como pela heterose, houve uma coincidência pela análise de imagem e a fenotipagem tradicional, tendo um alto coeficiente de correlação de Sperman acima de 95%.

O valor de um híbrido é tipicamente influenciado pelo efeito aditivo de cada alelo na característica, bem como por sua contribuição individual e pelas interações resultantes de dominância e epistasia. Esses dados desempenham um papel crucial na orientação dos processos seletivos, em programas de melhoramento genético, especialmente na seleção de genitores (LIU *et al.*, 2021). Houve concordância na estimativa desses parâmetros, para a análise das massas observadas e estimadas pela análise de imagens, comprovando a alta performance da metodologia para prever as massas frescas dos frutos do tomateiro, que pode ser confirmado com a alta correlação de Sperman (FIGURA 3 e 4). Elas (as massas) indicam que a precisão da medição do reconhecimento de imagem é equivalente à da medição manual. A obtenção de uma segmentação precisa do objeto de interesse é um passo importante na resolução dos desafios associados à classificação, reconhecimento e detecção de frutas e vegetais (VITE-CHÁVEZ *et al.*, 2023). Nesse contexto, o pacote *ExplImage* no software R, destacou-se pela eficiência na realização dessa tarefa, possibilitando de forma precisa a estimativa da área dos frutos, para posterior ajuste do modelo de regressão, para a estimação da massa fresca, apresentando também a vantagem de ser uma metodologia simples de executar e de fácil acesso.

Algumas técnicas de análise de imagens, para estimar a massa fresca de frutos, têm sido utilizadas, como a tomografia computadorizada (TC) de raios X, em frutos de romã (ARENDSE *et al.*, 2016), perceptrons multicamadas multiescala (MLP) e redes neurais convolucionais (CNNs) em maçãs (BARGOTI *et al.*, 2017), CNNs em mirtilos (NI *et al.*, 2020), combinação de diferentes métodos de extração de características da imagem com Mask-RCNN em tomates (ZHU *et al.*, 2022), redes neurais artificiais em frutos e vegetais assimétricos (HUYNH *et al.*, 2020) e análise de regressão, na predição de peso de batata-doce com o número de pixels da imagem (FERNANDES *et al.*, 2022).

Nas avaliações das pesquisas, a classificação dos frutos quanto ao formato e à mensuração da massa fresca, em geral, ocorre manualmente ou por equipamento. Por equipamento, um dispositivo com esteira transportadora é utilizado, para alimentar e alinhar os frutos, normalmente classificando-os por diâmetro. A metodologia desenvolvida por análise de imagens, destacando sua eficiência, pode ser integrada a essas esteiras, simplificando a mensuração dos frutos e essa estratégia extrapolada para outras culturas com frutos. As imagens são armazenadas, em um banco de dados, para consultas e comparações futuras, possibilitando o processamento simultâneo de várias delas e a fenotipagem de características qualitativas e quantitativas.

Este estudo foi formulado, a partir da hipótese de que a análise de imagens pode ser uma ferramenta promissora, para prever a massa fresca dos frutos de tomates, proporcionando uma abordagem inovadora e não invasiva na fenotipagem desses frutos. Ao analisar os resultados obtidos, confirmamos a eficiência dessa metodologia e, a partir dos dados, ressaltamos que avanços significativos poderão ser alcançados, no desenvolvimento industrial de equipamentos com esteiras e câmeras, a fim de reduzir o gasto de tempo, mão de obra e recursos financeiros em programas de melhoramento de larga escala.

4.3.5 CONCLUSÃO

A aplicação da visão computacional é eficiente na predição da massa fresca dos frutos do tomateiro, demonstrando consistência no ranqueamento dos diferentes genótipos para a CEC, CGC e heterose. Isso indica correspondência na tomada de decisões associadas à seleção de genitores e de combinações híbridas em programas de melhoramento. Este trabalho indica que a visão computacional pode ser utilizada no desenvolvimento de equipamentos, a fim de tornar a fenotipagem mais rápida, eficiente e menos dependente de mão de obra e recursos financeiros.

4.3.6 REFERÊNCIAS

ARENDSE, E.; FAWOLE, O. A.; MAGWAZA, L. S.; OPARA, U. L. Non-destructive characterization and volume estimation of pomegranate fruit external and internal morphological fractions using X-ray computed tomography. **Journal of Food Engineering**, v. 186, p. 42-49, 2016. DOI <https://doi.org/10.1016/j.jfoodeng.2016.04.011>

AZEVEDO, Alcinei Místico. 2022. "ExplImage: Ferramenta para análise de imagens em experimentos. Disponível em: < <https://cran.r-project.org/web/packages/ExplImage/index.html>>.

BALDISSERA, J. D. C.; VALENTINI, G.; COAN, M. M. D.; GUIDOLIN, A. F.; COIMBRA, J. L. M. Genetics factors related with the inheritance in autogamous plant populations. **Revista de Ciências Agroveterinárias**, v. 13, n. 2, p. 181-189, 2014. Disponível em: <https://www.revistas.udesc.br/index.php/agroveterinaria/article/view/5632>. Acesso em: 28 jan. 2024.

BARGOTI, S.; UNDERWOOD, J. P.; Image segmentation for fruit detection and yield estimation in apple orchards. **Journal of Field Robotics**, v. 34, n. 6, p. 1039-1060, 2017. DOI: <https://doi.org/10.1002/rob.21699>

BHARGAVA, A.; BANSAL, A.; Fruits and vegetables quality evaluation using computer vision: A review. **Journal of King Saud University-Computer and Information Sciences**, v. 33, n. 3, p. 243-257, 2021. DOI <https://doi.org/10.1016/j.jksuci.2018.06.002>

BHATTA, M.; SANDRO, P.; SMITH, M. R.; DELANEY, O.; VOSS-FELS, K. P.; GUTIERREZ, L.; HICKEY, L. T. Need for speed: manipulating plant growth to accelerate breeding cycles. **Current Opinion in Plant Biology**, v. 60, p. 101986, 2021. DOI [10.1016/j.pbi.2020.101986](https://doi.org/10.1016/j.pbi.2020.101986)

BORÉM, A.; MIRANDA, G. V.; FRITSCHÉ-NETO, R. **Melhoramento de plantas**. 7. Ed. Viçosa: Ed. UFV, 2017, 543 p.

COSTA, C.; SCHURR, U.; LORETO, F.; MENESATTI, P.; CARPENTIER, S. Plant phenotyping research trends, a science mapping approach. **Frontiers in plant science**, v. 9, p. 1933, 2019. DOI [10.3389/fpls.2018.01933](https://doi.org/10.3389/fpls.2018.01933)

CHANDRA, A. L.; DESAI, S. V.; GUO, W.; BALASUBRAMANIAN, V. N. Computer vision with deep learning for plant phenotyping in agriculture: A survey. **arXiv preprint arXiv:2006.11391**, 2020. DOI <https://doi.org/10.48550/arXiv.2006.11391>

CRUZ, C. D. Genes: a software package for analysis in experimental statistics and quantitative genetics. **Acta Scientiarum. Agronomy**, v. 35, p. 271-276, 2013. DOI <https://doi.org/10.4025/actasciagron.v35i3.21251>

CRUZ, C.D.; REGAZZI, A.J.; CARNEIRO, P.C.S. **Modelos Biométricos Aplicados ao Melhoramento Genético**. 4. ed, v. 1, Viçosa: Editora UFV, 2012. 514p

CRUZ, C. D.; VENCOVSKY, R. Comparação de alguns métodos de análise dialélica. **Revista Brasileira de Genética= Brazilian Journal of Genetics**, v. 12, n. 2, p. 425-38, 1989.

FALCONER, D.S., MACKAY, T.F.C. **Introduction to quantitative genetics**. 4.ed. England: Longman, 1996. 463p.

FERNANDES, A. C. G.; VALADARES, N. R.; RODRIGUES, C. H. O.; ALVES, R. A.; AZEVEDO, A. M.; GUEDES, L. L. M.; DE PAULA, L. S.G. Feasibility of computational vision in the genetic improvement of the production of sweet potato roots. **Horticultura Brasileira**, v. 40, n. 4, 2022.

GEBREMEDHIN, A.; BADENHORST, P. E.; WANG, J.; SPANGENBERG, G. C.; SMITH, K. F. Prospects for measurement of dry matter yield in forage breeding programs using sensor technologies. **Agronomy**, v. 9, n. 2, p. 65, 2019. DOI <https://doi.org/10.3390/agronomy9020065>

GRIFFING, B. Concept of general and specific combining ability in relation to diallel crossing systems. **Australian journal of biological sciences**, v. 9, n. 4, p. 463-493, 1956.

HUYNH, T.; TRAN, L.; DAO, S. Real-time size and mass estimation of slender axisymmetric fruit/vegetable using a single top view image. **Sensors**, v. 20, n. 18, p. 5406, 2020. DOI <https://doi.org/10.3390/s20185406>

KHAN, A.; HASSAN, T.; SHAFAY, M.; FAHMY, I.; WERGHI, N.; MUNDIGANSALAGE, S.; HUSSAIN, I. Tomato maturity recognition with convolutional transformers. **Scientific Reports**, v. 13, n. 1, p. 22885, 2023. DOI <https://doi.org/10.1038/s41598-023-50129-w>

LIU, Z.; JIANG, J.; REN, A.; XU, X.; ZHANG, H.; ZHAO, T.; JIANG, X.; SUN, W.; LI, J.; YANG, H. Heterosis and combining ability analysis of fruit yield, early maturity, and quality in tomato. **Agronomy**, v. 11, n. 4, p. 807, 2021. DOI <https://doi.org/10.3390/agronomy11040807>

MATOS, R. D.; RESENDE, J. T. V. D.; ZEIST, A. R.; CORTE, L. E. D.; DA-SILVA, P. R.; ZEFFA, D. M. Performance of the double-cross tomato hybrids from a partial diallel. **Ciência e Agrotecnologia**, v. 45, p. e027320, 2021. DOI <https://doi.org/10.1590/1413-7054202145027320>

NI, X.; LI, C.; JIANG, H.; TAKEDA, F. Deep learning image segmentation and extraction of blueberry fruit traits associated with harvestability and yield. **Horticulture research**, v. 7, 2020. DOI <https://doi.org/10.1038/s41438-020-0323-3>

PELT, D. M.; SETHIAN, J. A. A mixed-scale dense convolutional neural network for image analysis. **Proceedings of the National Academy of Sciences**, v. 115, n. 2, p. 254-259, 2018. DOI <https://doi.org/10.1073/pnas.1715832114>

R CORE TEAM. R: A Language and Environment for Statistical Computing. Vienna, Austria: R Foundation for Statistical Computing, 2019. Disponível em: <https://www.R-project.org/>.

REYNOLDS, M.; CHAPMAN, S.; CRESPO-HERRERA, L.; MOLERO, G.; MONDAL, S.; PEQUENO, D. N. L.; PINTO, F.; PINERA-CHAVEZ, F. J. ; POLAND, J. ; RIVERA-AMADO, C. ; SAINT PIERRE, C. ; SUKUMARAN, S. 2020). Breeder friendly phenotyping. **Plant Science**, v. 295, p. 110396, 2020. DOI <https://doi.org/10.1016/j.plantsci.2019.110396>

ROSA, A. H. B.; PERILLO, L. N.; NEVES, F. S.; RIBEIRO, D. B. Butterflies collected using malaise traps as useful bycatches for ecology and conservation. **Journal of Threatened Taxa**, v. 11, n. 9, p. 14235-14237, 2019. DOI <https://doi.org/10.11609/jott.4919.11.9.14235-14237>

SIERRA-OROZCO, E.; SHEKASTEBAND, R.; ILLA-BERENGUER, E.; SNOUFFER, A.; VAN DER KANAAP, E.; LEE, T. G. Identification and characterization of GLOBE, a major gene controlling fruit shape and impacting fruit size and marketability in tomato. **Horticulture Research**, v. 8, 2021. DOI <https://doi.org/10.1038/s41438-021-00574-3>

VITE-CHÁVEZ, O.; FLORES-TRONCOSO, J.; OLIVEIRA-REYNA, R.; MUNOZ-MINJARES, J. U. Improvement Procedure for Image Segmentation of Fruits and Vegetables Based on the Otsu Method. **Image Analysis and Stereology**, v. 42, n. 3, p. 185-196, 2023. DOI <https://doi.org/10.5566/ias.2939>

WALLACE, J. G.; RODGERS-MELNICK, E.; BUCKLER, E. S, Edward S. On the road to breeding 4.0: unraveling the good, the bad, and the boring of crop quantitative genomics. **Annual review of genetics**, v. 52, p. 421-444, 2018. DOI <https://doi.org/10.1146/annurev-genet-120116-024846>

WASHBURN, J. D.; MEJIA-GUERRA, M. K.; RAMSTEIN, G.; KREMLING, K. A.; VALLURU, R.; BUCKLER, E. S. Evolutionarily informed deep learning methods for predicting relative transcript abundance from DNA sequence. **Proceedings of the National Academy of Sciences**, v. 116, n. 12, p. 5542-5549, 2019. DOI <https://doi.org/10.1073/pnas.1814551116>

ZHU, G.; WANG, S.; HUANG, Z.; ZHANG, S.; LIAO, Q.; ZHANG, C. Rewiring of the fruit metabolome in tomato breeding. **Cell**, v. 172, n. 1, p. 249-261. e12, 2018. DOI [10.1016/j.cell.2017.12.019](https://doi.org/10.1016/j.cell.2017.12.019)

ZHU, Y.; GU, Q.; ZHAO, Y.; WAN, H.; WANG, R.; ZHANG, X.; CHENG, Y. Quantitative extraction and evaluation of tomato fruit phenotypes based on image recognition. **Frontiers in Plant Science**, v. 13, p. 859290, 2022. DOI <https://doi.org/10.3389/fpls.2022.859290>

5 CONSIDERAÇÕES FINAIS

A combinação de análise de imagens com visão computacional e aprendizado profundo oferece uma alternativa valiosa para aprimorar a qualidade dos frutos de tomateiro, diminuindo as subjetividades nas análises e reduzindo significativamente o tempo de fenotipagem. Ao alcançar uma fenotipagem mais precisa e eficiente dos frutos de tomate, este estudo contribui para a prevenção de erros e aumenta a eficiência na utilização de dados relevantes para futuros trabalhos de melhoramento não apenas do tomateiro, mas também de outras culturas.

As metodologias de aquisição das imagens foram de fácil execução e se mostraram eficazes, mas a preparação das imagens para o processamento computacional neste estudo foi desafiadora, uma das etapas mais laboriosas e que demandaram mais tempo, mas ainda são tentativas e erros que estão sendo ajustadas e logo terá eficiência nesta etapa também. Trabalhos podem ser feitos para otimização desta etapa.

Este estudo abre caminho para futuras pesquisas, sugerindo a aplicação da análise computacional de imagens para identificar a qualidade e produtividade por fruto em diversas culturas. Além disso, destaca o potencial uso de esteiras, drones e robôs na fenotipagem, indicando uma ampla gama de possibilidades para aprimorar a eficiência na avaliação de características morfoagronômicas.