



MARCELO HENRIQUE AVELAR MENDES

**INTEGRAÇÃO DE HORTICULTURA DE PRECISÃO E
INTELIGÊNCIA ARTIFICIAL NA CLASSIFICAÇÃO PÓS-
COLHEITA DE MORANGOS**

**LAVRAS – MG
2025**

MARCELO HENRIQUE AVELAR MENDES

**INTEGRAÇÃO DE HORTICULTURA DE PRECISÃO E INTELIGÊNCIA
ARTIFICIAL NA CLASSIFICAÇÃO PÓS-COLHEITA DE MORANGOS**

Tese apresentada à Universidade Federal de Lavras,
como parte das exigências do Programa de Pós-
Graduação em Agronomia/Fitotecnia, área de
concentração em Produção Vegetal, para a obtenção
do título de Doutor.

Profa. Dra. Luciane Vilela Resende
Orientadora

Prof. Dr. Adão Felipe dos Santos
Coorientador

**LAVRAS – MG
2025**

**Ficha Catalográfica elaborada pelo Sistema de Geração
de Ficha Catalográfica da Biblioteca Universitária da UFLA, com dados
informados pelo(a) próprio(a) autor(a).**

Mendes, Marcelo Henrique Avelar .

Integração de horticultura de precisão e inteligência artificial na classificação pós colheita de morangos / Marcelo Henrique Avelar Mendes. - 2025.

61 p. : il.

Orientadora: Luciane Vilela Resende

Coorientador: Adão Felipe dos Santos

Tese (Doutorado) - Universidade Federal de Lavras, 2025.

Bibliografia.

1. Aprendizado supervisionado. 2. Imagens digitais. 3. Predição de características.
4. Horticultura de precisão. I. Resende, Luciane Vilela . II. dos Santos, Adão Felipe .
III. Universidade Federal de Lavras. IV. Título.

MARCELO HENRIQUE AVELAR MENDES

**INTEGRAÇÃO DE HORTICULTURA DE PRECISÃO E INTELIGÊNCIA
ARTIFICIAL NA CLASSIFICAÇÃO PÓS-COLHEITA DE MORANGOS
INTEGRATION OF PRECISION HORTICULTURE AND ARTIFICIAL
INTELLIGENCE IN POST-HARVEST SORTING OF STRAWBERRIES**

Tese apresentada à Universidade Federal de Lavras,
como parte das exigências do Programa de Pós-
Graduação em Agronomia/Fitotecnia, área de
concentração em Produção Vegetal, para a obtenção
do título de Doutor.

APROVADO, em 27 de junho de 2025.

Dra. Luciane Vilela Resende – UFLA

Dr. Sebastião Márcio de Azevedo – UFLA

Dr. Adão Felipe dos Santos – UFLA

Dra. Daniela da Hora Farias – EPAMIG

Dr. Rouverson Pereira da Silva – UNESP

Documento assinado digitalmente



LUCIANE VILELA RESENDE

Data: 16/07/2025 16:01:12-0300

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

Profa. Dra. Luciane Vilela Resende
Orientadora

Prof. Dr. Adão Felipe dos Santos
Coorientador

**LAVRAS – MG
2025**

*“Dedico esta tese à minha família, à minha esposa e a todos que,
com amor e fé, me apoiaram e caminharam ao meu lado
ao longo desta jornada.”*

AGRADECIMENTOS

Em primeiro lugar, agradeço a Deus, por me conceder forças, sabedoria e saúde para seguir até aqui.

À minha família, em especial à minha mãe, Marlene Avelar de Andrade, ao meu pai, Emilio Américo de Castro Mendes (in memoriam), e ao meu irmão, Hudson Carlos de Andrade, pela base sólida que me proporcionaram por meio da educação, apoio incondicional, força, carinho e exemplo de integridade. Sou profundamente grato por tudo que fizeram por mim ao longo desta caminhada.

À minha esposa, Angélica Rodrigues Magalhães, minha companheira de vida, expresso minha profunda gratidão pelo amor constante, paciência nos momentos difíceis e apoio incondicional que tornaram possível a realização desta jornada.

À minha querida orientadora, professora e amiga Profa. Dra. Luciane Vilela Resende, minha sincera gratidão pela excelente orientação, apoio e confiança ao longo de todo o percurso. Seus ensinamentos foram fundamentais para o desenvolvimento deste trabalho e para minha formação como pesquisador.

Ao meu coorientador Prof. Dr. Adão Felipe dos Santos pela paciência, disposição, atenção e ensinamentos.

Ao professor Dr. Wilson Magela pela colaboração, espaço, atenção, dedicação e amizade.

À técnica Stefany Martins da Silva Lino e a todos do setor de Olericultura, pela amizade, colaboração, receptividade acolhedora e prontidão em auxiliar em todas as etapas do projeto.

Aos membros da banca, à Universidade Federal de Lavras, ao Departamento de Agricultura e ao Programa de Pós-Graduação em Fitotecnia, pelas contribuições valiosas, pelo excelente ensino, ética e ensinamentos ao longo do curso.

Aos colegas de pesquisa e amigos da vida, pela parceria, ajuda, trocas de ideias e pelo apoio nos momentos difíceis. As conversas, os cafés e os risos foram essenciais.

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

À Fundação de Amparo à Pesquisa do Estado de Minas Gerais (FAPEMIG) e ao Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq), pelo apoio por meio de bolsas de estudo e infraestrutura para a realização deste trabalho.

*“Quem tenta incessantemente alcança e
no balanço das andanças aprendi sobreviver” – (Marcelo D2).*

RESUMO

A predição e classificação da qualidade de frutos de morango atualmente dependem, majoritariamente, de métodos manuais e destrutivos, o que compromete a eficiência e a padronização no processo de pós-colheita. O uso de imagens digitais e algoritmos de aprendizado de máquina surge como alternativa promissora para automatizar essas etapas de forma precisa e não invasiva. O presente trabalho teve como objetivo avaliar modelos baseados em aprendizado de máquina, para desenvolver métodos não destrutivos que permitam mensurar a massa dos frutos de morango e realizar sua classificação comercial. Para isso, frutos de oito genótipos de morangueiro foram produzidos em casa de vegetação, colhidos e analisados entre setembro e outubro de 2023, na UFLA, Lavras/MG. Cada fruto foi fotografado em estúdio padronizado com dispositivo móvel e as imagens passaram por processamento para segmentação, remoção do fundo e extração de atributos morfológicos (ex: área, perímetro, fator de forma). Em paralelo, os frutos foram avaliados em laboratório quanto a massa, diâmetros e parâmetros de qualidade em escala. Foram aplicados oito algoritmos de aprendizado supervisionado – SVM, RF, DT, KNN, MLP, XGB, LGBM e MLR – para predição da massa e classificação dos frutos de morango. As análises foram realizadas em Python, no Google Colab. Os modelos foram treinados com 70% dos dados e testados com os 30% restantes, com a aplicação de validação cruzada com K-fold. Para predição da massa os modelos foram avaliados pelas métricas R^2 , RMSE, MAE e d, para a classificação seguindo as métricas Acurácia, F1- score, Recall e Precision. Para a predição da massa dos frutos, na fase de teste, o modelo SVM demonstrou os melhores resultados para ambos os conjuntos de dados. O KNN apresentou bom desempenho quanto ao conjunto de dados manuais, assim como, o XGB apresentou desempenho parecido ao SVM para o conjunto de dados de imagens. O modelo LGBM obteve os melhores desempenhos para treino, mas apresentou queda no teste, sugerindo sobre ajuste. Para ambos os conjuntos de dados o modelo MLP teve o desempenho menos satisfatório. O uso dos modelos para a classificação dos frutos atingiu acurácia variando de 98% a 96%, para os dados manuais, com destaque para os modelos, SVM, XGB, LGBM e DT, e para os dados de imagens todos modelos tiveram acurácia acima de 0,91, exceto o MLP, que apresentou 0,83. Embora os dados manuais tenham proporcionado maior estabilidade e acerto, aos modelos testados, as variáveis extraídas das imagens demonstraram potencial para predição e classificação dos frutos de morango. Isso reforça o uso de imagens padronizadas como ferramenta não destrutiva, eficiente e escalável para estimar a massa dos frutos e classificá-los, abrindo caminhos para aplicações práticas em monitoramento automatizado, com ganhos em agilidade e redução de custos, contribuindo para avanços na agricultura de precisão e suporte na automação de predição de massa para a cultura do morango.

Palavras-chaves: aprendizado supervisionado; imagens digitais; predição de características; horticultura de precisão.

ABSTRACT

Strawberry fruit quality prediction and grading currently rely primarily on manual and destructive methods, compromising efficiency and standardization in the post-harvest process. The use of digital images and machine learning algorithms emerges as a promising alternative for automating these steps accurately and non-invasively. This study aimed to evaluate machine learning-based models to develop non-destructive methods for measuring strawberry fruit weight and commercial grading. To this end, fruits of eight strawberry genotypes were produced in a greenhouse, harvested, and analyzed between September and October 2023 at UFLA, Lavras, Minas Gerais. Each fruit was photographed in a standardized studio with a mobile device, and the images were processed for segmentation, background removal, and extraction of morphological attributes (e.g., area, perimeter, shape factor). In parallel, the fruits were evaluated in the laboratory for weight, diameter, and quality parameters at scale. Eight supervised learning algorithms—SVM, RF, DT, KNN, MLP, XGB, LGBM, and MLR—were applied to predict strawberry mass and classify it. Analyses were performed in Python on Google Colab. The models were trained on 70% of the data and tested on the remaining 30% using k-fold cross-validation. For mass prediction, the models were evaluated using R^2 , RMSE, MAE, and d, and for classification, using the metrics accuracy, F1-score, recall, and precision. For fruit mass prediction, in the testing phase, the SVM model demonstrated the best results for both datasets. KNN performed well on the manual dataset, and XGB performed similarly to SVM for the image dataset. The LGBM model achieved the best performance in training but showed a decline in testing, suggesting overfitting. For both datasets, the MLP model performed less well. The use of models for fruit classification achieved accuracy ranging from 98% to 96% for manual data, with emphasis on the SVM, XGB, LGBM, and DT models. For image data, all models achieved accuracy above 0.91, except for MLP, which achieved 0.83. Although manual data provided greater stability and accuracy for the models tested, the variables extracted from the images demonstrated potential for predicting and classifying strawberry fruits. This reinforces the use of standardized images as a non-destructive, efficient, and scalable tool for estimating fruit mass and classifying them, paving the way for practical applications in automated monitoring, with gains in agility and cost reduction, contributing to advances in precision agriculture and supporting the automation of mass prediction for strawberry crops.

Keywords: supervised learning; digital images; trait prediction; precision horticulture.

INDICADORES DE IMPACTO

A presente pesquisa apresenta impactos em potencial de natureza tecnológica, econômica e social ao desenvolver e aplicar *machine learning* para estimar a massa de frutos de morango a partir de dados morfofisiológicos e imagens digitais. Esta abordagem contribui para a automatização, precisão na classificação e comercialização da fruta, promovendo ganhos significativos em eficiência e redução de custos no processo produtivo. Os impactos tecnológicos se evidenciam na proposta de um sistema preditivo de baixo custo que pode ser incorporado a plataformas de avaliação não destrutiva, promovendo inovação no setor agrícola. De acordo com a Política Nacional de Extensão, os impactos estão associados às áreas temáticas de Tecnologia e Produção, Trabalho e Meio Ambiente, por contribuírem para a modernização da agricultura, aumento da eficiência produtiva e redução do desperdício de alimentos. O estudo está alinhado com diversos Objetivos de Desenvolvimento Sustentável (ODS) das Nações Unidas, especialmente o ODS 2 (Fome Zero e Agricultura Sustentável), ODS 8 (Trabalho decente e crescimento econômico), ODS 9 (Indústria, Inovação e Infraestrutura) e ODS 12 (Consumo e Produção Responsáveis), ao propor uma alternativa tecnológica acessível, baseada em ciência de dados, que pode melhorar a rentabilidade da cadeia do morango, com impactos potenciais em outras culturas agrícolas.

IMPACT INDICATORS

This research presents potential technological, economic, and social impacts by developing and applying machine learning to estimate strawberry fruit weight based on morphophysiological data and digital images. This approach contributes to automation and accuracy in fruit classification and marketing, promoting significant efficiency gains and cost reductions in the production process. The technological impacts are evident in the proposed low-cost predictive system that can be incorporated into non-destructive evaluation platforms, fostering innovation in the agricultural sector. According to the National Extension Policy, the impacts are associated with the thematic areas of Technology and Production, Labor, and the Environment, as they contribute to the modernization of agriculture, increased production efficiency, and reduced food waste. The study is aligned with several United Nations Sustainable Development Goals (SDGs), especially SDG 2 (Zero Hunger and Sustainable Agriculture), SDG 8 (Decent Work and Economic Growth), SDG 9 (Industry, Innovation and Infrastructure) and SDG 12 (Responsible Consumption and Production), by proposing an accessible technological alternative, based on data science, that can improve the profitability of the strawberry chain, with potential impacts on other agricultural crops.

SUMÁRIO

1	INTRODUÇÃO	13
2	REFERENCIAL TEÓRICO	15
2.1	Origem e Diversidade do Morangueiro	15
2.2	Importância Econômica	17
2.3	Agricultura digital e tecnologias emergentes aplicadas à horticultura	19
2.4	Abordagens modernas para fenotipagem e avaliação física do morangueiro	22
2.5	Aprendizado de máquina na agricultura	24
2.6	Modelos de aprendizado supervisionado aplicados à estimativa de variáveis agronômicas	25
3	MATERIAL E MÉTODOS	30
3.1	Condução de plantas em casa de vegetação	30
3.2	Obtenção das imagens	30
3.3	Processamento das imagens	31
3.4	Análises morfológicas	33
3.5	Correlação de Spearman	33
3.6	Modelagem preditiva e classificação	33
4	RESULTADOS	36
5	DISCUSSÃO	46
6	CONCLUSÃO	51
	REFERÊNCIAS	52

1 INTRODUÇÃO

O morangueiro (*Fragaria ananassa*) é uma das plantas frutíferas de pequeno porte mais relevantes e valorizadas mundialmente, tanto por seu valor econômico quanto pelos potenciais benefícios à saúde. A produção global chegou a 14,5 milhões de toneladas anuais, sendo cultivado em quase todos os países do globo, gerando um valor econômico total da cultura em todo o mundo de US\$ 14 bilhões, evidenciando sua importância no cenário agrícola internacional (Guevara-Téran *et al.*, 2023; Menzel, 2024).

No Brasil, a cultura do morango apresenta expressiva relevância econômica e social, posicionando o país como o 9º produtor mundial e o principal da América do Sul. A produção nacional gira em torno de 150.000 toneladas, distribuídas em aproximadamente 4.200 hectares. O estado de Minas Gerais lidera a produção nacional, concentrando cerca de 60% do total produzido, com destaque para a região Sul do estado, responsável por aproximadamente 95% da produção mineira (Brandt *et al.*, 2022; Nunes *et al.*, 2022; Delazeri *et al.*, 2024).

O morango apresenta múltiplas formas de aproveitamento, podendo ser consumido tanto *in natura* quanto utilizado no preparo de geleias, doces e diversos outros produtos, o que contribui para aumentar ainda mais seu valor comercial e funcional. A lucratividade do cultivo de morango está diretamente relacionada ao rendimento das plantas e à qualidade dos frutos colhidos. Para agregar valor comercial e atender às diferentes exigências do mercado consumidor, a classificação pós-colheita é uma etapa muito importante, onde a massa do fruto, associado ao tamanho e à aparência são critérios essenciais de avaliação.

A massa dos frutos está diretamente relacionada à sua composição nutricional, ao comportamento pós-colheita à estimativa da produtividade e à aceitação pelos consumidores. Além disso, serve como referência para a classificação dos frutos em comerciais e não comerciais, influenciando na padronização e no valor de mercado.

O método de determinação da massa dos frutos é comumente realizado por meio de balanças de precisão. Embora esse método forneça medidas diretas e com boa precisão, possuem limitações significativas. O manuseio dos morangos durante a mensuração, pode causar danos mecânicos nos frutos, tornando-os suscetíveis a infecções e acelerando o processo de deterioração, o que resulta em perdas econômicas.

Nesse contexto, métodos não destrutivos baseados em visão computacional, aliados a algoritmos de aprendizado de máquina e ciência de dados, representam uma alternativa eficiente e estratégica para estimar a massa dos frutos e automatizar sua classificação,

contribuindo diretamente para a modernização dos processos de manejo na cadeia produtiva do morango (Nagaki *et al.*, 2025).

Algoritmos de aprendizado de máquina constroem modelos preditivos a partir de dados mensurados e rotulados que contêm características relevantes dos objetos de interesse, e têm sido amplamente aplicados na agricultura, inclusive para estimar a massa de frutos, oferecendo vantagens como rapidez, múltiplas avaliações simultâneas e decisões em tempo real (Basak *et al.*, 2022; Barbosa Júnior *et al.*, 2025; Chen *et al.*, 2025).

Nas últimas décadas, a aplicação de técnicas de Inteligência Artificial na cultura do morango tem avançado consideravelmente, especialmente em áreas como detecção de doenças (Li *et al.*, 2025), estimativa da biomassa vegetativa (Dalid *et al.*, 2025), estimativa de parâmetros relacionados à maturidade, colheita e qualidade de frutos (Elsayed *et al.*, 2025) além de servirem como ferramentas promissoras em programas de melhoramento genético (Zheng, Abd-Elrahman & Whitaker, 2021). Entretanto, há poucas abordagens focadas na estimativa precisa da massa individual dos frutos e sua classificação comercial com base em dados extraídos por visão computacional.

Acredita-se que modelos de aprendizado de máquina treinados com dados morfológicos extraídos de imagens são capazes de prever com precisão a massa dos frutos de morango e classificá-los corretamente em categorias comerciais, de forma não destrutiva, contribuindo para a modernização da cadeia produtiva do morango e a redução de perdas econômicas relacionadas à manipulação manual.

Portanto, este estudo tem como objetivo avaliar modelos baseados em aprendizado de máquina, para desenvolver métodos não destrutivos que permitam prever a massa dos frutos de morango e realizar sua classificação comercial.

2 REFERENCIAL TEÓRICO

2.1 Origem e Diversidade do Morangueiro

O morango cultivado (*Fragaria ananassa* Duch., $2n = 8x = 56$), pertencente à família Rosaceae, tem origem em um híbrido interespecífico entre as espécies selvagens octoploides *Fragaria virginiana*, do leste dos Estados Unidos, e *Fragaria chiloensis*, endêmica do Chile. A introdução dessas espécies na Europa foi essencial para sua hibridização, viabilizada pelo cultivo conjunto, algo impossível nas Américas devido ao isolamento geográfico. A espécie *F. virginiana* foi levada à França no século XVII, enquanto *F. chiloensis* foi introduzida por Amédée Frézier em 1714. Em 1766, Nicolas Duchesne realizou cruzamentos entre essas espécies, resultando no surgimento do morangueiro moderno, que, desde então, passou a ser cultivado amplamente em diferentes regiões do mundo (Darrow, 1966; Wang *et al.*, 2021; Hummer *et al.*, 2023; Chen *et al.*, 2024).

Os programas de melhoramento formal iniciaram-se na Inglaterra (1817) e nos EUA (1929) visando características específicas como remontagem e adaptação climática, além do objetivo de melhorar os frutos com foco em frutos maiores, mais firmes e com maior resistência ao transporte (Darrow, 1966; Gaston *et al.*, 2020; Porter *et al.*, 2023). No Brasil, não se sabe ao certo quando a cultura foi introduzida. O primeiro programa de melhoramento foi iniciado pelo Instituto Agrônomo de Campinas (IAC) em 1941, o que resultou no lançamento de uma cultivar chamada Campinas, trazendo maior produtividade e expansão de cultivo nos estados de São Paulo, Minas Gerais e Rio Grande do Sul e logo se expandiu em quase todo o país (Camargo *et al.*, 2010; Antunes & Peres, 2013).

Embora não se saiba ao certo o número total de cultivares de *Fragaria* × *ananassa* no mundo, estima-se que existam mais de mil, com variações em características como fotoperíodo, produtividade e resistência a doenças. Estudos têm investigado o desempenho dessas cultivares sob diferentes fatores ambientais e genéticos, incluindo estresses, compostos orgânicos voláteis e condições de cultivo, pois certas cultivares respondem melhor a ambientes específicos, afetando atributos como cor e aroma dos frutos (Gil *et al.*, 2024).

Os avanços genéticos promovidos pelo melhoramento foram determinantes para o aumento da produção de morangos nas últimas décadas. No entanto, no Brasil ainda há muito a ser desenvolvido. Apesar dos ganhos em produção, rendimento e adaptação ao fornecimento anual, o morango ainda é pouco estudado em comparação a culturas anuais de produção de grãos (Antunes *et al.*, 2021; Feldmann *et al.*, 2024; Zhong, 2024).

A dependência de material genético estrangeiro eleva os custos da produção de morangos no Brasil, devido à importação de mudas e pagamento de *royalties*. Além disso, as mudas chegam tardiamente, e podem introduzir pragas e patógenos no país, o que reforça a necessidade de fortalecer programas nacionais de melhoramento (Zeist & Resende, 2019; Santos *et al.*, 2021).

Esses programas têm como objetivos principais melhorar características agrônômicas e de qualidade da fruta, como resistência a pragas, adaptabilidade a sistemas semi-hidropônicos, produtividade, precocidade, tolerância ao calor, firmeza, cor, aroma e doçura (Santos *et al.*, 2021). No entanto, reunir todas essas características em uma única cultivar é desafiador, devido ao genoma octaploide e à forte interação com o ambiente (Barth *et al.*, 2022).

Com o objetivo de adaptar o cultivo de morango às condições edafoclimáticas do Sul de Minas Gerais, a Universidade Federal de Lavras (UFLA) iniciou, em 2011, um programa de melhoramento genético baseado em cruzamentos com cultivares amplamente utilizadas no Brasil. Atualmente, a UFLA mantém um banco de germoplasma com materiais promissores, superiores às cultivares comerciais mais usadas no país (Pereira *et al.*, 2022).

A coleta e preservação de germoplasma, junto à avaliação da variabilidade genética, são fundamentais para o sucesso de qualquer programa de melhoramento, pois a diversidade do banco de germoplasma influencia diretamente o desenvolvimento de novos genótipos (Mishra *et al.*, 2015; Corrêa *et al.*, 2021).

2.2 Importância Econômica

O morango (*Fragaria ananassa* Duch.) é uma fruta de alto valor agregado e grande importância socioeconômica, destacando-se na fruticultura por sua relevância nutricional, comercial e como fonte de renda para agricultores familiares, com forte presença nos mercados locais e internacionais (Faedo *et al.*, 2024).

A produção mundial de morango vem crescendo significativamente nos últimos anos, chegando a um aumento de sua demanda em aproximadamente 300%, de 3,49 milhões de toneladas em 1994 para 10,48 milhões de toneladas em 2023, tornando-se uma das mais importantes culturas frutíferas cultivadas em todo o mundo (FAO, 2023). A produção mundial total da fruta, gera um valor econômico total de US\$ 14 bilhões em todo o mundo (Guevara-Téran *et al.*, 2023; Menzel, 2024).

O cultivo do morango está concentrado no hemisfério norte, responsável por 98% da produção mundial, sendo 43% na Ásia, 27% nas Américas e 24% na Europa. A China lidera com 3,39 milhões de toneladas anuais, seguida pelos EUA (1,21 milhões) e México (0,84

milhões). Na África, o Egito é o principal produtor (470.913 toneladas), e na América do Sul, o Brasil lidera com mais de 200 mil toneladas por ano, representando mais de um terço da produção sul-americana e ocupando a oitava posição no ranking mundial (ATLAS BIG, 2021; Antunes *et al.*, 2021; Guevara-Téran *et al.*, 2023; Da Silva *et al.*, 2024).

No Brasil o morango a principal espécie de pequenas frutas vermelhas em várias regiões do país, com grande impacto socioeconômico, principalmente em pequenas áreas com agricultura familiar, com cerca de 5.300 hectares, apresentando uma produção de mais de 218.000 toneladas. Os empreendimentos agrícolas voltados para o cultivo de morangos geralmente têm áreas que variam de 0,5 a 2,5 hectares. No entanto, em casos de monocultura, essas áreas podem superar 15 hectares. Quanto à produção, os rendimentos médios de morangos vão de 15 a 51 toneladas por hectare. Já o consumo anual de morangos por pessoa varia de 1,5 kg a 5 kg, dependendo da região e dos hábitos de consumo (De Souza *et al.*, 2019; Antunes *et al.*, 2021; Chiomenro *et al.*, 2021; Santos *et al.*, 2021; Faedo *et al.*, 2024).

A extensão territorial brasileira é propícia para o cultivo do morango, sendo assim, doze estados brasileiros são responsáveis pela produção comercial de morangos. Porém, cerca de 75% dessa produção é oriundo dos estados de Minas Gerais onde a produção é 74.000 toneladas em 2.000 ha, Paraná (21.450 toneladas, 650 ha) e Rio Grande do Sul (20.350 toneladas, 550 ha). A produtividade média do morango no país fica em torno de 38,5 t/ha, com diferenças entre as regiões e o sistema de cultivo adotado. Esse rendimento ainda é bem menor quando comparado a outros países, como os Estados Unidos da América (67,9 t/ha) e a Espanha (52,8 t/ha) (De Souza *et al.*, 2019; Chiomenro *et al.*, 2021; Santos *et al.*, 2021; Da Silva *et al.*, 2024).

No Brasil, cerca de 70% dos cultivos de morango são realizados com as cultivares Albion, Camarosa e San Andreas (Martins de Lima *et al.*, 2021). Mas, atualmente, com o uso de novas cultivares, principalmente de dia neutro, o aumento da área plantada, o uso de plantas saudáveis e a alta profissionalização no cultivo do morangueiro, fizeram com que a produtividade da cultura aumentasse no país (Fagherazzi *et al.*, 2016).

Os principais fatores para exploração comercial desta cultura são a alta lucratividade, quando comparada a outras espécies como milho e soja, e o amplo conhecimento e aceitação pelo consumidor devido à diversidade de opções de comercialização e processamento do fruto, uma vez que o produto pode ser utilizado tanto para comercialização *in natura* quanto para processamento na forma de doces, iogurtes, sucos, licores e outros produtos que aproveitem as características marcantes do fruto, como aroma, cor e sabor (Dos Santos *et al.*, 2021).

O morangueiro também apresenta valor ornamental e seus frutos possuem elevado valor nutricional, apresentando altos teores de vitaminas, fibras alimentares, oligoelementos,

minerais, além de ser antioxidantes e antitumorais. Além de ser de extrema importância comercial, o morango é uma planta modelo para a pesquisa de frutíferas devido ao seu pequeno tamanho, ciclo curto de crescimento do fruto e alta eficiência de transformação, sendo muito importante na pesquisa científica em todo o mundo (Ni *et al.*, 2021; Wang *et al.*, 2021).

2.3 Agricultura digital e tecnologias emergentes aplicadas à agricultura

A crescente integração de tecnologias digitais na vida social e econômica está promovendo uma transformação significativa, especialmente na agricultura. Os *smartphones* alimentados por novas tecnologias de sensoriamento, inteligência artificial (IA) e algoritmos de *machine learning* (ML) criam uma nova percepção entre as pessoas e os sistemas para a resolução de problemas. Essas inovações estão redefinindo os processos de produção em escala global, impulsionando a eficiência, a sustentabilidade e a modernização do setor agrícola (Mendes *et al.*, 2020; Hackfort, 2021). A transição para a agricultura digital é essencial não apenas para aprimorar a gestão e aumentar os rendimentos das culturas, mas também para contribuir com objetivos globais, como a segurança alimentar e a mitigação das mudanças climáticas (Ezziyyani *et al.*, 2023; Abdo-Peralta *et al.*, 2024).

Os termos "agricultura digital", "agricultura inteligente" e "agricultura 4.0" são frequentemente usados de forma intercambiável para descrever a crescente aplicação de tecnologias baseadas em dados na agricultura e na produção de alimentos (Hackfort, 2021).

Agricultura digital refere-se à digitalização dos processos agrícolas, ou seja, à coleta, armazenamento, processamento e análise de dados agrícolas por meio de tecnologias digitais, tais como sensores, drones, imagens de satélite, *softwares* de gestão agrícola, *big data*, *Internet das Coisas* (IoT) e conectividade, com foco em transformar dados analógicos em digitais e integrá-los em sistemas de apoio à decisão (Dhal *et al.*, 2024). A agricultura digital engloba dois aspectos principais: a digitalização técnica, que envolve a conversão de informações analógicas em dados digitais, e a digitalização social, que diz respeito à incorporação de tecnologias computacionais nos processos agrícolas e no cotidiano do setor (Hackfort, 2021).

A agricultura inteligente aborda a implementação do uso de plataformas móveis, sensores, tecnologias digitais, computação em nuvem, robótica, IA e IoT, conectadas para tomada de decisão baseada em dados, com foco em eficiência, sustentabilidade e produtividade, otimização do uso de recursos (água, fertilizantes, energia), aumento da produtividade e redução de impactos ambientais (Mohamed *et al.*, 2021).

Já o termo Agricultura 4.0 refere-se à aplicação dos princípios da Indústria 4.0 ao setor agrícola, como a quarta evolução na tecnologia agrícola, ou seja, a integração dessas tecnologias

na agricultura. Se torna um termo mais atraente ao mercado que também denomina a agricultura inteligente ou a agricultura digital (Abbasi, Martinez e Ahmad, 2022).

Os dados coletados na agricultura digital, tanto manualmente quanto por equipamentos, podem apresentar diferentes formatos, com conteúdo específico, estruturas heterogêneas e em grandes volumes. Esses dados geralmente precisam ser pré-processados e transformados em um formato uniforme e apropriado para serem usados pelos algoritmos computacionais. Esses algoritmos desenvolvidos a partir desses bancos de dados criam *insights*, que podem ser análise de padrão ou tendência para ajudar os pesquisadores e agricultores a tomar decisões informadas ou pode ser operações automatizadas realizadas por uma máquina (Ingram *et al.*, 2022; Dara, Hazrati Fard & Kaur, 2022).

Atualmente as pesquisas acadêmicas voltadas a agricultura têm gerado um enorme base de dados devido as inúmeras características que se mensuram em campo. Essas características são difíceis de serem avaliadas, demanda tempo e as vezes utiliza-se de métodos destrutivos, devido à indisponibilidade de técnicas não invasivas, que sejam acessíveis e fáceis de usar. Com isso, existe a necessidade de se desenvolver técnicas acessíveis que possam medir de forma não destrutiva as características de crescimento e desenvolvimento, principalmente em programas de melhoramento de plantas, possibilitando avanços rápidos no aprimoramento de aspectos como forma, tamanho e cor de produtos agrícolas (Li, Cheng *et al.*, 2020; Zingaretti, Monfort & Pérez-Enciso, 2021).

As imagens digitais estão entre os dados mais acessíveis e amplamente disponíveis, desempenhando um papel crucial na avaliação de características morfológicas das plantas. A análise de imagem tem sido usada para avaliar de forma não destrutiva a área foliar de mudas, qualidade da planta, incidência de doenças, vigor da muda e qualidade pós-colheita em frutas e vegetais. Nos últimos anos, *smartphones* com câmeras de alta qualidade e *softwares* como o Matlab e o R, têm facilitado a captura e o processamento preciso de imagens de grandes grupos de plantas (Li, Cheng *et al.*, 2020; Zingaretti, Monfort & Pérez-Enciso, 2021).

Câmeras acopladas a drones ou *smartphones* têm possibilitado a obtenção de imagens em alta resolução, que, processadas por algoritmos inteligentes, permitem identificar plantas daninhas (Khan *et al.*, 2021), fazem o reconhecimento rápido e preciso de doenças (Li *et al.*, 2025), quantificam os efeitos causados por eventos climáticos (Sunil *et al.*, 2025), têm a capacidade de medir a biomassa vegetativa em escala (Dalid *et al.*, 2025), fazem a classificação de diferentes estágios fenológicos (Bushakra *et al.*, 2025), além de servirem como ferramentas promissoras em programas de melhoramento genético (Zheng, Abd-Elrahman & Whitaker, 2021).

O processamento das imagens através de algoritmos de aprendizado de máquina tem sido notavelmente bem-sucedido em tarefas de classificação, reconhecimento de objetos e segmentação semântica usando as informações contidas em imagens RGB e multiespectrais. Modelos de aprendizado de máquina vêm sendo amplamente utilizados na análise de dados agrícolas, por meio da extração de descritores visuais (como textura, forma, cor e padrões espaciais) fundamentais para a análise morfológica e fenotípica. Entretanto, a eficácia desses métodos depende da qualidade dos dados de entrada e da escolha precisa dos descritores, o que se torna particularmente desafiador em cenários com alta variabilidade ambiental e ruído visual (Li *et al.*, 2025).

2.4 Aprendizado de máquina na agricultura

A agricultura digital vem se consolidando como uma nova disciplina científica que utiliza métodos intensivos e baseados em dados para aumentar a produtividade, melhorar a eficiência no uso de insumos e, ao mesmo tempo, reduzir os impactos ambientais (Attri *et al.*, 2024). Esse novo paradigma gera um volume massivo de informações (comumente denominado *big data*) que exige não apenas grande capacidade de armazenamento, mas também ferramentas eficazes para edição, processamento, análise e interpretação (Benos *et al.*, 2021).

O *big data* agrícola caracteriza-se por apresentar cinco atributos fundamentais: volume, variedade, velocidade, veracidade e valor. Essas características tornam as abordagens convencionais de análise de dados insuficientes para atender às crescentes demandas da agricultura inteligente. A incapacidade de extrair informações úteis e acionáveis a partir de dados complexos de campo representa um dos principais gargalos na aplicação prática desse novo modelo agrícola (Benos *et al.*, 2021).

Aliado ao avanço da tecnologia de *big data* e a computação de alto desempenho, surge o aprendizado de máquina (*Machine Learning* - ML), um subconjunto da inteligência artificial (IA), para abrir novas possibilidades para desvendar, quantificar e compreender processos intensivos em dados em ambientes operacionais agrícolas. O ML pode ser definido como o campo científico que confere às máquinas a capacidade de aprender a partir de dados, sem a necessidade de programação explícita para cada tarefa (Liakos *et al.*, 2018; Attri *et al.*, 2024; Benos *et al.*, 2021).

O *Machine Learning* é um processo que envolve a utilização de algoritmos que permitem aos computadores realizarem previsões ou tomar decisões com base em múltiplas entradas de dados. De modo geral, os dados representam conjuntos de exemplos que podem ser

rotulados (dados de entrada com uma resposta conhecida), utilizados em tarefas de aprendizado supervisionado, ou não rotulados, aplicáveis ao aprendizado não supervisionado, em que o modelo identifica padrões e estruturas por conta própria. O sucesso dos algoritmos está diretamente relacionado à qualidade, diversidade e volume de dados disponíveis. Na agricultura, no entanto, a obtenção de grandes volumes de dados confiáveis ainda representa um dos principais desafios técnicos, dificultando a generalização dos modelos e limitando sua adoção em ambientes reais (Araújo *et al.*, 2023).

Nesse contexto, a análise de dados desempenha papel central. Trata-se de um processo que envolve coleta, transformação, limpeza, modelagem e interpretação, com o objetivo de extrair conhecimento aplicável. Ferramentas analíticas modernas são capazes de examinar grandes conjuntos de dados multimodais (como imagens, vídeos, textos e sinais digitais) e revelar padrões ocultos, correlações e tendências que escapariam à observação humana. A integração de sensores embarcados em plataformas móveis ou fixas permite a coleta contínua de dados em tempo real, contribuindo para a detecção precoce de doenças em folhas, frutos e raízes, avaliação da qualidade e saúde dos cultivos, identificação de estresses abióticos e manejo inteligente da irrigação (Akhter & Sofi, 2022).

Além disso, o aprendizado de máquina está intrinsecamente ligado aos avanços da agricultura de precisão, pois permite o monitoramento espacial e temporal de variáveis críticas do ambiente e da planta, com maior acurácia e capacidade de expansão. Dados utilizados nesses sistemas podem ser oriundos de diversas fontes: sensores ópticos embarcados em drones, imagens multiespectrais e hiperespectrais, dados meteorológicos, sensores de solo, além de variáveis fenotípicas e morfológicas. Essa heterogeneidade exige ferramentas robustas para integração e análise, reforçando o papel do ML como um elo entre dados brutos e decisões agronômicas informadas (Condran *et al.*, 2022).

Nesse cenário, é importante investigar a avaliação de modelos de ML no ambiente agrícola. Os algoritmos de ML comumente usados no domínio agrícola são árvore de decisão (DT), k-vizinho mais próximo (kNN), florestas aleatórias (RF), máquina de vetores de suporte (SVM), análise lineares múltiplas (*Multiple linear*) e redes neurais (NN). Além desses temos, algoritmo de reforço de gradiente extremo (XGB), e o LGBM modelo de árvore de decisão de reforço de gradiente. Essas técnicas podem ser categorizadas por seu nível de interpretabilidade, ou seja, a característica inerente à técnica de aprendizado que permite ao usuário explicar o classificador resultante (Condran *et al.*, 2022; Bogdanovski *et al.*, 2023; Yoon *et al.*, 2023; Aldossary *et al.*, 2024).

2.5 Modelos de aprendizado supervisionado

Os algoritmos de aprendizado de máquina (ML) têm sido amplamente empregados em diversas aplicações, devido à sua capacidade de lidar com dados complexos e não estruturados de forma precisa e eficiente. Esses métodos destacam-se por estabelecerem relações não paramétricas e não lineares entre variáveis independentes e dependentes, tornando-se especialmente úteis em contextos com elevada variabilidade e complexidade (Sarker, 2021; Tufail *et al.*, 2023). A análise de grandes volumes de dados, comum em ambientes reais, é uma das áreas em que os algoritmos de ML têm se mostrado superior aos métodos estatísticos convencionais (Rahamani *et al.*, 2021).

Esses algoritmos possibilitam a predição de resultados com alta acurácia, mesmo sem a necessidade de programação explícita, por meio de procedimentos fundamentados em análises estatísticas. Além disso, o ML compartilha princípios com a mineração de dados, como a identificação de padrões e a atualização contínua das saídas conforme novas entradas são fornecidas. Outro destaque é sua contribuição para o avanço de tecnologias modernas, como assistentes virtuais, os quais utilizam modelos de aprendizado profundo para interpretar linguagem natural e fornecer respostas contextualizadas (Jagtap *et al.*, 2022; Kusal *et al.*, 2022; Kwaghtyo & Eke, 2023). Diante desse panorama, torna-se relevante compreender os principais algoritmos supervisionados empregados na modelagem preditiva em aplicações agrônômicas, como Random Forest (RF), Support Vector Machine (SVM), Decision Tree (DT), entre outros.

O Random Forest (RF) é um algoritmo de aprendizado de máquina, amplamente utilizado em problemas de classificação devido à sua robustez frente a ruídos nos dados e à capacidade de lidar com a multicolinearidade entre variáveis. Essas características o tornam particularmente eficaz em tarefas de seleção de atributos, permitindo a eliminação de informações redundantes e a obtenção de alta acurácia com um número reduzido de variáveis, especialmente em dados espectrais (Huang *et al.*, 2022). Como o nome sugere, o RF é uma floresta de árvores de decisão, que constrói múltiplas árvores a partir de subconjuntos de dados e variáveis selecionados aleatoriamente por meio da técnica de *bagging* (*bootstrap aggregating*). A predição final é determinada pela votação da maioria das árvores. As amostras não utilizadas em cada árvore são empregadas na validação cruzada interna, o que contribui para seu bom desempenho generalizado (Islam *et al.*, 2021). Além disso, o RF se destaca pela velocidade operacional e versatilidade, sendo aplicado também em domínios como segurança cibernética, para classificação e detecção de comportamentos maliciosos (Peppes *et al.*, 2021).

O Support Vector Machine (SVM) é um algoritmo supervisionado de aprendizado de máquina que busca encontrar um hiperplano em um espaço N-dimensional (sendo N o número de variáveis) capaz de separar os dados de forma adequada entre diferentes classes. Entre suas principais vantagens, destacam-se a forte capacidade de generalização e a precisão elevada, mesmo em cenários com conjuntos de treinamento limitados. No entanto, o modelo apresenta limitações diante de problemas de alta dimensionalidade, presença de ruídos e dados rotulados de forma incorreta (Peppes *et al.*, 2021). Em contextos de classificação baseada em pixels, o SVM tem se mostrado eficaz por sua tolerância à baixa amostragem e sua capacidade de generalizar bem em problemas não lineares com dados de baixa dimensão. Apesar disso, sua performance pode ser significativamente afetada pela qualidade dos dados de entrada e pela sensibilidade no ajuste de hiperparâmetros, dado que o modelo depende fortemente da seleção de amostras representativas e confiáveis (Kok *et al.*, 2021).

O algoritmo Decision Tree (DT) é um método de aprendizado supervisionado não paramétrico amplamente utilizado para tarefas de classificação. Sua estrutura baseia-se em regras do tipo "*if-then-else*", com o objetivo de classificar uma instância de dados na classe mais adequada, com base em seus atributos. A lógica fundamental por trás do DT consiste em mapear todos os possíveis caminhos de decisão na forma de uma estrutura hierárquica semelhante a um fluxograma. Cada nó interno representa um teste sobre um atributo, cada ramo indica o resultado do teste, e os nós terminais contêm os rótulos de classe. Durante o processo de predição, os dados percorrem um caminho de decisão até que uma classe seja atribuída no nó final. Esse modelo apresenta a vantagem de lidar com dados completos e incompletos, sendo eficaz em diferentes tipos de conjuntos de dados, inclusive no contexto agrícola (Waleed *et al.*, 2021; Peppes *et al.*, 2021).

O algoritmo K-Nearest Neighbours (KNN) é um dos classificadores supervisionados mais estudados em aprendizado de máquina. Trata-se de um método não paramétrico, simples, intuitivo e eficaz, utilizado tanto para tarefas de classificação quanto de regressão. O KNN funciona classificando uma nova amostra com base na classe predominante entre as k amostras mais próximas no conjunto de treinamento. A proximidade é determinada por métricas de distância, sendo as mais comuns as funções Euclidiana, de Manhattan e de Minkowski. A eficácia do modelo está diretamente relacionada à escolha de dois elementos-chave: o número de vizinhos (k) e a métrica de distância (D). Embora diversas medidas possam ser utilizadas, a distância Euclidiana é a mais frequentemente adotada na prática (Peppes *et al.*, 2021; Amkor & El Barbri, 2023).

O Multi-Layer Perceptron (MLP) é um tipo de rede neural artificial do tipo *feedforward*, amplamente utilizado em tarefas de classificação e regressão. Essa arquitetura é composta por três a cinco camadas de neurônios: uma camada de entrada, uma ou mais camadas ocultas totalmente conectadas e uma camada de saída. Cada neurônio na camada de entrada representa uma variável de entrada, enquanto as camadas ocultas processam os sinais, executando as operações não lineares responsáveis pelo aprendizado. A camada de saída gera os valores previstos, que podem ser aplicados, por exemplo, à estimativa de rendimento agrícola. A verdadeira capacidade computacional do MLP está associada ao número de unidades ocultas presentes entre as camadas. Durante o treinamento, é empregada a técnica de retropropagação de erro, um método de aprendizado supervisionado que ajusta iterativamente os pesos da rede para minimizar o erro entre os valores previstos e observados (Zhu *et al.*, 2024; Ahmed, 2023).

O eXtreme Gradient Boosting (XGB) é um algoritmo de aprendizado de máquina baseado em *boosting*, que combina múltiplas árvores de decisão para formar um modelo mais robusto e preciso. Sua principal vantagem está na capacidade de reduzir o sobreajuste e o esforço computacional, ao mesmo tempo em que melhora a generalização do modelo. O XGB foi projetado para evitar *over-fitting*, ele constrói árvores sequencialmente, com cada nova árvore corrigindo os erros da anterior, em um processo iterativo que se encerra com base em critérios como a estabilização do erro quadrático médio (RMSE). Recurso como a parada antecipada evita ajustes excessivos aos dados de treinamento, enquanto a validação cruzada auxilia na escolha dos melhores hiperparâmetros, tornando o XGB uma das técnicas mais eficazes e amplamente utilizadas em problemas complexos de classificação (Huang *et al.*, 2022; Ren *et al.*, 2023)

O Light Gradient Boosting Machine (LGBM) é um modelo de aprendizado de máquina baseado em árvores de decisão, pertencente à família dos métodos Gradient Boosting Decision Tree (GBDT). Ele se destaca pelo uso eficiente de amostragem e compactação de Gradient-based One-Side Sampling (GOSS) e Exclusive Feature Bundling (EFB), estratégias que aumentam a eficiência do treinamento. O GOSS descarta instâncias com pequenos gradientes, otimizando o uso dos dados mais informativos, enquanto o EFB agrupa variáveis mutuamente exclusivas para reduzir a dimensionalidade, acelerando o treinamento com mínima perda de acurácia. Além disso, o LGBM oferece suporte integrado para a avaliação da importância das variáveis, o que permite identificar quais atributos mais influenciam a predição do modelo (Aziz *et al.*, 2022; Bogdanovski *et al.*, 2023).

A Regressão Linear Múltipla (MLR) é um dos métodos estatísticos mais amplamente utilizados em tarefas de predição, devido à sua estrutura simples, facilidade de cálculo e

interpretação direta dos coeficientes. Trata-se de uma extensão da regressão linear simples, em que uma variável dependente é modelada em função de múltiplas variáveis independentes. A principal vantagem da MLR está na sua capacidade de quantificar a correlação entre variáveis preditoras e a resposta, especialmente quando o número de características é inferior ao número de observações. No entanto, um de seus principais limites é a suposição de linearidade entre as variáveis, o que a torna ineficaz para modelar relações não lineares com variáveis ambientais ou outras de natureza complexa. Embora métodos modernos como a seleção de características possam melhorar sua acurácia e precisão, a MLR ainda não é ideal em contextos que exigem flexibilidade para capturar padrões mais sofisticados nos dados (Xie *et al.*, 2021; Catalano *et al.*, 2022).

A avaliação do desempenho dos modelos na maioria das vezes é avaliada por meio de métricas amplamente utilizadas, incluindo raiz do erro quadrático médio (RMSE), erro absoluto médio (MAE) e o coeficiente de determinação (R^2). Enquanto RMSE e MAE quantificam o erro entre os valores previstos e observados, sendo mais desejáveis quanto menores, o R^2 mede a proporção da variância explicada pelo modelo, variando de 0 a 1, onde valores mais altos indicam melhor ajuste. Contudo, embora o R^2 forneça uma visão geral da capacidade explicativa, ele não detecta viés nas previsões e, por isso, deve ser interpretado em conjunto com as métricas de erro. Assim, a combinação dessas métricas fornece uma base sólida para comparar e selecionar os modelos com melhor desempenho preditivo (Chicco *et al.*, 2021; Liew *et al.*, 2022).

O índice de concordância de Willmott (d) é uma medida adimensional, variando de 0 a 1, que indica o grau de concordância entre os valores observados e previstos. Quanto mais próximo de 1, maior a concordância. Esse índice é sensível à presença de *outliers* e é frequentemente utilizado para avaliar a acurácia dos modelos, especialmente no que se refere à representatividade do erro relativo médio. Além disso, o valor de d também pode fornecer indícios sobre possíveis vieses nas previsões do modelo (Dos Santos *et al.*, 2022; Wilmot *et al.*, 2023).

2.6 Abordagens modernas para avaliação de frutos de morango

A adoção de tecnologias digitais atrelado ao uso de modelos de aprendizado de máquina tem viabilizado abordagens não destrutivas de avaliação, o que é especialmente relevante no cultivo do morango, onde diversos atributos são fundamentais para determinar a qualidade e o valor comercial dos frutos. Essas técnicas tornam possível o monitoramento dessas características ao longo do desenvolvimento da planta, sem interferir no seu ciclo produtivo,

favorecendo decisões mais precisas e rápidas no manejo, além de contribuir para a padronização e rastreabilidade da produção (Palumbo *et al.*, 2022; Basak *et al.*, 2022).

No cultivo do morangueiro, as características morfológicas de plantas inteiras ou órgãos específicos da planta são características fenotípicas fundamentais e importantes que refletem diretamente o estado de saúde do crescimento da planta e o suprimento de nutrientes e são altamente correlacionadas com a biomassa vegetativa e o rendimento final. A medição de características fenotípicas relevantes como altura da planta, cobertura vegetal, índice de área foliar (IAF), formato da folha, diâmetro de coroa, comprimento do pecíolo, tamanho da folha, do fruto e da flor, massa, formato e cor do fruto, entre outros, são comuns entre produtores e pesquisadores, servindo como marcadores fenotípicos empregados para monitorar o equilíbrio do crescimento da planta e gerenciar as condições de cultivo (Kirk *et al.*, 2021; Ndikumana *et al.*, 2024; Zheng *et al.*, 2024).

A caracterização dos frutos (tamanho, cor, formato, uniformidade, firmeza e brilho) é importante para o sucesso varietal, devido à associação direta com a qualidade e o valor do produto no mercado, estando diretamente associado a produtividade e previsões de produção comercial, assim como a percepção do consumidor sobre a qualidade da fruta (Feldman *et al.*, 2020; Li *et al.*, 2020; James *et al.*, 2022; Ndikumana *et al.*, 2024).

Os métodos tradicionais para mensurar características de frutos e plantas dependem de amostragem manual, visual ou de sensores proximais, que são altamente criticadas por serem subjetivas, destrutivas e propensas a erros, além de serem trabalhosos e demorados. Com isso, pesquisadores de todo o mundo buscam automatizar uma ferramenta para medição de fenótipos, que seja rápida, robusta e econômica, para acelerar o processo de seleção artificial e melhorar o gerenciamento e manipulação dos dados na cultura do morango (Zingaretti, Monfort & Pérez-Enciso, 2021; Ndikumana *et al.*, 2024; Zheng *et al.*, 2024).

A coleta manual de dados, está sendo substituída por tecnologias modernas, como sensores miniaturizados e de baixo custo, câmeras digitais, *smartphones*, drones e espectrometria de massa, entre outros. Essas ferramentas permitem a obtenção de centenas de dados fenotípicos de maneira objetiva, eficiente e econômica (Zingaretti, Monfort & Pérez-Enciso, 2021).

O maior desafio no momento é adquirir rapidamente dados de fenotipagem de plantas em larga escala com alta dimensionalidade, densidade e precisão. Muitas questões ainda permanecem sobre como definir e extrair características complexas de forma eficiente, bem como melhorar a precisão e o rendimento. A aquisição precisa e rápida de características

hereditárias de interesse é crítica para melhorar a seleção no melhoramento do morangueiro (Zheng, Abd-Elrahman & Whitaker, 2021).

Os avanços em sensores ópticos e tecnologias de imagem possibilitam a fenotipagem de alto rendimento. Em particular, as câmeras RGB têm demonstrado grande potencial na análise de recursos fenótipos e genéticos, graças à sua facilidade de uso, eficiência computacional e capacidade de fornecer informações detalhadas e analisáveis (Yang *et al.*, 2024).

O Sistema automatizado permite a coleta de dados em larga escala, sem depender diretamente de avaliadores humanos, em qualquer época do ano. A captura repetida de dados quantitativos ao longo da estação de cultivo ajudaria a identificar quais variedades ou genótipos individuais apresentam maior potencial. Essas informações poderiam ser usadas para direcionar a avaliação detalhada, aumentando tanto a intensidade da seleção quanto a precisão do processo de melhoramento genético (James *et al.*, 2022).

Alguns autores já veem utilizando modelos de aprendizado de máquina para a avaliação dos frutos de morango, como para fazer a estimativa de parâmetros relacionados à maturidade e colheita (Elsayed *et al.*, 2025), a qualidade e monitoramento pós-colheita (Zheng, Abd-Elrahman & Whitaker, 2021), formato dos frutos (Feldman *et al.*, 2020), predição da massa dos frutos (Basak *et al.*, 2022) entre outras características de importância para a cultura.

Apesar do grande potencial, a adoção da agricultura digital na cultura do morango ainda enfrenta desafios, como o custo inicial de implementação, a necessidade de conectividade em áreas rurais e principalmente a capacitação dos técnicos e produtores. No entanto, experiências bem-sucedidas em centros de pesquisa e propriedades rurais demonstram que a digitalização pode contribuir significativamente para o aumento da produtividade, da sustentabilidade e da qualidade da produção de morangos (Birner, Daum & Pray, 2021; Kolling & Rampim, 2021).

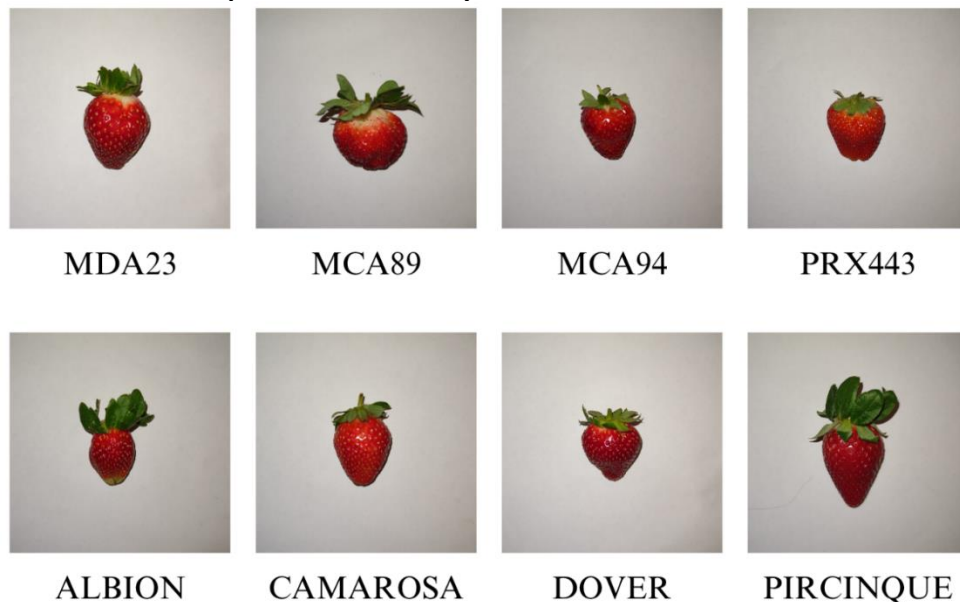
3 MATERIAL E MÉTODOS

3.1 Condução de plantas em casa de vegetação

O experimento foi realizado em casa de vegetação no município de Lavras (21°14'S, 45°00'W, a 918,8 m de altitude), no estado de Minas Gerais, Brasil. O clima da região, segundo a classificação climática de Köppen-Geiger, é considerado subtropical de altitude, com inverno seco e verão chuvoso (Alvares *et al.*, 2013).

Os genótipos utilizados para a colheita dos frutos para posterior avaliação foram: “Albion”, “Camarosa”, “Dover”, “Pircinque”, “MCA89”, “MCA94”, “MDA23” e “PRX443”, conforme exemplos na Figura 1.

Figura 1 – Imagens dos frutos dos genótipos utilizados para predição e classificação por meio de modelo de aprendizado de máquina.



Fonte: elaboração própria.

As plantas foram cultivadas no sistema semi-hidropônico, com fertirrigação, espaçadas em 25 cm entre plantas, conduzidas em delineamento em blocos casualizados (DBC), com quatro repetições, totalizando 32 parcelas (unidades experimentais) com seis plantas por parcela (unidades de observação). Em cada colheita (ocorrida escalonadamente em função da produção e maturidade dos frutos), foram selecionados cinco frutos, aleatoriamente, de cada unidade experimental para avaliação.

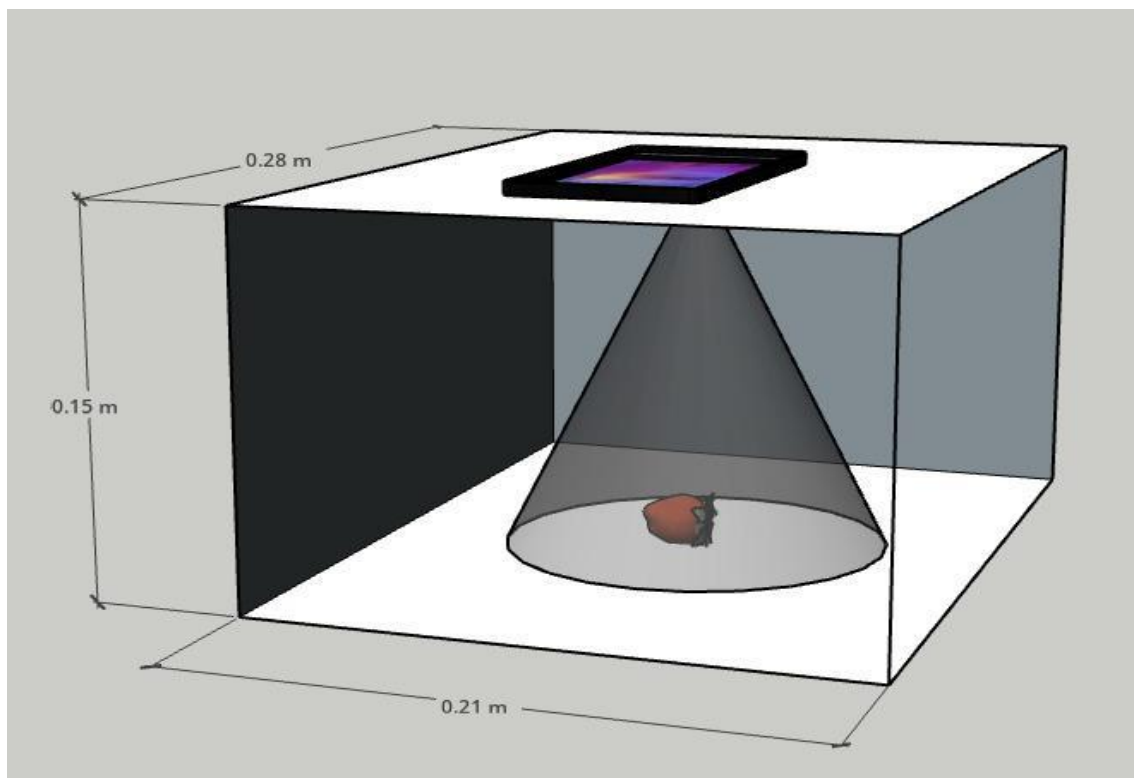
Durante o processo de colheita, todos os frutos visualmente maduros, em estágio de maturação comercial (75% de coloração vermelha) (Kilic *et al.*, 2021), foram colhidos manualmente, pela parte da manhã e armazenados em bandejas. Após, os frutos foram pegos

aleatoriamente, identificados e levados para avaliação laboratorial e de imagem, resultando em 1760 amostras avaliadas (220 frutos de oito genótipos).

3.2 Obtenção das imagens

As imagens foram capturadas em um estúdio fotográfico confeccionado com caixa de papelão ondulado de 15 cm de altura, 21 cm de largura e 28 cm de comprimento (para isolamento da luz externa), com fundo branco e iluminação apenas do flash da câmera (Figura 2). Cada amostra foi identificada com um ID para posterior análise. Todas as imagens foram capturadas pela mesma pessoa, mesmo dispositivo e mesmo ambiente, para diminuir o erro experimental.

Figura 2 – Diagrama esquemático do sistema de aquisição de imagens para frutos de morango.



Fonte: elaboração própria.

Foi utilizada câmera digital de dispositivo móvel (Redmi Note 10 Pro, Modelo M2101K6G) para capturar imagens, salvas em formato RGB, que representa a composição de cores baseada nos canais vermelho (Red), verde (Green) e azul (Blue). Esse modelo de cor aditivo é amplamente utilizado em dispositivos eletrônicos, permitindo representar diversas cores a partir da combinação das intensidades dos canais R, G e B, normalmente em uma escala de 0 a 255 para cada componente (Fan *et al.*, 2023).

O dispositivo é equipado com uma câmera traseira quad 108 + 8 + 5 + 2 MP com as seguintes características: abertura de f/1.9, foco automático, HDR, objetiva grande angular de 26mm, panorama, *phase detection* autofocus, sensor de profundidade, sensor de tamanho 1/4, tamanho de pixel 1.12 μm . Foi utilizado a câmera principal do dispositivo, no modo 108 megapixels com ultra HD ativado, usando zoom 1x, tempo de exposição de 1/30s, garantindo imagens de 12000 x 9000 pixels e resolução de 72 dpi. As imagens foram salvas no formato .JPG, com tamanho de aproximadamente 6.5 Mb.

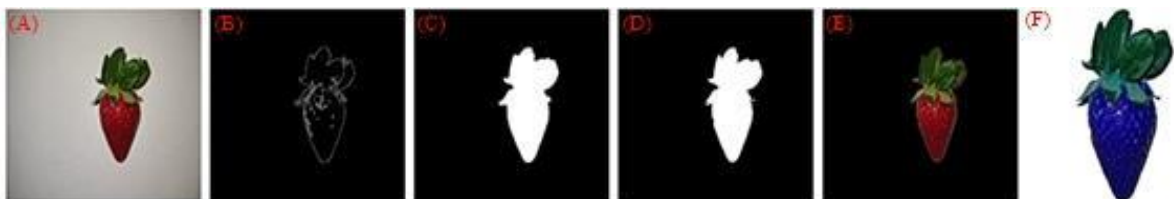
3.3 Processamento das imagens

As imagens passaram por um fluxo de processamento composto pelas etapas de reconhecimento, segmentação, remoção do fundo e extração de características dos frutos (Figura 2). Todos os procedimentos foram realizados na plataforma Google Colab, utilizando linguagem Python 3.10 com as bibliotecas OpenCV 4.8.0, NumPy, Pandas, Matplotlib e tqdm.

Inicialmente, todas as imagens foram redimensionadas para 640 x 640 pixels, garantindo padronização espacial para uma segmentação precisa. Em seguida, aplicou-se conversão para escala de cinza e suavização com filtro Gaussiano (5 x 5), a fim de reduzir ruídos e variações de iluminação. A segmentação foi realizada por meio do algoritmo Canny (limiares de 10 e 250), com refinamento das bordas via operações morfológicas de fechamento (elemento estruturante elíptico 5 x 5), repetidas duas vezes para garantir contornos contínuos.

Os contornos foram extraídos com a função *findContours*, sendo selecionado o maior contorno, considerado o fruto avaliado. A partir dele, foi gerada uma máscara binária aprimorada com operações de fechamento e abertura, suavizando imperfeições e eliminando resíduos. Para refinar a segmentação, aplicou-se a transformada de distância euclidiana na máscara, seguida de limiarização e segmentação por *watershed*, visando uma delimitação precisa entre fruto e fundo. A máscara final foi convertida em canal alfa, gerando imagens com fundo totalmente transparente. O fruto foi recortado com base na sua *bounding box* e centralizado sobre novo fundo transparente de 640x 640 pixels, mantendo sua proporção (Figura 3). As imagens resultantes foram salvas em formato PNG com canal alfa (RGBA).

Figura 3 – Visão geral do fluxo de trabalho aplicado ao processamento de imagens.



Legenda: carregamento da imagem original (A), extração do contorno (B), seguido pela aplicação de

técnicas de segmentação (C). Posteriormente, realiza-se o refinamento da máscara (D), isolamento da fruta (E) e recorte do fruto (F).

Fonte: Google Colab.

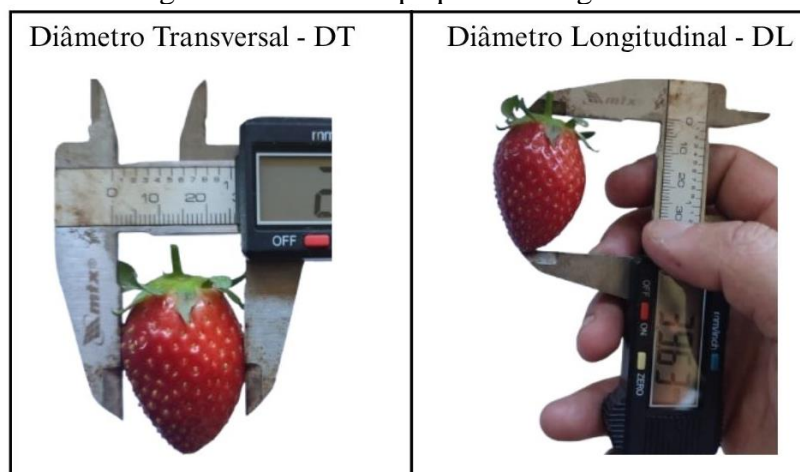
As características de cor extraídas incluíram a média e o desvio padrão dos valores R, G e B, considerando apenas os pixels dos frutos. Para a extração das características morfológicas, os canais RGBA foram separados, utilizando-se o canal alfa como máscara para delimitar a área do fruto. As características morfológicas incluíram: área (px²), perímetro (px), fator de forma ($4\pi \cdot \text{área} / \text{perímetro}$), excentricidade (a partir do ajuste de elipse ao contorno), altura (px), largura (px) via *bounding box* e razão largura/altura, como indicador de proporção.

Os dados extraídos foram organizados em um *Dataframe* do Pandas, ordenado pelo ID da imagem (nome do arquivo) e exportado em formato Excel.

3.4 Análises morfológicas quantitativas e qualitativas

Após a captura da imagem, cada fruto foi pesado (balança analítica) e mensurado manualmente o diâmetro transversal (DT) e diâmetro longitudinal (DL), com auxílio de paquímetro digital (Figura 4). Posteriormente os frutos foram classificados de acordo com os descritores morfológicos propostos por UPOV (2012), conforme a Tabela 1.

Figura 4 – Representação das avaliações de diâmetro longitudinal e diâmetro transversal dos frutos de morango com auxílio de paquímetro digital.



Fonte: elaboração própria.

Tabela 1 – Características dos frutos avaliados segunda os descritores morfológicos propostos por UPOV (2012) em escala de notas.

Característica	Forma de avaliação
Tamanho	1- Muito pequeno; 3- pequeno; 5- médio; 7- grande; 9- muito grande
Brilho	1- Sem brilho; 2- fraco; 3- moderado; 4- médio; 5- forte
Cavidade	1- Ausente; 2- pequeno; 3- médio; 4- grande 5- muito grande
Superfície	1- Uniforme; 2- ligeiramente irregular; 3- irregular; 4- fortemente irregular; 5- desigual
Cor	1- Rosa claro; 2- laranja claro 3- laranja médio; 4- vermelho fraco; 5- vermelho médio; 6- vermelho escuro; 7- vermelho negro
Uniformidade de Cor	1- Uniforme; 2- ligeiramente irregular; 3- irregular; 4- fortemente irregular; 5- desigual
Defeitos	1- Sem defeitos; 3- levemente danificado 5- defeitos graves

Fonte: Adaptada de UPOV (2012).

Esses dados foram integrados às informações extraídas das imagens por meio da coluna "ID", que associou cada fruto às suas respectivas características, para a realização das análises.

3.5 Correlação de Spearman

A análise de correlação de Spearman foi realizada para observar as relações entre as variáveis dos dois conjuntos de dados (extraídos das imagens e mensurados manualmente), e verificar a força e a direção entre os principais parâmetros. Os resultados são apresentados em um gráfico de matriz de correlação (*heatmap*), oferecendo uma visão clara das interdependências entre as variáveis.

3.6 Modelagem preditiva e classificação

Para a predição da massa dos frutos de morango, as variáveis foram selecionadas de acordo com a análise de correlação, selecionando as correlações significativas para inclusão potencial nos modelos preditivos. Assim, foram utilizados dois conjuntos de variáveis predictoras: análises manuais (DT, DL e tamanho do fruto); e análises de imagem (área (px²), perímetro (px), altura e largura (px)). As variáveis foram combinadas com a variável resposta massa dos frutos. Cada conjunto foi composto por 1760 observações (frutos).

Com base nisso, os conjuntos de dados foram divididos em 70% para treinamento e 30% para teste emparelhado, com validação cruzada do tipo *K-fold* ($k = 5$). Em cada iteração, um dos subconjuntos foi utilizado para teste, enquanto os demais serviram para o treinamento, garantindo que todos os dados fossem utilizados em ambas as etapas ao longo das cinco repetições (k) (Barboza *et al.*, 2025). O método utilizado foi *train_test_split* do pacote *scikit-learn*, com estratificação e randomização. A análise empregou oito algoritmos de aprendizado de máquina, que estão apresentados na Tabela 2.

Tabela 2 – Parâmetros principais utilizados na configuração dos modelos de aprendizado de máquina para predição e classificação de frutos de morango.

Modelo	Sigla	Parâmetros
Support Vector Machine	SVM	padrão via SVR; kernel = 'rbf'; C=1.0; epsilon = 0.1; gamma = 'scale';
Random Forest	RF	n_estimators = 100; random_state = 42; criterion = 'gini'.
Decision Tree	DT	max_depth = none; random_state = 42; criterion = 'gini'.
K-Nearest Neighbours	KNN	n_neighbors = 5; weights = 'uniform'; algorithm = 'auto'; leaf_size = 30; metric = 'minkowski', p = 2.
Multi-Layer Perceptron	MLP	hidden_layer_sizes = 100; activation = 'relu'; alpha = 0.0001; learning_rate = 'constant'; max_iter = 500; random_state = 42.
eXtreme Gradient Boosting	XGB	n_estimators = 100; learning_rate = 0.3; max_depth = 6; eval_metric = 'mlogloss'; use_label_encoder = false; verbosity = 1.
Light Gradient Boosting Machine	LGBM	boosting_type = 'gbdt'; num_leaves = 31; learning_rate = 0.1; n_estimators = 100; random_state = none.
Multilinear Regression	MLR	fit_intercept = true; copy_X = true; n_jobs = none; positive = false

Fonte: elaboração própria.

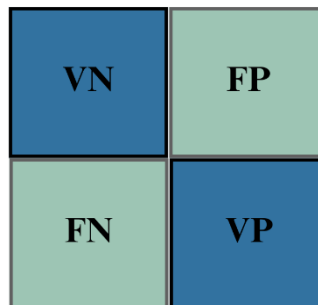
Para avaliação da performance dos modelos foram utilizadas as métricas R^2 (coeficiente de determinação), MAE (erro absoluto médio) e RMSE (raiz quadrada do erro quadrático médio), calculadas com as funções *r2_score*, *mean_absolute_error* e *mean_squared_error* do *scikit-learn* (Manvitha *et al.*, 2023).

Para a classificação dos frutos em comerciais ou não comerciais a variável alvo foi criado com base em critérios objetivos relacionados à qualidade dos frutos, sendo comercial frutos que apresentassem massa $\geq 10g$ e não comerciais $< 10g$ (Zanin *et al.*, 2019), posteriormente transformada em variável binária (1 para comercial, 0 para não comercial).

Para a classificação binária dos frutos, utilizamos novamente os oito algoritmos supervisionados: SVM, RF, DT, KNN, MLP, XGB, LGBM e MLR, com a mesma distribuição dos dados para treinamento (70%) e teste (30%). As métricas para avaliar o desempenho dos modelos foram: Acurácia, indica proporção de predições corretas positivas e negativas em relação ao total de observações; Precisão: indica a proporção de predições positivas que foram corretamente classificadas como positivos; Recall: mede a capacidade do modelo em identificar corretamente todas as instâncias reais da classe positiva; e F1-score: é a média harmônica entre precisão e recall e indica o equilíbrio entre evitar falsos positivos e falsos negativos (Nithya *et al.*, 2024).

Para calcular as métricas, é necessário construir a matriz de confusão, usadas para visualizar o desempenho do algoritmo. As matrizes de confusão foram geradas para cada modelo, separadamente para cada conjunto de dados. As células em cada matriz representam as classificações: no canto superior esquerdo, os verdadeiros negativos – VN (não comercial correta); no superior direito, os falsos positivos – FP (não comercial como comercial); no inferior esquerdo, os falsos negativos – FN (comercial como não comercial); e no inferior direito, os verdadeiros positivos – VP (comercial correta), conforme Figura 5 (Nithya *et al.*, 2024). Assim, quanto maiores os números na diagonal principal, melhor o desempenho da classificação (Yuan *et al.*, 2023).

Figura 5 – Esquema das matrizes de confusão dos modelos de aprendizado supervisionado para a classificação de frutos comerciais ou não comerciais



Legenda: VN: Verdadeiros negativos; FP: Falsos positivos; FN: Falsos negativos e VP: Verdadeiros positivos.

Fonte: elaboração própria.

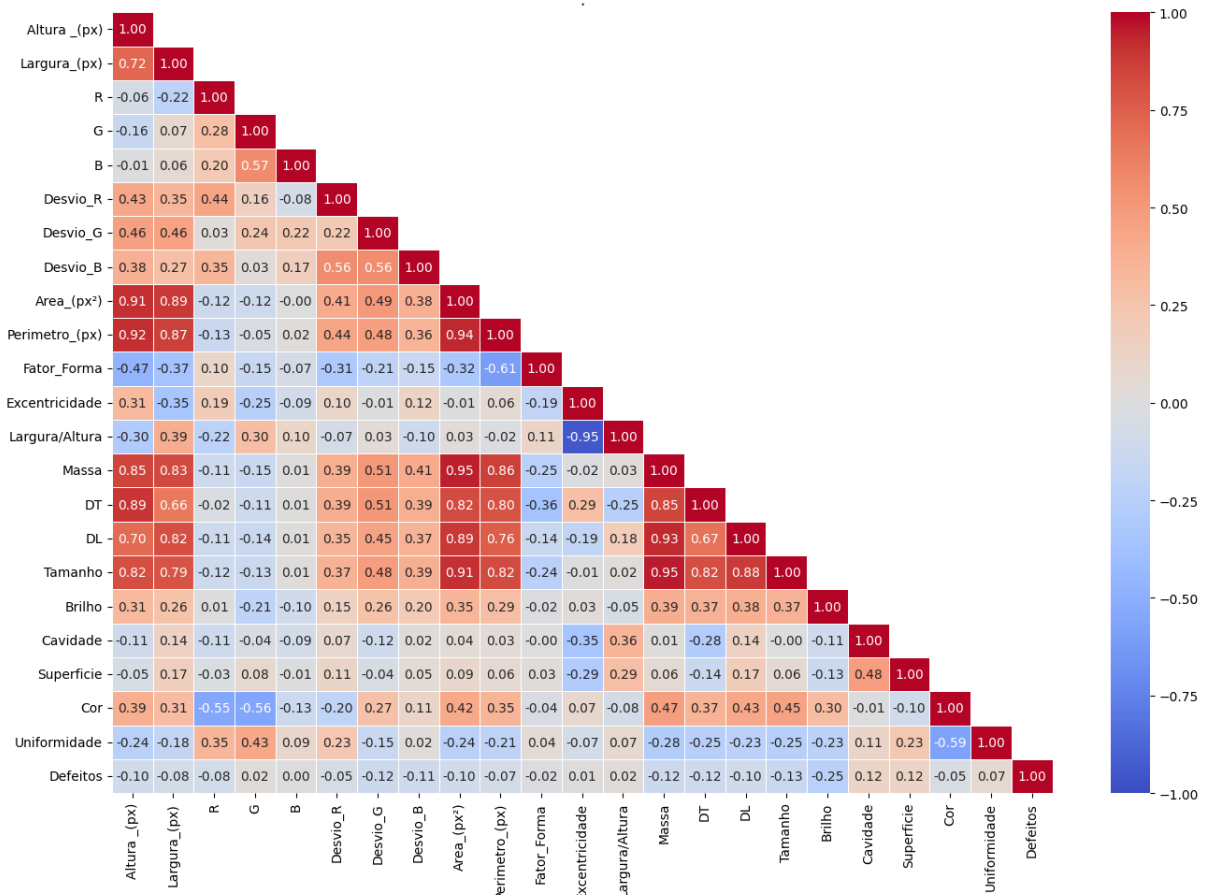
Assim, quanto maiores os números na diagonal principal, melhor o desempenho da classificação (Yuan *et al.*, 2023). Os resultados foram visualizados por meio de gráficos e tabelas que permitiram avaliar visualmente o comportamento dos algoritmos, a tendência e dispersão dos erros, bem como a proximidade das predições em relação aos valores reais.

Detalhes técnicos: Todas as análises foram realizadas utilizando a linguagem de programação Python (versão 3.10) no ambiente Google Colab e *software* RStudio[®]. Os dados foram importados a partir do Google Drive, com planilhas no formato .xlsx. As análises foram executadas em um notebook com Windows 10 Home Single Language, processador Intel[®] Core[™] i5-5200U, 4.00 GB de RAM, SSD de 256 GB e placa de vídeo Intel[®] HD Graphics 5500 (128 MB).

4 RESULTADOS

A análise de correlação indica a relação entre as variáveis extraídas das imagens e os parâmetros de qualidade da fruta. O gradiente de cor, passando de azul para vermelho, indica a força da correlação: azul representa uma correlação negativa alta (próxima de -1), branco significa uma correlação fraca (próxima de 0) e vermelho denota uma correlação positiva alta (próxima de 1) (Figura 6).

Figura 6 – Análise de correlação de Spearman entre os dados das imagens e os parâmetros de qualidade dos frutos.



Fonte: elaboração própria.

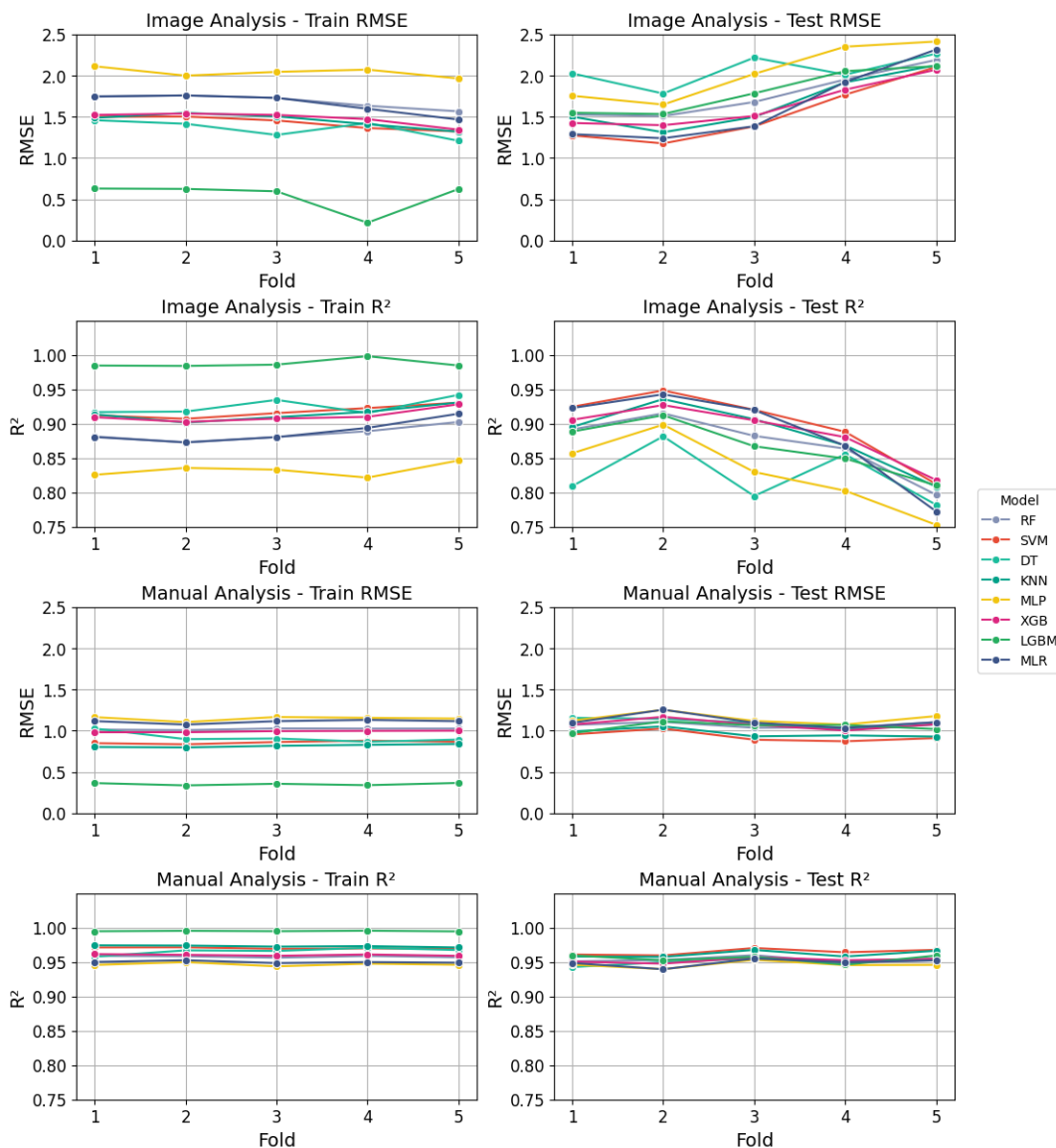
A análise de correlação de Spearman foi realizada com o objetivo de investigar as correlações lineares entre variáveis morfológicas, espectrais e qualitativas dos frutos, com destaque para aquelas que podem influenciar a massa, uma variável-chave em estudos de produtividade e classificação comercial.

Notavelmente, a massa apresenta correlações positivas muito fortes com características físicas e dimensionais do fruto DL (0,93), DT (0,85), tamanho (0,95), área (px²) (0,95), perímetro (px) (0,86), altura (px) (0,85) e largura (px) (0,83), evidenciando que frutos maiores

e mais volumosos tendem a apresentar maior peso. Em contrapartida, não foram observadas correlações significativas entre a massa e as componentes de cor (R, G, B), sugerindo que a cor dos frutos tem pouca ou nenhuma relação linear com a massa dos frutos. Esses resultados foram oportunos para identificar as características significativas para inclusão na predição da massa dos frutos através dos modelos de aprendizado de máquina.

Com respeito a avaliação dos modelos (Figura 7), durante todo o processamento de dados, apresentaram melhor desempenho na fase de treinamento, comparativamente a fase de teste, onde os valores de RMSE são menores e o ajuste (R^2) maior. Observou-se, nessa fase, que o modelo LGBM se distancia dos demais com o menor valor de RMSE e maior valor de R^2 , para ambos os conjuntos de dados, assim como o MLP tem maiores RMSE e menores R^2 .

Figura 7 – Desempenho dos modelos de aprendizado de máquina na predição da massa de frutos de morango via validação cruzada em cinco *fold*s, no conjunto de análises manuais e de imagens.

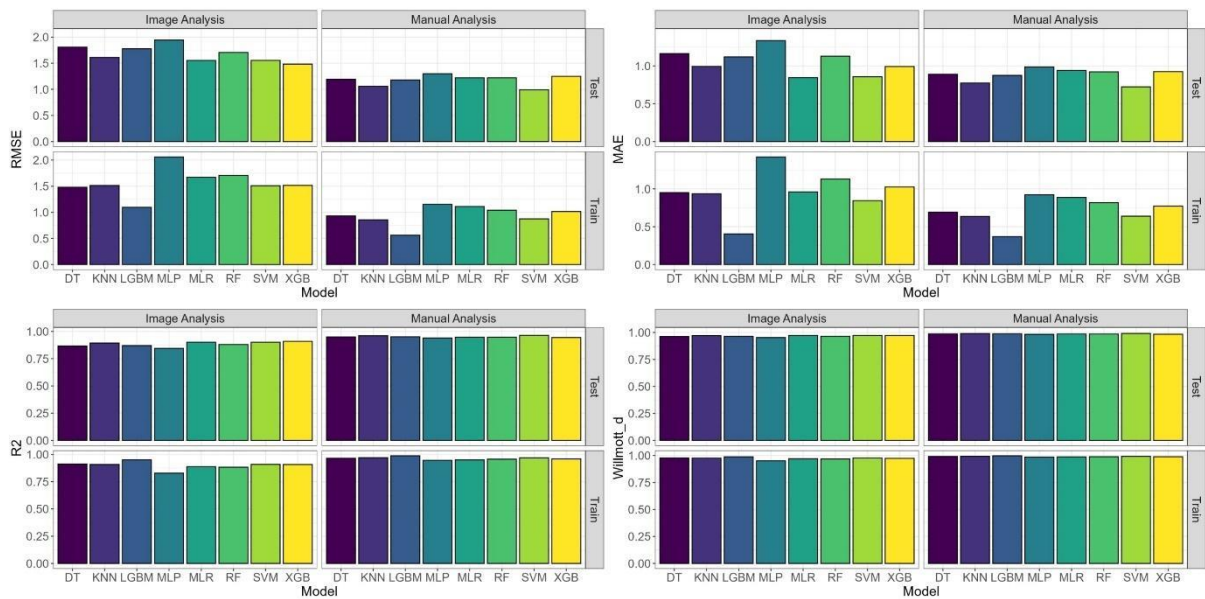


Fonte: elaboração própria.

Na análise dos modelos testados observou-se que, com os dados manuais, os modelos mantêm desempenho estável entre os *folds*, tanto em termos de erro (RMSE) quanto de coeficiente de determinação (R^2). Em contraste, os modelos treinados com dados extraídos de imagem apresentaram maior variabilidade entre os *folds*, especialmente nos conjuntos de teste, onde se observa que a partir do terceiro *fold* os valores de RMSE aumentaram e os de R^2 diminuíram (Figura 7).

Com relação aos conjuntos de dados, os modelos propostos apresentaram elevados valores de R^2 e índice d (Figura 8), combinados a baixos valores de erros (RMSE e MAE), indicando boa capacidade preditiva para a massa dos frutos de morango.

Figura 8 – Desempenho comparativo dos modelos de aprendizado de máquina aplicados aos dados manuais e de imagem, avaliados nos conjuntos de treinamento e teste.



Fonte: elaboração própria.

Observou-se similaridade entre os resultados na fase de treino e teste, sugerindo que os modelos testados estão bem generalizados. Contudo, as performances apresentaram diferenças entre os dois conjuntos de dados avaliados (manual e imagem) destacando variações na adaptabilidade, eficiência e precisão dos modelos na tarefa de predição (Figura 8).

Quando se analisa os dados de teste, o modelo SVM foi o que melhor se ajustou aos dados, com métricas de $R^2 = 0,96$; RMSE = 0,99; MAE = 0,72 e d = 0,99, para os dados manuais e $R^2 = 0,90$; RMSE = 1,55; MAE = 0,86; d = 0,97, para os dados extraídos das imagens. Embora o método manual apresente menor erro absoluto médio (MAE = 0,72), o método por imagem (MAE = 0,86) se mostrou competitivo, com uma diferença de apenas 16,28%.

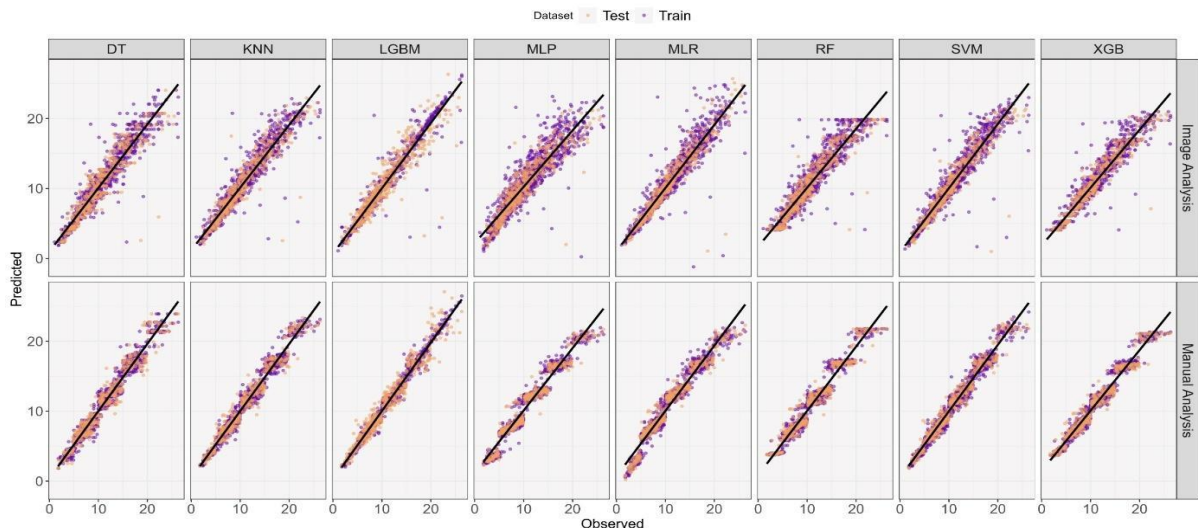
Ao conjunto de dados manuais o KNN ($R^2 = 0,96$; RMSE = 1,06; MAE = 0,77; $d = 0,99$), também apresentou bom ajuste, assim como para os dados de imagens, o modelo XGB ($R^2 = 0,91$; RMSE = 1,48; MAE = 0,99; $d = 0,97$), obteve desempenho similar ao SVM (Figura 8).

Por outro lado, o MLP foi o modelo que apresentou o menor desempenho de ajuste, especialmente na fase de teste, com os menores valores de R^2 e os maiores erros preditivos (Figura 8). Ao compararmos os valores de MAE entre os dois tipos de dados, observa-se que, o uso dos dados de imagem resultou em um aumento de erro de 25,56% em relação aos dados manuais, para esse modelo.

O modelo LGBM apresentou as melhores métricas na fase de treinamento ($R^2 = 0,99$; RMSE = 0,56; MAE = 0,37 e $d = 1,00$), mas perdeu desempenho no teste ($R^2 = 0,87$; RMSE = 1,78; MAE = 1,12; $d = 0,96$), sugerindo sobre ajuste dos dados. Esse comportamento sugere que o modelo aprendeu excessivamente os padrões específicos do conjunto de treino, com menor capacidade de generalização. Os resultados estão de acordo com os apresentados na Figura 7 e indicam um aumento de 66,96% no erro absoluto médio (MAE) entre as fases de treinamento e teste.

A avaliação visual da capacidade preditiva de cada modelo em relação às abordagens de análise manual e análise de imagem foi realizada por meio de gráficos de dispersão observado *versus* predito (Figura 9). A linha de 1:1, em cada gráfico serve como referência para a predição perfeita, onde um alinhamento ideal dos pontos indica que os valores preditos correspondem aos valores observados.

Figura 9 – Análise comparativa do desempenho preditivo dos modelos, observado *versus* predito, nos dois conjuntos de dados (imagem e manual).



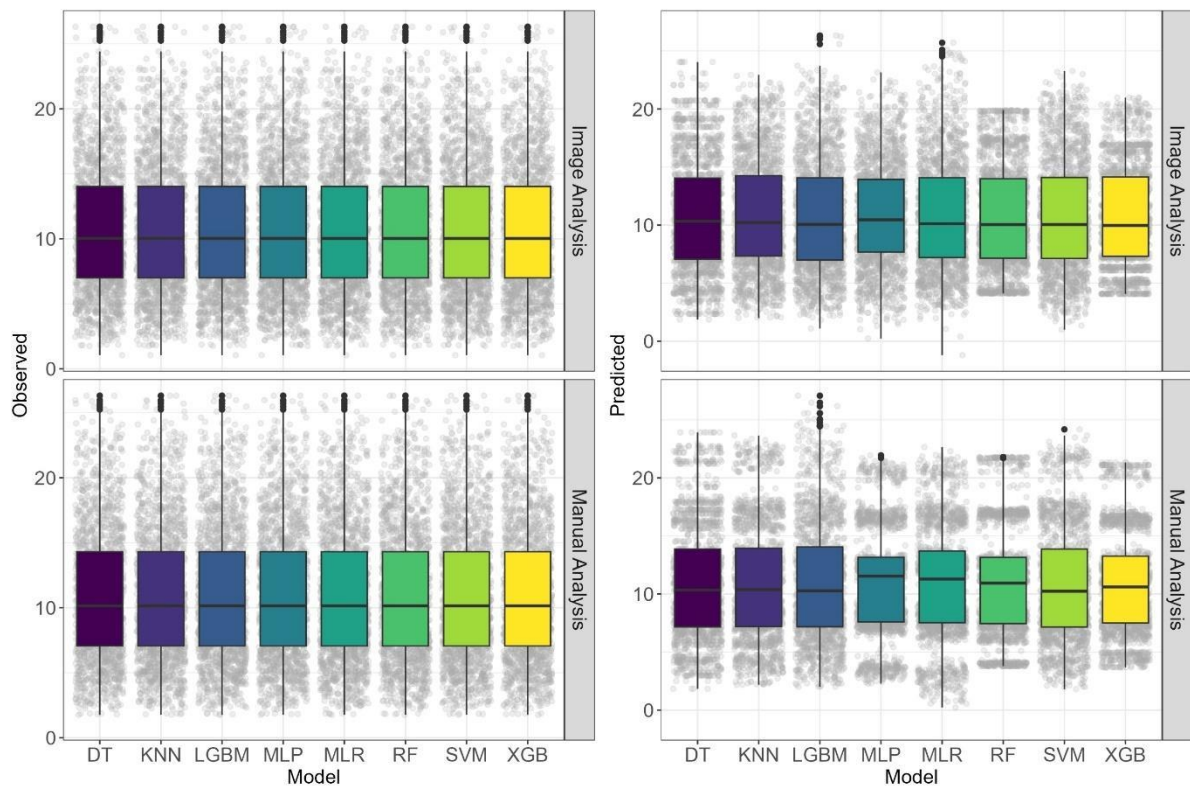
Fonte: elaboração própria.

Os modelos KNN, LGBM, SVM e XGB exibiram maior precisão preditiva. Seus pontos de dados foram caracterizados por uma concentração elevada e baixa dispersão em torno da linha 1:1, indicando um alinhamento superior entre valores observados e preditos. Para esses modelos, o comportamento dos pontos do conjunto de teste foi muito similar ao do conjunto de treinamento, indicando uma boa capacidade de generalização e evidência de bom ajuste. Não foram identificados vieses sistemáticos pronunciados (tendências consistentes de superestimação ou subestimação) em faixas específicas de valores (Figura 9).

Para a maioria dos modelos o conjunto de dados extraídos das imagens gerou resultados preditivos visualmente comparáveis ao conjunto de dados manual, demonstrando a eficácia da abordagem automatizada e potencial de uso como ferramenta de automação da predição da massa para a cultura do morango.

A distribuição estatística dos valores observados e dos valores preditos por cada modelo de aprendizado de máquina, sob as abordagens de análise de imagem e análise manual, estão representados na Figura 10.

Figura 10 – Comparação da distribuição de valores observados e preditos por modelos de aprendizado de máquina, nos dois conjuntos de dados.



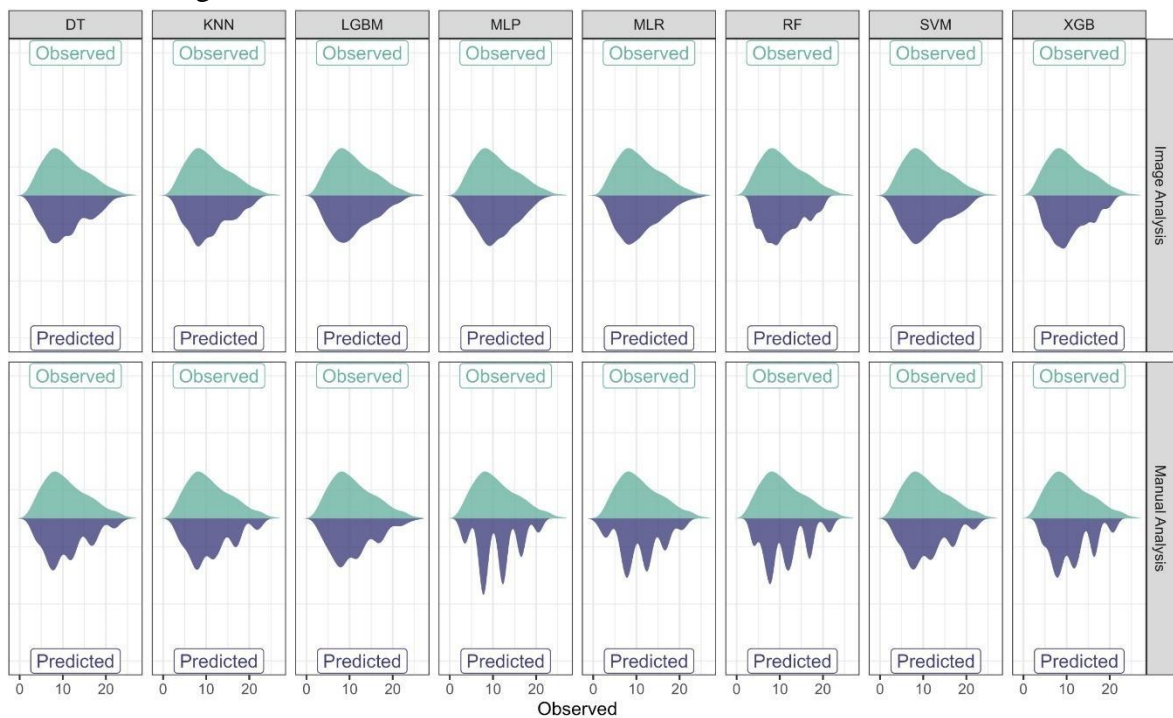
Fonte: elaboração própria.

À esquerda observa-se os box plots que representam os valores observados das análises de imagem e manual e a direita os preditos. Os pontos cinzas adjacentes a cada box plot representam os dados individuais, fornecendo uma visão de dispersão dos valores (Figura 10).

Observa-se que os box plots são muito próximos, para os dois conjuntos de dados (manual e imagem), o que demonstra que a distribuição dos dados reais se assemelha para ambas as abordagens. Quando analisado os dados de amplitude interquartil (altura da caixa), extensão dos *whiskers* e densidade de pontos, os modelos KNN, LGBM, SVM e XGB, observou-se box plots preditos bem alinhados com a distribuição dos valores observados, sugerindo alta fidelidade e replicando bem a dispersão quanto à distribuição real dos dados.

Comparação detalhada da densidade de distribuição dos valores observados (metade superior em verde) e dos valores preditos (metade inferior em roxo) para cada modelo de aprendizado de máquina, pode ser observado na Figura 11.

Figura 11 – Comparação da densidade de distribuição de valores observados e preditos por modelos de aprendizado de máquina em dois conjuntos de dados, manual e imagem.



Fonte: elaboração própria.

A variável alvo é representada pela distribuição dos dados observados e serve como padrão para o qual as predições dos modelos devem se espelhar, sendo assim, quanto mais próximos em termos de forma, altura do pico, largura da base e extensão das caudas forem as curvas de predição em relação as curvas observadas, melhor é o ajuste do modelo aos dados (Figura 11).

Para os modelos KNN, LGMB e SVM, em ambos os conjuntos de dados, as curvas de densidade dos valores preditos são mais semelhantes às curvas dos valores observados. Eles conseguem reproduzir com certa precisão a posição do pico principal da distribuição, a largura da base (indicando a dispersão geral) e a forma das "caudas" (representando a frequência de valores mais extremos). Entretanto, os modelos MLP, MLR e RF, principalmente no conjunto manual, apresentaram curvas preditas com mais picos, ou múltiplas ondas, indicando compressão dos valores preditos e certa falha em capturar a dispersão real e a suavidade da distribuição observada. Isso sugere que suas predições tendem a se agrupar em torno de um número limitado de valores (Figura 11).

O desempenho dos modelos de aprendizado supervisionado na classificação de frutos comerciais quanto a massa, utilizando os dois conjuntos de dados, manual e imagem, estão representados na Tabela 3.

Tabela 3 – Desempenho dos modelos de aprendizado supervisionado na classificação de frutos comerciais quanto a massa, utilizando os dois conjuntos de dados, campo e imagem.

Modelo	Manual				Imagem			
	Acurácia	F1 - score	Recall	Precisão	Acurácia	F1 - score	Recall	Precisão
MLR	0,96	0,96	1,00	0,93	0,92	0,93	0,94	0,92
MLP	0,98	0,98	1,00	0,95	0,83	0,81	0,71	0,96
DT	0,98	0,98	0,98	0,97	0,91	0,91	0,91	0,92
SVM	0,98	0,98	0,99	0,97	0,92	0,92	0,92	0,93
RF	0,97	0,97	0,98	0,97	0,91	0,92	0,93	0,91
XGB	0,98	0,98	0,98	0,97	0,91	0,92	0,93	0,91
KNN	0,96	0,96	0,97	0,95	0,92	0,92	0,94	0,90
LGBM	0,98	0,98	0,99	0,98	0,91	0,91	0,92	0,90

Fonte: elaboração própria.

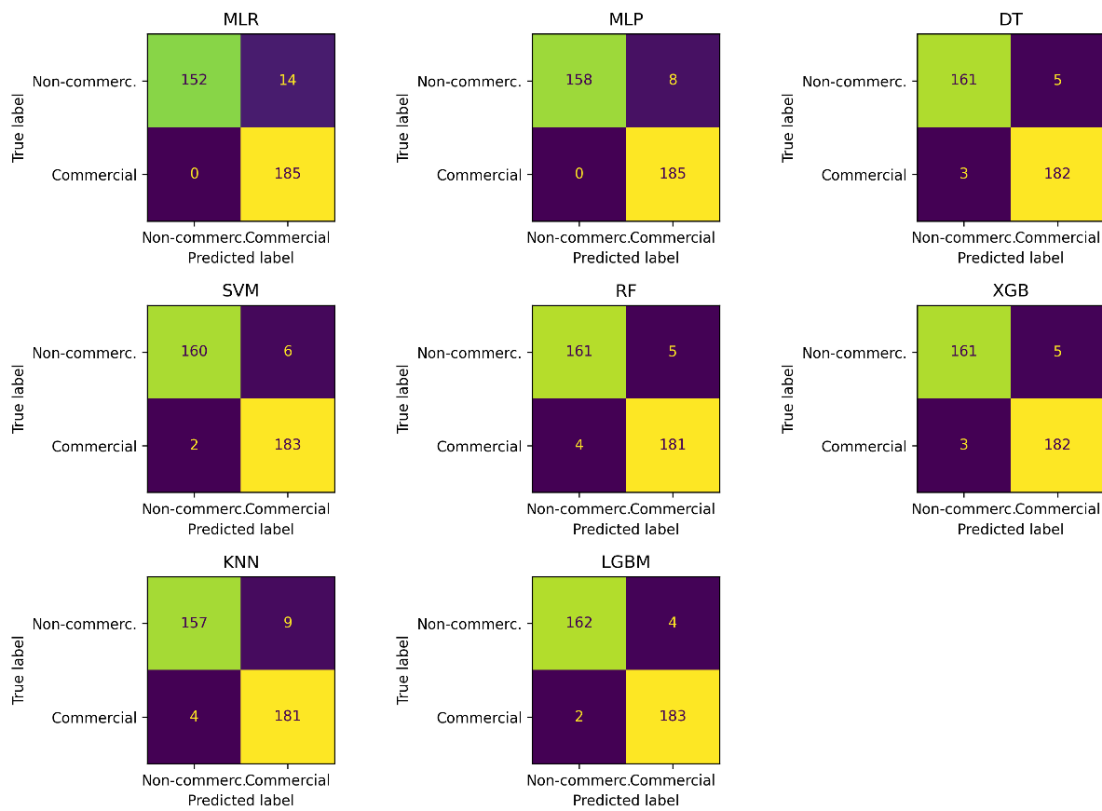
Todos os modelos apresentaram ajuste de classificação com os dados manuais, com as métricas acima de 0,96, indicando que os modelos apresentaram conformidade, exatidão, eficiência e equilíbrio, o que sugere o seu uso para classificar frutos de morango. Ressalta-se os modelos MLP, DT, SVM, XGB e LGBM, que atingiram 0,98 de acurácia e significativas métricas (Tabela 3).

Apesar de pequena queda de desempenho dos modelos quando testados com os dados extraídos das imagens, os valores de acurácia, F1-score, recall e precisão foram altos, acima de

0,91 para todos os modelos, exceto para o modelo MLP, que apresentou 0,83, 0,81 e 0,71 de acurácia, F1-score e recall, respectivamente, demonstrando menor potencial de uso na classificação de frutos de morango com esse conjunto de dados.

As matrizes de confusão de cada modelo avaliado nos conjuntos de dados manual e imagem, estão apresentadas nas Figuras 12 e 13. Em cada matriz, o eixo vertical representa a classe real (True label) e o eixo horizontal a classe predita (Predicted label), sendo as classes não-comercial ("Non-commerc.") e comercial (Commercial) (Figuras 12 e 13).

Figura 12 – Matrizes de confusão dos modelos de aprendizado supervisionado para a classificação de frutos comerciais ou não comerciais com o conjunto de dados manual.

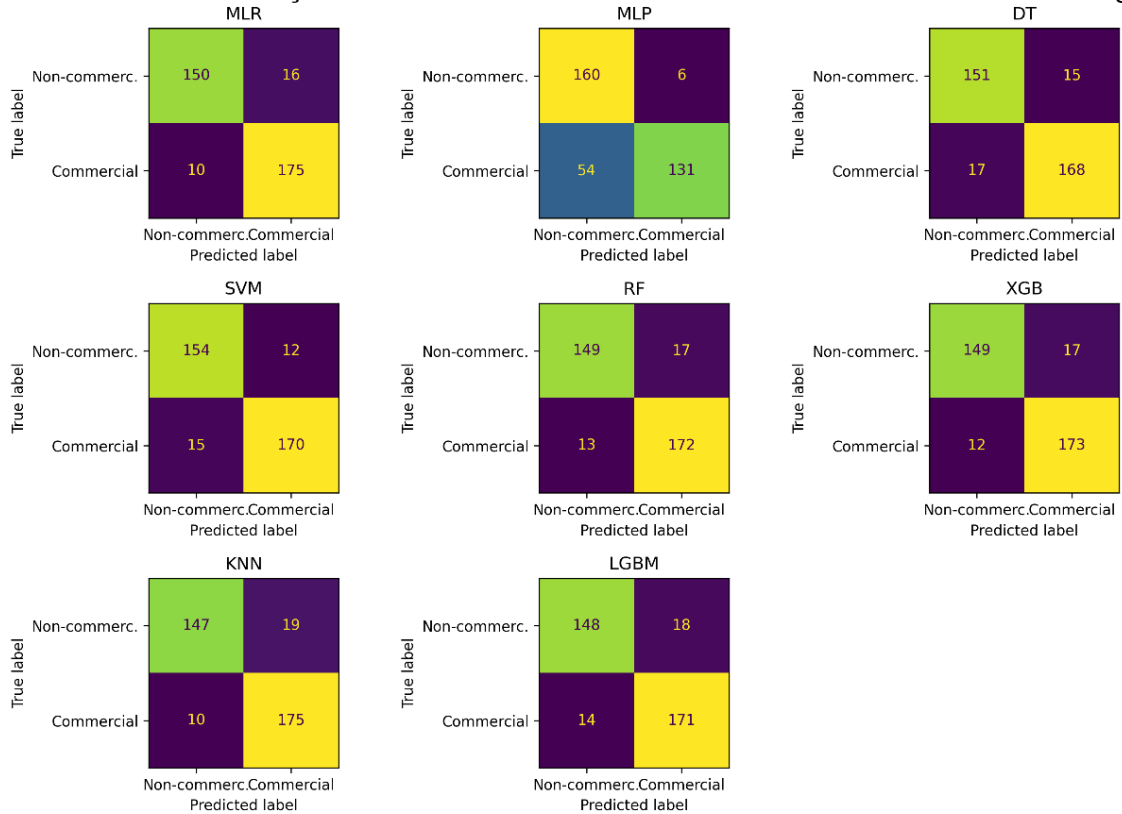


Fonte: elaboração própria.

No conjunto de dados manuais (Figura 12), observou-se que os modelos apresentaram desempenhos semelhantes na tarefa de classificação dos frutos. Modelos como DT, SVM, RF, XGB e LGBM, demonstraram taxas de acerto e comportamento de erros próximas. Já os modelos MLR, MLP e KNN, apresentaram tendência maior de erro na forma de falsos positivos, ou seja, casos em que frutos não comerciais foram classificados como comerciais, embora a quantidade total de erros ainda seja baixa.

A análise das matrizes de confusão referente ao conjunto de dados das imagens (Figura 13), revela que os modelos erraram mais, apresentando perfis de erros distintos para cada modelo, fornecendo informações importantes além das métricas já apresentadas.

Figura 13 – Matrizes de confusão dos modelos de aprendizado supervisionado para a classificação de frutos comerciais com as características extraídas das imagens.



Fonte: elaboração própria.

Os modelos SVM, RF e XGB destacam-se por alcançarem equilíbrio favorável entre falsos positivos e falsos negativos. Os erros ficaram em níveis relativamente baixos, indicando performance mais robusta e balanceada e com isso, eficácia em identificar corretamente a classe comercial e evitar classificar indevidamente a classe não comercial. Os modelos DT e LGBM, também apresentaram o mesmo comportamento de equilíbrio descrito anteriormente. O MLR foi o que menos apresentou erros, porém mantém com a tendência de falsos positivos (Figura 13).

O modelo MLP exibe número baixo de falsos positivos (6), indicando que quando prevê os frutos como comercial é altamente preciso (Precisão = 0,96). Entretanto, simultaneamente apresenta um número relativamente alto de falsos negativos (54). Isso demonstra que o MLP tem capacidade limitada de identificar a maioria das observações comerciais reais,

classificando-as erroneamente como não comerciais, resultando em um recall consideravelmente baixo (0,71) para essa classe (Figura 10).

Já os modelos MLR e o KNN mostram uma tendência inversa ao MLP em termos de erros. Ambos têm um número baixo de falsos negativos (10), mostrando boa capacidade de cobrir a classe comercial (alto Recall). Porém, apresentam número maior de falsos positivos (16 e 19, respectivamente), sugerindo que eles tendem a realizar mais identificações indevidas de frutos não comerciais como comerciais. Diante disso, a análise das matrizes de confusão é essencial para escolher o melhor modelo, já que a decisão ideal depende de qual erro é mais custoso ou indesejável na prática (Tabela 3 e Figura 13).

5 DISCUSSÃO

Os resultados da análise de correlação evidenciam que as variáveis extraídas das imagens, são altamente associadas à massa de frutos, demonstrando forte potencial para aplicações em sistemas de predição não destrutiva baseados em visão computacional. As características físicas das frutas guardam uma relação direta com sua massa e volume, o que permite utilizá-las como indicadores de qualidade. Segundo Mansuri *et al.* (2022), a estimativa precisa da massa de frutas requer a consideração conjunta de diversas propriedades físicas, evidenciando que modelos baseados em múltiplos atributos fornecem melhores resultados. As correlações fracas entre massa e as bandas RGB são esperadas, pois a cor superficial dos frutos, não reflete sua estrutura física. Enquanto as variáveis RGB capturam apenas aspectos visuais, a massa está relacionada ao volume e ao tamanho, melhor representados por medidas morfológicas como área, altura, largura e diâmetros.

A comparação entre os dois conjuntos de dados, demonstra que o conjunto manual oferece base mais robusta e estável para a modelagem preditiva da massa dos frutos. Isso é evidenciado pela menor variação entre os *folds* e pelos melhores valores médios de RMSE e R^2 , indicando maior precisão e capacidade de generalização dos modelos. Ainda assim, as variáveis extraídas das imagens também apresentaram resultados promissores, com desempenho consistente, embora ligeiramente inferior.

A segmentação das imagens, com o algoritmo aprimorado de Canny, foi muito importante, pois contribuiu para melhor detecção de bordas e melhor efeito antirruído. Estes dados corroboram com os observado por Yu *et al.*, 2021 na detecção de bordas de produtos agrícolas. Essa abordagem resultou em uma predição de massa mais acurada, uma vez que a área está diretamente relacionada ao tamanho dos morangos. As variáveis de imagem consideradas, contribuíram para a melhoria dos modelos, reduzindo os erros de predição e reforçando a importância da área (px^2) como variável-chave na predição da massa de forma não destrutiva (Barbosa Júnior *et al.*, 2025).

Considerando que a massa está relacionada ao tamanho, densidade e forma do fruto (sendo que tamanho e forma podem ser estimados por imagens), o uso de imagens associado ao aprendizado supervisionado de máquina permite prever a massa de forma não destrutiva (Jeong *et al.*, 2024).

O uso de modelos de aprendizado de máquina demonstrou-se eficaz para a predição da massa de frutos de morango. O modelo SVM apresentou o melhor ajuste para ambos os conjuntos de dados, com R^2 de 0,96 e 0,90, RMSE de 0,99 e 1,55, MAE de 0,72 e 0,86 e índice

de concordância (d) de 0,99 e 0,97, na fase de teste. O ajuste superior do SVM pode ser atribuído à sua robustez no tratamento de dados e à alta capacidade de generalização, mesmo diante de padrões complexos. Ele é eficaz em conjuntos de dados pequenos, com alta dimensionalidade e estrutura linear, o que justifica seu bom desempenho frente a modelos como MLP e XGB (Zheng *et al.*, 2022; Kapari *et al.*, 2024).

Em contraste, o modelo MLP apresentou desempenho inferior nos dois conjuntos de dados em comparação aos demais modelos, o que também foi observado por Raj *et al.*, (2022) ao estimar o teor de água em morango e Liu *et al.*, (2014) ao determinar a qualidade e maturação de frutos. Esses resultados são possivelmente devido à sua maior sensibilidade à parametrização e à necessidade de grandes volumes de dados para um treinamento eficaz, o que pode ter comprometido sua capacidade de generalização neste estudo (Madhavi *et al.*, 2022; Raj *et al.*, 2022).

O algoritmo K-Nearest Neighbors (KNN) apresentou desempenho elevado, especialmente no conjunto de dados manuais ($R^2 = 0,96$; RMSE = 1,06; MAE = 0,77; d = 0,99), na fase de teste. O ajuste do modelo aos dados pode ser atribuído à sua natureza não paramétrica, que permite flexibilidade na adaptação aos dados. Além disso, a simplicidade e consistência das variáveis manuais favoreceram previsões mais precisas pelos vizinhos mais próximos, equilibrando precisão e generalização (Saglam & Cetin, 2022; Sudipa *et al.*, 2024).

Ao conjunto de dados de imagens o modelo XGB demonstrou ajuste significativo ($R^2 = 0,91$, RMSE = 1,48, MAE = 0,99 e d = 0,97, no teste), possivelmente devido à sua estrutura sequencial de árvores de decisão, que capta padrões complexos. Sua alta precisão, velocidade de treinamento, tolerância a ruídos e alta dimensionalidade explicam sua eficácia na predição da massa via visão computacional (Barboza *et al.*, 2025; Wu *et al.*, 2025).

Vale ressaltar que o modelo LGBM teve ajuste superior em ambos os conjuntos de dados na fase de treino, mas sofreu queda de desempenho na fase de teste, sugerindo sobreajuste. Isso pode ser devido a este modelo se ajustar excessivamente às particularidades do conjunto de treino, capturando ruídos e padrões específicos que não se repetem em dados novos. Apesar disso, destaca-se por sua eficiência, baixo consumo de memória e rápido treinamento, que o tornam mais preciso quando bem ajustado (Madhavi *et al.*, 2022; Bogdanovski *et al.*, 2023).

As métricas R^2 , RMSE, MAE e o índice de concordância de Willmott (d-Willmott) foram fundamentais para avaliar a qualidade e robustez dos modelos. O R^2 indica o quanto as previsões se aproximam dos valores reais, refletindo a capacidade explicativa do modelo. O RMSE, sensível a grandes desvios, mede a magnitude dos erros e identifica modelos mais vulneráveis a *outliers*. O MAE expressa, de forma intuitiva, a média dos desvios absolutos,

revelando a consistência das estimativas. E o d-Willmott, mede o grau de concordância entre valores preditos e observados, indicando a precisão do modelo. A análise conjunta dessas métricas proporciona uma avaliação abrangente do desempenho, auxiliando na escolha do modelo mais adequado (Madhavi *et al.*, 2022; Amaro *et al.*, 2025).

A utilização da validação cruzada do tipo K-fold ($k = 5$), na análise dos modelos, contribuiu para reduzir o risco de sobreajuste, promovendo uma aprendizagem mais abrangente e adaptável, proporcionando uma estimativa mais robusta e confiável do desempenho dos modelos em dados não vistos, sugerindo que o desempenho possa ser mantido ou aprimorado em bases de dados maiores (Sudipa *et al.*, 2024).

A análise de box plot permite avaliar a fidelidade da distribuição dos valores preditos em relação aos valores observados, ou seja, se os modelos conseguem replicar a variabilidade dos dados. Já as curvas de densidade permitem avaliar a similaridade entre as distribuições dos valores preditos e observados, com foco na forma da curva, e não apenas nos valores centrais ou dispersão. Assim, os modelos SVM, XGB, KNN e LGBM avaliados através do desempenho preditivo, distribuição e densidade dos valores observados e preditos, apresentaram maior precisão, ajuste e fidelidade na reprodução da dispersão e distribuição dos dados, destacando-se como ferramentas práticas para prever com precisão as massas dos frutos de morango.

As métricas de classificação em conjunto das matrizes de confusão, são essenciais para selecionar o modelo mais adequado, pois permite avaliar o impacto dos diferentes tipos de erro (Choi *et al.*, 2021). Espera-se que os modelos consigam reconhecer o maior número possível de verdadeiros positivos e verdadeiros negativos, e que o número de falsos positivos e negativos possa ser evitado ao máximo (Wenchao & Zhi, 2022). Quanto maior as métricas melhor o desempenho de classificação do modelo (Yuan *et al.*, 2023). Quando a prioridade é minimizar falsos negativos e garantir a detecção da maioria dos frutos comerciais, modelos com alto recall são mais adequados. Por outro lado, em situações onde é fundamental evitar falsos positivos (como o risco de classificar frutos não comerciais como comerciais), modelos com elevada precisão tornam-se mais indicados. Seguindo esse raciocínio os modelos SVM, XGB e RF se destacaram pela maior eficiência e também por oferecer um equilíbrio consistente entre precisão e recall, mantendo um bom controle tanto sobre falsos positivos quanto sobre falsos negativos.

Esses resultados corroboram com os achados de estudos anteriores em modelagem agrícola. Basak *et al.*, (2022) e Oliveira *et al.*, (2023), demonstraram que o modelo SVM, também foi eficaz e se ajustou bem aos dados na predição da massa do fruto de morango, com métricas (R^2 acima de 0,85 e baixos RMSE) próximas ao observado neste estudo. Os modelos

XGB e LGBM, foram eficientes para a predição da produtividade de mirtilo (Niedbala *et al.*, 2022) e o modelo KNN obteve eficácia na discriminação do estágio de amadurecimento de frutos de morango (Ibba *et al.*, 2021). Para a classificação de frutos de morango, esses modelos também mostraram bom ajuste para detecção de frutos podres (Şengöz *et al.*, 2025), identificação de estágios de maturação e variedades (Huang *et al.*, 2025) e identificação de doenças (Pertiwi *et al.*, 2023).

A classificação dos morangos é feita principalmente de forma manual, que é subjetiva e propensa a erros, ao mesmo tempo que é demorado e pode causar danos aos frutos (Yuan *et al.*, 2023). Portanto, a utilização de imagens representa uma alternativa eficiente aos métodos tradicionais, como inspeções visuais e análises laboratoriais que demandam alta mão de obra. Além de reduzir custos e tempo essa abordagem viabilizar a coleta de dados em larga escala, sendo muito relevante no contexto de automação agrícola (Barbosa Júnior *et al.*, 2025). Predizer a massa sem causar danos aos morangos pode minimizar o processo de reclassificação e reduzir as etapas de colheita e embalagem (Jeong *et al.*, 2024). A classificação da qualidade dos frutos representa um elo importante na cadeia de colheita e comercialização, sendo que a comercialização segmentada por categorias de qualidade pode potencializar os lucros (Yuan *et al.*, 2023).

Essas tecnologias estão cada vez mais eficientes para avaliar a qualidade da fruta e prever a aceitação do consumidor devido à capacidade de extrair informações extras e correlacionar e avaliar multivariáveis em conjunto. Estudos futuros devem explorar uma gama mais ampla de variedades e diferentes estações do ano, juntamente com o desenvolvimento de modelos mais robustos (Castillo-Girones *et al.*, 2025).

No presente estudo, independentemente de as sépalas dos frutos de morango estarem abertas ou fechadas, os frutos foram recortados com precisão e a medição de suas características foi conduzida com sucesso. Isso demonstra a robustez do método proposto na extração precisa das características das imagens dos frutos, independentemente da apresentação das sépalas. A presença e frescor das sépalas é importante pois além de ser um critério de avaliação, os frutos são colhidos, pesados, armazenados (com as sépalas tendem a ter maior tempo de prateleira) e comercializados com as sépalas (Wang *et al.*, 2024; Ye *et al.*, 2024). Assim, o método proposto oferece desempenho consistente, independentemente de serem frutos de materiais genéticos diferentes ou da condição dos morangos, fornecendo medições confiáveis e estáveis para a predição da massa dos morangos (Jeong *et al.*, 2024).

Apesar dos resultados promissores obtidos neste estudo, algumas limitações devem ser consideradas. O tamanho relativamente pequeno da base de dados pode ter comprometido o

desempenho de modelos mais sensíveis à quantidade de amostras, como o MLP. Além disso, a qualidade das imagens pode ter sido influenciada por fatores como iluminação e ângulo de captura, que pode gerar ruídos nas variáveis extraídas. A utilização de base de dados de um número maior de áreas cultivadas sob diferentes condições bióticas e abióticas deve ser utilizado, possibilitando a criação de modelos ainda mais avançados e eficientes (Oliveira *et al.*, 2023; Huang *et al.*, 2025).

Ainda assim, os resultados demonstram o grande potencial dos modelos preditivos baseados em aprendizado de máquina, aliados a medições não destrutivas, como ferramentas eficazes e econômicas para prever a massa de frutos de morango. Dessa forma, esta pesquisa contribui para o avanço de soluções tecnológicas aplicadas à agricultura digital, com impactos diretos na eficiência e sustentabilidade das práticas agrícolas. O método proposto demonstra capacidade de fornecer informações que não são acessíveis por técnicas clássicas de controle de produtividade e qualidade, podendo ser expandido para a análise não destrutiva de outros parâmetros qualitativos, como maturação, firmeza, teor de sólidos solúveis e fibras em morangos, bem como aplicado à avaliação de qualidade em outras culturas agrícolas (Weng *et al.*, 2020; Oliveira *et al.*, 2023).

6 CONCLUSÃO

A predição da massa e a classificação dos frutos de morango por meio de modelos de aprendizado de máquina apresentaram resultados satisfatórios com ambos os conjuntos de dados avaliados. Os modelos SVM, XGB, KNN e LGBM obtiveram melhor ajuste na predição da massa, enquanto SVM, XGB e RF se destacaram na tarefa de classificação. Embora as variáveis agronômicas manuais tenham proporcionado maior estabilidade e precisão, os dados extraídos de imagens demonstraram-se promissores, evidenciando seu potencial como uma abordagem não destrutiva, eficiente e escalável. A utilização de imagens permite o monitoramento automatizado da massa e a classificação dos frutos, contribuindo para a otimização da cadeia produtiva, redução de custos operacionais e maior agilidade na tomada de decisões. Esses modelos podem ser integrados a sistemas inteligentes de colheita e classificação, como esteiras automatizadas, colhedoras mecanizadas ou drones, viabilizando aplicações em tempo real com baixo custo e alta eficiência. Ademais, os resultados obtidos estão em consonância com os Objetivos de Desenvolvimento Sustentável, ao promoverem práticas agrícolas mais sustentáveis (ODS 2 e 12), incentivarem a inovação no setor (ODS 9) e contribuírem, ainda que indiretamente, para a mitigação de impactos ambientais (ODS 13).

REFERÊNCIAS

- ABBASI, Rabiya; MARTINEZ, Pablo; AHMAD, Rafiq. The digitization of agricultural industry—a systematic literature review on agriculture 4.0. **Smart Agricultural Technology**, v. 2, p. 100042, 2022. <https://doi.org/10.1016/j.atech.2022.100042>
- ABDO-PERALTA, Paula, *et al.* Implementation of an Enhanced Edge Computing System for the Optimization of Strawberry Crop in Greenhouses: a smart agriculture approach. **Agronomy**, [S.L.], v. 14, n. 12, p. 3030, 19 dez. 2024. MDPI AG. <http://dx.doi.org/10.3390/agronomy14123030>.
- AHMED, Shakeel. A software framework for predicting the maize yield using modified multi-layer perceptron. **Sustainability**, v. 15, n. 4, p. 3017, 2023. <https://doi.org/10.3390/su15043017>
- AKHTER, Ravesa; SOFI, Shabir Ahmad. Precision agriculture using IoT data analytics and machine learning. **Journal of King Saud University-Computer and Information Sciences**, v. 34, n. 8, p. 5602-5618, 2022. <https://doi.org/10.1016/j.jksuci.2021.05.013>
- ALDOSSARY, Abdulrahman *et al.* In silico chemical experiments in the age of AI: From quantum chemistry to machine learning and back. **Advanced Materials**, v. 36, n. 30, p. 2402369, 2024. <https://doi.org/10.1002/adma.202402369>
- ALVARES, Clayton Alcarde *et al.* Köppen's climate classification map for Brazil. **Meteorologische zeitschrift**, v. 22, n. 6, p. 711-728, 2013. <https://doi.org/10.1127/0941-2948/2013/0507>
- AMARO, Rafaella Pironato *et al.* Performance evaluation of Sentinel-2 imagery, agronomic and climatic data for sugarcane yield estimation. **Computers and Electronics in Agriculture**, v. 237, p. 110522, 2025. <https://doi.org/10.1016/j.compag.2025.110522>
- AMKOR, Ali; EL BARBRI, Nouredine. Artificial intelligence methods for classification and prediction of potatoes harvested from fertilized soil based on a sensor array response. **Sensors and Actuators A: Physical**, v. 349, p. 114106, 2023. <https://doi.org/10.1016/j.sna.2022.114106>
- ANTUNES, L. E. C.; PERES, N. A. Strawberry production in Brazil and South America. **International Journal of Fruit Science**, v. 13, n. 1-2, p. 156-161, 2013. <https://doi.org/10.1080/15538362.2012.698147>
- ANTUNES, L. E. C.; REISSER JÚNIOR, C.; BONOW, S. **Morango: produção aumenta ano a ano**. 2021.
- ARAÚJO, Sara Oleiro *et al.* Machine learning applications in agriculture: current trends, challenges, and future perspectives. **Agronomy**, v. 13, n. 12, p. 2976, 2023. <https://doi.org/10.3390/agronomy13122976>
- ATLAS BIG. **World strawberry production by country**. 2020. Disponível em: <https://www.atlasbig.com/en-us/countries-strawberry-production>. Acesso em: 25 mar. 2025.

ATTRI, Ishana *et al.* Machine learning in agriculture: a review of crop management applications. **Multimedia Tools and Applications**, v. 83, n. 5, p. 12875-12915, 2024. <https://doi.org/10.1007/s11042-023-16105-2>

AZIZ, Rabia Musheer *et al.* LGBM: a machine learning approach for Ethereum fraud detection. **International Journal of Information Technology**, v. 14, n. 7, p. 3321-3331, 2022. <https://doi.org/10.1007/s41870-022-00864-6>

BARBOSA JÚNIOR, M. R. *et al.* Image-based and ML-driven analysis for assessing blueberry fruit quality. **Heliyon**, [S.l.], v. 11, e42288, 2025. <https://doi.org/10.1016/j.heliyon.2025.e42288>.

BARBOZA, Thiago Orlando Costa *et al.* Application of artificial intelligence for identification of peanut maturity using climatic variables and vegetation indices. **Precision Agriculture**, v. 26, n. 3, p. 1-30, 2025. <https://doi.org/10.1007/s11119-025-10237-1>

BARTH, E. *et al.* Métodos de análise multivariada melhoram a seleção de genótipos de morango com baixa exigência de frio. **Scientific Reports**, v. 12, p. 11458, 2022. <https://doi.org/10.1038/s41598-022-15688-4>.

BASAK, Jayanta Kumar *et al.* Non-destructive estimation of strawberry fruit weight using machine learning models. **Agronomy**, v. 12, n. 10, p. 2487, 2022. <https://doi.org/10.3390/agronomy12102487>

BENOS, Lefteris *et al.* Machine learning in agriculture: A comprehensive updated review. **Sensors**, v. 21, n. 11, p. 3758, 2021. <https://doi.org/10.3390/s21113758>

BIRNER, Regina; DAUM, Thomas; PRAY, Carl. Who drives the digital revolution in agriculture? A review of supply-side trends, players and challenges. **Applied economic perspectives and policy**, v. 43, n. 4, p. 1260-1285, 2021. <https://doi.org/10.1002/aepp.13145>

BOGDANOVSKI, Oliver Persson *et al.* Yield prediction for winter wheat with machine learning models using Sentinel-1, topography, and weather data. **Agriculture**, v. 13, n. 4, p. 813, 2023. <https://doi.org/10.3390/agriculture13040813>

BRANDT, Guilherme Q. *et al.* Productivity and analysis of morphological characters of experimental strawberry genotypes. **Horticultura Brasileira**, v. 40, n. 4, p. 426-431, 2022. <https://doi.org/10.1590/s0102-0536-20220411>

BUSHAKRA, Jill M. *et al.* A strawberry (*Fragaria L.*) crop ontology to enable standardized phenotyping for strawberry breeding and research. **HortScience**, v. 60, n. 5, p. 784-793, 2025. <https://doi.org/10.21273/HORTSCI18420-24>

CAMARGO, Leticia Kurchaidt Pinheiro *et al.* Performance and average mass of strawberry fruit obtained from different cropping systems. **Ambiência**, v. 6, n. 2, p. 281-288, 2010.

CASTILLO-GIRONES, Salvador *et al.* Advanced evaluation of strawberry quality, consumer preference, and cultivar discrimination through spectral imaging and neural networks. **Food Control**, v. 175, p. 111339, 2025. <https://doi.org/10.1016/j.foodcont.2025.111339>

CATALANO, Christian *et al.* Anomaly detection in smart agriculture systems. **Computers in Industry**, v. 143, p. 103750, 2022. <https://doi.org/10.1016/j.compind.2022.103750>

CHEN, Xi *et al.* Genetic Variation, Polyploidy, Hybridization Influencing the Aroma Profiles of Rosaceae Family. **Genes**, v. 15, n. 10, p. 1339, 2024. <https://doi.org/10.3390/genes15101339>

CHEN, Yiming *et al.* Estimating strawberry weight for grading by picking robot with point cloud completion and multimodal fusion network. **Scientific Reports**, v. 15, n. 1, p. 11227, 2025. <https://doi.org/10.1038/s41598-025-92641-1>

CHICCO, Davide *et al.* The coefficient of determination R-squared is more informative than SMAPE, MAE, MAPE, MSE and RMSE in regression analysis evaluation. **Peerj computer science**, v. 7, p. e623, 2021. <https://doi.org/10.7717/peerj-cs.623>

CHIOMENRO, J. L. T. *et al.* Horticultural potential of nine strawberry cultivars by greenhouse production in Brazil: a view through multivariate analysis. **Scientia Horticulturae**, v. 279, 2021. <https://doi.org/10.1016/j.scienta.2020.109738>.

CHOI, Jung-Gu *et al.* Machine learning framework for multi-level classification of company revenue. **IEEE Access**, v. 9, p. 96739-96750, 2021. <https://doi.org/10.1109/ACCESS.2021.3088874>

CONDRAN, Sarah *et al.* Machine learning in precision agriculture: a survey on trends, applications and evaluations over two decades. **IEEE Access**, v. 10, p. 73786-73803, 2022. <https://doi.org/10.1109/ACCESS.2022.3188649>

CORRÊA, Jessica Vanessa Wosniak *et al.* ISSR analysis reveals high genetic variation in strawberry three-way hybrids developed for tropical regions. **Plant Molecular Biology Reporter**, p. 1-11, 2021. <https://doi.org/10.1007/s11105-020-01270-7>

DA SILVA, *et al.* Estudo de mercado sobre morango *in natura* e polpa. **Revista do Encontro de Gestão e Tecnologia**, v. 1, n. 07, p. e289-e289, 2024. <https://doi.org/10.5281/zenodo.13323794>

DA SILVA, Lidiane Rodrigues *et al.* Strawberries in a warming world: examining the ecological niche of *Fragaria* × *ananassa* Duch–Across different climate scenarios. **Journal of Berry Research**, v. 14, n. 3, p. 193-208, 2024. <https://doi.org/10.3233/JBR-240012>

DALID, Cheryl *et al.* Genetic analysis of predicted vegetative biomass and biomass-related traits from digital phenotyping of strawberry. **The Plant Genome**, v. 18, n. 2, p. e70018, 2025. <https://doi.org/10.1002/tpg2.70018>

DARA, Rozita; HAZRATI FARD, Seyed Mehdi; KAUR, Jasmin. Recommendations for ethical and responsible use of artificial intelligence in digital agriculture. **Frontiers in Artificial Intelligence**, v. 5, p. 884192, 2022. <https://doi.org/10.3389/frai.2022.884192>

DARROW, G. M. **The Strawberry: History, Breeding and Physiology**. New York: Holt, Rinehart and Winston. v. 37, 1966.

DE SOUZA, D. C. *et al.* Variabilidade genética entre cultivares comerciais e híbridos experimentais de morangueiro com ênfase em análise de múltiplos fatores. **Magistra**, v. 30, p. 48-59, 2019.

DELAZERI, Elói Evandro *et al.* Physical and quality fruit parameters of new strawberry genotypes. **Pesquisa Agropecuária Brasileira**, v. 59, p. e03462, 2024. <https://doi.org/10.1590/S1678-3921.pab2024.v59.03462>

DHAL, Sambandh *et al.* Internet of Things (IoT) in digital agriculture: An overview. **Agronomy Journal**, v. 116, n. 3, p. 1144-1163, 2024.

DOS SANTOS, Erli Pinto *et al.* A Machine Learning approach to reconstruct cloudy affected vegetation indices imagery via data fusion from Sentinel-1 and Landsat 8. **Computers and Electronics in Agriculture**, v. 194, p. 106753, 2022. <https://doi.org/10.1016/j.compag.2022.106753>

DOS SANTOS, Marllon Fernando Soares *et al.* Agronomic performance of new strawberry cultivars in southern Brazil. **Revista de Ciências Agroveterinárias**, v. 20, n. 2, p. 149-158, 2021. <https://doi.org/10.5965/223811712022021149>

ELSAYED, Salah *et al.* Hyperspectral technology and machine learning models to estimate the fruit quality parameters of mango and strawberry crops. **PloS one**, v. 20, n. 2, p. e0313397, 2025. <https://doi.org/10.1371/journal.pone.0313397>

EZZIYYANI, Mohammed *et al.* Contribution to the development of a technological platform for analysis in precision agriculture for the biovigilance of cryptogamic diseases in strawberry (*Fragaria × ananassa*). **In: International Conference on Advanced Intelligent Systems for Sustainable Development, Cham: Springer Nature Switzerland**, 2022. p. 637-644.

FAEDO, Leonardo *et al.* The use of mineral dynamised high dilutions for natural plant biostimulation; effects on plant growth, crop production, fruit quality, pest and disease incidence in agroecological strawberry cultivation. **Biological Agriculture & Horticulture**, v. 40, n. 4, p. 267-287, 2024. <https://doi.org/10.1080/01448765.2024.2396894>

FAGHERAZZI, A. F.; *et al.* Strawberry production progress in Brazil. **Acta Horticulturae**, n. 1156, p. 937-940, 2017. <https://doi.org/10.17660/actahortic.2017.1156.138>.

FAN, Kexin *et al.* Engineering Strategies for Advancing Optical Signal Outputs in Smartphone-Enabled Point-of-Care Diagnostics. **Advanced Intelligent Systems**, v. 5, n. 6, p. 2200285, 2023. <https://doi.org/10.1002/aisy.202200285>

FAO. **Crops and livestock products: production quantities of strawberries in world**. 2023. Disponível em: <https://www.fao.org/faostat/en/#data/QCL/visualize>. Acesso em: 25 mar. 2025.

FELDMANN, Mitchell J. *et al.* Genetic gains underpinning a little-known strawberry green revolution. **Nature Communications**, v. 15, n. 1, p. 2468, 2024. <https://doi.org/10.1038/s41467-024-46421-6>

FELDMANN, Mitchell J. *et al.* Multi-dimensional machine learning approaches for fruit shape phenotyping in strawberry. **GigaScience**, v. 9, n. 5, p. g1aa030, 2020. <https://doi.org/10.1093/gigascience/g1aa030>

GASTON, Amelia *et al.* Applying the Solanaceae strategies to strawberry crop improvement. **Trends in plant science**, v. 25, n. 2, p. 130-140, 2020. <https://doi.org/10.1016/j.tplants.2019.10.003>

GIL *et al.* Evolution of the fruit ripening and development of the strawberry 'Aroma'; through transcriptional, physiological and chemicals analysis **Food Measure**, v. 18, p. 3160–3175, 2024. <https://doi.org/10.1007/s11694-024-02393-1>

GUEVARA-TERÁN, Mabel *et al.* Influence of altitude on the physicochemical composition and antioxidant capacity of strawberry: A preliminary systematic review and meta-analysis. **Phytochemistry Reviews**, v. 22, n. 6, p. 1567-1584, 2023. <https://doi.org/10.1007/s11101-022-09834-z>

HACKFORT S. Padrões de desigualdades na agricultura digital: uma revisão sistemática da literatura. **Sustentabilidade**. 2021; 13(22):12345. <https://doi.org/10.3390/su132212345>

HUANG, Jing *et al.* Classification of Strawberry Maturity Stages and Varieties Using Neural Networks Based on Volatile Organic Compounds. **Foods**, v. 14, n. 2, p. 169, 2025. <https://doi.org/10.3390/foods14020169>

HUANG, Linsheng *et al.* Combining random forest and XGBoost methods in detecting early and mid-term winter wheat stripe rust using canopy level hyperspectral measurements. **Agriculture**, v. 12, n. 1, p. 74, 2022. <https://doi.org/10.3390/agriculture12010074>

HUMMER, Kim E. *et al.* Phenotypic characterization of a strawberry (*Fragaria* × *ananassa* Duchesne ex Rosier) diversity collection. **Plants, People, Planet**, v. 5, n. 2, p. 209-224, 2023. <https://doi.org/10.1002/ppp3.10316>

IBBA, Pietro *et al.* Supervised binary classification methods for strawberry ripeness discrimination from bioimpedance data. **Scientific reports**, v. 11, n. 1, p. 11202, 2021. <https://doi.org/10.1038/s41598-021-90471-5>

INGRAM, Julie *et al.* Quais são as questões prioritárias de pesquisa para a agricultura digital?. **Land Use Policy**, v. 114, p. 105962, 2022. <https://doi.org/10.1016/j.landusepol.2021.105962>

ISLAM, Nahina *et al.* Early weed detection using image processing and machine learning techniques in an Australian chilli farm. **Agriculture**, v. 11, n. 5, p. 387, 2021. <https://doi.org/10.3390/agriculture11050387>

JAGTAP, Santosh T. *et al.* Towards application of various machine learning techniques in agriculture. **Materials Today: Proceedings**, v. 51, p. 793-797, 2022. <https://doi.org/10.1016/j.matpr.2021.06.236>

JAMES, Katherine Margaret Frances *et al.* High-throughput phenotyping for breeding targets—Current status and future directions of strawberry trait automation. **Plants, People, Planet**, v. 4, n. 5, p. 432-443, 2022. <https://doi.org/10.1002/ppp3.10275>

JEONG, Haejun *et al.* Automated Technology for Strawberry Size Measurement and Weight Prediction Using AI. **IEEE Access**, v. 12, p. 14157-14167, 2024. <https://doi.org/10.1109/ACCESS.2024.3356118>

KAPARI, Mpho *et al.* Comparing machine learning algorithms for estimating the maize crop water stress index (CWSI) using UAV-acquired remotely sensed data in smallholder croplands. **Drones**, v. 8, n. 2, p. 61, 2024. <https://doi.org/10.3390/drones8020061>

KHAN, Shahbaz *et al.* Deep learning-based identification system of weeds and crops in strawberry and pea fields for a precision agriculture sprayer. **Precision Agriculture**, v. 22, n. 6, p. 1711-1727, 2021. <https://doi.org/10.1007/s11119-021-09808-9>

KILIC, Neslihan *et al.* Efeito de fertilizantes orgânicos, inorgânicos e suas combinações nos parâmetros de qualidade dos frutos de morango. **Horticulturae**, v. 7, n. 10, p. 354, 2021. <https://doi.org/10.3390/horticulturae7100354>

KIRK, Daniel *et al.* Precision nutrition: A systematic literature review. **Computers in Biology and Medicine**, v. 133, p. 104365, 2021. <https://doi.org/10.1016/j.compbiomed.2021.104365>

KOK, Zhi Hong *et al.* Support vector machine in precision agriculture: A review. **Computers and Electronics in Agriculture**, v. 191, p. 106546, 2021. <https://doi.org/10.1016/j.compag.2021.106546>

KOLLING, Caio Ericles; RAMPIM, Leandro. Agricultura de precisão e digital: Perspectivas e desafios dos produtores rurais do estado do paran . **Uning  Review**, v. 36, p. eURJ3981-eURJ3981, 2021. <https://doi.org/10.46311/2178-2571.36.eURJ3981>

KUSAL, Sheetal *et al.* Agentes conversacionais baseados em IA: uma revis o de escopo, das tecnologias  s dire oes futuras. **IEEE Access**, v. 10, p. 92337-92356, 2022. <https://doi.org/10.1109/ACCESS.2022.3201144>

KWAGHTYO, Dekera Kenneth; EKE, Christopher Ifeanyi. Smart farming prediction models for precision agriculture: a comprehensive survey. **Artificial Intelligence Review**, v. 56, n. 6, p. 5729-5772, 2023. <https://doi.org/10.1007/s10462-022-10266-6>

LI, Cheng *et al.* Measuring plant growth characteristics using smartphone-based image analysis technique in controlled environment agriculture. **Computers and Electronics in Agriculture**, v. 168, p. 105123, 2020. <https://doi.org/10.1016/j.compag.2019.105123>

LI, Xiang *et al.* StrawberryNet: Fast and Precise Recognition of Strawberry Disease Based on Channel and Spatial Information Reconstruction. **Agriculture**, v. 15, n. 7, p. 779, 2025. <https://doi.org/10.3390/agriculture15070779>

LIAKOS, Konstantinos G. *et al.* Machine learning in agriculture: A review. **Sensors**, v. 18, n. 8, p. 2674, 2018. <https://doi.org/10.3390/s18082674>

LIEW, Yi Wei *et al.* Potential of biochar as soil amendment: prediction of elemental ratios from pyrolysis of agriculture biomass using artificial neural network. **Water, Air, & Soil Pollution**, v. 233, n. 2, p. 54, 2022. <https://doi.org/10.1007/s11270-022-05510-2>

LIU, Changhong *et al.* Application of multispectral imaging to determine quality attributes and ripeness stage in strawberry fruit. **PloS one**, v. 9, n. 2, p. e87818, 2014. <https://doi.org/10.1371/journal.pone.0087818>

MADHAVI, Bolappa Gamage Kaushalya *et al.* Prediction of strawberry leaf color using RGB mean values based on soil physicochemical parameters using machine learning models. **Agronomy**, v. 12, n. 5, p. 981, 2022. <https://doi.org/10.3390/agronomy12050981>

MANSURI, Shekh Mukhtar *et al.* Computer vision model for estimating the mass and volume of freshly harvested Thai apple ber (*Ziziphus mauritiana* L.) and its variation with storage days. **Scientia Horticulturae**, v. 305, p. 111436, 2022. <https://doi.org/10.1016/j.scienta.2022.111436>

MANVITHA, M. Sai *et al.* A predictive analysis on CO₂ emissions in automobiles using machine learning techniques. **In: 2023 International Conference on Intelligent Data Communication Technologies and Internet of Things (IDCIoT)**. IEEE, 2023. p. 394-401. <https://doi.org/10.1109/IDCIoT56793.2023.10053539>.

MARTINS DE LIMA, Juliana *et al.* Planting density interferes with strawberry production efficiency in southern Brazil. **Agronomy**, v. 11, n. 3, p. 408, 2021. <https://doi.org/10.3390/agronomy11030408>

MENDES, Jorge *et al.* Smartphone applications targeting precision agriculture practices—A systematic review. **Agronomy**, v. 10, n. 6, p. 855, 2020. <https://doi.org/10.3390/agronomy10060855>

MENZEL, Christopher Michael. Temperatures above 30°C decrease leaf growth in strawberry under global warming. **The Journal of Horticultural Science and Biotechnology**, p. 1-24, 2024. <https://doi.org/10.1080/14620316.2024.2360452>

MISHRA, Pranav Kumar *et al.* Genetic variability, heritability, and genetic advance in strawberry (*Fragaria* × *ananassa* Duch.). **Turkish Journal of Agriculture and Forestry**, v. 39, n. 3, p. 451-458, 2015. <https://doi.org/10.3906/tar-1408-99>

MOHAMED, Elsayed Said *et al.* Smart farming for improving agricultural management. **The Egyptian Journal of Remote Sensing and Space Science**, v. 24, n. 3, p. 971-981, 2021. <https://doi.org/10.1016/j.ejrs.2021.08.007>

NAGAKI, Yukihiisa *et al.* Ranking Prediction of Fresh Produce Quality: A Case Study of Strawberry Ripeness. **Food and Bioprocess Technology**, p. 1-10, 2025. <https://doi.org/10.1007/s11947-025-03847-0>

NDIKUMANA, Jean Nepo *et al.* Development of a deep-learning phenotyping tool for analyzing image-based strawberry phenotypes. **Frontiers in Plant Science**, v. 15, p. 1418383, 2024. <https://doi.org/10.3389/fpls.2024.1418383>

- NI, Jianguo *et al.* E-AlexNet: avaliação da qualidade de morango com base em aprendizado de máquina. **Journal of Food Measurement and Characterization**, v. 15, n. 5, p. 4530-4541, 2021. <https://doi.org/10.1007/s11694-021-01010-9>
- NIEDBAŁA, Gniewko *et al.* Prediction of blueberry (*Vaccinium Corymbosum* L.) yield based on artificial intelligence methods. **Agriculture**, v. 12, n. 12, p. 2089, 2022. <https://doi.org/10.3390/agriculture12122089>
- NITHYA, V.; JOSEPHINE, M. S.; JEYABALARAJA, V. Comparative approach on crop detection using machine learning and deep learning techniques. **International Journal of System Assurance Engineering and Management**, v. 15, n. 9, p. 4636-4648, 2024. <https://doi.org/10.1007/s13198-024-02483-9>
- NUNES, Néria S. *et al.* Post-harvest of strawberry accessions in the South Minas Gerais. **Horticultura Brasileira**, v. 40, n. 2, p. 221-225, 2022. <https://doi.org/10.1590/s0102-0536-20220212>
- OLIVEIRA, Larissa Silva de *et al.* Estimation of strawberry crop productivity by machine learning algorithms using data from multispectral images. **Agronomy**, v. 13, n. 5, p. 1229, 2023. <https://doi.org/10.3390/agronomy13051229>
- PALUMBO, Michela *et al.* Rapid and non-destructive techniques for the discrimination of ripening stages in Candonga Strawberries. **Foods**, v. 11, n. 11, p. 1534, 2022. <https://doi.org/10.3390/foods11111534>
- PEPPES, Nikolaos *et al.* Performance of machine learning-based multi-model voting ensemble methods for network threat detection in agriculture 4.0. **Sensors**, v. 21, n. 22, p. 7475, 2021. <https://doi.org/10.3390/s21227475>
- PEREIRA, M. A. *et al.* Productivity of experimental strawberry clones for Alfenas-MG. **Colloquium Agrariae**, 2022. p. 46-52. <https://doi.org/10.5747/ca.2022.v18.n1.a478>
- PERTIWI, Setyo *et al.* Deep Learning Model for Identification of Diseases on Strawberry (*Fragaria* sp.) Plants. **International Journal on Advanced Science, Engineering & Information Technology**, v. 13, n. 4, 2023. <https://doi.org/10.18517/ijaseit.13.4.19018>
- PORTER, Mark *et al.* Strawberry breeding for improved flavor. **Crop Science**, v. 63, n. 4, p. 1949-1963, 2023. <https://doi.org/10.1002/csc2.21012>
- RAHMANI, Amir Masoud *et al.* Machine learning (ML) in medicine: review, applications, and challenges. **Mathematics**, v. 9, n. 22, p. 2970, 2021. <https://doi.org/10.3390/math9222970>
- RAJ, Rahul *et al.* Strawberry water content estimation and ripeness classification using hyperspectral sensing. **Agronomy**, v. 12, n. 2, p. 425, 2022. <https://doi.org/10.3390/agronomy12020425>

- REN, Yangguang *et al.* Slurry-ability mathematical modeling of microwave-modified lignite: A comparative analysis of multivariate non-linear regression model and XGBoost algorithm model. **Energy**, v. 281, p. 128143, 2023. <https://doi.org/10.1016/j.energy.2023.128143>
- SAGLAM, Cevdet; CETIN, Necati. Prediction of pistachio (*Pistacia vera* L.) mass based on shape and size attributes by using machine learning algorithms. **Food Analytical Methods**, p. 1-12, 2022. <https://doi.org/10.1007/s12161-021-02154-6>
- SANTOS, M. F. S. dos *et al.* Agronomic performance of new strawberry cultivars in southern Brazil. **Revista de Ciências Agroveterinárias**, Lages, v. 20, n. 2, p. 149–158, 2021. <https://doi.org/10.5965/223811712022021149>.
- SARKER, Iqbal H. Machine learning: Algorithms, real-world applications and research directions. **SN computer science**, v. 2, n. 3, p. 160, 2021. <https://doi.org/10.1007/s42979-021-00592-x>
- ŞENGÖZ, Nilgün *et al.* XGBoost Tabanlı Derin Öğrenme Algoritması ile Açıklanabilir Yapay Zeka Modellerinin Kullanımı: Çürük Meyvelerin Tespiti. **Journal of Natural & Applied Sciences**, v. 29, n. 1, 2025. <https://doi.org/10.19113/sdufenbed.1575098>
- SUDIPA, I. Gede Iwan *et al.* Leveraging k-nearest neighbors for enhanced fruit classification and quality assessment. **Indonesian Journal of Data and Science**, v. 5, n. 1, p. 30-36, 2024. <https://doi.org/10.56705/ijodas.v5i1.125>
- SUNIL, G. C. *et al.*, Evaluation of multispectral imaging for freeze damage assessment in strawberries using AI-based computer vision technology. **Smart Agricultural Technology**, v. 10, 2025. <https://doi.org/10.1016/j.atech.2025.100851>.
- TUFAIL, Shahid *et al.* Advancements and challenges in machine learning: A comprehensive review of models, libraries, applications, and algorithms. **Electronics**, v. 12, n. 8, p. 1789, 2023. <https://doi.org/10.3390/electronics12081789>
- UPOV – INTERNATIONAL UNION FOR THE PROTECTION OF NEW VARIETIES OF PLANTS. **Strawberry: guidelines for the conduct of tests for distinctness, uniformity and stability**. Geneva, 2012. 36 p.
- WALEED, Muhammad *et al.* Classification of agriculture farm machinery using machine learning and internet of things. **Symmetry**, v. 13, n. 3, p. 403, 2021. <https://doi.org/10.3390/sym13030403>
- WANG, Lei *et al.* β -aminobutyric acid effectively postpones senescence of strawberry fruit by regulating metabolism of NO, H₂S, ascorbic acid, and ABA. **Horticulturae**, v. 10, n. 3, p. 218, 2024. <https://doi.org/10.3390/horticulturae10030218>
- WANG, Yuanhua *et al.* FaWRKY11 transcription factor positively regulates resistance to *Botrytis cinerea* in strawberry fruit. **Scientia Horticulturae**, v. 279, p. 109893, 2021. <https://doi.org/10.1016/j.scienta.2020.109893>

- WENCHAO, Xu; ZHI, Yan. Research on strawberry disease diagnosis based on improved residual network recognition model. **Mathematical Problems in Engineering**, v. 2022, n. 1, p. 6431942, 2022. <https://doi.org/10.1155/2022/6431942>
- WENG, Shizhuang *et al.* Non-destructive detection of strawberry quality using multi-features of hyperspectral imaging and multivariate methods. **Sensors**, v. 20, n. 11, p. 3074, 2020. <https://doi.org/10.3390/s20113074>
- WILMOT, Hélène *et al.* The use of a genomic relationship matrix for breed assignment of cattle breeds: comparison and combination with a machine learning method. **Journal of animal science**, v. 101, p. skad172, 2023. <https://doi.org/10.1093/jas/skad172>
- WU, Linxia *et al.* Geographical Origin Identification of Sweet Cherry Based on Quality Traits Combined with Dd-Simca and Xgboost. **Available at SSRN**. n. 5260062. 2025. <http://dx.doi.org/10.2139/ssrn.5260062>
- XIE, Dong *et al.* Estimating the contribution of environmental variables to water quality in the postrestoration littoral zones of Taihu Lake using the APCS-MLR model. **Science of the Total Environment**, v. 857, p. 159678, 2023. <https://doi.org/10.1016/j.scitotenv.2022.159678>
- YANG, Ningyuan *et al.* Detection of color phenotype in strawberry germplasm resources based on field robot and semantic segmentation. **Computers and Electronics in Agriculture**, v. 226, p. 109464, 2024. <https://doi.org/10.1016/j.compag.2024.109464>
- YE, Rong *et al.* CR-YOLOv9: improved YOLOv9 multi-stage strawberry fruit maturity detection application integrated with CRNET. **Foods**, v. 13, n. 16, p. 2571, 2024. <https://doi.org/10.3390/foods13162571>
- YOON, Byung C. *et al.* Incorporating algorithmic uncertainty into a clinical machine deep learning algorithm for urgent head CTs. **PLOS one**, v. 18, n. 3, p. e0281900, 2023. <https://doi.org/10.1371/journal.pone.0281900>
- YU, Xiaokang *et al.* Edge detection of agricultural products based on morphologically improved canny algorithm. **Mathematical Problems in Engineering**, v. 2021, n. 1, p. 6664970, 2021. <https://doi.org/10.1155/2021/6664970>
- YUAN, Shi-Qi *et al.*, Research on Strawberry Quality Grading Based on Object Detection and Stacking Fusion Model. **IEEE Access**, v. 11, p. 137475-137484, 2023. <https://doi.org/10.1109/ACCESS.2023.3339572>
- ZANIN, Daniel Suek *et al.* Agronomic performance of cultivars and advanced selections of strawberry in the South Plateau of Santa Catarina State. **Revista Ceres**, v. 66, n. 3, p. 159-167, 2019. <https://doi.org/10.1590/0034-737X201966030001>
- ZEIST, A. R & RESENDE, J. T. V. Strawberry breeding in Brazil: current momentum and perspectives. **Horticultura Brasileira**, v. 37, p. 007-016, 2019. doi: <http://dx.doi.org/10.1590/S0102-053620190101>.

ZHENG, Caiwang *et al.* Prediction of strawberry dry biomass from UAV multispectral imagery using multiple machine learning methods. **Remote Sensing**, v. 14, n. 18, p. 4511, 2022. <https://doi.org/10.3390/rs14184511>

ZHENG, Caiwang *et al.* Strawberry canopy structural parameters estimation and growth analysis from UAV multispectral imagery using a geospatial tool. **Computers and Electronics in Agriculture**, v. 226, p. 109440, 2024. <https://doi.org/10.1016/j.compag.2024.109440>

ZHENG, Caiwang; ABD-ELRAHMAN, Amr; WHITAKER, Vance. Remote sensing and machine learning in crop phenotyping and management, with an emphasis on applications in strawberry farming. **Remote Sensing**, v. 13, n. 3, p. 531, 2021. <https://doi.org/10.3390/rs13030531>

ZHONG, Ben JL. The fruit revolution in the genomic era: unveiling the journey of strawberry genetic improvement. **International Journal of Horticulture**, v. 14, n. 2, p. 72, 2024. <https://doi.org/10.5376/ijh.2024.14.0009>

ZHU, Liming *et al.* Regional assessment of soil moisture active passive enhanced L3 Soil Moisture Product and Its Application in Agriculture. **Remote Sensing**, v. 16, n. 7, p. 1225, 2024. <https://doi.org/10.3390/rs16071225>

ZINGARETTI, Laura M.; MONFORT, Amparo; PÉREZ-ENCISO, Miguel. Automatic fruit morphology phenome and genetic analysis: An application in the octoploid strawberry. **Plant Phenomics**, 2021. <https://doi.org/10.34133/2021/9812910>