

UNIVERSIDADE TECNOLÓGICA FEDERAL DO PARANÁ

PATRICK GUILHERME ROZA

**USO DE INFRAVERMELHO PRÓXIMO E VISÃO COMPUTACIONAL
COMBINADOS COM APRENDIZADO DE MÁQUINA PARA CLASSIFICAÇÃO DE
FUBÁ ORGÂNICO ADULTERADO**

PATO BRANCO

2025

PATRICK GUILHERME ROZA

**USO DE INFRAVERMELHO PRÓXIMO E VISÃO COMPUTACIONAL
COMBINADOS COM APRENDIZADO DE MÁQUINA PARA CLASSIFICAÇÃO DE
FUBÁ ORGÂNICO ADULTERADO**

**USE OF NEAR INFRARED AND COMPUTATIONAL VISION COMBINED WITH
MACHINE LEARNING FOR CLASSIFICATION OF ADULTERATED ORGANIC
CORNMEAL**

Trabalho de conclusão de curso de graduação
apresentada como requisito para obtenção do título de
Bacharel em química da Universidade Tecnológica
Federal do Paraná (UTFPR).

Orientador: Vanderlei Aparecido de Lima.

PATO BRANCO

2025



[4.0 Internacional](https://creativecommons.org/licenses/by-nc-nd/4.0/)

Esta licença permite download e compartilhamento do trabalho desde que sejam atribuídos créditos ao(s) autor(es), sem a possibilidade de alterá-lo ou utilizá-lo para fins comerciais. Conteúdos elaborados por terceiros, citados e referenciados nesta obra não são cobertos pela licença.



Ministério da Educação
Universidade Tecnológica Federal do Paraná
Campus Pato Branco
Departamento de Química
Bacharelado em Química



FOLHA DE APROVAÇÃO

USO DE INFRAVERMELHO PRÓXIMO E VISÃO COMPUTACIONAL
COMBINADOS COM APRENDIZADO DE MÁQUINA PARA CLASSIFICAÇÃO DE
FUBÁ ORGÂNICO ADULTERADO
por

PATRICK GUILHERME ROZA

Este Trabalho de Conclusão de Curso foi apresentado em 26 de junho de 2025 como requisito parcial para a obtenção do título de Bacharel em Química Industrial. O candidato foi arguido pela Banca Examinadora composta pelos professores abaixo assinados. Após deliberação, a Banca Examinadora considerou o trabalho aprovado.

Sirlei Dias Teixeira
Membro titular

Marina Leite Mitterer Daltoe
Membro titular

Ricardo da Silva Furlan
Membro titular

Nota: O Documento original e assinado pela Banca Examinadora encontra-se no SEI processo 23064.030157/2025-87 e documento 5012276.

AGRADECIMENTOS

Meus agradecimentos serão dedicados à minha família pelo apoio incondicional, incentivo incondicional, acreditarem em mim em todos os momentos e tornar possível chegar onde cheguei.

Aos meus professores e colegas da universidade, que contribuíram para meu crescimento e formação ao longo do curso.

Agradeço de maneira especial ao meu orientador Vanderlei Aparecido de Lima pela paciência, dedicação e pelas valiosas orientações durante a elaboração deste e dos demais trabalhos que realizamos juntos durante minha jornada na UTFPR.

Também gostaria de destacar a valiosa ajuda da professora Dra. Larissa Macedo Dos Santos Tonial e ao Felipe Augusto Bueno Rossi com o equipamento de espectroscopia de infravermelho próximo, seu tempo e ajuda foram de grande importância neste trabalho.

RESUMO

A crescente demanda por produtos orgânicos e seu alto valor agregado têm impulsionado casos de adulteração alimentar, comprometendo a autenticidade e a confiança do consumidor. Este trabalho propôs o desenvolvimento e avaliação de dois métodos não destrutivos para a detecção de fraudes em fubá orgânico adulterado com farinha de trigo: visão computacional por análise dos canais de cor (RGB/HSV) e espectroscopia no infravermelho próximo (NIR), ambos combinados com algoritmo de aprendizado de máquina, o *Random Forest* (RF). Foram preparadas amostras de fubá com dois níveis de adulteração (20% e 30%) e outro conjunto de amostras com fubá puro sem adulteração. As amostras foram foto-registradas e submetidas à extração de atributos cromáticos e leituras espectrais, em infravermelho próximo (NIR). Os modelos de classificação foram construídos nas linguagens Python e R, com validação para atestar a eficácia do modelo de classificação. As métricas de avaliação dos modelos foram apresentadas na forma de matriz de confusão e F1-score. O modelo baseado em atributos de cor apresentou acurácia de até 75% no conjunto de teste, com destaque para os canais de saturação e vermelho como os mais relevantes. O modelo espectroscópico alcançou acurácia superior a 99%, com elevada precisão e demonstrando excelente capacidade de generalização. Os resultados comprovam a eficácia das técnicas propostas, que se mostraram viáveis, acessíveis e promissoras para a autenticação de produtos alimentícios, podendo ser aplicadas em rotinas de controle de qualidade na indústria ou em inspeções regulatórias em alimentos.

Palavras-chave: Adulteração alimentar; Automação de análise; Agricultura orgânica; Classificação supervisionada; Autenticação de alimentos; Espectroscopia NIR.

ABSTRACT

The growing demand for organic products and their high added value have driven an increase in food adulteration cases, compromising product authenticity and consumer trust. This study proposed the development and evaluation of two non-destructive methods for detecting fraud in organic cornmeal adulterated with wheat flour: computer vision through analysis of color channels (RGB/HSV) and near-infrared (NIR) spectroscopy, both combined with a machine learning algorithm, Random Forest (RF). Samples of cornmeal with two levels of adulteration (20% and 30%) were prepared, along with a set of pure, non-adulterated cornmeal samples. The samples were photographed, and both color attributes and NIR spectral readings were extracted. The classification models were developed using Python and R, with validation to assess their classification performance. The evaluation metrics were presented using confusion matrices and F1-scores. The color-based model achieved an accuracy of up to 75% on the test set, with the saturation and red channels standing out as the most relevant. The spectroscopic model achieved an accuracy greater than 99%, demonstrating high precision and excellent generalization capability. The results confirm the effectiveness of the proposed techniques, which proved to be viable, accessible, and promising for the authentication of food products and can be applied in routine quality control in the food industry or regulatory inspections.

The growing demand for organic products and their high added value have driven cases of food adulteration, compromising authenticity and consumer confidence. This work proposed the development and evaluation of two non-destructive methods for detecting fraud in organic cornmeal adulterated with wheat flour: computer vision by color channel analysis (RGB/HSV) and near-infrared spectroscopy (NIR), both combined with machine learning algorithms. Cornmeal samples with three levels of adulteration (0%, 20% and 30%) were prepared and subjected to the extraction of chromatic and spectral attributes. The classification models were built in Python and R, with validation by accuracy, confusion matrix and F1-score. The model based on color attributes showed accuracy of up to 75% in the test set, with emphasis on the saturation and red channels as the most relevant. The spectroscopic model achieved accuracy above 99%, with high precision and demonstrating excellent generalization capacity. The results prove the effectiveness of the proposed techniques, which proved to be viable, accessible and promising for the authentication of food products, and can be applied in quality control routines in the industry or in regulatory

Keywords: Food adulteration; Analysis automation; Organic farming; Supervised grading; Food authentication; NIR spectroscopy

LISTA DE ILUSTRAÇÕES

Figura 1. Amostra de fubá acondicionada em béquer, utilizada para aquisição de imagens sob condições padronizadas de iluminação e fundo neutro, com o objetivo de extração de atributos cromáticos por meio de visão computacional.	27
Figura 2. Exemplos de imagens recortadas (ROIs) de amostras de fubá puro (OFT100FB) utilizadas na análise dos canais de cor RGB/HSV	28
Figura 3. Matriz de confusão dos dados de treino para a classificação de amostras de fubá com diferentes níveis de adulteração com farinha de trigo utilizando o algoritmo, Random Forest.....	32
Figura 4. Matriz de confusão dos dados de teste para a classificação de amostras de fubá com diferentes níveis de adulteração com farinha de trigo utilizando o algoritmo, Random Forest.....	33
Figura 5. Contribuição das componentes de cor R, H, S e V na classificação de fubá adulterado com farinha de trigo.	34
Figura 6. Valores médios das componentes de cor H, S e V nas amostras de fubá com diferentes teores de adulteração.....	36
Figura 7. Espectros de absorvância no infravermelho próximo (NIR) de amostras de fubá com diferentes níveis de adulteração por farinha de trigo, no intervalo de 900 a 1700 nm ⁻¹	37
Figura 8. Matriz de confusão para o conjunto de teste referente à classificação de amostras de fubá com diferentes níveis de adulteração por farinha de trigo, utilizando o algoritmo Random Forest aplicado aos dados espectroscópicos.....	40
Figura 9. Matriz de confusão para o conjunto de treino referente à classificação de amostras de fubá com diferentes níveis de adulteração por farinha de trigo, utilizando o algoritmo Random Forest aplicado aos dados espectroscópicos.....	41

LISTA DE TABELAS

Tabela 1 - Regiões espectrais no NIR e seus respectivos grupos funcionais e modos vibracionais associados.	38
Tabela 2 - Indicadores de precisão, revocação e F1-score para o modelo treinado (conjunto teste)	38
Tabela 3 - Indicadores de precisão, revocação e F1-score para o modelo treinado (conjunto treino).....	39

LISTA DE ABREVIATURAS E SIGLAS

ANN	Artificial Neural Network (Rede Neural Artificial)
HSV	Hue, Saturation, Value (Matiz, Saturação e Brilho)
iPCA	Interval Principal Component Analysis (Análise de Componentes Principais por Intervalos)
ML	Machine Learning (Aprendizado de Máquina)
UTFPR	Universidade Tecnológica Federal do Paraná
NIR	Near Infrared (Infravermelho Próximo)
PCA	Principal Component Analysis (Análise de Componentes Principais)
PLS	Partial Least Squares (Mínimos Quadrados Parciais)
RGB	Red, Green, Blue (Vermelho, Verde e Azul)
SVM	Support Vector Machine (Máquina de Vetores de Suporte)
VIF	Variance Inflation Factor (Fator de Inflação da Variância)

LISTA DE SÍMBOLOS

R	Intensidade do canal vermelho da imagem
G	Intensidade do canal verde da imagem
B	Intensidade do canal azul da imagem
H	Matiz (Hue)
S	Saturação (Saturation)
V	Valor de brilho (Value)
nm^{-1}	Número de onda, unidade usada em espectroscopia
O–H, C–H, N–H	Grupos funcionais presentes nas amostras analisadas
%	Porcentagem de adulteração nas amostras

SUMÁRIO

1 INTRODUÇÃO	13
2 OBJETIVOS.....	17
2.1 Geral	17
2.2 Específicos	17
3 REFERENCIAL TEÓRICO.....	18
3.1 Fubá orgânico.....	18
3.2 Fraude em alimentos.....	19
3.3 Métodos para detecção em alimentos.....	20
3.4 Padrões RGB de imagens.....	21
3.5 Espectroscopia de infravermelho próximo (NIR)	22
3.6 Análise de dados multivariados.....	24
4 MATERIAIS E MÉTODOS	26
4.1 Modelo de classificação de matrizes de fubá adulterado com farinha de trigo por meio de visão computacional.....	26
4.2 Modelo de classificação de matrizes de fubá adulterado com farinha de trigo por meio de espectroscopia infravermelho próximo (NIR).....	30
5 RESULTADOS E DISCUSSÃO	31
5.1 Análise por visão computacional	31
5.2 Análise por espectroscopia NIR	37
6 CONCLUSÃO	43
7 REFERÊNCIAS.....	45

1 INTRODUÇÃO

O fubá orgânico é uma farinha derivada do milho cultivado sem o uso de aditivos químicos, como pesticidas sintéticos, fertilizantes artificiais ou organismos geneticamente modificados (CAMPOS et al., 2018). No cultivo orgânico, utilizam-se fertilizantes naturais, como a compostagem. A produção do fubá envolve etapas como a limpeza dos grãos, moagem, que preserva os nutrientes e evita a contaminação química e, por fim, o peneiramento, que garante uma textura uniforme e remove impurezas. Para que o produto final seja sustentável, seguro e saudável para o consumo, é fundamental seguir rigorosamente os padrões da agricultura orgânica (KUEPER, 2002).

O fubá orgânico contribui para práticas agrícolas mais sustentáveis, como rotação de culturas, conservação do solo e uso eficiente da água, promovendo a conservação do ecossistema agrícola, redução de contaminação de solo e água (KAMIYAMA, 2017).

O fubá orgânico atende à demanda de consumidores que priorizam a saúde e buscam produtos naturais, sem agrotóxicos e aditivos químicos. Esse perfil de consumo tem impulsionado o crescimento do mercado de orgânicos, que vem se expandindo significativamente nos últimos anos. Produtos com certificação orgânica geram maior confiança e os consumidores estão dispostos a pagar mais por essa garantia (RIBEIRO, 2023).

O preço mais elevado do fubá orgânico em relação ao convencional se deve a diversos fatores. Entre eles, destacam-se as práticas agrícolas diferenciadas, como a adubação verde, a ausência de agrotóxicos e fertilizantes sintéticos, e os custos com certificação orgânica. Além disso, aspectos logísticos e a alta demanda do mercado contribuem para o valor agregado desses produtos (REGANOLD, 2015).

No entanto, a valorização dos produtos orgânicos também abre espaço para fraudes. A alta demanda e o custo elevado incentivam práticas fraudulentas que comprometem a confiança dos consumidores e a credibilidade do setor (ANAGAW, 2024). A prática de adulteração de alimentos orgânicos consiste em vender alimentos convencionais, rotulados como orgânicos, ou contaminação intencional ou não, de fubá orgânico com produtos não orgânicos (PRADO et al., 2021). Segundo Bannor et al. (2023), os produtos orgânicos estão entre os mais visados para adulteração, sendo

o grupo "organic foods" um dos principais alvos de fraudes alimentares, como rotulagem incorreta e falsificação de origem.

Os consumidores pagam preços elevados por produtos orgânicos e, muitas vezes, não recebem a qualidade esperada. Ao mesmo tempo, produtores que seguem rigorosamente os padrões orgânicos podem enfrentar queda nas vendas e perda de confiança no mercado. Isso gera uma necessidade crescente de inspeções e regulamentações, que resultam em custos adicionais para os consumidores. Todos esses fatores contribuem para um grave prejuízo econômico (HENSON, 2005; JOHNSON, 2014).

Além dos impactos econômicos, há também consequências ambientais significativas. Como os produtos adulterados não seguem os padrões orgânicos, ocorre o uso de insumos industriais que prejudicam o solo, contaminam os recursos hídricos e afetam negativamente os ecossistemas (MATTHEW, 2023).

A fraude alimentar representa uma ameaça significativa à saúde pública, podendo causar riscos diretos, indiretos e técnicos aos consumidores. Entre os efeitos adversos relatados estão reações alérgicas graves devido à rotulagem incorreta de frutos do mar, distúrbios metabólicos causados por adulterações com xaropes em mel, e problemas cardiovasculares e hepáticos associados à substituição de óleos nobres por óleos de menor qualidade. Além disso, substâncias tóxicas como bisfenol A, presentes em embalagens adulteradas, foram associadas a infertilidade, obesidade precoce, ansiedade e câncer. A ingestão de produtos com contaminantes como paracetamol em bebidas energéticas também resultou em sintomas neurológicos e cardiovasculares, especialmente em jovens adultos. Essas fraudes, motivadas por ganhos econômicos, colocam em risco a segurança dos alimentos e exigem maior vigilância e tecnologias analíticas eficazes para proteção da saúde da população (Aslam et al., 2024).

De acordo com Udayanga et al. (2023), em seu estudo sobre um sistema de imagem multiespectral de duplo modo, a indústria alimentícia vem integrando tecnologias de imagem baseadas em padrões de cor (RGB e além) com aprendizado de máquina para detectar adulterações. O trabalho apresenta um sistema que combina modos de refletância e transmitância para examinar espécimes alimentícios, incluindo especiarias como cúrcuma, alcançando até 99% de precisão na identificação de adulteração. Essa abordagem óptica, suplementada por análises cromatográficas (e.g., LC-MS/MS), PCR e espectrometria isotópica, reforça a

tendência de adoção de ferramentas visuais não invasivas para autenticação alimentar moderna. A detecção de fraudes em alimentos pela cor pode ser feita por captura de imagens (captação de imagens por fotos) e algoritmos de aprendizagem de máquina (SILVA, 2021).

A ferramenta “Color Picker” possui funções de leitura e processamento de imagens, assim como o ChemoStat, um software livre. De acordo com Helfer (2015), o ChemoStat pode ser utilizado para análise de espectros no infravermelho, bem como na análise de imagens. Nesse software, as imagens podem ser importadas nos formatos “.bmp”, “.jpg” e “.png”, permitindo a análise de histogramas RGB ou de valores individuais de componentes como R, G, B, R relativo, B relativo, G relativo, H (tonalidade), S (saturação), V (valor de iluminância), L (brilho) e I (intensidade). O ChemoStat oferece ferramentas como análise de agrupamento hierárquico (HCA), análise de componentes principais (PCA), análise de componentes principais por intervalos (iPCA), além de diversas técnicas de tratamento de dados. Por ser gratuito, leve e de simples manuseio, o ChemoStat se mostra um software confiável.

A extração dos padrões de cores gera um banco de dados que necessita ser tratado por modelos multivariados. Os tratamentos de dados multivariados combinam um conjunto de técnicas estatísticas que analisam dados que tem têm mais de uma variável dependente (HAIR, 2019). Essa abordagem é muito utilizada em problemas de química, biologia, economia e ciências humanas para interpretar dados extraídos de pesquisas (JOHNSON, 2007). Além de métodos estatísticos para analisar dados, podem se utilizar também linguagens de programação (KHARE, 2024).

Nas ciências de modo geral, duas linguagens de programação são proeminentes estão, as linguagens Python e R. Python, é uma linguagem de alto nível, e é uma linguagem simples, e que apresentam ampla gama de bibliotecas e ferramentas para análises de dados, usados e é usado para manipulação, visualização, modelagem de dados. Além disso, as linguagens Python e R, têm disponíveis algoritmos de aprendizado de máquina para análise de dados (PHUING, 2020). R também é uma linguagem de programação de alto nível, projetado para análises de dados e computação estatística possuindo grande variedade de pacotes e gráficos, modelagem estatística, visualização e relatórios sendo muito utilizado na academia, análise de dados empresariais e ciência de dados (SUN, 2021).

Hoje, a identificação dessas fraudes e adulterações é viabilizada pelo avanço da capacidade computacional, impulsionada pelo uso de algoritmos de aprendizagem

de máquina. Computadores são treinados por meio de algoritmos para realizar análises independentes e tomar decisões. Tais algoritmos são essenciais para o processamento de grandes volumes de dados, automatização de processos, análise de informações e aumento da eficiência de softwares (JORDAM, 2015).

Nesse contexto, as técnicas de espectroscopia ganharam destaque nas análises químicas, com sendo o infravermelho próximo (NIR) sendo uma das mais eficazes para identificar e quantificar espécies químicas em diversas matrizes. A leitura no espectrofotômetro NIR abrange comprimentos de onda entre aproximadamente 780 nm e 2500 nm (FENG, 2019). Equipamentos de NIR vêm sendo amplamente aplicados em áreas como análises químicas, agricultura de precisão, controle de qualidade industrial e, de forma crescente, na detecção de fraudes em alimentos (VOCCIO, 2024; ZHONG, 2024; TATA, 2024).

O presente estudo propõe a avaliação de duas abordagens analíticas para a detecção de adulteração em amostras de fubá orgânico com adição de farinha de trigo. A primeira abordagem baseia-se na análise dos padrões cromáticos das amostras, por meio da quantificação dos valores dos canais RGB extraídos de imagens digitais, enquanto a segunda utiliza espectroscopia no infravermelho próximo (NIR), uma técnica não destrutiva amplamente aplicada na caracterização de alimentos. Os dados gerados por ambas as técnicas foram processados e modelados por algoritmos de aprendizado de máquina, com implementação nas linguagens R e Python, visando à construção de modelos preditivos robustos para a identificação de fraudes alimentares.

2 OBJETIVOS

2.1 Geral

Avaliar a eficiência do uso de ferramentas de visão computacional, aprendizado de máquina e infravermelho próximo para a identificação de fraudes em amostras de fubá orgânico adulterado com farinha de trigo.

2.2 Específicos

- Analisar as amostras por meio de espectroscopia de infravermelho próximo (NIR) para construção de modelos preditivos de adulteração de fubá orgânico com farinha de trigo.
- Modelar o banco de dados por meio de aprendizado de máquina (ML) utilizando a linguagem de programação Python, por meio de algoritmos de aprendizado de máquina, para identificação das fraudes em fubá orgânico.
- Avaliar as métricas de desempenho dos algoritmos utilizados nas modelagens de dados por infravermelho próximo (NIR) e canais RGB das amostras de fubá adulteradas.

3 REFERENCIAL TEÓRICO

3.1 Fubá orgânico

A Revista FT (2023) destaca que o uso indiscriminado de agrotóxicos no Brasil contamina o solo, a água e comunidades rurais próximas, com evidências de intoxicação decorrentes da pulverização aérea, o que evidencia falhas na aplicação das normas e na fiscalização.

Nesse contexto, o consumo de produtos orgânicos no Brasil tem crescido de forma significativa, impulsionado principalmente por preocupações com a saúde e a preservação ambiental. Estudos científicos reforçam esse cenário: Deliberador et al. (2024), em pesquisa com consumidores brasileiros, mostraram que a preocupação ambiental e a consciência sobre saúde são fortes determinantes da intenção de comprar produtos orgânicos

Em relação à disposição em pagar preços mais altos por orgânicos, uma revisão sistemática concluiu que fatores como preocupação com saúde, percepção de qualidade, valores ambientais e localidade influenciam significativamente essa disposição (HU et al., 2024). Adicionalmente, pesquisa quantitativa realizada no Brasil apontou que grande parte dos consumidores está disposta a pagar até 10% a mais por alimentos orgânicos (WATANABE et al., 2023).

Em nível internacional, Huo et al. (2023) analisaram o impacto da renda e da qualidade ambiental na disposição de pagar mais por alimentos sustentáveis, concluindo que consumidores, especialmente em regiões mais conscientes ambientalmente, valorizam os alimentos orgânicos como escolhas éticas e saudáveis. Essas evidências demonstram que a crescente preferência por alimentos orgânicos está fundamentada em dados concretos de comportamento de consumo, refletindo uma mudança cultural em direção à saúde e à sustentabilidade.

O cultivo de produtos orgânicos dispensa o uso de agrotóxicos e agentes químicos, o que torna o processo mais complexo e em contrapartida energeticamente favorável (CHMELÍKOVÁ et al., 2024). O cultivo de milho orgânico passa por um complexo e frágil conjunto de técnicas como adubação equilibrada, como pode-se averiguar na obra "*Milho orgânico: técnicas de cultivo*", (2017). Nas recomendações pelo autor da obra, o solo não deve ser demasiadamente manipulado para não prejudicar as atividades microbianas que ocorrem no solo. No cultivo de milho orgânico, são utilizados fosfatos, farinha de osso, rochas, e minerais são empregados

como fontes de Ca, Mg, P₂O₅ e K₂O. Essas atividades são realizadas, entre outras técnicas, a fim de garantir a sustentabilidade, produtividade e qualidade do milho.

Após o milho ser colhido ele passa por uma seleção, a fim de selecionar os grãos de melhor qualidade, em seguida é limpo e descarregado em moinho para obtenção da farinha de milho que segue para equipamentos de separação e peneiramento a fim de separar a casca e do gérmen do milho. Esse processo de moagem do milho é utilizado para obter farinha, fubá e amido (CARDOSO, 2012; IDUGEL, 2023).

3.2 Fraude em alimentos

Fraude em alimentos pode ser definida como o ato intencional de adulterar, falsificar ou deturpar ingredientes, rótulos ou processos de produção com o objetivo de obter vantagem econômica (LAWRENCE, 2024). Essa prática representa uma ameaça à integridade do setor alimentício, comprometendo a segurança do consumidor, a credibilidade de produtores legítimos e a sustentabilidade do mercado. No contexto dos produtos orgânicos, essa problemática é ainda mais crítica, visto que esses alimentos possuem maior valor agregado e demandam práticas específicas de cultivo, certificação e logística, tornando-os alvos frequentes de adulteração (SACCHI; ROMANELLO; CANAVARI, 2024).

Casos recentes evidenciam a diversidade de produtos afetados. Por exemplo, a substituição parcial de noz-moscada por casca moída foi detectada com alta precisão utilizando espectroscopia no infravermelho próximo (NIR) aliada a algoritmos como redes neurais artificiais (ANN) e máquinas de vetores de suporte (SVM), alcançando acurácias superiores a 95% (RAHMAN et al., 2023). Essa prática, além de enganar o consumidor, altera o valor nutricional e funcional do produto, gerando prejuízos econômicos e sanitários.

A adulteração de leite com água é outro exemplo comum e potencialmente perigoso, especialmente quando afeta o teor nutricional e aumenta os riscos microbiológicos. Em estudo recente, sensores portáteis de NIR aliados a algoritmos de aprendizado de máquina foram capazes de detectar com precisão esse tipo de fraude, demonstrando a viabilidade de métodos não destrutivos e acessíveis para controle de qualidade em campo (KUMAR et al., 2024).

Produtos de origem vegetal, como a cúrcuma, também são frequentemente adulterados com amido para aumentar o volume comercializado. Essa adulteração,

visualmente imperceptível, foi identificada por meio de espectroscopia NIR portátil e modelos XGBoost, com excelente desempenho na classificação entre amostras puras e adulteradas (SINGH et al., 2024). Essa metodologia permitiu ainda identificar regiões espectrais críticas associadas a grupos funcionais característicos dos adulterantes.

No setor de grãos, a fraude em arroz por mistura com grãos de menor qualidade ou quebrados foi detectada por sensores ópticos de baixo custo combinados com algoritmos Random Forest, reforçando que mesmo tecnologias acessíveis podem ter grande impacto na fiscalização da cadeia de alimentos (SHAH et al., 2022).

Fraudes em carnes processadas também têm sido objeto de estudo. Em amostras de salsichas, a identificação de espécies diferentes daquelas informadas no rótulo foi possível por meio da espectroscopia NIR combinada com análise discriminante linear (LDA), atingindo acurácia superior a 97% (ZHANG et al., 2025). Este tipo de adulteração pode ter implicações legais, culturais e religiosas, além de impactar diretamente a confiança do consumidor.

Esses exemplos reforçam que a fraude alimentar não é apenas uma preocupação teórica, mas uma realidade presente e complexa, que exige a integração de tecnologias modernas de detecção. A combinação de espectroscopia no infravermelho com algoritmos de aprendizado de máquina tem se mostrado extremamente eficaz e promissora, tanto por sua rapidez e sensibilidade quanto pela possibilidade de aplicação em tempo real e sem destruição das amostras. Essas evidências fortalecem a importância de estudos como o presente, que propõem modelos eficientes e replicáveis para autenticação de alimentos, especialmente no contexto sensível e valorizado dos produtos orgânicos.

3.3 Métodos para detecção em alimentos

Na literatura há muitos artigos publicados sobre metodologias para identificação de fraudes em alimentos. Nos últimos anos, têm aumentado significativamente os estudos nessa área, dando a impressão que os trabalhos mais influentes foram publicados nos últimos cinco anos. De maneira semelhante a esse fato, um notável crescimento de pesquisas em relação a produtos orgânicos tem sido realizado. Dentre os produtos mais visados por fraudadores estão: óleo de oliva, frutos do mar, e suplementos alimentares (Richard Kwasi Bannor, Kingsley Kofi Arthur, Dixon

Oppong, Helena Oppong-Kyeremeh, 2023). Nesse artigo, os autores também afirmam a necessidade de mais pesquisas na área, pois apesar de existirem muitos artigos na área de adulteração, ainda seria interessante a criação de novas ferramentas para análises de alimentos fraudados.

Darra et al. (2017) avaliou a adulteração do melaço de romã, um xarope tradicional do oriente médio, com melaço de tâmara, um xarope de baixo valor econômico. Em sua pesquisa foi utilizado métodos espectroscópicos UV/VIS e infravermelho (ATR-FTIR) e cromatográfico (HPLC) para detectar adulterações do melaço de romã.

Espectroscopia de ressonância magnética nuclear (RMN) também se mostra uma ferramenta de grande importância para análise da caracterização química de produtos alimentícios. A origem, a qualidade e ainda, a autenticidade de alimentos pode ser aferida por RMN (Qian Qu, Lan Jin, 2022).

Outra ferramenta importante no controle de qualidade de alimentos é a espectroscopia por infravermelho próximo (NIR). Um estudo realizado por Ainara López-Maestresalasa, Kizkitza Insaustib, Carmen Jaréna, Claudia Pérez-Roncala, Olaia Urrutiab, Mari Jose Beriainb, Silvia Arazuria (2019), com o objetivo de diferenciar carnes moída, de cordeiro e bovina, por meio do infravermelho concluiu que a aplicação do NIR se mostrou eficaz, distinguindo-se a carne pura da adulterada com ótima acurácia e precisão, além de ser um método simples e não destrutivo.

3.4 Padrões RGB de imagens

O primeiro padrão de cores foi criado em 1931 pelo comitê CIE (Commission Internationale de l'Éclairage) chamado de CIE-RGB, que estabeleceu um padrão de cores primárias sendo as cores vermelho, verde e azul (RGB). Atualmente cada pixel da tela de um dispositivo eletrônico como aparelho celular, alterna o valor dessas três cores conforme necessidade para a formação de uma imagem (Scurri, 2002, livro "Fundamento de Imagem Digital"). A extração de características como número total de objetos, propriedades geométricas e propriedades de cor: matiz, saturação e luminosidade são possíveis por algoritmos de aprendizado de máquina (ML). Além dessas características, ainda é possível também extrair os dados da imagem em forma de histograma.

O método de extração de padrão RGB por tons de cinza é utilizado para simplificar a análise de imagem, convertendo uma imagem colorida em uma imagem

em tons de cinza para facilitar a análise de seus componentes visuais. Este processo pode ser feito combinando os valores dos três canais de intensidade em um único valor. Após essa extração de tons de cinza de uma imagem, um histograma é criado com a representação das distribuições das intensidades de cinza da imagem, mostrando a frequência de cada nível de cinza, e a partir disto é possível extrair padrões e características da imagem, como contraste, bordas e texturas, que pode ser utilizado para diversas aplicações como visão computacional, processamento de imagem e reconhecimento de objetos (Rafael C. Gonzalez e Richard E. Woods, 2018).

Scurri (2002) comenta sobre alguns dos melhores softwares de edição e análise de imagens, citando o Adobe Photoshop como o programa mais confiável, na versão 5,0, para editoração de imagens digitais, possuindo diversas funções e sendo capaz de extrair padrões RGB em suas ferramentas de edição. O programa GIMP também é citado no campo das edições de imagens digitais como sendo muito popular e um software gratuito.

No campo dos softwares de análise da imagem está disponível de forma gratuita, o ChemoStat que é um programa desenvolvido para análises de dados envolvendo quimiometria, uma ciência que utiliza a aplicação de métodos matemáticos e estatísticos no planejamento de experimentos e análise e interpretação de dados químicos, combinando estatística, química e ciência da computação para resolver problemas complexos (FILZMOSE, VARMUZA., 2016). Outros programas de fácil acesso e manuseio que oferecesse potencial para análise de dados e de imagens são o Matlab® e Pirouette®, destinado à alunos, professores e pesquisadores, abrangendo a análise exploratória de dados. Porém esses dois últimos são softwares pagos e o acesso é restrito devido a seu elevado custo.

3.5 Espectroscopia de infravermelho próximo (NIR)

A espectroscopia no infravermelho (IR) é uma técnica analítica não destrutiva amplamente utilizada para investigar a estrutura molecular de substâncias orgânicas e inorgânicas. Por meio da interação da radiação infravermelha com as moléculas, é possível obter informações sobre as vibrações específicas das ligações químicas, resultando em espectros característicos que funcionam como "impressões digitais químicas" dos compostos. Essa técnica é amplamente empregada nas faixas do infravermelho médio (MIR), próximo (NIR) e por transformada de Fourier (FTIR),

sendo essencial na análise de alimentos, tecidos biológicos, solos, entre outros materiais (FERGUSON et al., 2022; SACCHI; ROMANELLO; CANAVARI, 2024).

Com a evolução das técnicas de aquisição espectral, o desafio passou a ser o tratamento eficiente do grande volume de dados gerados. A alta dimensionalidade e a presença de variáveis altamente correlacionadas tornam a análise complexa. Para lidar com isso, métodos de análise multivariada como a Análise de Componentes Principais (PCA), Regressão por Mínimos Quadrados Parciais (PLS), versões iterativas como i-PLS e técnicas de seleção de variáveis como UVE (eliminador de variáveis irrelevantes) são amplamente aplicados para reduzir dimensionalidade e extrair informações relevantes (SCIENCEDATA, 2023; LAS BRUMAS, 2023). Essas técnicas permitem que os dados espectrais sejam interpretados de forma mais eficiente e usados como base para modelos preditivos.

A incorporação de métodos de aprendizado de máquina (machine learning) tem elevado significativamente o potencial da espectroscopia IR. Classificadores como Máquinas de Vetores de Suporte (SVM), Redes Neurais Artificiais (ANN), Florestas Aleatórias (RF), Regressão Logística e algoritmos como XGBoost são utilizados para desenvolver modelos que identificam padrões complexos em dados espectrais e realizam classificações ou previsões com alta acurácia (FERGUSON et al., 2022; PUNJABI et al., 2025). Estudos mostram que essas abordagens podem ser aplicadas com sucesso na detecção precoce de câncer, análise de células sanguíneas, caracterização de solos e até na simulação de espectros com base em estruturas moleculares teóricas.

Por fim, a integração entre espectroscopia no infravermelho, análise multivariada e aprendizado de máquina representa uma tendência consolidada na ciência de dados aplicada à química, agricultura, saúde e ciências ambientais. Além de otimizar o uso de dados espectrais, essa combinação permite desenvolver ferramentas rápidas, precisas e não destrutivas para tomada de decisão e controle de qualidade em processos industriais e laboratoriais. Com a disseminação de bibliotecas espectrais públicas e o avanço dos algoritmos de IA, espera-se um crescimento ainda maior dessas aplicações nos próximos anos (REZENDE et al., 2023; SACCHI; ROMANELLO; CANAVARI, 2024).

3.6 Análise de dados multivariados

A análise de dados multivariados é uma área da estatística voltada para o estudo simultâneo de múltiplas variáveis, buscando entender padrões, correlações e estruturas subjacentes nos dados. Técnicas consagradas como PCA (Análise de Componentes Principais), PLS (Regressão por Mínimos Quadrados Parciais), CCA (Análise de Correlação Canônica), LDA (Análise Discriminante Linear) e suas variantes (como SPCA, DPLS etc.) são empregadas para reduzir a dimensionalidade, extrair componentes relevantes e selecionar variáveis importantes em grandes conjuntos de dados, tornando-os mais interpretáveis e manejáveis (LIPOVETSKY, 2022).

A análise multivariada de dados, como a Análise de Componentes Principais (PCA) e o método supervisionado OPLS-DA, demonstra sua importância em sistemas complexos de alimentos, saúde e meio ambiente ao identificar agrupamentos naturais e eliminar redundâncias em conjuntos de dados químicos. Como demonstrado por TARAPOULOUZI et al. (2021), essa aplicação é ilustrada no estudo que utilizou técnicas quimiométricas avançadas para discriminar azeites gregos com base em assinaturas isotópicas ($\delta^{13}\text{C}$, $\delta^{18}\text{O}$, $\delta^2\text{H}$). A PCA foi empregada inicialmente para explorar padrões nos dados, enquanto a OPLS-DA permitiu uma classificação robusta, alcançando 93,75% de acurácia no treinamento e 100% no teste, validando a eficácia dessas abordagens estatísticas. Além disso, o estudo destaca como fatores ambientais, como clima e solo, influenciam as assinaturas isotópicas, conectando-se a questões de autenticidade, saúde e sustentabilidade.

A Regressão por Mínimos Quadrados Parciais (PLSR) é amplamente aplicada na quantificação de adulterantes em alimentos, combinada com dados espectroscópicos. Por exemplo, em sucos comerciais adulterados com sacarina, utilizou-se FT-NIR com PLS-DA para classificar amostras e PLSR para estimar o teor de adulterante, obtendo $R^2 = 97,04\%$ e $\text{RMSECV} = 0,88\%$ (MABOOD et al., 2020).

A integração de técnicas multivariadas com espectroscopia (MIR/NIR) e métodos como MCR-ALS e PLS-R tem se consolidado como estratégia valiosa para análises químicas de extratos naturais. Conforme relatado por Câmara et al. (2022), essa abordagem permitiu prever com alta precisão a atividade antioxidante de extratos de cravo e romã ($R^2 \geq 0,962$, $\text{RMSEP} \leq 3,38\%$), além de possibilitar a identificação dos perfis espectrais puros dos compostos ativos e interferentes, o que enriquece a

interpretação analítica. A combinação entre modelagem preditiva (PLS-R) e decomposição espectral (MCR-ALS) provê uma solução robusta para lidar com ruídos, redundâncias e interferência nos dados, ao mesmo tempo em que reduz custos e impactos ambientais, fatores fundamentais para a aplicabilidade industrial e acadêmica moderna.

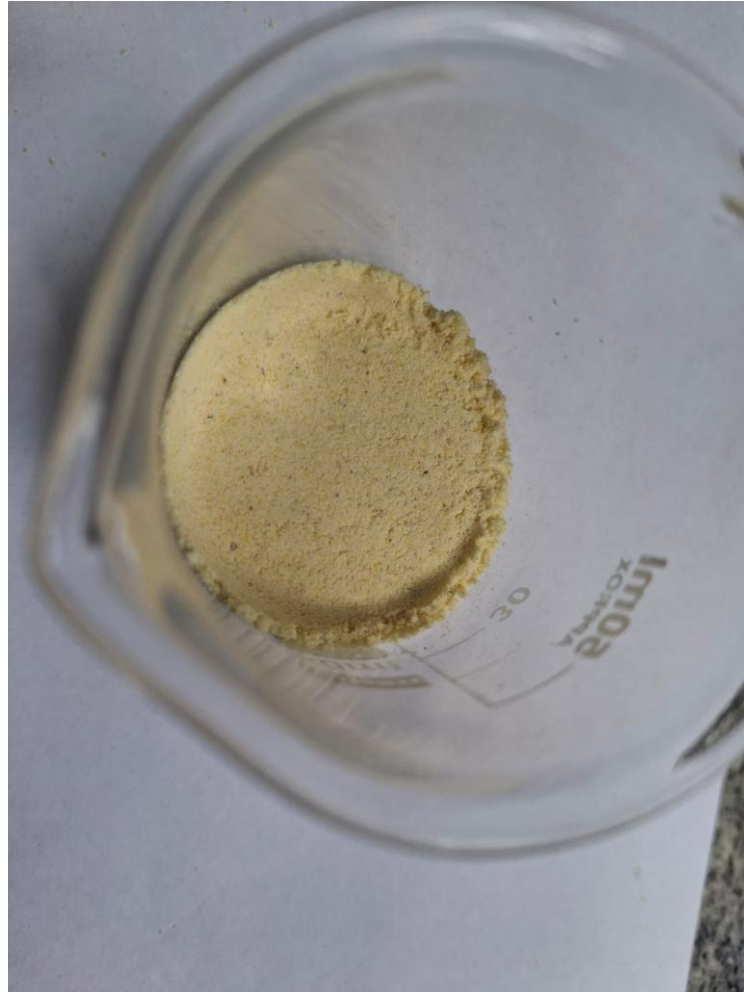
4 MATERIAIS E MÉTODOS

4.1 Modelo de classificação de matrizes de fubá adulterado com farinha de trigo por meio de visão computacional

As amostras de fubá (75) foram preparadas em três proporções distintas de adulteração com farinha de trigo: 0FT100FB (fubá puro), 20FT80FB (20% de farinha de trigo) e 30FT70FB (30% de farinha de trigo). Essas porcentagens foram escolhidas por representarem níveis realistas de fraude, difíceis de detectar visualmente, mas com impacto mensurável nas propriedades cromáticas e espectrais do fubá. Essas proporções permitiram a formação de classes bem definidas, facilitando a aplicação de algoritmos de aprendizado de máquina e garantindo a construção de modelos preditivos robustos para identificação de adulterações.

As imagens dessas amostras foram obtidas sob condições padronizadas de iluminação e fundo neutro, por meio de um aparelho celular marca Samsung A35, e as regiões de interesse (ROIs) foram selecionadas para extração das informações de cor.

Figura 1. Amostra de fubá acondicionada em béquer, utilizada para aquisição de imagens sob condições padronizadas de iluminação e fundo neutro, com o objetivo de extração de atributos cromáticos por meio de visão computacional.



Fonte: Elaborado pelo autor (2025)

Os recortes das imagens (ROIs) foram feitos no software Gimp versão 2.10.32. A extração das variáveis preditoras foi realizada com base nos canais RGB (Red, Green, Blue) e convertida para o espaço de cor HSV (Hue, Saturation, Value), totalizando seis atributos: R, G, B, H, S e V, no software ChemoStat2025. A organização dos dados e a construção dos modelos foram realizadas utilizando a linguagem de programação Python (versão 3.11) no ambiente Google Colaboratory.

Figura 2. Exemplos de imagens recortadas (ROIs) de amostras de fubá puro (0FT100FB) utilizadas na análise dos canais de cor RGB/HSV



Fonte: Elaborado pelo autor (2025)

O fator de inflação de variância (VIF), uma métrica estatística utilizada para verificar a presença de multicolinearidade entre variáveis preditoras. Valores elevados indicam redundância, o que pode comprometer a estabilidade dos modelos. Assim, variáveis com VIF acima de um determinado limite foram eliminadas, mantendo apenas aquelas com menor correlação entre si para treinar o algoritmo de classificação (SHRESTHA, 2020).

Foi aplicada a técnica de aumento amostral (ou data augmentation) com o objetivo de expandir o conjunto de dados de imagens utilizadas na modelagem por visão computacional. Essa técnica consiste em gerar novas instâncias de dados a partir das amostras originais, por meio de recortes múltiplos das regiões de interesse (ROIs), simulando variações reais nas características visuais sem alterar a classe da amostra. Com isso, o número inicial de 75 imagens foi ampliado para 225, aumentando a diversidade dos dados disponíveis para o treinamento dos modelos de classificação.

A finalidade dessa abordagem foi melhorar a robustez e a capacidade de generalização do modelo de aprendizado de máquina, reduzindo o risco de sobreajuste (overfitting), um problema comum em conjuntos de dados pequenos. Ao expor o algoritmo a mais variações visuais dentro das mesmas classes, o modelo se torna mais eficaz na identificação de padrões relevantes e na correta classificação de novas amostras, mesmo diante de pequenas variações na iluminação, textura ou coloração (SHORTEN; KHOSHGOFTAAR, 2019).

As variáveis selecionadas foram R, H, S e V e permaneceram como variáveis preditoras. A variável Classe, contendo as três classes (0FT100FB, 20FT80FB e 30FT70FB) das matrizes de fubá adulterado com farinha de trigo, foi a variável alvo, na construção dos modelos de classificação.

Foram utilizados os seguintes pacotes Python: (1) pandas para manipulação e organização de dados em tabelas, (2) numpy para cálculos numéricos e vetoriais, (3) matplotlib e seaborn para geração de gráficos e visualizações (incluindo matriz de confusão), (4) scikit-learn (sklearn) para divisão dos dados, normalização, métricas de avaliação e busca em grade (GridSearchCV) e (5) sklearn.ensemble.RandomForestClassifier para construção do modelo preditivo com o algoritmo Random Forest.

Os dados foram divididos em 70% para treinamento e 30% para teste, utilizando a técnica de hold-out, uma abordagem comum em tarefas de aprendizado de máquina supervisionado. Essa técnica consiste em separar uma parte dos dados para que o modelo aprenda os padrões (treinamento) e outra parte, não vista durante o treinamento, para avaliar sua capacidade de generalização (teste).

A escolha da proporção 70/30 busca equilibrar a quantidade de dados disponíveis para aprendizado com a necessidade de uma avaliação confiável do desempenho do modelo. Com 70% dos dados, o algoritmo tem acesso a uma amostra suficiente para capturar os padrões relevantes, enquanto os 30% restantes permitem testar o modelo com dados novos, garantindo uma validação realista e minimizando o risco de sobreajuste (Siva Kumar et al., 2024).

O algoritmo RandomForestClassifier foi aplicado para classificação multiclasse. O ajuste de hiperparâmetros foi realizado por meio da busca em grade com validação cruzada (GridSearchCV). A acurácia foi utilizada como métrica principal de desempenho, e os resultados foram avaliados por meio de matrizes de confusão e relatórios de classificação. Gráficos de importância das variáveis foram utilizados para identificar os atributos cromáticos mais relevantes para a classificação das classes de fubás adulterados com farinha de trigo.

4.2 Modelo de classificação de matrizes de fubá adulterado com farinha de trigo por meio de espectroscopia infravermelho próximo (NIR)

Além de dados de visão computacional, a presente investigação utilizou espectroscopia na região do infravermelho próximo (NIR) combinada com algoritmos de aprendizado de máquina para classificar amostras de fubá com diferentes níveis de adulteração por farinha de trigo. O conjunto de dados foi composto por 75 amostras espectrais originais, posteriormente ampliadas para 225 amostras por meio de replicação, visando aumentar a robustez do modelo.

A matriz espectral foi composta por variáveis representando os números de onda em nm^{-1} e as respectivas intensidades de absorbância. As regiões espectrais que apresentaram maior instabilidade nos sinais (baixo sinal-ruído) foram excluídas da modelagem, com o objetivo de melhorar a qualidade dos dados e a precisão dos modelos preditivos. Essa prática é comum em espectroscopia aplicada a alimentos, pois evita que informações distorcidas prejudiquem a extração de variáveis relevantes, especialmente em análises de classificação com aprendizado de máquina.

As amostras foram analisadas em equipamento de espectroscopia no infravermelho próximo (NIR), gerando os espectrogramas correspondentes. As características espectrais foram extraídas diretamente das colunas do conjunto de dados, sendo a variável *Class* utilizada para representar os dois níveis de adulteração (20FT80FB e 30FT70FB), além do grupo controle (0FT100FB — fubá puro). Os dados foram divididos em conjuntos de treino e teste, com 70% das amostras destinadas ao treino e 30% ao teste, aplicando-se estratificação das classes.

Para a modelagem preditiva, foi empregado o algoritmo Random Forest Classifier da biblioteca scikit-learn, com otimização de hiperparâmetros via GridSearchCV. O grid incluiu variações nos parâmetros *n_estimators*, *max_depth*, *min_samples_split*, *min_samples_leaf* e *bootstrap*, com validação cruzada de 5 vezes (*cv=5*). A métrica de avaliação utilizada na busca foi a acurácia. A performance do modelo final foi avaliada com base na matriz de confusão e relatório de classificação (precisão, revocação, F1-score).

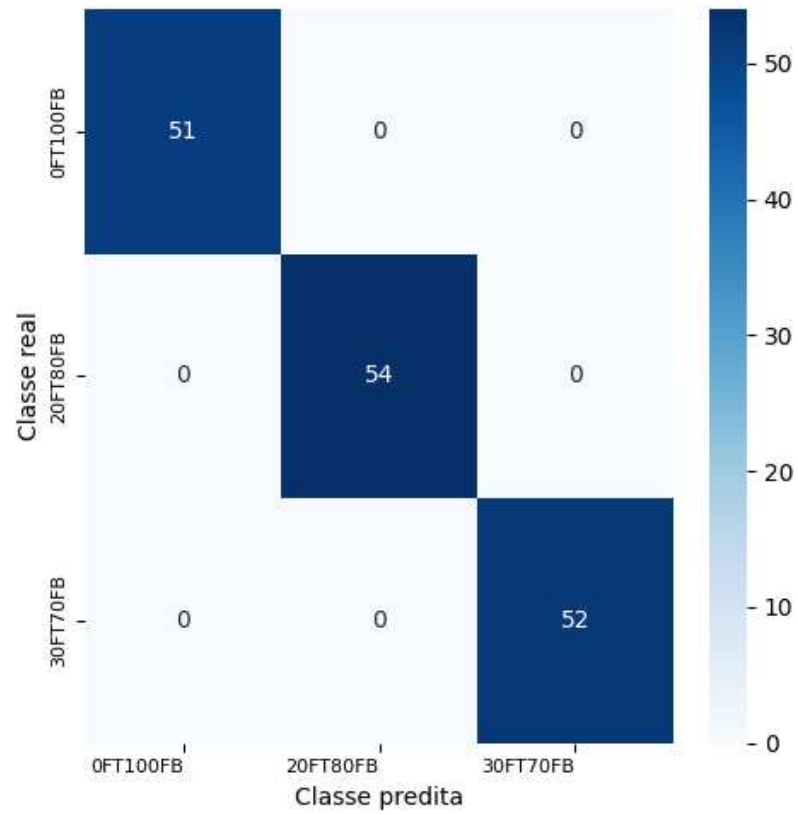
5 RESULTADOS E DISCUSSÃO

5.1 Análise por visão computacional

Os modelos treinados com os atributos de cor (R, H, S, V) demonstraram bom desempenho na classificação das amostras, indicando que a adulteração do fubá com farinha de trigo altera de maneira detectável as características cromáticas do produto.

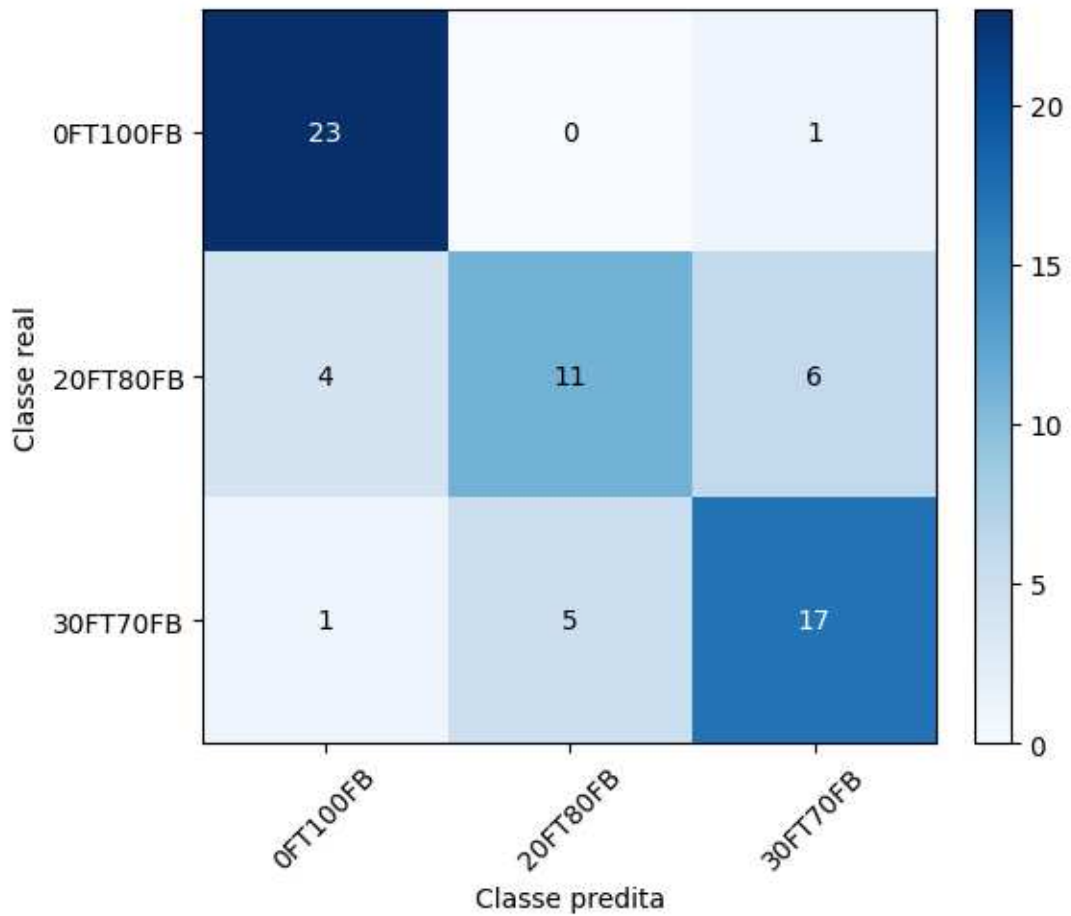
A acurácia média obtida pelo modelo otimizado foi satisfatória tanto no conjunto de treinamento (100%) quanto no teste (acurácia 75%), sugerindo capacidade de generalização e baixo risco de sobreajuste (Figuras 1 e 2). As matrizes de confusão revelaram que a classe 0FT100FB (fubá puro) foi a mais facilmente reconhecida, com altos índices de acerto (Figura 1). As classes 20FT80FB e 30FT70FB apresentaram maior taxa de confusão entre si, indicando que a adulteração em pequenas proporções pode gerar padrões cromáticos similares.

Figura 3. Matriz de confusão dos dados de treino para a classificação de amostras de fubá com diferentes níveis de adulteração com farinha de trigo utilizando o algoritmo, Random Forest.



Fonte: Elaborado pelo autor (2025)

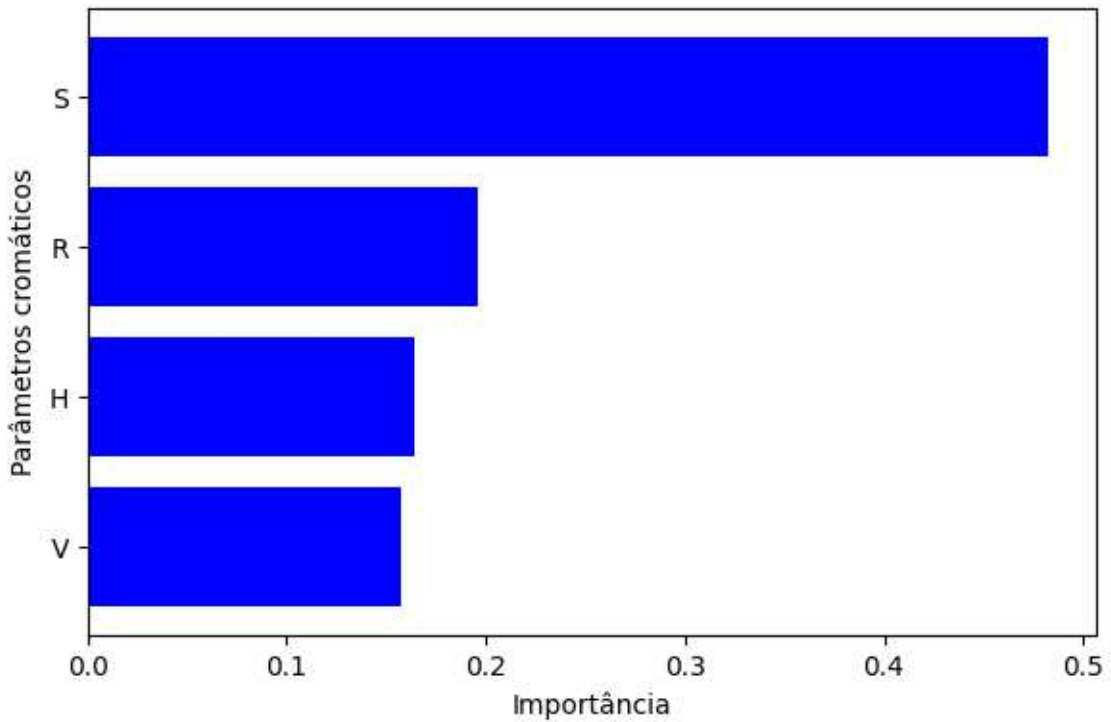
Figura 4. Matriz de confusão dos dados de teste para a classificação de amostras de fubá com diferentes níveis de adulteração com farinha de trigo utilizando o algoritmo, Random Forest



Fonte: Elaborado pelo autor (2025)

A análise da importância das variáveis (Figura 3) indicou que os atributos mais relevantes para o modelo foram S (saturação) e o R (vermelho), o que está de acordo com a expectativa de que a percepção visual da cor amarela do fubá é fortemente influenciada pela intensidade luminosa e pelos tons avermelhados.

Figura 5. Contribuição das componentes de cor R, H, S e V na classificação de fubá adulterado com farinha de trigo.



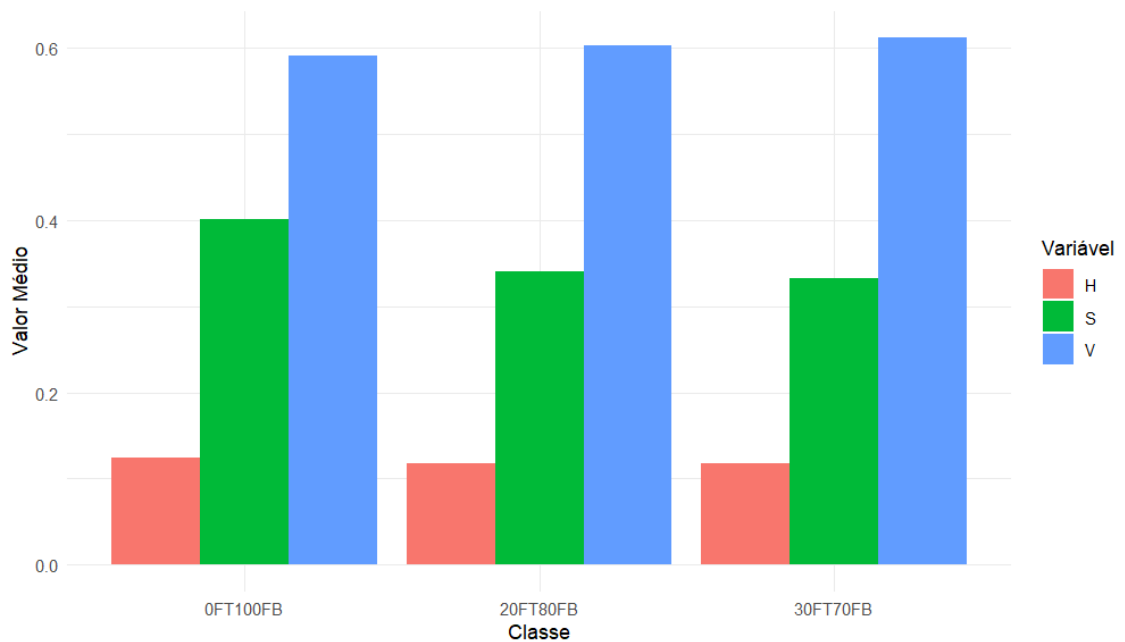
Fonte: Elaborado pelo autor (2025)

Os valores médios dos tons de cinza do canal R (*Red*) para as amostras com 100% de fubá (0FT100FB) foram de $151 \pm 8,22$; para as amostras com 20% de farinha de trigo (20FT80FB), $154 \pm 8,40$; e para as amostras com 30% de farinha de trigo (30FT70FB), $156 \pm 8,28$. Esses valores estão relativamente próximos entre si, refletindo uma pequena variação na intensidade do vermelho entre as amostras. Os valores médios dessa variável R (*Red*) mostraram uma leve elevação com o aumento da porcentagem de adulteração: as amostras puras de fubá (0FT100FB) apresentaram média de 151, enquanto as amostras com 20% e 30% de farinha de trigo (20FT80FB e 30FT70FB) apresentaram médias de 154 e 156, respectivamente. Essa tendência sugere que a adição de farinha de trigo, por possuir coloração mais clara, influencia diretamente a intensidade do canal vermelho das imagens, ainda que de forma sutil. Apesar das diferenças numéricas pequenas, essa variável se mostrou relevante para auxiliar na discriminação das classes. No entanto, essa diferença, ainda que sutil, contribuiu significativamente para a discriminação das amostras durante o processo de classificação, especialmente quando combinada com outras variáveis cromáticas.

Além do canal R, as variáveis H, S e V, correspondentes aos parâmetros *Hue* (matiz), *Saturation* (saturação) e *Value* (brilho) do espaço de cor HSV, também desempenharam um papel importante na caracterização das amostras. As componentes do espaço de cor HSV: H (Hue ou matiz), S (Saturation ou saturação) e V (Value ou brilho), contribuíram de maneira significativa para a diferenciação das amostras. A variável H manteve valores médios relativamente constantes entre as classes: 0,125 para 0FT100FB e 0,118 tanto para 20FT80FB quanto para 30FT70FB. Isso indica que a matiz, ou seja, a tonalidade dominante da cor, não foi significativamente alterada com a adulteração, mantendo-se estável entre os diferentes níveis de mistura.

Em contrapartida, a variável S apresentou uma redução progressiva conforme aumentou a porcentagem de farinha de trigo: 0,401 para 0FT100FB, 0,341 para 20FT80FB e 0,333 para 30FT70FB (Figura 4). Essa queda na saturação sugere que a adição de farinha de trigo diluiu a intensidade da cor das amostras, tornando-as visualmente menos vivas, o que pode estar relacionado à aparência mais opaca da farinha em comparação ao fubá puro.

Figura 6. Valores médios das componentes de cor H, S e V nas amostras de fubá com diferentes teores de adulteração.



Fonte: Elaborado pelo autor (2025)

De fato, a variável S foi a característica de cor mais importante para a discriminação das amostras de fubá com diferentes níveis de adulteração.

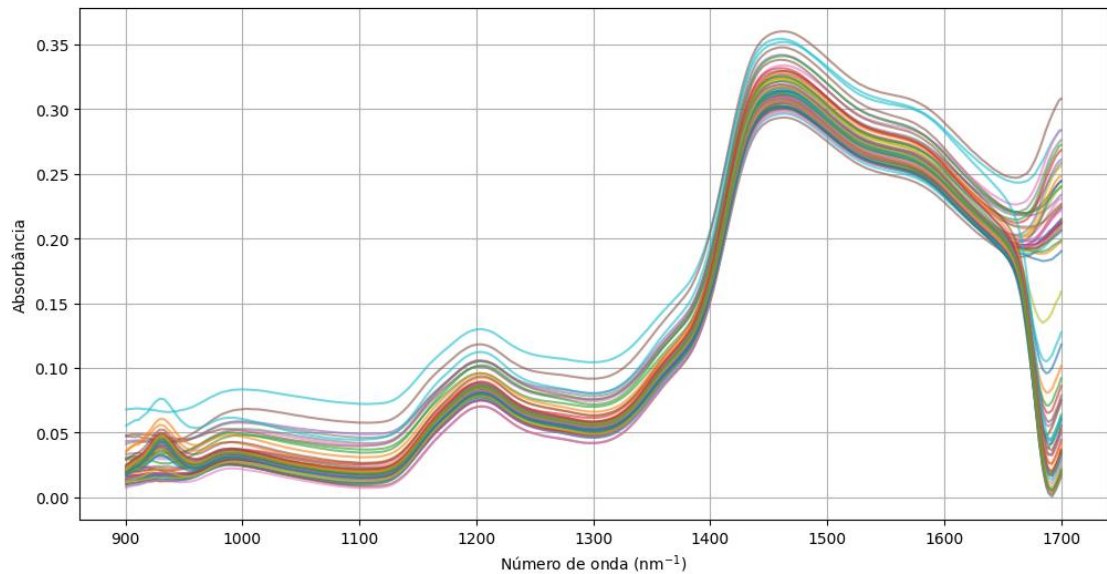
Por fim, a variável V apresentou um aumento gradativo com a adulteração: 0,591 para 0FT100FB, 0,603 para 20FT80FB e 0,612 para 30FT70FB. Esse comportamento é coerente com o fato de a farinha de trigo ser mais clara, o que contribui para o aumento do brilho geral das imagens.

Portanto, embora a variável H tenha se mantido estável, os parâmetros S e V mostraram variações sistemáticas entre as classes, sendo úteis para a distinção entre as amostras. Esses resultados demonstram que as informações cromáticas extraídas das imagens, especialmente as componentes do espaço de cor HSV, podem ser eficazes para identificar e classificar níveis de adulteração em amostras de fubá, mesmo quando as diferenças visuais são pouco perceptíveis a olho nu. Tais resultados reforçam o potencial do uso de visão computacional e aprendizado de máquina na detecção rápida e não destrutiva de fraudes alimentares.

5.2 Análise por espectroscopia NIR

A aplicação de algoritmos de aprendizado de máquina sobre dados espectrais obtidos por espectroscopia (Figura 5) no infravermelho próximo (NIR) demonstrou elevada acurácia na identificação de amostras de fubá com diferentes níveis de adulteração por farinha de trigo.

Figura 7. Espectros de absorbância no infravermelho próximo (NIR) de amostras de fubá com diferentes níveis de adulteração por farinha de trigo, no intervalo de 900 a 1700 nm^{-1} .



Fonte: Elaborado pelo autor (2025)

O espectro exibido na região entre 900 e 1700 nm^{-1} apresentou grupos funcionais, especialmente associados a ligações O-H, C-H e N-H, possivelmente provenientes de água, polissacarídeos, celulose, amido, lipídios e proteínas (Tabela 1).

Tabela 1. Regiões espectrais no NIR e seus respectivos grupos funcionais e modos vibracionais associados.

Região (nm ⁻¹)	Grupo funcional provável	Tipo de vibração associada
~930–960	O–H (água, álcool)	2º sobretom de estiramento O–H
~1150–1250	C–H (alquilas, lipídios)	2º sobretom de estiramento C–H
~1350–1450	O–H (forte, ampla banda)	1º sobretom de estiramento O–H (água, álcool)
~1450–1550	N–H ou C–H	Combinação de deformações + estiramentos
~1600–1700	C–H ou combinação C–H/N–H	Combinações harmônicas diversas (alquilas, proteínas)

Fonte: Elaborado pelo autor (2025)

A região mais intensa (em torno de 1400–1500 nm) provavelmente está associada a sobretom de O–H, indicando presença de água e carboidratos (amido), presentes nas amostras de fubá e fubá adulterado com farinha de trigo. As diferenças entre curvas podem estar relacionadas à variação da composição química, como teores de proteína, amido ou umidade, entre as amostras puras de fubá e adulteradas com farinha de trigo.

A matriz de confusão do modelo otimizado revelou uma elevada taxa de acerto na classificação das amostras. As classes 0FT100FB, 20FT80FB e 30FT70FB foram corretamente identificadas com alta precisão, conforme refletido nos valores de precisão, revocação (proporção de amostras positivas corretamente identificadas pelo modelo) e f1-score superiores a 0,95 para a maioria das classes, no conjunto de teste (Tabela 2).

Tabela 2. Indicadores de precisão, revocação e F1-score para o modelo treinado (conjunto teste)

Classe	Precisão	Revocação	F1-Score	Amostras (suporte)
0FT100FB	1.00	0.95	0.97	20
20FT80FB	0.95	1.00	0.97	19
30FT70FB	1.00	1.00	1.00	29
Acurácia global			0.99	68
Média macro	0.98	0.98	0.98	68
Média ponderada	0.99	0.99	0.99	68

Fonte: Elaborado pelo autor (2025)

Os mesmos indicadores de qualidade do modelo foram calculados para os dados do conjunto treino (Tabela 3). Os resultados mostraram que o modelo capturou de forma eficiente o padrão espectroscópico das amostras, com acurácia global igual a 1,0 (100% de acerto).

Tabela 3. . Indicadores de precisão, revocação e F1-score para o modelo treinado (conjunto treino)

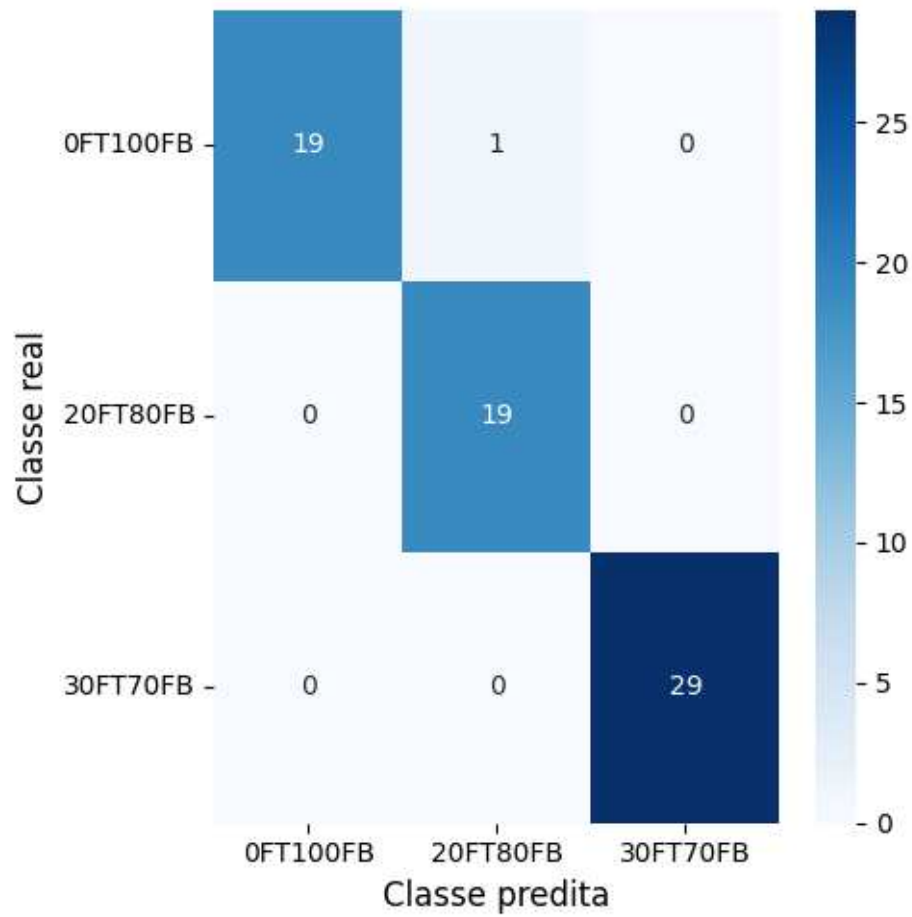
Classe	Precisão	Revocação	F1-Score	Amostras (suporte)
0FT100FB	1.00	1.00	1.00	54
20FT80FB	1.00	1.00	1.00	56
30FT70FB	1.00	1.00	1.00	47
Acurácia global			1.00	157
Média macro	1.00	1.00	1.00	157
Média ponderada	1.00	1.00	1.00	157

Fonte: Elaborado pelo autor (2025)

Além das acurácias obtidas nos conjuntos de treino e teste, também foi calculada a acurácia por validação cruzada (cross-validation), que resultou em 99,11% ($\pm 1,78\%$). O desvio padrão mostra o quanto esses resultados variaram entre os testes, neste caso, uma variação pequena, o que indica que o modelo foi estável. Embora o valor “99,11% $\pm 1,78\%$ ” possa dar a impressão de que a acurácia ultrapassou 100%, isso não acontece na prática. A acurácia máxima possível continua sendo 100%; o que esse número mostra é que o modelo teve ótimo desempenho e resultados muito próximos entre si. As acurácias dos modelos podem ser melhor visualizadas pela matriz de confusão.

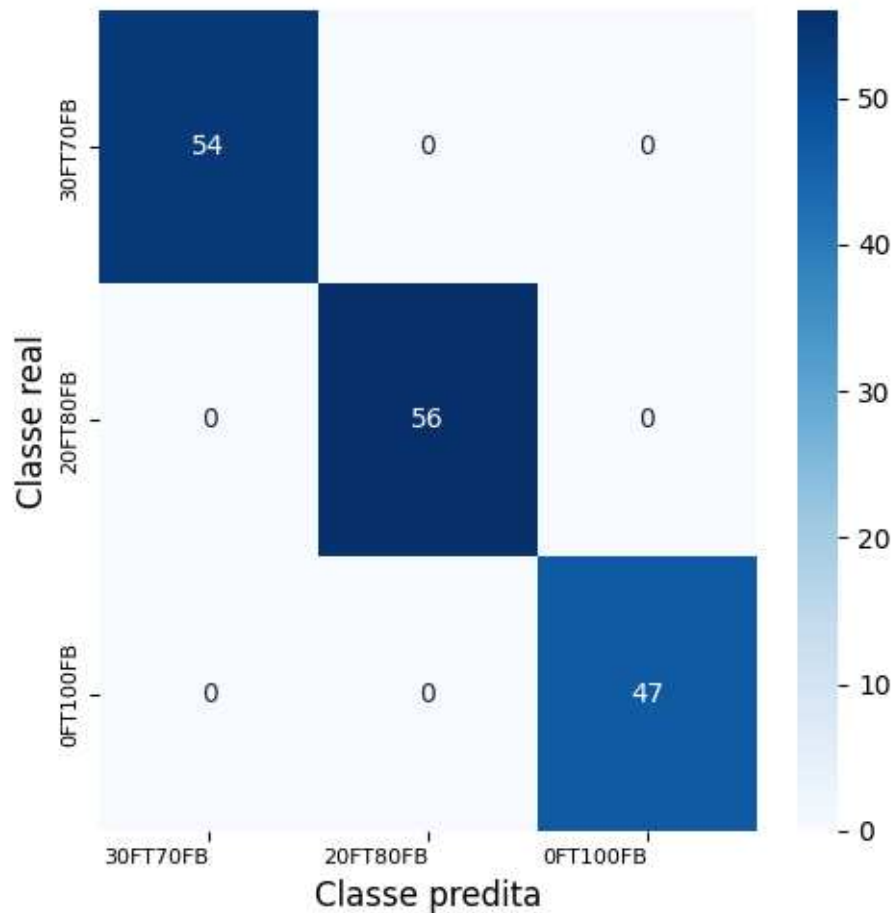
A matriz de confusão é uma ferramenta amplamente utilizada na avaliação de modelos de classificação, pois permite visualizar, de forma clara, o desempenho preditivo em relação às classes reais. Por meio dessa representação, é possível identificar não apenas a acurácia geral do modelo, mas também os acertos e erros específicos para cada categoria. No presente estudo, a matriz de confusão foi utilizada (Figuras 6 e 7) para avaliar a capacidade do algoritmo Random Forest em classificar corretamente as amostras de fubá com diferentes níveis de adulteração por farinha de trigo, com base nos dados espectroscópicos obtidos por NIR.

Figura 8. Matriz de confusão para o conjunto de teste referente à classificação de amostras de fubá com diferentes níveis de adulteração por farinha de trigo, utilizando o algoritmo Random Forest aplicado aos dados espectroscópicos.



Fonte: Elaborado pelo autor (2025)

Figura 9. Matriz de confusão para o conjunto de treino referente à classificação de amostras de fubá com diferentes níveis de adulteração por farinha de trigo, utilizando o algoritmo Random Forest aplicado aos dados espectroscópicos.



Fonte: Elaborado pelo autor (2025)

A análise da matriz de confusão revelou que o modelo Random Forest apresentou excelente desempenho na classificação das três classes de amostras avaliadas (0FT100FB, 20FT80FB e 30FT70FB), com acurácias individuais superiores a 95%. Observou-se que a classe intermediária (20FT80FB) obteve revocação máxima (1,00%), indicando que todas as suas amostras foram corretamente classificadas. As classes 0FT100FB e 30FT70FB também apresentaram resultados muito próximos da perfeição, com alta precisão e F1-score, o que evidencia a eficácia do modelo em distinguir níveis distintos de adulteração, mesmo com espectros visualmente semelhantes.

A ampliação do conjunto de dados (de 75 para 225 amostras) foi uma importante estratégia para melhorar a capacidade de generalização do modelo, reduzindo o risco de sobreajuste e aumentando a robustez das predições. A análise também revelou que, mesmo diante de diferenças visuais sutis nos espectros, o

algoritmo Random Forest foi capaz de identificar padrões complexos nos dados, reforçando o potencial do uso de aprendizado de máquina aliado à espectroscopia NIR como ferramenta analítica na detecção de fraudes em produtos alimentícios

6 CONCLUSÃO

Com base na seção de resultados e discussões deste trabalho, é possível concluir que os objetivos propostos foram em maioria alcançados e que as metodologias aplicadas — tanto de visão computacional baseada em canais de cor RGB/HSV quanto de espectroscopia no infravermelho próximo (NIR) — mostraram-se eficazes na identificação de adulterações em amostras de fubá orgânico com farinha de trigo. As análises indicaram que as alterações nas propriedades ópticas e espectrais dos produtos são detectáveis e classificáveis por meio de algoritmos de aprendizado de máquina, com boa acurácia em diferentes conjuntos de dados.

No que diz respeito à abordagem por visão computacional, o modelo baseado nos atributos cromáticos (R, H, S e V) obteve bons níveis de precisão, especialmente ao identificar o fubá puro. A saturação (S) e o canal vermelho (R) destacaram-se como os atributos mais sensíveis à adulteração, apresentando variações sistemáticas e coerentes conforme o nível de substituição do fubá por farinha de trigo. A leve elevação da intensidade do canal R e a redução da saturação sugerem que a farinha adulterante dilui as características visuais típicas do fubá, tornando possível sua detecção automática por modelos de classificação.

A modelagem dos dados espectroscópicos do NIR revelou resultados ainda mais promissores. Com espectros característicos bem definidos entre as amostras adulteradas e não adulteradas, e grupos funcionais identificáveis nas faixas de 900 a 1700 nm^{-1} , os modelos de aprendizado de máquina aplicados a esses dados alcançaram acurácia superior a 99%, tanto em testes quanto em validação cruzada. Isso demonstra a potência do NIR como ferramenta não destrutiva e altamente sensível na análise de alimentos, especialmente quando combinado a técnicas de inteligência artificial como algoritmo Random Forest.

Além das métricas elevadas de acurácia, os resultados indicaram baixa taxa de confusão entre classes, sobretudo no modelo espectroscópico, que obteve precisão e F1-score máximos em praticamente todas as classes de adulteração. A capacidade de generalização dos modelos foi reforçada pela técnica de expansão amostral e validação cruzada, o que assegura a robustez dos modelos para aplicações práticas, como inspeções em linha de produção ou controle de qualidade laboratorial.

Outro aspecto relevante foi a facilidade de implementação das ferramentas utilizadas. Softwares como ChemoStat e bibliotecas de código aberto em Python permitiram análises complexas com baixo custo e acessibilidade, o que reforça o caráter prático e aplicável do estudo. Esses recursos viabilizam a replicação da metodologia em diferentes contextos da cadeia de alimentos, favorecendo uma resposta mais rápida e eficiente à crescente preocupação com fraudes em produtos orgânicos.

Dessa forma, este trabalho demonstra que a combinação entre espectroscopia NIR, visão computacional e aprendizado de máquina é uma estratégia poderosa para a detecção de adulterações em produtos alimentícios, unindo sensibilidade analítica, automação e baixo custo. Essa abordagem pode ser expandida futuramente para outras matrizes alimentares, contribuindo significativamente para a segurança do consumidor, o fortalecimento do mercado de produtos orgânicos e o combate a práticas fraudulentas que ameaçam a integridade e a confiança no setor de alimentos.

Por fim, recomenda-se o uso complementar das técnicas RGB e NIR para maximizar a eficácia das análises, aproveitando as vantagens de cada abordagem: a agilidade da visão computacional e a profundidade da análise espectral. Também é recomendável que estudos futuros explorem amostras com teores ainda mais variados de adulteração, ampliem a base de dados e testem outros algoritmos, como redes neurais profundas, para aperfeiçoar ainda mais a detecção automática de fraudes em alimentos.

7 REFERÊNCIAS

ANAGAW, Y. Y. *et al.* Food adulteration: causes, risks, and detection techniques. **SAGE Open Medicine**, 11 abr. 2024.

BANNOR, R. K. *et al.* A comprehensive systematic review and bibliometric analysis of food fraud from a global perspective. **Journal of Agriculture and Food Research**, Sunyani, 14, dez. 2023.

BANNOR, R. K. *et al.* *A comprehensive systematic review and bibliometric analysis of food fraud from a global perspective*. Journal of Agriculture and Food Research, v. 14, p. 100686, 2023. Disponível em: <https://doi.org/10.1016/j.jafr.2023.100686>. Acesso em: 12 jun. 2025.

BISHOP, C. M. **Pattern recognition and machine learning**. New York: Springer, 2006.

BO, Y. *et al.* High-resolution pure shift NMR spectroscopy offers better metabolite discrimination in food quality analysis. **Food Research International**, v. 125, nov. 2019.

BURNS, D. A.; CIURCZAK, E. W. **Handbook of near-infrared analysis**. 3^a. ed. [S.I.]: CRC, v. 35.

CÂMARA, A. B. F. *et al.* Multivariate assessment for predicting antioxidant activity from clove and pomegranate extracts by MCR-ALS and PLS models combined to IR spectroscopy. **Food Chemistry**, v. 384, p. 132321, 2022. Disponível em: <https://doi.org/10.1016/j.foodchem.2022.132321>. Acesso em: 13 jun. 2025.

CAMPOS, V. D.; CASSOL, K. P.; WIZNIEWSKY, R. F. A sustentabilidade nos territórios do milho crioulo: Olhares para Anchieta/SC e Ibarama/RS. **Revista Grifos**, 44, 27 jul. 2018., p. 145-172

CAPELLESSOI, A. J.; CAZELLAI, A. Indicador de sustentabilidade dos agroecossistemas: estudo de caso em áreas de cultivo de milho. **Ciência Rural**, Santa Maria, 43, dez 2013., p. 2297-2303

CAPOBIANGO, M. *et al.* Extração química e enzimática das proteínas do fubá de milho. **Food Science and Technology**, v. 26, n. 4, p. 884–890, out. 2006.

CARDOSO, W. S. *et al.* Indústria do milho. [S.I.]: [s.n.], 2012., p. 175-195.

CARVALHO, D. V. **Defesa alimentar e fraude alimentar no âmbito dos sistemas de gestão da segurança alimentar: linhas de orientação**. 2021. Dissertação (Mestrado em Biotecnologia e Inovação) – Universidade Católica Portuguesa, Lisboa, 2021.

CHEN, *et al.* A review of multivariate analyses in imaging genetics. Journal, ano. Disponível em: PMC3972473. Acesso em: 12 jun. 2025.

CHMELLKOVA, L. *et al.* Energy-use efficiency of organic and conventional plant production systems in Germany. **Scientific Reports**, v. 14, n. 1806, 2024. Disponível em: <https://www.nature.com/articles/s41598-024-51768-3>. Acesso em: 12 jun. 2025

CHOLLET F. Deep learning with python. 2 ed. New York: Manning Publications Co. 2021. 477 p. Disponível em: <https://sourestdeeds.github.io/pdf/Deep%20Learning%20with%20Python.pdf>. Acesso em: 06 Jun. 2024.

Como é produzida a farinha de Milho? **Grupo Idugel**, 15 fev. 2023. Disponível em: <https://idugel.com.br/como-e-produzida-a-farinha-de-milho/>. Acesso em: 10 jun. 2024.

CRUZ, C. Panorama do consumo de orgânicos no Brasil. **Organis**, 2023. Disponível em: <https://ciorganicos.com.br/wp-content/uploads/2017/10/ciorganicos-20230616-141015.pdf>. Acesso em: 10 jun. 2024.

DARRA, E. *et al.* Food fraud detection in commercial pomegranate molasses syrups by UVeVIS spectroscopy, ATR-FTIR spectroscopy and HPLC methods. **Food Control**, Beirut, 78, 01 mar. 2017., p. 132-137

DAVIS, H. **Visual Basic.NET for Windows**. ilustrada. ed. [S.l.]: Peachpit Press, 2003.

DELIBERADOR, L. R. *et al.* The influence of organic food purchase intention on household food waste: insights from Brazil. *Sustainability*, v. 16, n. 9, p. 3795, 2024. Disponível em: <https://doi.org/10.3390/su16093795>. Acesso em: 12 jun. 2025

ESTEKI, M.; REGUEIRO, M.; GANDARA, S. Tackling Fraudsters with Global Strategies to Expose Fraud in the Food Chain. **Comprehensive Reviews in Food Science and Food Safety**, 2, mar. 2019., p. 425-440

EVERSTINE, K.; SPINK, J.; KENNEDY, S. Economically motivated adulteration (EMA) of food: common characteristics of EMA incidents. **J Food Prot**, v. 76. n. 4, p. 723-35, abr. 2013.

FERGUSON, B. *et al.* Infrared micro-spectroscopy coupled with multivariate and machine learning techniques for cancer classification in tissue. **Analyst**, v. 147, n. 21, p. 4177–4192, 2022. DOI: 10.1039/d2an00620a. acesso em: 13 jun. 2025.

GONZALEZ, R. C.; WOODS, R. E. **Digital Image Processing**. 4^a. ed. [S.l.]: Pearson, 2018. 206 p.

HAIR JR, F. *et al.* **MULTIVARIATE DATA ANALYSIS**. 8^a. ed. [S.l.]: Cengage, 2019. 813 p.

HARKALY, A. Alta demanda dos alimentos orgânicos pode influenciar o aumento de fraudes. **Portal e-food**, 2023. Disponível em: <https://portalefood.com.br/gestao/alta-demanda-dos-alimentos-organicos-pode-influenciar-o-aumento-de-fraudes/>. Acesso em: 05 jun. 2024.

HELPER, G. A. *et al.* ChemosStat, um software gratuito para análise exploratória de dados multivariados. **Química Nova**, v. 38, n. 4, p. 575–579, maio 2015.

HENSON, S.; REARDON, T. Private agri-food standards: implications for food policy and the agri-food system. **Food Policy**, 30, 2005., p. 241-253

HU, T. et al. Examining consumers' willingness to pay premium price for organic food in China. *Humanities and Social Sciences Communications*, v. 11, art. 27, 2024. Disponível em: <https://doi.org/10.1017/s1742170522000035>. Acesso em: 12 jun. 2025

Huo, H. et al. Consumers' willingness to pay for dairy and plant-based milk alternatives: environmental and quality attributes matter. *Frontiers in Psychology*, 2023. Disponível em: <https://doi.org/10.1080/10454446.2023.2185118>. Acesso em: 12 jun. 2025

JAIN, K. A. **Fundamentals of digital image processing**. New Jersey: Prentice-Hall, 1989.

JOHANN, L.; DALMORO, M; MACIEL, J. M. **Alimentos orgânicos: dinâmica na produção e comercialização**. 1. ed. Lajeado: Univates, 2019. 191 p.

JOHNSON, R. Food Fraud and "Economically Motivated Adulteration" of Food and Food Ingredients. **Congressional Research Service**, 10 maio 2014., p. 1-40

JOHNSON, R.; WICHERN, D. W. **Applied multivariate statistical analysis**. 6^a. ed. Nova Jersey: Pearson Prentice Hall, 2007.

KAMIYAMA, A. **Produto orgânico: Vamos falar sobre comercialização?** Rio de Janeiro: Sociedade Nacional de Agricultura, 2017.

KHARE, V; DWIVEDI, S. K.; BHATIA, M. Chapter 4 - Multivariate data analytics and cognitive analytics. In: KHARE, V.; DWIVEDI, S. K.; BHATIA, M. **Cognitive Science, Computational Intelligence, and Data Analytics**. [S.I.]: Morgan Kaufmann, 2024., p. 175-219.

KUEPPER, G. ORGANIC FIELD CORN PRODUCTION. **NCAT Agriculture Specialist**, Jan. 2002., p. 01-16

LAS BRUMAS. Participatory Guarantee Systems, a more inclusive organic certification alternative? *Frontiers in Sustainable Food Systems*, v. 7, 2023. Disponível em: <https://www.frontiersin.org/articles/10.3389/fsufs.2023.1178394/full>. Acesso em: 12 jun. 2025.

LAS BRUMAS. Participatory Guarantee Systems, a more inclusive organic certification alternative? *Frontiers in Sustainable Food Systems*, v. 7, 2023. Disponível em: <https://www.frontiersin.org/articles/10.3389/fsufs.2023.1178394/full>. Acesso em: 12 jun. 2025.

LI, Y. *et al.* The application of machine learning to air pollution research: A bibliometric analysis. *Ecotoxicology and Environmental Safety*, Beijing, 257, abr. 2023., p. 114-911

LIMA, N. W. F.; LEAL, P. J. B.; ALMEIDA, A. R. C. Fraude nos alimentos e suas implicações para a saúde humana: uma revisão de literatura. *Scientific Reports*, Mossoró, v. 1, n. 1, p. 1-18, 2024. Disponível em: [URL do artigo, se disponível]. Acesso em: 12 jun. 2025

LIPOVETSKY, S. Multivariate statistical methods: A brief review on their modifications and applications. **Multivariate Analysis and Statistics**, v. 15, 2022. Disponível em: <https://doi.org/10.3233/MAS-220017>. Acesso em: 13 jun. 2025.

MAESTRESALASA, A. Ó. *et al.* Detection of minced lamb and beef fraud using NIR spectroscopy. **Food Control**, Arrosadia, 98, 2019., p. 465-473

MITCHELL, M. **Machine learning**. Portland: McGraw-Hill Science, 1997.

MOORE, J. C.; SPINK, J.; LIPP, M. Ingredient Fraud and Economically Motivated Adulteration from 1980 to 2010. **Journal of Food Science**, 77, 01 abr. 2012., p. 117-126

OFFERMANN, F.; NIEBERG, H. **Economic Performance of Organic Farms in Europe**. Ettlingen: 1. ed. 2000. 198 p.

PHUONG T. N. **Comparison between two programming languages R and Python**. 2020. Tese (Bacharelado em matemática, ciência da computação e estatística) – Ludwig maximilians university minich, Munique, 2020.

PRADO, S. P. T.; RODRIGUES, M. L.; AQUINO, C. I.; OKADA, I. A.; IHA, M. H. Verificação de adulterações em cúrcuma, gengibre, noz-moscada, páprica, pimenta-do-reino e colorífico, comercializados no estado de São Paulo, Brasil. (Parte II). **Revista do Instituto Adolfo Lutz**, São Paulo, v. 80, p. 1-10, 2021. Disponível em: <https://doi.org/10.53393/rial.2021.v.80.37287>. Acesso em: 12 jun. 2025.

PUNJABI, N. *et al.* Infrared spectrum analysis of organic molecules with neural networks using standard reference data sets in combination with real-world data. **Journal of Cheminformatics**, v. 17, 2025. DOI: 10.1186/s13321-025-00728-1. Acesso em: 13 jun. 2025.

QU, Q.; JIN, L. Application of nuclear magnetic resonance in food analysis. **Food Science and Technology**, v. 42, 2022.

RACHINENI, K. *et al.* Facile and rapid detection of adulteration in mustard oils: NMR and unsupervised machine learning. **Food Control**, Mumbai, 150, ago. 2023.

RANDOLPH, N. *et al.* **Professional visual studio 2010**. [S.I.]: Wiley Publishing, 2010.

RASTOGI, S. *et al.* RGB colorimetric method-based detection of oxytocin in food samples using cysteamine functionalized gold nanoparticles. **Analytical Biochemistry**, New Delhi, 656, 2022., p. 114-886

REGANOLD, J. P.; WACHTER, J. Organic agriculture in the twenty-first century. **Nature Plants**, v. 2, fev. 2016., p. 15-221

REID, L. M.; O'DONNELL, C. P.; DOWNEY, G. Recent technological advances for the determination of food authenticity. **Trends in Food Science & Technology**, Dublin, 17, 2006., p. 344-353

RENCHER, A. C. **Methods of multivariate analysis**. 2^a. ed. [S.I.]: Wileyintersciece, 2002.

REZENDE, A. L. et al. Aplicação de técnicas multivariadas e IA na análise de espectros de infravermelho para determinação de matéria orgânica em solos. **Revista Brasileira de Ciência do Solo**, v. 47, 2023. Disponível em: <https://www.scielo.br/j/rbcs/a/xyz>. Acesso em: 12 jun. 2025.

RIBEIRO, J. E. Meta-analysis of consumer willingness to pay for organic food. **Research Square**, Canterbury, 21 jul. 2023., p. 33

SACCHI, G.; ROMANELLO, M. L.; CANAVARI, M. *The future of organic certification: potential impacts of the inclusion of Participatory Guarantee Systems in the European organic regulation*. *Agricultural and Food Economics*, v. 12, n. 1, p. 1-16, 2024. Disponível em: <https://link.springer.com/article/10.1186/s40100-024-00263-2>. Acesso em: 12 jun. 2025

SACCHI, G.; ROMANELLO, M. L.; CANAVARI, M. The future of organic certification: potential impacts of the inclusion of Participatory Guarantee Systems in the European organic regulation. **Agricultural and Food Economics**, v. 12, n. 1, p. 1-16, 2024. Disponível em: <https://link.springer.com/article/10.1186/s40100-024-00263-2>. Acesso em: 12 jun. 2025.

SÁEZ-HERNÁNDEZ, R. *et al.* A fast and non-invasive imaging procedure to fight red tuna fraud. **LWT - Food Science and Technology**, Valencia, 186, 2023., p. 115-231

SANTOS, N. C. B. D.; TIVELLI, S. W. **Milho orgânico técnicas de cultivo**. Rio de Janeiro: Sociedade Nacional de Agricultura, 2017.

SCURI, A. E.; **Fundamentos da Imagem Digital**. Rio do Janeiro: Tecgraf/PUC-Rio, 2002. 95 p.

SHORTEN, Connor; KHOSHGOFTAAR, Taghi M. **A survey on Image Data Augmentation for Deep Learning**. *Journal of Big Data*, [S. l.], v. 6, n. 60, p. 1–48, 2019. Disponível em: <https://doi.org/10.1186/s40537-019-0197-0>. Acesso em: 30 jun. 2025.

SHRESTHA, N. Detecting multicollinearity in regression analysis. **American Journal of Applied Mathematics and Statistics**, v. 8, n. 2, p. 39–42, 2020. Disponível em: <https://doi.org/10.12691/ajams-8-2-1>. Acesso em: 30 jun. 2025.

SIDDIQUE, A. B. *et al.* Prediction of magnetic nature of oxide compositions by using machine learning models. **Computational Condensed Matter**, 40, mar. 2024.

SILVA, A. F. S. Metodologias analíticas alternativas para a avaliação de autenticidade de alimentos explorando análises hiperespectrais e fotometria por imagens digitais. 2021. Tese (Doutorado em Química na Agricultura e no Ambiente) – Centro de Energia Nuclear na Agricultura, Universidade de São Paulo, Piracicaba, 2021. Disponível em: 10.11606/T.64.2021.tde-05102022-153348. Acesso em: 12 jun. 2024.

Siva Kumar, M., Parthasarathy, S., & Padmapriya, S. T. (2024). **Trade-off between training and testing ratio in machine learning for medical image processing**. Manuscript submitted for publication.

SKEET, J. **C# in depth**. 4^a. ed. [S.l.]: Hanning, 2019.

SOUZA, J. L.; CASALI, V. W. D.; SANTOS, R. H. S.; CECON, P. R. Balanço e análise da sustentabilidade energética na produção orgânica de hortaliças. **Horticultura Brasileira**, v. 26, n. 4, p. 433-440, out.-dez. 2008.

SPINK, J; MOYER, D. C. Defining the Public Health Threat of Food Fraud. **Journal of Food Science**, 76, 09 nov. 2011., p. 157-163

SUN, J.; TANG, M. The programming languages: introduction of R. **Systems Medicine**, v. 01, p. 1-8, 2021.

TARAPOULOUZI, M. et al. Chemometric Discrimination of the Geographical Origin of Three Greek Cultivars of Olive Oils by Stable Isotope Ratio Analysis. **Foods**, v. 10, n. 2, p. 336, 2021. Disponível em: <https://doi.org/10.3390/foods10020336>. Acesso em: 13 jun. 2025.

TATA, *et al.* Empowering veterinary clinical diagnosis in industrial poultry production by ambient mass spectrometry and chemometrics: a new approach for precise poultry farming. **Poultry Science**, v. 103, n. 6^a, jun. 2024.

TAYLOR, M. Mitigating the Risks of Food Fraud in an Inflationary Environment. **food safety tech**, 13 mar. 2023. Disponível em: https://foodsafetytech.com/feature_article/mitigating-the-risks-of-food-fraud-in-an-inflationary-environment/. Acesso em: 01 jun. 2024.

TREWAVAS. A. **Urban myths of organic farming: organic agriculture as an ideology, but can it meet today's needs?** Nature. Vol 410. p. 409-410. 22 mar 2001.

UDAYANGA, Darsha; SERASINGHE, Ashan; DASSANAYAKE, Supun; GODALIYADDA, Roshan; HERATH, H. M. V. R.; EKANAYAKE, M. P. B.; MALSHAN, H. L. P. Dual mode multispectral imaging system for food and agricultural product quality estimation. *arXiv*, 4 out. 2023. Disponível em: <https://arxiv.org/pdf/2310.03110>. Acesso em: 12 jun. 2025.

VARMUZA, K.; FILZMOSER, P. **Introduction to Multivariate Statistical Analysis in Chemometrics**. 1. ed. [S.l.]: [s.n.], 2009.

VITORINO, R. *et al.* Diagnostic and monitoring applications using near infrared (NIR) spectroscopy in cancer and other diseases. **Photodiagnosis and Photodynamic Therapy**, jul. 2023., p. 103-633

VOCCIO, R. *et al.* Combining PLS-DA and SIMCA on NIR data for classifying raw materials for tyre industry: A hierarchical classification model. **Chemometrics and Intelligent Laboratory Systems**, v. 250, 2024.

WATANABE, E. A. M. et al. The Consumption of Fresh Organic Food: Premium Pricing and the Predictors of Willingness to Pay. **Journal of Food Products Marketing**, v. 29, n. 2-3, p. 41-55, 2023. Disponível em: <https://doi.org/10.1080/10454446.2023.2185118>. Acesso em: 12 jun. 2025.

ZHONG, K. *et al.* A colorimetric and NIR fluorescent probe for ultrafast detecting bisulfite and organic amines and its applications in food, imaging, and monitoring fish freshness. **Food Chemistry**, v. 438, abr. 2024.