



UNIVERSIDADE
ESTADUAL DE LONDRINA

VINICIUS AVANZI BARBOSA MASCARELI

**DA CIÊNCIA DE ALIMENTOS À CIÊNCIA DE DADOS
AUTOMATIZADA: ABORDAGENS ANALÍTICAS
MODERNAS PARA QUALIDADE E CLASSIFICAÇÃO DE
VINAGRES**

VINICIUS AVANZI BARBOSA MASCARELI

**DA CIÊNCIA DE ALIMENTOS À CIÊNCIA DE DADOS
AUTOMATIZADA: ABORDAGENS ANALÍTICAS
MODERNAS PARA QUALIDADE E CLASSIFICAÇÃO DE
VINAGRES**

Tese apresentada ao Programa de Pós-graduação em
Ciência de Alimentos da Universidade Estadual de
Londrina - UEL, como requisito parcial para a
obtenção do título de Doutor.

Orientador: Prof^ª. Wilma Aparecida Spinosa

Coorientador: Prof. Diego Galvan

Londrina
2026

Ficha de identificação da obra elaborada pelo autor, através do Programa de Geração Automática do Sistema de Bibliotecas da UEL

Mascareli, VINICIUS AVANZI BARBOSA MASCARELI.

DA CIÊNCIA DE ALIMENTOS À CIÊNCIA DE DADOS AUTOMATIZADA:
ABORDAGENS ANALÍTICAS MODERNAS PARA QUALIDADE E
CLASSIFICAÇÃO DE VINAGRES / VINICIUS AVANZI BARBOSA
MASCARELI Mascareli. - Londrina, 2026.
176 f.

Orientador: WILMA APARECIDA SPINOSA.

Coorientador: DIEGO GALVAN.

Tese (Doutorado em Ciência de Alimentos) - Universidade Estadual de Londrina, Centro de Ciências Agrárias, Programa de Pós-Graduação em Ciência de Alimentos, 2026.

Inclui bibliografia.

1. Autenticação de alimentos - Tese. 2. Fingerprinting - Tese. 3. Ressonância magnética nuclear de baixo campo - Tese. 4. Machine learning - Tese. I. APARECIDA SPINOSA, WILMA. II. GALVAN, DIEGO. III. Universidade Estadual de Londrina. Centro de Ciências Agrárias. Programa de Pós-Graduação em Ciência de Alimentos. IV. Título.

CDU 641.1

VINICIUS AVANZI BARBOSA MASCARELI

**DA CIÊNCIA DE ALIMENTOS À CIÊNCIA DE DADOS AUTOMATIZADA:
ABORDAGENS ANALÍTICAS MODERNAS PARA QUALIDADE E
CLASSIFICAÇÃO DE VINAGRES**

Tese apresentada ao Programa de Pós-graduação
em Ciência de Alimentos da Universidade
Estadual de Londrina - UEL, como requisito
parcial para a obtenção do título de Doutor.

BANCA EXAMINADORA

Prof^ª. Dr^ª. Wilma Aparecida Spinosa
Universidade Estadual de Londrina - UEL

Prof^ª. Dr^ª. Claudimara da Silva Portela
Universidade Estadual de Londrina - UEL

Prof^ª. Dr^ª. Karina Benassi Angilelli
Universidade Estadual de Londrina - UEL

Prof^ª. Dr^ª. Érica Signori Romagnoli
Universidade Estadual de Londrina - UEL

Dr^ª. Jaqueline Munise Guimarães da Silva
Universidade Estadual de Londrina - UEL

Londrina, 21 de agosto de 2026.

AGRADECIMENTOS

A Prof^ª. Wilma Spinosa, minha orientadora e querida amiga, que acompanhou e me apoiou na realização deste trabalho e da minha carreira. Obrigado por todos os momentos e por acreditar em mim.

Ao Prof. Diego Galvan, meu coorientador e querido amigo pelo apoio e todo o suporte excepcional dado para a realização desta obra.

Aos membros da banca avaliadora, Claudimara S. Portela, Érica S. Romagnoli, Karina B. Angilelli, José R. Silva, Jaqueline M. G. da Silva e Ana Clara L. Pavanello.

Aos professores e educadores que contribuíram com todo o conhecimento durante a minha vida. Vocês sempre foram e serão a minha esperança de um mundo melhor. Em especial as docentes: Wilma A. Spinosa, Sandra H. Prudencio, Karina B. Angilelli, Marta T. Benassi, Hágata C. da Silva e Roberta A. Medeiros que inspiraram minha jornada.

A todos que contribuíram com o desenvolvimento do trabalho, em especial, Jelmir, Prof. Sylvio, Ingrid e Viviane.

Ao laboratório de Análise de Alimentos da UEL e as minhas amigas: Viviane, Eline, Rebeca, Claudimara, Ana Clara, Jaquellyne e Thayná.

As matriarcas: minha Mãe Rosane, Neiva, Dona Germana e Patrícia. As minhas Sisters: Bibi, Bicuda, Cassi, Esther, Flávia, Isabella, Matheus e Thay. A todos os meus amigos por apoiarem a minha jornada, vocês são minha família.

E por fim, agradeço toda a espiritualidade que sempre cuidou de mim!

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

*If you feel insignificant, you better think again
Better wake up because you're part of something way bigger
You're part of something way bigger
Not just a speck in the universe
Not just some words in a bible verse
You are the living word
Ahh, you're part of something way bigger
Bigger than you, bigger than we
Bigger than the picture they framed us to see
But now we see it
And it ain't no secret, no*

Beyoncé

MASCARELI, Vinicius Avanzi Barbosa. **Da ciência de alimentos à ciência de dados automatizada: abordagens analíticas modernas para qualidade e classificação de vinagres**. 2026. 176 f. Tese (Doutorado em Ciência de Alimentos) – Universidade Estadual de Londrina, Londrina, 2026.

RESUMO GERAL

A autenticação de alimentos representa um desafio crescente para a ciência de alimentos em decorrência da complexidade das matrizes alimentícias, da globalização das cadeias produtivas e da ocorrência de fraudes economicamente motivadas. Nesse contexto, abordagens baseadas em *fingerprinting* associadas à quimiometria e ao aprendizado de máquina têm se consolidado como alternativas aos métodos analíticos convencionais por permitirem a exploração simultânea de um grande conjunto de informações químicas. Entre essas técnicas, a espectroscopia de Ressonância Magnética Nuclear de Hidrogênio de Baixo Campo (^1H LF-NMR) destaca-se pela rapidez de análise, reduzido preparo de amostras e baixo custo operacional, embora sua menor resolução espectral imponha desafios relacionados à modelagem e à interpretação dos dados. Assim, foram desenvolvidas estratégias analíticas para classificação, autenticação e interpretação química de vinagres de maçã brasileiros utilizando *fingerprints* espectrais obtidos por ^1H LF-NMR, associados à Análise Discriminante por Mínimos Quadrados Parciais (PLS-DA) e à Inteligência Artificial Explicável (XAI). Foram analisadas 564 amostras, compreendendo vinagres comerciais autênticos e amostras adulteradas experimentalmente em diferentes cenários de fraude. Após alinhamento espectral, correção de linha de base e padronização dos dados, os modelos PLS-DA apresentaram desempenho consistente, alcançando aproximadamente 81% de acurácia, com macro-F1 de 0,79 para a discriminação entre vinagres orgânicos autênticos e adulterados e 0,80 para a diferenciação entre vinagres de maçã autênticos e adulterados com vinagre de álcool. Para ampliar a interpretabilidade dos modelos, foi aplicado o *Spectral Model eXplainer* (SMX), um framework de Inteligência Artificial Explicável capaz de identificar e quantificar a contribuição de diferentes regiões espectrais para o processo de classificação. As análises de fidelidade apresentaram área sob a curva (AUC) normalizada de 0,3723 para o cenário de autenticação de vinagres orgânicos e 0,1388 para a diferenciação entre vinagres de maçã autênticos e adulterados com vinagre de álcool, demonstrando elevada concordância entre as explicações produzidas e o comportamento dos classificadores. Em conjunto, os resultados demonstraram que a integração entre ^1H LF-NMR, quimiometria, aprendizado de máquina e Inteligência Artificial Explicável constitui uma estratégia analítica robusta para autenticação de alimentos, combinando elevado desempenho de classificação e interpretação quimicamente fundamentada dos modelos preditivos. A estrutura metodológica desenvolvida apresenta potencial para aplicação em outras matrizes alimentícias e diferentes técnicas de *fingerprinting*, contribuindo para o avanço de abordagens orientadas por dados na autenticação, rastreabilidade e controle de qualidade de alimentos.

Palavras-chave: Autenticação de alimentos; *Fingerprinting*; Ressonância magnética nuclear de baixo campo; *Machine learning*; Inteligência artificial explicável.

MASCARELI, Vinicius Avanzi Barbosa. **From Food Science to Automated Data Science: Modern Analytical Approaches for Vinegar Quality and Classification**. 2026. 176 p. Thesis (Doctorate in Food Science) – Universidade Estadual de Londrina, Londrina, 2026.

GENERAL ABSTRACT

Food authentication has become an increasingly important challenge in food science due to the complexity of food matrices, the globalization of food supply chains, and the growing incidence of economically motivated food fraud. In this context, fingerprinting-based approaches combined with chemometrics and machine learning have emerged as promising alternatives to conventional analytical methods by enabling the simultaneous exploitation of large amounts of chemical information. Among these techniques, low-field proton nuclear magnetic resonance (^1H LF-NMR) spectroscopy stands out because of its rapid analysis, minimal sample preparation, and low operational cost, although its lower spectral resolution poses challenges for data modeling and interpretation. Accordingly, analytical strategies were developed for the classification, authentication, and chemical interpretation of Brazilian apple vinegars using spectral fingerprints obtained by ^1H LF-NMR combined with Partial Least Squares Discriminant Analysis (PLS-DA) and eXplainable Artificial Intelligence (XAI). A total of 564 samples, comprising authentic commercial vinegars and experimentally adulterated samples representing different fraud scenarios, were analyzed. Following spectral alignment, baseline correction, and data standardization, the PLS-DA models achieved robust predictive performance, reaching approximately 81% accuracy, with macro-F1 scores of 0.79 for discriminating authentic and adulterated organic apple vinegars and 0.80 for differentiating authentic apple vinegars from samples adulterated with spirit vinegar. To improve model interpretability, the Spectral Model eXplainer (SMX) was developed as an Explainable Artificial Intelligence framework capable of identifying and quantifying the contribution of distinct spectral regions to the classification process. Fidelity analysis yielded normalized area under the curve (AUC) values of 0.3723 for the authentication of organic apple vinegars and 0.1388 for the discrimination of authentic and spirit vinegar-adulterated apple vinegars, demonstrating strong agreement between the generated explanations and the behavior of the predictive models. Overall, the results demonstrated that the integration of ^1H LF-NMR, chemometrics, machine learning, and Explainable Artificial Intelligence constitutes a robust analytical strategy for food authentication by combining high predictive performance with chemically meaningful interpretation of classification models. The methodological framework developed in this study has the potential to be extended to other food matrices and fingerprinting techniques, contributing to the advancement of data-driven approaches for food authentication, traceability, and quality control.

Key-words: Food authentication; Fingerprinting; Low-field nuclear magnetic resonance; Machine learning; Explainable artificial intelligence.

LISTA DE FIGURAS

CAPÍTULO II - DETECÇÃO DE FRAUDES EM VINAGRE DE MAÇÃ BRASILEIRO UTILIZANDO FINGERPRINTING POR RMN DE BAIXO CAMPO E MACHINE LEARNING

- Figura 1** – Visão esquemática do conjunto de amostras de vinagre, incluindo amostras originais e adulterações simuladas com vinagre de maçã tradicional e vinagre de álcool57
- Figura 2** – Regiões selecionadas dos espectros médios unidimensionais de LF-NMR para as classes de amostras MA (vinagre de maçã adulterado com 5, 15 ou 25% de vinagre de álcool), AM (amostras de vinagre de maçã orgânico) e AD (vinagre de maçã orgânico adulterado com 5, 6, 30 ou 31% de vinagre de maçã tradicional). As regiões espectrais destacadas correspondem a: A, 1,8–0,8 ppm, associada a álcoois, polióis, ácidos orgânicos e aminoácidos; B, 3,0–1,8 ppm, principalmente relacionada a ácidos orgânicos; e C, 4,0–3,0 ppm, predominantemente associada a carboidratos63
- Figura 3** – PCA – EXP1: Vinagre de maçã orgânico adulterado com vinagre de maçã tradicional (AM vs AD) – três quadrantes superiores. EXP2: Vinagre de maçã adulterado com vinagre de álcool (MA vs AM) – três quadrantes inferiores.....71
- Figura 4** – Matriz de confusão do Experimento 1 (AD vs AM).....75
- Figura 5** – Matriz de confusão do Experimento 1 (AD vs AM).....77

CAPÍTULO III - SPECTRAL MODEL EXPLAINER (SMX): INTELIGÊNCIA ARTIFICIAL EXPLICÁVEL NA CLASSIFICAÇÃO DE VINAGRES BRASILEIROS DE MAÇÃ POR RMN DE ¹H DE BAIXO CAMPO

Figura 1 – Intervalo espectral selecionado dos espectros médios de RMN de ^1H em baixo campo (LF-NMR) das classes MA (vinagre de maçã adulterado com 5, 15 ou 25% de vinagre de álcool), AM (vinagre de maçã orgânico) e AD (vinagre de maçã orgânico adulterado com 5, 6, 30 ou 31% de vinagre de maçã convencional).....131

Figura 2 – Mapeamento da relevância dos domínios espectrais sobre os espectros alinhados de ^1H LF-NMR, gerado pelo *Spectral Model eXplainer*. Painel superior: discriminação entre vinagre de maçã orgânico (AM) e vinagre de maçã orgânico adulterado (AD). Painel inferior: discriminação entre vinagre de maçã (AM) e vinagre de maçã adulterado com vinagre de álcool (MA). Os nove domínios espectrais enriquecidos em metabólitos são destacados por diferentes tonalidades, cuja intensidade é proporcional ao valor da métrica LRC, refletindo a contribuição relativa de cada domínio para as decisões de classificação.137

Figura 3 – Ranqueamento dos nove domínios espectrais enriquecidos em metabólitos com base na métrica LRC, gerado pelo SMX. Painel superior: discriminação entre vinagre de maçã orgânico (AM) e vinagre de maçã orgânico adulterado (AD). Painel inferior: discriminação entre vinagre de maçã (AM) e vinagre de maçã adulterado com vinagre de álcool (MA). Os valores correspondem à contribuição relativa de cada domínio espectral para o comportamento do modelo, expressa como porcentagem da pontuação total de LRC.....140

Figura 4 – Avaliação da fidelidade das explicações geradas pelo *Spectral Model eXplainer* por meio do mascaramento progressivo de domínios espectrais ranqueados. Painel superior: modelo de autenticação AD vs. AM; Painel inferior: modelo de autenticação AM vs. MA. Escores de fidelidade calculados a partir dos deslocamentos acumulados nas predições do modelo produzidos pela perturbação sequencial dos domínios espectrais ordenados de acordo com o ranqueamento de relevância estabelecido pelo SMX.149

Figura 5 – Comparação entre os ranqueamentos de perturbação orientados pelo *Spectral Model eXplainer* e pela Importância da Variável na Projeção para

os modelos de autenticação. Painel superior: modelo de autenticação AD vs.
AM; Painel inferior: modelo de autenticação AM vs. MA.154

LISTA DE TABELAS

CAPÍTULO II - DETECÇÃO DE FRAUDES EM VINAGRE DE MAÇÃ BRASILEIRO UTILIZANDO FINGERPRINTING POR RMN DE BAIXO CAMPO E MACHINE LEARNING

Tabela 1 – Comparação das estratégias de pré-processamento para o Experimento 1 (AD vs AM, sub-banda de 0–4,5 ppm, conjunto de dados limpo) utilizando validação cruzada aninhada com agrupamento (*loop* externo de 10 *folds*; *loop* interno de 5 *folds*) 72

Tabela 2 – Comparação de pré-processamento para MA vs AM (validação cruzada aninhada com agrupamento, *loop* externo de 10 *folds*; *loop* interno de 5 *folds*) 73

Tabela 3 – Melhor desempenho obtido para ambos os experimentos (validação cruzada estratificada de 10 *folds*, pré-processamento com alinhamento) 74

CAPÍTULO III - SPECTRAL MODEL EXPLAINER (SMX): INTELIGÊNCIA ARTIFICIAL EXPLICÁVEL NA CLASSIFICAÇÃO DE VINAGRES BRASILEIROS DE MAÇÃ POR RMN DE ¹H DE BAIXO CAMPO

Tabela 1 – Composição do conjunto de dados de vinagres utilizado nas análises de autenticação e explicabilidade 122

Tabela 2 – Definição dos supostos domínios espectrais de metabólitos empregados no framework SMX e sua correspondente interpretação química recorrentes na literatura 132

Tabela 3 – Desempenho preditivo dos modelos de PLS-DA empregados nas análises de explicabilidade 134

SUMÁRIO

CAPÍTULO I

1	INTRODUÇÃO GERAL	16
2	OBJETIVOS	18
2.1	OBJETIVO GERAL	18
2.2	OBJETIVOS ESPECÍFICOS	18
3	REVISÃO BIBLIOGRÁFICA	19
3.1	VINAGRE	19
3.1.1	Qualidade, Classificação e Composição dos Vinagres.....	22
3.1.2	Legislação do Brasil	23
3.2	VINAGRE DE MAÇÃ.....	24
3.3	VINAGRES ORGÂNICOS	26
3.4	AUTENTICAÇÃO EM ALIMENTOS.....	27
3.4.1	Fraude em Vinagres.....	29
3.5	MÉTODOS ANALÍTICOS PARA DETECÇÃO DE FRAUDE EM ALIMENTOS E BEBIDAS	30
3.6	MÉTODOS ANALÍTICOS ALTERNATIVOS PARA DETECÇÃO DE FRAUDE EM ALIMENTOS E BEBIDAS.....	32
3.6.1	Ressonância Magnética Nuclear de Baixo Campo (¹ H LF-NMR).....	32
3.7	QUIMIOMETRIA	34
3.7.1	Análise Discriminante por Mínimos Quadrados Parciais (PLS-DA).....	36
3.8	INTELIGÊNCIA ARTIFICIAL E <i>MACHINE LEARNING</i> NA CIÊNCIA DE ALIMENTOS	37
3.8.1	<i>eXplainable Artificial Intelligence</i> (XAI).....	39
4	REFERÊNCIAS	42

CAPÍTULO II: DETECÇÃO DE FRAUDES EM VINAGRE DE MAÇÃ BRASILEIRO UTILIZANDO FINGERPRINTING POR RMN DE BAIXO CAMPO E MACHINE LEARNING

1	INTRODUÇÃO	52
2	MATERIAL E MÉTODOS	55
2.1	VINAGRES AUTÊNTICOS E ADULTERADOS.....	55
2.2	SESSÃO EXPERIMENTAL	57

2.3	AQUISIÇÃO DE DADOS DE ^1H LF-NMR E EXPORTAÇÃO ESPECTRAL	58
2.4	TRATAMENTO E PRÉ-PROCESSAMENTO DOS DADOS	59
2.5	EXTRAÇÃO SUPERVISIONADA DE CARACTERÍSTICAS E MODELO DE CLASSIFICAÇÃO	60
2.6	MATRIZES DE CONFUSÃO E INSPEÇÃO DE ERROS.....	61
3	RESULTADOS E DISCUSSÃO	62
3.1	<i>FINGERPRINTS</i> ESPECTRAIS DE RMN DE ^1H DE BAIXO CAMPO DE AMOSTRAS DE VINAGRE.....	62
3.2	EXPLORAÇÃO DOS DADOS.....	69
3.3	ESTRATÉGIA DE AUTENTICAÇÃO ORIENTADA POR DADOS E EFEITO DO PRÉ-PROCESSAMENTO ESPECTRAL	72
3.4	AUTENTICAÇÃO DE VINAGRES E DETECÇÃO DE ADULTERAÇÃO	74
3.4.1	EXP. 1: Vinagre de maçã orgânico adulterado com vinagre de maçã tradicional (AD vs AM).....	74
3.4.2	EXP. 2: Vinagre de maçã adulterado com vinagre de álcool (MA vs AM)	76
3.5	PADRÕES DE ERRO, COMPORTAMENTO DEPENDENTE DA CLASSE E INTERPRETAÇÃO QUÍMICA	78
4	CONCLUSÃO.....	80
5	AGRADECIMENTOS	81
6	REFERÊNCIAS	82
	MATERIAL SUPLEMENTAR	90

CAPÍTULO III: *SPECTRAL MODEL EXPLAINER* (SMX): INTELIGÊNCIA ARTIFICIAL EXPLICÁVEL NA CLASSIFICAÇÃO DE VINAGRES BRASILEIROS DE MAÇÃ POR RMN DE ^1H DE BAIXO CAMPO

1	INTRODUÇÃO	116
2	MATERIAL E MÉTODOS	120
2.1	AMOSTRAS E SESSÃO EXPERIMENTAL	120
2.2	AQUISIÇÃO E PRÉ-PROCESSAMENTO DE <i>FINGERPRINTS</i> ESPECTRAIS.....	122
2.3	PROCESSAMENTO ESPECTRAL E PREPARAÇÃO DOS DADOS	123
2.4	MODELO DE CLASSIFICAÇÃO PARA AS ANÁLISES DE EXPLICABILIDADE	124
2.5	<i>SPECTRAL MODEL EXPLAINER</i> (SMX)	125
2.5.1	Formulação Matemática do <i>Spectral Model eXplainer</i> (SMX).....	126

2.6	<i>AVALIAÇÃO DA FIDELIDADE DAS EXPLICAÇÕES.....</i>	128
3	RESULTADOS E DISCUSSÃO	129
3.1	<i>ORGANIZAÇÃO QUÍMICA DOS DOMÍNIOS ESPECTRAIS DE RMN DE BAIXO CAMPO ...</i>	129
3.2	<i>MODELOS DE CLASSIFICAÇÃO UTILIZADOS NAS ANÁLISES DE EXPLICABILIDADE</i>	133
3.3	<i>EXPLICAÇÕES ESPECTRAIS GERADAS PELO SMX</i>	136
3.4	<i>INTERPRETAÇÃO QUÍMICA DOS DOMÍNIOS ESPECTRAIS RELEVANTES.....</i>	142
3.5	<i>AVALIAÇÃO DA FIDELIDADE DAS EXPLICAÇÕES ESPECTRAIS</i>	148
3.5.1	Comparação entre o SMX e a Interpretação Quimiométrica Convencional	152
3.5.2	Implicações para a Autenticação Explicável de Alimentos e Bebidas.....	156
4	CONCLUSÃO.....	158
5	AGRADECIMENTOS	160
6	REFERÊNCIAS	160
	CONCLUSÃO GERAL	167
	ANEXO I - GLOSSÁRIO	168
	ANEXO II – LISTA DE SIGLAS E ABREVIATURAS.....	174

CAPÍTULO I

1. INTRODUÇÃO GERAL

A crescente complexidade das cadeias produtivas de alimentos, associada à globalização dos mercados e ao aumento da demanda por produtos diferenciados, impõe novos desafios relacionados à garantia da qualidade, autenticidade e rastreabilidade dos alimentos. Paralelamente, consumidores tornam-se cada vez mais exigentes quanto à origem, aos métodos de produção e às características nutricionais dos produtos, enquanto órgãos reguladores e a indústria necessitam de métodos analíticos capazes de detectar fraudes de maneira rápida, confiável e economicamente viável. Nesse contexto, a ciência de alimentos vem passando por uma importante transformação, migrando de abordagens baseadas na quantificação individual de compostos químicos para estratégias capazes de interpretar sistemas químicos complexos por meio da integração entre instrumentação analítica e ciência de dados (Butler *et al.*, 2021; Gao *et al.*, 2019; Hong *et al.*, 2017; Robson *et al.*, 2021).

Historicamente, o controle de qualidade de alimentos foi fundamentado em análises direcionadas, nas quais poucos marcadores químicos eram utilizados para verificar conformidade com padrões legais de identidade e qualidade, como por exemplo, acidez volátil mínima ou teor de cinzas. Embora essas metodologias permaneçam essenciais para fins regulatórios, sua capacidade de avaliar autenticidade frequentemente é limitada, principalmente quando diferentes produtos apresentam composição semelhante ou quando adulterações preservam os parâmetros físico-químicos convencionais. Como consequência, técnicas analíticas não direcionadas passaram a ocupar posição de destaque na química de alimentos, permitindo caracterizar simultaneamente centenas de variáveis químicas que, em conjunto, constituem uma assinatura molecular única de cada produto (Callao; Ruisánchez, 2018; Gao *et al.*, 2019; Hong *et al.*, 2017).

Entre essas abordagens, o conceito de *fingerprinting* espectral tornou-se particularmente relevante. Em vez de identificar individualmente todos os compostos presentes em uma matriz, o *fingerprinting* utiliza o perfil espectral completo como uma representação global da composição química do alimento. Essa estratégia permite capturar diferenças sutis decorrentes da matéria-prima, do processamento, das condições ambientais e de possíveis adulterações, sendo especialmente adequada para matrizes naturalmente

complexas e altamente variáveis, como produtos fermentados (Callao; Ruisánchez, 2018; Avanzi Barbosa Mascareli *et al.*, 2023; Ríos-Reina *et al.*, 2021).

Nesse cenário, a espectroscopia de Ressonância Magnética Nuclear (RMN) destaca-se como uma das ferramentas mais completas para obtenção de informações químicas. Tradicionalmente empregada em equipamentos de alto campo, a RMN consolidou-se como técnica robusta para metabolômica e caracterização de alimentos. Entretanto, avanços recentes na tecnologia de ímãs permanentes possibilitaram o desenvolvimento de instrumentos de RMN de baixo campo (*Low-Field Nuclear Magnetic Resonance* – LF-NMR), significativamente mais compactos, econômicos e operacionais, tornando sua aplicação cada vez mais atrativa para laboratórios industriais e rotinas de controle de qualidade. Apesar da menor resolução espectral em comparação à RMN de alto campo, estudos recentes demonstram que a combinação entre *fingerprints* de LF-NMR e ferramentas estatísticas multivariadas é capaz de produzir modelos altamente discriminativos para autenticação de alimentos (De Aguiar *et al.*, 2022; Galvan *et al.*, 2021; Gouilleux *et al.*, 2016; Grootveld *et al.*, 2019).

A evolução paralela da ciência de dados ampliou ainda mais o potencial dessas técnicas. Métodos quimiométricos tradicionais, como Análise de Componentes Principais (PCA) e Análise Discriminante por Mínimos Quadrados Parciais (PLS-DA), passaram a coexistir com algoritmos modernos de aprendizado de máquina, capazes de explorar padrões complexos presentes em grandes conjuntos de dados. Esses métodos transformaram o paradigma da análise química, permitindo que decisões anteriormente dependentes da interpretação humana passassem a ser conduzidas por modelos matemáticos treinados diretamente sobre os dados experimentais (Barros Neto; Scarminio; Bruns, 2006; Ferreira, 2015; Kato *et al.*, 2019; Nobari Moghaddam *et al.*, 2022).

Entretanto, o crescente desempenho preditivo desses modelos trouxe um novo desafio: sua interpretabilidade. Muitos algoritmos de inteligência artificial são frequentemente tratados como “caixas-pretas”, fornecendo excelentes resultados de classificação sem explicar adequadamente quais informações químicas fundamentam suas decisões. Essa limitação é particularmente relevante em aplicações científicas e regulatórias, nas quais a compreensão do comportamento do modelo é tão importante quanto sua capacidade preditiva. Nos últimos anos, a Inteligência Artificial Explicável (*eXplainable Artificial Intelligence* – XAI) emergiu como uma importante área de pesquisa dedicada a tornar modelos complexos mais transparentes, permitindo relacionar regiões espectrais ou variáveis químicas específicas ao processo de tomada de decisão dos algoritmos (Ahmed *et*

al., 2025; Arrieta *et al.*, 2020; Arrighi *et al.*, 2026; Ribeiro *et al.*, 2026).

O vinagre representa um excelente modelo para investigação dessas abordagens. Trata-se de um produto fermentado de ampla relevância econômica, cuja composição química é influenciada por fatores como matéria-prima, condições de fermentação, processamento e armazenamento. Particularmente, vinagres de maçã e vinagres orgânicos possuem elevado valor agregado e, conseqüentemente, maior suscetibilidade à adulteração e à fraude econômica. A complexidade natural dessas matrizes torna insuficiente a utilização exclusiva de parâmetros físico-químicos convencionais para autenticação, reforçando a necessidade de metodologias capazes de explorar toda a informação química disponível (Bekatorou, 2019; Cavdaroglu; Ozen, 2022; Ho *et al.*, 2017; Liu *et al.*, 2019; Ríos-Reina *et al.*, 2021).

Dentro desse contexto, este trabalho propôs integrar instrumentação analítica moderna, quimiometria, aprendizado de máquina e inteligência artificial explicável para desenvolver novas estratégias de autenticação de vinagres baseadas em *fingerprints* espectrais de LF-NMR. Assim, este trabalho posiciona-se na interface entre química analítica, ciência de alimentos e inteligência artificial, evidenciando como abordagens orientadas por dados podem contribuir para a modernização dos sistemas de controle de qualidade, oferecendo métodos mais robustos, escaláveis e interpretáveis para autenticação de alimentos e bebidas (Ahmed *et al.*, 2025; Butler *et al.*, 2021; Gao *et al.*, 2019; Ribeiro *et al.*, 2026).

2. OBJETIVOS

2.1. Objetivo Geral

Desenvolver e avaliar uma abordagem analítica integrada baseada em *fingerprinting* espectral por Ressonância Magnética Nuclear de Baixo Campo de hidrogênio (^1H LF-NMR), métodos quimiométricos, *machine learning* e Inteligência Artificial Explicável (XAI) para a classificação e interpretação química de vinagres de maçã, contribuindo para o desenvolvimento de metodologias robustas, interpretáveis e aplicáveis ao controle de qualidade de alimentos.

2.2. Objetivos Específicos

- Desenvolver uma metodologia de *fingerprinting* espectral utilizando espectroscopia de

^1H LF-NMR para caracterização química não direcionada de diferentes tipos de vinagres comerciais.

- Construir um conjunto de dados representativo contendo amostras autênticas e adulteradas de vinagres, contemplando diferentes matérias-primas, sistemas de produção e cenários de fraude.
- Investigar estratégias de pré-processamento espectral e tratamento de dados capazes de aumentar a robustez e a capacidade discriminativa dos *fingerprints* obtidos por LF-NMR.
- Desenvolver e validar modelos quimiométricos supervisionados para classificação e autenticação de vinagres, avaliando seu desempenho por meio de protocolos rigorosos de validação cruzada.
- Avaliar a capacidade da combinação entre *fingerprints* de LF-NMR e aprendizado de máquina na detecção de diferentes tipos de adulteração em vinagres de maçã.
- Aplicar técnicas de Inteligência Artificial Explicável (XAI) para interpretar o comportamento dos modelos preditivos, identificando as regiões espectrais mais relevantes para a tomada de decisão e relacionando às características químicas das amostras.
- Demonstrar a viabilidade da utilização de metodologias orientadas por dados como ferramentas para modernização do controle de qualidade e autenticação de alimentos, promovendo maior confiabilidade, transparência e escalabilidade na análise de produtos fermentados.

3. REVISÃO BIBLIOGRÁFICA

3.1. VINAGRE

O vinagre é um produto fermentado de natureza aquosa e caráter ácido, obtido por meio de duas etapas bioquímicas sucessivas: a fermentação alcoólica, seguida da fermentação acética. Sua produção pode ser realizada a partir de diferentes matérias-primas ricas em carboidratos fermentescíveis, incluindo frutas e mel, fontes naturais de glicose e frutose, bem como cereais, tubérculos e outros vegetais ricos em amido, os quais são previamente convertidos em açúcares fermentáveis (Bekatorou, 2019; Zhang *et al.*, 2020).

A identidade química e legal do vinagre é determinada principalmente por seu teor de acidez volátil, expresso em ácido acético. De acordo com a legislação brasileira, o produto deve apresentar, no mínimo, 4 g de ácido acético por 100 g de vinagre para ser comercializado sob essa denominação (Brasil, 2012a). Historicamente, o vinagre acompanha a humanidade há mais de três milênios, tendo sido empregado não apenas como condimento, mas também como conservante de alimentos, agente de limpeza e produto de higiene. Nas últimas décadas, entretanto, seu consumo passou a ser impulsionado também pelo crescente interesse em alimentos associados à promoção da saúde, em virtude da presença de compostos bioativos e dos potenciais efeitos fisiológicos atribuídos ao seu consumo (Bekatorou, 2019; Román-Camacho *et al.*, 2023).

A expansão desse mercado reflete essa mudança no perfil de consumo. Segundo estimativas da Global Market Insights, o mercado mundial de vinagres movimentou aproximadamente US\$ 6,64 bilhões em 2022, com expectativa de crescimento anual superior a 3,2% até 2032, impulsionado pela demanda crescente por produtos naturais, funcionais e de maior valor agregado (Kunal Ahuja; Himanshu Kabra, 2022).

Embora a água e o ácido acético constituam os componentes majoritários do vinagre, sua composição química é significativamente mais complexa. Diversos metabólitos presentes em menores concentrações, como ácidos orgânicos, aminoácidos, açúcares residuais, melanoidinas, compostos fenólicos, polifenóis e outros metabólitos derivados da fermentação, contribuem para suas propriedades sensoriais, nutricionais e funcionais (Kahraman *et al.*, 2021a; Xie *et al.*, 2022). A diversidade e a abundância desses compostos dependem diretamente da matéria-prima utilizada, das condições de fermentação e das etapas tecnológicas empregadas durante o processamento.

Entre os diferentes tipos de vinagre, aqueles produzidos a partir de frutas apresentam, em geral, maior riqueza composicional e maior concentração de compostos bioativos quando comparados aos vinagres elaborados a partir de cereais ou cana-de-açúcar. Essa característica resulta em perfis sensoriais mais complexos e maior potencial funcional, atributos que têm ampliado o interesse dos consumidores por esses produtos. Nesse contexto, o vinagre de maçã destaca-se como uma das variedades de maior relevância comercial, sendo amplamente valorizado por suas propriedades sensoriais e pela associação com compostos antioxidantes e outras substâncias bioativas (Kahraman *et al.*, 2021a; Ousaaid *et al.*, 2021).

Do ponto de vista sensorial, a elevada concentração de ácido acético confere ao vinagre sua característica de pungência e o sabor ácido característico. Entretanto, a percepção sensorial final resulta da interação entre esse ácido orgânico e diversos compostos minoritários

responsáveis pelo aroma, sabor e complexidade do produto, incluindo ésteres, como acetato de etila e acetato de isoamila; álcoois, como 1-hexanol e 3-metil-1-butanol; aldeídos, como acetaldeído e hexanal; além de outros ácidos orgânicos, como os ácidos málico, láctico e succínico, e compostos fenólicos (Bekatorou, 2019).

A obtenção do vinagre envolve duas etapas fermentativas consecutivas. Inicialmente, ocorre a fermentação alcoólica, na qual leveduras, principalmente *Saccharomyces cerevisiae*, convertem açúcares fermentescíveis em etanol sob condições predominantemente anaeróbias (Bekatorou, 2019; Hutchinson *et al.*, 2019; Román-Camacho *et al.*, 2023). Na etapa subsequente, denominada fermentação acética, bactérias do ácido acético (BAA), pertencentes principalmente aos gêneros *Acetobacter* e *Komagataeibacter*, oxidam o etanol na presença de oxigênio, produzindo ácido acético como principal metabólito (Bekatorou, 2019; Hutchinson *et al.*, 2019; Román-Camacho *et al.*, 2023).

Industrialmente, a produção de vinagre é realizada predominantemente por dois sistemas tecnológicos: o método tradicional, também conhecido como método Orleans, e o método submerso. No método tradicional, as bactérias acéticas desenvolvem-se na interface entre a superfície do líquido e o oxigênio atmosférico, formando uma película superficial responsável pela oxidação do etanol. Embora esse processo apresente menor produtividade e maior tempo de fermentação, é reconhecido por produzir vinagres de elevada qualidade sensorial, caracterizados por aromas mais complexos e sabores mais refinados (Bekatorou, 2019; Hutchinson *et al.*, 2019; Román-Camacho *et al.*, 2023).

Em contraste, o método submerso caracteriza-se pela manutenção das bactérias completamente dispersas no meio hidroalcoólico, com fornecimento contínuo de oxigênio por sistemas de aeração. Essa configuração proporciona maior taxa de transferência de oxigênio, reduz significativamente o tempo de fermentação e aumenta o rendimento produtivo. Como consequência, tornou-se o sistema predominante na produção industrial em larga escala. Entretanto, a elevada velocidade do processo tende a resultar em vinagres com perfil aromático menos complexo e maior predominância da pungência característica do ácido acético quando comparados aos obtidos pelo método tradicional (Bekatorou, 2019; Hutchinson *et al.*, 2019).

Após a fermentação, etapas adicionais podem ser incorporadas ao processo produtivo para agregar valor ao produto final. Entre elas, destaca-se a maturação em barris de madeira, prática amplamente utilizada na produção de vinagres especiais. Durante o envelhecimento, ocorrem reações químicas que promovem a formação de ésteres e outros compostos aromáticos, aumentando a complexidade sensorial, suavizando a percepção da acidez e conferindo maior valor comercial ao produto (Li *et al.*, 2021).

3.1.1. Qualidade, Classificação e Composição dos Vinagres

A qualidade dos vinagres é frequentemente associada às suas características sensoriais, especialmente aroma e sabor. Entretanto, sua identidade e classificação também são regulamentadas por critérios legais que estabelecem padrões mínimos de composição e qualidade. No Brasil, a classificação dos vinagres baseia-se principalmente na origem da matéria-prima utilizada na fermentação, abrangendo categorias como vinagres de álcool, frutas, cereais, vegetais e mel, além de produtos mistos (obtidos a partir de mais de uma matéria-prima), compostos (adicionados de sucos de frutas, vegetais ou mel), condimentados e aqueles classificados como duplos ou triplos, de acordo com a acidez volátil mínima, superior a 8 g 100 mL⁻¹ e 12 g 100 mL⁻¹, respectivamente (Brasil, 2012a).

A matéria-prima empregada e o processo de produção exercem influência direta sobre a composição química do vinagre, resultando em perfis metabólicos distintos entre as diferentes categorias comerciais. Embora a água e o ácido acético sejam os constituintes predominantes, o produto apresenta uma matriz química altamente complexa, composta por ácidos orgânicos voláteis e não voláteis, aldeídos, cetonas, ésteres, álcoois superiores e diversos metabólitos oriundos das fermentações alcoólica e acética. Entre os ácidos orgânicos, o ácido acético representa o principal componente responsável pela acidez característica do produto, enquanto o ácido láctico figura entre os metabólitos não voláteis mais abundantes. Estudos metabolômicos demonstram que determinados tipos de vinagre podem conter mais de 200 compostos químicos identificáveis, evidenciando a elevada complexidade dessas matrizes fermentadas (Ho *et al.*, 2017; Liu *et al.*, 2019).

Além dos metabólitos relacionados ao processo fermentativo, os vinagres constituem uma importante fonte de compostos bioativos, incluindo polifenóis, melanoidinas, aminoácidos, açúcares residuais, vitaminas e minerais. A concentração desses constituintes é influenciada por diversos fatores, como a natureza da matéria-prima, as condições de fermentação, as etapas de processamento e o período de envelhecimento do produto. Em um estudo comparativo envolvendo 23 tipos de vinagre, Liu *et al.* (2019) observaram que os vinagres de vinho branco e de maçã apresentaram as maiores concentrações de ácidos orgânicos, enquanto o Vinagre Balsâmico de Modena destacou-se pelo elevado teor de compostos fenólicos. Esses resultados reforçam que a composição química dos vinagres varia substancialmente em função de sua origem e do processo tecnológico empregado, influenciando diretamente suas propriedades sensoriais, funcionais e seu valor comercial (Liu *et al.*, 2019; Xia *et al.*, 2020).

3.1.2. Legislação do Brasil

A definição dos padrões de identidade, qualidade e classificação dos fermentados acéticos é de competência do Ministério da Agricultura e Pecuária (MAPA), que estabelece os requisitos técnicos necessários para a produção e comercialização desses produtos no território nacional. As exigências regulatórias diferem conforme a categoria de vinagre, contemplando parâmetros específicos de composição, identidade e qualidade. Os principais critérios aplicáveis aos fermentados acéticos encontram-se sintetizados no **Quadro 1**, enquanto as especificações particulares referentes ao vinagre de vinho são apresentadas no **Quadro 2** (Brasil, 2012a; Brasil, 2018).

Quadro 1. Parâmetros dos fermentados acéticos no Brasil

Parâmetros	Álcool		Fruta		Cereal		Vegetal/Mel	
	Mín.	Máx.	Mín.	Máx.	Mín.	Máx.	Mín.	Máx.
Acidez volátil (g 100mL ⁻¹)	4,00	-	4,00	-	4,00	-	4,00	-
Álcool (%v/v) a 20°C	-	1,00	-	1,00	-	1,00	-	1,00
Cinzas (g L ⁻¹)	-		1,00	5,00	1,00	5,00	1,00	5,00
Extrato seco reduzido (g L ⁻¹)	-		6,00	-	7,00	-	7,00	-
Sulfatos (g L ⁻¹ s. de potássio)	-		-	1,00	-	1,00	-	1,00
Aspecto	Ausência de elementos estranhos a sua natureza e composição.							
Cheiro	Característico.							
Gosto	Ácido.							
Cor	De acordo com a matéria-prima de origem e composição.							

Fonte: Adaptado de; BRASIL, 2012a.

Quadro 2. Parâmetros exclusivamente para o vinagre de vinho

Parâmetros	Mínimo	Máximo
Acidez volátil (g 100mL ⁻¹)	5,00	-
Álcool (%v/v) a 20°C	-	1,00
Cinzas (g L ⁻¹)	1,00	-

Sulfatos (g L ⁻¹ s. de potássio)	-	1,20
Cloretos (g L ⁻¹ c. de sódio)	-	1,00
Extrato seco reduzido (g L ⁻¹)	7,00	-

Fonte: Adaptado de; Brasil, 2018.

Além dos critérios físico-químicos, a legislação vigente estabelece requisitos específicos relacionados à identidade e à comercialização dos fermentados acéticos. Dentre essas exigências, destacam-se a obrigatoriedade de informar no rótulo a utilização de corantes, aromatizantes e outros aditivos permitidos, bem como restrições quanto às características das embalagens, que não devem remeter à apresentação de produtos farmacêuticos. Em conjunto, essas determinações contribuem para a padronização da identidade do produto, assegurando transparência na rotulagem e proteção ao consumidor (Brasil, 2018).

3.2.VINAGRE DE MAÇÃ

O vinagre de maçã destaca-se como uma das variedades de vinagre de maior relevância comercial e científica, em virtude de suas características sensoriais, composição química e crescente associação com alegações de promoção da saúde. Tradicionalmente empregado como condimento culinário, seu uso expandiu-se nas últimas décadas para aplicações relacionadas ao bem-estar, medicina complementar e higiene doméstica. Esse produto é obtido pela fermentação acética da cidra de maçã, previamente produzida por fermentação alcoólica do suco da fruta, resultando em uma matriz fermentada rica em metabólitos derivados de ambas as etapas fermentativas (Bekatorou, 2019; Ousaaïd *et al.*, 2020).

Assim como ocorre com outros tipos de vinagre, sua composição é constituída predominantemente por água e ácido acético. Entretanto, a utilização da maçã como matéria-prima confere um perfil químico mais complexo e maior valor agregado em comparação aos vinagres produzidos a partir de álcool ou cereais. Além dos constituintes majoritários, o vinagre de maçã apresenta uma composição química diversificada, incluindo ácidos orgânicos, como os ácidos málico, cítrico, láctico, succínico e tartárico, e compostos fenólicos, como os ácidos clorogênico, gálico, cafeico e p-cumárico, além de catequina, epicatequina, quercetina e floridzina. Também estão presentes aminoácidos, açúcares residuais, vitaminas, minerais e outros metabólitos provenientes da matéria-prima e das transformações ocorridas durante os

processos fermentativos. Esses compostos contribuem tanto para a complexidade aromática e gustativa quanto para as propriedades funcionais atribuídas ao produto (Kahraman *et al.*, 2021b; Štornik *et al.*, 2016).

O interesse crescente pelo consumo de vinagre de maçã também está relacionado aos seus potenciais efeitos fisiológicos. Estudos clínicos e experimentais sugerem que seu consumo pode contribuir para a melhora do controle glicêmico em indivíduos com diabetes mellitus tipo 2, favorecer a modulação do metabolismo lipídico, auxiliar na redução do peso corporal e exercer efeitos benéficos sobre a saúde cardiovascular. Embora muitos desses resultados ainda necessitem de validação em estudos clínicos de longo prazo, as evidências disponíveis têm impulsionado o interesse científico e comercial por esse produto (El Abdali *et al.*, 2023; Morgan *et al.*, 2016; Ousaaïd *et al.*, 2020; Perumpuli *et al.*, 2022).

Grande parte desses efeitos tem sido atribuída à ação do ácido acético, principal ácido orgânico presente no vinagre, o qual pode aumentar a sensibilidade à insulina, reduzir a velocidade de absorção de carboidratos e modular a produção hepática de glicose, contribuindo para a diminuição da resposta glicêmica pós-prandial. Paralelamente, os compostos fenólicos e outros antioxidantes presentes na matriz podem atuar na neutralização de espécies reativas de oxigênio, reduzindo o estresse oxidativo e, conseqüentemente, auxiliando na prevenção de doenças crônicas relacionadas ao processo inflamatório e ao envelhecimento celular (Kara *et al.*, 2021; El Abdali *et al.*, 2023).

Além das evidências científicas recentes, o vinagre de maçã possui longa tradição na medicina popular, sendo historicamente empregado como coadjuvante no tratamento de afecções como dor de garganta, distúrbios digestivos e alterações cutâneas. Embora muitas dessas aplicações ainda careçam de comprovação clínica robusta, elas contribuíram para consolidar sua imagem como alimento funcional e impulsionaram sua valorização comercial (El Abdali *et al.*, 2023).

O aumento da demanda, aliado ao maior valor econômico do vinagre de maçã, principalmente na categoria orgânica, ampliou sua vulnerabilidade a práticas fraudulentas. Entre as fraudes mais frequentes destacam-se a substituição parcial por vinagres de menor valor comercial, a rotulagem enganosa quanto à origem da matéria-prima e a utilização de práticas de produção não permitidas pela legislação. No Brasil, por exemplo, a adição de açúcares durante a produção do vinagre de maçã é expressamente proibida, sendo caracterizada como fraude por comprometer a identidade e a autenticidade do produto (Brasil, 2020). Nesse contexto, o desenvolvimento de metodologias analíticas robustas capazes de verificar a autenticidade e detectar adulterações tornou-se uma necessidade crescente para a indústria,

órgãos reguladores e consumidores.

3.3.VINAGRES ORGÂNICOS

Nas últimas décadas, o aumento da preocupação dos consumidores com saúde, sustentabilidade e qualidade dos alimentos impulsionou significativamente a demanda por produtos orgânicos em diferentes segmentos do mercado alimentício. De maneira geral, alimentos orgânicos são obtidos por sistemas de produção que restringem o uso de fertilizantes sintéticos, agrotóxicos, organismos geneticamente modificados e outros insumos químicos, priorizando práticas agrícolas sustentáveis voltadas à conservação dos recursos naturais, ao bem-estar dos produtores e à oferta de alimentos de elevada qualidade (Kushwah *et al.*, 2019).

No Brasil, a produção orgânica é regulamentada pelo Ministério da Agricultura e Pecuária (MAPA), que define como orgânicos os produtos oriundos de sistemas de produção agropecuária ou de processos extrativistas sustentáveis capazes de preservar a integridade dos ecossistemas e promover o uso responsável dos recursos naturais. Além dos requisitos relacionados às práticas de cultivo e processamento, a comercialização desses produtos depende da conformidade com os mecanismos oficiais de controle e certificação estabelecidos pela legislação brasileira (Brasil, 2009).

Esse cenário também se reflete no mercado de vinagres, especialmente no segmento de vinagres de maçã orgânicos, cuja demanda tem aumentado em função da crescente valorização de alimentos produzidos segundo princípios de sustentabilidade ambiental e menor utilização de insumos químicos. Estudos comparativos sugerem que o sistema de cultivo e as condições de processamento podem influenciar a microbiota presente durante a fermentação acética. Nesse contexto, vinagres produzidos a partir de matéria-prima orgânica apresentaram maior diversidade microbiana em comparação aos convencionais, característica que pode contribuir para diferenças na composição química, no perfil metabólico e nas propriedades sensoriais do produto final (Bouatenin *et al.*, 2021; Jaikang *et al.*, 2019; Štornik, Skok & Trček, 2016).

Além dos aspectos relacionados à sustentabilidade, o vinagre de maçã orgânico tem despertado interesse devido aos potenciais efeitos benéficos atribuídos ao consumo de vinagre de maçã, incluindo possíveis contribuições para o controle glicêmico em indivíduos com diabetes *mellitus* tipo 2. Entretanto, as evidências disponíveis indicam que esses efeitos estão associados principalmente à composição do vinagre e à presença de compostos bioativos, não havendo consenso de que a certificação orgânica, por si só, resulte em benefícios fisiológicos superiores aos observados em produtos convencionais (Hong *et al.*, 2017; Morgan & Mosawy,

2016).

O maior custo de produção decorrente das exigências de certificação, rastreabilidade e manejo agrícola sustentável torna os produtos orgânicos mais valorizados comercialmente quando comparados aos seus equivalentes convencionais. Essa diferença de valor agregado aumenta a suscetibilidade desse mercado a práticas fraudulentas, incluindo a comercialização de produtos rotulados como orgânicos sem que atendam aos requisitos legais de certificação. No setor de vinagres, esse cenário reforça a importância do desenvolvimento de metodologias analíticas capazes de verificar, de forma objetiva e confiável, a autenticidade da origem orgânica e assegurar a conformidade com a legislação vigente (Brasil, 2020).

3.4. AUTENTICAÇÃO EM ALIMENTOS

Nos últimos anos, conceitos como *Food Fraud*, *Food Authentication*, *Food Integrity* e *Food Control* têm assumido papel central nas pesquisas voltadas à segurança e à qualidade dos alimentos. O crescente processo de globalização das cadeias produtivas, aliado à maior complexidade dos sistemas de distribuição e comercialização, aumentou a necessidade de métodos capazes de garantir a autenticidade e a integridade dos produtos alimentícios. Nesse contexto, a fraude em alimentos é definida como qualquer adulteração, substituição, falsificação ou deturpação intencional de alimentos, ingredientes, embalagens ou informações de rotulagem, realizada com o objetivo de obter vantagem econômica ilícita (FAO, 2021).

As fraudes alimentares podem assumir diferentes formas, incluindo substituição parcial ou total de ingredientes, diluição de produtos, adição de substâncias não autorizadas, adulteração da composição química, falsificação da origem geográfica ou botânica, utilização indevida de certificações, além de declarações enganosas relacionadas à qualidade ou ao processo de produção. Embora muitas dessas práticas tenham motivação exclusivamente econômica, suas consequências extrapolam as perdas financeiras, comprometendo a confiança do consumidor, a competitividade do setor produtivo e, em diversas situações, a saúde pública (FAO, 2021).

Diversos episódios ocorridos nas últimas décadas evidenciam o impacto das fraudes alimentares em escala global. Um dos casos mais emblemáticos foi registrado na Espanha, na década de 1980, quando óleo industrial de colza desnaturado foi comercializado fraudulentamente como azeite de oliva, ocasionando milhares de casos de intoxicação e centenas de óbitos. Outro episódio amplamente documentado ocorreu na China, em 2008, quando a adição deliberada de melamina ao leite destinado ao consumo infantil resultou em

centenas de milhares de intoxicações e graves complicações renais, tornando-se um dos maiores escândalos da indústria alimentícia moderna (Gelpí *et al.*, 2002; Gossner *et al.*, 2009).

No Brasil, casos de grande repercussão também demonstram a vulnerabilidade das cadeias produtivas à fraude. Entre eles destacam-se a Operação Leite Compensado, que revelou a adulteração de leite por meio da adição de água e compostos químicos destinados a mascarar alterações na composição, e a Operação Carne Fraca, responsável por identificar irregularidades envolvendo a utilização de ingredientes não autorizados, falhas de inspeção e práticas fraudulentas em produtos cárneos comercializados por grandes empresas do setor (Andreatta *et al.*, 2019; Quevedo-Silva, Freire & Spanhol-Finocchio, 2019).

A identificação dessas fraudes representa um desafio significativo, uma vez que muitas adulterações não produzem alterações perceptíveis nas características sensoriais do alimento. Consequentemente, inspeções visuais ou avaliações organolépticas frequentemente são insuficientes para detectar modificações intencionais na composição dos produtos. Dessa forma, a autenticação de alimentos depende cada vez mais da aplicação de metodologias analíticas capazes de caracterizar a composição química das amostras e identificar desvios em relação aos padrões esperados. Nesse cenário, técnicas instrumentais modernas associadas à quimiometria e ao aprendizado de máquina têm se consolidado como ferramentas promissoras para a detecção rápida, objetiva e confiável de fraudes alimentares (Butler *et al.*, 2021; Kazazić *et al.*, 2021; Robson *et al.*, 2021).

Sob a perspectiva analítica, a investigação da autenticidade de um alimento pode envolver diferentes abordagens, dependendo da natureza da fraude. Em muitos casos, busca-se verificar a origem geográfica, botânica, zoológica ou o sistema de produção do alimento por meio de estratégias de classificação e reconhecimento de padrões baseadas em análises multivariadas. Em outros, o objetivo consiste na identificação e quantificação de compostos adicionados indevidamente ou da substituição parcial de ingredientes, exigindo métodos analíticos qualitativos e quantitativos de elevada sensibilidade e especificidade, tais como cromatografia líquida de alta eficiência ou cromatografia gasosa (Kazazić *et al.*, 2021).

No caso específico do vinagre de maçã, o crescente valor agregado do produto, especialmente na categoria orgânica, associado à dificuldade de verificação da autenticidade apenas por inspeção visual ou documental, torna esse alimento particularmente suscetível a diferentes modalidades de fraude. Esse cenário evidencia a necessidade do desenvolvimento de metodologias analíticas rápidas, não destrutivas e de alta confiabilidade, capazes de explorar diferenças globais na composição química e fornecer suporte objetivo à autenticação e ao controle de qualidade, aspectos que constituem o foco da presente tese.

3.4.1. Fraude em Vinagres

A elevada diversidade de matérias-primas, processos fermentativos e tecnologias empregadas na produção de vinagres resulta em perfis químicos característicos para cada categoria de produto. Essas diferenças composicionais possibilitam a avaliação da autenticidade por meio de abordagens analíticas qualitativas e quantitativas, capazes de identificar alterações na composição química decorrentes de adulterações ou desvios do processo produtivo. Além de proteger o consumidor, a autenticação contribui para preservar as características sensoriais, nutricionais e o valor comercial dos diferentes tipos de vinagre. Como exemplo, vinagres produzidos a partir de frutas apresentam, em geral, maior concentração e diversidade de compostos fenólicos quando comparados aos obtidos de álcool ou cereais, característica frequentemente explorada como marcador de autenticidade (Boffo *et al.*, 2009).

A simplicidade dos requisitos regulatórios mínimos para comercialização de vinagres em diversos países, baseados principalmente na concentração mínima de ácido acético, favorece a ocorrência de determinadas modalidades de fraude. Entre elas, destaca-se a adição de ácido acético sintético (ácido acético glacial) para corrigir ou elevar artificialmente a acidez de produtos de baixa qualidade, bem como a produção de vinagres formulados exclusivamente pela diluição desse composto em água. Dependendo da legislação vigente, essas práticas podem configurar adulteração por descaracterizarem a origem fermentativa do produto (Cavdaroglu & Ozen, 2022).

Outra fraude amplamente relatada consiste na substituição parcial de vinagres produzidos a partir de matérias-primas de maior valor agregado por vinagres obtidos de álcool etílico ou de cana-de-açúcar. Essa adulteração é particularmente relevante em vinagres de vinho e de frutas, nos quais a adição de vinagre de álcool permite aumentar o volume comercializado com menor custo de produção, comprometendo a composição química e reduzindo o valor econômico do produto. Além disso, também são descritas fraudes relacionadas à falsa declaração da origem geográfica, ao uso indevido de denominações de origem protegida, à rotulagem incorreta quanto ao tempo de envelhecimento e à utilização de alegações enganosas sobre o processo produtivo (Consonni *et al.*, 2008a; Jamin & Thomas, 2018; Ríos-Reina *et al.*, 2021).

No caso dos vinagres de frutas, outra prática fraudulenta consiste na adição de açúcares ao mosto durante a etapa de fermentação alcoólica, especialmente quando a matéria-prima apresenta baixa concentração de açúcares fermentescíveis. Essa estratégia aumenta

artificialmente o rendimento da produção de etanol e, conseqüentemente, a quantidade final de ácido acético obtida após a fermentação acética, descaracterizando o processo produtivo e infringindo a legislação brasileira aplicável aos vinagres de frutas (Brasil, 2020).

Nas últimas décadas, avanços em técnicas analíticas instrumentais têm impulsionado o desenvolvimento de metodologias destinadas à autenticação de vinagres, abrangendo diferentes estratégias baseadas em espectroscopia, cromatografia, ressonância magnética nuclear, metabolômica e análise isotópica. Apesar desse progresso, permanecem desafios importantes relacionados à detecção de adulterações em produtos de maior valor agregado, à identificação de fraudes envolvendo sistemas de produção específicos, como alimentos orgânicos, e ao desenvolvimento de métodos rápidos, não destrutivos e capazes de avaliar simultaneamente múltiplos atributos de autenticidade. Nesse contexto, o vinagre de maçã orgânico representa uma matriz particularmente desafiadora, uma vez que diferenças sutis em sua composição química exigem abordagens analíticas de elevada sensibilidade associadas a ferramentas quimiométricas e de aprendizado de máquina para a extração de informações relevantes. Dessa forma, o desenvolvimento de metodologias baseadas em *fingerprinting* químico e inteligência artificial constitui uma alternativa promissora para fortalecer o controle de qualidade, a rastreabilidade e a autenticação desses produtos.

3.5. MÉTODOS ANALÍTICOS PARA DETECÇÃO DE FRAUDE EM ALIMENTOS E BEBIDAS

O desenvolvimento de metodologias analíticas para autenticação de alimentos tem evoluído significativamente nas últimas décadas, impulsionado pelo aumento da complexidade das cadeias produtivas e pela crescente ocorrência de fraudes alimentares. Tradicionalmente, técnicas cromatográficas e espectrométricas constituem a base das análises de autenticidade, destacando-se a Cromatografia Líquida de Alta Eficiência (CLAE), a Cromatografia Gasosa (CG) e a Espectrometria de Massas (EM). Essas técnicas apresentam elevada sensibilidade, seletividade e capacidade de identificação molecular, sendo amplamente empregadas na caracterização da composição química de alimentos e na detecção de compostos marcadores de adulteração (Callao & Ruisánchez, 2018; Hong *et al.*, 2017).

Embora essas abordagens sejam consideradas métodos de referência em diversas aplicações, sua utilização frequentemente envolve preparo extensivo das amostras, análises demoradas e elevado custo operacional. Em resposta a essas limitações, observa-se um interesse crescente por metodologias capazes de fornecer informações químicas de forma rápida,

minimamente destrutiva e com reduzida necessidade de preparo da amostra. Nesse contexto, técnicas espectroscópicas e espectrométricas, como a Ressonância Magnética Nuclear (RMN), a espectroscopia no infravermelho (FT-IR e FT-NIR), a espectroscopia Raman e outras abordagens de *fingerprinting* químico, têm se consolidado como alternativas promissoras para autenticação de alimentos (Callao & Ruisánchez, 2018; Hong *et al.*, 2017).

Paralelamente à evolução instrumental, a estratégia analítica também sofreu mudanças importantes. Em vez da quantificação isolada de compostos específicos, cresce o emprego de abordagens não direcionadas, nas quais o conjunto completo das informações espectrais ou cromatográficas é utilizado para representar a composição global do alimento. Esse conceito, conhecido como *chemical fingerprinting* ou *spectral fingerprinting*, permite explorar simultaneamente centenas ou milhares de variáveis químicas, fornecendo uma assinatura analítica característica de cada amostra. A interpretação dessas assinaturas depende da aplicação de ferramentas estatísticas multivariadas, quimiometria e, mais recentemente, algoritmos de aprendizado de máquina, capazes de identificar padrões complexos relacionados à origem, autenticidade e qualidade dos alimentos (Callao & Ruisánchez, 2018; Gao *et al.*, 2019).

Em virtude dessas características, métodos voltados à autenticação de alimentos devem combinar rapidez, sensibilidade, especificidade, mínimo preparo da amostra, reprodutibilidade e viabilidade para análises de rotina. Nesse contexto, técnicas espectroscópicas associadas à quimiometria têm se consolidado como ferramentas para caracterização e detecção de adulterações, destacando-se a RMN, pela obtenção de *fingerprints* químicos e perfis metabólicos; FT-IR e NIR, pela caracterização rápida e não destrutiva da composição global da matriz; e UV-Vis, pela resposta de espécies cromóforas, incluindo compostos fenólicos. Essas abordagens têm sido amplamente exploradas na classificação, discriminação e autenticação de alimentos e bebidas. (Gao *et al.*, 2019; Avanzi Barbosa Mascareli *et al.*, 2023).

No caso específico dos vinagres, os métodos tradicionalmente empregados para investigação de autenticidade baseiam-se principalmente em técnicas cromatográficas acopladas a diferentes sistemas de detecção, incluindo Cromatografia Líquida e Cromatografia Gasosa acopladas à Espectrometria de Massas (LC-MS e GC-MS), além de Cromatografia Líquida com detector por Arranjo de Diodos (HPLC-DAD). Essas metodologias possibilitam a identificação e quantificação de compostos específicos relacionados à matéria-prima, ao processo fermentativo ou à ocorrência de adulterações, apresentando elevada robustez analítica (Cavdaroglu & Ozen, 2022; Consonni *et al.*, 2008b; Ríos-Reina *et al.*, 2021).

Nos últimos anos, entretanto, técnicas espectroscópicas baseadas em fingerprinting têm

despertado interesse crescente para autenticação de vinagres, especialmente por permitirem análises rápidas, não destrutivas e com reduzido preparo das amostras. Nessa abordagem, o espectro completo da amostra é utilizado como uma assinatura química (spectral fingerprint), cuja interpretação é realizada por métodos quimiométricos e algoritmos de aprendizado de máquina. Essa estratégia vem sendo aplicada com sucesso utilizando diferentes plataformas analíticas, incluindo RMN de alto campo (^{13}C e ^1H), espectroscopia no infravermelho próximo (FT-NIR), infravermelho médio (FT-IR) e espectroscopia UV-Vis, permitindo discriminar diferentes tipos de vinagre, identificar adulterações e avaliar atributos relacionados à autenticidade e à qualidade do produto (Cavdaroglu & Ozen, 2022; Consonni *et al.*, 2008b; Grassi & Alamprese, 2024; Xie *et al.*, 2012).

Dentre essas tecnologias, a Ressonância Magnética Nuclear destaca-se por fornecer informações abrangentes sobre a composição molecular da amostra de forma altamente reprodutível e quantitativa, sem necessidade de separação cromatográfica. Quando associada à quimiometria e ao aprendizado de máquina, a RMN permite explorar diferenças sutis na composição química dos alimentos, constituindo uma das abordagens mais promissoras para autenticação baseada em *fingerprinting*. Essa combinação representa o fundamento metodológico da presente tese, na qual espectros de RMN de baixo campo foram integrados a métodos quimiométricos e técnicas de inteligência artificial para classificação, autenticação e interpretação química de vinagres.

3.6. MÉTODOS ANALÍTICOS ALTERNATIVOS PARA DETECÇÃO DE FRAUDE EM ALIMENTOS E BEBIDAS

3.6.1. Ressonância Magnética Nuclear de Baixo Campo (^1H LF-NMR)

A Ressonância Magnética Nuclear (RMN) constitui uma das técnicas analíticas mais versáteis para caracterização molecular, sendo amplamente empregada em química, bioquímica, ciência dos alimentos, metabolômica e desenvolvimento farmacêutico. Dentre os diferentes núcleos passíveis de análise, o hidrogênio (^1H) é o mais frequentemente investigado devido à sua elevada abundância natural, alta sensibilidade magnetogírica e ampla distribuição em moléculas orgânicas. Tradicionalmente, a RMN de alto campo (^1H HF-NMR), operando em frequências iguais ou superiores a 400 MHz, representa a principal plataforma para elucidação estrutural e análises metabolômicas de alta resolução. Entretanto, avanços recentes no desenvolvimento de magnetos permanentes permitiram o surgimento da Ressonância

Magnética Nuclear de Baixo Campo ^1H LF-NMR, cujos equipamentos operam, em geral, entre 43 e 90 MHz, oferecendo uma alternativa mais compacta, robusta e economicamente acessível (De Aguiar *et al.*, 2022; Galvan *et al.*, 2021b; Grootveld *et al.*, 2019).

Ao contrário dos espectrômetros convencionais, que utilizam magnetos supercondutores resfriados por hélio líquido e nitrogênio líquido, os equipamentos de baixo campo empregam magnetos permanentes, frequentemente baseados na geometria de Halbach, um arranjo de ímãs com orientações de magnetização espacialmente distribuídas de modo a concentrar e homogeneizar o campo magnético na região interna do conjunto, enquanto reduz sua intensidade externamente. Essa configuração elimina a necessidade de sistemas criogênicos, reduz significativamente os custos de aquisição e manutenção e possibilita a construção de instrumentos compactos, ampliando o potencial de utilização da técnica em laboratórios industriais, plantas de produção e aplicações *at-line* e *on-line* (De Aguiar *et al.*, 2022; Galvan *et al.*, 2021b; Magalhães de Aguiar *et al.*, 2020).

Apesar das diferenças instrumentais, a ^1H LF-NMR fundamenta-se nos mesmos princípios físicos da RMN convencional. Quando submetidos a um campo magnético externo, os núcleos de hidrogênio, que possuem momento magnético nuclear diferente de zero, distribuem-se entre dois estados de energia correspondentes aos spins paralelo e antiparalelo em relação ao campo aplicado. A diferença energética entre esses estados é proporcional à intensidade do campo magnético e determina a frequência de ressonância descrita pela equação de Larmor (Galvan, 2019; Gouilleux *et al.*, 2016; Guthrie, 2014; Pavia, 2014).

A aplicação de um pulso de radiofrequência na frequência de ressonância promove a transição dos núcleos entre esses estados energéticos. Após a interrupção do pulso, os spins retornam ao equilíbrio termodinâmico por meio dos processos de relaxação longitudinal (T_1) e transversal (T_2), emitindo um sinal eletromagnético conhecido como *Free Induction Decay* (FID). A transformação matemática desse sinal, geralmente por meio da Transformada de Fourier, produz o espectro de RMN, no qual os deslocamentos químicos refletem o ambiente eletrônico dos núcleos de hidrogênio presentes na amostra (Guthrie, 2014; Pavia, 2014).

Embora a redução da intensidade do campo magnético resulte em menor dispersão dos deslocamentos químicos e maior sobreposição de sinais espectrais quando comparada à RMN de alto campo, essa limitação é frequentemente compensada pela aplicação de ferramentas quimiométricas e algoritmos de aprendizado de máquina. Nessas abordagens, o espectro completo passa a ser tratado como uma assinatura química (*spectral fingerprint*), permitindo explorar diferenças globais na composição das amostras sem a necessidade de identificar individualmente cada metabólito (Galvan *et al.*, 2021b; Grootveld *et al.*, 2019).

Essa característica tornou a ^1H LF-NMR particularmente atrativa para aplicações em ciência de alimentos. A técnica permite análises rápidas, não destrutivas, altamente reprodutíveis e com reduzido preparo das amostras, possibilitando tanto abordagens direcionadas, voltadas à quantificação de compostos específicos, quanto estratégias não direcionadas baseadas em *chemical fingerprinting*. Essas últimas têm demonstrado elevado potencial para avaliação da autenticidade, classificação de alimentos, monitoramento de processos fermentativos e detecção de fraudes (De Aguiar *et al.*, 2022; Galvan *et al.*, 2021b). Diversos estudos recentes demonstram a aplicabilidade da ^1H LF-NMR em diferentes matrizes alimentícias, incluindo óleos vegetais, bebidas, produtos lácteos e alimentos fermentados. Galvan *et al.* (2021), por exemplo, demonstraram que espectros obtidos em baixo campo, associados à quimiometria, foram capazes de prever parâmetros de qualidade de óleos comestíveis com desempenho comparável ao obtido utilizando RMN de alto campo, evidenciando o potencial da técnica como ferramenta alternativa para aplicações em controle de qualidade industrial.

No caso específico dos vinagres, entretanto, a aplicação da ^1H LF-NMR permanece relativamente pouco explorada quando comparada a outras plataformas analíticas, como cromatografia, espectrometria de massas e RMN de alto campo. Essa lacuna representa uma oportunidade relevante para o desenvolvimento de metodologias baseadas em *spectral fingerprinting*, especialmente quando associadas à quimiometria e ao aprendizado de máquina, permitindo a diferenciação entre tipos de vinagre, a identificação de adulterações e a autenticação de produtos de maior valor agregado, como os vinagres de maçã orgânicos.

3.7. QUIMIOMETRIA

O desenvolvimento da quimiometria está intimamente relacionado aos avanços da instrumentação analítica e da capacidade computacional observados a partir da década de 1970. A crescente geração de conjuntos de dados químicos de alta dimensionalidade impulsionou o surgimento dessa área interdisciplinar, que integra conhecimentos de química, estatística, matemática e ciência da computação com o objetivo de planejar experimentos, extrair informações relevantes e construir modelos capazes de descrever ou prever propriedades de sistemas químicos complexos (Barros Neto, Scarminio & Bruns, 2006; Ferreira, 2015).

Atualmente, a quimiometria constitui um dos principais pilares da química analítica moderna, desempenhando papel fundamental na interpretação de dados provenientes de técnicas instrumentais de alta complexidade, como espectroscopia, cromatografia,

espectrometria de massas e ressonância magnética nuclear. Nessas aplicações, o interesse frequentemente deixa de estar restrito à interpretação individual de variáveis químicas, passando a explorar relações multivariadas existentes entre centenas ou milhares de sinais analíticos obtidos simultaneamente (Barros Neto, Scarminio & Bruns, 2006; Ferreira, 2015).

De maneira geral, as aplicações quimiométricas podem ser agrupadas em três grandes áreas: planejamento experimental (*Design of Experiments* – DoE), reconhecimento de padrões (*Pattern Recognition*) e calibração multivariada. Na ciência de alimentos, particularmente em estudos de autenticação, controle de qualidade e detecção de fraudes, predominam as abordagens de reconhecimento de padrões e calibração multivariada, uma vez que essas estratégias permitem explorar a variabilidade química inerente às diferentes matrizes alimentícias e estabelecer modelos capazes de discriminar amostras autênticas e adulteradas (Barros Neto, Scarminio & Bruns, 2006; Ferreira, 2015; Varmuza & Filzmoser, 2009; Lavine & Workman, 2008).

No contexto do reconhecimento de padrões, diferentes algoritmos podem ser empregados conforme o objetivo da análise. Métodos exploratórios, como a Análise de Componentes Principais (*Principal Component Analysis* – PCA), permitem reduzir a dimensionalidade dos dados preservando a maior parte da variabilidade original, facilitando a visualização de agrupamentos, tendências e possíveis amostras discrepantes. Por outro lado, técnicas de calibração multivariada, como a Regressão por Componentes Principais (*Principal Component Regression* – PCR) e os Quadrados Mínimos Parciais (*Partial Least Squares* – PLS), são amplamente utilizadas para estabelecer relações quantitativas entre conjuntos de variáveis preditoras e respostas analíticas de interesse (Olivieri, 2018; Varmuza & Filzmoser, 2009).

Os métodos quimiométricos também podem ser classificados de acordo com a disponibilidade de informações sobre as amostras durante o treinamento dos modelos. Nos métodos não supervisionados, a estrutura dos dados é investigada sem conhecimento prévio das classes ou categorias às quais pertencem as amostras, sendo empregados principalmente em análises exploratórias e reconhecimento de padrões naturais. Em contraste, os métodos supervisionados utilizam informações previamente conhecidas sobre a identidade das amostras para construir modelos preditivos capazes de realizar classificação ou regressão. Entre esses algoritmos destacam-se a Análise Discriminante por Quadrados Mínimos Parciais (PLS-DA), Máquinas de Vetores de Suporte (SVM), *Random Forest*, Redes Neurais Artificiais e diversos outros métodos de aprendizado de máquina amplamente empregados na autenticação de alimentos (Ferreira, 2015; Olivieri, 2018).

A crescente adoção de técnicas espectroscópicas de alta produtividade consolidou a quimiometria como ferramenta indispensável para análise de dados de *fingerprinting* químico. Nessa abordagem, o espectro completo é utilizado como representação global da composição química da amostra, permitindo que algoritmos multivariados identifiquem padrões dificilmente perceptíveis por inspeção visual ou por análises univariadas convencionais. Consequentemente, a combinação entre técnicas instrumentais e métodos quimiométricos tornou-se uma das principais estratégias para classificação, autenticação e detecção de fraudes em alimentos, fornecendo modelos robustos capazes de explorar informações distribuídas ao longo de todo o perfil espectral.

3.7.1. Análise Discriminante por Mínimos Quadrados Parciais (PLS-DA)

A Análise Discriminante por Quadrados Mínimos Parciais (*Partial Least Squares Discriminant Analysis* – PLS-DA) é um dos métodos supervisionados mais empregados em quimiometria para problemas de classificação, particularmente em aplicações envolvendo dados espectrais de alta dimensionalidade. Sua ampla utilização em espectroscopia, metabolômica e autenticação de alimentos decorre da capacidade de lidar com conjuntos de dados caracterizados por elevado número de variáveis, forte colinearidade entre preditores e número relativamente reduzido de amostras, condições frequentemente encontradas em análises químicas modernas (Ferreira, 2015; Mutz *et al.*, 2022; Olivieri, 2018).

De forma geral, os métodos supervisionados de classificação podem ser agrupados em duas categorias principais: classificação de uma classe (*One-Class Classification* – OCC) e classificação discriminante (*Discriminant Classification* – DC). Nos modelos OCC, o objetivo consiste em modelar exclusivamente a classe de interesse, permitindo avaliar se uma nova amostra pertence ou não a essa categoria, sem a necessidade de representar explicitamente todas as demais classes existentes. Em contraste, os métodos de classificação discriminante utilizam informações de múltiplas classes durante o treinamento para construir fronteiras de decisão capazes de maximizar a separação entre os diferentes grupos de amostras (Ferreira, 2015; Mutz *et al.*, 2022; Olivieri, 2018).

A PLS-DA pertence à categoria dos métodos de classificação discriminante e representa uma adaptação da regressão por Quadrados Mínimos Parciais (*Partial Least Squares* – PLS) para problemas de classificação categórica. Enquanto o modelo PLS convencional estabelece relações entre uma matriz de variáveis preditoras (X) e uma variável resposta contínua (Y), a PLS-DA substitui essa resposta por uma matriz categórica, geralmente codificada por variáveis

binárias, representando a classe à qual cada amostra pertence. O algoritmo projeta simultaneamente as matrizes X e Y em um novo espaço de variáveis latentes (*Latent Variables* – LVs), calculadas de modo a maximizar a covariância entre ambas as matrizes. Dessa forma, obtêm-se componentes capazes de concentrar a variabilidade dos dados espectrais mais relevante para a discriminação entre as classes (Ferreira, 2015; Olivieri, 2018).

O desempenho do modelo depende diretamente da escolha adequada do número de variáveis latentes, normalmente determinada por procedimentos de validação cruzada, buscando minimizar o erro de classificação e evitar sobreajuste (*overfitting*). Após a construção do modelo, novas amostras são projetadas no espaço das variáveis latentes e classificadas de acordo com regras de decisão previamente estabelecidas, possibilitando a atribuição objetiva de cada amostra à classe de maior probabilidade (Mutz *et al.*, 2022; Olivieri, 2018).

Em virtude dessas características, a PLS-DA consolidou-se como uma das principais ferramentas quimiométricas para autenticação e controle de qualidade de alimentos. Diversos estudos demonstram sua aplicação na identificação de adulterações em diferentes matrizes alimentícias, incluindo páprica, leite, cerveja, azeites e bebidas fermentadas, apresentando elevado desempenho na discriminação entre produtos autênticos e fraudulentos (Barbosa *et al.*, 2020; Lelis *et al.*, 2022; Mutz *et al.*, 2022). No caso específico dos vinagres, Consonni *et al.* (2008) empregaram a PLS-DA para caracterizar o Vinagre Balsâmico Tradicional de Modena, enquanto Peng *et al.* (2019) utilizaram a técnica para detectar adulterações em vinagres envelhecidos de Shanxi, evidenciando o potencial desse algoritmo para exploração de dados espectrais complexos e autenticação de produtos de elevado valor agregado.

3.8. INTELIGÊNCIA ARTIFICIAL E *MACHINE LEARNING* NA CIÊNCIA DE ALIMENTOS

Nas últimas décadas, o avanço da instrumentação analítica e das tecnologias de aquisição de dados promoveu uma mudança significativa na forma como informações químicas são exploradas na ciência de alimentos. Nesse cenário, a Inteligência Artificial (IA) e, em particular, o Aprendizado de Máquina (*Machine Learning*), consolidaram-se como ferramentas fundamentais para extração de conhecimento, reconhecimento de padrões e tomada de decisão baseada em dados (Kato *et al.*, 2019; Grassi & Alamprese, 2024).

A Inteligência Artificial compreende um conjunto de métodos computacionais desenvolvidos para executar tarefas que tradicionalmente requerem inteligência humana, incluindo reconhecimento de padrões, aprendizagem, raciocínio e tomada de decisão. Dentre

suas diversas áreas, o *Machine Learning* representa uma das abordagens mais difundidas na química analítica e na ciência de alimentos, caracterizando-se pelo desenvolvimento de algoritmos capazes de aprender relações existentes nos dados a partir de exemplos, sem a necessidade de regras explicitamente programadas para cada situação (Kato *et al.*, 2019).

Em aplicações voltadas à autenticação de alimentos, os algoritmos de aprendizado de máquina apresentam vantagens importantes em relação aos métodos estatísticos tradicionais, especialmente quando empregados em conjuntos de dados complexos, altamente correlacionados e compostos por grande número de variáveis. Esses modelos são capazes de identificar relações não lineares, explorar interações multivariadas e reconhecer padrões sutis frequentemente imperceptíveis por abordagens univariadas ou por inspeção visual dos dados, proporcionando elevado desempenho na classificação de alimentos, detecção de adulterações, predição de propriedades físico-químicas e avaliação de qualidade (Da Silva Pereira *et al.*, 2024b; Grassi *et al.*, 2024).

No contexto da ciência de alimentos, diferentes abordagens de modelagem multivariada e aprendizado de máquina têm sido empregadas para classificação, autenticação e predição de propriedades. Entre os métodos supervisionados destacam-se a Análise Discriminante por Quadrados Mínimos Parciais (PLS-DA), as Máquinas de Vetores de Suporte (*Support Vector Machines* – SVM), Random Forest, métodos baseados em *gradient boosting* e Redes Neurais Artificiais (*Artificial Neural Networks* – ANN). Em análises exploratórias, abordagens não supervisionadas, como a Análise de Componentes Principais (PCA) e métodos de agrupamento, são frequentemente utilizadas para investigar padrões, similaridades e estruturas intrínsecas dos dados (Kato *et al.*, 2019).

Nos últimos anos, essas abordagens passaram a desempenhar papel central em estudos de autenticação de alimentos. Nesses estudos, a combinação entre técnicas instrumentais e modelos computacionais tem permitido alcançar elevados índices de acurácia, sensibilidade e especificidade, frequentemente superando métodos estatísticos convencionais na discriminação entre produtos autênticos e adulterados (Grassi *et al.*, 2024; Grassi & Alamprese, 2024).

No caso específico dos vinagres, técnicas de *chemical fingerprinting* associadas à análise multivariada têm sido aplicadas à classificação e autenticação desses produtos. Avanzi Barbosa Mascareli *et al.* (2023), por exemplo, avaliaram espectros de ^1H NMR, FT-IR, NIR e UV-Vis associados a ComDim e PLS-DA para classificar vinagres segundo a matéria-prima, o tempo de envelhecimento e o sistema de cultivo, obtendo os melhores desempenhos com ^1H NMR/PLS-DA. Grassi e Alamprese (2024) empregaram UV-Vis, FT-IR, FT-NIR e NIR portátil associados à análise discriminante linear (LDA) para autenticação de vinagres de vinho

branco adulterados com vinagre de álcool. Mais recentemente, Grassi *et al.* (2024) demonstraram a aplicação da RMN de baixo campo associada à LDA na discriminação entre vinagres de vinho branco autênticos e adulterados com 5–25% (v/v) de vinagre de álcool. Esses estudos evidenciam o potencial de abordagens espectrais não direcionadas associadas à quimiometria para classificação e detecção de fraudes em vinagres.

Embora esses modelos frequentemente apresentem excelente desempenho preditivo, muitos algoritmos modernos são considerados modelos de baixa interpretabilidade, dificultando a compreensão dos mecanismos responsáveis pelas decisões de classificação. Essa limitação representa um desafio importante em aplicações científicas e regulatórias, nas quais não basta apenas obter elevada acurácia; é igualmente necessário compreender quais informações químicas sustentam as decisões do modelo. Como consequência, observa-se um crescente interesse pelo desenvolvimento de métodos de Inteligência Artificial Explicável (*eXplainable Artificial Intelligence* – XAI), capazes de fornecer interpretações quantitativas sobre a contribuição das variáveis para as predições realizadas pelos algoritmos (Da Silva Pereira *et al.*, 2024b; Grassi *et al.*, 2024).

3.8.1. *eXplainable Artificial Intelligence* (XAI)

A crescente incorporação de algoritmos de Inteligência Artificial (IA) e Aprendizado de Máquina (*Machine Learning* – ML) em aplicações científicas tem impulsionado avanços significativos na capacidade de análise e interpretação de conjuntos de dados complexos. Entretanto, muitos dos modelos de maior desempenho preditivo, especialmente aqueles baseados em técnicas de aprendizado profundo (*Deep Learning*) e algoritmos não lineares, apresentam reduzida interpretabilidade, sendo frequentemente classificados como modelos *black-box*. Nesses casos, embora o desempenho de classificação ou predição seja elevado, os mecanismos responsáveis pelas decisões do modelo permanecem pouco compreendidos, dificultando sua validação científica e limitando sua adoção em aplicações regulatórias e industriais (Arrieta *et al.*, 2020).

Nesse contexto, surgiu o conceito de Inteligência Artificial Explicável (*eXplainable Artificial Intelligence* – XAI), definido como o conjunto de métodos capazes de tornar transparentes as decisões produzidas por modelos de IA fornecendo explicações compreensíveis sobre como as variáveis de entrada influenciam os resultados obtidos. Diferentemente dos modelos convencionais de aprendizado de máquina, cujo foco principal é a maximização do desempenho preditivo, as abordagens de XAI buscam conciliar elevada

capacidade de classificação com interpretabilidade, permitindo compreender quais características dos dados fundamentam cada decisão do algoritmo (Arrieta *et al.*, 2020).

Na ciência de alimentos, a aplicação de XAI vem adquirindo importância crescente em razão da elevada complexidade dos dados gerados por técnicas analíticas modernas, como espectroscopia, cromatografia, espectrometria de massas e ressonância magnética nuclear. Essas técnicas produzem grandes volumes de informações químicas que, quando analisadas por algoritmos de aprendizado de máquina, podem resultar em modelos altamente precisos, porém de difícil interpretação. A incorporação de métodos de XAI permite superar essa limitação ao identificar quais variáveis, regiões espectrais ou atributos químicos exercem maior influência sobre as decisões do modelo, tornando os resultados mais transparentes, interpretáveis e cientificamente fundamentados (Ahmed *et al.*, 2025; Arrieta *et al.*, 2020; Arrighi *et al.*, 2026; Da Silva Pereira *et al.*, 2024; Ribeiro *et al.*, 2026).

Essa característica apresenta especial relevância em aplicações relacionadas ao controle de qualidade, segurança dos alimentos e detecção de fraudes. Em estudos de autenticação, por exemplo, modelos explicáveis possibilitam identificar desvios sutis nos perfis físico-químicos e espectrais associados a adulterações, substituições de matérias-primas ou alterações no processamento industrial. Além de aumentar a confiabilidade dos modelos preditivos perante órgãos reguladores e a indústria, essa capacidade permite estabelecer relações diretas entre os resultados computacionais e os mecanismos químicos responsáveis pela diferenciação das amostras, fortalecendo a robustez científica das metodologias desenvolvidas (Arrieta *et al.*, 2020; Da Silva Pereira *et al.*, 2024).

Além das aplicações em autenticação, a Inteligência Artificial Explicável tem contribuído para o monitoramento e otimização de processos tecnológicos na indústria de alimentos. A interpretação das variáveis mais relevantes para o desempenho dos modelos possibilita identificar parâmetros críticos de processamento, como tempo, temperatura, pressão e composição das matérias-primas, favorecendo ajustes operacionais, redução de desperdícios, aumento da eficiência produtiva e melhoria da qualidade dos produtos finais (Ahmed *et al.*, 2025; Arrieta *et al.*, 2020; Arrighi *et al.*, 2026; Da Silva Pereira *et al.*, 2024; Ribeiro *et al.*, 2026).

Outra área de crescente interesse corresponde ao desenvolvimento de alimentos funcionais e personalizados. Nesse contexto, métodos de XAI têm sido empregados na interpretação de dados provenientes de abordagens ômicas, incluindo genômica, metabolômica e proteômica, permitindo identificar biomarcadores, metabólitos e compostos bioativos associados às propriedades nutricionais e funcionais dos alimentos. Essa capacidade amplia a

compreensão dos mecanismos biológicos envolvidos, favorecendo o desenvolvimento de novos produtos e contribuindo para decisões mais fundamentadas em pesquisa e desenvolvimento (Baldrich *et al.*, 2021).

Em aplicações espectroscópicas, a utilização de XAI assume importância ainda maior devido ao elevado número de variáveis que compõem os espectros analíticos. Diferentemente das abordagens tradicionais, que frequentemente utilizam apenas métricas globais de desempenho, as técnicas de explicabilidade permitem quantificar a contribuição individual de cada variável espectral para o processo de classificação. Como consequência, torna-se possível identificar regiões específicas do espectro, sinais característicos ou grupos de metabólitos responsáveis pela diferenciação entre classes de alimentos, estabelecendo uma conexão direta entre o comportamento matemático dos algoritmos e a interpretação química das amostras (Ahmed *et al.*, 2025; Arrighi *et al.*, 2026; Ribeiro *et al.*, 2026).

Essa integração entre espectroscopia, quimiometria, *machine learning* e Inteligência Artificial Explicável representa uma evolução natural das abordagens multivariadas clássicas empregadas na química analítica. Ao combinar elevado desempenho preditivo com transparência e interpretabilidade, os métodos de XAI aumentam a confiabilidade dos modelos, favorecem sua aceitação em aplicações regulatórias e ampliam o potencial de utilização da Inteligência Artificial em controle de qualidade, autenticação de alimentos e detecção de fraudes (Arrieta *et al.*, 2020; Baldrich *et al.*, 2021; Da Silva Pereira *et al.*, 2024).]

Nesse contexto, a autenticação de vinagres constitui um cenário particularmente relevante para a integração dessas abordagens, considerando a complexidade composicional da matriz, a diversidade de matérias-primas e processos produtivos e a possibilidade de adulterações economicamente motivadas. Embora técnicas espectroscópicas associadas à quimiometria tenham demonstrado elevado potencial para a caracterização e classificação desses produtos, ainda há espaço para abordagens que conciliem instrumentação acessível, análise rápida, modelagem robusta e interpretação química das decisões dos modelos. Assim, a associação da RMN de baixo campo ao *fingerprinting* espectral, a métodos quimiométricos, *machine learning* e a ferramentas de XAI apresenta-se como uma estratégia promissora para o desenvolvimento de métodos de autenticação mais rápidos, robustos e interpretáveis. É nesse cenário que se insere o presente trabalho, buscando não apenas discriminar e detectar adulterações em vinagres de maçã, mas também compreender quais informações espectrais sustentam as decisões dos modelos desenvolvidos.

4. REFERÊNCIAS

AHMED, M. T.; AHMED, M. W.; KAMRUZZAMAN, M. A systematic review of explainable artificial intelligence for spectroscopic agricultural quality assessment. *Computers and Electronics in Agriculture*, v. 235, p. 110354, 2025. DOI: <https://doi.org/10.1016/j.compag.2025.110354>.

ANDREATTA, T.; COSTA, N. L.; SANTOS, I. F. dos; BINELLO, L. A Operação “Leite Compensado” e as percepções dos consumidores de leite no município de Panambi/RS. *Nucleus*, v. 16, n. 1, p. 45–56, 2019. DOI: <https://doi.org/10.3738/1982.2278.2884>.

ARRIETA, A. B. *et al.* Explainable artificial intelligence (XAI): concepts, taxonomies, opportunities and challenges toward responsible AI. *Information Fusion*, v. 58, p. 82–115, 2020.

ARRIGHI, L.; DE MORAES, I. A.; ZULLICH, M.; SIMONATO, M.; BARBIN, D. F.; BARBON JUNIOR, S. Explainable artificial intelligence techniques for interpretation of food models: a review. *Artificial Intelligence Review*, v. 59, n. 8, art. 176, 2026. DOI: <https://doi.org/10.1007/s10462-026-11553-2>.

ARRIGHI, L.; PENNELLA, L.; TAVARES, G. M.; BARBON JUNIOR, S. Decision Predicate Graphs: enhancing interpretability in tree ensembles. *arXiv*, 2024. Disponível em: <http://arxiv.org/abs/2404.02942>.

AVANZI BARBOSA MASCARELI, V. A. B.; GALVAN, D.; CRAVEIRO DE ANDRADE, J.; LELIS, C.; CONTE-JUNIOR, C. A.; LOPES, G. M. G.; MACEDO JÚNIOR, F. C. de; SPINOSA, W. A. Spectralprint techniques coupled with chemometric tools for vinegar classifications. *Food Chemistry*, v. 410, p. 135373, 2023. DOI: <https://doi.org/10.1016/j.foodchem.2022.135373>.

BALDRICH, E. *et al.* Artificial intelligence in food science: applications, recent advances, and future prospects. *Trends in Food Science & Technology*, v. 113, p. 109–127, 2021.

BARBOSA, S.; SAURINA, J.; PUIGNOU, L.; NÚÑEZ, O. Classification and authentication of paprika by UHPLC-HRMS fingerprinting and multivariate calibration methods (PCA and PLS-DA). *Foods*, v. 9, n. 4, p. 486, 2020. DOI: <https://doi.org/10.3390/foods9040486>.

BARROS NETO, B. de; SCARMINIO, I. S.; BRUNS, R. E. 25 anos de quimiometria no Brasil. *Química Nova*, v. 29, n. 6, p. 1401–1406, 2006. DOI: <https://doi.org/10.1590/S0100-40422006000600042>.

BEKATOROU, A. *Advances in vinegar production.*: CRC Press, 2019. DOI: <https://doi.org/10.1201/9781351208475>.

BOFFO, E. F.; TAVARES, L. A.; FERREIRA, M. M. C.; FERREIRA, A. G. Classification of Brazilian vinegars according to their ^1H NMR spectra by pattern recognition analysis. *LWT – Food Science and Technology*, v. 42, n. 9, p. 1455–1460, 2009. DOI: <https://doi.org/10.1016/j.lwt.2009.05.008>.

BOUATENIN, K. M. J.-P.; KOUAMÉ, K. A.; GUEU-KEHI, M. E.; DJÉNI, N. T.; DJÈ, K. M. Organic production of vinegar from mango and papaya. *Food Science & Nutrition*, v. 9, n. 1, p. 190–196, 2021. DOI: <https://doi.org/10.1002/fsn3.1981>.

BRASIL. Ministério da Agricultura, Pecuária e Abastecimento. Instrução Normativa nº 17, de 28 de maio de 2009. *Diário Oficial da União*: Brasília, DF, 2009.

BRASIL. Ministério da Agricultura, Pecuária e Abastecimento. Instrução Normativa nº 16, de 19 de março de 2020. *Diário Oficial da União*: Brasília, DF, 2020.

BUTLER, F.; LUIJCKX, N. L.; MARVIN, H. J. P.; BOUZEMBRAK, Y.; MOJTAHED, V. Role of analytical testing for food fraud risk mitigation: a commentary on implementation of analytical fraud testing. *Current Research in Food Science*, v. 4, p. 301–307, 2021. DOI: <https://doi.org/10.1016/j.crfs.2021.03.013>.

CALLAO, M. P.; RUISÁNCHEZ, I. An overview of multivariate qualitative methods for food fraud detection. *Food Control*, v. 86, p. 283–293, 2018. DOI: <https://doi.org/10.1016/j.foodcont.2017.11.034>.

CAVDAROGLU, C.; OZEN, B. Detection of vinegar adulteration with spirit vinegar and acetic acid using UV–visible and Fourier transform infrared spectroscopy. *Food Chemistry*, v. 379, p. 132150, 2022. DOI: <https://doi.org/10.1016/j.foodchem.2022.132150>.

CONSONNI, R.; CAGLIANI, L. R.; BENEVELLI, F.; SPRAUL, M.; HUMPFER, E.; STOCCHERO, M. NMR and chemometric methods: a powerful combination for characterization of Balsamic and Traditional Balsamic Vinegar of Modena. *Analytica Chimica Acta*, v. 611, n. 1, p. 31–40, 2008a. DOI: <https://doi.org/10.1016/j.aca.2008.01.065>.

CONSONNI, R.; CAGLIANI, L. R.; RINALDINI, S.; INCERTI, A. Analytical method for authentication of Traditional Balsamic Vinegar of Modena. *Talanta*, v. 75, n. 3, p. 765–769, 2008b. DOI: <https://doi.org/10.1016/j.talanta.2007.12.005>.

DA SILVA PEREIRA, E.; CRUZ-TIRADO, J. P.; LOURENÇO CRIPPA, B.; MARTINS MORASI, R.; MILAGRES DE ALMEIDA, J.; FERNANDES BARBIN, D.; BARBON JUNIOR, S.; CRISTINA CIRONE SILVA, N. Portable near infrared (NIR) spectrometer coupled with machine learning to classify milk with subclinical mastitis. *Food Control*, v. 163, p. 110527, 2024. DOI: <https://doi.org/10.1016/j.foodcont.2024.110527>.

DE AGUIAR, L. M.; GALVAN, D.; BONA, E.; COLNAGO, L. A.; KILLNER, M. H. M. Data fusion of middle-resolution NMR spectroscopy and low-field relaxometry using the Common Dimensions Analysis (ComDim) to monitor diesel fuel adulteration. *Talanta*, v. 236, p. 122838, 2022. DOI: <https://doi.org/10.1016/j.talanta.2021.122838>.

EL ABDALI, Y. *et al.* Exploring the bioactive compounds in some apple vinegar samples and their biological activities. *Plants*, v. 12, n. 22, p. 3850, 2023. DOI: <https://doi.org/10.3390/plants12223850>.

FERREIRA, M. M. C. *Quimiometria: conceitos, métodos e aplicações.*: Editora da Unicamp, 2015. DOI: <https://doi.org/10.7476/9788526814714>.

GALVAN, D. *Modelagem da reação de transesterificação utilizando catalisador inorgânico.* 2019.

GALVAN, D.; TANAMATI, A. A. C.; CASANOVA, F.; DANIELI, E.; BONA, E.; KILLNER, M. H. M. Compact low-field NMR spectroscopy and chemometrics applied to the analysis of edible oils. *Food Chemistry*, v. 365, p. 130476, 2021. DOI: <https://doi.org/10.1016/j.foodchem.2021.130476>.

GAO, B.; HOLROYD, S. E.; MOORE, J. C.; LAURVICK, K.; GENDEL, S. M.; XIE, Z. Opportunities and challenges using non-targeted methods for food fraud detection. *Journal of Agricultural and Food Chemistry*, v. 67, n. 31, p. 8425–8430, 2019. DOI: <https://doi.org/10.1021/acs.jafc.9b03085>.

GELPÍ, E. *et al.* The Spanish toxic oil syndrome 20 years after its onset: a multidisciplinary review of scientific knowledge. *Environmental Health Perspectives*, v. 110, n. 5, p. 457–464, 2002. DOI: <https://doi.org/10.1289/ehp.110-1240833>.

GOSSNER, C. M.-E.; SCHLUNDT, J.; BEN EMBAREK, P.; HIRD, S.; LO-FO-WONG, D.; BELTRAN, J. J. O.; TEOH, K. N.; TRITSCHER, A. The melamine incident: implications for international food and feed safety. *Environmental Health Perspectives*, v. 117, n. 12, p. 1803–1808, 2009. DOI: <https://doi.org/10.1289/ehp.0900949>.

GOUILLEUX, B.; CHARRIER, B.; AKOKA, S.; FELPIN, F.-X.; RODRIGUEZ-ZUBIRI, M.; GIRAudeau, P. Ultrafast 2D NMR on a benchtop spectrometer: applications and perspectives. *TrAC Trends in Analytical Chemistry*, v. 83, p. 65–75, 2016. DOI: <https://doi.org/10.1016/j.trac.2016.01.014>.

GRASSI, S.; ALAMPRESE, C. Spectroscopic non-targeted techniques in combination with linear discriminant analysis for wine vinegar authentication. *Food and Bioprocess Technology*, v. 17, n. 2, p. 479–488, 2024. DOI: <https://doi.org/10.1007/s11947-023-03143-9>.

GRASSI, S.; BORGONOVO, G.; GENNARO, M.; ALAMPRESE, C. NMR-based approach

to detect white wine vinegar fraud. *Food Chemistry*, v. 456, p. 139953, 2024. DOI: <https://doi.org/10.1016/j.foodchem.2024.139953>.

GROOTVELD, M.; PERCIVAL, B.; GIBSON, M.; OSMAN, Y.; EDGAR, M.; MOLINARI, M.; MATHER, M. L.; CASANOVA, F.; WILSON, P. B. Progress in low-field benchtop NMR spectroscopy in chemical and biochemical analysis. *Analytica Chimica Acta*, v. 1067, p. 11–30, 2019. DOI: <https://doi.org/10.1016/j.aca.2019.02.026>.

GUTHRIE, R. D. Introduction to Spectroscopy (Pavia, Donald; Lampman, Gary M.; Kriz, George S., Jr.). *Journal of Chemical Education*, v. 56, n. 10, p. A323, 2014. DOI: <https://doi.org/10.1021/ed056pA323.2>.

HO, C. W.; LAZIM, A. M.; FAZRY, S.; ZAKI, U. K. H. H.; LIM, S. J. Varieties, production, composition and health benefits of vinegars: a review. *Food Chemistry*, v. 221, p. 1621–1630, 2017. DOI: <https://doi.org/10.1016/j.foodchem.2016.10.128>.

HONG, E.; LEE, S. Y.; JEONG, J. Y.; PARK, J. M.; KIM, B. H.; KWON, K.; CHUN, H. S. Modern analytical methods for the detection of food fraud and adulteration by food category. *Journal of the Science of Food and Agriculture*, v. 97, n. 12, p. 3877–3896, 2017. DOI: <https://doi.org/10.1002/jsfa.8364>.

HUTCHINSON, U. F.; JOLLY, N. P.; CHIDI, B. S.; NGONGANG, M. M.; NTWAMPE, S. K. O. Vinegar engineering: a bioprocess perspective. *Food Engineering Reviews*, v. 11, n. 4, p. 290–305, 2019. DOI: <https://doi.org/10.1007/s12393-019-09196-x>.

JAIKANG, J.; RUNGSARDTHONG, V.; THUMTHANARUK, B.; PUTTANLEK, C.; UTTAPAP, D.; PLATHANAPORN, V.; VATANYOOPAISARN, S. Production of vinegar from organic broken rice noodles. *IOP Conference Series: Earth and Environmental Science*, v. 346, n. 1, p. 012045, 2019. DOI: <https://doi.org/10.1088/1755-1315/346/1/012045>.

JAMIN, E.; THOMAS, F. SNIF-NMR applications in an economic context: fraud detection in food products. In: WEBB, G. A. (ed.). *Modern magnetic resonance*. Cham: Springer International Publishing, 2018. p. 1405–1416. DOI: https://doi.org/10.1007/978-3-319-28388-3_103.

KAHRAMAN, H. A.; TUTUN, H.; KAYA, M. M.; TUTUN, S.; USLUER, M. S.; RUGJİ, J.; YURDAKUL, O. Total phenolic content, antiradical, antimicrobial and antibiofilm properties of grape and apple vinegar. *Journal of Advances in VetBio Science and Techniques*, v. 6, n. 2, p. 150–158, 2021. DOI: <https://doi.org/10.31797/vetbio.960155>.

KARA, M.; ASSOUGUEM, A.; AL KAMALY, O. M.; BENMESSAOUD, S.; IMTARA, H.; MECHCHATE, H.; HANO, C.; ZERHOUNI, A. R.; BAHHOU, J. The impact of apple variety and the production methods on the antibacterial activity of vinegar samples. *Molecules*, v. 26, n. 18, p. 5437, 2021. DOI: <https://doi.org/10.3390/molecules26185437>.

KATO, T.; MASTELINI, S. M.; CAMPOS, G. F. C.; BARBON, A. P. A. da C.; PRUDENCIO, S. H.; SHIMOKOMAKI, M.; SOARES, A. L.; BARBON, S. White striping degree assessment using computer vision system and consumer acceptance test. *Asian-Australasian Journal of Animal Sciences*, v. 32, n. 7, p. 1015–1026, 2019. DOI: <https://doi.org/10.5713/ajas.18.0504>.

KAZAZIĆ, S.; GAJDOŠ-KLJUSURIĆ, J.; RADELJEVIĆ, B.; PLAVLJANIĆ, D.; ŠPOLJARIĆ, J.; LJUBIĆ, T.; BILIĆ, B.; MIKULEC, N. Comparison of GC and NIR spectra as a rapid tool for food fraud detection: case of butter adulteration with different fat types. *Journal of Food Processing and Preservation*, v. 45, n. 9, 2021. DOI: <https://doi.org/10.1111/jfpp.15732>.

KUNAL AHUJA; HIMANSHU KABRA. Vinegar market size: by product type, source, flavor, distribution channel, application, growth prospects, regional outlook and forecast. 2022.

KUSHWAH, S.; DHIR, A.; SAGAR, M.; GUPTA, B. Determinants of organic food consumption: a systematic literature review on motives and barriers. *Appetite*, v. 143, p. 104402, 2019. DOI: <https://doi.org/10.1016/j.appet.2019.104402>.

LAVINE, B.; WORKMAN, J. Chemometrics. *Analytical Chemistry*, v. 80, n. 12, p. 4519–4531, 2008. DOI: <https://doi.org/10.1021/ac800728t>.

LELIS, C. A.; GALVAN, D.; TESSARO, L.; DE ANDRADE, J. C.; MUTZ, Y. S.; CONTE-JUNIOR, C. A. Fluorescence spectroscopy in tandem with chemometric tools applied to milk quality control. *Journal of Food Composition and Analysis*, v. 109, p. 104515, 2022. DOI: <https://doi.org/10.1016/j.jfca.2022.104515>.

LI, H.; MING, X.; LIU, Z.; XU, L.; XU, D.; HU, L.; MO, H.; ZHOU, X. Accelerating vinegar aging by combination of ultrasonic and magnetic field assistance. *Ultrasonics Sonochemistry*, v. 78, p. 105708, 2021. DOI: <https://doi.org/10.1016/j.ultsonch.2021.105708>.

LIU, Q.; TANG, G.-Y.; ZHAO, C.-N.; GAN, R.-Y.; LI, H.-B. Antioxidant activities, phenolic profiles, and organic acid contents of fruit vinegars. *Antioxidants*, v. 8, n. 4, p. 78, 2019. DOI: <https://doi.org/10.3390/antiox8040078>.

MAGALHÃES DE AGUIAR, L.; BONA, E.; COLNAGO, L. A.; ROHWEDDER, J. J. R.; KILLNER, M. H. M. Application of low-field and medium-resolution ¹H NMR spectroscopy combined with chemometric methods for automotive gasoline quality control. *Fuel*, v. 282, p. 118684, 2020. DOI: <https://doi.org/10.1016/j.fuel.2020.118684>.

MORGAN, J.; MOSAWY, S. The potential of apple cider vinegar in the management of type

2 diabetes. *International Journal of Diabetes Research*, v. 5, n. 6, p. 129–134, 2016. DOI: <https://doi.org/10.5923/j.diabetes.20160506.02>.

MUTZ, Y. S.; DO ROSARIO, D.; SILVA, L. R. G.; GALVAN, D.; JANEGITZ, B. C.; FERREIRA, R. de Q.; CONTE-JUNIOR, C. A. A single screen-printed electrode in tandem with chemometric tools for the forensic differentiation of Brazilian beers. *Scientific Reports*, v. 12, n. 1, p. 5630, 2022. DOI: <https://doi.org/10.1038/s41598-022-09632-9>.

NOBARI MOGHADDAM, H.; TAMIJI, Z.; AKBARI LAKEH, M.; KHOSHAYAND, M. R.; HAJI MAHMOODI, M. Multivariate analysis of food fraud: a review of NIR-based instruments in tandem with chemometrics. *Journal of Food Composition and Analysis*, v. 107, p. 104343, 2022. DOI: <https://doi.org/10.1016/j.jfca.2021.104343>.

OLIVIERI, A. C. The inverse least-squares model. In: OLIVIERI, A. C. *Practical three-way calibration*. [S. l.: s. n.], 2018. DOI: https://doi.org/10.1007/978-3-319-97097-4_3.

OUSAAID, D.; LAAROUSSI, H.; BAKOUR, M.; ELGHOUZI, A.; ABOULGHAZI, A.; LYOUSSE, B.; ELARABI, I. Beneficial effects of apple vinegar on hyperglycemia and hyperlipidemia in hypercaloric-fed rats. *Journal of Diabetes Research*, v. 2020, p. 1–7, 2020. DOI: <https://doi.org/10.1155/2020/9284987>.

OUSAAID, D.; LAAROUSSI, H.; BAKOUR, M.; ENNAJI, H.; LYOUSSE, B.; EL ARABI, I. Antifungal and antibacterial activities of apple vinegar of different cultivars. *International Journal of Microbiology*, v. 2021, p. 1–6, 2021. DOI: <https://doi.org/10.1155/2021/6087671>.

PAVIA, D. L. *Introduction to spectroscopy*. 5. ed. Washington: Cengage Learning, 2014.

PENG, T.-Q.; YIN, X.-L.; SUN, W.; DING, B.; MA, L.-A.; GU, H.-W. Developing an excitation-emission matrix fluorescence spectroscopy method coupled with multi-way classification algorithms for the identification of the adulteration of Shanxi aged vinegars. *Food Analytical Methods*, v. 12, n. 10, p. 2306–2313, 2019. DOI: <https://doi.org/10.1007/s12161-019-01586-5>.

PERUMPULI, B.; DILRUKSHI, N. Vinegar: a functional ingredient for human health. *International Food Research Journal*, v. 29, n. 5, p. 959–974, 2022. DOI: <https://doi.org/10.47836/ifrj.29.5.01>.

QUEVEDO-SILVA, F.; FREIRE, O.; SPANHOL-FINOCCHIO, C. P. “Carne Fraca”. *British Food Journal*, v. 122, n. 2, p. 722–735, 2019. DOI: <https://doi.org/10.1108/BFJ-07-2019-0491>.

RIBEIRO, J. V.; GONCALVES, R. F.; MELQUIADES, F. L.; BARBON JUNIOR, S. Spectral Model eXplainer: a chemically-grounded explainability framework for spectral-based machine learning models. *arXiv*, 2026. Disponível em: <http://arxiv.org/abs/2605.02684>.

RÍOS-REINA, R.; CAMIÑA, J. M.; CALLEJÓN, R. M.; AZCARATE, S. M. Spectralprint techniques for wine and vinegar characterization, authentication and quality control: advances and projections. *TrAC – Trends in Analytical Chemistry*, v. 134, p. 116121, 2021. DOI: <https://doi.org/10.1016/j.trac.2020.116121>.

ROBSON, K.; DEAN, M.; HAUGHEY, S.; ELLIOTT, C. A comprehensive review of food fraud terminologies and food fraud mitigation guides. *Food Control*, v. 120, p. 107516, 2021. DOI: <https://doi.org/10.1016/j.foodcont.2020.107516>.

ROMÁN-CAMACHO, J. J.; GARCÍA-GARCÍA, I.; SANTOS-DUEÑAS, I. M.; GARCÍA-MARTÍNEZ, T.; MAURICIO, J. C. Latest trends in industrial vinegar production and the role of acetic acid bacteria: classification, metabolism, and applications—a comprehensive review. *Foods*, v. 12, n. 19, p. 3705, 2023. DOI: <https://doi.org/10.3390/foods12193705>.

ŠTORNIK, A.; SKOK, B.; TRČEK, J. Comparison of cultivable acetic acid bacterial microbiota between organic and conventional apple cider vinegar. *Food Technology and Biotechnology*, v. 54, 2016. DOI: <https://doi.org/10.17113/ftb.54.01.16.4082>.

XIA, T.; ZHANG, B.; DUAN, W.; ZHANG, J.; WANG, M. Nutrients and bioactive components from vinegar: a fermented and functional food. *Journal of Functional Foods*, v. 64, p. 103681, 2020. DOI: <https://doi.org/10.1016/j.jff.2019.103681>.

XIE, H.; BU, L.; PENG, X.; LI, Z. Ultraviolet spectroscopy method for classifying vinegars. *Advanced Materials Research*, v. 346, p. 865–874, 2012. DOI: <https://doi.org/10.4028/www.scientific.net/AMR.346.865>.

XIE, Z.; KOYSOMBOON, C.; ZHANG, H.; LU, Z.; ZHANG, X.; CHEN, F. Vinegar volatile organic compounds: analytical methods, constituents, and formation processes. *Frontiers in Microbiology*, v. 13, 2022. DOI: <https://doi.org/10.3389/fmicb.2022.907883>.

ZHANG, X.-L.; ZHENG, Y.; XIA, M.-L.; WU, Y.-N.; LIU, X.-J.; XIE, S.-K.; WU, Y.-F.; WANG, M. Knowledge domain and emerging trends in vinegar research: a bibliometric review of the literature from WoSCC. *Foods*, v. 9, n. 2, p. 166, 2020. DOI: <https://doi.org/10.3390/foods9020166>.

CAPÍTULO II

ARTIGO CIENTÍFICO 1

Detecção de fraudes em vinagre de maçã brasileiro utilizando *fingerprinting* por RMN de baixo campo e *machine learning*

**Deteção de fraudes em vinagre de maçã brasileiro utilizando *fingerprinting* por RMN
de baixo campo e machine learning**

Vinícius Avanzi Barbosa Avanzi Barbosa Mascareli^{a,1}, Diego Galvan^b, Jelmir Craveiro de Andrade^{c,d}, Ingrid Alves de Moraes^e, Viviane Lopes Leite da Costa^a, Carlos Adam Conte-Junior^c, Sylvio Barbon Junior^e, Wilma Aparecida Spinosa^a

^a Departamento de Ciência e Tecnologia de Alimentos, Universidade Estadual de Londrina (UEL), Londrina, PR 86.057-970, Brasil.

^b Departamento de Química, Universidade Federal de Santa Catarina (UFSC), Florianópolis, SC, 88.040-900, Brasil.

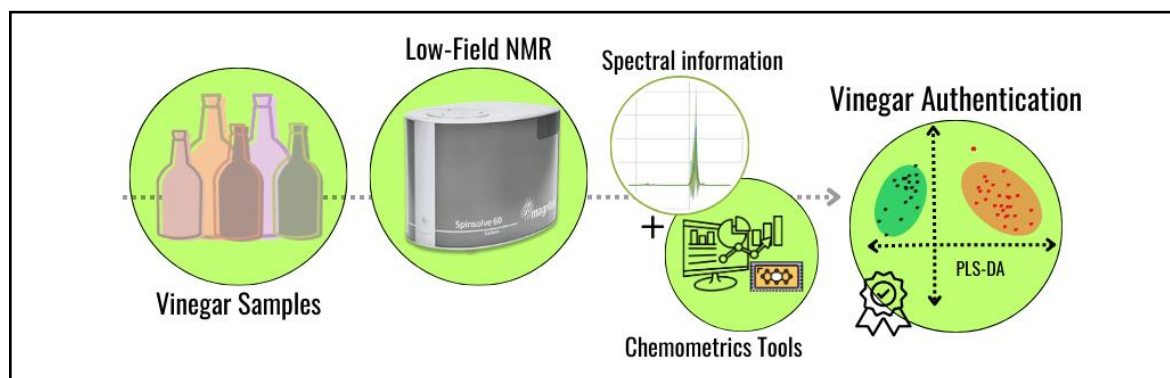
^c Centro Laboratorial Analítico e Molecular (CLAM), Instituto de Química (IQ), Universidade Federal do Rio de Janeiro (UFRJ), Rio de Janeiro, RJ 21.941-909, Brasil.

^d Núcleo de Análise de Alimentos (NAL), Laboratório de Apoio ao Desenvolvimento Tecnológico (LADETEC), Universidade Federal do Rio de Janeiro (UFRJ), Rio de Janeiro, RJ 21.941-598, Brasil.

^e Departamento de Engenharia e Arquitetura, Universidade de Trieste (UNITS), 34127 Trieste, Itália.

¹Autor correspondente: *V. A. B. Avanzi Barbosa Mascareli. E-mail: vini.mascareli@gmail.com ou vini.mascareli@uel.br ORCID: <https://orcid.org/0000-0002-8327-8790>. Departamento de Ciência e Tecnologia de Alimentos, Universidade Estadual de Londrina, Londrina, Brasil.

Resumo Gráfico



Resumo

A autenticação de vinagres é desafiada pela complexidade química das matrizes fermentadas e pelos fortes incentivos econômicos à fraude, particularmente em produtos associados à origem de frutas e à produção orgânica. Neste estudo, foi avaliada uma estratégia não direcionada baseada em *fingerprinting* espectral por ressonância magnética nuclear de hidrogênio de baixo campo (^1H LF-NMR) combinada com modelagem quimiométrica supervisionada para diferenciar vinagres autênticos de amostras adulteradas simuladas, estabelecendo a LF-NMR como uma técnica confiável para autenticação de vinagres. Foi analisado um conjunto de dados composto por 564 amostras, incluindo 139 vinagres comerciais (vinagre de maçã orgânico, vinagre de maçã tradicional e vinagre de álcool) e 425 amostras adulteradas preparadas em concentrações de 5–31% (v/v) utilizando vinagre de maçã tradicional ou vinagre de álcool, abrangendo diferentes cenários de fraude. Os espectros adquiridos sem pré-tratamento na região de 0,0–4,5 ppm foram processados utilizando *pipelines* estruturados de pré-processamento e avaliados sob validação cruzada aninhada. O alinhamento espectral mostrou-se crítico, melhorando significativamente a robustez e a capacidade de generalização em comparação com espectros não corrigidos ou corrigidos por linha de base, especialmente sob variabilidade composicional natural. Os classificadores PLS-DA alcançaram acurácias gerais de

aproximadamente 0,81 e valores de macro-F1 entre 0,79–0,80, sendo as amostras adulteradas identificadas de forma mais confiável (F1 de até 0,85) do que os vinagres de maçã autênticos, os quais apresentaram maior heterogeneidade relacionada à diversidade da matéria-prima e às propriedades fermentativas. Até o momento, este é o primeiro estudo sistemático abordando a autenticação de vinagre de maçã orgânico utilizando ^1H LF-NMR sob variabilidade agrônômica realista. *Fingerprints* de LF-NMR adequadamente pré-processados e alinhados mantiveram poder discriminatório suficiente para triagem confiável, apoiando seu uso como uma ferramenta prática, econômica e escalável, alinhada aos princípios da química verde e ao controle rotineiro da qualidade de alimentos.

Palavras-chave: *spectralprint*; autenticação de alimentos; técnica espectroscópica; análise não direcionada; produto orgânico; discriminação por mínimos quadrados parciais; bebida fermentada; quimiometria.

1. Introdução

O vinagre é um produto fermentado amplamente consumido e historicamente relevante, obtido por meio de fermentações alcoólica e acética sucessivas de matérias-primas ricas em açúcares fermentáveis ou amido, como frutas, cereais, vegetais e mel (Bekatorou, 2019; Perumpuli & Dilrukshi, 2022). Sua composição química é dominada por água e ácido acético, acompanhados por uma ampla variedade de constituintes minoritários, incluindo ácidos orgânicos, aminoácidos, açúcares residuais, compostos fenólicos e produtos derivados de reações relacionadas ao processamento (Ho *et al.*, 2017; Xia *et al.*, 2020). A natureza e a abundância relativa desses componentes são fortemente influenciadas pela matéria-prima, pelas condições de fermentação e pelas etapas tecnológicas envolvidas na produção, moldando, em última instância, os atributos sensoriais, bem como o valor comercial e funcional do produto final (Q. Liu *et al.*, 2019; Perumpuli & Dilrukshi, 2022; Yıkmış *et al.*, 2021).

Nos últimos anos, o mercado de vinagres apresentou crescimento substancial, impulsionado pela expansão de produtos percebidos como naturais, orgânicos e associados à origem ou à saúde. Essa tendência promoveu a diversificação dos tipos de vinagre, aumentando particularmente a relevância de mercado dos vinagres de frutas e dos vinagres produzidos sob alegações de produção orgânica (Kunal Ahuja & Himanshu Kabra, 2022). Nesse contexto, o vinagre de maçã orgânico emergiu como um dos produtos de maior valor agregado, devido ao maior custo de sua matéria-prima em comparação ao vinagre de álcool, ao apelo relacionado à sustentabilidade, ao perfil sensorial diferenciado e à presença de compostos bioativos (El Abdali et al., 2023; Kahraman et al., 2021; Morgan & Mosawy, 2016; Ousaaïd et al., 2020, 2021; Štornik et al., 2016). Esses atributos aumentam a atratividade para o consumidor, mas simultaneamente ampliam a vulnerabilidade a práticas fraudulentas (Cavdaroglu & Ozen, 2023).

As diferenças de preço, especialmente entre vinagres convencionais e aqueles associados a alegações específicas relacionadas à matéria-prima, sistema de produção ou certificação orgânica, criam fortes incentivos econômicos para adulteração e rotulagem enganosa (Fang et al., 2022; Grassi et al., 2024; Hsieh et al., 2013; Nie et al., 2019; Song et al., 2022; Xiong et al., 2023). Tais práticas representam um desafio significativo para a indústria alimentícia, órgãos reguladores e consumidores. Fraudes em vinagres podem ocorrer por diferentes vias, incluindo diluição com vinagres de menor valor comercial, adição não autorizada de açúcares durante a fermentação de vinagres de frutas e declarações enganosas quanto à origem ou métodos de produção. No caso específico do vinagre de maçã, a adição de açúcares ao mosto fermentável é ilegal no Brasil e explicitamente proibida pela legislação vigente, ressaltando a necessidade de abordagens analíticas confiáveis capazes de detectar tais irregularidades (Brasil, Ministério da Agricultura, 2020; Hsieh et al., 2013).

Tradicionalmente, os padrões legais de identidade e qualidade para vinagres baseiam-se

principalmente em parâmetros físico-químicos clássicos, como acidez volátil e teor residual de álcool. Embora essas métricas sejam essenciais para conformidade regulatória, frequentemente são insuficientes para garantir a autenticidade do produto, uma vez que não preservam a identidade química global associada à matéria-prima e ao processo de produção (Avanzi Barbosa Mascareli et al., 2023). Consequentemente, tem crescido o interesse por estratégias analíticas não direcionadas capazes de avaliar o vinagre como um sistema químico complexo, considerando simultaneamente múltiplos constituintes e suas inter-relações (Gao et al., 2019; Grassi & Alamprese, 2024a).

As técnicas espectroscópicas têm ganhado destaque nesse contexto porque permitem a obtenção de perfis químicos globais, ou *fingerprints* espectrais, com mínima preparação de amostra e sem necessidade de quantificação prévia de marcadores específicos (Loffredi et al., 2021; Ríos-Reina et al., 2019; Shi et al., 2025). Entre essas técnicas, a espectroscopia de ressonância magnética nuclear de baixo campo (LF-NMR) apresenta vantagens notáveis, incluindo alta reprodutibilidade, análise não destrutiva e capacidade de fornecer simultaneamente informações estruturais e composicionais. Mais recentemente, a LF-NMR emergiu como uma alternativa promissora à RMN de alto campo, devido aos menores custos operacionais, ausência de necessidade de criogenia e disponibilidade de instrumentos compactos baseados em ímãs permanentes, características que favorecem sua implementação em ambientes industriais e de controle de qualidade de rotina (Conte et al., 2025; Galvan et al., 2021, 2023; Grassi et al., 2024; Grootveld et al., 2019; Tehari et al., 2026).

Apesar da resolução espectral intrinsecamente inferior da LF-NMR, essa limitação não impede sua aplicação em abordagens de *fingerprinting*, particularmente quando combinada com métodos quimiométricos (Galvan et al., 2023). Métodos multivariados, como análise de componentes principais e modelos supervisionados baseados em mínimos quadrados parciais, têm sido amplamente aplicados em estudos de autenticação de alimentos e detecção de fraudes

(Barbosa et al., 2020; Callao & Ruisánchez, 2018; Nobari Moghaddam et al., 2022; Oliveri & Downey, 2012; Tehari et al., 2026). Entretanto, sua aplicação em vinagres, especialmente no contexto da autenticação de vinagre de maçã e vinagre orgânico, ainda permanece relativamente limitada na literatura.

Neste trabalho, foi avaliada a viabilidade da espectroscopia de ressonância magnética nuclear de baixo campo combinada com técnicas supervisionadas de *machine learning* para discriminação e autenticação de vinagres comerciais. O estudo concentra-se na diferenciação entre vinagres autênticos e amostras adulteradas simuladas, estabelecendo os *fingerprints* espectrais de ^1H LF-NMR de baixo campo como uma plataforma analítica não direcionada confiável para autenticação de vinagres. A abordagem proposta integra *pipelines* estruturados de pré-processamento, incluindo alinhamento espectral crítico, com estratégias rigorosas de validação cruzada aninhada, garantindo classificação robusta, objetiva e orientada por dados. Além do desempenho dos modelos, este trabalho fornece a primeira caracterização sistemática da autenticidade de vinagre de maçã orgânico utilizando ^1H LF-NMR sob variabilidade composicional realista, validando a técnica como uma ferramenta prática de triagem alinhada aos princípios da química verde e às demandas operacionais do controle de qualidade rotineiro de alimentos.

2. Material e métodos

2.1. Vinagres autênticos e adulterados

Um total de 139 amostras de vinagres provenientes de diferentes lotes de produção foram incluídos neste estudo. Dentre essas, 46 amostras foram fornecidas pela Tecnologia em Saúde Indústria de Alimentos Ltda. (Assis, São Paulo, Brasil), enquanto as 93 amostras restantes foram adquiridas comercialmente em estabelecimentos varejistas localizados em

diferentes cidades brasileiras: Londrina (Paraná), Assis (São Paulo), Campo Grande (Mato Grosso do Sul) e Florianópolis (Santa Catarina).

Todas as amostras foram transportadas para Londrina, Paraná, e armazenadas na Universidade Estadual de Londrina sob condições controladas, permanecendo seladas em temperatura de $\sim 4^{\circ}\text{C}$ e protegidas da luz até análises posteriores.

O conjunto de dados completo (disponível como material suplementar SM1) compreendeu vinagres produzidos a partir de diferentes tipos de matérias-primas, agrupados em três categorias principais: vinagre de maçã tradicional, vinagre de álcool e vinagre de maçã orgânico. Em relação ao sistema de produção da matéria-prima, 51 amostras corresponderam ao cultivo orgânico, enquanto 88 amostras foram classificadas como não orgânicas.

Quanto às amostras adulteradas, foram considerados dois tipos de simulações de adulteração. O primeiro envolveu a adulteração de vinagres de maçã orgânicos com vinagres de maçã tradicionais, enquanto o segundo consistiu na adulteração de vinagres de maçã com vinagre de álcool.

O primeiro esquema de adulteração (Tipo 1) foi realizado utilizando dez amostras certificadas de vinagre de maçã orgânico provenientes de diferentes lotes (aproximadamente 25% do total de amostras de vinagre de maçã orgânico), adulteradas com cinco vinagres de maçã tradicionais comerciais (aproximadamente 10% do total de amostras de vinagre de maçã tradicional) nos níveis de 5, 6, 30 e 31% (v/v), utilizando micropipetas calibradas. Consequentemente, um total de 200 amostras adulteradas foi gerado nessa categoria, denominada AD ($10 \text{ vinagres de maçã orgânicos} \times 5 \text{ adulterantes} \times 4 \text{ níveis de adulteração}$).

O segundo esquema de adulteração (Tipo 2) foi conduzido utilizando quinze amostras certificadas de vinagre de maçã provenientes de diferentes lotes (aproximadamente 20% do total de amostras de vinagre de maçã), adulteradas com cinco vinagres comerciais de álcool (aproximadamente 10% do total de amostras de vinagre de álcool) nas concentrações de 5, 15

e 25% (v/v), também utilizando micropipetas calibradas. Como resultado, um total de 225 amostras adulteradas foi produzido para esse segundo tipo de adulteração, denominado MA (15 vinagres de maçã $\times$ 5 adulterantes $\times$ 3 níveis de adulteração).

No total, o conjunto final de dados compreendeu 564 amostras (originais + adulteradas), conforme ilustrado no diagrama esquemático apresentado na (Fig. 1).

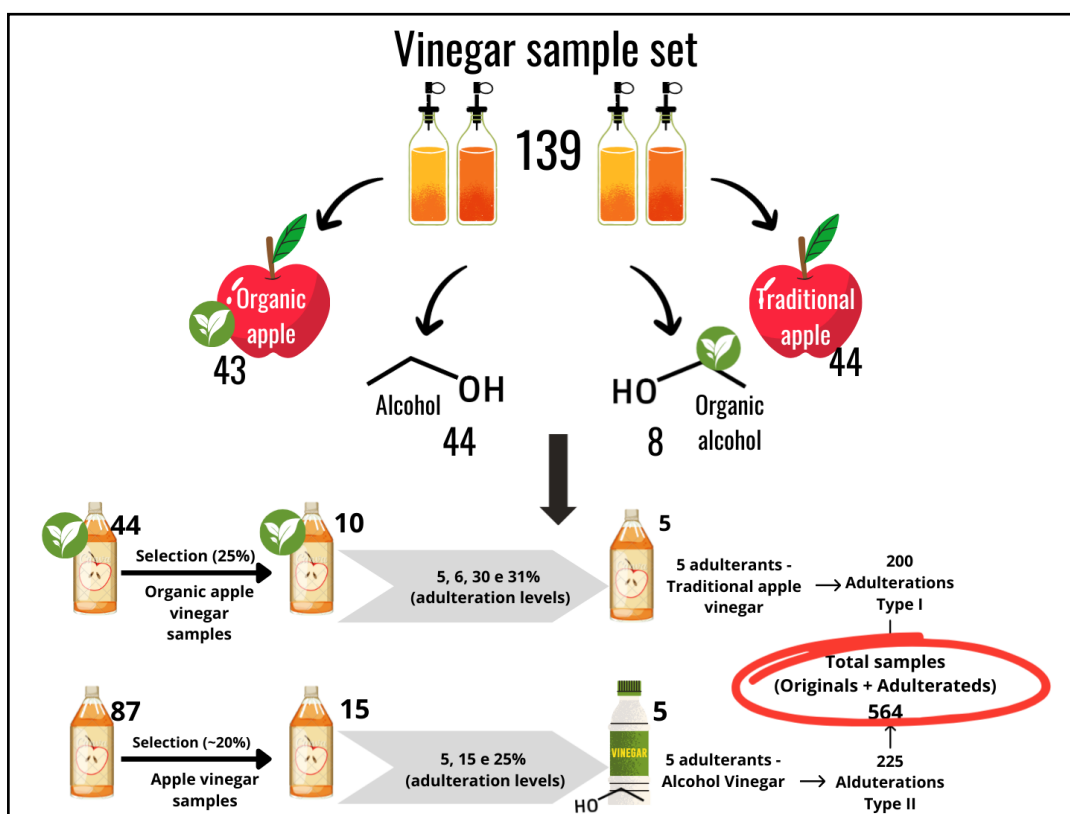


Figura 1. Visão esquemática do conjunto de amostras de vinagre, incluindo amostras originais e adulterações simuladas com vinagre de maçã tradicional e vinagre de álcool.

2.2. Sessão experimental

Este estudo abordou a autenticação de vinagres utilizando *fingerprints* espectrais obtidos por ^1H LF-NMR combinados com abordagens supervisionadas de *machine learning*. Foram investigadas duas tarefas de classificação binária: vinagre de maçã orgânico *versus* vinagre de maçã orgânico adulterado com vinagre de maçã tradicional (AM vs AD), e vinagre

de maçã *versus* vinagre de maçã adulterado com vinagre de álcool (AM vs MA).

Os dados espectrais foram adquiridos a partir de amostras de vinagre pertencentes a essas categorias, conforme descrito na Seção 2.1, compreendendo tanto amostras autênticas quanto adulteradas simuladas. As amostras autênticas originaram-se de diferentes marcas, lotes de produção e sistemas produtivos (orgânico e não orgânico), obtidas tanto por doação direta da indústria quanto adquiridas em mercados varejistas de diferentes regiões brasileiras, garantindo variabilidade representativa dos produtos comerciais.

Para a tarefa AM vs AD, foram consideradas amostras certificadas de vinagre de maçã orgânico e suas respectivas contrapartes adulteradas preparadas com vinagre de maçã tradicional em diferentes níveis volumétricos. Para a tarefa AM vs MA, foram utilizadas amostras de vinagre de maçã e suas respectivas adulterações com vinagre de álcool em diferentes concentrações.

Todas as amostras foram transportadas para a Universidade Estadual de Londrina, mantidas seladas, armazenadas em $\sim 4^{\circ}\text{C}$ e protegidas da luz até a aquisição espectral, a fim de minimizar variabilidades relacionadas ao armazenamento. Os espectros de ^1H LF-NMR foram posteriormente utilizados como variáveis de entrada para modelos supervisionados de classificação, desenvolvidos para discriminar amostras autênticas de adulteradas em cada tarefa de classificação.

2.3. Aquisição de dados de ^1H LF-NMR e exportação espectral

Para a aquisição espectral, 500 μL de cada amostra foram diretamente transferidos para tubos de RMN de 5 mm, sem adição de solventes ou solventes deuterados. Antes da coleta dos dados, as amostras foram equilibradas em estufa termostaticada por 10 min para estabilização da temperatura entre 26,5 e 28,5 $^{\circ}\text{C}$, correspondente à temperatura da sonda.

Os espectros de ^1H LF-NMR foram adquiridos em um espectrômetro de bancada

Spinsolve 60 MHz (Magritek GmbH, Aachen, Alemanha), equipado com magneto permanente e operando a 60 MHz para ^1H , utilizando tubos de RMN de 5 mm. As aquisições foram realizadas entre 26,5 e 28,5 °C, no modo “1D PROTON CDEC”, utilizando pulso de radiofrequência de 90°. Foram coletadas 16 varreduras, com tempo de aquisição de 3,2 s e atraso de repetição de 15 s, resultando em aproximadamente 4 min de análise por amostra. A largura espectral foi ajustada para 328 ppm, com frequência central em 110 ppm.

O processamento espectral foi realizado utilizando o *software* TopSpin (Bruker BioSpin, Alemanha), e os espectros foram posteriormente exportados em formato espectral bruto para análises posteriores dos dados.

2.4. Tratamento e pré-processamento dos dados

Antes da análise do conjunto de dados, a região espectral entre 0,0 e 4,5 ppm foi selecionada por concentrar ressonâncias intensas e quimicamente informativas nas matrizes de vinagre. Esse intervalo é predominantemente caracterizado por sinais de compostos de baixa massa molecular provenientes da matéria-prima e dos processos fermentativos, incluindo acetato/ácido acético, etanol, lactato/ácido láctico, succinato/ácido succínico, citrato/ácido cítrico e malato/ácido málico, além de aminoácidos, como alanina, e carboidratos residuais. Esses constituintes estão presentes em diferentes proporções nas amostras autênticas e adulteradas e contribuem significativamente para o *fingerprint* espectral obtido por LF-NMR. Restringir a análise a essa região aumenta a sensibilidade às diferenças composicionais associadas à dinâmica fermentativa, ao mesmo tempo em que reduz a influência de regiões espectrais menos informativas ou mais suscetíveis a ruído (Avanzi Barbosa Mascareli et al., 2023; Boffo et al., 2009; Caligiani et al., 2007; Grassi et al., 2024).

Os *outliers* foram identificados na sub-banda de 0,0–4,5 ppm utilizando um modelo *Isolation Forest* treinado sobre a matriz espectral (não supervisionado) (Ceschin et al., 2026).

O *Isolation Forest* foi configurado com 300 árvores e *contamination*='auto'. Apenas as amostras previstas como *inliers* foram mantidas para criação dos conjuntos de dados “limpos” utilizados na avaliação dos modelos. Para evitar vazamento de informação, todas as etapas de pré-processamento foram implementadas como transformadores do *scikit-learn* e aplicadas dentro de cada *fold* de treinamento por meio de *pipelines*.

Três estratégias de pré-processamento foram comparadas ao Original (sem correção): (1) transformação identidade (intensidades brutas da sub-banda). (2) Alinhamento (referenciado ao treinamento, inspirado no *icoshift*): os espectros foram alinhados dentro de cada *fold* externo de validação cruzada à média espectral do conjunto de treinamento, aplicando-se uma busca de deslocamento que maximiza a correlação, restrita a ± 10 pontos de dados (isto é, uma janela limitada de deslocamento), evitando vazamento de informação do conjunto teste. O deslocamento que maximizou a similaridade normalizada em relação ao espectro de referência foi aplicado com preenchimento das bordas. (3) Correção de linha de base (ALS): estimativa da linha de base por mínimos quadrados assimétricos aplicada a cada espectro individualmente e subtraída do sinal original. Os parâmetros utilizados foram: $\lambda = 1 \times 10^5$, $p = 0,01$ e *niter* = 10 (Avanzi Barbosa Mascareli et al., 2023; Savorani et al., 2010).

Todas as análises foram realizadas em Python 3.11.2 utilizando bibliotecas científicas padrão (NumPy 1.24.0, pandas 2.2.3, SciPy 1.14.1, matplotlib 3.7.5) e scikit-learn 1.4.2 para modelagem, validação cruzada aninhada com agrupamento e métricas de desempenho.

Após o pré-processamento, todos os pipelines aplicaram StandardScaler (padronização z-score) às variáveis espectrais (ou características transformadas) dentro de cada *fold*.

2.5. Extração supervisionada de características e modelo de classificação

Um modelo supervisionado de variáveis latentes foi implementado utilizando *Partial Least Squares* como extrator de características (PLS-DA) (da Silva Pereira et al., 2024; Moraes

et al., 2025). Os rótulos das classes foram codificados como valores numéricos (0/1) para ajuste do modelo PLS. O modelo PLS-DA ajustado foi então utilizado para transformar os espectros em variáveis latentes (*scores*).

A seleção do modelo concentrou-se no número de variáveis latentes do PLS. Dois procedimentos complementares foram utilizados. Para cada estratégia de pré-processamento, o parâmetro *n_components* variou de 1 a 12, e o desempenho foi estimado utilizando validação cruzada estratificada de 10 *folds*. A principal métrica utilizada foi o F1-score macro-médio (média $\pm$ desvio padrão entre os *folds*). Essa varredura foi empregada para visualizar a relação entre complexidade do modelo e desempenho.

A segunda abordagem utilizou validação cruzada aninhada com agrupamento para obter uma estimativa não enviesada do desempenho preditivo, prevenindo simultaneamente vazamento de informação entre amostras de mesma origem. No loop externo (10 *folds*), as divisões foram criadas agrupando amostras que compartilhavam o mesmo lote-base de vinagre, de forma que todas as misturas derivadas de um determinado vinagre-base fossem mantidas dentro de um único *fold* para avaliação da capacidade de generalização. No loop interno (5 *folds*), os hiperparâmetros foram ajustados sob a mesma restrição de agrupamento, garantindo que a seleção do modelo também respeitasse a estrutura agrupada dos dados.

Dentro de cada divisão externa de treinamento, um *GridSearchCV* otimizou o parâmetro *n_components* entre 1 e 12 variáveis latentes, utilizando macro-F1 como critério de seleção.

Todas as análises foram realizadas em Python utilizando as bibliotecas científicas padrão e versões mencionadas na Seção 2.4.

2.6. Matrizes de confusão e inspeção de erros

Para fornecer previsões *fold-wise* fora da amostra para inspeção qualitativa, uma etapa final de avaliação foi realizada utilizando validação cruzada estratificada de 10 *folds* com loops

explícitos de validação. Para a estratégia de pré-processamento de melhor desempenho e o valor selecionado de $n_components$ (com base na validação cruzada aninhada), as predições de cada *fold* de teste foram agregadas.

Todas as análises foram realizadas em Python utilizando as bibliotecas científicas padrão mencionadas na Seção 2.4.

3. Resultados e Discussão

3.1. *Fingerprints* espectrais de RMN de ^1H de baixo campo de amostras de vinagre

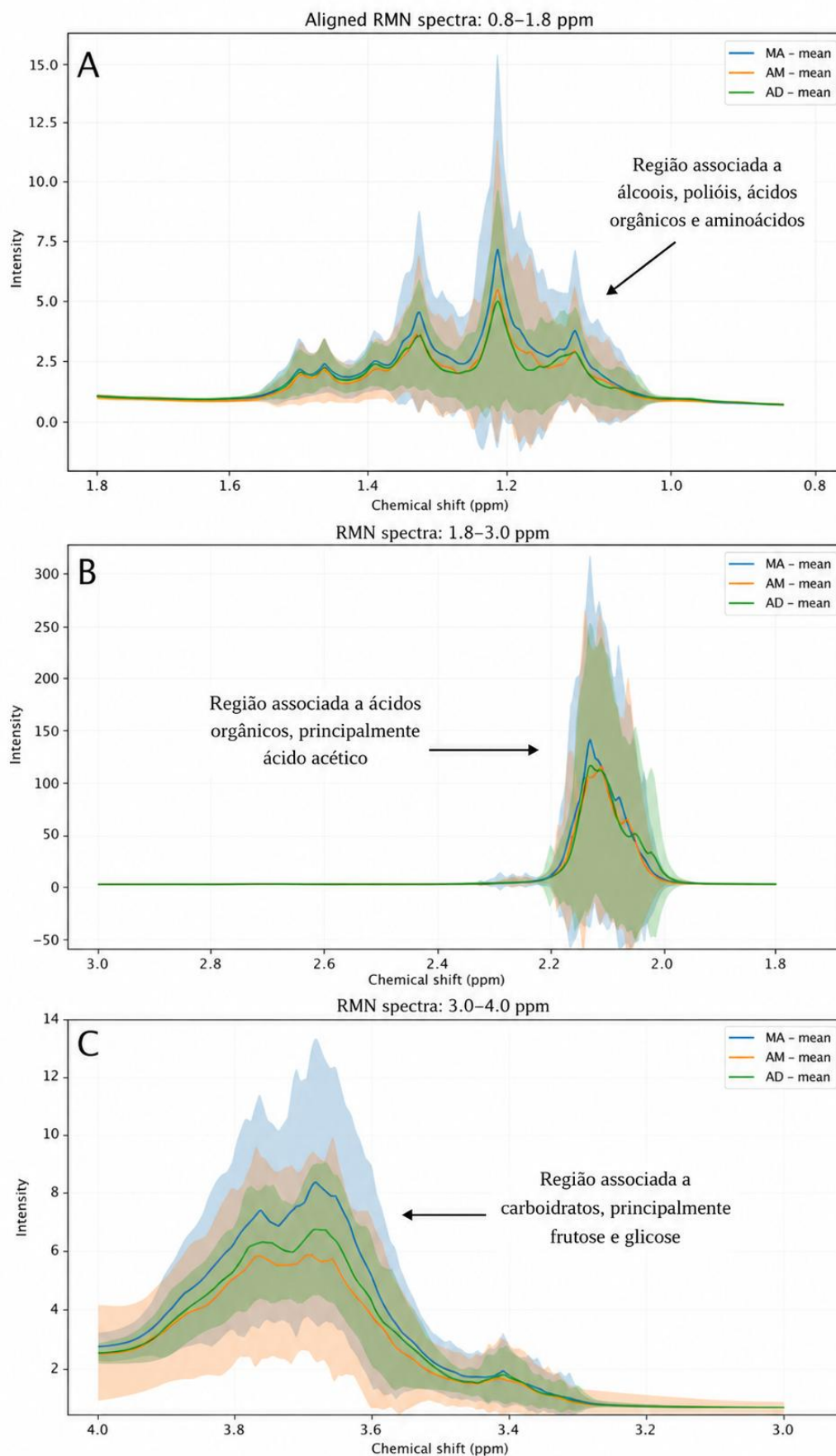


Figura 2. Regiões selecionadas dos espectros médios unidimensionais de LF-NMR para

as classes de amostras MA (vinagre de maçã adulterado com 5, 15 ou 25% de vinagre de álcool), AM (amostras de vinagre de maçã orgânico) e AD (vinagre de maçã orgânico adulterado com 5, 6, 30 ou 31% de vinagre de maçã tradicional). As regiões espectrais destacadas correspondem a: A, 1,8–0,8 ppm, associada a álcoois, polióis, ácidos orgânicos e aminoácidos; B, 3,0–1,8 ppm, principalmente relacionada a ácidos orgânicos; e C, 4,0–3,0 ppm, predominantemente associada a carboidratos.

A Figura 2 apresenta os espectros médios de RMN de ^1H de baixo campo das amostras de vinagre agrupadas por classe (AM, AD e MA). Como esperado para produtos fermentados derivados de uma matéria-prima comum, os perfis espectrais globais são altamente semelhantes entre as classes; entretanto, diferenças sutis, mas quimicamente relevantes, podem ser observadas em regiões específicas. Essas variações refletem diferenças na composição subjacente de carboidratos, ácidos orgânicos, álcoois e metabólitos derivados da fermentação.

Semelhanças espectrais globais associadas a variações sutis dependentes da classe têm sido consistentemente relatadas em estudos utilizando RMN de ^1H de alto e baixo campo voltados à caracterização e autenticação de vinagres. Investigações sobre vinagres de vinho demonstraram que, apesar da pronunciada sobreposição dos perfis espectrais globais, análises multivariadas podem explorar pequenas, porém sistemáticas, diferenças de intensidade associadas a metabólitos fermentativos e substratos residuais. Esses achados reforçam a adequação do *fingerprinting* por RMN de baixo campo como uma abordagem não direcionada para avaliação da autenticidade composicional em matrizes fermentadas complexas (Avanzi Barbosa Mascareli et al., 2023; Grassi et al., 2024).

Os espectros de RMN de ^1H de baixo campo podem ser amplamente divididos em três regiões quimicamente relevantes. A primeira região, compreendida aproximadamente entre 3,0 e 4,0 ppm, está predominantemente associada às ressonâncias de carboidratos, particularmente

de açúcares residuais como glicose e frutose. Mesmo após a fermentação, os açúcares podem não ser completamente consumidos e permanecer detectáveis, contribuindo para o *fingerprint* espectral. Esse comportamento é consistente com estudos de RMN de alto campo em vinagres, os quais indicam que os carboidratos representam uma importante fração contribuinte do perfil químico, juntamente com outros metabólitos de baixa massa molecular, como etanol e ácidos orgânicos (Avanzi Barbosa Mascareli et al., 2023; Boffo et al., 2009; Caligiani et al., 2007; Grassi et al., 2024).

No contexto da adulteração, essa região é particularmente sensível aos efeitos de diluição e mistura, uma vez que vinagres produzidos sob diferentes regimes fermentativos ou oriundos de distintos sistemas de produção podem apresentar diferentes concentrações de açúcares. Estudos prévios de autenticação baseados em RMN relataram que alterações relativas nas ressonâncias associadas aos carboidratos desempenham papel relevante na distinção entre vinagres de vinho autênticos e aqueles adulterados com outros tipos de vinagre, como o vinagre de álcool (Grassi et al., 2024).

A segunda região, localizada aproximadamente entre 1,8 e 3,0 ppm, é atribuída principalmente a ácidos orgânicos, incluindo os ácidos acético, málico, cítrico e succínico. O ácido acético, principal produto da fermentação acética, contribui fortemente para essa região, enquanto os ácidos málico e cítrico são indicativos da origem frutal dos vinagres de maçã (Boffo et al., 2009; Caligiani et al., 2007). As variações dentro dessa janela espectral são particularmente informativas, uma vez que as proporções relativas desses ácidos podem se alterar dependendo das condições de processamento, qualidade da matéria-prima e presença de adulterantes. Estudos multivariados baseados em RMN identificam consistentemente os ácidos orgânicos entre as variáveis mais relevantes para discriminação e autenticação de vinagres (Avanzi Barbosa Mascareli et al., 2023).

Os ácidos orgânicos têm sido repetidamente identificados como contribuintes

discriminativos-chave em estudos de autenticação de vinagres, independentemente da técnica espectroscópica empregada (Cavdaroglu & Ozen, 2022). Tanto a RMN de alto campo quanto métodos espectroscópicos complementares relatam que variações nos perfis de ácido acético, málico e cítrico refletem diferenças na origem da matéria-prima, vias fermentativas e práticas tecnológicas. A relevância dessa região observada no presente estudo é, portanto, consistente com a literatura mais ampla, confirmando que a composição de ácidos orgânicos constitui um eixo central de diferenciação química em matrizes de vinagre (Avanzi Barbosa Mascareli et al., 2023; Caligiani et al., 2007; Grassi et al., 2024; Grassi & Alamprese, 2024a).

A região de baixo deslocamento químico, estendendo-se aproximadamente de 0,8 a 1,8 ppm, contém um conjunto denso de sinais sobrepostos provenientes de álcoois, polióis, ácidos orgânicos, aminoácidos e compostos voláteis relacionados à fermentação, incluindo etanol, acetoína, 2,3-butanodiol, ácido láctico, alanina, leucina, valina e acetato de etila (Avanzi Barbosa Mascareli et al., 2023; Caligiani et al., 2007). Embora a resolução espectral seja intrinsecamente limitada em instrumentos de baixo campo, essa região integra informações de múltiplos metabólitos e reflete amplamente a dinâmica fermentativa e a complexidade metabólica das matrizes de vinagre. Estudos de RMN de alto campo corroboram que essa região carrega informações composicionais substanciais, mesmo na ausência de etapas de extração ou concentração (Avanzi Barbosa Mascareli et al., 2023; Boffo et al., 2009; Caligiani et al., 2007).

Embora as atribuições individuais de compostos sejam limitadas em baixos campos magnéticos, estudos anteriores demonstram que essa região espectral integra informações provenientes de múltiplas vias relacionadas à fermentação. Assim, ela reflete desfechos metabólicos cumulativos, em vez de marcadores isolados, o que se alinha bem aos princípios das abordagens não direcionadas de *fingerprinting*. A contribuição dessa região para a discriminação das amostras destaca a capacidade da RMN de baixo campo em capturar complexidade bioquímica significativa, mesmo sob resolução espectral restrita (Grassi et al.,

2024).

A composição química de produtos orgânicos, como o vinagre de maçã orgânico, pode apresentar variabilidade substancial como resultado da interação entre condições ambientais e práticas de manejo ao longo do tempo. Flutuações sazonais e interanuais em compostos fenólicos, açúcares, aminoácidos e metabólitos voláteis são amplamente atribuídas às condições climáticas ao longo do ciclo produtivo, incluindo variações de precipitação, temperatura e radiação solar, bem como à dinâmica temporal do solo, particularmente em relação ao teor de matéria orgânica e à disponibilidade de nutrientes (Barański et al., 2014; Cruz-Carrión et al., 2023; Cubero-Leon et al., 2018).

Essa variabilidade intrínseca também tem sido relatada como um fator crítico que dificulta a autenticação de produtos orgânicos em geral. Estudos observam que vinagres orgânicos podem apresentar maior dispersão composicional do que seus equivalentes convencionais, não necessariamente em decorrência de adulteração, mas como consequência de insumos agrícolas e práticas fermentativas menos padronizadas. As observações do presente estudo são, portanto, consistentes com relatos da literatura que enfatizam a necessidade de interpretar resultados de autenticação de vinagres orgânicos dentro do contexto de heterogeneidade química natural (Avanzi Barbosa Mascareli et al., 2023).

Esses fatores modulam processos fisiológicos e metabólicos das plantas, influenciando diretamente a biossíntese e o acúmulo de metabólitos primários e secundários. Em maçãs, matéria-prima do vinagre de maçã, alterações nos perfis fenólicos têm sido associadas a diferenças nas condições climáticas do ano de colheita e no microclima do pomar, sendo estresses térmicos e hídricos frequentemente relacionados ao aumento das concentrações fenólicas como parte de uma resposta adaptativa. De maneira semelhante, a composição de açúcares reflete tanto características genéticas quanto variabilidade ambiental que afeta a atividade fotossintética e a partição de carboidratos, enquanto os perfis de aminoácidos estão

fortemente relacionados à disponibilidade de nitrogênio e ao estado nutricional do solo. Embora alguns estudos sugiram que sistemas de produção orgânica possam apresentar maior variabilidade composicional para determinados compostos em comparação aos sistemas convencionais, esse comportamento não é consistente para todas as classes químicas. Ainda assim, o papel central do clima e da dinâmica do solo fornece uma estrutura robusta para interpretação da variabilidade observada na composição química dos vinagres de maçã orgânicos (Barański et al., 2014; Cruz-Carrión et al., 2023; Cubero-Leon et al., 2018; Martí et al., 2018; Petkovsek et al., 2010).

A literatura complementar sobre abordagens espectroscópicas não direcionadas para análise de vinagres reforça que a discriminação entre produtos autênticos e adulterados é primariamente conduzida por diferenças composicionais globais, e não por marcadores químicos isolados. Técnicas como espectroscopia no infravermelho médio e UV-Vis, por exemplo, alcançaram elevada acurácia na detecção de adulterações ao capturar variações composicionais gerais (Grassi & Alamprese, 2024a; Manolache et al., 2025). No contexto da RMN, estudos recentes aplicando especificamente RMN de baixo campo à autenticação de vinagres demonstraram que regiões espectrais entre aproximadamente 3,75 e 0,90 ppm abrangem variabilidade química suficiente para permitir discriminação robusta entre vinagres autênticos e adulterados, mesmo quando o adulterante é outro tipo de vinagre, como o vinagre de álcool (Grassi et al., 2024).

As diferenças dependentes da classe observadas ao longo das três regiões espectrais não podem ser diretamente atribuídas a compostos individuais quando se utiliza instrumentação de baixo campo. Ainda assim, essas diferenças são quimicamente significativas. Alterações nas intensidades da região de carboidratos podem refletir diferentes graus de consumo de açúcares ou efeitos de diluição associados à adulteração (Grassi et al., 2024); variações nas ressonâncias de ácidos orgânicos capturam mudanças em metabólitos derivados da fermentação; e

modificações na região mais complexa de baixo campo refletem alterações em subprodutos fermentativos e compostos voláteis relacionados às condições de processamento e às características da matéria-prima.

Essa abordagem composta de *fingerprinting*, baseada em variações sutis, porém sistemáticas, nas intensidades relativas dos sinais, é consistente com paradigmas quimiométricos não direcionados, nos quais modelos multivariados exploram variações correlacionadas ao longo de todo o espectro, em vez de depender de marcadores específicos individuais (Z. Liu et al., 2022; Sobolev et al., 2015).

Em conjunto, as observações do presente estudo e a literatura de suporte confirmam que os *fingerprints* espectrais de RMN de baixo campo constituem assinaturas quimicamente interpretáveis da composição de vinagres. Embora as ressonâncias individuais nem sempre possam ser resolvidas ou atribuídas de maneira inequívoca, os padrões espectrais integrados capturam diferenças químicas significativas entre amostras autênticas e adulteradas, representando uma estratégia viável para discriminação e autenticação de produtos fermentados complexos, como diferentes tipos de vinagre de maçã.

3.2. Exploração dos dados

Uma avaliação exploratória preliminar foi realizada utilizando Análise de Componentes Principais (PCA) como abordagem não supervisionada para investigar possíveis tendências intrínsecas de agrupamento nos conjuntos de dados espectrais antes da classificação supervisionada. Os modelos de PCA foram avaliados independentemente para a discriminação entre vinagres de maçã orgânicos autênticos e amostras adulteradas com vinagre de maçã convencional (AM vs AD), bem como para vinagres de maçã adulterados com vinagre de álcool (AM vs MA), considerando diferentes estratégias de pré-processamento, nomeadamente os espectros originais, a sub-banda espectral selecionada (0–4,5 ppm) e o mesmo intervalo

espectral após remoção de outliers (Grassi et al., 2024).

De maneira geral, a discriminação entre as classes avaliadas utilizando a abordagem não supervisionada de PCA foi sutil, uma vez que uma considerável sobreposição entre as amostras foi observada em todas as condições testadas (Fig. 3). Embora algumas tendências visuais de agrupamento possam ser percebidas em determinados gráficos de scores, particularmente nas projeções de PC1 versus PC2, esses padrões não estiveram claramente associados à discriminação pretendida entre as classes. Em vez disso, as distribuições observadas provavelmente refletem outras fontes de variabilidade química não relacionadas ao status de adulteração.

Esse comportamento sugere que diversos fatores podem exercer influência mais pronunciada sobre os perfis de *fingerprinting* por RMN dos vinagres do que os padrões específicos de adulteração investigados neste estudo. A heterogeneidade composicional natural associada à variabilidade da matéria-prima, práticas de produção, condições de fermentação, efeitos sazonais, armazenamento e outras características intrínsecas da matriz pode contribuir de maneira mais significativa para a variância espectral global capturada pelos primeiros componentes principais. Consequentemente, a estrutura dominante de variância descrita por PC1 e PC2 foi insuficiente para fornecer separação clara das classes tanto para AM vs AD quanto para AM vs MA (Fig. 3).

De forma geral, esses achados reforçam que a informação discriminante relevante associada à adulteração foi sutil e provavelmente incorporada em estruturas latentes de ordem superior, não adequadamente capturadas pelo modelo não supervisionado de PCA. Portanto, abordagens quimiométricas supervisionadas tornaram-se necessárias para extrair informações relacionadas às classes e melhorar o desempenho da discriminação.

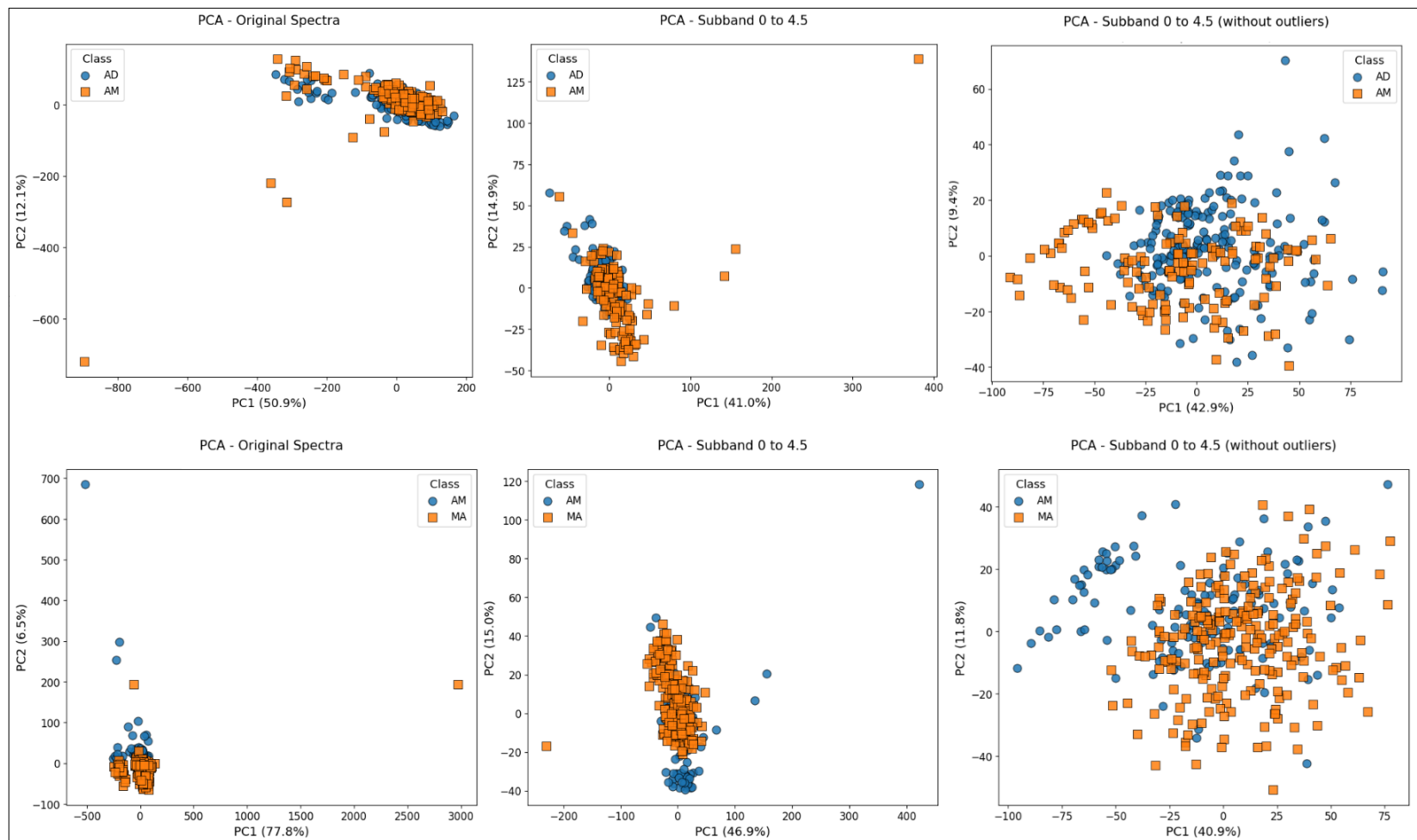


Figura 3. PCA – EXP1: Vinagre de maçã orgânico adulterado com vinagre de maçã tradicional (AM vs AD) – três quadrantes superiores.

EXP2: Vinagre de maçã adulterado com vinagre de álcool (MA vs AM) – três quadrantes inferiores

3.3. Estratégia de autenticação orientada por dados e efeito do pré-processamento espectral

A autenticação de vinagres comerciais foi abordada utilizando uma estratégia orientada por dados baseada em *fingerprints* espectrais de RMN de ^1H de baixo campo combinados com modelagem quimiométrica supervisionada. Dois cenários de classificação binária foram investigados, nomeadamente vinagre de maçã orgânico *versus* vinagre de maçã orgânico adulterado com vinagre de maçã tradicional (AM vs AD), e vinagre de maçã *versus* vinagre de maçã adulterado com vinagre de álcool (AM vs MA). Para ambas as tarefas, modelos de classificação foram desenvolvidos utilizando *pipelines* estruturados de pré-processamento e avaliados sob esquemas rigorosos de validação cruzada aninhada, a fim de garantir estimativas não enviesadas de desempenho. As métricas são reportadas sob validação cruzada aninhada com agrupamento (vinagre-base não visto), o que previne vazamento de informação entre amostras de mesma origem.

Uma avaliação comparativa das estratégias de pré-processamento espectral revelou influência marcante sobre o desempenho da classificação. Três abordagens foram avaliadas: espectros brutos sem correção (Original), alinhamento espectral e correção de linha de base utilizando mínimos quadrados assimétricos (ALS). Em ambos os cenários experimentais, o alinhamento espectral superou consistentemente as demais estratégias, resultando em maiores valores de *F1-score* macro-médio e acurácia de classificação sob validação cruzada aninhada. Em contraste, a correção de linha de base por ALS resultou em desempenho inferior e menos estável, enquanto o uso de espectros não corrigidos produziu resultados intermediários.

Para a tarefa AM vs AD, o *pipeline* com alinhamento alcançou o maior desempenho na validação cruzada aninhada (macro-F1 = $0,77 \pm 0,07$; acurácia = $0,79 \pm 0,06$), superando claramente tanto os *pipelines* Original quanto aqueles corrigidos por ALS (Tabela 1).

Tabela 1. Comparação das estratégias de pré-processamento para o Experimento 1 (AD vs AM, sub-banda de 0–4,5 ppm, conjunto de dados limpo) utilizando validação cruzada aninhada com agrupamento (loop externo de 10 folds; loop interno de 5 folds)

Pré-processamento	Melhor varredura (n, F1)	F1 da validação cruzada aninhada (média $\pm$ DP)	Acurácia da validação cruzada aninhada (média $\pm$ DP)	Precisão da validação cruzada aninhada (média $\pm$ DP)	Melhor n aninhado (moda)
Original	n=11, F1=0,7718	0,7300 $\pm$ 0,0683	0,7464 $\pm$ 0,0597	0,7402 $\pm$ 0,0607	7
Alinhado	n=9, F1=0,7938	0,7708 $\pm$ 0,0669	0,7858 $\pm$ 0,0594	0,7824 $\pm$ 0,0619	8
Linha de base (ALS)	n=7, F1=0,7353	0,7324 $\pm$ 0,1039	0,7496 $\pm$ 0,0966	0,7457 $\pm$ 0,1041	6

Uma tendência semelhante foi observada para a tarefa AM vs MA, na qual o alinhamento novamente forneceu as métricas de generalização mais confiáveis (macro-F1 = $0,73 \pm 0,08$; acurácia = $0,75 \pm 0,07$), enquanto a correção de linha de base apresentou o menor desempenho (Tabela 2). Esses resultados demonstram que pequenos desalinhamentos espectrais dentro da região de 0,0–4,5 ppm podem afetar substancialmente o aprendizado e a generalização do modelo, e que a correção desses deslocamentos é crítica para autenticação robusta baseada em *fingerprinting*.

De maneira geral, a superioridade consistente do alinhamento espectral em ambos os problemas de classificação indica que o pré-processamento desempenha papel decisivo em modelos de autenticação baseados em RMN de baixo campo, particularmente quando diferenças químicas sutis entre matrizes de vinagre intimamente relacionadas estão sendo exploradas.

Tabela 2. Comparação de pré-processamento para MA vs AM (validação cruzada aninhada com agrupamento, *loop* externo de 10 folds; *loop* interno de 5 folds)

Pré-processamento	Melhor varredura (n, F1)	F1 da validação cruzada aninhada	Acurácia da validação cruzada aninhada	Precisão da validação cruzada aninhada	Melhor n aninhado (moda)
Original	n=7,	0,7261 $\pm$ 0,0922	0,7517 $\pm$ 0,0811	0,7425 $\pm$ 0,0884	7

Pré-processamento	Melhor varredura (n, F1)	F1 da validação cruzada aninhada	Acurácia da validação cruzada aninhada	Precisão da validação cruzada aninhada	Melhor n aninhado (moda)
	F1=0,7487				
Alinhado	n=8, F1=0,7954	0,7314 ± 0,0768	0,7543 ± 0,0664	0,7391 ± 0,0728	4
Linha de base (ALS)	n=8, F1=0,6874	0,6929 ± 0,0846	0,7183 ± 0,0740	0,7025 ± 0,0802	6

3.4. Autenticação de vinagres e detecção de adulteração

A Tabela 3 resume o melhor desempenho de classificação obtido para ambos os cenários de autenticação utilizando espectros alinhados de RMN de ^1H de baixo campo e validação cruzada estratificada de 10 *folds*. Acurácias globais comparáveis ($\sim 0,81$) e F1-scores macro-médios ($\sim 0,79$ – $0,80$) foram alcançados em ambos os experimentos, indicando que os *fingerprints* espectrais de LF-NMR contêm informação química suficiente para sustentar a discriminação entre amostras autênticas e adulteradas de vinagre sob diferentes estratégias de adulteração.

Tabela 3. Melhor desempenho obtido para ambos os experimentos (validação cruzada estratificada de 10 *folds*, pré-processamento com alinhamento)

Experimento	Tarefa	Amostras utilizadas	Acurácia	Macro-Precisão	Macro-Recall	Macro-F1	Observações
Exp. 1	AD vs AM	331	0,81	0,80	0,79	0,79	AD F1=0,84; AM F1=0,74; 64 erros
Exp. 2	MA vs AM	358	0,81	0,80	0,79	0,80	MA F1=0,85; AM F1=0,74; 67 erros

3.4.1. EXP. 1: Vinagre de maçã orgânico adulterado com vinagre de maçã tradicional (AD vs AM)

Para a discriminação entre vinagre de maçã orgânico autêntico (AM) e vinagre de maçã orgânico adulterado com vinagre de maçã tradicional (AD), o modelo supervisionado baseado em PLS alcançou acurácia global de 0,81 e macro-F1 de 0,79. Apesar desse desempenho geral

satisfatório, as métricas por classe revelaram pronunciada assimetria. As amostras adulteradas (AD) foram identificadas com maior confiabilidade ($\text{recall} = 0,87$; $\text{F1} = 0,84$), enquanto os vinagres de maçã orgânicos autênticos (AM) apresentaram menor recall (0,71) e F1-score (0,74).

A matriz de confusão (Fig. 4) indica que a maioria das classificações incorretas ocorreu na direção $\text{AM} \rightarrow \text{AD}$, significando que uma fração substancial das amostras orgânicas autênticas foi erroneamente classificada como adulterada. Esse comportamento pode ser diretamente relacionado às características químicas e composicionais do conjunto de amostras. Vinagres de maçã orgânicos e tradicionais compartilham a mesma origem botânica e vias fermentativas semelhantes, resultando em *fingerprints* de LF-NMR altamente sobrepostos, dominados por carboidratos, ácidos orgânicos e metabólitos derivados da fermentação na região de 0,0–4,5 ppm.

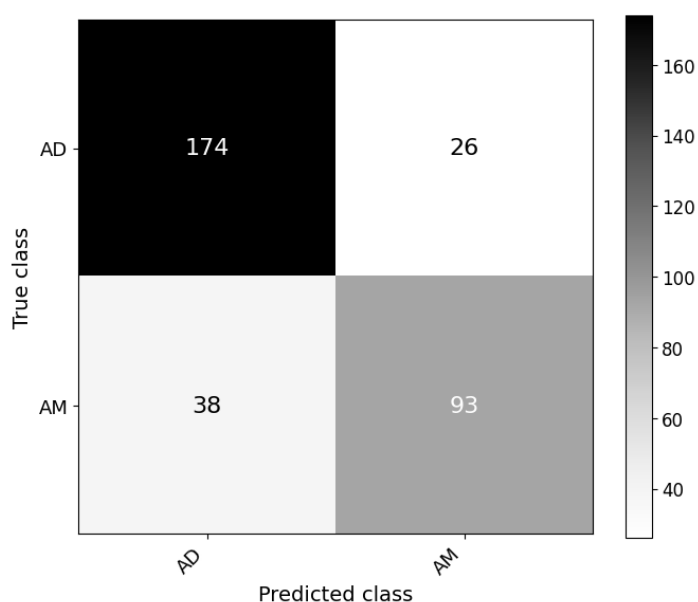


Figura 4. Matriz de confusão do Experimento 1 (AD vs AM)

Do ponto de vista químico, a adulteração de vinagre de maçã orgânico com vinagre de maçã tradicional não introduz novas classes de compostos, mas modifica as concentrações

relativas de metabólitos já presentes em ambas as matrizes. Consequentemente, a classe AD tende a formar um agrupamento mais compacto no espaço de *fingerprinting*, refletindo deslocamentos composicionais sistemáticos (por exemplo, diluição ou redistribuição de açúcares e ácidos orgânicos), enquanto a classe AM apresenta maior variabilidade intrínseca. Essa variabilidade provavelmente decorre de diferenças nas condições do pomar, ano de colheita, parâmetros fermentativos e práticas de produção orgânica, todos fatores que influenciam a abundância de metabólitos sem gerar marcadores distintivos. Como resultado, amostras autênticas de AM se sobrepõem parcialmente às amostras AD, aumentando a probabilidade de falsos positivos de adulteração.

Padrões assimétricos semelhantes de classificação incorreta foram descritos em estudos anteriores envolvendo adulteração de vinagres derivados de frutas, nos quais a mistura com produtos quimicamente semelhantes leva à sobreposição parcial no espaço de *fingerprinting*. Nesses casos, as amostras adulteradas frequentemente formam agrupamentos mais compactos devido a deslocamentos composicionais sistemáticos, enquanto as amostras autênticas mantêm dispersão mais ampla associada à variabilidade natural. Os presentes resultados estão intimamente alinhados com esse comportamento, reforçando a plausibilidade química dos desfechos de classificação observados (Grassi et al., 2024).

Sob a perspectiva de autenticação, essa estrutura assimétrica de erros é conservadora e desejável em contextos de triagem, pois prioriza a detecção de potenciais adulterações em detrimento do risco de falsos negativos.

3.4.2. EXP. 2: Vinagre de maçã adulterado com vinagre de álcool (MA vs AM)

No segundo cenário de autenticação, vinagre de maçã adulterado com vinagre de álcool (MA) foi discriminado de vinagre de maçã autêntico (AM) com acurácia global de 0,81 e macro-F1 de 0,80. Assim como no experimento anterior, as amostras adulteradas foram mais

facilmente reconhecidas (recall MA = 0,87; F1 = 0,85), enquanto AM novamente apresentou menor recall (0,72) e F1-score (0,74).

Embora o vinagre de álcool represente um produto quimicamente distinto, derivado de substratos alcoólicos e não de frutas, o padrão de classificação incorreta observado indica que persiste uma sobreposição espectral parcial com o vinagre de maçã no domínio de RMN de baixo campo. A matriz de confusão (Fig. 5) mostra que as classificações incorretas ocorreram predominantemente novamente na direção AM $\rightarrow$ MA, sugerindo que um subconjunto das amostras de vinagre de maçã compartilha similaridades composicionais com o vinagre de álcool dentro da janela espectral avaliada.

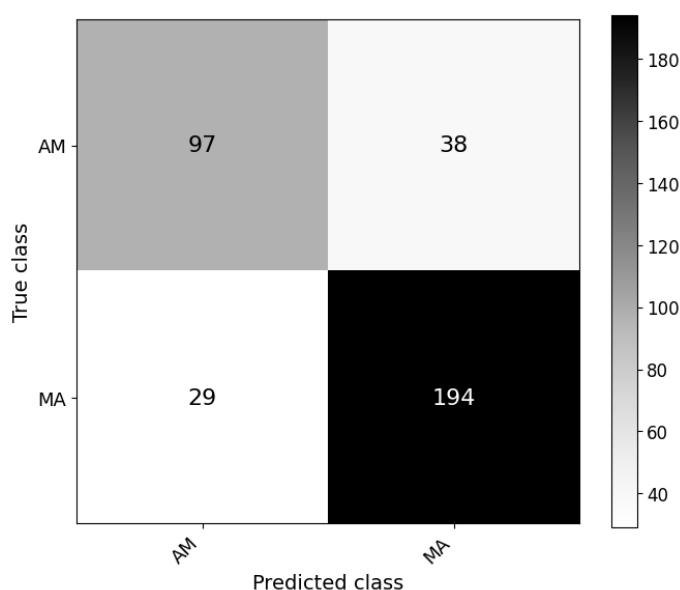


Figura 5. Matriz de confusão do Experimento 1 (MA vs AM)

Essa sobreposição pode ser racionalizada pela predominância de produtos fermentativos de baixa massa molecular, notadamente ácido acético, ressonâncias relacionadas ao etanol e outros ácidos orgânicos, comuns a ambas as matrizes e que contribuem fortemente para o sinal de LF-NMR entre 0,0 e 4,5 ppm. Embora metabólitos específicos de frutas, como ácido málico ou cítrico, contribuam para a identidade do vinagre de maçã, seus sinais podem estar

parcialmente mascarados ou diluídos em espectros de baixo campo, limitando seu poder discriminativo no nível de ressonâncias individuais. Consequentemente, as amostras MA formam uma classe relativamente coerente devido ao perfil químico consistente do vinagre de álcool, enquanto as amostras AM novamente apresentam variabilidade mais ampla relacionada à diversidade da matéria-prima e à heterogeneidade do processamento (Avanzi Barbosa Mascareli et al., 2023; Boffo et al., 2009; Caligiani et al., 2007; Grassi et al., 2024).

Investigações anteriores baseadas em RMN comparando vinagres de frutas e vinagres de álcool também relatam que, apesar das claras diferenças na origem da matéria-prima, a sobreposição espectral em baixo campo persiste devido à predominância de ácido acético e outros produtos fermentativos de baixa massa molecular. Como resultado, a discriminação depende de padrões multivariados sutis, em vez da presença ou ausência de ressonâncias específicas. Os achados do presente estudo corroboram essa interpretação e demonstram que os *fingerprints* de RMN de baixo campo mantêm poder discriminativo suficiente mesmo sob essas condições quimicamente desafiadoras (Cavdaroglu & Ozen, 2023).

3.5. Padrões de erro, comportamento dependente da classe e interpretação química

Uma interpretação conjunta de ambos os experimentos de classificação, sustentada pelas métricas de desempenho da Tabela 3 e pelas matrizes de confusão das Figuras 4 e 5, revela um padrão quimicamente consistente: os vinagres autênticos (AM) representam a classe mais heterogênea e desafiadora, enquanto as amostras adulteradas (AD e MA) são classificadas com maior facilidade. Em ambos os cenários, AM apresentou sistematicamente menores valores de recall e F1-score (Tabela 3) e foi responsável pela maioria dos eventos de classificação incorreta, predominantemente na direção AM → adulterado, indicando maior dispersão intraclasse e sobreposição parcial na dimensão de *fingerprinting* por LF-NMR.

Esse comportamento indica que AM ocupa uma região mais ampla no espaço latente

(maior dispersão intraclasse), consistente com sua maior variabilidade composicional no espaço de *fingerprinting* de LF-NMR. No cenário AD vs AM, AM se sobrepõe a AD devido aos metabólitos derivados da maçã compartilhados e à química fermentativa semelhante. No cenário MA vs AM, AM também se sobrepõe parcialmente a MA em decorrência dos produtos finais comuns da fermentação que dominam a região espectral de baixo campo. Assim, AM efetivamente conecta o espaço químico entre matrizes de vinagre derivadas de frutas e derivadas de álcool. Esse posicionamento sustenta o *fingerprinting* por LF-NMR principalmente como uma abordagem de triagem para autenticação de vinagres. Na prática, amostras limítrofes de AM (ou seja, espectros próximos ao limite de decisão) podem requerer análises confirmatórias utilizando métodos direcionados, como RMN de maior campo, cromatografia ou abordagens isotópicas, dependendo do contexto regulatório ou industrial.

Do ponto de vista quimiométrico, essa sobreposição é inerente à abordagem não direcionada de *fingerprinting*, que explora variações correlacionadas ao longo de todo o espectro em vez de depender de compostos marcadores exclusivos. A variabilidade relacionada à cultivar da maçã, cinética fermentativa, tempo de maturação e diferenças entre lotes amplia ainda mais a distribuição da classe AM, reforçando sua sobreposição parcial com ambas as classes adulteradas (Avanzi Barbosa Mascareli et al., 2023; Petkovsek et al., 2010).

As melhorias observadas associadas ao alinhamento espectral confirmam que estratégias de pré-processamento podem reduzir variâncias artificiais e aumentar a robustez do modelo, mas não conseguem superar completamente a similaridade química intrínseca e a variabilidade biológica presentes nos conjuntos de amostras. Em termos práticos, esses achados demonstram que o *fingerprinting* por RMN de baixo campo combinado com quimiometria supervisionada é particularmente eficaz para detectar padrões sistemáticos de adulteração, enquanto a classificação de vinagres de maçã autênticos requer interpretação cautelosa devido à sua diversidade composicional natural.

De maneira geral, os resultados destacam tanto o potencial analítico quanto as limitações fundamentais do *fingerprinting* baseado em LF-NMR para autenticação de vinagres, ressaltando a importância da integração entre o conhecimento químico dos conjuntos de amostras, a modelagem quimiométrica e a análise detalhada de erros em aplicações de detecção de fraude em alimentos.

Essas tendências estão de acordo com observações relatadas em estudos de autenticação espectroscópica não direcionada, nos quais a discriminação frequentemente reflete a estrutura composicional global capturada por *fingerprints* multivariados, em vez da dependência de marcadores individuais (Avanzi Barbosa Mascareli et al., 2023; Cavdaroglu & Ozen, 2022, 2023). A RMN de ^1H de baixo campo, quando combinada com pré-processamento apropriado e modelagem quimiométrica supervisionada, emerge como uma ferramenta prática e quimicamente interpretável para autenticação de vinagres, particularmente adequada para aplicações de triagem de alto rendimento, nas quais robustez e mínima preparação de amostra são essenciais (Cavdaroglu & Ozen, 2023; Grassi et al., 2024; Grassi & Alamprese, 2024b).

4. Conclusão

Este estudo demonstra que o *fingerprinting* espectral por RMN de ^1H de baixo campo combinado com quimiometria supervisionada fornece uma estratégia robusta e quimicamente interpretável para a autenticação de vinagres de maçã e a detecção de adulterações. Os resultados de classificação refletem não apenas o desempenho do modelo, mas também a complexidade composicional inerente do vinagre como matriz alimentar fermentada.

O vinagre de maçã orgânico apresenta pronunciada variabilidade natural impulsionada pela diversidade da matéria-prima, condições ambientais, dinâmica do solo e práticas fermentativas. Essa variabilidade é diretamente capturada pelos *fingerprints* não direcionados de LF-NMR, os quais são dominados por metabólitos de baixa massa molecular comuns a

diferentes tipos de vinagre. Consequentemente, a sobreposição espectral parcial entre amostras autênticas e adulteradas é quimicamente esperada e reflete similaridades genuínas, e não limitações metodológicas.

Mesmo sob essas condições, a abordagem orientada por dados proposta identificou consistentemente diferenças composicionais sistemáticas associadas à adulteração, apresentando desempenho estável e padrões de erro coerentes dependentes da classe em diferentes esquemas de validação. As amostras adulteradas formaram agrupamentos mais compactos e reprodutíveis no espaço de *fingerprinting*, enquanto os vinagres de maçã autênticos exibiram maior dispersão intraclasse, em concordância com sua heterogeneidade biológica e tecnológica.

Os resultados também enfatizam o papel crítico do alinhamento espectral na redução de variâncias artificiais e na melhoria da generalização dos modelos, ao mesmo tempo em que destacam os limites intrínsecos impostos pela similaridade entre matrizes em análises não direcionadas. De maneira geral, os achados confirmam que os *fingerprints* espectrais de RMN de ^1H de baixo campo constituem assinaturas químicas relevantes para autenticação de vinagres, sustentando sua aplicação como uma ferramenta rápida, economicamente viável e escalável para triagem rotineira de produtos fermentados complexos.

5. Agradecimentos

Os autores também agradecem ao Laboratório (LAABBM) – IQ-UFRJ pelo suporte nas análises de LF-NMR. O apoio financeiro fornecido pelo Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) – número de concessão [307846/2020-2]; e pela Universidade Federal de Santa Catarina (UFSC). Este estudo foi financiado em parte pela Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Código de Financiamento 001.0

6. Referências

- Avanzi Barbosa Mascareli, V., Galvan, D., Craveiro de Andrade, J., Lelis, C., Adam Conte-Junior, C., Michelino Gaeta Lopes, G., César de Macedo Júnior, F., & Aparecida Spinosa, W. (2023). Spectralprint techniques coupled with chemometric tools for vinegar classifications. *Food Chemistry*, 410, 135373. <https://doi.org/10.1016/j.foodchem.2022.135373>
- Barański, M., Średnicka-Tober, D., Volakakis, N., Seal, C., Sanderson, R., Stewart, G. B., Benbrook, C., Biavati, B., Markellou, E., Giotis, C., Gromadzka-Ostrowska, J., Rembiałkowska, E., Skwarło-Sońta, K., Tahvonen, R., Janovská, D., Niggli, U., Nicot, P., & Leifert, C. (2014). Higher antioxidant and lower cadmium concentrations and lower incidence of pesticide residues in organically grown crops: a systematic literature review and meta-analyses. *British Journal of Nutrition*, 112(5), 794–811. <https://doi.org/10.1017/S0007114514001366>
- Barbosa, S., Saurina, J., Puignou, L., & Núñez, O. (2020). Classification and Authentication of Paprika by UHPLC-HRMS Fingerprinting and Multivariate Calibration Methods (PCA and PLS-DA). *Foods*, 9(4), 486. <https://doi.org/10.3390/foods9040486>
- Bekatorou, A. (2019). *Advances in Vinegar Production* (A. Bekatorou, Ed.). CRC Press. <https://doi.org/10.1201/9781351208475>
- Boffo, E. F., Tavares, L. A., Ferreira, M. M. C., & Ferreira, A. G. (2009). Classification of Brazilian vinegars according to their ¹H NMR spectra by pattern recognition analysis. *LWT - Food Science and Technology*, 42(9), 1455–1460. <https://doi.org/10.1016/j.lwt.2009.05.008>
- Brasil, Ministério da Agricultura, P. e A. (2020). Instrução Normativa, no16, de 19 de Março de 2020 e no 17 de 28 de maio de 2009. In *Diário Oficial da República Federal do Brasil* (8.918; p. 2). 2009.

- Caligiani, A., Acquotti, D., Palla, G., & Bocchi, V. (2007). Identification and quantification of the main organic components of vinegars by high resolution ^1H NMR spectroscopy. *Analytica Chimica Acta*, 585(1), 110–119. <https://doi.org/10.1016/j.aca.2006.12.016>
- Callao, M. P., & Ruisánchez, I. (2018). An overview of multivariate qualitative methods for food fraud detection. *Food Control*, 86, 283–293. <https://doi.org/10.1016/j.foodcont.2017.11.034>
- Cavdaroglu, C., & Ozen, B. (2022). Detection of vinegar adulteration with spirit vinegar and acetic acid using UV–visible and Fourier transform infrared spectroscopy. *Food Chemistry*, 379, 132150. <https://doi.org/10.1016/j.foodchem.2022.132150>
- Cavdaroglu, C., & Ozen, B. (2023). Applications of UV–Visible, Fluorescence and Mid-Infrared Spectroscopic Methods Combined with Chemometrics for the Authentication of Apple Vinegar. *Foods*, 12(6), 1139. <https://doi.org/10.3390/foods12061139>
- Ceschin, M., Arrighi, L., Longo, L., & Barbon Junior, S. (2026). Extending Decision Predicate Graphs for Comprehensive Explanation of Isolation Forest (pp. 271–293). https://doi.org/10.1007/978-3-032-08324-1_12
- Conte, T. L., de Andrade, J. C., da Silva Borges, R. L., Effting, L., Ravaglia, L. M., Alcantara, G. B., Conte-Junior, C. A., Bona, E., & Galvan, D. (2025). Benchtop NMR spectroscopy and multivariate calibration for fast prediction of oxidation stability in commercial biodiesel. *Microchemical Journal*, 212, 113149. <https://doi.org/10.1016/j.microc.2025.113149>
- Cruz-Carrión, Á., Ruiz de Azua, Ma. J., Muguerza, B., Mulero, M., Bravo, F. I., Arola-Arnal, A., & Suarez, M. (2023). Organic vs. Non-Organic Plant-Based Foods—A Comparative Study on Phenolic Content and Antioxidant Capacity. *Plants*, 12(1), 183. <https://doi.org/10.3390/plants12010183>
- Cubero-Leon, E., De Rudder, O., & Maquet, A. (2018). Metabolomics for organic food

- authentication: Results from a long-term field study in carrots. *Food Chemistry*, 239, 760–770. <https://doi.org/10.1016/j.foodchem.2017.06.161>
- da Silva Pereira, E., Cruz-Tirado, J. P., Lourenço Crippa, B., Martins Morasi, R., Milagres de Almeida, J., Fernandes Barbin, D., Barbon Junior, S., & Cristina Cirone Silva, N. (2024). Portable near infrared (NIR) spectrometer coupled with machine learning to classify milk with subclinical mastitis. *Food Control*, 163, 110527. <https://doi.org/10.1016/j.foodcont.2024.110527>
- El Abdali, Y., Saghrouchni, H., Kara, M., Mssillou, I., Allali, A., Jordan, Y. A. Bin, Kafkas, N. E., El-Assri, E.-M., Nafidi, H.-A., Bourhia, M., Almaary, K. S., Eloutassi, N., & Bouia, A. (2023). Exploring the Bioactive Compounds in Some Apple Vinegar Samples and Their Biological Activities. *Plants*, 12(22), 3850. <https://doi.org/10.3390/plants12223850>
- Fang, C.-J., You, H.-C., Huang, Z.-L., Hsu, C.-L., Tsai, C.-F., Lin, Y.-T., Kao, Y.-M., Tseng, S.-H., Wang, D.-Y., & Su, N.-W. (2022). Stable carbon isotopic characterization of rice vinegar protein as an intrinsic reference for discriminating the authenticity of brewed rice vinegar. *Journal of Food and Drug Analysis*, 30(1), 38–45. <https://doi.org/10.38212/2224-6614.3398>
- Galvan, D., de Aguiar, L. M., Bona, E., Marini, F., & Killner, M. H. M. (2023). Successful combination of benchtop nuclear magnetic resonance spectroscopy and chemometric tools: A review. *Analytica Chimica Acta*, 1273, 341495. <https://doi.org/10.1016/j.aca.2023.341495>
- Galvan, D., Tanamati, A. A. C., Casanova, F., Danieli, E., Bona, E., & Killner, M. H. M. (2021). Compact low-field NMR spectroscopy and chemometrics applied to the analysis of edible oils. *Food Chemistry*, 365, 130476. <https://doi.org/10.1016/j.foodchem.2021.130476>
- Gao, B., Holroyd, S. E., Moore, J. C., Laurvick, K., Gendel, S. M., & Xie, Z. (2019). Opportunities and Challenges Using Non-targeted Methods for Food Fraud Detection.

- Journal of Agricultural and Food Chemistry, 67(31), 8425–8430.
<https://doi.org/10.1021/acs.jafc.9b03085>
- Grassi, S., & Alamprese, C. (2024b). Spectroscopic Non-targeted Techniques in Combination with Linear Discriminant Analysis for Wine Vinegar Authentication. *Food and Bioprocess Technology*, 17(2), 479–488. <https://doi.org/10.1007/s11947-023-03143-9>
- Grassi, S., Borgonovo, G., Gennaro, M., & Alamprese, C. (2024). NMR-based approach to detect white wine vinegar fraud. *Food Chemistry*, 456, 139953. <https://doi.org/10.1016/j.foodchem.2024.139953>
- Grootveld, M., Percival, B., Gibson, M., Osman, Y., Edgar, M., Molinari, M., Mather, M. L., Casanova, F., & Wilson, P. B. (2019). Progress in low-field benchtop NMR spectroscopy in chemical and biochemical analysis. *Analytica Chimica Acta*, 1067, 11–30. <https://doi.org/10.1016/j.aca.2019.02.026>
- Ho, C. W., Lazim, A. M., Fazry, S., Zaki, U. K. H. H., & Lim, S. J. (2017). Varieties, production, composition and health benefits of vinegars: A review. *Food Chemistry*, 221, 1621–1630. <https://doi.org/10.1016/j.foodchem.2016.10.128>
- Hsieh, C. W., Li, P. H., Cheng, J. Y., & Ma, J. T. (2013). Using SNIF-NMR method to identify the adulteration of molasses spirit vinegar by synthetic acetic acid in rice vinegar. *Industrial Crops and Products*, 50, 904–908. <https://doi.org/10.1016/j.indcrop.2013.08.014>
- Kahraman, H. A., Tutun, H., Kaya, M. M., Tutun, S., Usluer, M. S., Rugji, J., & Yurdakul, O. (2021). Total phenolic content, antiradical, antimicrobial and antibiofilm properties of grape and apple vinegar. *Journal of Advances in VetBio Science and Techniques*, 6(2), 150–158. <https://doi.org/10.31797/vetbio.960155>
- Kunal Ahuja, & Himanshu Kabra. (2022). Vinegar Market Size - By Product Type (White Vinegar, Apple Cider Vinegar, Balsamic Vinegar, Red Wine Vinegar, Rice Vinegar),

- Source (Organic, Synthetic), Flavor, Distribution Channel, Application, Growth Prospects, Regional Outlook & Forecast, 2023 – 203. <https://www.gminsights.com/industry-analysis/vinegar-market>
- Liu, Q., Tang, G.-Y., Zhao, C.-N., Gan, R.-Y., & Li, H.-B. (2019). Antioxidant Activities, Phenolic Profiles, and Organic Acid Contents of Fruit Vinegars. *Antioxidants*, 8(4), 78. <https://doi.org/10.3390/antiox8040078>
- Liu, Z., Zhang, M., Chen, P., Harnly, J. M., & Sun, J. (2022). Mass Spectrometry-Based Nontargeted and Targeted Analytical Approaches in Fingerprinting and Metabolomics of Food and Agricultural Research. *Journal of Agricultural and Food Chemistry*, 70(36), 11138–11153. <https://doi.org/10.1021/acs.jafc.2c01878>
- Loffredi, E., Grassi, S., & Alamprese, C. (2021). Spectroscopic approaches for non-destructive shell egg quality and freshness evaluation: Opportunities and challenges. *Food Control*, 129, 108255. <https://doi.org/10.1016/j.foodcont.2021.108255>
- Manolache, F. A., Poteraş, C. B., Iordache, T. A., & Todaşcă, M. C. (2025). MODERN ANALYTICAL TECHNIQUES FOR VINEGAR AUTHENTICATION AND QUALITY ASSESSMENT: A COMPREHENSIVE REVIEW. In *Revue Roumaine de Chimie* (Vol. 70, Numbers 9–10, pp. 573–583). Publishing House of the Romanian Academy. <https://doi.org/10.33224/rrech.2025.70.9-10.02>
- Martí, R., Leiva-Brondo, M., Lahoz, I., Campillo, C., Cebolla-Cornejo, J., & Roselló, S. (2018). Polyphenol and l-ascorbic acid content in tomato as influenced by high lycopene genotypes and organic farming at different environments. *Food Chemistry*, 239, 148–156. <https://doi.org/10.1016/j.foodchem.2017.06.102>
- Moraes, I. A., Kaufmann, K. C., Barbon Junior, S., Neves, M. G., Villa, J. E. L., & Barbin, D. F. (2025). Authentication of coriander oil and adulterant identification using electronic nose and spectroscopic techniques. *Food Chemistry*, 483, 144196.

<https://doi.org/10.1016/j.foodchem.2025.144196>

- Morgan, J., & Mosawy, S. (2016). The Potential of Apple Cider Vinegar in the Management of Type 2 Diabetes. *International Journal of Diabetes Research*, 5(6), 129–134. <https://doi.org/10.5923/j.diabetes.20160506.02>
- Nie, J., Li, Y., Xing, J., Chao, J., Qin, X., & Li, Z. (2019). Comparison of two types of vinegar with different aging times by NMR-based metabolomic approach. *Journal of Food Biochemistry*, 43(5), e12835. <https://doi.org/10.1111/jfbc.12835>
- Nobari Moghaddam, H., Tamiji, Z., Akbari Lakeh, M., Khoshayand, M. R., & Haji Mahmoodi, M. (2022). Multivariate analysis of food fraud: A review of NIR based instruments in tandem with chemometrics. *Journal of Food Composition and Analysis*, 107, 104343. <https://doi.org/10.1016/j.jfca.2021.104343>
- Oliveri, P., & Downey, G. (2012). Multivariate class modeling for the verification of food-authenticity claims. *TrAC - Trends in Analytical Chemistry*, 35, 74–86. <https://doi.org/10.1016/j.trac.2012.02.005>
- Ousaaïd, D., Laaroussi, H., Bakour, M., ElGhouizi, A., Aboulghazi, A., Lyoussi, B., & ElArabi, I. (2020). Beneficial Effects of Apple Vinegar on Hyperglycemia and Hyperlipidemia in Hypercaloric-Fed Rats. *Journal of Diabetes Research*, 2020, 1–7. <https://doi.org/10.1155/2020/9284987>
- Ousaaïd, D., Laaroussi, H., Bakour, M., Ennaji, H., Lyoussi, B., & El Arabi, I. (2021). Antifungal and Antibacterial Activities of Apple Vinegar of Different Cultivars. *International Journal of Microbiology*, 2021, 1–6. <https://doi.org/10.1155/2021/6087671>
- Perumpuli, B., & Dilrukshi, N. (2022). Vinegar: A functional ingredient for human health. *International Food Research Journal*, 29(5), 959–974. <https://doi.org/10.47836/ifrj.29.5.01>
- Petkovsek, M. M., Slatnar, A., Stampar, F., & Veberic, R. (2010). The influence of organic/integrated production on the content of phenolic compounds in apple leaves and

- fruits in four different varieties over a 2-year period. *Journal of the Science of Food and Agriculture*, 90(14), 2366–2378. <https://doi.org/10.1002/jsfa.4093>
- Ríos-Reina, R., Callejón, R. M., Savorani, F., Amigo, J. M., & Cocchi, M. (2019). Data fusion approaches in spectroscopic characterization and classification of PDO wine vinegars. *Talanta*, 198, 560–572. <https://doi.org/10.1016/j.talanta.2019.01.100>
- Savorani, F., Tomasi, G., & Engelsen, S. B. (2010). icoshift: A versatile tool for the rapid alignment of 1D NMR spectra. *Journal of Magnetic Resonance*, 202(2), 190–202. <https://doi.org/10.1016/j.jmr.2009.11.012>
- Shi, S., Zhang, K., Tian, N., Jin, Z., Liu, K., Huang, L., Tian, X., Cao, C., Zhang, Y., & Jiang, Y. (2025). Spectroscopic techniques combined with chemometrics for rapid detection of food adulteration: Applications, perspectives, and challenges. *Food Research International*, 211, 116459. <https://doi.org/10.1016/j.foodres.2025.116459>
- Sobolev, A., Mannina, L., Proietti, N., Carradori, S., Daglia, M., Giusti, A., Antiochia, R., & Capitani, D. (2015). Untargeted NMR-Based Methodology in the Study of Fruit Metabolites. *Molecules*, 20(3), 4088–4108. <https://doi.org/10.3390/molecules20034088>
- Song, J., Qiu, H., Du, P., Mou, F., Nie, Z., Zheng, Y., & Wang, M. (2022). Polyphenols extracted from Shanxi-aged vinegar exert hypolipidemic effects on OA-induced HepG2 cells via the PPAR α -LXR α -ABCA1 pathway. *Journal of Food Biochemistry*, 46(2). <https://doi.org/10.1111/jfbc.14029>
- Štornik, A., Skok, B., & Trček, J. (2016). Comparison of Cultivable Acetic Acid Bacterial Microbiota between Organic and Conventional Apple Cider Vinegar. *Food Technology and Biotechnology*, 54. <https://doi.org/10.17113/ftb.54.01.16.4082>
- Tehari, S., de Andrade, J. C., de Oliveira, A. T., da Silva Gomes, M. F., Yamamoto, K. C., & Conte-Junior, C. A. (2026). Authentication and traceability of Amazonian stingray meat via low field ^1H NMR metabolic fingerprinting and chemometrics. *Microchemical*

Journal, 222, 117163. <https://doi.org/10.1016/j.microc.2026.117163>

Xia, T., Zhang, B., Duan, W., Zhang, J., & Wang, M. (2020). Nutrients and bioactive components from vinegar: A fermented and functional food. *Journal of Functional Foods*, 64, 103681. <https://doi.org/10.1016/j.jff.2019.103681>

Xiong, C., Wang, Q., & Xu, J. (2023). Tracing the geographical origin of grain vinegar using the stable isotope ratio mass spectrometry and chemometrics. *Journal of Food Composition and Analysis*, 123. <https://doi.org/10.1016/j.jfca.2023.105554>

Yıkımsı, S., Aksu, F., Altunatmaz, S. S., & Çöl, B. G. (2021). Ultrasound Processing of Vinegar: Modelling the Impact on Bioactives and Other Quality Factors. *Foods*, 10(8), 1703. <https://doi.org/10.3390/foods10081703>

MATERIAL SUPLEMENTAR

SM1. Conjunto amostral completo e planejamento de adulterações

A. Conjunto Am – Amostras originais

ID	Tipo	Orgânica ou Tradicional
Am 01	Vinagre de Maçã	O
Am 02	Vinagre de Maçã	O
Am 03	Vinagre de Maçã	O
Am 04	Vinagre de Maçã	O
Am 05	Vinagre de Álcool	T
Am 06	Vinagre de Maçã	O
Am 07	Vinagre de Maçã	T
Am 08	Vinagre de Maçã	T
Am 09	Vinagre de Maçã	O
Am 10	Vinagre de Maçã	T
Am 11	Vinagre de Maçã	T
Am 12	Vinagre de Álcool	T
Am 13	Vinagre de Álcool	T
Am 14	Vinagre de Maçã	O
Am 15	Vinagre de Maçã	O
Am 16	Vinagre de Maçã	O
Am 17	Vinagre de Maçã	O
Am 18	Vinagre de Maçã	T
Am 19	Vinagre de Maçã	T

Am 20	Vinagre de Maça	O
Am 21	Vinagre de Maça	T
Am 22	Vinagre de Maça	O
Am 23	Vinagre de Maça	T
Am 24	Vinagre de Maça	O
Am 25	Vinagre de Maça	T
Am 26	Vinagre de Maça	T
Am 27	Vinagre de Álcool	T
Am 28	Vinagre de Álcool	T
Am 29	Vinagre de Álcool	T
Am 30	Vinagre de Maça	T
Am 31	Vinagre de Maça	O
Am 32	Vinagre de Álcool	O
Am 33	Vinagre de Maça	T
Am 34	Vinagre de Maça	T
Am 35	Vinagre de Álcool	T
Am 36	Vinagre de Álcool	T
Am 37	Vinagre de Álcool	T
Am 38	Vinagre de Álcool	T
Am 39	Vinagre de Álcool	T
Am 40	Vinagre de Maça	O
Am 41	Vinagre de Maça	T
Am 42	Vinagre de Maça	T
Am 43	Vinagre de Álcool	T
Am 44	Vinagre de Álcool	T
Am 45	Vinagre de Álcool	T
Am 46	Vinagre de Álcool	T
Am 47	Vinagre de Maça	T
Am 48	Vinagre de Maça	O
Am 49	Vinagre de Maça	T

Am 50	Vinagre de Maça	O
Am 51	Vinagre de Maça	T
Am 52	Vinagre de Álcool	T
Am 53	Vinagre de Álcool	T
Am 54	Vinagre de Maça	O
Am 55	Vinagre de Maça	O
Am 56	Vinagre de Álcool	T
Am 57	Vinagre de Álcool	T
Am 58	Vinagre de Maça	T
Am 59	Vinagre de Maça	T
Am 60	Vinagre de Maça	T
Am 61	Vinagre de Álcool	T
Am 62	Vinagre de Álcool	T
Am 63	Vinagre de Maça	T
Am 64	Vinagre de Maça	T
Am 65	Vinagre de Maça	O
Am 66	Vinagre de Maça	O
Am 67	Vinagre de Álcool	T
Am 68	Vinagre de Maça	T
Am 69	Vinagre de Maça	T
Am 70	Vinagre de Maça	T
Am 71	Vinagre de Maça	T
Am 72	Vinagre de Álcool	T
Am 73	Vinagre de Maça	O
Am 74	Vinagre de Maça	T
Am 75	Vinagre de Maça	O
Am 76	Vinagre de Maça	O
Am 77	Vinagre de Maça	O
Am 78	Vinagre de Maça	O
Am 79	Vinagre de Maça	O

Am 80	Vinagre de Maçã	T
Am 81	Vinagre de Maçã	O
Am 82	Vinagre de Maçã	O
Am 83	Vinagre de Maçã	O
Am 84	Vinagre de Maçã	O
Am 85	Vinagre de Maçã	O
Am 86	Vinagre de Maçã	O
Am 87	Vinagre de Maçã	O
Am 88	Vinagre de Maçã	O
Am 89	Vinagre de Maçã	O
Am 90	Vinagre de Maçã	O
Am 91	Vinagre de Maçã	O
Am 92	Vinagre de Maçã	O
Am 93	Vinagre de Maçã	O
Am 94	Vinagre de Maçã	O
Am 95	Vinagre de Maçã	T
Am 96	Vinagre de Maçã	O
Am 97	Vinagre de Maçã	O
Am 98	Vinagre de Álcool	O
Am 99	Vinagre de Álcool	O
Am 100	Vinagre de Álcool	O
Am 101	Vinagre de Álcool	O
Am 102	Vinagre de Álcool	O
Am 103	Vinagre de Álcool	O
Am 104	Vinagre de Álcool	O
Am 105	Vinagre de Álcool	T
Am 106	Vinagre de Álcool	T
Am 107	Vinagre de Álcool	T
Am 108	Vinagre de Álcool	T
Am 109	Vinagre de Álcool	T

Am 110	Vinagre de Álcool	T
Am 111	Vinagre de Álcool	T
Am 112	Vinagre de Álcool	T
Am 113	Vinagre de Álcool	T
Am 114	Vinagre de Álcool	T
Am 115	Vinagre de Álcool	T
Am 116	Vinagre de Álcool	T
Am 117	Vinagre de Álcool	T
Am 118	Vinagre de Álcool	T
Am 119	Vinagre de Álcool	T
Am 120	Vinagre de Maçã	T
Am 121	Vinagre de Maçã	T
Am 122	Vinagre de Maçã	T
Am 123	Vinagre de Maçã	T
Am 124	Vinagre de Maçã	T
Am 125	Vinagre de Maçã	T
Am 126	Vinagre de Álcool	T
Am 127	Vinagre de Álcool	T
Am 128	Vinagre de Álcool	T
Am 129	Vinagre de Álcool	T
Am 130	Vinagre de Maçã	T
Am 131	Vinagre de Maçã	T
Am 132	Vinagre de Maçã	T
Am 133	Vinagre de Maçã	T
Am 134	Vinagre de Maçã	T
Am 135	Vinagre de Maçã	T
Am 136	Vinagre de Maçã	T
Am 137	Vinagre de Álcool	T
Am 138	Vinagre de Álcool	T

Am 139

Vinagre de Álcool

T

B. Conjunto AD – Amostras de vinagre orgânico de maçã, adulterados com vinagres convencionais de maçã

Amostra Original – Vinagre Orgânico de Maçã - (AM)	Adulterante – Vinagre Tradicional de Maçã (AM)	Nível de Adulteração	Amostra Adulterada Gerada (AD) /10 mL
AM 73	AM 19	Conc. 5% (0,5 mL)	AD 01
		Conc. 6% (0,6 mL)	AD 02
		Conc. 30% (3,0 mL)	AD 03
		Conc. 31% (3,1 mL)	AD 04
AM 76	AM 25	Conc. 5% (0,5 mL)	AD 05
		Conc. 6% (0,6 mL)	AD 06
		Conc. 30% (3,0 mL)	AD 07
		Conc. 31% (3,1 mL)	AD 08
AM 90	AM 34	Conc. 5% (0,5 mL)	AD 09
		Conc. 6% (0,6 mL)	AD 10
		Conc. 30% (3,0 mL)	AD 11
		Conc. 31% (3,1 mL)	AD 12
AM 96	AM 130	Conc. 5% (0,5 mL)	AD 13
		Conc. 6% (0,6 mL)	AD 14
		Conc. 30% (3,0 mL)	AD 15
		Conc. 31% (3,1 mL)	AD 16
AM 87	AM 80	Conc. 5% (0,5 mL)	AD 17

		Conc. 6% (0,6 mL)	AD 18
		Conc. 30% (3,0 mL)	AD 19
		Conc. 31% (3,1 mL)	AD 20
AM 89	AM 19	Conc. 5% (0,5 mL)	AD 21
		Conc. 6% (0,6 mL)	AD 22
		Conc. 30% (3,0 mL)	AD 23
		Conc. 31% (3,1 mL)	AD 24
AM 93	AM 25	Conc. 5% (0,5 mL)	AD 25
		Conc. 6% (0,6 mL)	AD 26
		Conc. 30% (3,0 mL)	AD 27
		Conc. 31% (3,1 mL)	AD 28
AM 75	AM 34	Conc. 5% (0,5 mL)	AD 29
		Conc. 6% (0,6 mL)	AD 30
		Conc. 30% (3,0 mL)	AD 31
		Conc. 31% (3,1 mL)	AD 32
AM 94	AM 130	Conc. 5% (0,5 mL)	AD 33
		Conc. 6% (0,6 mL)	AD 34
		Conc. 30% (3,0 mL)	AD 35
		Conc. 31% (3,1 mL)	AD 36
AM 85	AM 80	Conc. 5% (0,5 mL)	AD 37
		Conc. 6% (0,6 mL)	AD 38
		Conc. 30% (3,0 mL)	AD 39
		Conc. 31% (3,1 mL)	AD 40
AM 73	AM 25	Conc. 5% (0,5 mL)	AD 41
		Conc. 6% (0,6 mL)	AD 42
		Conc. 30% (3,0 mL)	AD 43

		Conc. 31% (3,1 mL)	AD 44
AM 76	AM 34	Conc. 5% (0,5 mL)	AD 45
		Conc. 6% (0,6 mL)	AD 46
		Conc. 30% (3,0 mL)	AD 47
		Conc. 31% (3,1 mL)	AD 48
AM 90	AM 130	Conc. 5% (0,5 mL)	AD 49
		Conc. 6% (0,6 mL)	AD 50
		Conc. 30% (3,0 mL)	AD 51
		Conc. 31% (3,1 mL)	AD 52
AM 96	AM 80	Conc. 5% (0,5 mL)	AD 53
		Conc. 6% (0,6 mL)	AD 54
		Conc. 30% (3,0 mL)	AD 55
		Conc. 31% (3,1 mL)	AD 56
AM 87	AM 19	Conc. 5% (0,5 mL)	AD 57
		Conc. 6% (0,6 mL)	AD 58
		Conc. 30% (3,0 mL)	AD 59
		Conc. 31% (3,1 mL)	AD 60
AM 89	AM 25	Conc. 5% (0,5 mL)	AD 61
		Conc. 6% (0,6 mL)	AD 62
		Conc. 30% (3,0 mL)	AD 63
		Conc. 31% (3,1 mL)	AD 64
AM 93	AM 34	Conc. 5% (0,5 mL)	AD 65
		Conc. 6% (0,6 mL)	AD 66
		Conc. 30% (3,0 mL)	AD 67
		Conc. 31% (3,1 mL)	AD 68

AM 75	AM 130	Conc. 5% (0,5 mL)	AD 69
		Conc. 6% (0,6 mL)	AD 70
		Conc. 30% (3,0 mL)	AD 71
		Conc. 31% (3,1 mL)	AD 72
AM 94	AM 80	Conc. 5% (0,5 mL)	AD 73
		Conc. 6% (0,6 mL)	AD 74
		Conc. 30% (3,0 mL)	AD 75
		Conc. 31% (3,1 mL)	AD 76
AM 85	AM 19	Conc. 5% (0,5 mL)	AD 77
		Conc. 6% (0,6 mL)	AD 78
		Conc. 30% (3,0 mL)	AD 79
		Conc. 31% (3,1 mL)	AD 80
AM 73	AM 34	Conc. 5% (0,5 mL)	AD 81
		Conc. 6% (0,6 mL)	AD 82
		Conc. 30% (3,0 mL)	AD 83
		Conc. 31% (3,1 mL)	AD 84
AM 76	AM 130	Conc. 5% (0,5 mL)	AD 85
		Conc. 6% (0,6 mL)	AD 86
		Conc. 30% (3,0 mL)	AD 87
		Conc. 31% (3,1 mL)	AD 88
AM 90	AM 80	Conc. 5% (0,5 mL)	AD 89
		Conc. 6% (0,6 mL)	AD 90
		Conc. 30% (3,0 mL)	AD 91
		Conc. 31% (3,1 mL)	AD 92
AM 96	AM 19	Conc. 5% (0,5 mL)	AD 93
		Conc. 6% (0,6 mL)	AD 94

		Conc. 30% (3,0 mL)	AD 95
		Conc. 31% (3,1 mL)	AD 96
AM 87	AM 25	Conc. 5% (0,5 mL)	AD 97
		Conc. 6% (0,6 mL)	AD 98
		Conc. 30% (3,0 mL)	AD 99
		Conc. 31% (3,1 mL)	AD 100
AM 89	AM 34	Conc. 5% (0,5 mL)	AD 101
		Conc. 6% (0,6 mL)	AD 102
		Conc. 30% (3,0 mL)	AD 103
		Conc. 31% (3,1 mL)	AD 104
AM 93	AM 130	Conc. 5% (0,5 mL)	AD 105
		Conc. 6% (0,6 mL)	AD 106
		Conc. 30% (3,0 mL)	AD 107
		Conc. 31% (3,1 mL)	AD 108
AM 75	AM 80	Conc. 5% (0,5 mL)	AD 109
		Conc. 6% (0,6 mL)	AD 110
		Conc. 30% (3,0 mL)	AD 111
		Conc. 31% (3,1 mL)	AD 112
AM 94	AM 19	Conc. 5% (0,5 mL)	AD 113
		Conc. 6% (0,6 mL)	AD 114
		Conc. 30% (3,0 mL)	AD 115
		Conc. 31% (3,1 mL)	AD 116
AM 85	AM 25	Conc. 5% (0,5 mL)	AD 117
		Conc. 6% (0,6 mL)	AD 118
		Conc. 30% (3,0 mL)	AD 119
		Conc. 31% (3,1 mL)	AD 120

AM 73	AM 130	Conc. 5% (0,5 mL)	AD 121
		Conc. 6% (0,6 mL)	AD 122
		Conc. 30% (3,0 mL)	AD 123
		Conc. 31% (3,1 mL)	AD 124
AM 76	AM 80	Conc. 5% (0,5 mL)	AD 125
		Conc. 6% (0,6 mL)	AD 126
		Conc. 30% (3,0 mL)	AD 127
		Conc. 31% (3,1 mL)	AD 128
AM 90	AM 19	Conc. 5% (0,5 mL)	AD 129
		Conc. 6% (0,6 mL)	AD 130
		Conc. 30% (3,0 mL)	AD 131
		Conc. 31% (3,1 mL)	AD 132
AM 96	AM 25	Conc. 5% (0,5 mL)	AD 133
		Conc. 6% (0,6 mL)	AD 134
		Conc. 30% (3,0 mL)	AD 135
		Conc. 31% (3,1 mL)	AD 136
AM 87	AM 34	Conc. 5% (0,5 mL)	AD 137
		Conc. 6% (0,6 mL)	AD 138
		Conc. 30% (3,0 mL)	AD 139
		Conc. 31% (3,1 mL)	AD 140
AM 89	AM 130	Conc. 5% (0,5 mL)	AD 141
		Conc. 6% (0,6 mL)	AD 142
		Conc. 30% (3,0 mL)	AD 143
		Conc. 31% (3,1 mL)	AD 144
AM 93	AM 80	Conc. 5% (0,5 mL)	AD 145

		Conc. 6% (0,6 mL)	AD 146
		Conc. 30% (3,0 mL)	AD 147
		Conc. 31% (3,1 mL)	AD 148
AM 75	AM 19	Conc. 5% (0,5 mL)	AD 149
		Conc. 6% (0,6 mL)	AD 150
		Conc. 30% (3,0 mL)	AD 151
		Conc. 31% (3,1 mL)	AD 152
AM 94	AM 25	Conc. 5% (0,5 mL)	AD 153
		Conc. 6% (0,6 mL)	AD 154
		Conc. 30% (3,0 mL)	AD 155
		Conc. 31% (3,1 mL)	AD 156
AM 85	AM 34	Conc. 5% (0,5 mL)	AD 157
		Conc. 6% (0,6 mL)	AD 158
		Conc. 30% (3,0 mL)	AD 159
		Conc. 31% (3,1 mL)	AD 160
AM 73	AM 80	Conc. 5% (0,5 mL)	AD 161
		Conc. 6% (0,6 mL)	AD 162
		Conc. 30% (3,0 mL)	AD 163
		Conc. 31% (3,1 mL)	AD 164
AM 76	AM 19	Conc. 5% (0,5 mL)	AD 165
		Conc. 6% (0,6 mL)	AD 166
		Conc. 30% (3,0 mL)	AD 167
		Conc. 31% (3,1 mL)	AD 168
AM 90	AM 25	Conc. 5% (0,5 mL)	AD 169
		Conc. 6% (0,6 mL)	AD 170
		Conc. 30% (3,0 mL)	AD 170

		Conc. 31% (3,1 mL)	AD 172
AM 96	AM 34	Conc. 5% (0,5 mL)	AD 173
		Conc. 6% (0,6 mL)	AD 174
		Conc. 30% (3,0 mL)	AD 175
		Conc. 31% (3,1 mL)	AD 176
AM 87	AM 130	Conc. 5% (0,5 mL)	AD 177
		Conc. 6% (0,6 mL)	AD 178
		Conc. 30% (3,0 mL)	AD 179
		Conc. 31% (3,1 mL)	AD 180
AM 89	AM 80	Conc. 5% (0,5 mL)	AD 181
		Conc. 6% (0,6 mL)	AD 182
		Conc. 30% (3,0 mL)	AD 183
		Conc. 31% (3,1 mL)	AD 184
AM 93	AM 19	Conc. 5% (0,5 mL)	AD 185
		Conc. 6% (0,6 mL)	AD 186
		Conc. 30% (3,0 mL)	AD 187
		Conc. 31% (3,1 mL)	AD 188
AM 75	AM 25	Conc. 5% (0,5 mL)	AD 189
		Conc. 6% (0,6 mL)	AD 190
		Conc. 30% (3,0 mL)	AD 191
		Conc. 31% (3,1 mL)	AD 192
AM 94	AM 34	Conc. 5% (0,5 mL)	AD 193
		Conc. 6% (0,6 mL)	AD 194
		Conc. 30% (3,0 mL)	AD 195
		Conc. 31% (3,1 mL)	AD 196

AM 85	AM 130	Conc. 5% (0,5 mL)	AD 197
		Conc. 6% (0,6 mL)	AD 198
		Conc. 30% (3,0 mL)	AD 199
		Conc. 31% (3,1 mL)	AD 200

C. Conjunto MA – Amostras de vinagre de maçã, adulterados com vinagres de álcool

Amostra Original – Vinagre de Maçã (AM)	Adulterante – Vinagre de Álcool (AM)	Nível de Adulteração	Amostra Adulterada Gerada (MA) /10 mL
AM 80	AM 103	Conc. 5% (0,5 mL)	MA 01
		Conc. 15% (1,5 mL)	MA 02
		Conc. 25% (2,5 mL)	MA 03
AM 73	AM 110	Conc. 5% (0,5 mL)	MA 04
		Conc. 15% (1,5 mL)	MA 05
		Conc. 25% (2,5 mL)	MA 06
AM 85	AM 13	Conc. 5% (0,5 mL)	MA 07
		Conc. 15% (1,5 mL)	MA 08
		Conc. 25% (2,5 mL)	MA 09
AM 78	AM 28	Conc. 5% (0,5 mL)	MA 10
		Conc. 15% (1,5 mL)	MA 11
		Conc. 25% (2,5 mL)	MA 12
AM 77	AM 43	Conc. 5% (0,5 mL)	MA 13
		Conc. 15% (1,5 mL)	MA 14

		Conc. 25% (2,5 mL)	MA 15
AM 88	AM 103	Conc. 5% (0,5 mL)	MA 16
		Conc. 15% (1,5 mL)	MA 17
		Conc. 25% (2,5 mL)	MA 18
AM 74	AM 110	Conc. 5% (0,5 mL)	MA 19
		Conc. 15% (1,5 mL)	MA 20
		Conc. 25% (2,5 mL)	MA 21
AM 91	AM 13	Conc. 5% (0,5 mL)	MA 22
		Conc. 15% (1,5 mL)	MA 23
		Conc. 25% (2,5 mL)	MA 24
AM 95	AM 28	Conc. 5% (0,5 mL)	MA 25
		Conc. 15% (1,5 mL)	MA 26
		Conc. 25% (2,5 mL)	MA 27
AM 82	AM 43	Conc. 5% (0,5 mL)	MA 28
		Conc. 15% (1,5 mL)	MA 29
		Conc. 25% (2,5 mL)	MA 30
AM 92	AM 103	Conc. 5% (0,5 mL)	MA 31
		Conc. 15% (1,5 mL)	MA 32
		Conc. 25% (2,5 mL)	MA 33
AM 86	AM 110	Conc. 5% (0,5 mL)	MA 34
		Conc. 15% (1,5 mL)	MA 35
		Conc. 25% (2,5 mL)	MA 36
AM 84	AM 13	Conc. 5% (0,5 mL)	MA 37
		Conc. 15% (1,5 mL)	MA 38

		Conc. 25% (2,5 mL)	MA 39
AM 79	AM 28	Conc. 5% (0,5 mL)	MA 40
		Conc. 15% (1,5 mL)	MA 41
		Conc. 25% (2,5 mL)	MA 42
AM 81	AM 43	Conc. 5% (0,5 mL)	MA 43
		Conc. 15% (1,5 mL)	MA 44
		Conc. 25% (2,5 mL)	MA 45
AM 80	AM 110	Conc. 5% (0,5 mL)	MA 46
		Conc. 15% (1,5 mL)	MA 47
		Conc. 25% (2,5 mL)	MA 48
AM 73	AM 13	Conc. 5% (0,5 mL)	MA 49
		Conc. 15% (1,5 mL)	MA 50
		Conc. 25% (2,5 mL)	MA 51
AM 85	AM 28	Conc. 5% (0,5 mL)	MA 52
		Conc. 15% (1,5 mL)	MA 53
		Conc. 25% (2,5 mL)	MA 54
AM 78	AM 43	Conc. 5% (0,5 mL)	MA 55
		Conc. 15% (1,5 mL)	MA 56
		Conc. 25% (2,5 mL)	MA 57
AM 77	AM 103	Conc. 5% (0,5 mL)	MA 58
		Conc. 15% (1,5 mL)	MA 59
		Conc. 25% (2,5 mL)	MA 60
AM 88	AM 110	Conc. 5% (0,5 mL)	MA 61
		Conc. 15% (1,5 mL)	MA 62

		Conc. 25% (2,5 mL)	MA 63
AM 74	AM 13	Conc. 5% (0,5 mL)	MA 64
		Conc. 15% (1,5 mL)	MA 65
		Conc. 25% (2,5 mL)	MA 66
AM 91	AM 28	Conc. 5% (0,5 mL)	MA 67
		Conc. 15% (1,5 mL)	MA 68
		Conc. 25% (2,5 mL)	MA 69
AM 95	AM 43	Conc. 5% (0,5 mL)	MA 70
		Conc. 15% (1,5 mL)	MA 71
		Conc. 25% (2,5 mL)	MA 72
AM 82	AM 103	Conc. 5% (0,5 mL)	MA 73
		Conc. 15% (1,5 mL)	MA 74
		Conc. 25% (2,5 mL)	MA 75
AM 92	AM 110	Conc. 5% (0,5 mL)	MA 76
		Conc. 15% (1,5 mL)	MA 77
		Conc. 25% (2,5 mL)	MA 78
AM 86	AM 13	Conc. 5% (0,5 mL)	MA 79
		Conc. 15% (1,5 mL)	MA 80
		Conc. 25% (2,5 mL)	MA 81
AM 84	AM 28	Conc. 5% (0,5 mL)	MA 82
		Conc. 15% (1,5 mL)	MA 83
		Conc. 25% (2,5 mL)	MA 84
AM 79	AM 43	Conc. 5% (0,5 mL)	MA 85
		Conc. 15% (1,5 mL)	MA 86

		Conc. 25% (2,5 mL)	MA 87
AM 81	AM 103	Conc. 5% (0,5 mL)	MA 88
		Conc. 15% (1,5 mL)	MA 89
		Conc. 25% (2,5 mL)	MA 90
AM 80	AM 13	Conc. 5% (0,5 mL)	MA 91
		Conc. 15% (1,5 mL)	MA 92
		Conc. 25% (2,5 mL)	MA 93
AM 73	AM 28	Conc. 5% (0,5 mL)	MA 94
		Conc. 15% (1,5 mL)	MA 95
		Conc. 25% (2,5 mL)	MA 96
AM 85	AM 43	Conc. 5% (0,5 mL)	MA 97
		Conc. 15% (1,5 mL)	MA 98
		Conc. 25% (2,5 mL)	MA 99
AM 78	AM 103	Conc. 5% (0,5 mL)	MA 100
		Conc. 15% (1,5 mL)	MA 101
		Conc. 25% (2,5 mL)	MA 102
AM 77	AM 110	Conc. 5% (0,5 mL)	MA 103
		Conc. 15% (1,5 mL)	MA 104
		Conc. 25% (2,5 mL)	MA 105
AM 88	AM 13	Conc. 5% (0,5 mL)	MA 106
		Conc. 15% (1,5 mL)	MA 107
		Conc. 25% (2,5 mL)	MA 108
AM 74	AM 28	Conc. 5% (0,5 mL)	MA 109
		Conc. 15% (1,5 mL)	MA 110

		Conc. 25% (2,5 mL)	MA 111
AM 91	AM 43	Conc. 5% (0,5 mL)	MA 112
		Conc. 15% (1,5 mL)	MA 113
		Conc. 25% (2,5 mL)	MA 114
AM 95	AM 103	Conc. 5% (0,5 mL)	MA 115
		Conc. 15% (1,5 mL)	MA 116
		Conc. 25% (2,5 mL)	MA 117
AM 82	AM 110	Conc. 5% (0,5 mL)	MA 118
		Conc. 15% (1,5 mL)	MA 119
		Conc. 25% (2,5 mL)	MA 120
AM 92	AM 13	Conc. 5% (0,5 mL)	MA 121
		Conc. 15% (1,5 mL)	MA 122
		Conc. 25% (2,5 mL)	MA 123
AM 86	AM 28	Conc. 5% (0,5 mL)	MA 124
		Conc. 15% (1,5 mL)	MA 125
		Conc. 25% (2,5 mL)	MA 126
AM 84	AM 43	Conc. 5% (0,5 mL)	MA 127
		Conc. 15% (1,5 mL)	MA 128
		Conc. 25% (2,5 mL)	MA 129
AM 79	AM 103	Conc. 5% (0,5 mL)	MA 130
		Conc. 15% (1,5 mL)	MA 131
		Conc. 25% (2,5 mL)	MA 132
AM 81	AM 110	Conc. 5% (0,5 mL)	MA 133
		Conc. 15% (1,5 mL)	MA 134

		Conc. 25% (2,5 mL)	MA 135
AM 80	AM 28	Conc. 5% (0,5 mL)	MA 136
		Conc. 15% (1,5 mL)	MA 137
		Conc. 25% (2,5 mL)	MA 138
AM 73	AM 43	Conc. 5% (0,5 mL)	MA 139
		Conc. 15% (1,5 mL)	MA 140
		Conc. 25% (2,5 mL)	MA 141
AM 85	AM 103	Conc. 5% (0,5 mL)	MA 142
		Conc. 15% (1,5 mL)	MA 143
		Conc. 25% (2,5 mL)	MA 144
AM 78	AM 110	Conc. 5% (0,5 mL)	MA 145
		Conc. 15% (1,5 mL)	MA 146
		Conc. 25% (2,5 mL)	MA 147
AM 77	AM 13	Conc. 5% (0,5 mL)	MA 148
		Conc. 15% (1,5 mL)	MA 149
		Conc. 25% (2,5 mL)	MA 150
AM 88	AM 28	Conc. 5% (0,5 mL)	MA 151
		Conc. 15% (1,5 mL)	MA 152
		Conc. 25% (2,5 mL)	MA 153
AM 74	AM 43	Conc. 5% (0,5 mL)	MA 154
		Conc. 15% (1,5 mL)	MA 155
		Conc. 25% (2,5 mL)	MA 156
AM 91	AM 103	Conc. 5% (0,5 mL)	MA 157
		Conc. 15% (1,5 mL)	MA 158

		Conc. 25% (2,5 mL)	MA 159
AM 95	AM 110	Conc. 5% (0,5 mL)	MA 160
		Conc. 15% (1,5 mL)	MA 161
		Conc. 25% (2,5 mL)	MA 162
AM 82	AM 13	Conc. 5% (0,5 mL)	MA 163
		Conc. 15% (1,5 mL)	MA 164
		Conc. 25% (2,5 mL)	MA 165
AM 92	AM 28	Conc. 5% (0,5 mL)	MA 166
		Conc. 15% (1,5 mL)	MA 167
		Conc. 25% (2,5 mL)	MA 168
AM 86	AM 43	Conc. 5% (0,5 mL)	MA 169
		Conc. 15% (1,5 mL)	MA 170
		Conc. 25% (2,5 mL)	MA 171
AM 84	AM 103	Conc. 5% (0,5 mL)	MA 172
		Conc. 15% (1,5 mL)	MA 173
		Conc. 25% (2,5 mL)	MA 174
AM 79	AM 110	Conc. 5% (0,5 mL)	MA 175
		Conc. 15% (1,5 mL)	MA 176
		Conc. 25% (2,5 mL)	MA 177
AM 81	AM 13	Conc. 5% (0,5 mL)	MA 178
		Conc. 15% (1,5 mL)	MA 179
		Conc. 25% (2,5 mL)	MA 180
AM 80	AM 43	Conc. 5% (0,5 mL)	MA 181
		Conc. 15% (1,5 mL)	MA 182

		Conc. 25% (2,5 mL)	MA 183
AM 73	AM 103	Conc. 5% (0,5 mL)	MA 184
		Conc. 15% (1,5 mL)	MA 185
		Conc. 25% (2,5 mL)	MA 186
AM 85	AM 110	Conc. 5% (0,5 mL)	MA 187
		Conc. 15% (1,5 mL)	MA 188
		Conc. 25% (2,5 mL)	MA 189
AM 78	AM 13	Conc. 5% (0,5 mL)	MA 190
		Conc. 15% (1,5 mL)	MA 191
		Conc. 25% (2,5 mL)	MA 192
AM 77	AM 28	Conc. 5% (0,5 mL)	MA 193
		Conc. 15% (1,5 mL)	MA 194
		Conc. 25% (2,5 mL)	MA 195
AM 88	AM 43	Conc. 5% (0,5 mL)	MA 196
		Conc. 15% (1,5 mL)	MA 197
		Conc. 25% (2,5 mL)	MA 198
AM 74	AM 103	Conc. 5% (0,5 mL)	MA 199
		Conc. 15% (1,5 mL)	MA 200
		Conc. 25% (2,5 mL)	MA 201
AM 91	AM 110	Conc. 5% (0,5 mL)	MA 202
		Conc. 15% (1,5 mL)	MA 203
		Conc. 25% (2,5 mL)	MA 204
AM 95	AM 13	Conc. 5% (0,5 mL)	MA 205
		Conc. 15% (1,5 mL)	MA 206

		Conc. 25% (2,5 mL)	MA 207
AM 82	AM 28	Conc. 5% (0,5 mL)	MA 208
		Conc. 15% (1,5 mL)	MA 209
		Conc. 25% (2,5 mL)	MA 210
AM 92	AM 43	Conc. 5% (0,5 mL)	MA 211
		Conc. 15% (1,5 mL)	MA 212
		Conc. 25% (2,5 mL)	MA 213
AM 86	AM 103	Conc. 5% (0,5 mL)	MA 214
		Conc. 15% (1,5 mL)	MA 215
		Conc. 25% (2,5 mL)	MA 216
AM 84	AM 110	Conc. 5% (0,5 mL)	MA 217
		Conc. 15% (1,5 mL)	MA 218
		Conc. 25% (2,5 mL)	MA 219
AM 79	AM 13	Conc. 5% (0,5 mL)	MA 220
		Conc. 15% (1,5 mL)	MA 221
		Conc. 25% (2,5 mL)	MA 222
AM 81	AM 28	Conc. 5% (0,5 mL)	MA 223
		Conc. 15% (1,5 mL)	MA 224
		Conc. 25% (2,5 mL)	MA 225

CAPÍTULO III

ARTIGO CIENTÍFICO 2

Spectral Model eXplainer (SMX): Inteligência Artificial Explicável na Classificação de Vinagres Brasileiros de Maçã por RMN de ^1H de Baixo Campo

Spectral Model eXplainer (SMX): Inteligência Artificial Explicável na Classificação de Vinagres Brasileiros de Maçã por RMN de ^1H de Baixo Campo

Vinícius Avanzi Barbosa Mascareli^{a,1}, Diego Galvan^b, Jelmir Craveiro de Andrade^{c,d}, Ingrid Alves de Moraes^e, Viviane Lopes Leite da Costa^a, Carlos Adam Conte-Junior^c, Sylvio Barbon Junior^f, Wilma Aparecida Spinosa^a

^a Departamento de Ciência e Tecnologia de Alimentos, Universidade Estadual de Londrina (UEL), Londrina, PR 86.057-970, Brasil.

^b Departamento de Química, Universidade Federal de Santa Catarina (UFSC), Florianópolis, SC, 88.040-900, Brasil.

^c Centro Laboratorial Analítico e Molecular (CLAM), Instituto de Química (IQ), Universidade Federal do Rio de Janeiro (UFRJ), Rio de Janeiro, RJ 21.941-909, Brasil.

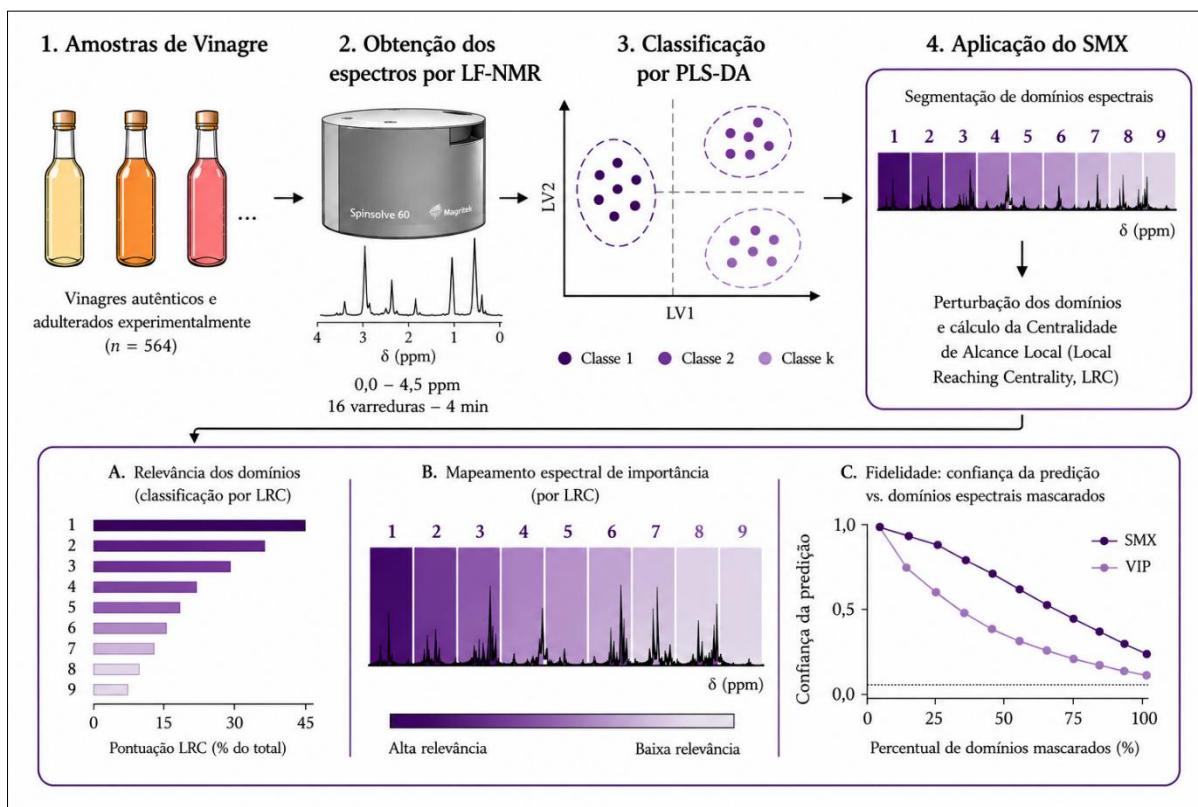
^d Núcleo de Análise de Alimentos (NAL), Laboratório de Apoio ao Desenvolvimento Tecnológico (LADETEC), Universidade Federal do Rio de Janeiro (UFRJ), Rio de Janeiro, RJ 21.941-598, Brasil.

^e Departamento de Engenharia de Alimentos, Escola de Engenharia de Alimentos, Universidade Federal de Goiás (UFG), Goiás, GO, Brasil.

^f Departamento de Engenharia e Arquitetura, Universidade de Trieste (UNITS), 34127 Trieste, Itália.

¹Autor correspondente: *V. A. B. Mascareli. E-mail: vini.mascareli@gmail.com ou vini.mascareli@uel.br ORCID: <https://orcid.org/0000-0002-8327-8790>. Departamento de Ciência e Tecnologia de Alimentos, Universidade Estadual de Londrina, Londrina, Brasil.

Resumo Gráfico



Resumo

A interpretabilidade permanece como um dos principais desafios para a adoção de modelos de *Machine Learning* na classificação e autenticação de alimentos baseadas em espectroscopia. Neste estudo, o *Spectral Model eXplainer* (SMX), uma estrutura de inteligência artificial explicável desenvolvida especificamente para dados espectroscópicos, foi aplicado a *fingerprints* obtidos por ressonância magnética nuclear de prótons de baixo campo (^1H LF-NMR) de vinagres comerciais brasileiros. Um total de 564 amostras autênticas e adulteradas experimentalmente foi avaliado em dois cenários de adulteração: vinagre de maçã orgânico adulterado com vinagre de maçã convencional e vinagre de maçã adulterado com vinagre de álcool. Modelos de autenticação foram desenvolvidos por Análise Discriminante por Mínimos Quadrados Parciais (PLS-DA) e, posteriormente, interpretados por meio do SMX, que quantificou a contribuição de domínios espectrais marcados por metabólitos utilizando análises

baseadas em perturbação e a métrica *Local Reaching Centrality* (LRC). O método revelou perfis de relevância distintos para os dois cenários de autenticação, destacando regiões espectrais associadas a metabólitos derivados da fermentação, ácidos orgânicos, compostos relacionados ao etanol e carboidratos residuais. As análises de fidelidade baseadas em perturbação demonstraram elevada concordância entre as explicações geradas e o comportamento dos classificadores, enquanto a comparação com a métrica *Variable Importance in Projection* (VIP) evidenciou que o SMX identificou de forma mais fiel os domínios espectrais responsáveis pelas decisões dos modelos. Mais do que ampliar a interpretabilidade, o SMX estabeleceu uma conexão direta entre o comportamento preditivo e regiões espectrais quimicamente significativas, fornecendo evidências mecanísticas sobre a base composicional da autenticidade dos vinagres. Esses resultados demonstram o potencial de abordagens de inteligência artificial explicável desenvolvidas especificamente para dados espectroscópicos em integrar modelagem preditiva e interpretação química, contribuindo para o desenvolvimento de estratégias de aprendizado de máquina mais transparentes, confiáveis e cientificamente fundamentadas para autenticação de alimentos.

Palavras-chave: *Machine Learning*; técnicas espectroscópicas; *fingerprinting*; PLS-DA; Autenticação de alimentos; Quimiometria; Interpretabilidade de modelos.

1. Introdução

A crescente demanda por alimentos e bebidas de maior valor agregado, incluindo produtos associados a certificações de origem, sistemas de produção orgânica e alegações relacionadas à saúde, tem ampliado a necessidade de ferramentas analíticas capazes de caracterizar de forma abrangente sua composição química e fornecer suporte confiável à avaliação da qualidade, autenticidade e integridade dos produtos (Lozano-Castellón et al., 2024;

Vinothkanna et al., 2024; Zhang et al., 2024).

Embora análises físico-químicas direcionadas e a quantificação de marcadores específicos permaneçam essenciais para diversas aplicações regulatórias e de controle de qualidade, essas abordagens frequentemente oferecem apenas uma visão parcial de matrizes alimentares complexas. Como alternativa, metodologias analíticas não direcionadas têm se consolidado na ciência de alimentos por explorarem simultaneamente um grande número de variáveis químicas presentes em uma única aquisição analítica. Nesse contexto, estratégias de *fingerprinting* associadas à quimiometria e ao *machine learning* possibilitam transformar espectros complexos em assinaturas químicas capazes de capturar diferenças sutis entre amostras, tornando-se ferramentas cada vez mais relevantes para classificação, caracterização, monitoramento da qualidade e estudos de autenticidade (Augusti et al., 2024; García-Pérez et al., 2024; Goyal et al., 2024; Mialon et al., 2023; Vinothkanna et al., 2024).

Entre as técnicas de *fingerprinting*, a espectroscopia por ressonância magnética nuclear (RMN) destaca-se pela capacidade de fornecer, de forma simultânea e não destrutiva, informações sobre uma ampla diversidade de constituintes moleculares, demandando preparo mínimo da amostra (Ciampa et al., 2022; Consonni et al., 2008; Grassi et al., 2024; Lolli & Caligiani, 2024; Tang et al., 2024).

Nos últimos anos, a disponibilidade de equipamentos de ressonância magnética nuclear de baixo campo (LF-NMR), baseados em ímãs permanentes compactos, ampliou significativamente a aplicação dessa tecnologia em laboratórios de pesquisa e controle de qualidade, impulsionada pela redução dos custos de aquisição e manutenção, simplificação operacional e eliminação da infraestrutura criogênica necessária aos sistemas de alto campo (Lolli & Caligiani, 2024; Michal, 2020; Yang et al., 2023). Embora a menor intensidade do campo magnético resulte em redução da resolução espectral, diversos estudos demonstram que os *fingerprints* obtidos por LF-NMR preservam informação química suficiente para sustentar

modelos robustos de classificação, autenticação e avaliação da qualidade quando combinados a métodos modernos de análise multivariada e *machine learning* (Conte et al., 2025; Galvan et al., 2021; Lolli & Caligiani, 2024; Michal, 2020; Tang et al., 2024).

Paralelamente ao avanço das técnicas analíticas, o *machine learning* tornou-se uma das principais ferramentas para exploração de conjuntos de dados espectroscópicos de alta dimensionalidade. Algoritmos supervisionados, como a Análise Discriminante por Mínimos Quadrados Parciais (PLS-DA), Máquinas de Vetores de Suporte (SVM), *Random Forest* (RF) e abordagens correlatas, vêm apresentando elevado desempenho em diferentes aplicações envolvendo classificação, autenticação e detecção de adulterações em alimentos e bebidas (Augusti et al., 2024; da Silva et al., 2025; Dimakopoulou-Papazoglou et al., 2025; Goyal et al., 2024; Shih et al., 2024; Vinothkanna et al., 2024).

Entretanto, o aumento da capacidade preditiva desses modelos foi acompanhado por uma redução de sua interpretabilidade. Em aplicações científicas e industriais, a elevada acurácia não é suficiente para justificar a adoção de um modelo quando não há compreensão sobre quais características químicas fundamentam suas decisões. Consequentemente, pesquisadores, laboratórios de controle de qualidade e órgãos reguladores têm demandado modelos capazes não apenas de produzir previsões confiáveis, mas também de fornecer explicações consistentes sobre as regiões espectrais e os padrões composicionais responsáveis pelo processo decisório (Ahmed et al., 2025; Arrighi et al., 2026; Yang et al., 2024).

Nesse contexto, a Inteligência Artificial Explicável (*eXplainable Artificial Intelligence* – XAI) emergiu como uma área dedicada ao desenvolvimento de métodos que aumentem a transparência e a interpretabilidade de modelos de aprendizado de máquina (Ahmed et al., 2025; Arrighi et al., 2026). No entanto, a maioria das abordagens de explicabilidade disponíveis foi originalmente concebida para dados tabulares, imagens ou processamento de linguagem natural, não considerando as características particulares dos dados espectroscópicos, marcados

por elevada correlação entre variáveis, continuidade dos sinais e forte organização em domínios espectrais. De maneira semelhante, ferramentas quimiométricas tradicionalmente empregadas para interpretação de modelos, como coeficientes de regressão, vetores de carga (*loadings*) e a métrica *Variable Importance in Projection* (VIP), frequentemente distribuem a relevância entre centenas de variáveis individuais, dificultando a associação entre a resposta do modelo e regiões espectrais quimicamente interpretáveis (Arrighi et al., 2024, 2026; Biancolillo et al., 2020; Ribeiro et al., 2026).

Essas limitações tornam-se ainda mais evidentes em espectros obtidos por LF-NMR, nos quais a menor resolução favorece a sobreposição de sinais e reforça a necessidade de métodos de explicabilidade desenvolvidos especificamente para esse tipo de informação analítica (Arrighi et al., 2024, 2026; Biancolillo et al., 2020; Ribeiro et al., 2026).

Apesar do crescente interesse pela aplicação de XAI em química analítica, estudos voltados à interpretação de modelos construídos a partir de *fingerprints* de LF-NMR ainda permanecem escassos (Arrighi et al., 2026; Parker et al., 2014; Ribeiro et al., 2026). Até onde se tem conhecimento, nenhum trabalho anterior investigou de forma sistemática a explicabilidade espectral em modelos de autenticação de vinagres comerciais utilizando essa plataforma analítica.

Entre os diferentes produtos passíveis de investigação por abordagens não direcionadas, os vinagres orgânico e convencional de maçã representam modelos de estudo particularmente relevantes devido às suas elevadas variabilidades composicionais, crescente demanda por produtos orgânicos, apelo a saúde e importância comercial. Nessas condições, abordagens baseadas em *fingerprinting* apresentam-se especialmente promissoras por avaliarem a composição química global do produto, em vez de dependerem exclusivamente de marcadores previamente definidos (Cavdaroglu & Ozen, 2023; Ousaaid et al., 2020; Štornik et al., 2016).

Para enfrentar essa lacuna, o presente estudo investiga a aplicabilidade do *Spectral*

Model eXplainer (SMX), uma estrutura de explicabilidade recentemente proposta e desenvolvida especificamente para conjuntos de dados espectroscópicos. Diferentemente das abordagens convencionais, o SMX opera por meio da perturbação sistemática de regiões espectrais previamente definidas utilizando estratégias de mascaramento, permitindo quantificar a contribuição de domínios marcados por metabólitos para a confiança e o processo decisório dos modelos (Ribeiro et al., 2026).

Dessa forma, o objetivo deste estudo foi avançar a aplicação da Inteligência Artificial Explicável na autenticação de alimentos por meio da utilização do *Spectral Model eXplainer* (SMX) em *fingerprints* obtidos por RMN de ^1H de baixo campo de vinagres comerciais brasileiros de maçã. Além de ampliar a interpretabilidade dos modelos, a abordagem estabelece uma conexão direta entre desempenho preditivo e interpretação química, oferecendo uma estratégia de explicabilidade desenvolvida especificamente para dados espectroscópicos e contribuindo para o desenvolvimento de sistemas analíticos mais transparentes, confiáveis e cientificamente fundamentados para a autenticação de alimentos e bebidas.

2. Material e Métodos

2.1. Amostras e Sessão Experimental

Foi utilizado um conjunto diversificado de vinagres contendo 139 amostras provenientes de diferentes marcas, lotes de produção e sistemas produtivos, incluindo vinagre de maçã orgânico certificado, vinagre de maçã convencional e vinagre de álcool. As amostras foram obtidas diretamente da indústria ou adquiridas em estabelecimentos comerciais localizados em diferentes regiões do Brasil. Com base na certificação e nas informações de rotulagem, 51 amostras foram classificadas como orgânicas e 88 como vinagres produzidos por sistemas convencionais.

Todas as amostras foram transportadas para a Universidade Estadual de Londrina e armazenadas em suas embalagens originais, lacradas, à temperatura de $\sim 4^{\circ}\text{C}$ e protegidas da exposição direta à luz até a realização das análises. Essas condições de armazenamento foram adotadas para preservar a integridade das amostras e minimizar variações decorrentes do armazenamento.

Para simular cenários realistas de adulteração motivada economicamente, foram estabelecidos dois esquemas independentes de adulteração. O primeiro consistiu na adulteração de vinagre de maçã orgânico com vinagre de maçã convencional, enquanto o segundo simulou a adição de vinagre de álcool, de menor valor comercial, ao vinagre de maçã. Esses cenários foram selecionados por representarem práticas de adulteração economicamente motivadas capazes de promover alterações composicionais de difícil detecção por métodos analíticos convencionais (Avanzi Barbosa Mascareli et al., 2023; Cavdaroglu & Ozen, 2023).

No primeiro cenário, dez amostras de vinagre de maçã orgânico provenientes de diferentes lotes de produção foram individualmente misturadas com cinco amostras de vinagre de maçã convencional nos níveis de adulteração de 5, 6, 30 e 31% (v/v) baseados nos limites descritos pela legislação, resultando em 200 amostras adulteradas (AD) (Brasil, Ministério da Agricultura, Pecuária e Abastecimento, 2009). No segundo cenário, quinze amostras de vinagre de maçã foram misturadas com cinco amostras de vinagre de álcool nos níveis de 5, 15 e 25% (v/v), gerando 225 amostras adulteradas (MA). Todas as misturas foram preparadas utilizando micropipetas calibradas, de forma a garantir a precisão volumétrica.

O conjunto de dados final, apresentado na Tabela 1, foi composto por 564 amostras, incluindo vinagres autênticos e adulterados experimentalmente. A partir desse conjunto, foram definidos dois problemas binários de autenticação: a discriminação entre vinagre de maçã orgânico autêntico e vinagre de maçã orgânico adulterado com vinagre de maçã convencional (AM vs. AD) e a discriminação entre vinagre de maçã autêntico e vinagre de maçã adulterado

com vinagre de álcool (AM vs. MA). Esses conjuntos de dados foram posteriormente empregados nas análises de classificação supervisionada e de explicabilidade dos modelos.

Tabela 1. Composição do conjunto de dados de vinagres utilizado nas análises de autenticação e explicabilidade

Classe de amostra	Quantidade
Classe AM (Vinagres Originais)	139
Classe AD	200
Classe MA	225
Total	564

2.2. Aquisição e Pré-processamento de *Fingerprints* Espectrais

A ^1H LF-NMR foi empregada para gerar *fingerprints* espectrais representativos da composição química global das amostras de vinagre (Conte et al., 2025; Galvan et al., 2023; Lolli & Caligiani, 2024; Michal, 2020). Para cada análise, 500 μL de amostra foram transferidos diretamente para um tubo de RMN de 5 mm, sem diluição, extração ou adição de solventes deuterados. Antes da aquisição espectral, as amostras foram equilibradas por 10 min em uma câmara com temperatura controlada, assegurando a estabilização térmica na faixa operacional do equipamento (26,5–28,5 $^{\circ}\text{C}$).

Os espectros foram adquiridos em um espectrômetro de RMN de bancada Spinsolve 60 MHz (Magritek GmbH, Aachen, Alemanha), equipado com magneto permanente e operando a 60 MHz para ^1H . As aquisições foram realizadas utilizando a sequência de pulsos 1D PROTON CDEC, com pulso de excitação de 90° . Para cada amostra, foram coletadas 16 varreduras (scans), empregando tempo de aquisição de 3,2 s e intervalo entre varreduras de 15 s, resultando

em tempo total de aproximadamente 4 min por espectro. Os dados foram registrados com largura espectral de 328 ppm e frequência central de 110 ppm.

Após a aquisição, os espectros foram processados no software TopSpin (Bruker BioSpin, Alemanha) e exportados em forma numérica para as análises computacionais subsequentes nas etapas de classificação supervisionada e nas análises de explicabilidade descritas a seguir.

2.3. Processamento Espectral e Preparação dos Dados

A região espectral compreendida entre 0,0 e 4,5 ppm foi selecionada para as análises subsequentes por concentrar os sinais de maior intensidade e relevância química presentes nos vinagres, abrangendo ressonâncias atribuídas aos metabólitos derivados da fermentação (Avanzi Barbosa Mascareli et al., 2023; Grassi et al., 2024). A restrição da análise a esse intervalo reduziu a influência de regiões espectrais com baixo conteúdo informativo, preservando, ao mesmo tempo, a variabilidade composicional relevante para os estudos de autenticação (Grassi et al., 2024).

Potenciais *outliers* foram identificados por meio do algoritmo não supervisionado *Isolation Forest*, aplicado ao intervalo espectral selecionado. As amostras classificadas como anômalas foram excluídas das análises subsequentes com o objetivo de aumentar a estabilidade dos modelos e a confiabilidade das análises de explicabilidade (Ceschin et al., 2026).

O alinhamento espectral foi adotado como estratégia de pré-processamento para as análises sequentes. Após o pré-processamento, as variáveis espectrais foram padronizadas por meio da normalização *z-score* antes do desenvolvimento dos modelos e das análises de explicabilidade.

Todas as análises foram realizadas utilizando o conjunto de dados espectrais

previamente alinhado e padronizado.

2.4. Modelo de Classificação para as Análises de Explicabilidade

Para investigar os determinantes espectrais subjacentes à autenticação dos vinagres, foram desenvolvidos modelos de classificação supervisionada utilizando a Análise Discriminante por Mínimos Quadrados Parciais (PLS-DA). Os espectros previamente alinhados e padronizados foram empregados como variáveis preditoras, enquanto as classes binárias foram definidas de acordo com cada cenário de autenticação (AM vs. AD e AM vs. MA).

A PLS-DA foi selecionada devido à sua ampla utilização em estudos de classificação baseados em espectroscopia e à sua capacidade de modelar eficientemente conjuntos de dados altamente colineares, projetando as variáveis espectrais em um espaço reduzido de variáveis latentes, preservando simultaneamente as informações relevantes para a discriminação entre as classes. Os modelos obtidos foram posteriormente utilizados como base preditiva para as análises de explicabilidade.

O desenvolvimento dos modelos foi realizado de forma independente para cada tarefa de autenticação. Para obter estimativas não viesadas do desempenho preditivo, todos os procedimentos de validação consideraram explicitamente a estrutura agrupada do conjunto de dados, garantindo que amostras derivadas do mesmo vinagre de origem permanecessem na mesma partição durante as etapas de treinamento e validação.

Foram selecionadas para as análises de explicabilidade as configurações de PLS-DA com alinhamento espectral que apresentaram o desempenho preditivo mais robusto. Os modelos finais foram compostos por oito variáveis latentes para a tarefa de classificação AM vs. AD e quatro variáveis latentes para a tarefa AM vs. MA, correspondendo às soluções modais

identificadas durante o procedimento de validação cruzada aninhada com agrupamento.

Os classificadores finais foram então empregados como modelos preditivos no *framework Spectral Model eXplainer (SMX)*. Após a etapa de seleção dos modelos, nenhuma otimização adicional foi realizada, uma vez que o objetivo deste estudo consistiu em interpretar as informações espectrais responsáveis pelas predições dos classificadores, e não em maximizar seu desempenho preditivo.

2.5. Spectral Model eXplainer (SMX)

As regiões espectrais que contribuíram para as predições dos modelos foram analisadas utilizando o *Spectral Model eXplainer (SMX)*, um *framework* de inteligência artificial explicável (XAI) pós-hoc, desenvolvido especificamente para conjuntos de dados espectroscópicos (Ahmed et al., 2025; Arrighi et al., 2026; Ribeiro et al., 2026).

O SMX foi aplicado de forma independente aos modelos finais de PLS-DA desenvolvidos para cada cenário de autenticação. As probabilidades de pertencimento às classes, preditas pelos classificadores, foram utilizadas como variável resposta no procedimento de explicabilidade. A relevância espectral foi quantificada por meio de uma estratégia baseada em perturbação, na qual domínios espectrais previamente definidos foram mascarados sequencialmente e substituídos por valores representativos estimados a partir do conjunto de treinamento. As variações resultantes nas probabilidades preditas das classes foram utilizadas para quantificar a contribuição de cada domínio espectral para as predições do modelo.

O intervalo espectral alinhado entre 0,8 e 4,0 ppm foi particionado em nove domínios espectrais relacionados a metabólitos: EtOH/BCAA (0,8–1,1 ppm), EtOH/Polióis (1,1–1,4 ppm), Lac/Ala (1,4–1,8 ppm), Acetato (1,8–2,2 ppm), Succ/Cit (2,2–2,6 ppm), Mal/Cit (2,6–3,0 ppm), Açúcares-L (3,0–3,4 ppm), Açúcares/Gly (3,4–3,8 ppm) e Açúcares-H (3,8–4,0

ppm). Os limites desses domínios foram estabelecidos com base nas atribuições de deslocamento químico previamente reportadas em estudos metabolômicos de vinagres (Boffo et al., 2009; Caligiani et al., 2007; Grassi et al., 2024; Nie et al., 2019).

Para cada domínio espectral, o SMX gerou grafos explicativos utilizando predicados derivados dos quantis de 0,25, 0,50 e 0,75 das respectivas pontuações regionais. A importância de cada domínio foi resumida por meio da métrica *Local Reaching Centrality* (LRC), calculada individualmente para cada domínio e utilizada para estabelecer um ranqueamento de sua contribuição relativa para as predições do classificador. Uma descrição matemática detalhada do *framework* SMX é apresentada na seção seguinte.

Para fins de comparação, também foram calculados os escores de *Variable Importance in Projection* (VIP) a partir dos respectivos modelos de PLS-DA, utilizando a estrutura de variáveis latentes ajustada.

2.5.1. Formulação Matemática do Spectral Model eXplainer (SMX)

O SMX foi implementado de acordo com a metodologia proposta por Ribeiro et al. (2026), com adaptações restritas à estratégia de segmentação espectral adotada no presente estudo. O *framework* foi aplicado de forma *post-hoc* aos modelos finais de PLS-DA, utilizando como entrada as impressões digitais espectrais previamente alinhadas. A formulação matemática do algoritmo SMX é resumida a seguir.

Inicialmente, a matriz espectral foi particionada em m regiões espectrais predefinidas, de acordo com o esquema de segmentação selecionado. Cada região espectral Z_m foi definida como:

$$Z_m = \{j: \lambda_{\text{start}}^{(m)} \leq \lambda_j \leq \lambda_{\text{end}}^{(m)}\},$$

em que λ_j representa o deslocamento químico associado à variável espectral j .

Em seguida, foi realizada a redução de dimensionalidade de forma independente para cada região espectral por meio da Análise de Componentes Principais (PCA). Apenas o primeiro componente principal (PC1) foi retido, gerando um escore representativo para cada região espectral:

$$t_i^{(m)} = \left(x_i^{(m)} - \bar{x}_m \right)^T w_1^{(m)},$$

em que $x_i^{(m)}$ corresponde ao vetor contendo as variáveis espectrais da amostra i na região m , $\bar{x}_m$ representa o vetor médio correspondente e $w_1^{(m)}$ corresponde ao vetor de cargas (*loadings*) associado ao primeiro componente principal. Esses escores regionais foram posteriormente utilizados para a geração dos predicados no *framework* SMX.

A relevância espectral foi estimada utilizando a estratégia de perturbação descrita por Ribeiro et al. (2026). Para cada região espectral avaliada, apenas as variáveis pertencentes ao intervalo selecionado foram substituídas pela mediana calculada a partir do conjunto de treinamento, enquanto todas as demais variáveis permaneceram inalteradas:

$$\tilde{x}_{ij} = \begin{cases} \text{median}(x_j), & j \in Z_m \\ x_{ij}, & j \notin Z_m \end{cases}$$

Os espectros perturbados foram então submetidos ao classificador PLS-DA previamente treinado, sendo o impacto da perturbação quantificado a partir da variação nas probabilidades preditas das classes:

$$Imp(P_j) = \frac{1}{|S|} \sum_{i \in S} \frac{1}{C} \sum_{c=1}^C |p_{ic} - p_{ic}^{\text{pert}}|,$$

em que p_{ic} e p_{ic}^{pert} representam, respectivamente, as probabilidades preditas antes e após

a perturbação, C corresponde ao número de classes e S representa o subconjunto de amostras que satisfaz o predicado correspondente.

Os escores de perturbação obtidos para todos os predicados foram integrados em um grafo direcionado e ponderado, conforme a implementação original do SMX. A relevância espectral foi resumida por meio da métrica *Local Reaching Centrality* (LRC), calculada como:

$$LRC(v) = \frac{1}{|V| - 1} \sum_{u \in V \setminus \{v\}} \frac{1}{|\pi(v, u)|} \sum_{k=1}^{|\pi(v, u)|} w_k^{(v, u)},$$

em que V representa o conjunto de nós do grafo, $\pi(v, u)$ corresponde ao menor caminho direcionado entre os nós v e u , e $w_k^{(v, u)}$ representa o peso associado a cada aresta que compõe esse caminho. Os valores de LRC resultantes foram utilizados para ranquear as regiões espectrais de acordo com sua contribuição relativa para as predições do classificador.

Todos os demais parâmetros de implementação, incluindo a geração dos predicados, o procedimento de *bagging* estocástico, a construção do grafo e a estimativa da centralidade, seguiram a implementação original descrita por Ribeiro et al. (2026).

2.6. Avaliação da Fidelidade das Explicações

A consistência dos métodos de explicação foi avaliada por meio de uma análise de fidelidade baseada em perturbação. Para cada modelo de autenticação, as regiões espectrais foram ranqueadas de acordo com os escores de relevância gerados pelo *Spectral Model eXplainer* (SMX) ou pelos valores agregados de *Variable Importance in Projection* (VIP) (Ahmed et al., 2025; Arrighi et al., 2026; Ribeiro et al., 2026).

A fidelidade foi avaliada por meio de experimentos de mascaramento progressivo. Em

cada iteração, a região espectral classificada como mais relevante foi substituída utilizando a mesma estratégia de perturbação empregada na análise do SMX, enquanto todas as demais variáveis espectrais foram mantidas inalteradas. Os espectros perturbados foram então submetidos ao respectivo modelo de PLS-DA, e as probabilidades preditas das classes foram recalculadas (Ahmed et al., 2025; Arrighi et al., 2026; Ribeiro et al., 2026).

O procedimento de mascaramento foi realizado de forma cumulativa, seguindo a ordem de relevância estabelecida por cada método de explicação. Após cada etapa de mascaramento, a variação na probabilidade predita da classe foi registrada e utilizada como medida quantitativa da fidelidade das explicações.

Para permitir a comparação direta entre os métodos de explicação, os valores de VIP foram agregados dentro dos mesmos domínios espectrais dos metabólitos definidos para o esquema de segmentação adotado no SMX. Os ranqueamentos regionais de VIP resultantes foram então avaliados utilizando o mesmo protocolo de perturbação empregado para o SMX. Todas as análises de fidelidade foram realizadas de forma independente para cada cenário de autenticação e para ambas as estratégias de segmentação espectral.

3. Resultados e Discussão

3.1. Organização Química dos Domínios Espectrais de RMN de Baixo Campo

Os espectros de RMN de ^1H de baixo campo dos vinagres apresentaram extensa sobreposição de ressonâncias devido à reduzida dispersão espectral característica dos instrumentos baseados em ímãs permanentes (Lolli & Caligiani, 2024; Michal, 2020). Embora essa limitação dificulte a atribuição e a interpretação de ressonâncias individuais, ela não impediu a extração de informações quimicamente relevantes, uma vez que as impressões

digitais metabolômicas são intrinsecamente multivariadas. Em vez de se basearem em picos isolados, os modelos de classificação exploraram variações coordenadas distribuídas em múltiplas regiões espectrais correlacionadas. Consequentemente, as estratégias de explicabilidade fundamentadas exclusivamente em variáveis individuais fornecem uma interpretação química limitada para matrizes fermentadas complexas (Arrighi et al., 2026; Ribeiro et al., 2026).

Para superar essa limitação, a interpretação espectral foi realizada utilizando uma abordagem orientada por domínios, na qual ressonâncias vizinhas foram agrupadas de acordo com suas relações químicas e bioquímicas. Essa estratégia está alinhada aos princípios da metabolômica não direcionada, na qual informações biologicamente relevantes emergem de perturbações coordenadas que afetam grupos de metabólitos, e não variáveis espectrais isoladas (Arrighi et al., 2026; Ribeiro et al., 2026).

Dessa forma, a organização da impressão digital espectral em domínios quimicamente coerentes estabelece uma conexão direta entre o comportamento do modelo e processos bioquímicos interpretáveis, fornecendo explicações mais consistentes com os princípios da química analítica do que aquelas obtidas por abordagens convencionais baseadas na interpretação de variáveis individuais (Arrighi et al., 2026; Ribeiro et al., 2026).

O intervalo espectral selecionado (0,8–4,0 ppm) compreende a maior parte das ressonâncias provenientes de metabólitos comumente encontrados em vinagres. Conforme ilustrado na Figura 2, podem ser reconhecidos três principais domínios bioquímicos. A região entre aproximadamente 3,0 e 4,0 ppm é predominantemente composta por ressonâncias associadas a carboidratos, incluindo açúcares residuais e outros compostos poli-hidroxilados. A região intermediária (1,8–3,0 ppm) está principalmente relacionada aos ácidos orgânicos produzidos durante os processos de fermentação alcoólica e acética, enquanto a região de menor deslocamento químico (0,8–1,8 ppm) contém uma mistura complexa de compostos derivados

do etanol, aminoácidos, polióis e outros metabólitos gerados ao longo do processo fermentativo (Avanzi Barbosa Mascareli et al., 2023; Boffo et al., 2009; Caligiani et al., 2007; Consonni, Cagliani, Benevelli, et al., 2008; Grassi et al., 2024; Grassi & Alamprese, 2024; Nie et al., 2019).

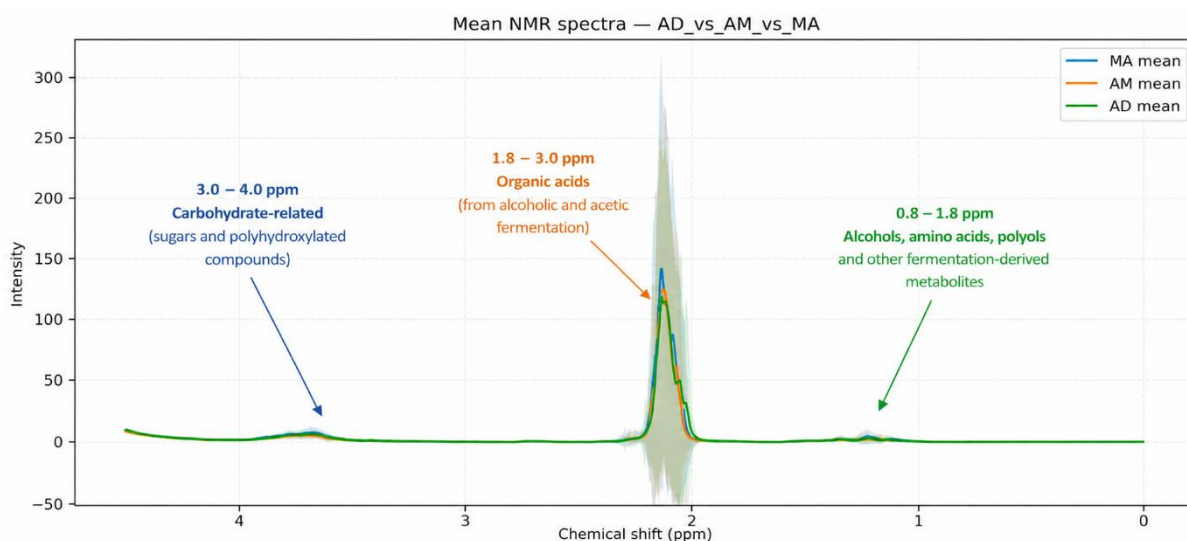


Figura 1. Intervalo espectral selecionado dos espectros médios de RMN de ^1H em baixo campo (LF-NMR) das classes MA (vinagre de maçã adulterado com 5, 15 ou 25% de vinagre de álcool), AM (vinagre de maçã orgânico) e AD (vinagre de maçã orgânico adulterado com 5, 6, 30 ou 31% de vinagre de maçã convencional).

Embora esses três domínios abrangentes forneçam uma visão geral útil da composição química dos vinagres, cada um deles ainda integra múltiplas contribuições metabólicas sobrepostas. A região correspondente aos carboidratos contém sinais provenientes de substratos fermentescíveis residuais e polióis, enquanto a região dos ácidos orgânicos engloba metabólitos como acetato, malato, citrato e succinato, os quais participam das principais vias metabólicas envolvidas nos processos fermentativos. De forma semelhante, a região de menor deslocamento químico reúne ressonâncias associadas ao etanol, aminoácidos, metabólitos microbianos e outros compostos formados ao longo da fermentação. Consequentemente, a interpretação nesse

nível permite identificar as principais classes bioquímicas que contribuem para a classificação, porém não possibilita discriminar os domínios composicionais específicos responsáveis pelo comportamento do modelo (Boffo et al., 2009; Caligiani et al., 2007; Consonni, Cagliani, Benevelli, et al., 2008; Grassi et al., 2024; Grassi & Alamprese, 2024; Nie et al., 2019).

Para aumentar a resolução interpretativa, cada um desses domínios bioquímicos abrangentes foi subdividido em nove domínios espectrais indicados por metabólitos (Tabela 2), representando regiões quimicamente coerentes, descritas de forma recorrente em estudos metabolômicos de vinagres. Esses domínios não devem ser interpretados como atribuições definitivas de metabólitos, uma vez que a sobreposição de sinais permanece uma característica intrínseca da RMN de ^1H em baixo campo (LF-NMR). Em vez disso, representam ambientes espectrais enriquecidos em compostos quimicamente relacionados, que, em conjunto, descrevem diferentes aspectos da composição química dos vinagres e da bioquímica dos processos fermentativos (Boffo et al., 2009; Caligiani et al., 2007; Consonni, Cagliani, Benevelli, et al., 2008; Grassi et al., 2024; Grassi & Alamprese, 2024; Nie et al., 2019).

Tabela 2. Definição dos supostos domínios espectrais de metabólitos empregados no *framework* SMX e sua correspondente interpretação química recorrentes na literatura

Domínio Espectral	Deslocamento Químico (ppm)	Principais Metabólitos
EtOH/BCAA	0,8–1,1	Etanol, leucina, isoleucina, valina
EtOH/Polióis	1,1–1,4	Compostos relacionados ao etanol e polióis
Lac/Ala	1,4–1,8	Lactato, alanina
Acetato	1,8–2,2	Ácido Acético/acetato

Domínio Espectral	Deslocamento Químico (ppm)	Principais Metabólitos
Succ/Cit	2,2–2,6	Succinato, citrato
Mal/Cit	2,6–3,0	Malato, citrato
Açúcares-L	3,0–3,4	Carboidratos de baixo deslocamento químico
Açúcares/Gly	3,4–3,8	Sinais relacionados a açúcares e glicerol
Açúcares-H	3,8–4,0	Carboidratos de alto deslocamento químico

Essa organização constitui a base conceitual do *framework* do SMX. Ao permitir que a relevância espectral seja analisada em múltiplos níveis, desde classes bioquímicas abrangentes até domínios espectrais enriquecidos em metabólitos, o *framework* estabelece uma conexão entre as predições dos modelos de aprendizado de máquina e sua interpretação metabolômica. Em vez de identificar variáveis espectrais isoladas como influentes, a estratégia proposta possibilita que análises subsequentes determinem como domínios espectrais coordenados governam as decisões de classificação, estabelecendo, assim, um *framework* quimicamente interpretável para a aplicação da inteligência artificial explicável à autenticação de alimentos e bebidas baseada em dados espectroscópicos.

3.2. Modelos de Classificação Utilizados nas Análises de Explicabilidade

Antes da interpretação do comportamento dos modelos por meio do SMX, o desempenho preditivo dos classificadores PLS-DA foi inicialmente avaliado, assegurando que as análises de explicabilidade fossem conduzidas sobre modelos com capacidade discriminativa

adequada. Os modelos desenvolvidos para os cenários de autenticação AM vs. AD e AM vs. MA apresentaram desempenho semelhante, alcançando acurácias globais de aproximadamente 81% e valores de macro-F1 de 0,79 e 0,80, respectivamente (Tabela 3).

Tabela 3. Desempenho preditivo dos modelos de PLS-DA empregados nas análises de explicabilidade

Cenário de Classificação		Precisão	Recall	F1-score	N amostral
AM vs MA	AM	0.77	0.72	0.74	135
	MA	0.84	0.87	0.85	223
	Média Macro	0.80	0.79	0.80	358
	Média Ponderada	0.81	0.81	0.81	358
	Acurácia	—	—	0.81	358
		—	—	—	—
AM vs AD	AM	0.78	0.71	0.74	131
	AD	0.82	0.87	0.84	200
	Média Macro	0.80	0.79	0.79	331
	Média Ponderada	0.81	0.81	0.80	331
	Acurácia	—	—	0.81	331

Ambos os classificadores foram capazes de capturar informações discriminantes a partir das impressões digitais de LF-NMR, fornecendo uma base suficientemente robusta para as análises subsequentes de explicabilidade. Como o objetivo principal deste estudo não foi maximizar o desempenho preditivo, mas compreender quais informações espectrais

fundamentam as decisões de classificação, as seções seguintes concentram-se nos padrões de relevância identificados pelo SMX e em sua interpretação química.

O desempenho alcançado pelos dois modelos de PLS-DA demonstram que as impressões digitais obtidas por LF-NMR preservam informações composicionais suficientes para aplicações em autenticação de alimentos e bebidas, apesar da menor resolução espectral inerente a essa técnica em comparação com instrumentos de alto campo (Michal, 2020; Lolli & Caligiani, 2024).

Esse comportamento evidencia que o poder discriminante da LF-NMR emerge da integração de múltiplas informações espectrais, reforçando sua adequação como uma plataforma rápida e não direcionada para avaliação da autenticidade de alimentos e bebidas, especialmente quando combinada a métodos quimiométricos de classificação (Conte et al., 2025; Galvan et al., 2023).

Os resultados específicos por classe também evidenciam uma assimetria consistente no desempenho preditivo. Em ambos os cenários de autenticação, as classes adulteradas apresentaram valores de F1-score superiores aos observados para a respectiva classe de vinagre de maçã autêntico. Esse comportamento sugere que os procedimentos de adulteração produziram modificações composicionais relativamente sistemáticas, enquanto os vinagres de maçã autênticos preservaram maior heterogeneidade intrínseca, decorrente de diferenças nas matérias-primas, nos lotes de produção, nas condições de fermentação e no processamento industrial. Essa variabilidade aumenta a sobreposição entre as classes e torna a categoria de vinagres autênticos mais difícil de ser modelada como um grupo quimicamente homogêneo.

Essas observações são particularmente relevantes para as análises de explicabilidade. Explicações confiáveis do comportamento de modelos de aprendizado de máquina exigem não apenas desempenho preditivo satisfatório, mas também evidências de que o classificador aprendeu diferenças composicionais genuinamente associadas às classes investigadas. Os

valores equilibrados de macro-F1, aliados ao desempenho consistente observado em ambos os cenários de autenticação, indicam que os modelos capturaram informações espectrais reprodutíveis, fornecendo uma base robusta para investigar como diferentes domínios espectrais enriquecidos em metabólitos influenciaram suas decisões de classificação.

3.3. Explicações Espectrais Geradas pelo SMX

Como dito anteriormente, para obter uma interpretação mais detalhada do comportamento dos modelos, o SMX foi aplicado utilizando a estratégia de segmentação fina, na qual a impressão de LF-NMR foi particionada em nove domínios espectrais que representam metabólitos, baseados em trabalhos descritos na literatura (Boffo et al., 2009; Caligiani et al., 2007; Consonni, Cagliani, Benevelli, et al., 2008; Grassi et al., 2024; Grassi & Alamprese, 2024; Nie et al., 2019).

Os mapas de relevância resultantes revelaram perfis de importância distintos para os dois cenários de autenticação (Figura 2), demonstrando que os determinantes espectrais da classificação dependeram do mecanismo de adulteração investigado. Em ambos os modelos, entretanto, a relevância esteve distribuída entre diversos domínios espectrais, em vez de concentrada em uma única região dominante. Esse padrão indica que as predições dos modelos resultaram da contribuição combinada de múltiplas características composicionais, em concordância com a natureza distribuída das informações codificadas em impressões digitais espectroscópicas não direcionadas (Ribeiro et al., 2026).

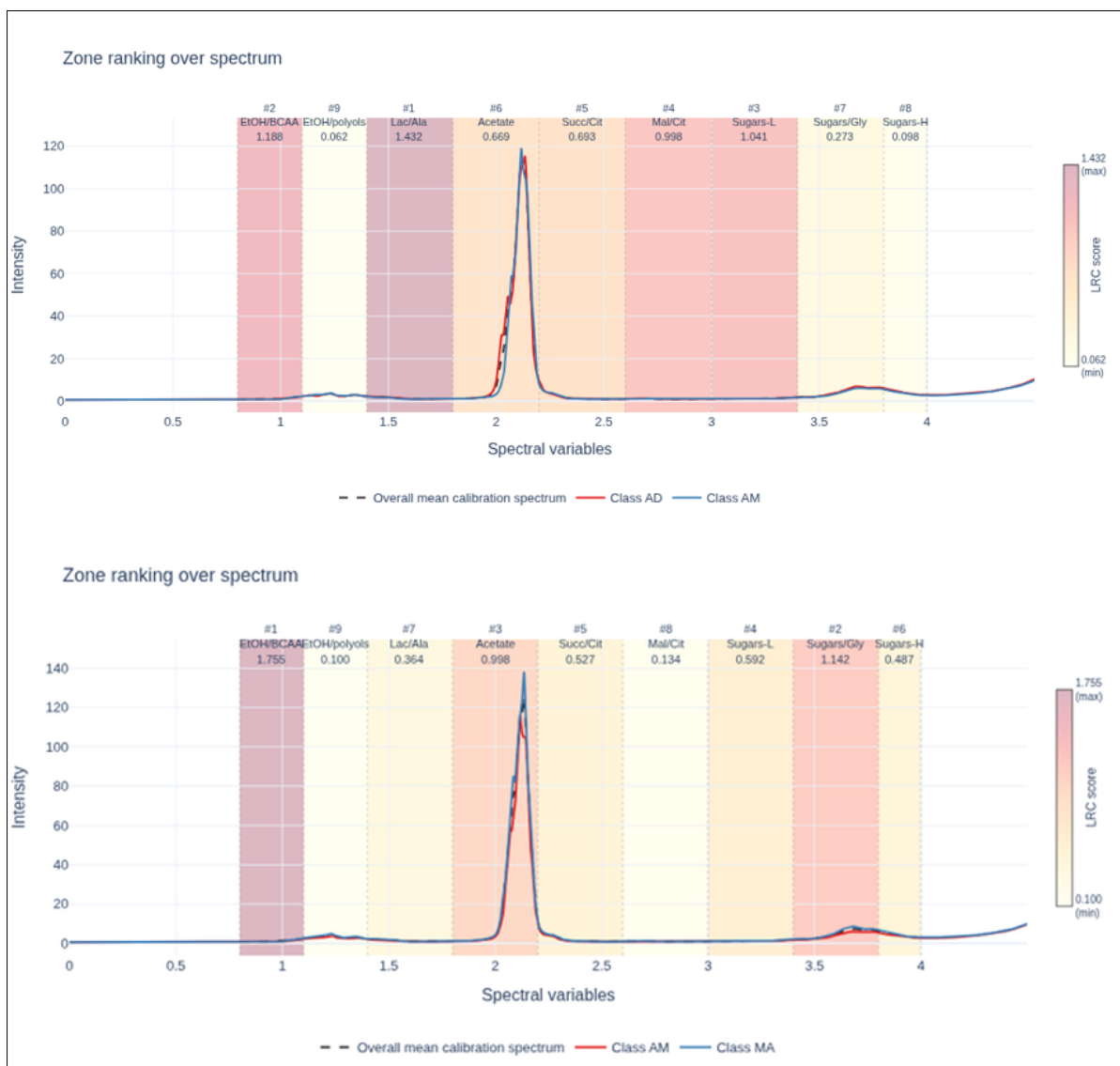


Figura 2. Mapeamento da relevância dos domínios espectrais sobre os espectros alinhados de ^1H LF-NMR, gerado pelo *Spectral Model eXplainer*. **Painel superior:** discriminação entre vinagre de maçã orgânico (AM) e vinagre de maçã orgânico adulterado (AD). **Painel inferior:** discriminação entre vinagre de maçã (AM) e vinagre de maçã adulterado com vinagre de álcool (MA). Os nove domínios espectrais enriquecidos em metabólitos são destacados por diferentes tonalidades, cuja intensidade é proporcional ao valor da métrica LRC, refletindo a contribuição relativa de cada domínio para as decisões de classificação.

Essa distribuição da relevância é consistente com os princípios fundamentais da

metabolômica não direcionada, na qual a discriminação entre amostras é, em geral, sustentada por variações coordenadas envolvendo múltiplos metabólitos correlacionados, e não por biomarcadores isolados. Comportamento semelhante tem sido relatado em estudos de autenticação baseados em impressões digitais obtidas por RMN, FTIR e NIR, nos quais o desempenho da classificação frequentemente depende de padrões metabólicos multivariados distribuídos ao longo de todo o perfil analítico (Biancolillo et al., 2020; García-Pérez et al., 2024).

Durante a produção de vinagre, a composição da matéria-prima, a fermentação alcoólica, a oxidação acética, a sucessão microbiana e o processo de maturação influenciam conjuntamente diversas vias bioquímicas interconectadas. Esses processos geram estruturas de covariância que são naturalmente capturadas pelas impressões digitais espectroscópicas e posteriormente exploradas por classificadores multivariados (Emwas et al., 2019; Grassi et al., 2024; Grassi & Alamprese, 2024b; Wishart, 2019). Sob essa perspectiva, as distribuições de relevância geradas pelo *Spectral Model eXplainer* indicam que os modelos capturaram padrões metabólicos quimicamente plausíveis, sugerindo que as decisões de classificação foram fundamentadas em informações composicionais distribuídas ao longo do espectro.

Para o modelo de autenticação AD vs. AM, a maior relevância foi atribuída ao domínio Lac/Ala (22,2%), seguido por EtOH/BCAA (18,4%) e Açúcares-L (16,1%) (Figura 3). Contribuições relevantes também foram observadas para os domínios Mal/Cit (15,5%), Succ/Cit (10,7%) e Acetato (10,4%), enquanto os demais domínios associados aos carboidratos apresentaram relevância comparativamente menor. De modo geral, esse ranqueamento indica que a discriminação entre vinagres de maçã orgânicos e convencionais foi sustentada pela contribuição conjunta de metabólitos derivados da fermentação, sinais associados a aminoácidos, carboidratos residuais e regiões espectrais relacionadas aos ácidos orgânicos.

Uma hierarquia de relevância distinta foi observada para o modelo de autenticação AM

vs. MA. Nesse caso, o domínio EtOH/BCAA apresentou a maior contribuição (28,8%), seguido por Açúcares/Gly (18,7%) e Acetato (16,4%). Contribuições intermediárias foram observadas para Açúcares-L (9,7%), Succ/Cit (8,6%) e Açúcares-H (8,0%), enquanto Mal/Cit (2,2%) e EtOH/Polióis (1,6%) apresentaram apenas contribuições marginais. Esses resultados indicam que a discriminação entre vinagre de maçã autêntico e amostras adulteradas com vinagre de álcool dependeu predominantemente de domínios espectrais associados a compostos relacionados ao etanol, ressonâncias provenientes de carboidratos e produtos da fermentação acética.

A comparação entre os dois modelos de autenticação revelou uma mudança expressiva nas informações espectrais que governaram a classificação. Enquanto o modelo AD vs. AM foi sustentado predominantemente pela contribuição de domínios espectrais associados a metabólitos derivados da fermentação, aminoácidos, carboidratos residuais e ácidos orgânicos, o modelo AM vs. MA apresentou maior relevância em domínios relacionados a compostos associados ao etanol, regiões enriquecidas em carboidratos e metabólitos relacionados ao acetato. Essa divergência demonstra que a base espectral da autenticação é dependente do cenário de adulteração e reflete a relação composicional entre o produto autêntico e o adulterante.

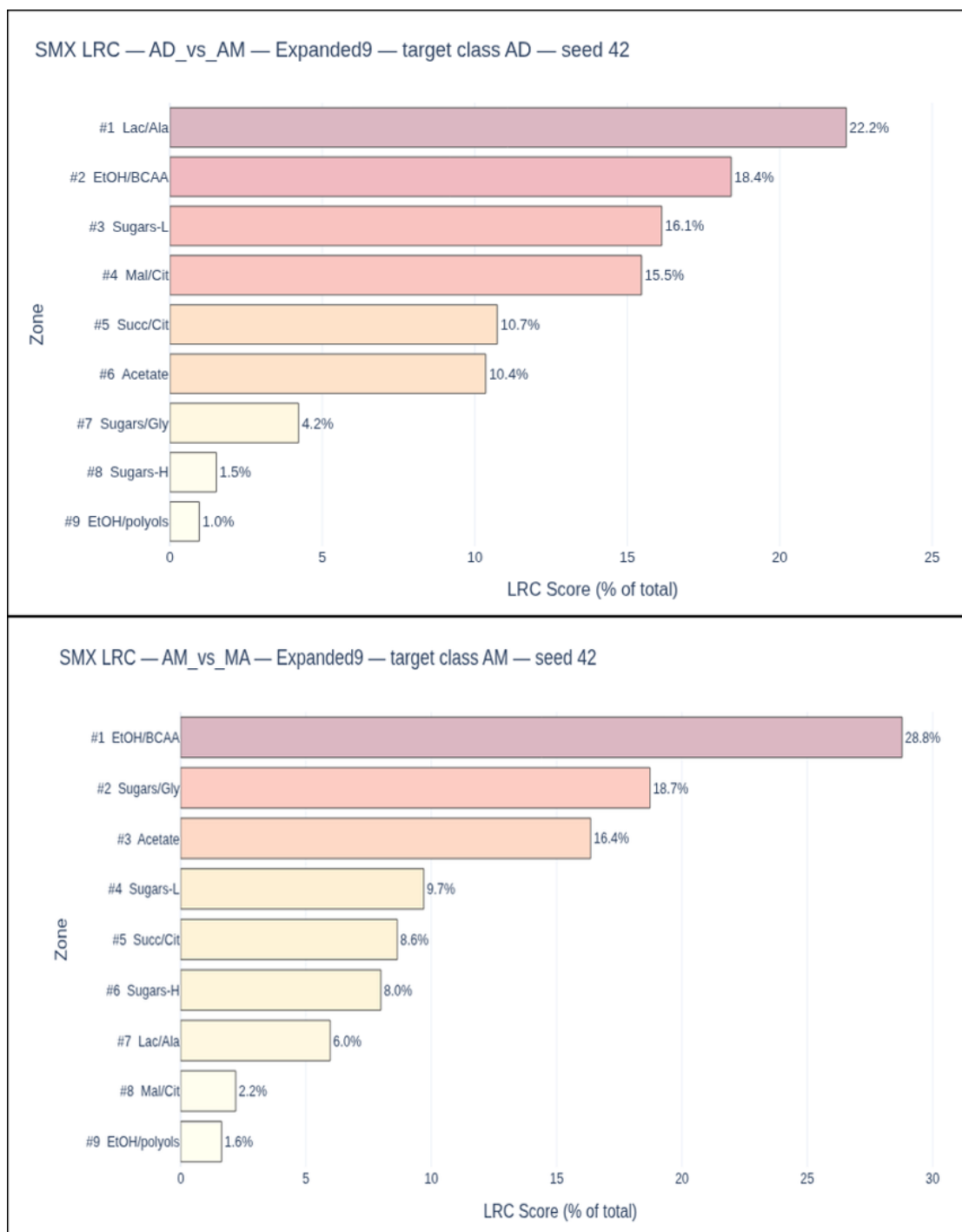


Figura 3. Ranqueamento dos nove domínios espectrais enriquecidos em metabólitos com base na métrica LRC, gerado pelo SMX. **Painel superior:** discriminação entre vinagre de maçã orgânico (AM) e vinagre de maçã orgânico adulterado (AD). **Painel inferior:** discriminação entre vinagre de maçã (AM) e vinagre de maçã adulterado com vinagre de álcool (MA). Os valores correspondem à contribuição relativa de cada domínio espectral para o comportamento do modelo, expressa como porcentagem da pontuação total de LRC.

Essa distinção é quimicamente consistente com a natureza dos dois mecanismos de fraude. A substituição do vinagre de maçã orgânico por vinagre de maçã convencional preserva o perfil metabólico geral derivado da fruta, sendo, portanto, esperado que produza diferenças relativamente sutis associadas às práticas agrícolas, à composição da matéria-prima, ao histórico fermentativo e às condições de processamento. Em contrapartida, a adulteração com vinagre de álcool introduz uma matriz quimicamente mais simples, contendo menores concentrações de metabólitos derivados da fruta e de produtos secundários da fermentação, resultando em alterações mais amplas na organização relativa da impressão digital espectral. Essas diferenças composicionais explicam por que cada cenário de adulteração foi caracterizado por um conjunto distinto de domínios espectrais relevantes, evidenciando que a contribuição relativa das diferentes regiões do espectro depende diretamente da natureza química do adulterante e de sua interação com a matriz original.

Respostas metabolômicas dependentes do cenário de adulteração também foram descritas em estudos de autenticidade envolvendo sucos de frutas, vinhos, vinagres e outros produtos fermentados, nos quais os padrões metabólicos responsáveis pela classificação variaram de acordo tanto com a identidade do adulterante quanto com a complexidade bioquímica da matriz original (Consonni et al., 2008; Mialon et al., 2023; Grassi & Alamprese, 2024).

Nesse contexto, os resultados obtidos pelo SMX reforçam que as decisões dos modelos foram fundamentadas na integração de informações químicas distribuídas ao longo da impressão digital espectral, em vez de dependerem de variáveis espectrais individuais. Essa característica é consistente com os princípios da metabolômica não direcionada e demonstra que a interpretação fornecida pelo SMX preserva a natureza multivariada da classificação, permitindo associar as decisões do modelo a padrões químicos plausíveis e quimicamente interpretáveis.

Sob a perspectiva da explicabilidade, a estratégia de segmentação aprimorou substancialmente a interpretabilidade ao localizar a relevância do modelo em domínios espectrais enriquecidos em metabólitos, em vez de intervalos espectrais amplos ou variáveis espectrais isoladas. Esse nível de refinamento permitiu traduzir o comportamento dos modelos de aprendizado de máquina em informações quimicamente interpretáveis, preservando a natureza multivariada da impressão digital espectroscópica. Mais importante, as diferentes hierarquias de relevância observadas entre os dois modelos demonstram que o *Spectral Model eXplainer* não fornece uma assinatura genérica da autenticidade de vinagres. Ao contrário, o *framework* evidencia as consequências bioquímicas específicas de cada mecanismo de adulteração, revelando como diferentes tipos de fraude reorganizam de maneira distinta as informações espectrais associadas ao metaboloma do vinagre. Esses resultados reforçam o potencial do SMX como uma ferramenta de inteligência artificial explicável para converter modelos quimiométricos em interpretações quimicamente fundamentadas, ampliando a transparência e a confiabilidade das abordagens metabolômicas aplicadas à autenticação de alimentos.

3.4. Interpretação Química dos Domínios Espectrais Relevantes

As explicações geradas pelo *Spectral Model eXplainer* fornecem informações valiosas sobre as características composicionais subjacentes à autenticação de vinagres ao relacionarem a relevância do modelo a domínios espectrais enriquecidos em metabólitos quimicamente interpretáveis. Embora os espectros de ^1H LF-NMR sejam inerentemente caracterizados por extensa sobreposição de sinais, os padrões de relevância identificados pelo *framework* revelaram associações consistentes entre os resultados da autenticação e conjuntos específicos de metabólitos derivados da fermentação. Dessa forma, a interpretação do comportamento dos modelos transcende a avaliação do desempenho preditivo, possibilitando associar as decisões de classificação a processos bioquímicos e tecnológicos plausíveis envolvidos na produção do

vinagre.

Para o cenário AD vs. AM, a maior relevância foi atribuída ao domínio Lac/Ala, seguido por EtOH/BCAA e Açúcares-L. Em conjunto, essas regiões corresponderam a mais da metade da relevância total identificada pelo SMX, indicando que a distinção entre vinagres de maçã orgânicos e convencionais está fortemente associada a domínios espectrais enriquecidos em metabólitos derivados da fermentação e carboidratos residuais. O domínio Lac/Ala compreende ressonâncias comumente atribuídas ao lactato, à alanina e a compostos estruturalmente relacionados, produzidos durante o metabolismo microbiano. Variações observadas nesse domínio podem refletir diferenças na cinética fermentativa, na dinâmica das comunidades microbianas, na utilização de substratos e nas práticas de produção. Considerando que os sistemas de produção orgânico e convencional frequentemente diferem quanto ao manejo agrícola, às características da matéria-prima e às condições de processamento, a elevada relevância dessa região sugere que a autenticação é influenciada por assinaturas metabólicas sutis acumuladas ao longo do processo produtivo.

A predominância do domínio Lac/Ala indica que alterações coordenadas em uma região espectral enriquecida em metabólitos relacionados ao metabolismo fermentativo exercem maior influência sobre a autenticação do que a abundância dos principais constituintes químicos do vinagre. Embora esse domínio inclua ressonâncias frequentemente atribuídas ao lactato e à alanina, sua relevância provavelmente reflete a contribuição conjunta de múltiplos metabólitos correlacionados, característica inerente às abordagens metabolômicas não direcionadas. O lactato e a alanina representam componentes importantes dessa região espectral, porém suas concentrações devem ser interpretadas no contexto de um conjunto mais amplo de metabólitos correlacionados, cuja organização conjunta é capturada pela impressão digital espectroscópica e explorada pelos modelos multivariados. Dessa forma, as variações observadas nesse domínio podem refletir diferenças na dinâmica da fermentação, composição microbiana, na

disponibilidade de oxigênio e na eficiência de utilização dos substratos. Consequentemente, as alterações detectadas nessa região espectral representam, com maior probabilidade, alterações coordenadas na organização metabolômica da matriz ao longo da produção do vinagre, em vez de modificações na concentração de compostos individuais, refletindo a resposta integrada de múltiplas vias metabólicas associadas aos processos fermentativos (Boffo et al., 2009; Nie et al., 2019; Grassi et al., 2024).

Essa observação está de acordo com estudos metabolômicos anteriores que demonstram que a autenticidade de vinagres é frequentemente determinada por alterações coordenadas em regiões espectrais enriquecidas em metabólitos derivados da fermentação, em vez de depender predominantemente dos constituintes majoritários, como o ácido acético, cujas concentrações tendem a permanecer relativamente semelhantes entre produtos comerciais que atendem aos padrões regulatórios (Caligiani et al., 2007; Consonni et al., 2008).

A elevada relevância do domínio EtOH/BCAA reforça essa interpretação. Essa região espectral é enriquecida em ressonâncias associadas ao etanol, aminoácidos de cadeia ramificada e outros metabólitos derivados do metabolismo fermentativo. Esse conjunto de ressonâncias está associado à atividade microbiana e às transformações bioquímicas que ocorrem durante a fermentação. A contribuição desse domínio para as predições do modelo sugere que diferenças no histórico fermentativo e na reorganização bioquímica da matriz constituem uma importante fonte de variação entre produtos orgânicos e convencionais. Entretanto, considerando a natureza não direcionada da abordagem empregada, a relevância observada provavelmente reflete a contribuição conjunta de múltiplos metabólitos correlacionados presentes nessa região espectral, e não exclusivamente dos compostos individualmente atribuídos. As ressonâncias atribuídas aos aminoácidos de cadeia ramificada estão associadas ao metabolismo microbiano e ao catabolismo de aminoácidos por leveduras e bactérias, enquanto os sinais relacionados ao etanol refletem etapas fundamentais da fermentação alcoólica e suas interações com os

processos subsequentes de acetificação. Dessa forma, as variações observadas nesse domínio espectral representam, mais provavelmente, alterações coordenadas na organização metabolômica da matriz estabelecidas ao longo da fermentação alcoólica e da acetificação, em vez de simples variações na concentração de etanol. Essa interpretação está em consonância com estudos de metabolômica baseada em RMN, nos quais padrões espectrais associados ao metabolismo fermentativo foram reconhecidos como importantes determinantes da diferenciação de alimentos fermentados (Boffo et al., 2009; Grassi et al., 2024).

Os domínios espectrais enriquecidos em ressonâncias associadas a carboidratos residuais também contribuíram substancialmente para a discriminação do modelo. A intensidade dessas ressonâncias depende da composição da fruta, da extensão da fermentação, do consumo microbiano de açúcares e de eventuais processos de diluição. Consequentemente, essas regiões espectrais capturam assinaturas espectrais integradas associadas à qualidade da matéria-prima, à eficiência dos processos fermentativos e à organização metabolômica estabelecida durante a produção do vinagre, justificando sua recorrente relevância em estudos metabolômicos aplicados à autenticação de vinagres (Caligiani et al., 2007; Nie et al., 2019; Grassi et al., 2024).

A relevância adicional atribuída aos domínios Mal/Cit, Succ/Cit e Acetato reforça a importância de regiões espectrais associadas às vias metabólicas dos ácidos orgânicos no processo de autenticação. Esses domínios compreendem ressonâncias frequentemente atribuídas a espécies químicas associadas ao metabolismo fermentativo, formadas durante as fermentações alcoólica e acética, refletindo o equilíbrio metabólico estabelecido ao longo da produção do vinagre. Em conjunto, a distribuição de relevância observada para o modelo AD vs. AM sugere que a autenticação é governada pela integração de múltiplos padrões espectrais associados ao metabolismo fermentativo, distribuídos entre diferentes vias bioquímicas.

Um padrão composicional distinto foi observado na tarefa de autenticação AM vs. MA.

Nesse caso, o domínio EtOH/BCAA representou a região espectral de maior relevância, seguido pelos domínios Açúcares/Gly e Acetato. Diferentemente do cenário AD vs. AM, no qual a relevância esteve predominantemente associada a domínios espectrais enriquecidos em ressonâncias associadas ao metabolismo fermentativo, a discriminação entre vinagre de maçã e amostras adulteradas com vinagre de álcool foi mais fortemente sustentada por domínios associados a compostos derivados do etanol, carboidratos e produtos da fermentação acética. Esse comportamento demonstra que a substituição parcial da matriz de vinagre de maçã por vinagre de álcool promove uma reorganização distinta da assinatura espectral, refletindo diferenças na composição química global da amostra. Assim, a explicabilidade fornecida pelo SMX evidencia que diferentes mecanismos de adulteração resultam em padrões específicos de relevância entre domínios espectrais, reforçando que a autenticação é sustentada pela organização multivariada do *fingerprint*.

A elevada relevância do domínio EtOH/BCAA é quimicamente plausível quando consideradas as diferenças fundamentais entre vinagres derivados de frutas e vinagres de álcool. Enquanto o vinagre de maçã é produzido a partir de uma matéria-prima quimicamente complexa, caracterizada por uma ampla diversidade de carboidratos, compostos nitrogenados, ácidos orgânicos e metabólitos secundários, o vinagre de álcool é obtido a partir de substratos etanólicos altamente refinados, apresentando, em geral, uma organização metabolômica da matriz menos complexa. Como consequência, a adição de vinagre de álcool promove uma reorganização da composição química global da matriz, modificando a distribuição relativa das ressonâncias presentes nesse domínio espectral e produzindo alterações detectáveis na assinatura espectral da matriz obtida por ^1H LF-NMR.

De maneira semelhante, a elevada relevância do domínio Açúcares/Gly provavelmente reflete modificações em ressonâncias associadas a carboidratos residuais, polióis e outras espécies químicas correlacionadas, decorrentes de alterações na organização metabolômica

característica do vinagre de maçã. O aumento da importância desse domínio é consistente com a simplificação composicional esperada quando componentes característicos da matriz frutífera são parcialmente substituídos pelos constituintes do vinagre de álcool, promovendo modificações na organização metabólica da matriz e, conseqüentemente, na assinatura espectral da matriz.

O domínio Acetato também apresentou contribuição expressiva para a classificação, indicando que alterações no equilíbrio entre produtos da fermentação acética e outras regiões espectrais associadas ao metabolismo fermentativo constituem uma fonte adicional de informação discriminante. Embora o acetato seja um dos principais constituintes químicos do vinagre, sua concentração isolada dificilmente seria suficiente para explicar o desempenho do modelo. Em vez disso, sua relevância parece emergir da interação com o conjunto de padrões espectrais que compõem a assinatura espectral da matriz.

Em conjunto, esses resultados demonstram que diferentes mecanismos de adulteração promovem perturbações distintas na organização metabólica da matriz do vinagre. Enquanto o modelo AD vs. AM foi predominantemente sensível a domínios espectrais enriquecidos em ressonâncias associadas ao metabolismo fermentativo, aos carboidratos residuais e às vias relacionadas aos ácidos orgânicos, o modelo AM vs. MA apresentou maior sensibilidade a domínios espectrais enriquecidos em ressonâncias associadas ao metabolismo do etanol, aos carboidratos residuais e aos produtos da fermentação acética.

De maneira mais ampla, esses resultados demonstram o potencial da XAI para integrar o elevado desempenho preditivo de modelos de aprendizado de máquina a interpretações quimicamente fundamentadas em estudos de autenticação de alimentos baseados em espectroscopia. Ao fornecer explicações associadas a domínios espectrais enriquecidos em ressonâncias, o *Spectral Model eXplainer* amplia a interpretabilidade de modelos quimiométricos sem comprometer sua capacidade discriminante, favorecendo uma

compreensão mais consistente dos processos químicos subjacentes às decisões do classificador.

Sob a perspectiva da metabolômica, esses resultados reforçam o conceito de que a autenticidade de alimentos constitui uma propriedade emergente da organização metabolômica da matriz. Nesse contexto, a capacidade do SMX de relacionar o comportamento do classificador a domínios espectrais enriquecidos em ressonâncias associadas a processos bioquímicos representa um avanço conceitual em relação às abordagens tradicionais de ranqueamento de variáveis. Em vez de apenas identificar variáveis estatisticamente relevantes, o *framework* evidencia como alterações coordenadas na organização metabolômica da matriz se refletem na assinatura espectral da matriz e, coletivamente, sustentam as decisões de autenticação do modelo. Dessa forma, as explicações produzidas permanecem consistentes tanto com a natureza multivariada da espectroscopia quanto com a interpretação em nível de sistemas que caracteriza a metabolômica moderna (Arrighi et al., 2026; Emwas et al., 2019; Wishart, 2019).

3.5. Avaliação da Fidelidade das Explicações Espectrais

A confiabilidade das explicações geradas pelo SMX foi avaliada por meio de uma estratégia de mascaramento progressivo de domínios espectrais, na qual essas regiões foram sequencialmente perturbadas de acordo com seu ranqueamento de relevância, enquanto as alterações nas predições do modelo eram monitoradas. Diferentemente das métricas convencionais de importância de variáveis, essa abordagem quantifica diretamente o grau de dependência das predições do classificador em relação às informações contidas em cada domínio espectral, fornecendo uma avaliação empírica da fidelidade das explicações.

Para o modelo de autenticação AD vs. AM, o mascaramento progressivo resultou em uma área sob a curva (AUC) normalizada de 0,3723, correspondente ao percentil de fidelidade de 95,0%, sendo classificado como Muito Alto (Figura 4, Superior). Os maiores deslocamentos

nas predições foram observados após a perturbação dos domínios Lac/Ala, EtOH/BCAA e Açúcares-L, em concordância com as regiões previamente identificadas pelo SMX como as mais relevantes para o processo de classificação. O aumento monotônico do deslocamento acumulado das predições demonstra que o ranqueamento de relevância refletiu adequadamente a hierarquia das informações espectrais exploradas pelo classificador, evidenciando uma forte correspondência entre as explicações produzidas e o processo decisório subjacente ao modelo.



Figura 4. Avaliação da fidelidade das explicações geradas pelo *Spectral Model eXplainer* por meio do mascaramento progressivo de domínios espectrais ranqueados. **Painel superior:**

modelo de autenticação AD vs. AM; **Painel inferior:** modelo de autenticação AM vs. MA. Escores de fidelidade calculados a partir dos deslocamentos acumulados nas predições do modelo produzidos pela perturbação sequencial dos domínios espectrais ordenados de acordo com o ranqueamento de relevância estabelecido pelo SMX.

Em contraste, o modelo de autenticação AM vs. MA apresentou uma AUC normalizada de 0,1388 e um percentil de fidelidade de 73,0%, correspondente a uma fidelidade Moderada (Figura 4, Inferior). Embora a perturbação dos domínios espectrais mais bem ranqueados tenha reduzido consistentemente a confiança das predições do modelo, o efeito acumulado foi substancialmente menor do que o observado para o modelo AD vs. AM. Esse comportamento sugere que a discriminação entre vinagre de maçã autêntico e amostras adulteradas com vinagre de álcool depende de informações químicas distribuídas entre múltiplos domínios espectrais enriquecidos em ressonâncias correlacionadas, em vez de estar concentrada em um número reduzido de regiões altamente dominantes.

Essa diferença na fidelidade das explicações é consistente com a natureza química dos dois problemas de autenticação. A adulteração de vinagre orgânico por vinagre convencional de maçã promove alterações mais localizadas na organização metabolômica da matriz, associadas às diferenças entre os sistemas de produção, resultando em modificações composicionais que podem ser capturadas por um número relativamente pequeno de domínios espectrais altamente informativos. Em contraste, a adulteração com vinagre de álcool promove uma reorganização mais ampla da organização metabolômica da matriz, afetando simultaneamente diferentes processos metabólicos e resultando em uma distribuição mais difusa da informação preditiva ao longo da assinatura espectral da matriz. Nessas condições, a perturbação de um único domínio espectral produz um impacto proporcionalmente menor sobre as predições do classificador.

Esses resultados reforçam um conceito importante da inteligência artificial explicável aplicada à química analítica: a fidelidade das explicações depende não apenas do algoritmo de interpretação empregado, mas também da forma como a informação química está organizada no conjunto de dados. Impressões digitais espectroscópicas obtidas de matrizes biológicas complexas frequentemente contêm informações metabolômicas altamente correlacionadas e distribuídas entre múltiplos domínios espectrais, de modo que diferentes problemas de classificação podem naturalmente apresentar distintos níveis de concentração das explicações. Consequentemente, uma menor fidelidade não deve ser necessariamente interpretada como uma limitação do método de explicabilidade, mas como uma evidência de que a informação preditiva está intrinsecamente distribuída ao longo da assinatura espectral da matriz (Ahmed et al., 2025; Ribeiro et al., 2026; Arrighi et al., 2026; Molnar, 2022; Wishart, 2019; Emwas et al., 2021).

Comportamento semelhante tem sido relatado em estudos recentes de inteligência artificial explicável aplicados a dados espectroscópicos, nos quais explicações altamente fiéis estiveram associadas a perturbações químicas mais localizadas, enquanto alterações metabolômicas mais distribuídas resultaram em perfis de relevância compartilhados entre múltiplos domínios espectrais enriquecidos em ressonâncias correlacionadas (Ahmed et al., 2025; Ribeiro et al., 2026; Arrighi et al., 2026; Molnar, 2022; Wishart, 2019; Emwas et al., 2021).

De modo geral, as análises de fidelidade demonstram que o SMX captura de forma confiável as informações espectrais que governam as predições do modelo, ao mesmo tempo em que evidencia como diferentes mecanismos de fraude modificam a organização metabolômica da matriz e sua representação na assinatura espectral da matriz. Além de validar a robustez das explicações produzidas, esses resultados demonstram que abordagens de explicabilidade baseadas em perturbação podem, por si só, fornecer informações relevantes sobre a complexidade dos sistemas químicos subjacentes, ampliando seu papel da simples

interpretação do comportamento de modelos de aprendizado de máquina para a caracterização de padrões metabolômicos associados à autenticidade de alimentos.

3.5.1. Comparação entre o SMX e a Interpretação Quimiométrica Convencional

A interpretação de modelos espectroscópicos de classificação tem sido tradicionalmente baseada em ferramentas quimiométricas, como a Importância da Variável na Projeção (Variable Importance in Projection – VIP), que estimam a contribuição de variáveis individuais para a estrutura de variáveis latentes de modelos baseados em PLS. Embora amplamente empregadas em estudos de autenticação de alimentos, as pontuações de VIP são calculadas no nível de variáveis individuais e, por esse motivo, podem tornar-se de difícil interpretação quando aplicadas a conjuntos de dados espectroscópicos altamente correlacionados e caracterizados por extensa sobreposição de sinais, como as impressões digitais obtidas por LF-¹H NMR.

Os resultados obtidos neste estudo evidenciam diferenças conceituais fundamentais entre o VIP e o *Spectral Model eXplainer*. Enquanto o VIP identifica variáveis que contribuem para a construção da estrutura de variáveis latentes do modelo, o SMX avalia diretamente a dependência das predições do classificador em relação a domínios espectrais, por meio de experimentos controlados de perturbação. Dessa forma, as explicações geradas pelo SMX permanecem diretamente relacionadas a regiões espectrais quimicamente coerentes, permitindo interpretações compatíveis com a organização metabolômica da matriz e, quando respaldadas pela literatura, com processos bioquímicos associados à fermentação. Essa abordagem produz explicações mais consistentes com a forma pela qual dados espectroscópicos são interpretados na química analítica e na metabolômica não direcionada (Emwas et al., 2019; Mialon et al., 2023; Nie et al., 2019).

Essas diferenças conceituais tornaram-se particularmente evidentes durante a análise de fidelidade (Figura 5). Em ambos os cenários de autenticação, as sequências de perturbação

orientadas pelo SMX produziram, de forma consistente, deslocamentos mais pronunciados nas predições do modelo do que aqueles obtidos a partir dos ranqueamentos baseados em VIP. Esse comportamento demonstra que os domínios espectrais priorizados pelo SMX apresentam maior correspondência com as informações efetivamente utilizadas pelo classificador durante seu processo de decisão. Em contraste, variáveis com elevados valores de VIP, embora relevantes para a construção da estrutura de variáveis latentes, produziram impactos comparativamente menores sobre as predições quando as informações espectrais correspondentes foram removidas.

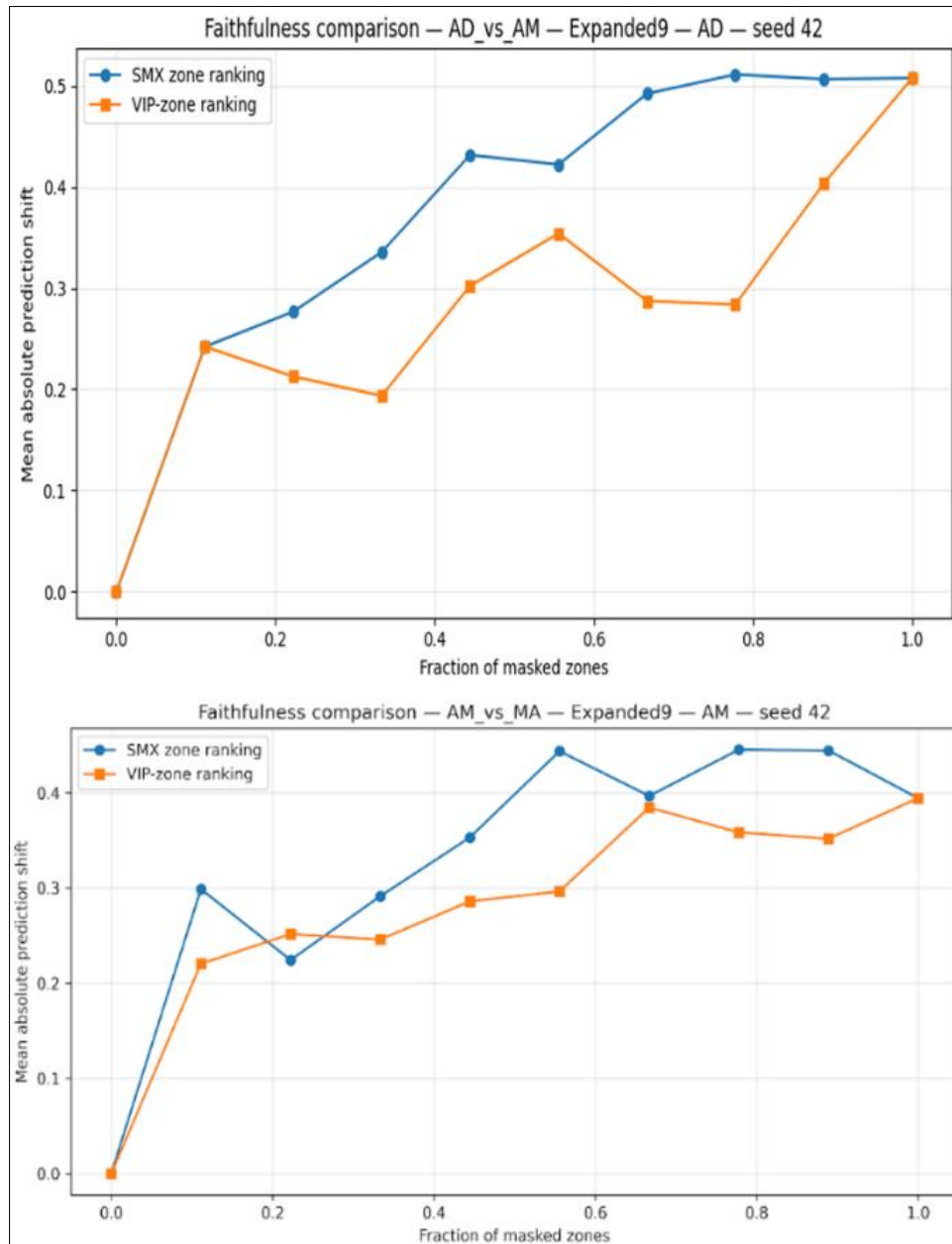


Figura 5. Comparação entre os ranqueamentos de perturbação orientados pelo *Spectral Model explainer* e pela Importância da Variável na Projeção para os modelos de autenticação. **Painel superior:** modelo de autenticação AD vs. AM; **Painel inferior:** modelo de autenticação AM vs. MA.

Essa distinção reflete os diferentes objetivos das duas abordagens. O VIP quantifica a contribuição estatística das variáveis para a construção das componentes latentes do modelo,

mas não avalia explicitamente sua influência sobre as previsões geradas pelo classificador. Em contraste, métodos de explicabilidade baseados em perturbação mensuram diretamente a sensibilidade das previsões a modificações controladas nos dados de entrada, permitindo que as explicações obtidas representem com maior fidelidade o mecanismo de decisão aprendido pelo modelo (Molnar, 2022; Ribeiro et al., 2026; Arrighi et al., 2026).

Dessa forma, o SMX não substitui as abordagens quimiométricas convencionais, mas fornece informações complementares ao relacionar o comportamento preditivo do modelo com domínios espectrais enriquecidos em ressonâncias e quimicamente coerentes.

Uma vantagem adicional do SMX reside em sua capacidade de interpretação em múltiplas escalas. Como a metodologia opera sobre domínios espectrais predefinidos, em vez de variáveis espectrais isoladas, a relevância pode ser interpretada em termos de regiões espectrais enriquecidas em ressonâncias, assinaturas composicionais e, quando suportado pela literatura, de conjuntos de metabólitos associados a determinadas rotas bioquímicas, em vez de pontos espectrais individuais. Essa característica é particularmente importante para impressões digitais obtidas por ^1H LF-NMR, nas quais a extensa sobreposição de ressonâncias limita a interpretação química de variáveis isoladas e a informação discriminativa emerge da variação coordenada entre múltiplas regiões espectrais da organização metabolômica da matriz.

Sob uma perspectiva mais ampla, esses resultados demonstram como a inteligência artificial explicável pode extrapolar a interpretação do comportamento de modelos computacionais e consolidar-se como uma ferramenta analítica para as ciências espectroscópicas. Ao traduzir decisões complexas de modelos de aprendizado de máquina em domínios espectrais quimicamente coerentes, o SMX estabelece uma ponte interpretável entre a modelagem preditiva e o conhecimento químico disponível na literatura. Essa capacidade não apenas facilita a comunicação entre cientistas de dados, químicos analíticos e profissionais envolvidos em processos regulatórios, como também aumenta a confiança na aplicação de

modelos de aprendizado de máquina destinados ao controle de qualidade industrial, à autenticação de alimentos e às investigações regulatórias. Em conjunto, esses resultados demonstram que métodos de explicabilidade baseados em perturbação representam um passo importante para o desenvolvimento de sistemas de inteligência artificial transparentes, cientificamente interpretáveis e operacionalmente confiáveis para aplicações em química analítica.

3.5.2. Implicações para a Autenticação Explicável de Alimentos e Bebidas

A rápida incorporação de métodos de aprendizado de máquina à autenticação de alimentos tem ampliado significativamente a capacidade de detectar fraudes, verificar autenticidade e classificar matrizes alimentícias complexas a partir de impressões digitais analíticas não direcionadas. Entretanto, a adoção prática dessas abordagens depende, cada vez mais, não apenas do desempenho preditivo dos modelos, mas também da capacidade de explicar como suas decisões são geradas. No contexto da química analítica, a transparência dos modelos tornou-se um requisito fundamental, uma vez que conclusões científicas, decisões regulatórias e procedimentos de controle de qualidade industrial exigem justificativas cientificamente interpretáveis, e não apenas elevada capacidade de predição (Molnar, 2022; Wishart, 2019).

O presente estudo demonstra que a explicabilidade pode transformar modelos espectroscópicos de classificação de ferramentas exclusivamente preditivas em instrumentos analíticos capazes de produzir interpretações quimicamente fundamentadas. O SMX gera explicações naturalmente compatíveis com a interpretação de impressões digitais espectroscópicas complexas. Essa estratégia baseada em domínios espectrais é particularmente vantajosa para ^1H LF -NMR (Emwas et al., 2021; Lolli & Caligiani, 2024).

Uma implicação importante dos resultados obtidos neste estudo é que a explicabilidade não deve ser encarada apenas como uma técnica *post hoc* de visualização, mas como uma

camada analítica adicional capaz de revelar como diferentes padrões espectrais contribuem para o comportamento do modelo. Os distintos perfis de relevância observados para os dois cenários de adulteração ilustram claramente esse conceito. Em vez de produzir ranqueamentos universais de importância, o SMX identificou assinaturas composicionais compatíveis com as diferentes alterações na organização metabólica da matriz introduzidas por cada mecanismo de fraude, estabelecendo uma conexão entre o comportamento do classificador e padrões espectrais quimicamente coerentes, interpretados à luz do conhecimento disponível sobre os processos fermentativos (Wishart, 2019; Emwas et al., 2021).

As análises de fidelidade reforçam ainda mais o valor analítico da estrutura proposta. A elevada concordância entre os ranqueamentos de relevância obtidos por perturbação e o comportamento efetivamente apresentado pelo modelo demonstra que as explicações produzidas pelo SMX representam de forma confiável as informações espectrais exploradas pelo classificador durante seu mecanismo de decisão aprendido. Diferentemente de abordagens convencionais de ranqueamento de variáveis, como a VIP, que quantifica a contribuição estatística das variáveis para a construção da estrutura de variáveis latentes, o SMX avalia diretamente a influência dos domínios espectrais sobre as predições por meio de perturbações controladas. Como consequência, as explicações produzidas representam com maior fidelidade as informações efetivamente responsáveis pelas decisões de classificação (Molnar, 2022; Ribeiro et al., 2026).

Embora demonstrado neste trabalho utilizando impressões digitais de ^1H LF -NMR de vinagres comerciais, o arcabouço conceitual que fundamenta o SMX é independente tanto da plataforma analítica quanto da matriz alimentícia investigada. Considerando que diversas técnicas, incluindo espectroscopia no infravermelho, infravermelho próximo, Raman, imageamento hiperespectral e RMN de alto campo, produzem impressões digitais altamente correlacionadas e caracterizadas por informações químicas distribuídas, o paradigma baseado

em domínios espectrais adotado pelo SMX constitui uma estratégia geral para integrar a inteligência artificial explicável à química analítica. Dessa forma, espera-se que essa abordagem possa ser aplicada a uma ampla variedade de problemas relacionados à autenticação, rastreabilidade, avaliação da qualidade e monitoramento de processos envolvendo conjuntos de dados químicos complexos.

4. Conclusão

Este estudo demonstra que a inteligência artificial explicável pode ampliar de maneira significativa o papel do aprendizado de máquina na autenticação espectroscópica de alimentos. Ao introduzir o *Spectral Model eXplainer*, demonstramos que modelos de classificação podem evoluir além de algoritmos estritamente preditivos para atuar como ferramentas analíticas capazes de fornecer interpretações quimicamente fundamentadas sobre seu próprio comportamento. Por meio da integração entre explicabilidade baseada em perturbação, segmentação de domínios espectrais quimicamente coerentes e análise de *Local Reaching Centrality*, o SMX permitiu identificar domínios espectrais enriquecidos em ressonâncias que exerceram maior influência sobre as decisões do classificador, estabelecendo uma conexão interpretável entre o comportamento do modelo e a organização metabolômica da matriz.

Aplicado a impressões digitais de ^1H LF-NMR de vinagres comerciais, o SMX discriminou com sucesso dois diferentes cenários de autenticação e revelou assinaturas composicionais específicas compatíveis com alterações distribuídas em regiões espectrais associadas a compostos derivados da fermentação, ácidos orgânicos, compostos relacionados ao etanol e regiões enriquecidas em carboidratos. O método também apresentou elevada fidelidade, superando de forma consistente a interpretação convencional baseada em VIP ao produzir explicações que representaram com maior fidelidade as informações espectrais efetivamente exploradas pelo classificador. Esses resultados indicam que abordagens de

explicabilidade baseadas em perturbação fornecem uma representação mais confiável e quimicamente coerente do mecanismo de decisão aprendido pelo modelo do que métricas tradicionais baseadas em variáveis latentes.

Além da autenticação de vinagres, a principal contribuição deste trabalho consiste no estabelecimento de um arcabouço conceitual para a aplicação de aprendizado de máquina interpretável em dados espectroscópicos. Em vez de tratar a explicabilidade como uma etapa complementar de visualização executada após o desenvolvimento do modelo, o SMX incorpora a interpretação química ao próprio fluxo de trabalho analítico, permitindo relacionar o comportamento preditivo do modelo a domínios espectrais quimicamente coerentes.

Embora demonstrado neste estudo utilizando ^1H LF-NMR em baixo campo, os princípios conceituais que fundamentam o SMX são independentes da plataforma analítica empregada e potencialmente aplicáveis a outras técnicas espectroscópicas. Da mesma forma, o método apresenta potencial de aplicação em problemas relacionados à autenticação de alimentos, metabolômica, monitoramento de processos e avaliação da qualidade. À medida que a inteligência artificial se torna cada vez mais integrada à química analítica, cresce também a necessidade de modelos transparentes, cientificamente interpretáveis e compatíveis com os requisitos regulatórios.

Em uma perspectiva mais ampla, este trabalho reforça a transição da inteligência artificial explicável de uma ferramenta voltada exclusivamente à interpretação computacional de modelos para uma metodologia cientificamente fundamentada de apoio à interpretação de dados químicos complexos. Ao estabelecer uma conexão interpretável entre o desempenho preditivo dos modelos e a organização química das impressões digitais espectroscópicas, o SMX propõe um novo paradigma para integrar a inteligência artificial explicável à química analítica, aproximando modelagem preditiva e interpretação química em um único fluxo de trabalho científico. Essa abordagem fornece uma base sólida para o desenvolvimento da

próxima geração de métodos de aprendizado de máquina que sejam simultaneamente transparentes, confiáveis, quimicamente interpretáveis e aplicáveis em ambientes industriais, contribuindo para consolidar a inteligência artificial explicável como um componente essencial da química analítica moderna e da ciência de alimentos.

5. Agradecimentos

Os autores também agradecem o apoio financeiro concedido, em parte, pela Coordenação de Aperfeiçoamento de Pessoal de Nível Superior – Brasil (CAPES) – Código de Financiamento 001. Este estudo foi financiado pelo Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) – processo nº 307846/2020-2.

6. Referências

- Ahmed, Md. T., Ahmed, M. W., & Kamruzzaman, M. (2025). A systematic review of explainable artificial intelligence for spectroscopic agricultural quality assessment. *Computers and Electronics in Agriculture*, 235, 110354. <https://doi.org/10.1016/j.compag.2025.110354>
- Arrighi, L., de Moraes, I. A., Zullich, M., Simonato, M., Barbin, D. F., & Barbon Junior, S. (2026). Explainable artificial intelligence techniques for interpretation of food models: a review. *Artificial Intelligence Review*, 59(8), 176. <https://doi.org/10.1007/s10462-026-11553-2>
- Arrighi, L., Pennella, L., Tavares, G. M., & Junior, S. B. (2024). *Decision Predicate Graphs: Enhancing Interpretability in Tree Ensembles*. <http://arxiv.org/abs/2404.02942>
- Augusti, R., Fulgêncio, A. C. C., Nogueira, H. M., Gomes, J. C. L., dos Santos, L. B., de Macedo, A. N., Porto, B. L. S., Sena, M. M., & Almeida, M. R. (2024). Enhancing food authentication screening through the integration of chemometrics and ambient ionization

- mass spectrometry: A comprehensive review. *Trends in Food Science & Technology*, 147, 104480. <https://doi.org/10.1016/j.tifs.2024.104480>
- Avanzi Barbosa Mascareli, V., Galvan, D., Craveiro de Andrade, J., Lelis, C., Adam Conte-Junior, C., Michelino Gaeta Lopes, G., César de Macedo Júnior, F., & Aparecida Spinosa, W. (2023). Spectralprint techniques coupled with chemometric tools for vinegar classifications. *Food Chemistry*, 410, 135373. <https://doi.org/10.1016/j.foodchem.2022.135373>
- Biancolillo, A., Marini, F., Ruckebusch, C., & Vitale, R. (2020). Chemometric Strategies for Spectroscopy-Based Food Authentication. *Applied Sciences*, 10(18), 6544. <https://doi.org/10.3390/app10186544>
- Boffo, E. F., Tavares, L. A., Ferreira, M. M. C., & Ferreira, A. G. (2009). Classification of Brazilian vinegars according to their ^1H NMR spectra by pattern recognition analysis. *LWT - Food Science and Technology*, 42(9), 1455–1460. <https://doi.org/10.1016/j.lwt.2009.05.008>
- Brasil, Ministério da Agricultura, Pecuária e Abastecimento (MAPA), 2009. Instrução Normativa No. 19, de 28 de maio de 2009. *Approves the mechanisms for control and information on organic quality* [in Portuguese]. Diário Oficial da União, Brasília, DF, 29 May 2009. Available at: <https://www.gov.br/agricultura/pt-br/assuntos/sustentabilidade/organicos/legislacao/portugues/instrucao-normativa-no-19-de-28-de-maio-de-2009-mecanismos-de-controle-e-formas-de-organizacao.pdf/view> (accessed 20 June 2026).
- Caligiani, A., Acquotti, D., Palla, G., & Bocchi, V. (2007). Identification and quantification of the main organic components of vinegars by high resolution ^1H NMR spectroscopy. *Analytica Chimica Acta*, 585(1), 110–119. <https://doi.org/10.1016/j.aca.2006.12.016>
- Cavdaroglu, C., & Ozen, B. (2023). Applications of UV–Visible, Fluorescence and Mid-

- Infrared Spectroscopic Methods Combined with Chemometrics for the Authentication of Apple Vinegar. *Foods*, 12(6), 1139. <https://doi.org/10.3390/foods12061139>
- Ceschin, M., Arrighi, L., Longo, L., & Barbon Junior, S. (2026). *Extending Decision Predicate Graphs for Comprehensive Explanation of Isolation Forest* (pp. 271–293). https://doi.org/10.1007/978-3-032-08324-1_12
- Ciampa, A., Danesi, F., & Picone, G. (2022). NMR-Based Metabolomics for a More Holistic and Sustainable Research in Food Quality Assessment: A Narrative Review. *Applied Sciences*, 13(1), 372. <https://doi.org/10.3390/app13010372>
- Consonni, R., Cagliani, L. R., Benevelli, F., Spraul, M., Humpfer, E., & Stocchero, M. (2008). NMR and Chemometric methods: A powerful combination for characterization of Balsamic and Traditional Balsamic Vinegar of Modena. *Analytica Chimica Acta*, 611(1), 31–40. <https://doi.org/10.1016/j.aca.2008.01.065>
- Consonni, R., Cagliani, L. R., Rinaldini, S., & Incerti, A. (2008). Analytical method for authentication of Traditional Balsamic Vinegar of Modena. *Talanta*, 75(3), 765–769. <https://doi.org/10.1016/j.talanta.2007.12.005>
- Conte, T. L., de Andrade, J. C., da Silva Borges, R. L., Effting, L., Ravaglia, L. M., Alcantara, G. B., Conte-Junior, C. A., Bona, E., & Galvan, D. (2025). Benchtop NMR spectroscopy and multivariate calibration for fast prediction of oxidation stability in commercial biodiesel. *Microchemical Journal*, 212, 113149. <https://doi.org/10.1016/j.microc.2025.113149>
- da Silva, L. P., Oliveira, M. D. L., & Villa, J. E. L. (2025). A comparison of decision tree-based algorithms for food discrimination using vibrational spectroscopy. *Food Chemistry*, 488, 144909. <https://doi.org/10.1016/j.foodchem.2025.144909>
- Dimakopoulou-Papazoglou, D., Serrano, S., Rodríguez, I., Ploskas, N., Koutsoumanis, K., & Katsanidis, E. (2025). FTIR spectroscopy combined with machine learning for the

- classification of Mediterranean honey based on origin. *Journal of Food Composition and Analysis*, 144, 107778. <https://doi.org/10.1016/j.jfca.2025.107778>
- Emwas, A.-H., Roy, R., McKay, R. T., Tenori, L., Saccenti, E., Gowda, G. A. N., Raftery, D., Alahmari, F., Jaremko, L., Jaremko, M., & Wishart, D. S. (2019). NMR Spectroscopy for Metabolomics Research. *Metabolites*, 9(7), 123. <https://doi.org/10.3390/metabo9070123>
- Galvan, D., de Aguiar, L. M., Bona, E., Marini, F., & Killner, M. H. M. (2023). Successful combination of benchtop nuclear magnetic resonance spectroscopy and chemometric tools: A review. *Analytica Chimica Acta*, 1273, 341495. <https://doi.org/10.1016/j.aca.2023.341495>
- Galvan, D., Tanamati, A. A. C., Casanova, F., Danieli, E., Bona, E., & Killner, M. H. M. (2021). Compact low-field NMR spectroscopy and chemometrics applied to the analysis of edible oils. *Food Chemistry*, 365, 130476. <https://doi.org/10.1016/j.foodchem.2021.130476>
- García-Pérez, P., Becchi, P. P., Zhang, L., Rocchetti, G., & Lucini, L. (2024). Metabolomics and chemometrics: The next-generation analytical toolkit for the evaluation of food quality and authenticity. *Trends in Food Science & Technology*, 147, 104481. <https://doi.org/10.1016/j.tifs.2024.104481>
- Goyal, R., Singha, P., & Singh, S. K. (2024). Spectroscopic food adulteration detection using machine learning: Current challenges and future prospects. *Trends in Food Science & Technology*, 146, 104377. <https://doi.org/10.1016/j.tifs.2024.104377>
- Grassi, S., & Alamprese, C. (2024a). Spectroscopic Non-targeted Techniques in Combination with Linear Discriminant Analysis for Wine Vinegar Authentication. *Food and Bioprocess Technology*, 17(2), 479–488. <https://doi.org/10.1007/s11947-023-03143-9>
- Grassi, S., Borgonovo, G., Gennaro, M., & Alamprese, C. (2024). NMR-based approach to detect white wine vinegar fraud. *Food Chemistry*, 456, 139953. <https://doi.org/10.1016/j.foodchem.2024.139953>

- Lolli, V., & Caligiani, A. (2024). How nuclear magnetic resonance contributes to food authentication: current trends and perspectives. *Current Opinion in Food Science*, 58, 101200. <https://doi.org/10.1016/j.cofs.2024.101200>
- Lozano-Castellón, J., Laveriano-Santos, E. P., Abuhabib, M. M., Pozzoli, C., Pérez, M., Vallverdú-Queralt, A., & Lamuela-Raventós, R. M. (2024). Proven traceability strategies using chemometrics for organic food authenticity. *Trends in Food Science & Technology*, 147, 104430. <https://doi.org/10.1016/j.tifs.2024.104430>
- Mialon, N., Roig, B., Capodanno, E., & Cadiere, A. (2023). Untargeted metabolomic approaches in food authenticity: A review that showcases biomarkers. *Food Chemistry*, 398, 133856. <https://doi.org/10.1016/j.foodchem.2022.133856>
- Michal, C. A. (2020). Low-cost low-field NMR and MRI: Instrumentation and applications. *Journal of Magnetic Resonance*, 319, 106800. <https://doi.org/10.1016/j.jmr.2020.106800>
- Molnar, C. (2022). Interpretable Machine Learning: A Guide for Making Black Box Models Explainable (2nd ed.). Munich, Germany: Christoph Molnar. Available at: <https://christophm.github.io/interpretable-ml-book/>
- Nie, J., Li, Y., Xing, J., Chao, J., Qin, X., & Li, Z. (2019). Comparison of two types of vinegar with different aging times by NMR-based metabolomic approach. *Journal of Food Biochemistry*, 43(5), e12835. <https://doi.org/10.1111/jfbc.12835>
- Ousaaïd, D., Laaroussi, H., Bakour, M., ElGhouizi, A., Aboulghazi, A., Lyoussi, B., & ElArabi, I. (2020). Beneficial Effects of Apple Vinegar on Hyperglycemia and Hyperlipidemia in Hypercaloric-Fed Rats. *Journal of Diabetes Research*, 2020, 1–7. <https://doi.org/10.1155/2020/9284987>
- Parker, T., Limer, E., Watson, A. D., Defernez, M., Williamson, D., & Kemsley, E. K. (2014). 60MHz ¹H NMR spectroscopy for the analysis of edible oils. *TrAC Trends in Analytical Chemistry*, 57, 147–158. <https://doi.org/10.1016/j.trac.2014.02.006>

- Ribeiro, J. V., Goncalves, R. F., Melquiades, F. L., & Junior, S. B. (2026). *Spectral Model eXplainer: a chemically-grounded explainability framework for spectral-based machine learning models*. <http://arxiv.org/abs/2605.02684>
- Shih, M., Yuan, Y., & Shi, G. (2024). Comparative analysis of LDA, PLS-DA, SVM, RF, and voting ensemble for discrimination origin in greenish-white to white nephrites using LIBS. *Journal of Analytical Atomic Spectrometry*, 39(6), 1560–1570. <https://doi.org/10.1039/D3JA00464C>
- Štornik, A., Skok, B., & Trček, J. (2016). Comparison of Cultivable Acetic Acid Bacterial Microbiota between Organic and Conventional Apple Cider Vinegar. *Food Technology and Biotechnology*, 54. <https://doi.org/10.17113/ftb.54.01.16.4082>
- Tang, F., Banker, T., Green, H. S., Wang, S. C., & Hatzakis, E. (2024). Analysis and authentication of avocado oil by low-field benchtop NMR spectroscopy and chemometrics. *Journal of Food Science*, 89(7), 4276–4285. <https://doi.org/10.1111/1750-3841.17142>
- Vinothkanna, A., Dar, O. I., Liu, Z., & Jia, A.-Q. (2024). Advanced detection tools in food fraud: A systematic review for holistic and rational detection method based on research and patents. *Food Chemistry*, 446, 138893. <https://doi.org/10.1016/j.foodchem.2024.138893>
- Wishart, D. S. (2019). Metabolomics for Investigating Physiological and Pathophysiological Processes. *Physiological Reviews*, 99(4), 1819–1875. <https://doi.org/10.1152/physrev.00035.2018>
- Yang, L., Chen, F., Chen, L., Zhang, Z., Chen, J., Wang, J., Cheng, X., Feng, J., Bao, Q., & Liu, C. (2023). An easy-built Halbach magnet for LF-NMR with high homogeneity using optimized target-field passive shimming method. *Journal of Magnetic Resonance*, 357, 107582. <https://doi.org/10.1016/j.jmr.2023.107582>

- Yang, Y., Zhang, L., Qu, X., Zhang, W., Shi, J., & Xu, X. (2024). Enhanced food authenticity control using machine learning-assisted elemental analysis. *Food Research International*, 198, 115330. <https://doi.org/10.1016/j.foodres.2024.115330>
- Zhang, S., Chen, J., Gao, F., Su, W., Li, T., & Wang, Y. (2024). Foodomics as a Tool for Evaluating Food Authenticity and Safety from Field to Table: A Review. *Foods*, 14(1), 15. <https://doi.org/10.3390/foods14010015>

CONCLUSÃO GERAL

A combinação entre espectroscopia de Ressonância Magnética Nuclear de Hidrogênio de Baixo Campo (^1H LF-NMR), quimiometria, aprendizado de máquina e Inteligência Artificial Explicável (XAI) mostrou-se adequada para a autenticação de vinagres de maçã. Os *fingerprints* espectrais obtidos permitiram discriminar diferentes cenários de adulteração por meio de modelos supervisionados, evidenciando que a informação química contida nos espectros foi suficiente para sustentar classificações robustas, mesmo utilizando uma plataforma instrumental de baixo campo.

A incorporação de uma abordagem baseada em Inteligência Artificial Explicável permitiu complementar o desempenho dos modelos de classificação por meio da identificação e interpretação das regiões espectrais mais relevantes para o processo decisório. A associação entre explicabilidade e *fingerprints* espectrais estabeleceu uma relação objetiva entre a resposta computacional dos modelos e características químicas da matriz analisada, fornecendo uma interpretação consistente dos fatores responsáveis pela diferenciação entre amostras autênticas e adulteradas. Esses resultados evidenciaram que métodos de XAI podem aumentar a confiabilidade da interpretação de modelos quimiométricos aplicados à autenticação de alimentos.

Em conjunto, os resultados demonstraram que a integração entre ^1H LF-NMR, quimiometria, aprendizado de máquina e Inteligência Artificial Explicável constituiu uma abordagem analítica capaz de combinar elevado desempenho de classificação com interpretabilidade química. Além de permitir a discriminação entre diferentes cenários de adulteração, a metodologia desenvolvida possibilitou compreender quais regiões espectrais sustentaram as decisões dos modelos e como essas informações se relacionaram com alterações na composição química das amostras, ampliando o potencial analítico de estratégias baseadas em *chemical fingerprinting*.

Embora o estudo tenha sido conduzido utilizando vinagres de maçã como modelo experimental e espectroscopia de ^1H LF-NMR como plataforma analítica, a estrutura metodológica apresentada pode ser adaptada a outras matrizes alimentícias e a diferentes técnicas de *fingerprinting* espectral. Dessa forma, a abordagem proposta fornece uma estratégia analítica aplicável a estudos de autenticação, rastreabilidade e controle de qualidade, especialmente em situações nas quais sejam desejáveis métodos rápidos, não destrutivos e acompanhados de interpretações quimicamente fundamentadas.

ANEXO I

GLOSSÁRIO

A

Acurácia (*Accuracy*) — Métrica de desempenho correspondente à proporção de amostras corretamente classificadas em relação ao número total de amostras avaliadas.

Alinhamento espectral — Etapa de pré-processamento destinada a corrigir pequenos deslocamentos dos sinais entre espectros, aumentando a correspondência entre variáveis espectrais equivalentes.

Análise de Componentes Principais (PCA, *Principal Component Analysis*) — Método multivariado não supervisionado de redução de dimensionalidade que transforma variáveis correlacionadas em componentes principais ortogonais, preservando progressivamente a variabilidade dos dados.

Análise Discriminante por Mínimos Quadrados Parciais (PLS-DA, *Partial Least Squares Discriminant Analysis*) — Método supervisionado de classificação baseado na projeção das variáveis originais em um espaço de variáveis latentes que maximiza sua relação com as classes de interesse.

Área sob a curva (AUC, *Area Under the Curve*) — Medida que resume quantitativamente o comportamento de uma curva. Neste trabalho, a AUC normalizada foi empregada na avaliação da fidelidade das explicações produzidas pelo SMX.

B

Bagging — Estratégia de aprendizado baseada na geração e combinação de subconjuntos ou modelos obtidos por reamostragem, utilizada para aumentar estabilidade e robustez. No SMX, integra o procedimento de geração e organização dos predicados.

C

Classificador — Modelo matemático treinado para atribuir uma observação a uma dentre duas ou mais classes previamente definidas.

ComDim (*Common Dimensions Analysis*) — Método multibloco de análise multivariada utilizado para identificar dimensões comuns compartilhadas por diferentes blocos de dados.

Correção de linha de base — Pré-processamento espectral destinado a remover variações sistemáticas ou tendências do sinal de fundo que não estão diretamente associadas às ressonâncias de interesse.

Correção por Mínimos Quadrados Assimétricos (ALS, *Asymmetric Least Squares*) — Método empregado para estimar e remover a linha de base de sinais espectrais por meio de mínimos quadrados com penalização assimétrica.

D

Deslocamento químico (*Chemical shift*) — Posição de uma ressonância em um espectro de RMN, geralmente expressa em partes por milhão (ppm), relacionada ao ambiente químico dos núcleos observados.

Domínio espectral — Intervalo definido do espectro agrupando variáveis espectrais relacionadas. No SMX, os domínios são tratados como unidades de interpretação para quantificar a relevância de diferentes regiões do espectro para a classificação.

E

Especificidade — Capacidade de um classificador de reconhecer corretamente amostras pertencentes à classe negativa, correspondendo à proporção de verdadeiros negativos corretamente identificados.

F

F1-score — Média harmônica entre precisão e recall, utilizada para avaliar simultaneamente a capacidade do modelo de evitar falsos positivos e de identificar corretamente as amostras de uma classe.

F1-score macro (*macro-F1*) — Média não ponderada dos valores de F1-score calculados individualmente para cada classe, atribuindo igual importância às classes independentemente do número de amostras.

Fidelidade (*Faithfulness/Fidelity*) — Grau em que uma explicação representa efetivamente o

comportamento do modelo que pretende explicar. No SMX, foi avaliada por meio do impacto do mascaramento progressivo de domínios espectrais sobre as predições do classificador.

Fingerprinting espectral — Abordagem analítica não direcionada que utiliza o perfil espectral global de uma amostra como sua assinatura química, sem exigir necessariamente a identificação ou quantificação individual de todos os compostos presentes.

Fold — Subconjunto dos dados utilizado durante procedimentos de validação cruzada. A cada iteração, diferentes *folds* são empregados para treinamento e validação do modelo.

Framework — estrutura organizada que integra métodos, procedimentos e ferramentas para a execução de uma abordagem ou tarefa específica.

I

Inteligência Artificial Explicável (XAI, *eXplainable Artificial Intelligence*) — Conjunto de métodos destinados a tornar compreensíveis e interpretáveis as decisões produzidas por modelos de inteligência artificial ou aprendizado de máquina.

Interpretabilidade — Grau em que o comportamento, as relações internas ou as decisões de um modelo podem ser compreendidos e relacionados às características das variáveis de entrada.

Isolation Forest — Algoritmo não supervisionado de detecção de anomalias que identifica observações atípicas com base na facilidade com que podem ser isoladas por particionamentos sucessivos dos dados.

Inlier — Observação considerada compatível com o padrão predominante do conjunto de dados e, portanto, não identificada como anômala pelo método de detecção de outliers.

L

LF-NMR (*Low-Field Nuclear Magnetic Resonance*) — Ressonância Magnética Nuclear de baixo campo, modalidade de RMN baseada em campos magnéticos de menor intensidade, frequentemente produzidos por magnetos permanentes.

LRC (*Local Reaching Centrality*) — Métrica de centralidade aplicada pelo SMX ao grafo de predicados para quantificar a influência relativa de diferentes domínios espectrais sobre o comportamento do classificador.

Loading — Coeficiente que expressa a contribuição de uma variável original para determinado componente principal ou variável latente em um modelo multivariado.

M

Machine Learning (Aprendizado de Máquina) — Campo da inteligência artificial dedicado ao desenvolvimento de modelos capazes de aprender padrões a partir de dados e utilizá-los para classificação, regressão, reconhecimento de padrões ou outras tarefas.

Mascaramento espectral (*Spectral masking*) — Procedimento de perturbação no qual os valores de determinada região espectral são substituídos por valores de referência para avaliar o impacto da remoção daquela informação sobre a predição do modelo.

Matriz de confusão — Representação tabular que compara as classes reais e preditas por um classificador, permitindo visualizar acertos e diferentes tipos de erro de classificação.

O

Outlier — Observação que apresenta comportamento substancialmente diferente do padrão predominante do conjunto de dados e que pode representar variabilidade extrema, erro experimental ou condição não representativa.

P

Padronização *z-score* — Transformação das variáveis pela subtração de sua média e divisão pelo desvio-padrão, resultando em variáveis centradas e escaladas.

PC1 (*First Principal Component*) — Primeiro componente principal obtido por PCA, correspondente à direção que explica a maior parcela possível da variância dos dados. No SMX, o PC1 de cada domínio espectral é empregado como representação reduzida daquela região.

Perturbação — Alteração controlada das variáveis de entrada utilizada para avaliar como uma região ou característica influencia a saída de um modelo.

Pipeline — Estrutura computacional que organiza sequencialmente etapas de pré-processamento, transformação e modelagem, garantindo que os mesmos procedimentos sejam aplicados adequadamente aos dados de treinamento e validação.

ppm (partes por milhão) — Unidade relativa utilizada para expressar deslocamentos químicos

em espectroscopia de RMN.

Precisão (Precision) — Proporção das amostras classificadas como pertencentes a determinada classe que realmente pertencem a essa classe.

Pré-processamento espectral — Conjunto de transformações aplicadas aos espectros antes da modelagem com o objetivo de reduzir variações indesejadas e melhorar a comparabilidade das amostras.

Q

Quimiometria — Área que aplica métodos matemáticos, estatísticos e computacionais à extração de informação química de dados experimentais multivariados.

R

Recall (Sensibilidade) — Proporção das amostras pertencentes a determinada classe que são corretamente identificadas pelo classificador.

RMN de ^1H — Ressonância Magnética Nuclear de hidrogênio, técnica espectroscópica baseada na interação dos núcleos de ^1H com um campo magnético e pulsos de radiofrequência, permitindo obter informações relacionadas ao ambiente químico desses núcleos.

S

Score — Coordenada de uma amostra no espaço definido pelos componentes principais ou variáveis latentes de um modelo multivariado.

SMX (*Spectral Model eXplainer*) — *Framework* de Inteligência Artificial Explicável desenvolvido para interpretar modelos baseados em dados espectrais, quantificando a contribuição de regiões espectrais para as previsões do classificador.

StandardScaler — Implementação computacional utilizada para padronizar variáveis pela transformação z-score.

V

Validação cruzada (*Cross-validation*) — Estratégia de avaliação na qual os dados são sucessivamente particionados em subconjuntos de treinamento e validação, permitindo estimar

a capacidade de generalização do modelo.

Validação cruzada aninhada (*Nested cross-validation*) — Procedimento de validação constituído por ciclos interno e externo, no qual o ciclo interno é utilizado para seleção ou otimização do modelo e o externo para estimar seu desempenho de generalização com menor viés.

Validação cruzada com agrupamento (*Group cross-validation*) — Estratégia na qual amostras relacionadas são mantidas no mesmo grupo durante a divisão entre treinamento e validação, reduzindo o risco de vazamento de informação.

Variável latente (LV) — Nova variável obtida como combinação das variáveis originais, utilizada para representar de forma reduzida estruturas relevantes presentes em dados multivariados.

VIP (*Variable Importance in Projection*) — Métrica derivada de modelos PLS que resume a contribuição das variáveis originais para a projeção e para a resposta modelada, sendo utilizada para identificar regiões ou variáveis de maior importância no modelo.

Z

z-score — Ver Padronização z-score.

ANEXO II

LISTA DE ABREVIATURAS E SIGLAS

¹H	Hidrogênio-1 (próton)
AD	Vinagre de maçã orgânico adulterado com vinagre de maçã convencional/tradicional
ALS	<i>Asymmetric Least Squares</i> (Mínimos Quadrados Assimétricos)
AM	Vinagre de maçã autêntico
ANN	<i>Artificial Neural Networks</i> (Redes Neurais Artificiais)
AUC	<i>Area Under the Curve</i> (Área sob a Curva)
BAA	Bactérias do Ácido Acético
BCAA	<i>Branched-Chain Amino Acids</i> (Aminoácidos de Cadeia Ramificada)
CLAM	Centro Laboratorial Analítico e Molecular
ComDim	<i>Common Dimensions Analysis</i> (Análise de Dimensões Comuns)
CV	<i>Cross-Validation</i> (Validação Cruzada)
EXP1	Experimento 1
EXP2	Experimento 2
FT-IR	<i>Fourier Transform Infrared Spectroscopy</i> (Espectroscopia no Infravermelho com Transformada de Fourier)
FT-NIR	<i>Fourier Transform Near-Infrared Spectroscopy</i> (Espectroscopia no Infravermelho Próximo com Transformada de Fourier)
IA	Inteligência Artificial
IQ	Instituto de Química
Lac/Ala	Domínio espectral associado principalmente a lactato e alanina
LADETEC	Laboratório de Apoio ao Desenvolvimento Tecnológico
LDA	<i>Linear Discriminant Analysis</i> (Análise Discriminante Linear)
LF-NMR	<i>Low-Field Nuclear Magnetic Resonance</i> (Ressonância Magnética Nuclear de Baixo Campo)
LRC	<i>Local Reaching Centrality</i> (Centralidade de Alcance Local)

MA	Vinagre de maçã adulterado com vinagre de álcool
Mal/Cit	Domínio espectral associado principalmente a malato e citrato
MAPA	Ministério da Agricultura e Pecuária
MHz	<i>Megahertz</i>
NAL	Núcleo de Análise de Alimentos
NIR	<i>Near-Infrared Spectroscopy</i> (Espectroscopia no Infravermelho Próximo)
NMR	<i>Nuclear Magnetic Resonance</i> (Ressonância Magnética Nuclear)
PC1	<i>First Principal Component</i> (Primeiro Componente Principal)
PCA	<i>Principal Component Analysis</i> (Análise de Componentes Principais)
PLS	<i>Partial Least Squares</i> (Mínimos Quadrados Parciais)
PLS-DA	<i>Partial Least Squares Discriminant Analysis</i> (Análise Discriminante por Mínimos Quadrados Parciais)
ppm	Partes por milhão
RF	<i>Random Forest</i>
RMN	Ressonância Magnética Nuclear
SMX	<i>Spectral Model eXplainer</i>
Succ/Cit	Domínio espectral associado principalmente a succinato e citrato
SVM	<i>Support Vector Machines</i> (Máquinas de Vetores de Suporte)
UEL	Universidade Estadual de Londrina
UFRJ	Universidade Federal do Rio de Janeiro
UFSC	Universidade Federal de Santa Catarina
UNITS	Universidade de Trieste
UV-Vis	Espectroscopia no Ultravioleta-Visível
VIP	<i>Variable Importance in Projection</i> (Importância da Variável na Projeção)
XAI	<i>eXplainable Artificial Intelligence</i> (Inteligência Artificial Explicável)
XGBoost	<i>Extreme Gradient Boosting</i>