



**UNIVERSIDADE ESTADUAL DA PARAÍBA
CAMPUS I – CAMPINA GRANDE
PRÓ-REITORIA DE PÓS-GRADUAÇÃO E PESQUISA
PROGRAMA DE PÓS GRADUAÇÃO EM QUÍMICA
MESTRADO EM QUÍMICA**



VITOR DANTAS ALVES

**MÉTODO DE CLASSIFICAÇÃO COM SELEÇÃO META-HEURÍSTICA DE
VARIÁVEIS PARA IDENTIFICAR ADIÇÃO DE LEITE BOVINO EM CAPRINO**

**CAMPINA GRANDE
2021**

VITOR DANTAS ALVES

MÉTODO DE CLASSIFICAÇÃO COM SELEÇÃO META-HEURÍSTICA DE VARIÁVEIS PARA IDENTIFICAR ADIÇÃO DE LEITE BOVINO EM CAPRINO

Trabalho de Dissertação apresentado ao Programa de Pós-Graduação em Química da Universidade Estadual da Paraíba, como requisito obrigatório à obtenção do título de Mestre em Química.

Área de concentração: Química Analítica.

Orientador: Prof. Dr. JOSÉ GERMANO VERAS NETO.

Coorientadora: Dra. ELAINY VIRGINIA DOS SANTOS PEREIRA.

**CAMPINA GRANDE - PB
2021**

É expressamente proibido a comercialização deste documento, tanto na forma impressa como eletrônica. Sua reprodução total ou parcial é permitida exclusivamente para fins acadêmicos e científicos, desde que na reprodução figure a identificação do autor, título, instituição e ano do trabalho.

A474m Alves, Vitor Dantas.

Método de classificação com seleção meta-heurística de variáveis para identificar adição de leite bovino em caprino [manuscrito] / Vitor Dantas Alves. - 2022.

55 p. : il. colorido.

Digitado.

Dissertação (Mestrado em Química - Mestrado) - Universidade Estadual da Paraíba, Centro de Ciências e Tecnologia, 2022.

"Orientação : Prof. Dr. José Germano Vêras Neto , Departamento de Química - CCT."

"Coorientação: Profa. Dra. Elaine Virginia dos Santos Pereira , UFPB - Universidade Federal da Paraíba"

1. Indústria alimentícia. 2. Autenticação do leite. 3. Quimiometria. 4. Leite caprino. 5. Controle de qualidade. I.

Título

21. ed. CDD 633.51

VITOR DANTAS ALVES

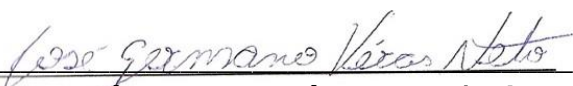
MÉTODO DE CLASSIFICAÇÃO COM SELEÇÃO META-HEURÍSTICA DE VARIÁVEIS PARA IDENTIFICAR ADIÇÃO DE LEITE BOVINO EM CAPRINO

Trabalho de Dissertação apresentado ao Programa de Pós-Graduação em Química da Universidade Estadual da Paraíba, como requisito obrigatório à obtenção do título de Mestre em Química.

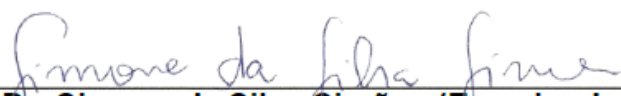
Área de concentração: Química Analítica.

Aprovada em: 26/ 01/ 2022.


BANCA EXAMINADORA



Prof. Dr. José Germano Vêras Neto (Orientador)
Universidade Estadual da Paraíba (UEPB)



Prof. Dr. Simone da Silva Simões (Examinadora interna)
Universidade Estadual da Paraíba (UEPB)



Prof. Dr. Frederico Luis Felipe Soares (Examinador externo)
Universidade Federal do Paraná (UFPR)

À minha avó Anália (*in memoriam*), por
toda contribuição na minha vida,
DEDICO.

AGRADECIMENTOS

Agradeço todos que fizeram parte da minha jornada de crescimento acadêmico e pessoal.

Principalmente a minha avó Anália (*in memoriam*) por todo incentivo e abrigo em todos os momentos que precisei.

À minha mãe Adriana e minha avó Gercina, que apesar das dificuldades conseguiram facilitar meus caminhos e me criarem em meio a tantas adversidades.

À minhas tias, Elissandra, Eliene e Elizangela, por fazerem parte dessa rede de apoio durante toda minha criação e estarem disponíveis todo momento para me ajudar.

À minhas irmãs, Laura e Lívia, pelo apoio e por me aguentarem todos os dias em meio a todas mudanças que passamos nesses últimos anos.

Aos meus amigos, Bruno, Carol e Manu, por estarem ao meu lado desde a graduação e mesmo em momentos de fragilidade segurarem minha mão.

Às amigadas que se iniciaram no laboratório, Mariana, Lavínia, João, Kilma, Giselle, Naara, Hilthon, Rômulo, Odilon, Danúbio... e aos demais que colaboraram e colaboram com o LQAQ e LABDEM.

Aos meus colegas de turma do mestrado, Gutierrez, Elane, Elizandro e Rénally, por enfrentarem o desafio das aulas no meio de uma pandemia comigo e por todos momentos de apoio mútuo.

Ao meu orientador Germano e coorientadora Elaine que aceitaram esse desafio comigo e por todo suporte durante a realização desse trabalho.

Aos professores do PPGQ – UEPB, Simone que me acompanha desde a iniciação científica e a partir das suas considerações e suporte vem me ajudado na minha construção acadêmica. Assim como, Conceição, Juracy, Mary e Ricardo pela destreza em suas aulas nos momentos de isolamento que passamos.

A todos pesquisadores e professores que impulsionaram a minha formação quimiométrica durante esse mestrado, David, Adriano, Paulo, Fabiola, Marco, além dos meus orientadores.

A David por toda paciência e agilidade na secretaria da coordenação do PPGQ e aos demais colaboradores que estiveram disponíveis para ajudar a tornar esse processo menos difícil.

A todos os colaboradores da UEPB e técnicos, juntamente a todos que fazem e fizeram parte da minha jornada acadêmica.

À FAPESP e a CAPES pelo financiamento que foi primordial para o andamento e conclusão desse trabalho.

RESUMO

O leite caprino é uma fonte rica em proteínas e carboidratos e com valorização internacional, principalmente em países europeus e asiáticos. Devido as características nutricionais e ao destaque econômico, esse produto pode sofrer adulterações em momentos após a coleta pela adição de leite de outra espécie que apresenta os mesmos componentes químicos majoritários, porém com características distintas em termos de quantidade e qualidade. Neste sentido, o principal adulterante de leite caprino é o leite bovino, por garantir dificuldades intrínsecas e extrínsecas de percepção da adulteração, mas com ganhos econômicos significativos. Vale destacar que qualquer adulteração de leite por adição de outra substância que altere propriedades organolépticas sem informações adequadas no rótulo, dentro do que permite a legislação nacional, é considerada como crime. Diversas pesquisas tem sido desenvolvidas para obtenção de métodos rápidos, baratos, mas com acurácia suficiente para atender legislação. Instrumentos portáteis são preferenciais nestes tipos de estudo pois são mais baratos que os de bancada e podem ser conduzidos para operações de fiscalização. A Espectrometria no infravermelho Próximo (NIRS) em associação com métodos quimiométricos podem auxiliar na elucidação de informações que promovam o reconhecimento necessários para distinguir amostras autênticas de adulteradas de forma expedita. Nesta perspectiva, este trabalho propõe a identificação de amostras de leite caprino com leite bovino usando espectrofotômetro portátil na região do NIR e métodos quimiométricos. A análise espectral quimiométrica associou de algoritmos meta-heurísticos e análise discriminante. O GA-LDA e o ACO-LDA mostraram uma eficácia de 100% de acurácia, exatidão e especificidade para distinção de amostras autênticas de leite caprino e amostras adulteradas mesmo com apenas 5% de adição de leite bovino. Neste sentido, o método desenvolvido permitiu identificar amostras de leite caprino adulteradas com leite bovino de forma rápida, não destrutiva e exata.

Palavras-chave: Autenticação de leite. Reconhecimento de padrões. Algoritmos Meta-heurísticos. Análise Discriminante Linear. NIR portátil.

ABSTRACT

Goat milk is a rich source of proteins and carbohydrates, with an international valorization, mainly in European and Asian countries. Because of its nutritional properties and economic value, this product may be adulterated after milking by the addition of different cattle's milk. These adulterants have similar chemical composition, but distinct quantity and qualitative aspects from goat milk. In this perspective, the main goat milk adulterant is cow milk, used to ensure a masked adulteration, and at the same time provide economic gains. According to Brazilian legislation, any adulteration of milk by adding substances that change its organoleptic properties without proper label indication is considered a crime. In a way to assist the inspection of food control official organizations, researches are developed to obtain fast, low cost, accurate and exact methods that attend legislation criteria. Portable instruments are suitable in these studies because of their low cost compared to a benchtop, allowing remote inspections. Near-infrared Spectroscopy (NIR), in association with chemometrics provides interpretation of information that allows distinguish authentic and adulterated samples. From this view, this work proposes the goat milk added with cow milk identification using portable NIR spectrometer along with chemometric methods. For that it was associated metaheuristic algorithms and discriminant analysis. GA-LDA and ACO-LDA showed efficient models with 100% of accuracy, sensitivity and specificity to discriminate goat milk authentic samples from adulterated, even with only 5% of cow milk added. Therefore, the developed models recognized goat milk samples adulterated by cow milk using fast, non-destructive and accurate methods.

Key-words: Milk adulteration. Pattern Recognition. Metaheuristic algorithms. Linear Discriminant Analysis. Portable NIR.

LISTA DE FIGURAS

Figura 1 – Espectros dos dados brutos de todas as amostras separados por classe.	32
Figura 2 – Espectros dos dados pré-processados com filtro de média móvel + (A) linha de base offset (BO), (B) correlação linear da linha de base (LCB), (C) SNV e (D) MSC.	33
Figura 3 – Escores da PCA dos dados brutos com visualização pela variável categórica.	35
Figura 4 – Gráfico de pesos das PC1 e PC4 dos dados brutos.	36
Figura 5 – Análise de possível amostra anômala 196 pelos gráficos de (A) escores dos dados brutos com e (B) influência em relação ao resíduo na PC4, e para amostra 195 em (C) escores e (D) influência.	37
Figura 6 – Análise da distribuição amostral na PCA sem possíveis amostras anômalas (195 e 196) pelos gráficos de (A) escores e (B) influência.	38
Figura 7 – Escores e pesos da PCA dos dados pré-processados com filtro de média móvel com 7 pontos associado a (A) linha de base offset, (B) correção linear de linha de base, (C) SNV e (D) MSC.	39
Figura 8 – Espectro médio com destaque as variáveis selecionadas pelo GA-LDA e ACO-LDA após pré-processamentos dos dados por MM+BO.	42
Figura 9 – Discriminação linear das amostras do conjunto de teste por GA-LDA (MM+BO).	44

LISTA DE TABELAS

Tabela 1 – Composição média dos leites mais consumidos por humanos (a cada 100g analisados).....	15
Tabela 2 – Lotes caprinos e bovinos que foram coletadas as amostras de leite.....	29
Tabela 3 – Modelos criados por GA-LDA e ACO-LDA.....	41
Tabela 4 – Matrizes de confusão dos modelos criados.....	46
Tabela 5 – Comparação dos modelos de referência (Pereira et al., 2020) com os modelos obtidos.....	48

LISTA DE ABREVIATURAS E SIGLAS

ACO	Otimização por Colônia de Formigas, do inglês <i>Ant Colony Optimization</i>
BO	Linha de base <i>offset</i> , do inglês <i>Baseline offset</i>
FD	Função discriminante, do inglês <i>Discriminant Function</i>
FT-NIR	Infravermelho próximo por transformada de Fourier, do inglês <i>Fourier-transform near-infrared</i>
GA	Algoritmo Genético, do inglês <i>Genetic Algorithm</i>
IDF	Federação Internacional de Laticínios, do inglês <i>International Dairy Federation</i>
ISO	Organização Internacional para Padronização, do inglês <i>International Organization for Standardization</i>
iSPA-PLS-DA	Seleção de intervalos do Algoritmo de Projeções Sucessivas em Análise Discriminante por Mínimos Quadrados Parciais, do inglês <i>Interval selection of Successive Projection Algorithm by Partial Least Squares Discriminant Analysis</i>
LCB	Correção linear de linha de base, do inglês <i>Linear Baseline Correction</i>
LDA	Análise Discriminante Linear, do inglês <i>Linear Discriminant Analysis</i>
LQAQ	Laboratório de Química Analítica e Quimiometria
MAPA	Ministério da Agricultura Pecuária e Abastecimento
MM	Média móvel, do inglês <i>Moving average</i>
MSC	Correção multiplicativa de espalhamento, do inglês <i>Multiplicative Scatter Correction</i>
NIR	Infravermelho próximo, do inglês <i>near infrared</i>
OC-PLS	Mínimos Quadrados Parciais por Uma Classe, do inglês <i>One Class Partial Least Squares</i>
PCA	Análise por Componentes Principais, do inglês <i>Principal Component Analysis</i>
PLS-DA	Análise Discriminante por Mínimos Quadrados Parciais, do inglês <i>Partial Least Squares Discriminant Analysis</i>
SG	Savitzky-Golay
SNV	Variância Normal Padrão, do inglês <i>Standard Normal Variate</i>

UEPB

Universidade Estadual da Paraíba

UHT

Temperatura ultra-alta, do inglês *ultra high temperature*

SUMÁRIO

1	INTRODUÇÃO	12
2	FUNDAMENTAÇÃO TEÓRICA	14
2.1	Contexto cultural e nutricional do leite caprino	14
2.2	Contexto econômico do leite caprino	16
2.3	Adulterações no leite caprino	17
2.4	Espectroscopia na região do infravermelho próximo	19
2.5	Quimiometria	21
	<i>2.5.1 Análise por componentes principais</i>	<i>21</i>
	<i>2.5.2 Análise discriminante linear</i>	<i>22</i>
	<i>2.5.3 Seleção de variáveis</i>	<i>24</i>
3	METODOLOGIA	28
3.1	Aquisição das amostras	28
3.2	Análise espectral	29
3.3	Análise dos dados	30
4	RESULTADOS	32
4.1	Medidas espectrais e pré-processamentos	32
4.2	Análise dos dados brutos e pré-processados por PCA	34
4.3	Modelos obtidos por GA-LDA e ACO-LDA	40
5	CONCLUSÕES	50
	REFERÊNCIAS	51

1 INTRODUÇÃO

O consumo de leite caprino vem aumentando nos últimos anos devido a suas propriedades nutricionais. Além do consumo como alternativa a ingestão de leite bovino para indivíduos intolerantes e/ou alérgicos a lactose. Essas propriedades nutricionais do leite caprino estão associadas à sua composição química *sui generis* (MUEHLHOFF; BENNET; MCMAHON, 2013; CLARK, 2017; MILLER; LU, 2019; SIQUEIRA, 2019).

A valorização da cultura leiteira caprina baseia-se principalmente no mercado asiático e europeu, com foco na produção de derivados como queijo e iogurte. Porém desde o final do século XX que o Brasil possui políticas de incentivos para produção de leite caprino por produtores rurais de pequeno e médio porte em vários estados nordestinos, por exemplo. Destaca-se neste sentido o estado da Paraíba devido à alta adaptabilidade da espécie aos climas da região (CLARK, 2017; JUNIOR; SIQUEIRA; STOCK; 2019).

Apesar do crescimento econômico da cadeia produtiva da pecuária caprina, como o leite possui destacadas propriedades nutricionais e o preço ao consumidor ainda é mais alto comparado ao leite bovino, é este produto um iminente alvo de fraude. A principal fraude que ocorre é a adição de quantidades consideráveis de leites com custo mais baixo para aumentar o volume final, acarretando prejuízos financeiros e nutricionais aos consumidores. Estas adições de produtos estranhos ao leite caprino são normalmente identificadas no momento da recepção do produto nas indústrias/cooperativas e nas fiscalizações que porventura aconteçam. Por outro lado, a adição de componentes já presentes na composição do leite promove um desafio para as análises usuais pela variabilidade associada a um material perecível.

Geralmente a adição de água, por ser o componente majoritário dos leites é uma das mais comuns adulterações do leite. Outro potencial adulteração é pela inserção de leites de outra espécie com menor valor agregado para aumentar volume, mantendo seus componentes principais. Essas práticas além de serem crime, podem causar malefícios a saúde do consumidor e danos econômicos na produção de derivados (MAPA, 2000).

Metodologias analíticas convencionais por via úmida ou utilizando citometria para determinação da composição das amostras de leite são empregadas para indicar possíveis adulterantes. Porém, estes procedimentos não apresentam rapidez necessária para detectar adulterações logo após a coleta, ou na aquisição do produtor rural. Além disso, são técnicas invasivas e destrutivas para as amostras e geram resíduos que devem ser tratados. Assim, uso de técnicas analíticas instrumentais com equipamentos portáteis tem sido desenvolvidas para viabilizar a investigação remota das amostras de leite para proporcionar resultado imediato sem causar danos ao analito (MAPA, 2018).

Neste sentido, a espectroscopia na região do infravermelho próximo (NIR) proporciona os benefícios associados à análise vibracional molecular de forma rápida, sem produzir resíduos e sem degradar a amostra de leite analisada. Além de que as bandas vibracionais obtidas na região NIR são de ligações químicas geralmente presentes em alimentos, possibilitando informações para o estudo da interferência na adição do leite bovino no leite caprino (PAQUINI, 2018; BEC; HUCK, 2019; ZAREEF et al, 2020).

Associada a espectroscopia NIR, ferramentas quimiométricas são extremamente úteis na investigação de presença de contaminantes em amostras de alimentos, e em de qualquer outra matriz amostral em que possa ser gerada uma grande quantidade de dados. Entretanto, parte dos dados obtidos podem proporcionar informações irrelevantes para identificar presença de contaminantes ou redundantes com outras variáveis informativas. Assim, o uso de algoritmos de seleção de variáveis meta-heurísticos (inspirados em comportamentos biológicos) promove conhecimento das variáveis que possuem informações relevantes para distinção de amostras puras e adulteradas. Nesse sentido, o objetivo do trabalho foi o desenvolvimento de modelos quimiométricos a partir de dados obtidos em espectrofotômetro portátil na região do infravermelho próximo para identificar adulterações no leite caprino por adição de leite bovino.

2 FUNDAMENTAÇÃO TEÓRICA

2.1 Contexto cultural e nutricional do leite caprino

O consumo de leite caprino se iniciou de forma doméstica para auxiliar na alimentação e nutrição de famílias em um contexto rural brasileiro. Tendo em vista que o uso do leite bovino possuía todo seu foco na comercialização e para produção de seus derivados. A partir do início do século XX, com início de estudos relacionados às contribuições nutricionais do leite de cabra e seus benefícios a saúde, proporcionando visibilidade ao potencial das produções leiteiras caprinas, pela investigação de aspectos sensoriais e químicos (CLARK, 2017).

Os estudos evidenciaram características específicas ao leite de cabra, além das sensoriais como sabor e odor característicos. Atributos foram evidenciados principalmente relacionados a melhor digestibilidade, maior alcalinidade, potencial hipoalergênico e terapêutico, quando comparado ao leite bovino. Essas propriedades foram fatores determinantes para o aumento no consumo do leite caprino promovendo sua maior valorização nutricional e, conseqüentemente, econômica (CLARK, 2017; MILLER; LU, 2019).

Os benefícios nutricionais são destacados diante de algumas divergências em relação as composições médias de leites de outras origens animais. Na **Tabela 1** pode-se perceber que os componentes majoritários dos leites de vaca, cabra, búfala e ovelha são os mesmos, porém, possuem proporções diferentes, com destaque para as porcentagens de gorduras totais (carboidratos) e proteínas (MUEHLHOFF; BENNET; MCMAHON, 2013; SIQUEIRA, 2019).

Tabela 1 – Composição média dos leites mais consumidos por humanos (a cada 100g analisados).

Constituinte	Vaca	Cabra	Búfala	Ovelha
Água	87,8	87,7	83,2	82,1
Gorduras totais	3,3	5,8	7,5	6,4
Proteínas totais	3,3	3,4	4,0	5,6
Lactose	4,7	4,4	4,4	5,1
Cinzas	0,7	0,8	0,8	0,9

Fonte: MUEHLHOFF, BENNET and MCMAHON (2013).

A porcentagem de gorduras totais é o fator que mais se evidencia pela diferença do teor presente entre leite caprino e no leite bovino. É, portanto, o índice de gorduras totais o componente primordial responsável pelo crescente apreço do leite caprino e na produção de seus derivados, especialmente como queijos. Apesar de possuir valores inferiores aos de leite bubalino e ovino, o leite caprino é mais acessível no Brasil comparado a esses. Entretanto, o consumo de leite bovino segue dominante no mercado devido a características sensoriais particulares ao caprino, podendo causar estranhamento ao primeiro contato (SIQUEIRA, 2019).

Outro fator relevante é a porcentagem de lactose, que tanto no leite caprino quanto no bubalino apresenta-se em valores inferiores comparado aos demais. Esse componente está diretamente relacionando a maior digestibilidade, com a presença de proteína β -caseína no leite caprino que faz com que as moléculas de lactose sejam quebradas com mais facilidade (PARK, 2017).

A diferenciação entre os leites de cada espécie pode ser dada pelos fatores citados anteriormente, relacionado aos componentes majoritários além da água. Carboidratos e proteínas podem indicar a espécie de origem do leite, tendo em vista não só as diferenças percentuais, mas como as particularidades intrínsecas a cada espécie (PARK, 2017).

A variabilidade da proporção dos componentes do leite pode ser relacionadas as condições na criação dos animais. Sendo em relação a alimentação, condições ambientais, condições de manejo e pasto, a estação do

ano, a localidade, além de características intrínsecas ao animal em específico como raça, saúde e estágio de lactação. Alguns fatores adicionais como condições de ordenha do leite, higiene do animal e ainda o condicionamento após a coleta podem afetar sua qualidade do leite (LAD et. al., 2017; JUNIOR; SIQUEIRA; STOCK; 2020).

2.2 Contexto econômico do leite caprino

As características nutricionais e químicas do leite caprino e de seus derivados reafirma a valorização econômica que vem conquistando nos últimos anos. Segundo dados da Organização das Nações Unidas para a Alimentação e a Agricultura (FAO, do inglês *Food and Agriculture of the United Nations*), entre os anos de 2009 e 2019, o maior produtor e consumidor de leite de cabra foi o continente asiático, participando com aproximadamente 58,4% da produção mundial. Seguido pelos continentes africano (23%), europeu (14,5%) e americano (4,1%). A divisibilidade dos climas nesses continentes endossa a alta adaptabilidade da pecuária caprina em diversas condições climáticas (FAOSTAT, 2021).

A valorização do leite caprino estende-se com destaque ao mercado europeu, países como França, Espanha, Grécia e Holanda possuem apreço cultural pelo leite em si e seus derivados. O sabor característico e os benefícios a saúde são atrativos adicionais ao interesse por seus derivados, com a produção de queijos e iogurtes para consumo local e para exportação, principalmente para o continente americano (MILLER; LU, 2019).

No continente americano, o Brasil ocupa lugar de destaque com produção média crescente desde o início da segunda década dos anos 2000. Em 2019 a média foi de 280 mil toneladas de leite caprino produzidos. Apesar de uma queda brusca no ano de 2014, de cerca de 40 mil toneladas, o país conseguiu recuperar seu potencial produtivo nos anos seguintes. Grande parte da produção vem de polos concentrados na região nordeste e sudeste do Brasil (MILLER; LU, 2019; FAOSTAT, 2021).

Os estados com maiores índices de produção do Brasil são Bahia, Paraíba, Minas Gerais, Pernambuco e Rio Grande do Norte, entre os anos de

2006 e 2017. A fronteira entre Bahia e Minas Gerais, assim como entre Pernambuco, Rio Grande do Norte e Paraíba, facilita a troca cultural em relação a cultura caprina, assim como a as condições de manejo e pasto serem parecidas entre esses estados. Principalmente no Nordeste, onde a pecuária caprina é fonte de subsistência para muitas famílias da região rural, além do contexto cultural forte associado ao consumo de produtos caprinos (JUNIOR; SIQUEIRA; STOCK; 2019).

Durante as décadas de 1980 e 1990, investimentos por meio de programas governamentais foram essenciais para incentivar essas famílias que sobreviviam da pecuária caprina na região nordeste. Essas contribuições viabilizaram que os pequenos produtores fossem modernizando a maneira de manejo e ordenha, e assim associando tecnologias para otimizar a produção de leite caprino. Paraíba, Pernambuco e Rio Grande do Norte puderam se beneficiar, desenvolvendo sua produção leiteira caprina ao ponto de chegar a exportação de derivados como queijo (JUNIOR; SIQUEIRA; STOCK; 2019).

As decorrências desses investimentos estão sendo observadas atualmente com a procura crescente do leite caprino e seus derivados para consumo e exportação, relacionado tanto aos benefícios nutricionais quanto ao sabor. O valor agregado acompanha esse crescimento e com isso está sujeito a fraudes e adulterações com intuito de aumentar o volume de produção e ampliar o rendimento econômico de forma ilícita.

2.3 Adulterações no leite caprino

No contexto de leite caprino, as adulterações são provocadas pela: adição de água; alterações em qualquer componente natural ou estranho a composição normal do leite; adição ou presença de substâncias conservadores ou inibidoras do crescimento de microrganismos; caso esteja cru e seja comercializado como pasteurizado, esterilizado ou UHT; adição de leite de outras espécies animais. Sendo motivadas pelo aumento no rendimento do leite, visando maior lucro (MAPA, 2000).

A adição de leite de outras espécies de valor agregado menor é viabilizada pela composição básica dos leites serem similares. Porém além de ser crime, a

adulteração de leites bovinos que apresentam valor econômico inferior ao leite caprino interfere nas proporções de cada componente do leite, divergindo tanto do leite adulterante quanto do leite cru original. Tal fato, pode causar ainda riscos à saúde de consumidores intolerantes ou alérgico a lactose que buscam o leite caprino como alternativa ao leite bovino.

As adulterações provocadas pela adição de leite bovino no leite caprino podem ser identificadas por divergências nos parâmetros estabelecidos pelo Ministério da Agricultura e Pecuária e Abastecimento (MAPA). Teores de gordura e proteínas totais, lactose anidra e sólidos não gordurosos e totais são analisados de acordo com a ISO 9622 / IDF 141 (2013) por medidas de absorção da radiação no infravermelho médio, além de uso de sensores de condutividade. Entretanto, estas medidas não são tão precisas, demandam a degradação da amostra de leite durante a análise (MAPA, 2018).

Outro método que pode indicar adulteração indicado pelo MAPA (2018) é o uso da contagem de células somáticas por mililitro de leite e a contagem de unidades formadoras de colônias por mililitro de leite feitas pelo método citométrico em fluxo, seguindo as ISO 13366-2/ IDF 148-2 e ISO 4833-1, respectivamente. Essa análise promove uma investigação sobre as condições microbiológicas das amostras de leite. Devido a sua alta perecibilidade, tende a crescer a quantidade de microrganismos presentes, estando também associado ao seu condicionamento.

Tentativas de desenvolvimento de metodologias capazes de detectar e quantificar adulterações tem sido frequentes. A perspectiva desses métodos propostos é minimizar erros durante a análise de matrizes tão semelhantes como os leites de diferentes espécies animais. A distinção de leite caprino em diversas porcentagens de adição do leite bovino, ambos em pó, foram aplicadas por espectroscopia de tempo de vida de fluorescência por meio da emissão e excitação (BRANDÃO et al., 2017).

Outras perspectivas recentes utilizando espectroscopia vibracional na região o infravermelho próximo (NIR) associada a ferramentas quimiométricas para autenticação de leite caprino podem ser apresentadas. Teixeira et al. (2020)

usaram espectroscopia NIR com transformada de Fourier (FT-NIR) para detectar diferentes proporções de adulterantes, como água, ureia, proteína “whey” bovina e leite bovino. Pereira et al. (2020) utilizaram espectroscopia no infravermelho próximo (NIR) para identificar adulterações de leite caprino em leite bovino.

A associação da espectroscopia, em destaque a NIR, na análise de adulterações de alimentos, principalmente na autenticação de leite caprino promove o uso de informações químicas para distinguir amostras autênticas de adulteradas. As técnicas quimiométricas auxiliam essa investigação com a construção de modelos que aplicam os conhecimentos para criar meios de identificar adulterações baseadas na informação analítica.

2.4 Espectroscopia na região do infravermelho próximo

A espectroscopia NIR fornece informações a partir do estado vibracional das moléculas após aplicação de um feixe de radiação eletromagnética na amostra entre os comprimentos de onda de 750 nm e 2500 nm. Por fim, resulta em um espectro que está associado às intensidades de sinal analítico demonstrado pelas ligações moleculares presentes na amostra (PASQUINI, 2018).

As interações entre a radiação eletromagnéticas e as moléculas ocorrem envolvendo sobreposições de excitação, combinações e ressonância dos modos vibracionais fundamentais. Sendo esses modos associados a vibrações altas anarmônicas de grupos funcionais moleculares contendo um átomo de massa molecular relativamente alta ligado ao átomo de hidrogênio X-H (onde X representa C, O, N e S). Além disso, vibrações de ligações químicas fortes entre esses átomos “pesados”, como C=O, possuem sinal analítico relevante na região do infravermelho próximo (PAQUINI, 2018; BEC; HUCK, 2019; ZAREEF et al, 2020).

As informações químicas associadas aos espectros NIR obtidos após análise podem facilitar a interpretação de diversas substâncias presentes em alimentos, polímeros, combustíveis, medicamentos, solos, entre outras. Além de ser uma técnica que possibilita a obtenção rápida de resultados, apresenta baixo custo operacional, de fácil manuseio, não é destrutiva ao analito, minimizando a

produção de resíduos. Outra vantagem apresentada pela técnica é a análise por meio de equipamentos portáteis, os quais possibilitam a coleta dos espectros de forma remota (PAQUINI, 2018; BEC; HUCK, 2019).

A análise dos dados espectrais NIR conduzida com auxílio de ferramentas quimiométricas favorece a investigação pelas informações químicas presente nos dados. Métodos estatísticos e matemáticos são aplicados para obtenção de resultados satisfatórios, considerando parâmetros para avaliação dos modelos, tendo em vista a elucidação de informações químicas presentes nos espectros, em forma de bandas alargadas e sobrepostas. Essas técnicas ajudam interpretação na das variáveis espectrais, assim como na criação de modelos que conseguem entender a distribuição das amostras buscando por tendências nos comportamentos. Esses modelos são desenvolvidos para quantificar e/ou classificar amostras a partir da informação química extraída dos espectros (PAQUINI, 2018; KAUFMANN, 2021).

A análise espectral NIR, com auxílio de técnicas quimiométricas, vem apresentando destaque em estudos que conseguem prever amostras adulteradas de leite, por meio da construção de modelos de classificação. Mesmo que a diferença seja mais evidenciada entre leite e adulterantes naturais vegetais, como provindos de proteínas de soja e ervilha. Apesar disso, o uso de leites de outras espécies como adulterantes podem ser indicados por alterações nas concentrações, apesar da similaridade em seus constituintes (WU et al., 2016; CHEN et al. 2017; MABOOD et al., 2017a; MABOOD et al., 2017b; KARUNATHILAKA et al., 2018; PEREIRA, 2020).

Para investigação de adulteração por leites de espécies animais distintas, como leite bovino adicionado no caprino, ferramentas quimiométricas podem ser essenciais para interpretação dos dados espectrais. Entretanto, algumas variáveis espectrais não são esclarecedoras desta contaminação e, portanto, precisam ser descartadas na construção de modelos. Neste sentido, métodos de seleção de variáveis promovem a extração da informação química relevante e redução do ruído para a construção de modelos mais ajustados em relação a modelos globais, que utilizam todas as variáveis (WU et al., 2016; CHEN et al. 2017).

2.5 Quimiometria

A quimiometria atribui à análise de dados uma visão estatística e matemática sobre conhecimentos químicos referentes às respostas analíticas. Técnicas instrumentais multivariadas, como as espectroscopias por exemplo, permitem obter uma grande quantidade de dados por amostra e por isso são as que mais beneficiam o analista quando utilizam a quimiometria na interpretação dos resultados, facilitando o estudo sobre as espécies químicas presentes (BEEBE et al, 1998; FERREIRA, 2016).

O estudo quimiométrico dos dados espectrais podem seguir diversas vertentes dependendo do objetivo de estudo. As técnicas de reconhecimento de padrões (classificação) destacam-se no caso da identificação de amostras adulteradas e diferenciação de amostras autênticas. Essas técnicas potencializam a análise de divergências no comportamento espectral das amostras e a tendência de aproximação e separação entre elas no espaço amostral (BEEBE et al, 1998; FERREIRA, 2016).

2.5.1 Análise por componentes principais

Um dos métodos mais utilizados em reconhecimento de padrões é a Análise por Componentes Principais (PCA, do inglês *Principal Component Analysis*). Esta técnica destaca-se por permitir ao analista realizar uma investigação exploratória minuciosa, com intuito de encontrar uma relação entre as variáveis que facilite a visualização da informação química presente em cada amostras. A PCA é considerada um método de análise exploratória que proporciona estudo dos dados apenas pelas informações intrínsecas as variáveis explicar a dispersão das amostras em um espaço dimensional menor do que o espaço original (BEEBE et al, 1998; FERREIRA, 2016).

Os dados espectrais organizam-se de modo que as amostras (n) e as variáveis (m) formam uma matriz de dados $\mathbf{X}_{(n \times m)}$. Essas variáveis são representadas pelos comprimentos de onda do espectro de cada amostra relacionadas a intensidade de sinal espectral. A atuação da PCA é na decomposição dos dados para que ocorra a redução do espaço dimensional de modo que, a matriz (\mathbf{X}) de dados seja decomposta em uma matriz de escores

(**T**) e em uma matriz de pesos (**P**), com presença de resíduos (**E**), **Equação 1** (BEEBE et al, 1998; FERREIRA, 2016).

$$\mathbf{X} = \mathbf{TL}^T + \mathbf{E} \quad (1)$$

Os escores (**T**) representam um novo conjunto de coordenadas associando direções com informações fortemente relevantes, permitindo a projeção de novos dados, componentes principais, em poucas dimensões. Assim, oferece uma melhor visualização dos dados, o que possibilita a indicação de tendências nos dados como a presença de agrupamentos (*clusters*) (BEEBE et al, 1998; FERREIRA, 2016).

Apesar de PCA ser bastante utilizada para identificar a formação de agrupamentos que facilitam o conhecimento sobre a semelhança e divergência no comportamento das amostras de acordo com as variáveis, apenas técnicas de reconhecimento de padrões supervisionadas conseguem distinguir os grupos pré-estabelecidos em que futuras amostras podem ser classificadas. Dentre as técnicas de reconhecimento de padrões supervisionadas destaca-se a Análise Discriminante Linear (LDA, do inglês *Linear Discriminant Analysis*).

2.5.2 Análise discriminante linear

A LDA estabelece uma superfície de decisão no espaço amostral na qual as amostras são classificadas em regiões específicas do plano das respostas. Esse limiar é regido pela **Equação 2**, quando $k = 0$, sendo $k > 0$ amostras pertencentes a uma classe e $k < 0$ referentes a outra classe (FERREIRA, 2016).

$$k(\mathbf{x}) = [a_0 \ a_1 \ a_2 \ \dots \ a_j] \begin{bmatrix} 1 \\ x_1 \\ x_2 \\ \dots \\ x_j \end{bmatrix} \quad (2)$$

A fronteira que é criada entre as amostras de acordo com as variáveis estabelece o limite entre as duas classes, em que a amostra só pode ser pertencente a apenas uma classe. De acordo com a distribuição dos dados das amostras, a LDA gera o discriminante, que funciona como classificador, baseado na matriz de dados **X** e nos conhecimentos prévios da classe de cada amostra para construção do modelo (FERREIRA, 2016; GOMES et al, 2022).

Por ser um modelo de classificação supervisionada, na LDA cada classe é dividida em um conjunto de treinamento e um conjunto de teste. Esses conjuntos contêm amostras que serão usadas para o desenvolvimento do modelo (treinamento) e as amostras que serão usadas para validação da classificação (teste) realizada pelo modelo. A separação pode ser feita por algoritmos como o Kennard-Stone (GOMES et al, 2022).

Com uso da LDA, modelos robustos criados para prever adulterações de alimentos, geralmente associado a seleção de variáveis. Essa robustez está associada aos parâmetros: Acurácia, Sensibilidade e Especificidade. Esses parâmetros são calculados pelo estudo do número de falsos positivos (FP), verdadeiros positivos (VP), falsos negativos (FN), verdadeiros negativos (VN) (LOPES et al. 2015; FERREIRA, 2016). Assim mostradas as **Equações 3, 4 e 5** a seguir:

$$\text{Acurácia} = \frac{VP+VN}{VP+FN+FP+VN} \times 100 \quad (3)$$

$$\text{Sensibilidade} = \frac{VP}{VP+FN} \times 100 \quad (4)$$

$$\text{Especificidade} = \frac{VN}{FP+VN} \times 100 \quad (5)$$

A acurácia é correspondente a proporção de classificações adequadas no conjunto de dados, representada pela **Equação 3**. A sensibilidade baseia-se na capacidade do modelo em estimar corretamente a classe das amostras em razão ao número de amostras total pertencente a classe modelada, como mostrada na **Equação 4**. Por outro lado, a especificidade determina a capacidade do modelo estimar corretamente as amostras que não pertencem a classe modelada, demonstrada na **Equação 5**.

Modelos que apresentam as métricas das figuras de mérito indicadas acima próximas a 100% apresentam capacidade preditivas consideradas como satisfatórias para diferenciar/classificar as amostras. No caso das adulterações, os modelos robustos conseguem distinguir a maior parte possível das amostras autênticas das amostras adulteradas.

Para atingir sucesso na classificação em modelos discriminatórios geralmente são utilizados algoritmos de seleção de variáveis devido a eliminação de variáveis pouco informativas e consequente aumento da importância das variáveis discriminantes. Portanto, as técnicas de seleção de variáveis permitem aumentar o poder discriminatório das amostras em comparação aos métodos que utilizam todas as variáveis (GOMES et al. 2022).

2.5.3 Seleção de variáveis

No sentido de tratamento de dados de alimentos, a seleção de variáveis é uma ferramenta poderosa na extração de informações presentes em grandes conjuntos de dados. Muitas variáveis são frequentes em análises espectrométricas, a seleção promove a identificação de informações úteis e significativas que representem o conjunto de dados. Por outro lado, minimiza as variáveis que não apresentam informações relevantes ou que possam ser consideradas como ruído durante a construção de modelos quimiométricos (XIAOBO et al., 2010; ALLEGRINI; OLIVIERI, 2011; NIAZI; LEARDI, 2012; GOMES et al, 2022).

A seleção de variáveis pode ser realizada de acordo com algoritmos para determinar as variáveis que possuem impacto maior na construção de modelos que consigam prever o comportamento das amostras. Algoritmos meta-heurísticos baseados em conceitos biológicos são utilizados para implementar comportamentos da natureza a variabilidade dos dados químicos, entre eles destacam-se o algoritmo genético (GA, do inglês *Genetic Algorithm*) e o algoritmo de colônia de formigas (ACO, do inglês *Ant Colony Optimization*) (ALLEGRINI; OLIVIERI, 2011; KATOCH; CHAUHAN; KUMAR, 2021).

O algoritmo de GA é baseado nos princípios da seleção natural e em conceitos genéticos para seleção de variáveis do conjunto de dados. Essa seleção acontece de forma binária, atribuindo dígito "1" as variáveis selecionadas e dígito "0" as variáveis não selecionadas. Assim são construídos vetores, associados aos cromossomos, com dígitos binários, interpretados como genes, para minimizarem geralmente o erro de predição médio de um conjunto de amostras. As métricas associadas a formação de um novo indivíduo é regida

pela **Equação 6** (ALLEGRINI; OLIVIERI, 2011; KATOCH; CHAUHAN; KUMAR, 2020).

$$R = \frac{(G+2\sqrt{g})}{3G} \quad (6)$$

O termo g é o numero gerações, G é o numero total de gerações evolucionarias por população. O R muda dinamicamente e aumenta com o crescimento do número de gerações evolucionárias. No estado inicial, a similaridade entre os indivíduos é baixa, então o valor de R também deve ser baixo para garantir que as novas populações não destruirão as condições genéticas ótimas dos indivíduos (KATOCH; CHAUHAN; KUMAR, 2020).

O conceito de que os indivíduos mais adaptáveis sobrevivem, é seguido pelo algoritmo genético à medida que os vetores vão sendo construídos. Mutações e combinações vão acontecendo nos cromossomos (vetores) até o alcance de um número de gerações preestabelecido. Ao fim, o cromossomo que melhor codificar e resolver o problema proposto, é incluído no modelo multivariado de análise (ALLEGRINI; OLIVIERI, 2011).

O algoritmo de ACO simula o comportamento de colônias de formigas na busca pelo melhor percurso até a fonte de alimento. As variáveis são identificadas por coordenadas dimensionais, de forma similar ao GA, para definir os caminhos possíveis para as formigas seguirem. Cada caminho é conectado a um número de variáveis selecionadas, as quais correspondem a um erro de predição do modelo (ALLEGRINI; OLIVIERI, 2011).

No ACO, a cada geração, as formigas depositam uma certa quantidade de feromônio para potencializar a redução dos valores de erro em cada caminho. Enquanto isso, as formigas encontram novos caminhos baseados no feromônio acumulado em cada coordenada, pela medida heurística dos caminhos e pela busca aleatória por todos percursos disponíveis. Essa busca é promovida pela combinação probabilística desses fatores, permitindo desvios dos caminhos mais prováveis A busca realizada pelas formigas é baseada de forma estocástica pela exalação de feromônio regida pelas Equações 7 e 8 (ALLEGRINI; OLIVIERI, 2011).

$$\tau_i^o(t+1) = (1 - \rho_2)\tau_i^o(t) + \rho_2\Delta\tau_i^o, o = 1, \dots, O$$

$$\Delta\tau_i^o = \sum_{k \in S_{ib}(t)} |1 - q_k(i)| \frac{Q}{f_o(q_k)}$$

O fator de evaporação do feromônio é representada por ρ_2 , em que o valor varia entre 0 e 1 ($0 < \rho_2 < 1$). $S_{ib}(t)$ é o termo referente a melhor solução de interação em um tempo t . A quantidade de feromônio liberada é obtida pelo termo $Q/f_o(q_k)$, quanto o q_k refere-se a melhor k -ésima interação, caso $q_k(i) = 0$ nenhum feromônio será exalado. O feromônio exalado pode maximizar o caminho até a informação relevante (He; Ma, 2014).

Ambos os algoritmos de seleção de variáveis podem ser utilizados na construção de modelos de classificação de amostras baseados nas variáveis originais do conjunto de dados. No contexto de investigar a presença ou ausência de adulteração no leite, a LDA consegue estabelecer um limiar entre as amostras autênticas e adulteradas. Sendo assim, considerado toda amostra com comportamento distinto ao autêntico, ser considerado como adulterada.

2.6 Aplicação da LDA e seleção de variáveis a dados espectrais para identificar adulterações

Dados espectrais são objetos de estudos de diversas pesquisas relacionadas a qualidade a autenticação de alimentos de origem animal e vegetal utilizando LDA e técnicas de seleção de variáveis. Devido ao alto número de variáveis dependentes entre si presentes nos dados espectrais, a escolha de regiões de interesse específicas pode eliminar essa correlação. Além disso, é possível reduzir a quantidade de variáveis para obter a informações de interesse sobre os dados de forma mais objetiva que os demais métodos de classificação (KAUFMANN, 2021).

A LDA promove a descrição se cada amostra pertence ou não a uma classe particular. No sentido de analisar adulterações esse fator a torna uma ferramenta poderosa, pois o simples reconhecimento de uma amostra encontra-se adulterada ou não, já deixa explícita a autenticidade do alimento diante das informações usadas para construir o modelo. Alimentos perecíveis como leite,

necessitam de análises rápidas, objetivas e confiáveis para detectar sua qualidade e autenticidade, sendo a espectroscopia NIR e modelos de classificação por análise discriminante linear associada a seleção de variáveis.

A partir de aplicações já citadas anteriormente, podemos perceber uma alta vertente de construções de modelos para prever adulterações utilizando métodos quimiométricos derivados da regressão por mínimos quadrados. De modo que, Pereira et al. 2020 conseguiu descrever a classificação de amostras de adulterações de leite caprino por leite bovino entre métodos como PLS-DA, OC-PLS e iSPA-PLS-DA. Sendo assim o estudo dessa seleção de variáveis por GA-LDA e ACO-LDA uma vertente ainda não explorada para minimizar as variáveis associadas a classificação dessas amostras.

A seleção de variáveis associada ao LDA tem sido fonte de diversos estudos na classificação de amostras alimentares. Carvalho et al. (2019) utilizaram o GA-LDA e os dados mostraram 100% de acurácia na distinção de amostras defeituosas de macadâmia utilizando espectroscopia NIR e ressonância magnética nuclear (NMR). Outra perspectiva é a classificação de genótipos e predição da composição de óleos de soja realizada por Leite et al. (2020) utilizando GA-LDA que rendeu resultados satisfatórios, com predição de aproximadamente 89% das amostras.

A seleção de variáveis é uma abordagem importante tanto para modelos quantitativos quanto para modelos de classificação, que não tem sido presente em programas computacionais comerciais. A maioria das ferramentas para seleção de variáveis são originadas de *toolboxes* e *softwares* feitos por pesquisadores e disponibilizados para serem utilizados por usuários com noções de programação, tornando o acesso a ferramenta limitado (GOMES et al, 2022).

3 METODOLOGIA

As etapas de aquisição de amostras e análise espectral foram realizadas por pesquisadores parceiros no Núcleo de Pesquisa e extensão em Alimentos (NUPEA), situado na UEPB, campus I, em Campina Grande. O equipamento portátil foi cedido pelo Laboratório de Automação e Instrumentação em Química Analítica e Quimiometria (LAQA), situado na UFPB, campus I, em João Pessoa-PB.

Todos os dados foram cedidos ao Laboratório de Química Analítica e Quimiometria (LQAQ) na perspectiva de serem desenvolvidos novos estudos, além dos já desenvolvidos pelos grupos de pesquisa do LAQA. Assim foram informadas as condições de aquisição e análise espectral das amostras de leite caprino e adulteradas por leite bovino.

3.1 Aquisição das amostras

As amostras de leite caprino foram obtidas por parceria com pequenos produtores de regiões de Campina Grande e do Cariri Paraibano, descritas na **Tabela 2**. As coletas foram realizadas em diversos produtores em outubro de 2019. Da mesma forma foi realizada para as amostras de leite bovino obtida por pequenos produtores das regiões de Campina Grande, Taperoá e Sobrado na Paraíba.

Os rebanhos eram compostos por animais de diferentes idades e períodos de lactação. Não houve regulação da alimentação oferecida aos animais, garantindo maior variabilidade química real ao leite produzido. Os sistemas de manejo variavam entre semi-extensivo, semi-intensivo e intensivo.

Os lotes caprinos foram 5 obtidos no município de Taperoá-PB, enquanto apenas 1 foi obtido no município de Campina Grande-PB. Os rebanhos possuíam cabras de raças puras: Alpina Americana, Anglo Nubiana, Moxotó, Murciana e Parda. Dos lotes de leite bovino, 3 foram provenientes do mesmo produtor no município de Taperoá-PB, porém seus lotes de raças diferentes, Sindi e Guzerá. O último lote bovino (Bovino 4) foi obtido em Sobrado-PB da raça Girolando.

Durante a coleta houve acompanhamento da ordenha do leite, e as amostras foram transportadas sob refrigeração em caixa isotérmica, para o Núcleo de Pesquisa e Extensão em Alimentos (NUPEA) da Universidade Estadual da Paraíba em Campina Grande. Onde as amostras foram pasteurizadas a 65 °C por 30 minutos e resfriadas a 4 °C.

Tabela 2 – Lotes caprinos e bovinos que foram coletadas as amostras de leite.

Lote	Tipo de criação	Raça	Localização geográfica
Caprino 1	Semi-extensivo	Moxotó	Taperoá/PB
Caprino 2	Semi-extensivo	Moxotó	Taperoá/PB
Caprino 3	Semi-extensivo	Murciana	Taperoá/PB
Caprino 4	Semi-extensivo	Alpina Americana	Taperoá/PB
Caprino 5	Semi-extensivo	Parda	Taperoá/PB
Caprino 6	Intensivo	Anglo Nubiana	Campina Grande/PB
Bovino 1	Intensivo	Sindi	Taperoá/PB
Bovino 2	Intensivo	Guzerá	Taperoá/PB
Bovino 3	Semi-Intensivo	Sindi/Guzerá	Taperoá/PB
Bovino 4	Semi-Intensivo	Girolando	Sobrado/PB

Fonte: elaborada pelo autor, 2021.

As amostras adulteradas de leite caprino foram produzidas perante a mistura individual e aleatória dos diferentes lotes de leites caprino e bovino puros em diferentes níveis. As adições foram feitas nos níveis: 5, 10, 15, 20, 30, 40, 50, 60, 75% de leite bovino (m/m) em um total de 10 g para cada amostra adulterada. Em seguida, foram homogeneizadas em tubos tipo Falcon de 50 mL por sucessivas inversões.

3.2 Análise espectral

Para obtenção dos espectros de 6 lotes de leite caprino, 4 lotes de leite bovino, assim como suas misturas, foram submetidos a análise em equipamento NIR portátil, modelo DLP NIRscan Nano Evaluation Module, marca Texas Instruments. Cada espectro foi obtido a partir de uma média de 32 varreduras. Como branco foi usado o espectro de uma pastilha de politetrafluoretileno (PTFE).

Os espectros NIR na região entre 900 e 1650 nm foram obtidos das amostras de leite caprino dos 6 lotes e combinações aleatórias entre eles resultando em 54 amostras. Assim como das amostras adulteradas, constituídas por mistura de leite caprino e bovino em diferentes níveis, resultando em 146 amostras.

3.3 Análise dos dados

As análises dos dados foram realizadas com uso dos softwares *Unscrambler versão 9.8* e em ambiente *Matlab versão R2011a*, com aplicação de Multivariate Data Handling (*data_hand_gui*) e Linear Discriminant Analysis Selection toolbox (*ldavs_gui*), elaborados pelo prof. Dr. Adriano de Araújo Gomes.

Técnicas de pré-processamento foram utilizadas para minimizar o espalhamento e variações sistemáticas da linha de base dos espectros. Dentre essas técnicas foram usadas, filtro média móvel, correlação de linha de base, correção linear da linha de base, SNV e MSC. Assim como combinações entre os pré-processamentos foram avaliadas.

A PCA foi usada métodos de classificação não supervisionado e a LDA como método supervisionado. Para esse último, cada classe foi dividida em conjunto de treinamento (70%) e teste (30%), por aplicação do algoritmo de Kennard-Stone (KS). De modo que o conjunto de treinamento possuía 38 amostras de leite caprino e 102 amostras adulteradas por leite bovino. Enquanto o conjunto de teste, 16 de leite caprino e 44 adulteradas por leite bovino.

Modelos quimiométricos GA-LDA utilizado com parâmetros padrão fixos de combinação dos operadores genéticos (crossover) em 60%, probabilidade de mutação em 5%, tamanho de população de 50 indivíduos, envolvendo-os em 100 gerações, com variância máxima de 15%. O ACO-LDA foi aplicado com parâmetros fixos de taxa de evaporação de feromônio em 65%, aplicado em 50 formigas, formando 100 colônias, sendo 0,35 formigas “cegas”, com variância máxima de 15%.

Ao término das desenvolvimento dos modelos de GA-LDA e ACO-LDA aplicados no conjunto de teste das amostras, em que foi avaliado os parâmetros

de acurácia, sensibilidade e especificidade para predição das em cada classe. Para cada algoritmo foram criados e testados 50 modelos em decuplicada, por ser algoritmos estocásticos. Sendo escolhido o melhor modelo diante dos parâmetros para representar a distinção entre as amostras.

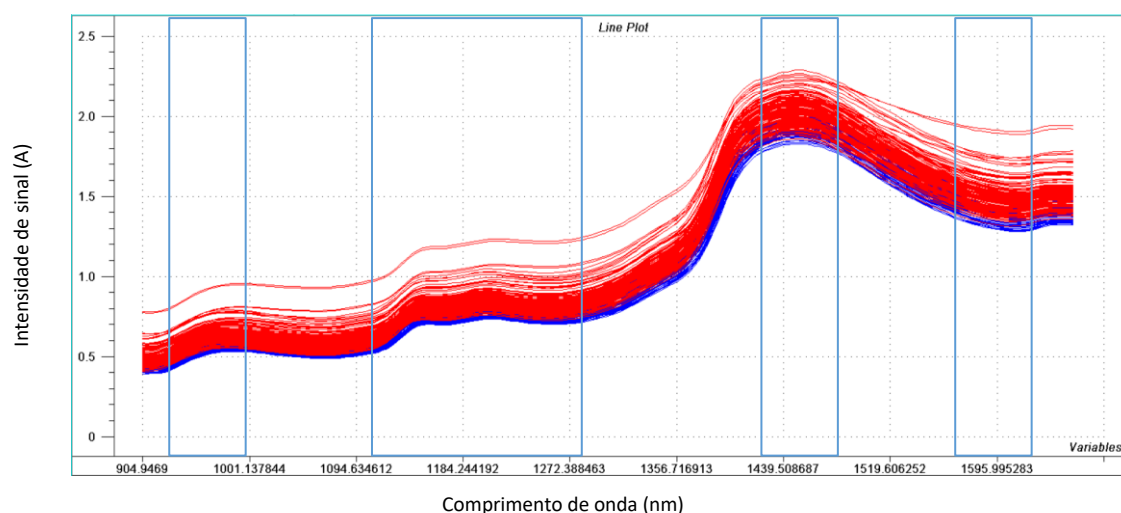
Os modelos classificação por LDA criados foram avaliados pela disposição das amostras nas matrizes de confusão. Essas matrizes de confusão comparam as classes verdadeiras (pré-estabelecidas) para as amostras com as classes estimadas, gerando a interpretação de falsos positivos (FP), falsos negativos (FN), verdadeiros positivos (VP) e verdadeiros negativos (VN). A partir desses valores podem ser avaliados os parâmetros de acurácia, sensibilidade e especificidade dos modelos. As variáveis selecionadas pelos algoritmos também foram avaliadas para identificar os modelos que mais conseguiram representar a variabilidade das amostras e distinguir as amostras perante as classes pré-estabelecidas.

4 RESULTADOS

4.1 Medidas espectrais e pré-processamentos

Os espectros das amostras de leite caprino e de adulterações destas com leite bovino na faixa de 5 a 75% (v/v) foram obtidos na região espectral de 904 a 1650 nm. Ao observar os espectros (dados brutos), **Figura 1**, das amostras de leite caprino, autênticas e adulteradas, é perceptível uma banda por volta de 1450 nm. Referente ao primeiro sobretom do estiramento vibracional da ligação O-H respectivo a água. Outra banda com menor intensidade de sinal, entre 1100 e 1200 nm, correspondente ao segundo sobretom de estiramento de ligação C-H relacionada aos carboidratos. Em aproximadamente 1630 nm há combinação harmônica de estiramentos N-H livres e amidas secundárias associada às proteínas presentes nas amostras.

Figura 1 – Espectros dos dados brutos de todas as amostras separados por classe.



Legenda: espectros vermelhos (amostras adulteradas por leite bovino); espectros azuis (amostras puras de leite caprino).

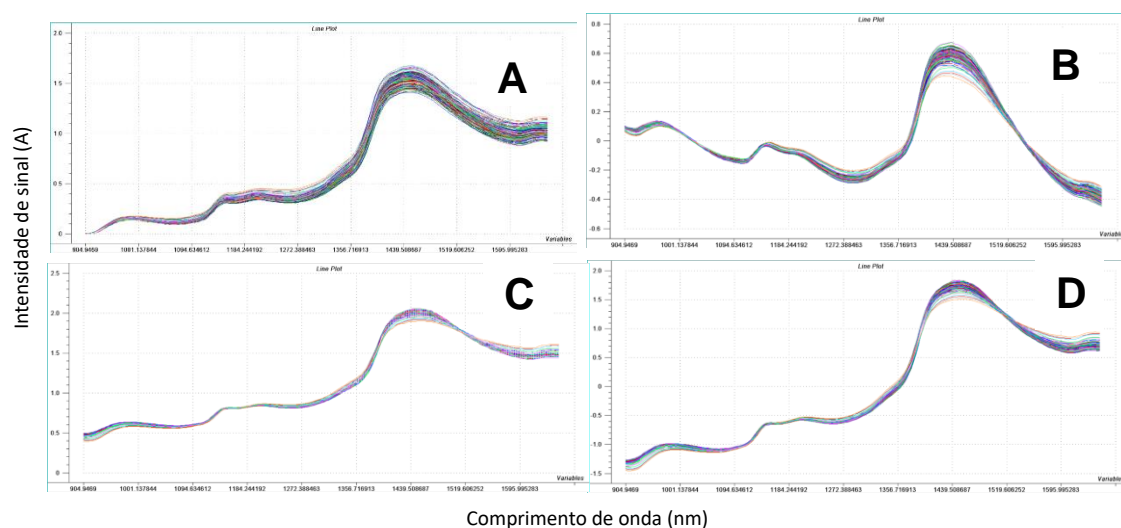
Fonte: elaborada pelo autor, 2021.

Visto que os espectros dos dados brutos apresentam um deslocamento da linha de base e ruídos relacionados, possivelmente, associada a alta sensibilidade analítica da espectroscopia NIR. Geralmente a sensibilidade é afetada por propriedades físico-químicas como tamanho de partículas, e mais comumente informações indesejadas. Portanto, fez-se necessário o uso de pré-processamentos dos dados (PASQUINI, 2018).

O filtro de média móvel foi aplicado em diversos comprimentos de filtro com intuito de verificar o efeito do pré-tratamento na elucidação das informações espectrais. Evitando que houvesse redução significativa constatada pelos métodos de classificação, tanto não supervisionados quanto supervisionados. Entre as diversas janelas (pontos) testadas, aquela com 7 pontos foi escolhida devido aos melhores resultados.

Nos espectros é visível o impacto dos pré-processamentos realizados após o filtro de média móvel, comparando aos espectros brutos. A intensidade de sinal foi reduzida e algumas bandas foram intensificadas. Ainda houve uma normalização de algumas amostras que possuíam intensidade de sinal elevado. Na **Figura 2**, nivelamento da linha de base pelo filtro de média móvel com 7 pontos associado individualmente a: linha de base *offset* (BO); correlação linear da linha de base (LCB); SNV; MSC.

Figura 2 – Espectros dos dados pré-processados com filtro de média móvel + (A) linha de base *offset* (BO), (B) correlação linear da linha de base (LCB), (C) SNV e (D) MSC.



Fonte: elaborada pelo autor, 2021

O uso de pré-tratamentos no conjunto de dados espectrais proporcionou uma melhor uniformidade na visualização das amostras. A observação torna-se mais minuciosa por meio da PCA, pelos grupos formados nos gráficos de escores. A análise baseia-se nos dados brutos e como acontecem mudanças nos agrupamentos dos dados em cada caso de pré-processamentos escolhidos.

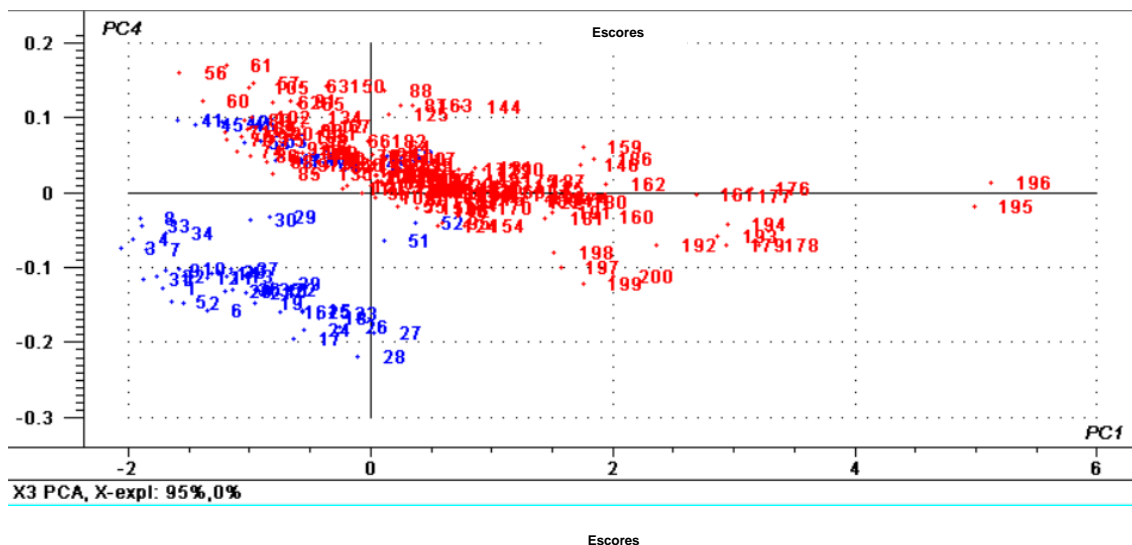
Ao associar um maior número de pré-tratamentos proporcionou elucidação de informações espectrais, além da redução dos ruídos espectrais. O uso da derivada de Savitzky e Golay (1964) mostrou ainda menos satisfatórios na interpretação dos dados pelos métodos quimiométricos, como os gráficos escores da PCA não apresentou quaisquer distinção entre agrupamentos, impossibilitando para busca da classificação das amostras nas classes pré-estabelecidas.

As análises dos dados foram realizadas usando os métodos de classificação não supervisionado e supervisionado. Para isto foram considerados os resultados obtidos pelos espectros brutos e após o filtro de média móvel juntamente a: linha de base offset; correção linear de linha de base; SNV; MSC.

4.2 Análise exploratória dos dados brutos e pré-processados por PCA

A partir da análise exploratória dos dados foi possível identificar dois agrupamentos baseados no gráfico de escores. Esses agrupamentos foram revelados pela relação entre a 1ª componente principal (PC1) e a 4ª componente principal (PC4) projetados no espaço amostral. As 2ª e 3ª componentes principais apresentaram resultados similares ao da 1ª PC, tanto para gráfico de pesos quanto para o de escores. Diante das PCs, é possível perceber um número alto de amostras com comportamento similar sendo agrupadas próximas e divergentes ao outro agrupamento formado com número menor de amostras, demonstrado na **Figura 3a**. Pelo conhecimento *a priori* sobre as amostras é possível induzir que o maior agrupamento sejam amostras adulteradas e o menor sejam amostras autênticas. Para auxiliar essa visualização foi criada uma variável categórica com a real instrução das amostras, de acordo com adulterações ou não, como mostrado na **Figura 3b**.

Figura 3 – Escores da PCA dos dados brutos com visualização pela variável categórica.



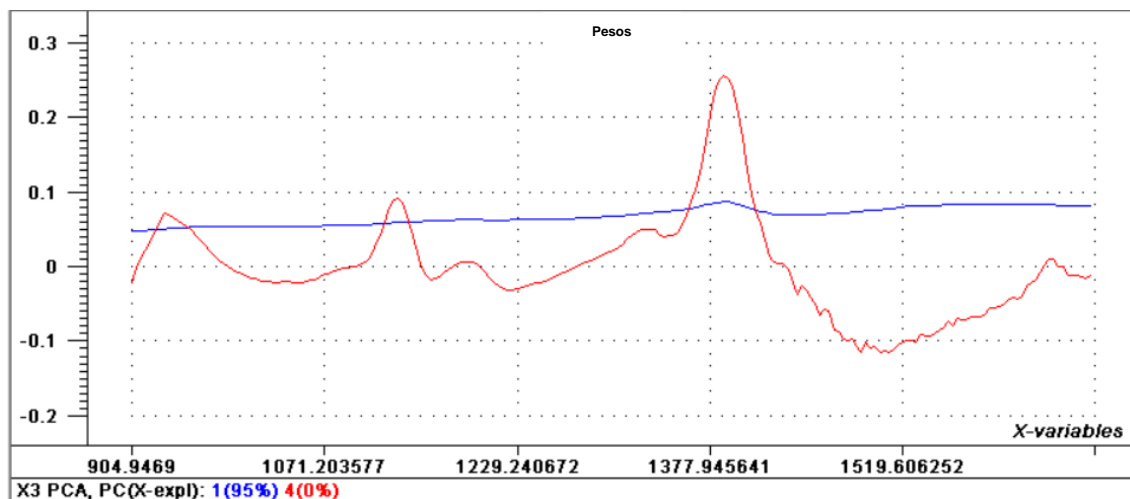
Legenda: amostras vermelhas (amostras adulteradas por leite bovino caprino); amostras azuis (amostras puras de leite).

Fonte: elaborada pelo autor, 2021.

O uso da variável categórica auxilia na verificação se os agrupamentos formados pela PCA condizem com a referência verdadeira sobre o conjunto de dados. Pode-se perceber que algumas amostras puras possuem comportamento que as agrupam mais próximas a misturas de leite caprino e leite bovino, do que ao agrupamento de amostras puras. Possivelmente relacionado as condições de manejo e as espécies que foram coletados os leites caprinos, ou ainda a mistura desses leites.

Outro fator importante reforçado pelo gráfico de pesos, é que apesar a variância explicada absoluta de PC1 é 95,312% e de PC4 é 0,421%, essa última aparenta apresentar um poder discriminante superior comparando com as demais componentes principais, visualizados na **Figura 4**. Os pesos da PC4 apresentam informações relevantes em relação a análise espectral dos dados. De forma que as variáveis espectrais utilizadas são relacionadas a variáveis que podem conter conhecimentos que contribuíram para formação dos agrupamentos no gráfico de escores. Ainda que esses agrupamentos não correspondem totalmente de acordo com a variável categórica, promove entendimento sobre a tendência de comportamento das amostras.

Figura 4 – Gráfico de pesos das PC1 e PC4 dos dados brutos.



Fonte: elaborada pelo autor, 2021.

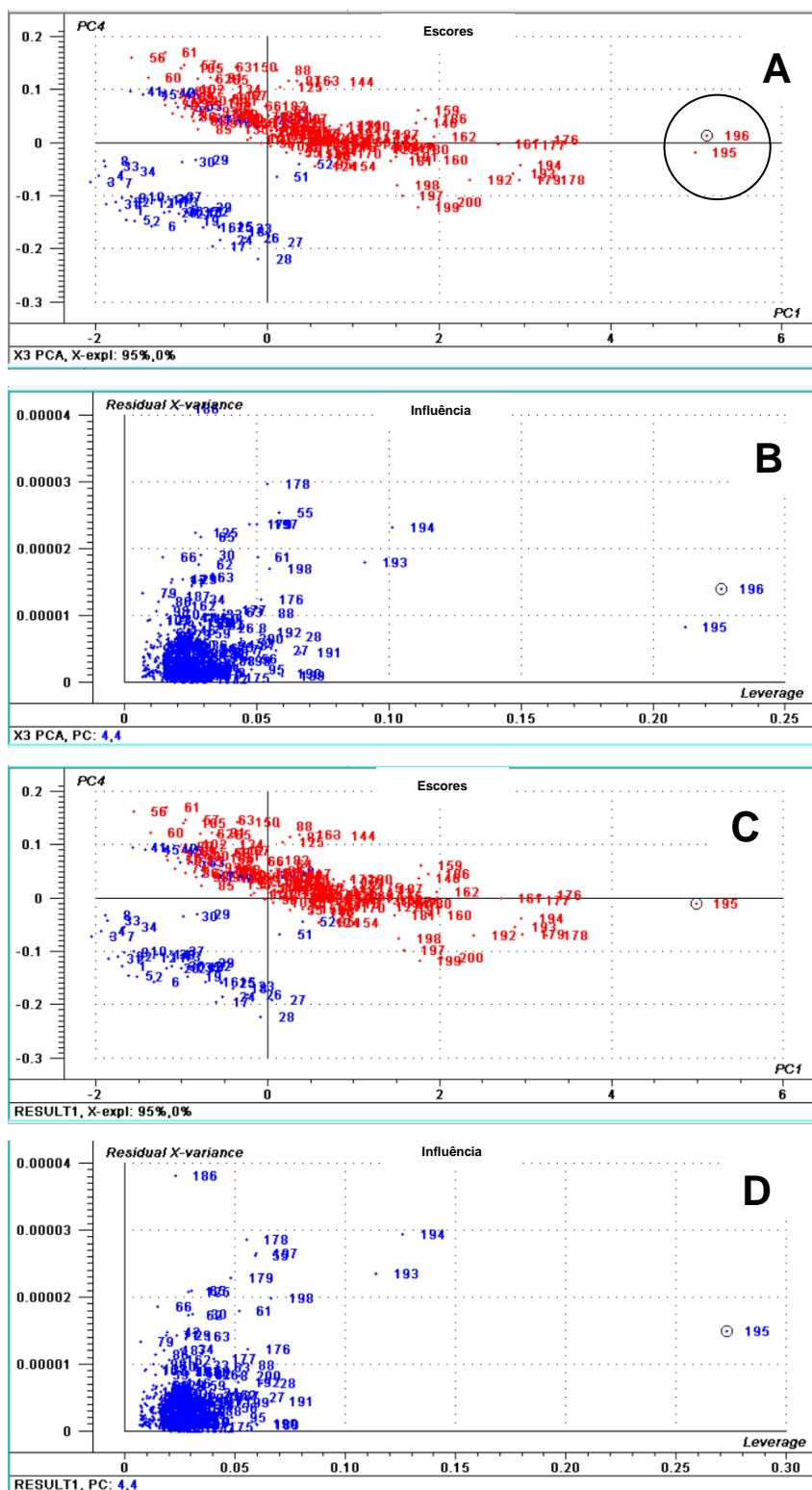
Como comentado anteriormente, gráfico de pesos elucidam melhor a relação de cada variável real com cada PC, na **Figura 4**. A PC4 apresenta picos referentes às ligações C-H presentes nas amostras, sendo o segundo sobretudo entre 1100 e 1200 nm e entre 1350 e 1400 nm. Essas variáveis auxiliaram na formação dos dois agrupamentos de amostras devido à presença de carboidratos, em teores distintos, nos leites caprino e bovinos (XIAOBO et al., 2010; PAQUINI, 2018; BEC; HUCK, 2019; ZAREEF et al, 2020).

Algumas das amostras apresentaram comportamentos um pouco distintos das demais, podendo ser potenciais amostras anômalas (outliers). Segundo Bro e Smilde (2014), o uso da variância dos resíduos é a métrica mais simples e comum para detectar amostras anômalas. De modo que a influência das amostras é medida diante das informações presentes que não são importantes para a construção do modelo de PCA, caso sejam valores altos, a amostra é considerada como anômala.

O gráfico de influência foi estudado para identificar se a anomalia na disposição dessas amostras diante do conjunto de dados prejudicava a explicação da distribuição das amostras. Em que a PCA foi recalculada, excluindo primeiramente a amostra 196, o que não apresentou nenhum impacto considerável na variância explicada dos dados ou na distribuição dos agrupamentos. A partir disso a amostra 195 foi excluída do cálculo e também

não houveram alterações significativas, como mostrados nos gráficos da **Figura 5**, a seguir.

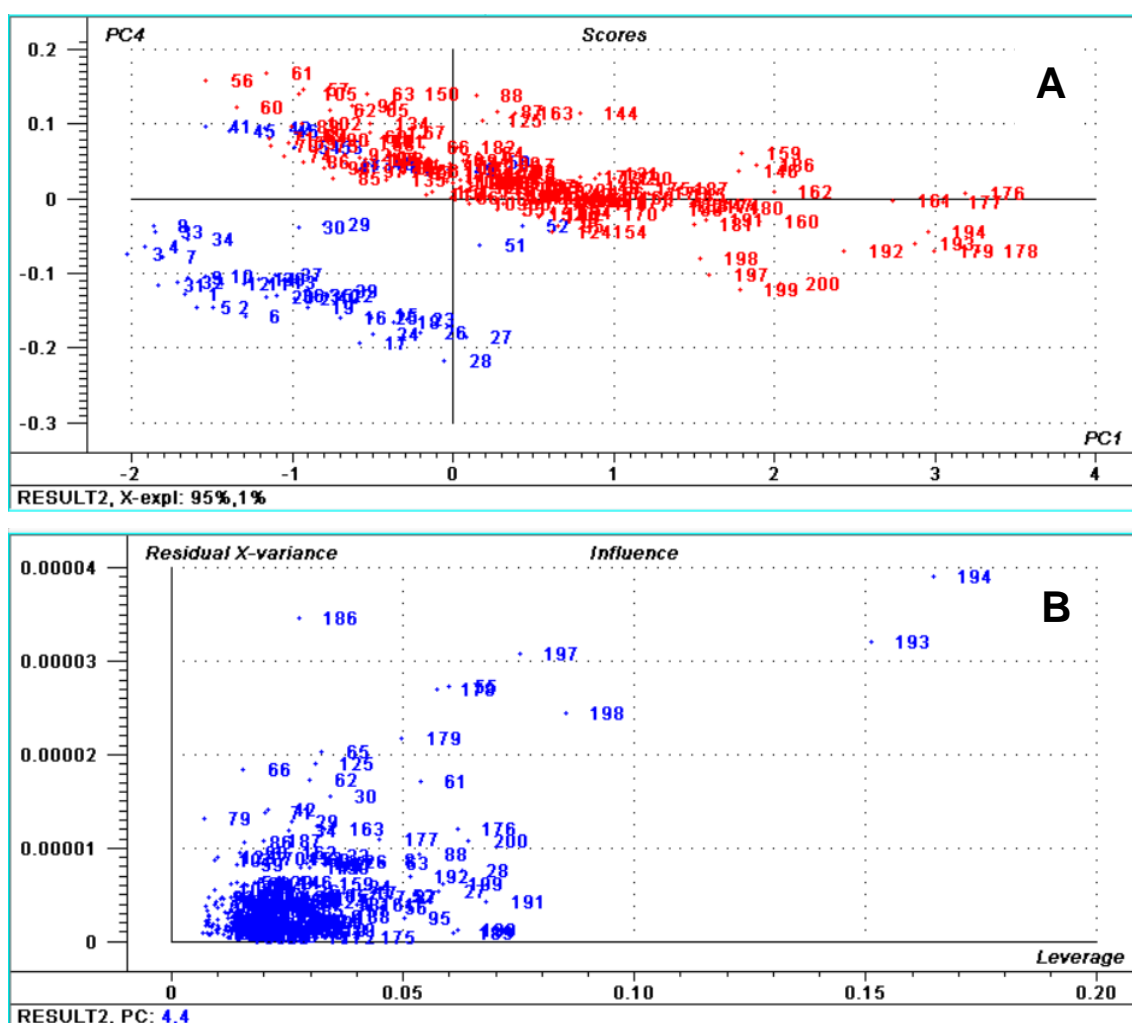
Figura 5 – Análise de possível amostra anômala 196 pelos gráficos de (A) escores dos dados brutos com e (B) influência em relação ao resíduo na PC4, e para amostra 195 em (C) escores e (D) influência.



Fonte: elaborada pelo autor, 2021.

A remoção das amostras 196 do cálculo da PCA foi avaliada pela alta influência, como podemos perceber na **Figura 5b e 5d**. Ainda que outras amostras possuíssem valores de resíduo elevados como a amostra 186. As amostras 195 e 196 destacam-se pelo diferencial em relação a influência e o resíduo considerável ao comparar com as demais amostras do conjunto de dados, referentes a **Figura 6b**. Com a remoção das duas amostras a distribuição do conjunto de dados manteve-se similar ao gráfico original (**Figura 5a**), como mostrado no gráfico de escores da **Figura 6a**, abaixo.

Figura 6 – Análise da distribuição amostral na PCA sem possíveis amostras anômalas (195 e 196) pelos gráficos de (A) escores e (B) influência.



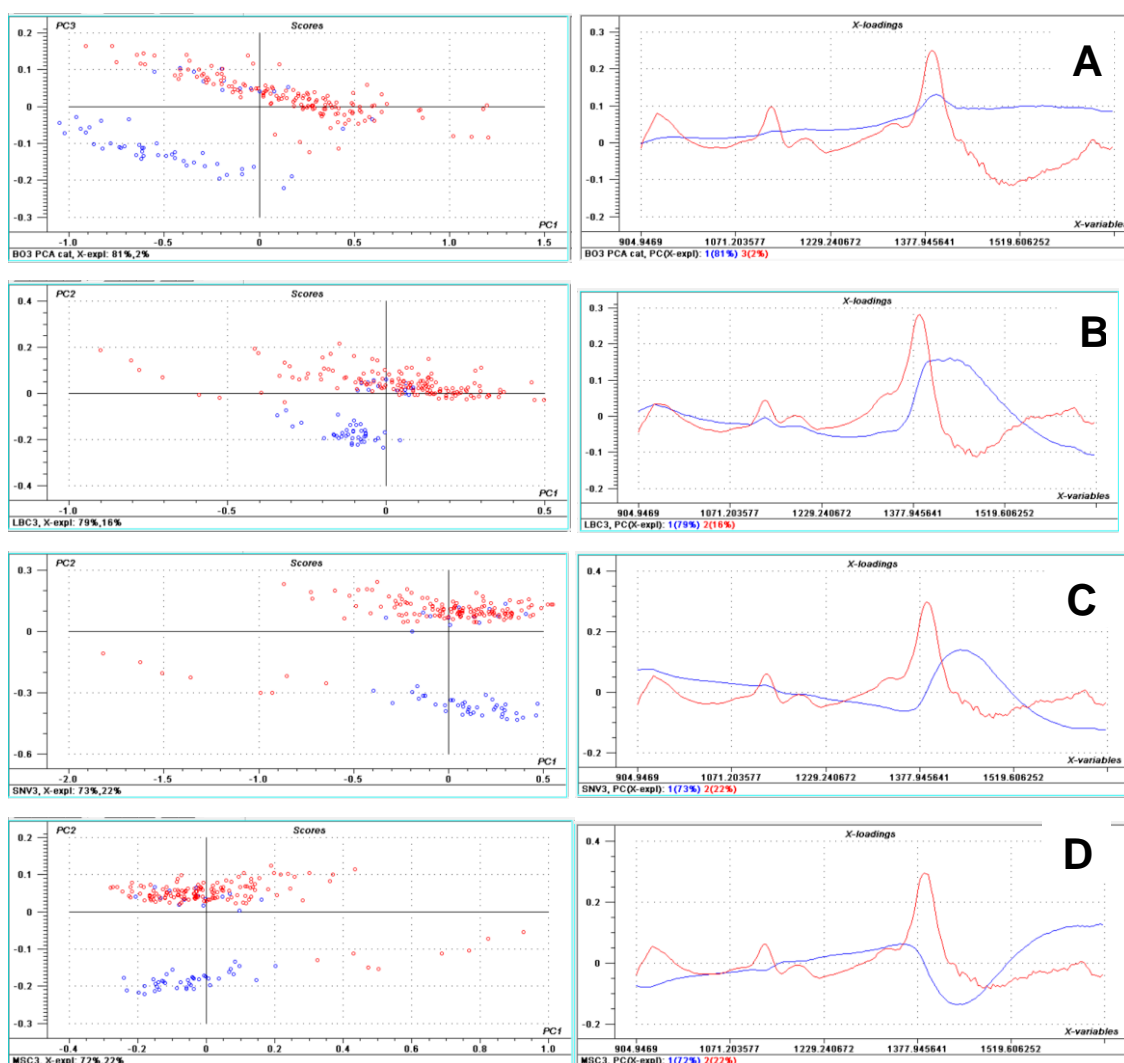
Fonte: elaborada pelo autor, 2021.

A presença de amostras anômalas foi descartada devido a análise do gráfico de influência e do comportamento das amostras no espectro original. As duas amostras (195 e 196) apresentam tendência a serem consideradas outliers, devido aos valores de influência (*leverage*) e resíduo (**Figura 6b**) considerados

simultaneamente altos, possivelmente causado por ruídos espectrais. Apesar disso há uma coerência comparando com outras amostras que apresentam resíduo mais alto e amostras que apresentam apenas da influência alta.

Como perceptível nos espectros pré-processados, na **Figura 2**, os pré-processamentos conseguiram reduzir consideravelmente as divergências espectrais e o resultado é notado também pela PCA. Além do agrupamento das amostras referente a variável categórica que distingue as amostras autênticas de leite caprino (azul) e as adulteradas por leite bovino (vermelho). A **Figura 7** a seguir representam os escores da PCA dos dados pré-processados.

Figura 7 – Escores e pesos da PCA dos dados pré-processados com filtro de média móvel com 7 pontos associado a (A) linha de base offset, (B) correção linear de linha de base, (C) SNV e (D) MSC.



Fonte: elaborada pelo autor, 2021.

A visualização dos escores após o uso dos pré-processamentos indicaram que algumas amostras do agrupamento de adulteradas (vermelho) apresentaram comportamento mais aproximado a amostras autênticas. Esse comportamento distinto é possivelmente ocasionado pela alta variabilidade das amostras coletadas de leite caprino e leite bovino. Condições como manejo e raça dos animais podem ter sido fatores que levaram à similaridade entre as informações analíticas coletadas para criação dos modelos PCA.

Ao considerar que os resultados apresentem uma tendência a separação entre as classes, possibilitando a distinção desses agrupamentos pela classificação supervisionada e o uso de seleção de variáveis. Essa seleção de variáveis auxilia na busca por informações espectrais pontuais que discrimine, por meio da LDA, as amostras autênticas das amostras contaminadas por leite bovino.

4.3 Modelos obtidos por GA-LDA e ACO-LDA

Os dados brutos e após aplicação dos pré-processamentos de filtro média móvel seguidos por linha de base *offset*, correlação linear de linha de base, SNV, MSC foram utilizados para análise discriminante linear (LDA) como técnica supervisionada de reconhecimento de padrões. Os algoritmos de seleção e variáveis usados foram o algoritmo genético (GA), o algoritmo por colônias de formigas (ACO).

Anteriormente a LDA, os dados foram separados em conjuntos de treinamento e teste utilizando o algoritmo de Kennard-Stone (KS). A separação das amostras aconteceu de modo que ambos conjuntos representassem as classes e as variabilidades das amostras. Sendo um total de 200 amostras, 54 autênticas de leite caprino e 146 adulteradas por adição de leite bovino em diversos níveis. Então, originou-se um conjunto de treinamento com 140 amostras e um conjunto de teste com 60 amostras.

As amostras de treinamento foram utilizadas para criação dos modelos GA-LDA e ACO-LDA, enquanto as amostras de teste serviram para avaliar a capacidade preditiva realizada por cada modelo. Cada modelo foi desenvolvido em decuplicata (dez repetições) nas mesmas condições de pré-processamentos.

Por fim, foi selecionado o modelo mais adequado entre os dez construídos de cada pré-processamento para a classificação das amostras.

Os modelos de classificação GA-LDA e ACO-LDA foram avaliados de acordo com a predição realizada na etapa de teste, de acordo com os parâmetros de taxa de acurácia, sensibilidade e especificidade das adequações nas classes construídas. Sendo classe 1, amostras de leite caprino autênticas, e classe 2, amostras adulteradas por adição de leite bovino, como mostrado da **Tabela 3**. Os modelos foram comparados entre si para encontrar o melhor modelo que descrevesse a predição das amostras de acordo com as variáveis selecionadas.

Tabela 3 – Valores de predição de modelos criados por GA-LDA e ACO-LDA utilizando full cross validation.

Pré-processamento	Classes	GA-LDA			ACO-LDA				
		VS	A (%)	S (%)	E (%)	VS	A (%)	S (%)	E (%)
Brutos	1	5	100	93,7	100	5	100	93,7	100
	2		97,8	100	93,7		97,8	100	93,7
MM+BO	1	8	100	100	100	7	100	100	100
	2		100	100	100		100	100	100
MM+LBC	1	7	100	87,5	100	6	100	93,7	100
	2		95,6	100	87,5		97,8	100	93,7
MM+MSC	1	6	100	87,5	100	6	100	87,5	100
	2		95,6	100	87,5		95,6	100	87,5
MM+SNV	1	7	100	87,5	100	4	100	87,5	100
	2		95,6	100	87,5		95,6	100	87,5

Legenda: VS: Variáveis Selecionadas; A: Acurácia; S: Sensibilidade; E: Especificidade.

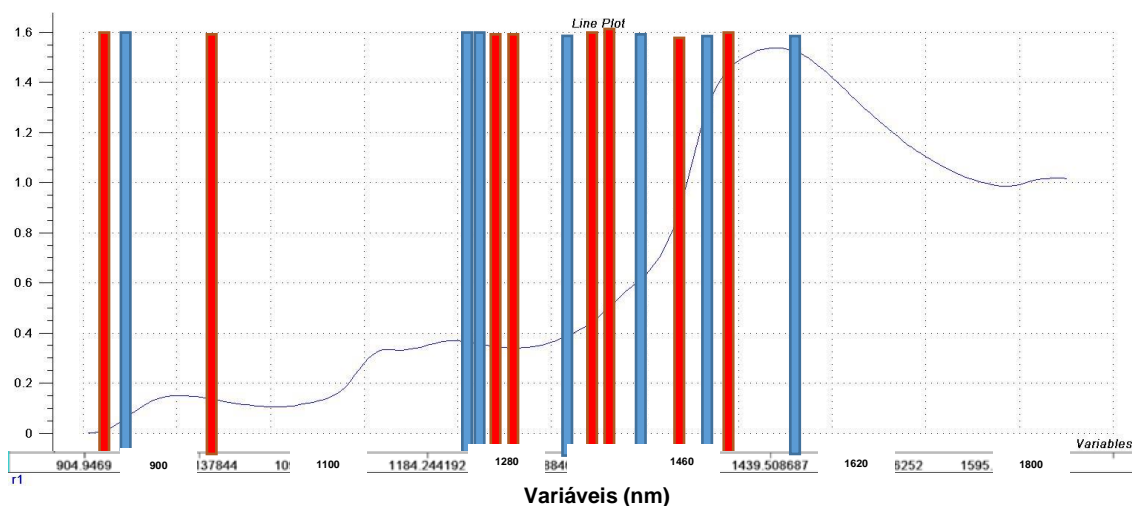
Fonte: elaborada pelo autor, 2021.

Os modelos construídos com os algoritmos de seleção de variáveis por GA e ACO com os dados brutos e usando pré-tratamento de linha de base *offset* apresentam resultados mais significativos. Essa escolha correspondeu a predição de amostras de acordo com as classes pré-estabelecidas para identificar a adulteração. Apesar de utilizar selecionar mais variáveis que os

demais modelos, é possível obter 100% de acurácia, especificidade e sensibilidade.

A **Figura 8** apresenta os oito comprimentos de onda (variáveis) selecionados pelo GA-LDA se encontram distribuídos no ao longo do espectro médio pré-processado com filtro de média móvel e linha de base *offset* (MM+BO). Assim como, sete variáveis selecionadas pelo ACO-LDA utilizando os dados pré-processados por filtro de média móvel e linha de base *offset* (MM+BO).

Figura 8 – Espectro médio com destaque as variáveis selecionadas pelo GA-LDA e ACO-LDA após pré-processamentos dos dados por MM+BO.



Legenda: Variáveis vermelha selecionadas pelo GA-LDA; variáveis azuis selecionadas pelo ACO-LDA.

Fonte: elaborada pelo autor, 2021.

O GA-LDA foi aplicado no banco de dados em dez repetições para verificar a eficiência do método em escolher variáveis distintas que apresentassem informações relevantes para a análise discriminante. O modelo escolhido entre os obtidos representou a tendência de resultados apresentado pelos demais para predição das amostras com uso de oito comprimento de ondas: 917, 1032, 1206, 1239, 1336, 1339, 1371 e 1402. Seus pesos são apresentados na **Figura8**.

O modelo criado com ACO-LDA também demonstrou resultado satisfatório com uso dos mesmos pré-processamentos (MM+BO) para prever as amostras com 100% de acurácia, sensibilidade e especificidade. As variáveis

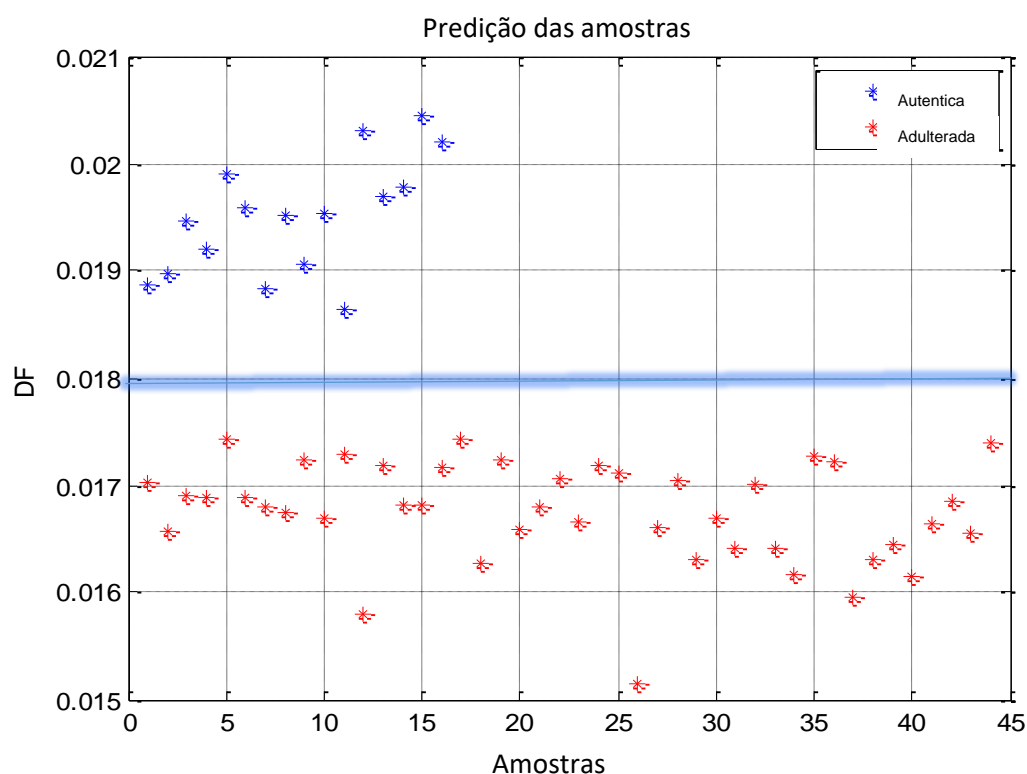
selecionadas foram 945, 1226, 1236, 1282, 1318, 1399 e 1456, referentes a mesma faixa de comprimentos de onda (900-1650 nm) que o modelo GA-LDA.

Os comprimentos selecionados do espectro MM+BO pelo modelo representam pesos diferentes na discriminação das amostras, como mostrado pela **Figura 8**. Assim como nos pesos da PCA, as variáveis selecionadas pelo GA tentam a proporcionar informações nas bandas espectrais próximas 1200 nm e entre 1350 e 1400 nm, correspondentes a estiramentos C-H associado a carboidratos das amostras. Com base na interpretação das variáveis, a adição de leite bovino no leite caprino promove alteração nesse componente e a divergência no comportamento espectral das amostras (XIAOBO et al., 2010; PAQUINI, 2018; BEC; HUCK, 2019; ZAREEF et al, 2020).

Outras variáveis relevantes ao modelo GA-LDA (MM+BO) é a 917 e ao modelo ACO-LDA é a 948, as quais representam a presença de sobretons do estiramento O-H nas amostras referente a água presente como componente majoritário das amostras analisadas. O leite independente da espécie apresenta grande quantidade água, esse componente é frequentemente utilizado como adulterante, porém essa possível contaminação foi controlada o acompanhamento da coleta e das condições de processamento e acondicionamento até as análises espectrais.

As variáveis selecionadas (VS) pelo GA para o modelo LDA, mostrada na **Figura 8**, viabilizaram a construção do modelo a partir do conjunto de treinamento. A partir disso foi possível visualizar na predição das amostras um limiar que distingue por completo as amostras do conjunto de teste em amostras autênticas de leite caprino e amostras adulteradas por leite bovino, na **Figura 9**, em azul e vermelho, respectivamente.

Figura 9 – Discriminação linear das amostras do conjunto de teste por GA-LDA (MM+BO).

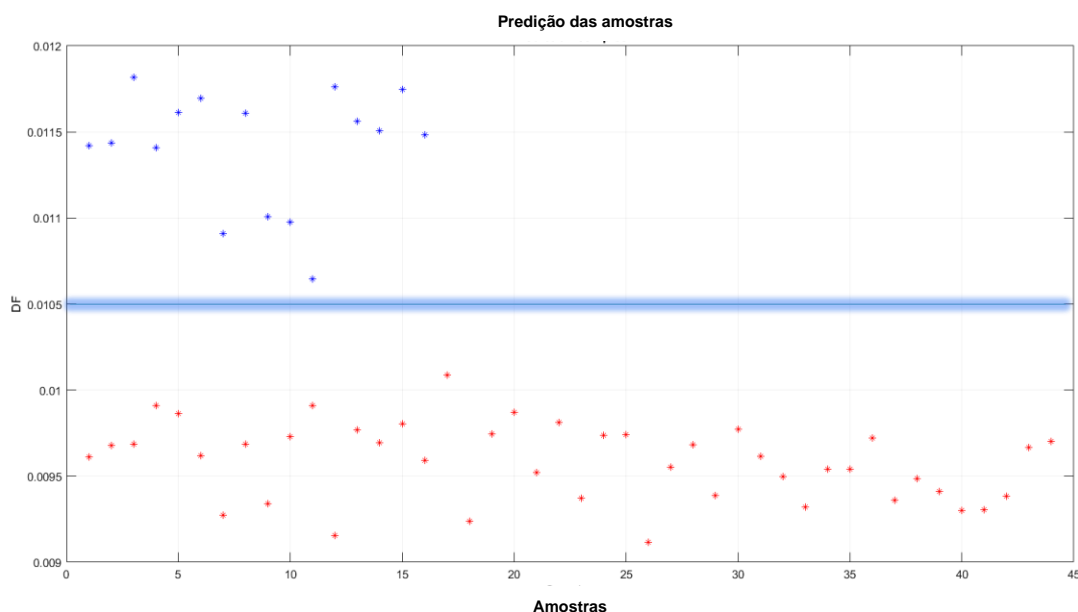


Fonte: elaborada pelo autor, 2021.

O modelo criado por GA-LDA com MM+BO selecionou mais variáveis que os modelos criados com os demais pré-processamentos. Porém conseguiu prever com acurácia e especificar de acordo com a classe pertencente. Além de possuir sensibilidade de adequar todas amostras do conjunto de teste formando duas classes distintas separadas pelo limiar discriminante (DF), destacado em azul nas **Figuras 9 e 10**. A criação do modelo com GA-LDA e ACO-LDA mostraram-se adequadas ao conjunto de dados para distinção das amostras e identificação das adulterações no leite caprino por leite bovino.

O comportamento preditivo do ACO-LDA (**Figura 10**) mostra-se similar ao poder discriminante das variáveis selecionadas pelo GA-LDA, reforçando a ideia que ainda que os dois algoritmos selecionem variáveis diferentes, a análise da distinção das amostras é regida pela LDA. De forma que, a seleção das 7 variáveis por ACO-LDA no espectro de dados (MM+BO) mostrou-se mais que eficiente para a distinção das amostras. Mostrada pela **Figura 10**, abaixo.

Figura 10 – Predição das amostras do conjunto de teste por ACO-LDA.



Fonte: elaborada pelo autor, 2021.

A predição das amostras distinguiu os dois grupos, autênticas e adulteradas, pela construção do limiar discriminante (DF). Percebe-se que há uma separação significativa das amostras do conjunto de predição. Essa constatação reafirma os resultados satisfatórios de acordo com os parâmetros analisados anteriormente do modelo BO criado por ACO-LDA.

As variáveis selecionadas pelos modelos dos dados pré-processados com MM+BO são correspondentes as bandas de sobretons do estiramento vibracional das ligações O-H e C-H presentes nas amostras de leite, como mostrado no gráfico de pesos da PCA. Essas variáveis conseguiram distinguir as amostras autênticas das amostras adulteradas de acordo com o teor de água e de carboidratos, componentes majoritários presentes nas amostras de leite. O teor de gordura (carboidratos) especialmente diferencia o leite caprino do bovino, pela discrepância percentual presente nas amostras, sendo 5,8% e 3,3%, respectivamente.

Os modelos criados a partir dos dados brutos e com seleção de variáveis demonstraram resultados satisfatórios. Os algoritmos GA-LDA e ACO-LDA dos dados brutos selecionaram cinco variáveis espectrais. Mesmo sendo um valor inferior ao modelo MM+BO, conseguiu prever a separação das amostras nas

classes com apenas um erro de classificação. Destacado pelas matrizes de confusão e pelos parâmetros estatísticos dos modelos.

Os demais pré-processamentos, LCB, SNV e MSC mostraram resultados similares, ainda que inferiores aos citados anteriormente, mas ainda consideráveis para distinção entre as classes das amostras. Cada modelo apresenta apenas 2 erros de classificação de amostras autênticas como amostras adulteradas. Como mostrado nas matrizes de confusão da **Tabela 4** abaixo:

Tabela 4 – Matrizes de confusão dos modelos criados para etapa de teste.

Pré-processamento	Classe verdadeira	Classe prevista	
		Autênticas	Adulteradas
Dados brutos	Autênticas	15	1
	Adulteradas	0	44
BO	Autênticas	16	0
	Adulteradas	0	44
LCB	Autênticas	14	2
	Adulteradas	0	44
SNV	Autênticas	14	2
	Adulteradas	0	44
MSC	Autênticas	14	2
	Adulteradas	0	44

Fonte: elaborada pelo autor, 2021.

A matriz de confusão mostra uma visão interessante que já era perceptível pelo gráfico de escores da PCA, algumas amostras da classe de leite caprinos autênticos tendem a agruparem-se com as amostras adulteradas. O pré-processamento de linha de base *offset* consegue minimizar essa tendência, porém os demais pré-processamentos (LCB, SNV e MSC) intensificam ao comparar com os dados brutos. Portanto, o modelo da linha de base *offset* o único que consegue separar totalmente as amostras do conjunto de teste nas classes.

Pereira et al. (2020) conseguiu modelos com o mesmo conjunto de dados para identificar a adulteração de leite caprino por leite bovino utilizando o método

dos mínimos quadrados parciais (PLS, do inglês *Partial Least Squares*) associado a análise one-class (OC-PLS, do inglês *One-class Partial Least Squares*) e análise discriminante (PLS-DA, do inglês *Partial Least Squares Discriminant Analysis*), além do uso do algoritmo de projeções sucessivas (iSPA-PLS-DA, do inglês *interval Successive Projection Algorithm with Partial Least Squares Discriminant Analysis*). Esses métodos de classificação utilizam a métrica de regressão do PLS para distinguir as amostras em classes distintas. Para o mesmo pré-processamento identificado como ideal para distinção das amostras usando GA-LDA e ACO-LDA, os resultados foram inferiores considerando os parâmetros de acurácia, sensibilidade e especificidade.

A seleção de variáveis por GA-LDA e ACO-LDA proporcionou melhores resultados na classificação do conjunto de dados comparado aos obtidos por Pereira et al. (2020). A identificação das variáveis que possuíam informações mais relevantes para distinção de amostras do grupo de leite caprino das amostras adulteradas por leite bovino aumentou a robustez dos modelos desenvolvidos pelos métodos dinâmicos.

Galvão et al. (2020) e Gomes et al. (2022) afirmam que a seleção das variáveis por iSPA segue uma vertente determinística em que são selecionados intervalos nos dados espectrais, a qual minimiza a multicolinearidade dos dados e evita variáveis redundantes. Por outro lado, na abordagem dinâmica utilizada pelos algoritmos GA e ACO, é associada uma avaliação mais minuciosa da resolução final, considerando a seleção de pontos de variáveis individuais realizada por esses algoritmos. Durante esse estudo cada modelo GA-LDA e ACO-LDA foi testado 10 vezes para estudar esse comportamento mais dinâmico dos algoritmos.

Com isso, a construção dos modelos foi realizada no intuito de extrair as variáveis mais relevantes para distinção de amostras autênticas e adulteradas. Ao comparar com a metodologia quimiométrica empregada por Pereira et al. (2020), é perceptível uma intensificação da importância das variáveis selecionadas para a confirmação da autenticidade das amostras de leite caprino. Na **Tabela 5** estão destacados os resultados obtidos em comparação a referência.

Tabela 5 – Comparação dos modelos de referência (Pereira et al., 2020) com os modelos obtidos.

Modelos (MM+BO)	Parâmetros estatísticos		
	A (%)	S (%)	E (%)
OC-PLS	69,1	92,9	66,7
PLS-DA	88,3	100/84,1	84,1/100
<i>i10</i> -SPA-PLS-DA	98,3	100/97,7	97,2/100
GA-LDA	100	100	100
ACO-LDA	100	100	100

Fonte: elaborada pelo autor, 2021.

Os parâmetros estatísticos indicam uma acurácia crescente ao implementar o PLS-DA, ainda mais associado a seleção de intervalos por SPA. Além disso, houve uma ascensão da sensibilidade e especificidade na predição das amostras diante das adulterações. Porém quando seletivo as amostras autênticas de leite caprino, não se mostrou tão específico as amostras adulteradas, vice-versa.

A seleção de variáveis realizada por SPA agregou na construção de um modelo com parâmetros estatísticos mais altos que os demais utilizados por Pereira et al. (2020). O SPA baseia-se na criação de um vetor de variáveis menor possível que minimize a colinearidade. Entretanto, as variáveis selecionadas podem possuir baixa relação entre sinal e ruído, impedindo a interpretação da informação química. Exigindo um estudo mais aprofundado com repetições, assim como as seleções de variáveis por GA e ACO.

Os algoritmos GA e ACO seguem uma vertente mais dinâmica na seleção de variáveis, a qual inicialmente é de forma aleatória. Contudo, as combinações são continuadas de acordo com a meta-heurística (fundamento biológico) de cada algoritmo. Por outro lado, o SPA é um método determinístico mais baseado nos resultados que são obtidos pelas projeções sucessivas nas variáveis. De forma que esses métodos apresentam formas distintas de compreender as variáveis escolhidas para construção dos modelos quimiométricos (PONTES, 2009; XIAOBO et al., 2010; GOMES et al. 2022).

A associação da seleção de variáveis com a LDA mostrou-se uma ferramenta promissora na criação de modelos quimiométricos para a análise da adulteração do leite caprino por leite bovino. Ainda que requeressem repetições durante o processo, os resultados com GA-LDA e ACO-LDA conseguiram representar melhor o comportamento das amostras, em relação as propostas utilizando algoritmos com base no PLS.

Os modelos foram criados seguindo os mesmos pré-processamentos, porém como podemos perceber a forma de análise dos dados pode evidenciar ou minimizar informações relevantes. A classificação realizada pelo *i10*-SPA-PLS-DA em 4 intervalos mostrou-se adequada para a detecção de adulterações em leite caprino por leite bovino. Nesse sentido, a visualização por análise discriminante linear proporcionou outra perspectiva com a seleção de variáveis. De modo que, os modelos GA-LDA e ACO-LDA conseguiram descrever a autenticidade das amostras de leite caprino.

As formas de visualizar os dados apresentam atributos e limitações, que podem ser associadas para o um conhecimento e extração de informações relevantes sobre as amostras. A distinção das amostras adulteradas reforça que o uso das ferramentas quimiométricas são um potencial necessário para interpretação dos dados espectrais NIR. Além de auxiliar na investigação remota pelo uso de instrumentação portátil.

5 CONCLUSÕES

As análises efetuadas por espectroscopia NIR proporcionaram informação química sobre o leite caprino e adulterações, realizadas para prever um contexto real pela adição de leite bovino. A espectroscopia NIR apresenta benefícios como análises rápidas, de forma não destrutiva e sem gerar resíduos. A instrumentação NIR portátil auxilia na análise remota de alimentos perecíveis como o leite, além da análise *in situ*, logo após a coleta.

Os componentes do leite foram interpretados pelas bandas espectrais presentes nos comprimentos de onda na região NIR. Os componentes majoritários nas amostras de leite, carboidratos e água, foram essenciais para a compreensão dos agrupamentos formados pelas amostras na PCA. A aplicação dos pré-processamentos espectrais por filtro de média móvel com 7 pontos e linha de base *offset* (MM+BO) reduziram ruídos instrumentais.

Os algoritmos meta-heurísticos de GA e ACO foram primordiais para seleção de variáveis de interesse que descreveram as amostras, distinguindo-as em duas classes distintas pela LDA. O desenvolvimento dos modelos que conseguiram indicar as amostras verdadeiramente caprinas e diferenciar das alteradas por leite bovino. Os modelos foram avaliados por parâmetros estatísticos e comparados aos modelos de referência, mostrando-se robustos para predição das adulterações por leite bovino.

Os modelos desenvolvidos com uso de GA-LDA e ACO-LDA conseguiram prever em 100% de acurácia, sensibilidade e especificidade, sem falsos positivos e/ou falsos negativos. Os resultados obtidos foram avanço na interpretação dos dados, complementando os estudos de Pereira et al. (2020) com uma perspectiva meta-heurística de seleção de variáveis.

Projeções futuras podem ser exploradas com usos de outros algoritmos de seleção de variáveis associada tanto a LDA quanto ao uso de outros métodos de classificação. A exploração das variáveis que possuem maior frequência e informações espectrais em novos modelos a serem desenvolvidos.

REFERÊNCIAS

- ALLEGRINI, F.; OLIVIERI, A. C. A new and efficient variable selection algorithm based on ant colony optimization. Applications to near infrared spectroscopy/partial least-squares analysis. **Analytica Chimica Acta**, v. 699, 2011.
- AL-MBAIDEEN, A.A. Application of moving average filter for quantitative analysis of the NIR spectra. **Journal of Analytical Chemistry**, v. 74, 2019.
- ALMEIDA, V.E.; FERNANDES, D.D.S.; DINIZ, P.H.G.D.; GOMES, A.A.; VERAS, G.; GALVÃO, R. K. H.; ARAUJO, M.C.U. Scores selection via Fisher's discriminant power in PCA-LDA to improve the classification of food data. **Food Chemistry**, v. 363, 2021.
- BEC, K.B.; HUCK, C.W. Advances in near-infrared spectroscopy and related computational methods. **Molecules**, v. 24, n. 23, 4370, 2019.
- BEEBE, K.R.; PELL, R.J.; SEASHOLTZ, M.B. Chemometrics: a practical guide. New York: John Wiley & Sons. 1998.
- BRANDAO, M.P.; NETO, M.G.; ANJOS, V.D.C.D.; BELL, M.J.V. Detection of adulteration of goat milk powder with bovine milk powder by front-face and time resolved fluorescence. **Food Control**, v. 81, p. 168-172, 2017.
- BRO, R.; SMILDE, A.K. Principal component analysis: tutorial review. **Analytical Methods**, v. 6, 2014.
- CARVALHO, L.C.; PEREIRA, F.M.V.; MORAIS, C.L.M.; LIMA, K.M.G.; TEIXEIRA, G.H.A. Assessment of macadamia kernel quality defects by means of near infrared spectroscopy (NIRS) and nuclear magnetic resonance (NMR). **Food Control**, v. 106, 2019.
- CHEN, H.; TAN, C.; LIN, Z.; WU, T. Detection of melamine adulteration in milk by near-infrared spectroscopy and one class partial least squares. **Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy**, v. 173, 2017.
- CLARK, S.; GARCIA, M.B.M.; A 100-year review: Advances in goat milk research. **Journal of Dairy Science**, v. 100, 2017.

COSTA, R.G.; MONTE, H.L.B.D.; FILHO, E.C.P.; JUNIOR, E.V.H.; CRUZ, G.R.B da; MENEZES, M.P.C. Typology and characterization of goat milk production systems in Cariris Paraibanos. **Revista Brasileira de Zootecnia**, v. 39, 2010.

FAOSTAT. Food and Agriculture Organization of the United Nations Statistics Division. 2021. Disponível em: <<http://faostat3.fao.org/browse/Q/QL/E>>. Acesso em: 19 de maio de 2021.

FERREIRA, M.M.C. Quimiometria: conceitos, métodos e aplicações. Campinas: Editora Unicamp. 2016.

GOMES, A.A.; AZCARATE, S.M.; DINIZ, P.H.G.D.; FERNANDES, D.D.S.; VERAS, G. Variable selection in the chemometric treatment of food data: A tutorial review. **Food Chemistry**, v. 370, 2022.

HE, Y.; MA, Z. Optimal design of linear sensor networks for plants: A multi-objective ant colony optimization approach. *Chemometrics and Intelligent Laboratory Systems*, v. 135, 2014.

HUANG. J.; ROMERO-TORRES, S.; MOSHGBAR, M. Practical considerations in data pré-processing for NIR and Raman spectroscopy. **American Pharmaceutical Review**, v. 13, 2010.

JUNIOR, I.J.D.; SIQUEIRA, K.B.; STOCK, L.A. Produção, composição e processamento de leite de cabra no Brasil. Juiz de Fora: EMBRAPA. 2020.

KARUNATHILAKA, S.R.; YAKES, B.J.; H.E., K.; CHUNG, J.K.; MASSOBA, M. Non-targeted NIR spectroscopy and SIMCA classification for commercial milk powder authentication: A study using eleven potential adulterants. **Heliyon**, v. 4, 2018.

KATOCH, S.; CHAUHAN, S.S.; KUMAR, V. A review on genetic algorithm: past, present, and future. **Multimedia Tools and Applications**, v. 80, 2020.

KAUFMANN, K.C.; SAMPAIO, K.A.; GARCIA-MARTÍN, J.F.; BARBIN, D.F. Identification of coriander oil adulteration using portable NIR spectrometer. **Food Control**, v. 132, 2020.

LAD, S.S; APARNATHI, K.D.; MEHTA, B.; VELPULA, S. Goat milk in human nutrition and health. **International Journal of Current Microbiology and Applied Sciences**, v. 6, 2017.

LEITE, D.C.; CORRÊA, A.A.P.; JÚNIOR, LC.C.; LIMA, K.M.G.; MORAIS, C.L.M.; VIANNA, V.F.; TEIXEIRA, G.H.A.; MAURO, A.O.D.; UNÊDA-TREVISOLI, S.H. Non-destructive genotypes classification and oil content prediction using near-infrared spectroscopy and chemometric tools in soybean breeding program. **Journal of Food Composition and Analysis**, v. 91, 2020.

LOPES, M.I.; CALLAO, M.P.; RUISANCHEZ, I. A tutorial on the validation of quantitative methods: From the univariate to the multivariate approach. **Analytica Chimica Acta**, v. 891, 2015.

MABOOD, F.; JABEEN, F.; AHMED, M.; HUSSAIN, J.; AL MASHAYKHI, S. A.A.A.; AL RUBAIEY, Z. M. A; FAROOQ, S.; BOQUÉ, R.; ALI, L.; HUSSAIN, Z.; AL-HARASSI, A.; KHAN, A. L.; NAUREEN, Z.; IDREES, M.; MANZOOR, S. Development of new NIR-spectroscopy method combined with multivariate analysis for detection of adulteration in camel milk with goat milk. **Food Chemistry**, v. 221, 2017a.

MABOOD, F.; JABEEN, F.; HUSSAIN, J.; AL-HARASSI, A.; HAMAED, A.; A-MASHAYKHI, S. A. A. A.; AL RUBAIEY, Z. M. A; MANZOOR, S.; KHAN, A.; HAQ, I. Q. M.; GILANI, S. A.; KHAN, A. FT-NIRS coupled with chemometric methods as a rapid alternative tool for the detection & quantification of cow milk adulteration in camel milk samples. **Vibrational Spectroscopy**, v. 92, 2017b.

MAPA – MINISTÉRIO DA AGRICULTURA PECUÁRIA E ABASTECIMENTO. **Instrução Normativa Nº 37, de 31 de outubro de 2000**. Aprovar o Regulamento Técnico de Identidade e Qualidade de Leite de Cabra, conforme consta do Anexo desta Instrução Normativa.

MAPA – MINISTÉRIO DA AGRICULTURA PECUÁRIA E ABASTECIMENTO. **Instrução Normativa Nº 76, de 26 de novembro de 2018**. Aprovar os Regulamentos Técnicos que fixam a identidade e as características de qualidade que devem apresentar o leite cru refrigerado, o leite pasteurizado, o leite pasteurizado tipo A, na forma desta Instrução Normativa e do Anexo único.

MILLER, B.A.; LU, C.D. Current status of global dairy goat production: an overview. **Asian-Australian Journal of Animal Sciences**, v. 32, 2019.

MUEHLHOFF, E.; BENNET, A.; MCMAHON, D. Milk and dairy products in human nutrition. Rome: FAOSTAT. 2013.

NIAZI, A.; LEARDI, R. Genetic algorithms in chemometrics. *Journal of Chemometrics*, v. 26, 2012.

PARK, Y.W. Hypo-allergenic and therapeutic significance of goat milk. **Small Ruminant Research**, v. 14, 1993.

PARK, Y.W.; HAENLEIN, G.F.W.; WENDORFF, W.L. Handbook of milk of non-bovine mammals. New York: John Wiley & Sons, 2017.

PASQUINI, C. Near infrared spectroscopy: A mature analytical technique with new perspectives – A review. **Analytica Chimica Acta**, v. 1026, 2018.

PEREIRA, E.V.S. Novas estratégias analíticas baseadas em espectroscopia no infravermelho próximo e imagens digitais para identificação e quantificação de adulterações em leite caprino. João Pessoa, Programa de Pós-Graduação em Ciência e Tecnologia de Alimentos, UFPB. (2020). p. 156: il (Tese de Doutorado).

PEREIRA, E.V.S.; FERNANDES, D.D.S.; ARAÚJO, M.C.U.; DINIZ, P.H.G.D.; MACIEL, M.I.S. Simultaneous determination of goat milk adulteration with cow milk and their fat and protein contents using NIR spectroscopy and PLS algorithms. **LWT – Food Science and Technology**, v. 127, 2020.

PONTES, M.J.C. Algoritmo das projeções sucessivas para a seleção de variáveis espectrais em problemas de classificação. João Pessoa, Programa de Pós-Graduação em Química, UFPB. (2009). p. 123: il (Tese de Doutorado).

SAVITSZKY, A.; GOLAY, M. J. E. Smoothing and differentiation of data by simplified least squares procedures. **Analytical Chemistry**, v. 36, 1964.

SIQUEIRA, K. B. O mercado consumidor de leite e derivados. Juiz de Fora: EMBRAPA. 2019.

TEIXEIRA, J.L.D.P.; CARAMÊS, E.T.D.S; BAPTISTA, D.P.; GIGANTE, M.L. PALLONE, J.A.L. Vibrational spectroscopy and chemometrics tools for

authenticity and improvement the safety control in goat milk. **Food Control**, v. 112, 2020.

XIAOBO, Z.; JIEWEN, Z.; POVEY, M.J.W.; HOLMES, M.; HANPIN, M. Variable selection methods in near-infrared spectroscopy. **Analytica Chimica Acta**, v. 667, 2010.

WU, T.; CHEN, H.; LIN, Z.; TAN, C. Identification and quantitation of melamine in milk by near-infrared spectroscopy and chemometrics. **Journal of Spectroscopy**, v. 16, 2016.

ZAREEF, M.; CHEN, Q.; HASSAN, M.M.; ARSLAN, M.; HASHIM, M.M.; AHMAD, W.; KUTSANEDZIE, F.Y.H.; AGYEKUM, A.A. An Overview on the Application of typical Non-linear Algorithms Coupled with NIR Spectroscopy in Food Analysis. **Food Engineering Reviews**, v. 12, 2020.