

**UNIVERSIDADE FEDERAL DE MINAS GERAIS**  
**Escola de Veterinária**  
**Programa de Pós-Graduação em Ciência Animal**

Juliana Soares Lima

**ESPECTROFOTOMETRIA FTIR (*Fourier Transform Infrared*) E TÉCNICAS DE  
APRENDIZADO DE MÁQUINA PARA A DETECÇÃO DE FRAUDE POR ADIÇÃO  
DE SORO DE QUEIJO AO LEITE CRU**

Belo Horizonte  
2021

Juliana Soares Lima

**ESPECTROFOTOMETRIA FTIR (*Fourier Transform Infrared*) E TÉCNICAS DE APRENDIZADO DE MÁQUINA PARA A DETECÇÃO DE FRAUDE POR ADIÇÃO DE SORO DE QUEIJO AO LEITE CRU**

Tese apresentada ao Programa de Pós-Graduação em Ciência Animal da Escola de Veterinária da Universidade Federal de Minas Gerais como requisito parcial para à obtenção do título de Doutor em Ciência Animal.

**Área de concentração:** Tecnologia e Inspeção de Produtos de Origem Animal.

**Orientador:** Leorges Moraes da Fonseca

**Coorientadores:** Mônica de Oliveira Leite e Habib Asseiss Neto

Belo Horizonte  
2021

L732e

Lima, Juliana Soares, 1982-

Espectrofotometria FTIR (Fourier Transform Infrared) e técnicas de aprendizado de máquina para a detecção de fraude por adição de soro de queijo ao leite cru/ Juliana Soares Lima. -2021.

123 f.:il

Orientador: Leorges Moraes da Fonseca

Coorientadores: Mônica de Oliveira Leite

Habib Asseiss Neto

Tese (Doutorado) apresentada ao Programa de Pós-Graduação em Ciência Animal, da Escola de Veterinária da Universidade Federal de Minas Gerais, como requisito para obtenção do Título de Doutora em Ciência Animal.

Área de concentração: Tecnologia e Inspeção de Produtos de Origem Animal.

Bibliografias: f. 110 a 121.

1. Leite – Análise - Teses - 2. Leite – Qualidade - Teses - 3. Laticínios – Teses – 4. Soro de queijo – Teses - I. Fonseca, Leorges Moraes da – II. Leite, Mônica de Oliveira - III. Neto, Habib Asseiss – IV. Universidade Federal de Minas Gerais, Escola de Veterinária – V. Título.

CDD – 637

Bibliotecária responsável Cristiane Patricia Gomes – CRB2569

Biblioteca da Escola de Veterinária, Universidade Federal de Minas Gerais



UNIVERSIDADE FEDERAL DE MINAS GERAIS  
**ESCOLA DE VETERINÁRIA**  
COLEGIADO DO PROGRAMA DE PÓS-GRADUAÇÃO EM CIÊNCIA ANIMAL

**FOLHA DE APROVAÇÃO**

**JULIANA SOARES LIMA**

Tese submetida à banca examinadora designada pelo Colegiado do Programa de Pós-Graduação em CIÊNCIA ANIMAL, como requisito para obtenção do grau de DOUTOR em CIÊNCIA ANIMAL, área de concentração em Tecnologia e Inspeção de Produtos de Origem Animal.

Aprovada em 31 de maio de 2021, pela banca constituída pelos membros:

Dr.(a). Leorges Moraes da Fonseca  
Presidente – Orientador(a)  
Dr.(a). Débora Cristina Sampaio de Assis  
Dr.(a). Sérgio Vale Aguiar Campos  
Dr.(a). Cláudia Freire de Andrade Morais Penna  
Dr.(a). Clarice Gebara Muraro Serrate Cordeiro  
Dr.(a). Marco Antônio Sloboda Cortez

---

Documento assinado eletronicamente por **Leorges Moraes da Fonseca, Professor do Magistério Superior**, em 31/05/2021, às 14:15, conforme horário oficial de Brasília, com fundamento no art. 5º do Decreto nº 10.543, de 13 de novembro de 2020.

---

Documento assinado eletronicamente por **Marco Antonio Sloboda Cortez, Usuário Externo**, em 31/05/2021, às 16:03, conforme horário oficial de Brasília, com fundamento no art. 5º do Decreto nº 10.543, de 13 de novembro de 2020.

---

Documento assinado eletronicamente por **Sergio Vale Aguiar Campos, Professor do Magistério Superior**, em 31/05/2021, às 16:58, conforme horário oficial de Brasília, com fundamento no art. 5º do Decreto nº 10.543, de 13 de novembro de 2020.

---

Documento assinado eletronicamente por **Claudia Freire de Andrade Morais Penna, Coordenador(a) de curso**, em 31/05/2021, às 17:36, conforme horário oficial de Brasília, com fundamento no art. 5º do Decreto nº 10.543, de 13 de novembro de 2020.

---

Documento assinado eletronicamente por **Clarice Gebara Muraro Serrate Cordeiro, Usuário Externo**, em 01/06/2021, às 18:05, conforme horário oficial de Brasília, com fundamento no art. 5º do Decreto nº 10.543, de 13 de novembro de 2020.

---

Documento assinado eletronicamente por **Debora Cristina Sampaio de Assis, Professora do Magistério Superior**, em 07/06/2021, às 19:32, conforme horário oficial de Brasília, com fundamento no art. 5º do Decreto nº 10.543, de 13 de novembro de 2020.



A autenticidade deste documento pode ser conferida no site  
[https://sei.ufmg.br/sei/controlador\\_externo.php?cao=documento\\_conferir&id\\_orgao\\_acesso\\_externo=0](https://sei.ufmg.br/sei/controlador_externo.php?cao=documento_conferir&id_orgao_acesso_externo=0), informando o código verificador **0743321** eo código CRC **F4FBADBC**.

## **Agradecimentos**

Obrigada Senhor, por mais essa vitória! Deus é bom o tempo todo, o tempo todo Deus é bom!

Agradeço primeiramente a minha família, especialmente minha mãe e meu pai, que acreditaram em mim e sempre me apoiaram nas minhas decisões. Ao meu marido Davi, que esteve ao meu lado por todos esses dias.

Agradeço ao meu orientador Leorges, e aos coorientadores Mônica e Habib pela paciência e boa vontade em me ajudar a fazer um trabalho relevante.

Deixo também um agradecimento a todos os professores do curso que de alguma forma contribuíram para a minha formação.

Agradeço a todos os funcionários do Laboratório de Análise da Qualidade do Leite, em especial à Marcinha, pela parceria durante todo o curso.

Quero agradecer a minha amiga Daniela Solo com quem desde o início deste trabalho pude discutir os problemas e pensar soluções, sem você nada disso teria acontecido.

Um agradecimento especial a todos os amigos que conheci durante estes anos de estudos e os amigos de velhas datas, que sempre me serviram de apoio nas horas que mais precisei.

Ao CNPq pelo suporte financeiro (167801/2017-1), assim como ao FINEP, MAPA, FAPEMIG(APQ-02740-17) e FEPE-UFMG (Proj. 882).

Por fim, um agradecimento especial ao meu filho Gael, meu maior incentivador.

Dedico esse projeto ao meu pai, que compartilhou da alegria da defesa, mas não está mais aqui para ver os frutos desse trabalho. De onde é que você esteja, meu muito obrigada por tudo. Valeu a pena tanto esforço!

## Resumo

A realização de fraudes em leite causa prejuízos econômicos e riscos à saúde dos consumidores. A adição de soro de queijo ao leite está entre as principais fraudes, cuja finalidade é aumentar o volume fornecido com custo reduzido. As provas para detecção de fraudes são muito laboriosas, caras e demoradas, dificultando a realização na frequência determinada pela legislação. A técnica de Espectroscopia no Infravermelho (FTIR) apresenta-se como uma alternativa promissora para identificação desse tipo de fraude, principalmente porque um número elevado de dados analíticos de leite podem ser obtidos e avaliados nos Laboratórios da Rede Brasileira de Laboratórios de Análise da Qualidade do Leite (RBQL), credenciados pelo Ministério da Agricultura, Pecuária e Abastecimento, a exemplo do Laboratório de Análise da Qualidade do Leite da Escola de Veterinária da Universidade Federal de Minas Gerais (LabUFMG). Entre os métodos utilizados para Aprendizado de Máquina têm-se as Redes Neurais Artificiais (RNA) e as Árvores de Decisão, que apresentam grande potencial de uso como ferramentas de avaliação de possíveis adulterações de leite com soro de queijo. O objetivo desse trabalho foi utilizar metodologias computacionais capazes de identificar adulteração no leite por adição de soro de queijo em amostras analisadas por espectroscopia FTIR. Foram utilizadas 585 amostras de leite cru diluídas com soro obtido da fabricação do queijo minas frescal nas concentrações: 0%, 1%, 2%, 5%, 10%, 15%, 20%, 25% e 30%. As amostras foram armazenadas a 7°C, 20°C e 30°C durante 0, 24, 48, 72 e 168 horas e analisadas em aparelho FTIR do LabUFMG. Adicionalmente, 585 amostras de leite cru autêntico, coletadas de tanques de refrigeração, foram usadas juntamente com as amostras anteriores, para treinamento, teste e validação dos modelos de “machine learning”. Diferentes classificadores foram utilizados para treinar modelos capazes de reconhecer as alterações provocadas pela adição de soro de queijo ao leite nas características de composição normal do leite. Os registros de composição e os espectros resultantes das análises do equipamento FTIR foram organizados de forma a servir como entrada para a tarefa de classificação em modelos de RNA e árvores de decisão. A classificação foi realizada usando dois tipos de dados, para três tipos de modelos: os componentes numéricos foram analisados por RNA e Árvore de Decisão e o espectro total do infravermelho foi analisado por meio de Rede Neural Convolutiva (CNN). A metodologia proposta foi capaz de gerar acurácias médias de 97,8%, quando se utilizam dados de composição na RNA, e de 96,2% nos testes por Árvore de decisão. Utilizando os dados espectrais (CNN) a taxa de acurácia total do modelo foi de 93,84%. O uso do FTIR associado a Redes Neurais Artificiais foi eficiente para diferenciar amostras com leite e adicionadas de soro de queijo.

Palavras-chave: Fraude. Soro de queijo. FTIR. Aprendizagem de máquina. Redes Neurais.

## *Abstract*

Fraud in milk causes economic damage and potential health risks to consumers. Cheese whey addition to milk is one of the major frauds, done to increase the volume provided at a reduced cost. Methods for fraud detection usually are laborious, expensive and time-consuming. The Infrared Spectroscopy (FTIR) technique presents itself as a promising alternative for the identification of this type of fraud, mainly because a large number of analytical milk data can be obtained and evaluated in the Laboratories of the Brazilian Network of Milk Quality Analysis Laboratories (RBQL), accredited by the Ministry of Agriculture, such as the Milk Quality Analysis Laboratory of the Veterinary School in the “Universidade Federal de Minas Gerais” (LabUFMG). Among the methods used for machine learning are the Artificial Neural Networks (ANN) and the Decision Trees, which have great potential for use as tools for evaluating possible adulterations of milk with cheese whey. The objective of this work was to use computational methodologies capable of identifying adulteration in milk by adding whey cheese to samples analyzed by FTIR spectroscopy. 585 samples of raw milk diluted with whey obtained from the manufacture of Fresh Minas cheese were used in concentrations: 0%, 1%, 2%, 5%, 10%, 15%, 20%, 25% and 30%. The samples were stored at 7°C, 20°C and 30°C for 0, 24, 48, 72 and 168 hours and analyzed in an FTIR device from LabUFMG. Compositional results of 585 samples of authentic bulk tank raw milk were added to the experimental samples and utilized for the “machine learning” models training, test, and validation. Different classifiers were used to recognize the compositional and or spectral changes caused by the addition of cheese whey to milk. The composition records and the spectra resulting from the analysis of the FTIR equipment were organized in order to serve as an entry for the classification task in the ANN models and Decision Trees. The classification was performed using two types of data, for three types of models: the numerical components were analyzed by ANN and Decision Tree and the total infrared spectrum was analyzed by Convolutional Neural Network (CNN). The proposed methodology was able to generate average accuracy of 97,8%, when using composition data in the ANN, and 96,2% using Decision Tree. Using the spectrum data (CNN), the total accuracy rate of the model was 93,84%. The use of FTIR associated with Artificial Neural Networks was efficient to differentiate samples with milk and added with cheese whey.

Keywords: Fraud. Cheese whey. FTIR. Machine learning. Neural Networks.

## LISTA DE FIGURAS

<b>Figura 2.1</b>	Local da ação da quimosina na k-caseína.....	26
<b>Figura 2.2</b>	Uma parte do espectro eletromagnético que mostra a relação do infravermelho vibracional com outros tipos de radiação.....	35
<b>Figura 2.3</b>	Esquema ilustrativo para o interferômetro de Michelson e do espectro resultante da aplicação da transformada de Fourier.....	37
<b>Figura 2.4</b>	Espectros de amostras de leite obtidos por FTIR.....	38
<b>Figura 2.5</b>	Etapas do processo de descoberta de conhecimentos em banco de dados (KDD).....	43
<b>Figura 2.6</b>	Hierarquia do aprendizado de acordo com o grau de supervisão.....	45
<b>Figura 2.7</b>	Exemplo de um classificador utilizando árvore de decisão.....	48
<b>Figura 2.8</b>	Exemplo de um caminho de classificação.....	48
<b>Figura 2.9</b>	Representação simplificada de uma rede neural artificial mostrando camada de entrada, oculta e saída.....	53
<b>Figura 2.10</b>	Estrutura da RNA usada na detecção de adição de soro de queijo no leite.....	53
<b>Figura 3.1</b>	Esquema da preparação das amostras com diluições (0%, 1%, 2%, 5%, 10%, 15%, 20%, 25% e 30%), armazenadas a 7, 20 e 30°C nos tempos 0h, 24h, 48h, 72h e 168h.....	61
<b>Figura 3.2</b>	Esquema da preparação das amostras para cromatografia com diluições (0%, 1%, 2% e 5%), armazenadas a 7, 20 e 30°C nos tempos 0h e 168h.....	62
<b>Figura 3.3</b>	Componentes numéricos gerados pelo equipamento FTIR em formato CVS de amostras aleatórias.....	66
<b>Figura 3.4</b>	Médias das concentrações dos componentes do soro em relação a concentração de soro de queijo adicionado ao leite.....	68
<b>Figura 3.5</b>	Teores (g/100g) de gordura, proteína, lactose e sólidos não gordurosos após adição de soro obtido da fabricação de queijo ao leite nas concentrações de 0%, 1%, 2%, 5%, 10%, 15%, 20%, 25% e 30% .....	70
<b>Figura 3.6</b>	Gráfico de dispersão das variáveis de composição do leite versus a concentração de soro.....	72
<b>Figura 3.7</b>	Relação e correlação entre a concentração de soro (0, 1%, 2%, 5%) e o índice de CMP.....	75
<b>Figura 3.8</b>	Relação entre os níveis de CMP, concentração de soro (0, 1%, 2% e 5%) e Contagem Padrão em Placas definida por UFC (log/ml).....	76
<b>Figura 4.1</b>	Esquema da preparação das amostras com diluições (0%, 1%, 2%, 5%, 10%, 15%, 20%, 25% e 30%), armazenadas a 7, 20 e 30°C nos tempos 0h, 24h, 48h, 72h e 168h.....	83
<b>Figura 4.2</b>	Modelo de curvas da adição de soro de queijo ao leite cru, analisados pelo FTIR. ....	88
<b>Figura 4.3</b>	Arquitetura da Rede Neural Convolutiva para classificação binária	

	de espectros infravermelhos puros.....	89
<b>Figura 4.4</b>	Árvore de decisão das variáveis de composição gerada pelo algoritmo de classificação. A visualização do nó está completa e parte é ampliada abaixo. Os nós cinzas apresentaram uma forte influência.....	92
<b>Figura 4.5</b>	Curva ROC do modelo apresentado mostrando a relação entre a taxa de verdadeiros positivos e a taxa de falsos positivos para um intervalo de limiares de probabilidades de classificação. A área sob a curva (AUC) possui o valor de 0,99 no treinamento e 0,98 para o teste.....	93
<b>Figura 4.6</b>	Desempenho do método de árvore de classificação CART® para detectar soro de queijo adicionado ao leite cru com base em dados de composição obtidos com o método do infravermelho com transformada de Fourier (FTIR).....	95
<b>Figura 4.7</b>	Resultados de acurácia, precisão, sensibilidade e especificidade para diferentes tratamentos de Redes Neurais Artificiais, em modelo binário para identificação de soro de queijo em leite .....	98
<b>Figura 4.8</b>	Exemplo de aquisição de dados convertidos em arquivo formato .csv com todos os componentes numéricos obtidos de amostras aleatórias..	100
<b>Figura 4.9</b>	Alterações provocadas no espectro de amostra com soro (5%), em relação ao espectro do leite puro.....	101
<b>Figura 4.10</b>	Gráfico da precisão e da perda do modelo da CNN nas etapas de treinamento e validação. O modelo foi treinado por 300 épocas. (a) Precisão do treinamento e validação. (b) Perda de treinamento e validação.....	102
<b>Figura 4.11</b>	Saliências calculadas pela CNN com o espectro de uma amostra com soro de queijo. Regiões mais escuras compreendem os pontos onde a rede considerou mais importante.....	103
<b>Figura 4.12</b>	Técnica de validação cruzada <i>k-fold</i> , com treinamento e teste com $k=10$ .....	104
<b>Figura 4.13</b>	Medidas de desempenho da RNA calculadas a partir da matriz de confusão.....	105
<b>Figura 4.14</b>	Curva ROC do modelo apresentado mostrando a relação entre a taxa de verdadeiros positivos e a taxa de falsos positivos para um intervalo de limiares de probabilidades de classificação. A área sob a curva (AUC) possui o valor de 0,988.....	106

## LISTA DE QUADROS

<b>Quadro 2.1</b>	Principais informações dos tipos de neurônios.....	52
<b>Quadro 2.2</b>	Matriz de confusão para um problema genérico de classificação de duas classes (ocorrência e não ocorrência de um evento).....	56
<b>Quadro 3.1</b>	Requisitos físico-químicos para o leite cru refrigerado de acordo com a Instrução Normativa n° 76 de 2018.....	66

## LISTA DE TABELAS

<b>Tabela 2.1</b>	Métodos relatados na literatura para detecção de adição de soro de queijo ao leite.....	32
<b>Tabela 3.1</b>	Estatística descritiva a partir dos dados de amostras analisadas no LabUFMG com adição de soro (n = 585).....	67
<b>Tabela 3.2</b>	Valores médios dos componentes do leite em relação a concentração de soro de queijo quando aplicado o Teste de Student-Newman-Keuls (Valor de $\alpha = 0,05$ ).....	69
<b>Tabela 3.3</b>	Resultados médios de CMP, obtidos por CLAE em amostras de leite adicionadas de soro com diferença de concentração, tempo e temperatura.....	74
<b>Tabela 3.4</b>	Valores médios da Contagem Bacteriana e do Índice de CMP em relação a concentração de soro de queijo (0, 1%, 2%, 5%) quando aplicado o Teste de Tukey (Valor de $\alpha = 0,05$ ).....	75
<b>Tabela 3.5</b>	Valores médios da Contagem Bacteriana e do Índice de CMP em relação a temperatura de incubação (7°C, 20°C e 30°C) quando aplicado o Teste de Tukey (Valor de $\alpha = 0,05$ ).....	76
<b>Tabela 4.1</b>	Matriz de confusão para uma árvore de classificação CART® com gordura, proteína, caseína, lactose, sólidos totais, sólidos não gordurosos e ponto de congelamento como recursos de entrada e saída binomial para leite cru e leite cru adicionado com soro de queijo (2%, 5%, 10%, 15%, 20%, 25% e 30%).....	94
<b>Tabela 4.2</b>	Características das amostras usadas para configurar a RNA.....	97
<b>Tabela 4.3</b>	Matriz de confusão para uma Rede Perceptron Multicamadas com gordura, proteína, caseína, lactose, sólidos totais, sólidos não gordurosos e ponto de congelamento como recursos de entrada e saída binomial para leite cru e leite cru adicionado com soro de queijo (2%, 5%, 10%, 15%, 20%, 25% e 30%).....	97

## SUMÁRIO

<b>CAPÍTULO 1 - INTRODUÇÃO .....</b>	<b>14</b>
1.1 Objetivo.....	17
<b>CAPÍTULO 2 - REFERENCIAL TEÓRICO .....</b>	<b>18</b>
2.1 Produção de queijo.....	18
2.2 Fraude em Leite por adição de soro .....	21
2.3 Ocorrências de fraude por soro de queijo .....	22
2.4 Métodos para detecção de fraude por adição de soro em leite.....	25
2.4.1 Cromatografia Líquida de Alta eficiência (CLAE) .....	25
2.4.2 Técnica da ninhidrina ácida (determinação do ácido siálico) .....	29
2.4.3 Eletroforese .....	31
2.5 Fundamentos da Espectroscopia no Infravermelho.....	33
2.5.1 Espectroscopia FTIR para análise de leite.....	39
2.6 Aprendizado de máquina.....	42
2.6.1 Técnicas de Aprendizado de máquina.....	44
2.6.1.1 Técnica de Árvores de decisão.....	47
2.6.1.2 Redes Neurais Artificiais (RNA).....	49
<b>CAPÍTULO 3 - ESPECTROFOTOMETRIA POR INFRAVERMELHO FTIR E CROMATOLOGRAFIA LIQUIDA DE ALTA EFICIÊNCIA (HPLC) NA IDENTIFICAÇÃO DE FRAUDES POR ADIÇÃO DE SORO DE QUEIJO AO LEITE.</b>	
3.1 Introdução.....	58
3.2 Material e Métodos.....	59
3.3 Resultados e Discussão.....	65
<b>CAPÍTULO 4- APRENDIZADO DE MÁQUINA PARA DETECÇÃO DE FRAUDE POR ADIÇÃO DE SORO OBTIDO DA FABRICAÇÃO DE QUEIJO AO LEITE.</b>	
4.1 Introdução.....	80
4.2 Material e Métodos.....	82

4.3 Resultados e Discussão.....	91
<b>CONCLUSÕES.....</b>	<b>109</b>
<b>CONSIDERAÇÕES FINAIS.....</b>	<b>109</b>
<b>REFERÊNCIA BIBIOGRÁFICA.....</b>	<b>110</b>
<b>COLABORAÇÕES INTERINSTITUCIONAIS PARA ELABORAÇÃO DO PROJETO.....</b>	<b>122</b>
<b>CONTRIBUIÇÕES ACADÊMICAS.....</b>	<b>123</b>

## CAPÍTULO 1

### INTRODUÇÃO

A demanda nacional por grandes volumes de leite de boa qualidade faz com que a pecuária leiteira seja bastante competitiva. Assim, é necessário quantificar e conhecer os parâmetros que podem interferir nos resultados obtidos na análise da qualidade do leite, diagnosticando os pontos que devem ser corrigidos, a fim de gerar ganhos efetivos na atividade. Existe na indústria láctea o pagamento diferenciado para produtores que forneçam leite com teores mais elevados de gordura e proteína, além de outros parâmetros de qualidade considerados essenciais. Destaca-se também a importância da análise do leite individual para monitoramento da qualidade e identificação de problemas críticos (Oliveira, 2011).

A qualidade do leite consumido no país é uma constante preocupação dos técnicos e autoridades ligadas à área de saúde e laticínios. Um dos problemas encontrados no leite é a ocorrência de diversas fraudes que causam prejuízos econômicos e riscos à saúde dos consumidores (Robim *et al.*, 2012).

Nas indústrias de laticínios, os principais prejuízos com as fraudes são a redução do rendimento de alguns produtos lácteos, a diminuição do valor nutricional, a alteração da qualidade dos produtos beneficiados e o risco aos consumidores em virtude da presença de substâncias que podem ser prejudiciais à saúde, tais como agente antimicrobianos, reconstituintes de densidade, neutralizante de acidez, entre outras. No Brasil, a adição de qualquer substância estranha à composição do leite cru constitui fraude (Brasil, 2017).

A prática de fraudes na área de laticínios é um problema recorrente. Dentre as principais fraudes destaca-se a adição de soro de queijo ao leite, cuja finalidade é aumentar o volume fornecido com custo reduzido (Aquino, 2013).

A adição de soro ao leite pode ser detectada e quantificada pela determinação do caseinomacropéptido (CMP), em que se tolera um índice de CMP de até 30mg/L para

o leite ser destinado ao abastecimento direto (Brasil, 2006). A cromatografia líquida de alta eficiência (CLAE), análise oficial utilizada para detecção de CMP, e a eletroforese capilar são ferramentas úteis e precisas para assegurar a qualidade do leite, auxiliando na detecção da fraude por soro de queijo (Cortez *et al.*, 2010). Porém, o método oficial tem como desvantagens o custo elevado das análises, a necessidade de realizar um procedimento de extração/preparo de amostras, o tempo elevado de realização das análises e a ocorrência de falso positivos pela ação de bactérias psicrotóxicas.

Já que o soro é um derivado lácteo, dependendo da quantidade adicionada e das manobras utilizadas para mascarar este tipo de fraude, os procedimentos analíticos oficiais de rotina realizados na análise de leite não têm sensibilidade para identificar a sua adição, levando, muitas vezes a avaliação equivocada da qualidade do leite que será destinado ao consumo (Fuente e Juarez, 2005).

O atual nível de desenvolvimento da pecuária leiteira no Brasil demanda a existência de recursos laboratoriais que proporcionem informações de qualidade em tempo adequado para o contínuo crescimento do setor. Nesse contexto, o Ministério da Agricultura, Pecuária e Abastecimento, criou em 2002, pela Instrução Normativa nº37/2002, a Rede Brasileira de Laboratórios de Controle da Qualidade do Leite (RBQL), que conta com equipamentos eletrônicos modernos, como o Espectrofotômetro do Infravermelho (FTIR), para análises rápidas e precisas de parâmetros químicos, com alta capacidade analítica.

Como a indústria necessita de respostas rápidas, técnicas que exigem menor manipulação, preparo, custo e tempo de análise reduzidos, além de bom grau de precisão têm sido objeto de estudos nos últimos anos. Nesse sentido, a técnica de FTIR apresenta-se como uma alternativa promissora para identificação de todo tipo de fraude, incluindo a adição de soro de queijo ao leite.

As análises de leite realizadas por instrumentos que utilizam FTIR produzem muitos dados, referentes aos resultados de análises realizadas diariamente e que poderiam ser adicionalmente utilizados em técnicas como aprendizagem de máquinas, uma área da inteligência artificial.

A aprendizagem de máquinas é incrivelmente poderosa para fazer previsões ou sugestões calculadas a partir de grandes quantidades de dados. Alguns dos exemplos mais comuns de aprendizagem de máquinas são os algoritmos da Netflix para fazer sugestões de filmes aos seus usuários com base em filmes já assistidos no passado ou algoritmos da Amazon que recomendam livros com base em compras anteriores. Um algoritmo pode aprender a detectar padrões a partir de um grande volume de dados. A operação é realizada com o mínimo de interferência humana.

Classificação é um método reconhecido para tomar decisões em novas situações. As técnicas de classificação são capazes de processar uma grande quantidade de dados e podem ser usadas para classificá-los com base no conjunto de treinamento e nas categorias das classes. O termo pode abranger qualquer contexto em que alguma decisão ou previsão seja feita com base em informações disponíveis (Nikam, 2015). Entre os diversos algoritmos empregados para se realizar a classificação, estão as Redes Neurais Artificiais e as Árvores de Decisão.

Uma rede neural artificial (RNA) é um modelo computacional que consiste em processamento de elementos simples (neurônios artificiais) que aplicam uma determinada função matemática para os dados, gerando uma única resposta. Eles são organizados em camadas e estão vinculados juntos. Um processo chamado treinamento ou aprendizagem realiza o ajuste desses dados. Ele também é responsável por extrair as características dos dados e armazenar o conhecimento das redes (Alves da Rocha *et al.*, 2015).

Outro meio eficiente para representar classificadores a partir de dados, é o uso de árvores de decisão. Esta técnica tem como pontos fortes a sua eficiência em termos de tempo de processamento e o fornecimento de um meio intuitivo de analisar os resultados, por apresentar como estrutura final do classificador uma forma de representação simbólica simples e normalmente bastante compreensível, o que facilita a compreensão do problema em análise (Garcia, 2003).

Diante do exposto, este trabalho visa utilizar o banco de dados criado por resultados obtidos da espectroscopia no infravermelho (FTIR) e treinar modelos com algoritmos de aprendizado de máquina e classificação. É um trabalho inédito, pois utiliza métodos de aprendizado de máquina em conjunto com espectroscopia de infravermelho para identificar adulteração do leite pela adição de soro de queijo com alta precisão e rapidez.

### **1.1 Objetivos**

- Utilizar metodologias computacionais capazes de identificar adulteração no leite por adição de soro de queijo em amostras analisadas por espectroscopia FTIR;
- Utilizar um banco de dados criado por resultados obtidos da espectroscopia FTIR a partir de amostras fraudadas com soro de queijo em diferentes concentrações e treinar modelos de aprendizado de máquina;
- Propor uma arquitetura de rede neural convolucional capaz de reconhecer e extrair informações a partir das coordenadas espectrais obtidas por FTIR, sem necessidade de pré-processamentos;
- Treinar modelos com algoritmos de classificação para detecção de soro de queijo adicionado ao leite cru utilizando os resultados da análise de composição como dados de entrada;
- Analisar os resultados obtidos na Cromatografia Líquida de alta eficiência para avaliar a presença de CMP nas amostras fraudadas com soro de queijo.

## CAPÍTULO 2 - REFERENCIAL TEÓRICO

### 2.1 A Produção de queijo e o soro

O queijo é um dos alimentos preparados mais antigos que a história da humanidade registra. Os principais componentes do queijo são: água, gordura, proteínas, lactose, ácido láctico, cloreto de sódio e sais diversos (resíduo mineral fixo) presentes em diferentes proporções e com finalidades específicas (Abreu, 2005).

A produção de queijo é basicamente um processo de concentração do leite no qual parte dos componentes sólidos, principalmente proteína e gordura, são aglomerados na coalhada enquanto as proteínas do soro, lactose e sólidos solúveis são removidos no soro (Kosikowski, 1979).

O regulamento Técnico de Identidade e Qualidade do Queijo define o Queijo Minas Frescal como aquele queijo fresco obtido pela coagulação do leite, podendo ser com coalho ou outras enzimas coagulantes, adicionado ou não de bactérias lácticas específicas, leite em pó, creme, sólidos de origem láctea, cloreto de sódio e cloreto de cálcio (Brasil, 1997).

Em 2020, o Ministério da Agricultura Pecuária e Abastecimento (MAPA) aprovou o Regulamento Técnico que fixa os padrões de identidade e qualidade para o soro de leite e o soro de leite ácido, que define o soro como o produto lácteo líquido extraído da coagulação do leite utilizado no processo de fabricação de queijos, caseína alimentar e produtos similares (Brasil, 2020).

O soro de queijo é um produto aquoso verde-amarelado proveniente da fabricação de queijos sendo produzido a partir da coagulação enzimática ou ácida do leite (Tullio, 2007). É a porção aquosa que se separa da massa durante a fabricação convencional de queijos, e retém cerca de 55% dos nutrientes do leite. Aproximadamente de 85 a 90% do volume de leite utilizado na fabricação de queijos resulta em soro (Kosikowski, 1979). É composto por aproximadamente 93% de água e os 7% restantes correspondem

a matéria seca representada pela lactose, proteínas, gordura e sais minerais (Fogaça, 2017).

Apesar de ter grande parte de sua composição de água, contém alto valor nutricional que é atribuído, principalmente, às proteínas que possuem aminoácidos essenciais. Estas proteínas apresentam importantes propriedades funcionais de solubilidade, formação e estabilidade de espuma, emulsibilidade, geleificação, formação de filmes e cápsulas protetoras, além de fazerem parte de um grupo proteico com características estruturais bastante diversificadas (Sgarbieri, 2005). Contém cerca da metade de extrato seco total do leite originalmente utilizado para a sua fabricação, apresentando em média, 6,35 a 6,50% de sólidos totais, sendo cerca de 4,85 a 4,90% de lactose; 0,75 a 0,80% de proteínas (principalmente albumina e globulinas); 0,04 a 0,5% de gordura, além de sais minerais (Kosikowski, 1979).

No leite são encontradas duas classes principais de proteínas: as caseínas e as proteínas do soro. O grupo das caseínas representa cerca de 80% do total das proteínas, enquanto as proteínas do soro, as quais estão incluídas neste grupo a lactalbumina, a lactoglobulina, a lactoferrina e as imunoglobulinas, perfazem os 20% restantes. Desta forma, esta relação entre caseína (80%) e proteínas do soro (20%) deve ser mantida no leite, caso não haja adição de proteínas do soro na sua formulação. Naqueles produtos nos quais esta relação está alterada pode-se dizer que existem fortes indícios de fraude (Sgarbieri, 2005).

A quantidade do soro gerada no processo de fabricação de queijo varia de acordo com o tipo de queijo produzido e as técnicas de fabricação empregadas (Richards, 1997), o que acarreta diversas mudanças nos teores de proteína, gordura e minerais (Aquino, 2013). Em média, para a fabricação de um quilo de queijo necessita-se de dez litros de leite e, dependendo do uso de água na fabricação de acordo com o tipo de queijo fabricado, recuperam-se de nove a doze litros de soro (Richards, 1997).

O Queijo Minas Frescal é um produto de grande interesse para as indústrias de laticínios por ter alto teor de umidade e elevado rendimento, variando entre 5 e 7 litros de leite por

quilo de queijo, resultando em um rápido retorno do investimento, preços mais acessíveis e grande aceitabilidade pelos consumidores (Hoffmann *et al.*, 2002).

São constantes os esforços para aproveitamento de resíduos agroindustriais em todo o mundo. Em particular o soro do queijo, pela abundância de produção, características nutricionais e elevada capacidade poluente, tem sido motivo de vários estudos (Silva e Herman-Gomez, 2000; Valduga *et al.*, 2006; Madrona *et al.*, 2009; Gabardo *et al.*, 2011). Nos países europeus e nos Estados Unidos, o soro é usado na alimentação humana, enquanto no Brasil ele ainda é subutilizado sendo, muitas vezes, destinado para alimentação animal. O pior destino que se dá a este produto é o descarte no meio ambiente (Santos e Ferreira, 2001).

Historicamente o soro de queijo já foi considerado um produto residual de baixo valor econômico. Seu grande volume de obtenção representa um dos maiores problemas para a indústria de laticínios. Em toda a extensão territorial brasileira, a maioria dos pequenos e médios produtores de queijo não consegue dar um destino nobre ao soro de queijo, havendo uma enorme deficiência de tecnologias que viabilizem o aproveitamento de suas proteínas. Por conter elevada quantidade de nutrientes, é considerado um material altamente poluente, com a demanda bioquímica de oxigênio (DBO) de 30.000 a 50.000 mL<sup>-1</sup> (Borges *et al.*, 2001). Contudo, faz-se necessário o estudo de formas para conservá-lo e utilizá-lo, minimizando problemas tanto de ordem ambiental como nutricional, em benefício da fauna, flora, animais domésticos e do próprio homem (Florencio *et al.*, 2013).

O volume de soro de queijo produzido anualmente no mundo é de aproximadamente 180 a 190 milhões de toneladas, sendo que 50% (aproximadamente 90 milhões de toneladas por ano) são descartadas, sem qualquer tipo de reuso. 25% (aproximadamente 45 milhões de toneladas/ano) são utilizadas diretamente na sua forma líquida, para a produção de ricota e de bebidas fermentadas a base de soro. Aproximadamente 30% (ou 13,5 milhões de toneladas/ano) da quantidade produzida no mundo são transformados em soro de queijo em pó, produto que pode ser vendido diretamente ou aproveitado para a produção de fórmulas infantis. O restante é purificado e vendido como lactose e seus derivados ou é utilizado para produção de concentrado das

proteínas do soro, podendo ser comercializados nos suplementos conhecidos como *whey protein* (Mollea; Marmo; Bosco, 2013).

As aplicações do soro são inúmeras, englobando as indústrias de lácteos, carnes, panificação (melhorar o valor nutricional, sabor e cor), chocolate, cereais, aperitivos, bebidas, material combustível, fertilizante e alimentação animal, entre outros. Em virtude do seu alto valor nutricional o soro de queijo se torna um substrato atraente para sua utilização em processos fermentativos para obtenção de vários produtos, como ácido láctico, butanol e etanol, entre outros (Florencio *et al.*, 2013).

## **2.2 Fraude em Leite por adição de soro**

No Brasil, segundo o Art. 501 do Regulamento da Inspeção Industrial e Sanitária de Produtos de Origem Animal/RIISPOA/2017, considera-se impróprio para qualquer tipo de aproveitamento o leite cru quando, na seleção da matéria prima, apresente resíduos de produtos inibidores, de neutralizantes de acidez, de reconstituintes de densidade ou do índice crioscópico, de conservadores, de agentes inibidores do crescimento microbiano ou de outras substâncias estranhas à sua composição (BRASIL, 2017). A legislação brasileira proíbe a adição de soro de queijo ao leite de consumo.

Moore *et al.* (2012) criaram um banco de dados sobre fraudes em alimentos a partir de artigos de revistas científicas publicados em inglês entre 1980 e 2010. Dentre as adulterações relatadas em artigos científicos, a adulteração em leite ocupou a segunda posição, atrás apenas da adulteração em azeite.

O soro de queijo é um produto interessante para quem visa o lucro ilícito, pois pode ser utilizado tanto para aumentar o volume do leite, como para aumentar as proteínas. Além disso, é economicamente atrativa sua adição ao leite fluido pelo fato de apresentar um custo reduzido ou nulo, já que é um produto quase sempre disponível pela alta quantidade gerada da fabricação de queijos, e pelo fato de ter baixo aproveitamento em derivados e subprodutos lácteos, além do custo do tratamento para o seu descarte ser alto (Carvalho *et al.*, 2007).

Leites para consumo têm a finalidade de servir de alimento. Ao serem fraudados, impõem prejuízo econômico ao consumidor e, principalmente, têm seu valor nutritivo limitado e suas intenções de uso desqualificadas. Nas fraudes com substituição de leite por soro, a cada 10% de soro adicionado há uma redução de aproximadamente 8% no teor de proteínas, principalmente das caseínas, e conseqüentemente, de fosfato de cálcio. Ao invés de 3,5%, o leite passa a ter aproximadamente 3,2% de proteínas; com 20% apenas 2,9%; com 30% apenas 2,6% e assim por diante. Essa fraude reveste-se de caráter criminoso ao, deliberadamente, privar níveis nutricionais esperados, necessários e pelos quais o consumidor está remunerando integralmente (Prata, 2002). Caracteriza, ainda, prática abusiva que lesa os direitos do consumidor e que pode, em casos de consumo prolongado, gerar desnutrição, principalmente em lactentes. No entanto, para uma grande diversidade de outros produtos como leites fermentados, confeitos, panificações e as denominadas bebidas lácteas, o soro de queijo é um ingrediente de uso comum (Oliveira, 2009).

Em testes realizados por Cortez *et al.* (2010), julgadores na análise sensorial detectaram diferença apenas com a adição de 45% de soro de queijo no leite. Esses autores avaliaram que a adição de soro ao leite fluido não acarretou modificações marcantes nos valores da crioscopia, mesmo com adição de 30% de soro, o que pode ser considerado uma grande preocupação em relação à detecção desta fraude por métodos mais simples.

### **2.3 Ocorrência de fraude por adição de soro de queijo ao leite**

Montáñez *et al.* (2000) realizaram um trabalho no México para detecção do glicomacropéptido (GMP) para identificação da adição fraudulenta de leite em pó com soro de queijo através da técnica de eletroforese SDS PAGE. Das 108 amostras analisadas encontraram 14,81% de amostras positivas, concluindo que as marcas analisadas não cumpriam os requerimentos legais.

Souza *et al.* (2000) investigaram a adulteração de bebidas lácteas distribuídas no Programa Nacional da Merenda Escolar (Brasil) por meio da análise de eletroforese em gel de poliacrilamida, e concluíram, que o sistema desenvolvido, seguido de análise densitométrica, foi capaz de identificar e quantificar adulteração por adição de soro em

misturas lácteas complexas. Cerca de 49% das bebidas lácteas sabor café e 29% das de sabor chocolate, de um total de 50 amostras analisadas por sabor, apresentaram adulteração. O soro foi adicionado para alcançar o total de proteínas especificado nos rótulos.

Benítez *et al.* (2001) estudaram a presença de soro de queijo em leite em pó por meio da determinação do CMP, por CLAE-FG. Das 13 amostras importadas, analisadas em Cuba, três apresentaram concentrações detectáveis de soro de queijo, havendo adição maior ou igual a 1% (v/v).

Urbán *et al.* (2002) estudaram a presença de CMP em leite fluido, cru e industrializado, vendido na cidade do México, por meio da eletroforese em gel de poliacrilamida. Os autores encontraram 13,8% das amostras de leite cru, 57,1% de leite pasteurizado, e 37,5% das amostras de leite UAT positivas para a presença de CMP em níveis maiores que 2%, o que confirmou a adulteração com soro de queijo.

Em 2007, a Polícia Federal flagrou duas cooperativas de laticínios no estado de Minas Gerais acusadas de adicionar soro de queijo, peróxido de hidrogênio e hidróxido de sódio ao leite longa vida integral, numa operação que ficou conhecida como “Ouro Branco” (Abrantes, 2014).

Ayala *et al.* (2009) detectaram a presença de soro de queijo em amostras de leites UHT (*Ultra High Temperature*) comercializadas na Cidade do México, pela técnica da eletroforese em gel de poliacrilamida (PAGE), CLAE e da UV-4<sup>a</sup>DS (espectro de absorção derivada da quarta). De acordo com os resultados obtidos, os autores concluíram que a porcentagem de amostras positivas com soro de queijo analisadas por eletroforese, CLAE e UV-4<sup>a</sup>DS foi de 70, 50 e 90%, respectivamente. As técnicas utilizadas relacionadas com a detecção e estimativa do CMP por eletroforese ou por CLAE e os resultados obtidos indicaram que a quantificação da relação proteínas de soro/proteína total (PS/PT) por UV-4<sup>a</sup>DS foi um indicador mais sensível a adições de soro de queijo.

Firmino *et al.* (2010) avaliaram a qualidade físico-química e a presença de adulterantes no leite cru da região de Rio Pomba - MG. Das 20 amostras analisadas, 48,3% foram positivas para a presença de formol, nitrato, cloreto, sacarose, urina e soro lácteo. Foi detectado percentual médio estimado de soro lácteo no leite variando de 0,81% a 11,74%. De acordo com os autores, essas supostas concentrações de soro encontradas em muitas amostras analisadas podem estar relacionadas com a ocorrência de mastite e com a ação dos micro-organismos psicrotrofos proteolíticos encontrados no leite desta região de Rio Pomba, MG.

Mendonça *et al.* (2017) avaliaram 15 amostras de leite comercializadas no município de Itabuna/BA, utilizando cinco marcas diferentes de leite em pó, leite UHT e leite pasteurizado, a fim de detectar a fraude por soro, por meio da presença de caseína glicomacropéptido (c-GMP). Foi utilizado um Kit rápido, STICK c-GMP®, que é um teste imunocromatográfico para detecção de c-GMP. Neste estudo foram encontradas três amostras de leite fora da especificação permitida para c-GMP, sendo que duas destas não deveriam estar sendo comercializadas para alimentação humana, devido à alta quantidade de soro presente, o que afeta consequentemente o valor nutricional do leite.

Souza *et al.* (2018) compararam três métodos analíticos para determinação de soro em leite cru refrigerado: CLAE, ninidrina ácida e colorimétrico adaptado. Foram coletadas 100 amostras de leite cru refrigerado de tanques de expansão. O leite foi adulterado com soro de queijo de composição padrão, nas proporções de 5%, 10%, 15% e 20%. As amostras foram submetidas à análise pelo método da ninidrina ácida, apresentando 10 (14,7%) amostras negativas e 58 (85,3%) positivas. O teor médio de ácido siálico encontrado na técnica da ninidrina foi de 5,58µg/mL, com valor mais frequente de 2,70 µg/mL. Das 68 amostras negativas pela cromatografia líquida de alta eficiência, duas foram positivas (2,94%) e 66 (97,06%) negativas, quando analisadas pelo método colorimétrico. A frequência relativa de amostras positivas foi de 32%, com a CLAE apresentando a maior média de soro (14,37%), seguida do método colorimétrico (5,28%) e o da ninidrina ácida (3,12%). A técnica de cromatografia líquida de alta eficiência diferiu dos métodos de ninidrina ácida e colorimétrico, enquanto os métodos da ninidrina e colorimétrico não diferiram entre si, podendo ambos serem utilizados

como metodologias de triagem. Entre as três técnicas, a cromatografia líquida de alta eficiência foi a metodologia mais sensível na detecção e quantificação do soro em leite cru refrigerado.

## **2.4 Métodos para detecção de fraude por adição de soro em leite**

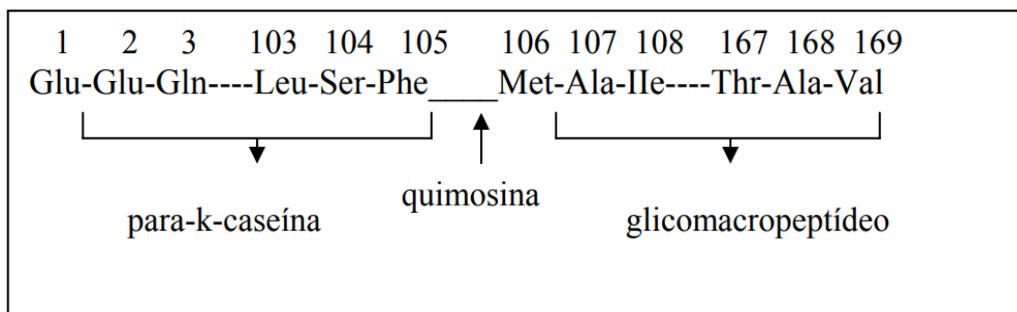
Existem vários métodos para a detecção da adição de soro de queijo ao leite, os quais se desenvolveram com a finalidade inicial de detectar a fraude de soro em pó ao leite em pó. Isto vem ao encontro das necessidades de países europeus, que possuem uma grande produção e comércio internacional de leite e soro em pó. No caso do Brasil, voltam-se as atenções para fraude de soro de queijo ao leite fluido, devido à sua maior importância no mercado (Silva, 1997).

As provas para detecção de fraudes são muito laboriosas e isso dificulta a realização rápida e na frequência determinada pela legislação. Por estes e outros motivos, várias ocorrências de fraudes passam despercebidas e boa parte do leite fraudado acaba chegando aos consumidores (Mereze *et al.*, 2015).

Existem diversas técnicas para detecção de adição de soro de queijo ao leite, mas o teste oficial e aprovado pelo MAPA (Brasil, 2018a) é o do método de determinação do índice de CMP no leite, utilizando-se a CLAE.

### **2.4.1 Cromatografia líquida de alta eficiência (CLAE)**

A adição fraudulenta de soro ao leite é oficialmente detectada e quantificada pela determinação do caseinomacropéptido (CMP). Este método consiste na detecção e quantificação de CMP por meio de cromatografia líquida de alta eficiência (CLAE) com separação em coluna de filtração em gel e detecção em ultravioleta. O CMP é resultante da hidrólise da molécula de caseína entre os aminoácidos 105 e 106 (fenilalanina e metionina) pela ação de enzima (quimosina). Desta cisão forma-se a para-k-caseína (1-105), que permanece na porção de caseína e o caseinomacropéptido (106-169), solúvel no soro (Carvalho *et al.*, 2007). A Figura 2.1 representa o local da ação da quimosina na k-caseína.

**Figura 2.1** - Local da ação da quimosina na k-caseína

FONTE: FENNEMA (2000).

O caseinomacropéptido é um dos inúmeros nomes para os peptídeos resultantes da hidrólise da k-caseína, sendo também conhecido como glicomacropéptido (GMP), devido ao seu elevado conteúdo de carboidratos; caseinoglicomacropéptido (CGMP); ou peptídeo derivado da caseína (CDP) (Tullio, 2007).

O polipeptídeo liberado durante a fase primária da coagulação pela ação da quimosina sobre a k-caseína é formado a partir da sequência de cerca de 64 aminoácidos, a quarta parte dos quais são hidroxiaminoácidos, regularmente distribuídos. Todos os carboidratos estão ligados covalentemente a este peptídeo, assim como o único grupo fosfato. O perfil dos aminoácidos e o seu conteúdo em açúcares tornam o CMP fortemente polar e, como consequência, ele é solúvel no soro após a separação da massa (López-Fandiño e Ramos, 1992).

A cromatografia consiste em um método físico-químico de separação dos componentes de uma mistura. Fundamenta-se na migração diferencial dos componentes, que ocorre devido a diferentes interações entre duas fases imiscíveis, a fase móvel e a fase estacionária. A cromatografia pode ser utilizada para a identificação de compostos por comparação (com padrões previamente existentes), para a purificação de compostos, separando-se as substâncias indesejáveis e para a separação dos componentes de uma mistura (Degani *et al.*, 1998).

A CLAE é muito usada na análise de detecção do CMP que apresenta massa molar em torno de 8000 Daltons, possibilitando sua separação dos demais peptídeos do leite, por meio de colunas de CLAE – filtração em gel. O método consiste num sistema de bombeamento isocrático, sendo a fase móvel constituída de uma solução tampão

fosfato, em pH 6,0. A coluna utilizada geralmente é a ZORBAX GF 250 (Du Pont), cujo princípio de separação baseia-se na exclusão molecular. A detecção se dará por espectrofotometria ultravioleta, no comprimento de onda de 205nm, sendo os compostos quantificados por integração eletrônica das áreas dos picos (Alvim, 1992).

A CLAE é um método sensível e rápido, porém requer mão de obra qualificada e equipamentos sofisticados, tornando-se um método de alto custo (Fukuda *et al.*, 2004; Carvalho *et al.*, 2007; Magalhães, 2008).

Lopez-Fandiño e Ramos (1993) e Recio *et al.* (1996) utilizaram CLAE em fase reversa para determinar o local da hidrólise da  $\kappa$ -caseína pela quimosina e enzimas microbianas. Eles foram capazes de obter a separação das variantes genéticas A e B da  $\kappa$ -caseína através da precipitação com TCA 4%. Todavia essa metodologia apresenta algumas dificuldades em seu emprego em escala industrial, devido ao custo relativamente elevado em sua implantação e a necessidade de pessoal treinado, por não ser uma técnica de fácil execução.

Mollé e Léonil (1995) realizaram o primeiro uso da CLAE em Fase Reversa acoplada à espectrometria de massa, para a análise e caracterização do CMP. A partir dos resultados obtidos eles concluíram que a utilização destes métodos combinados oferece uma alternativa muito promissora para a análise de misturas complexas, especialmente quando há alguma dificuldade em separação apenas por CLAE, sendo capaz de diferenciar o CMP de outros peptídeos originados a partir da hidrólise próximo à ligação 105-106 (pseudo-CMP).

De acordo com os resultados encontrados por Souza *et al.* (2018), a CLAE é a metodologia de maior sensibilidade na determinação do soro de queijo em leite, e os métodos de ninidrina ácida e colorimétrico adaptado são os mais empregados na triagem tendo em vista a detecção e a quantificação de soro de queijo em leite.

No entanto, proteases termoestáveis produzidas por micro-organismos psicrotóxicos podem, também, ser responsáveis por esse fenômeno mesmo após o tratamento térmico (Carvalho *et al.*, 2007). A proteólise causada por esses micro-organismos no leite,

principalmente *Pseudomonas* ssp., proporciona a degradação parcial da caseína liberando peptídeos contendo ácido siálico, que influencia nos métodos de detecção de adição de soro ao leite (Firmino *et al.*, 2010).

As proteinases psicrotróficas hidrolisam a caseína, levando a formação do CMP, hidrolisando próximo às ligações 105-106 (Fig 2.1). O pseudo-CMP pode diferir do CMP por um único aminoácido (metionina terminal no CMP, resíduo 106 – alanina terminal no pseudo-CMP, resíduo 107). O pseudo-CMP apresenta um aminoácido a menos que o CMP (fragmento 106-169 da k-caseína) (Recio *et al.*, 1996). Assim, a liberação deste pseudo-CMP pode ocorrer naturalmente no leite nas condições de estocagem e transporte, sendo que vários fatores, tais como tempo pós-ordenha e condições de transporte, interferem na sua liberação. Deste modo, as análises baseadas na identificação do CMP podem resultar na classificação errônea das amostras, ou seja, resultados falso-positivos, quanto à adição ao leite, de uma quantidade de soro de queijo (Magalhães, 2008).

Durante a pasteurização, grande parte dos micro-organismos psicrotróficos é destruída, mas este tratamento tem poucos resultados sobre a atividade das enzimas termoresistentes produzidas por estes micro-organismos (Cunha e Brandão, 2000). Mesmo após os micro-organismos psicrotróficos serem inativados pelo processo de pasteurização ou ultrapasteurização do leite a termoresistência de proteases bacterianas e a atividade residual elevada pode ser detectada após os tratamentos térmicos, uma vez que as enzimas produzidas continuam provocando alterações no produto (Tronco, 2008).

No caso de ser resultante da ação de enzimas produzidas por bactérias psicrotróficas, a presença de níveis elevados de CMP seria um indicativo de produto alterado. O artigo 504 do RIISPOA (Brasil, 2017) diz que são considerados alterados as matérias-primas ou os produtos que não apresentem condições higiênico-sanitárias adequadas ao fim a que se destinam e incorrem em risco à saúde pública, sendo considerados adulterados as matérias-primas ou os produtos de origem animal e identificados como fraudados ou falsificados.

A contagem padrão em placa (CPP) pode fornecer uma ideia do grau da contaminação do alimento e auxiliar na previsão do seu tempo útil de conservação (Silva, 2001). Entre os microrganismos encontrados estão os principais contaminantes do leite: *Bacillus*, *Enterobacteriaceae*, *Clostridium*, *Corynebacterium* e *Streptococcus* (Sangaletti *et al.*, 2009)

Oliveira *et al.* (2009) verificaram que o tempo de armazenamento do leite pode interferir no resultado do índice de CMP, podendo aumentar e levar a alterações nos resultados de CMP obtidos pela CLAE. Concluíram que a estocagem do leite em temperaturas inferiores a 7°C poderia reduzir o desenvolvimento de bactérias psicrotróficas com potencial proteolítico.

Prata (2009) relatou que o teor de caseinomacropéptido em leite e derivados tem assumido características de importante marcador das ações proteolíticas sofridas, seja na condição de matéria-prima ou mesmo no monitoramento de etapas importantes do processo tecnológico do leite.

De acordo com a Legislação brasileira vigente, quando o índice de CMP for de até 30 mg/L o leite poderá ser destinado ao abastecimento direto, quando estiver entre 30 mg/L e 75mg/L poderá ser destinado à produção de derivados lácteos e acima de 75 mg/L à alimentação animal ou indústria química em geral ou a outro destino a ser avaliado tecnicamente (Brasil, 2006).

#### **2.4.2 Técnica da ninidrina ácida (determinação de ácido siálico)**

Esta técnica baseia-se na identificação do ácido siálico no soro, liberado durante a elaboração de queijos por ação enzimática. O ácido siálico (N-acetilneuramínico), também denominado NANA, é um componente natural do leite, particularmente ligado ao GMP que é liberado da  $\kappa$ -caseína (Silva, 1997; Brasil, 2006). Na hidrólise inicial por ação da quimosina ocorre a formação do peptídeo conhecido como glicomacropéptido (GMP), que carrega todos os açúcares, sendo um deles o ácido siálico (Guinee e Wilkinson, 1992).

A análise do ácido siálico livre e ligado à glicoproteína do leite (ASL) embora seja um método de fácil realização, vem sendo estudado por alguns autores para avaliar a adição de soro ao leite. No entanto, este parâmetro pode sofrer influência de fatores externos que, em grande parte, não podem ser controlados e, como consequência, levam ao aumento da concentração de ácido siálico no leite (Aquino, 2013).

De acordo com Fukuda *et al.* (2004), o método do ASL além de ser de fácil execução, tem se mostrado eficiente na rotina laboratorial, não exigindo equipamentos sofisticados ou pessoal qualificado. Quando comparado outros métodos, apesar de possuir menor custo, a sua utilização apresenta algumas limitações como a alta sensibilidade ao oxigênio e o tempo requerido para as análises.

Este método só apresenta uma boa sensibilidade quando se tem a comparação com a amostra controle, ou seja, um leite comprovadamente não fraudado, pois há uma constante variação na concentração de ácido siálico ao longo da estocagem que não permite que sejam fixados valores específicos para identificação ou não da fraude com soro como foi feito por Prata (2002).

Segundo Silva (1997) este método necessita de um conhecimento prévio do teor de ácido siálico livre do leite, leite em pó, soro e soro em pó de produção nacional, seus valores médios e variações, tornando-se necessário estudar os efeitos do processamento tecnológico e a influência de quantidades residuais de lactose, que podem interferir no método. Além disso, adição de soro proveniente de coagulação não enzimática do leite não pode ser detectada por este método.

A técnica do ASL para a determinação quantitativa de ácido siálico foi homologada como metodologia oficial no Brasil, através da Instrução Normativa nº 68 de 14/12/06 do Ministério da Agricultura, Pecuária e Abastecimento (Brasil, 2006) que regulamentou os Métodos Analíticos Físico-Químicos para Controle de Leite e Produtos Lácteos. Porém, na atualização do Manual de métodos oficiais para análise de alimentos de origem animal, através da Instrução Normativa nº 30 de 26/06/2018 (Brasil, 2018a) não consta mais essa análise como método oficial.

### 2.4.3 Eletroforese

A eletroforese SDS-PAGE pode ser considerada um método eficaz para detecção deste tipo de fraude pela sensibilidade na identificação de proteínas que caracterizem a adição de soro ao leite, pela proporção de  $\alpha$ -caseína,  $\beta$ -caseína e a  $\beta$ -lactoglobulina (Aquino, 2013).

Esta técnica é utilizada para separar misturas complexas de proteínas, a partir de células, matrizes alimentares, frações subcelulares, ou imunoprecipitados, para investigar composições de subunidade ou para purificar as proteínas para utilização em outras aplicações (Schröder *et al.*, 2008). É amplamente utilizada em pesquisas com a finalidade de identificar fraudes em alimentos e a possível presença de proteínas estranhas à composição normal de produtos cárneos e lácteos (Aquino, 2013).

Uma das vantagens da utilização da eletroforese SDS-PAGE está relacionada com ausência da variação no perfil proteico ao longo da estocagem que ocorre em outras metodologias, como é o caso do ASL, que sofre constante influência da quantidade de GMP. Esta característica seria a principal razão para se inserir a eletroforese SDS-PAGE como análise oficial na identificação da fraude com soro de queijo. Além disso, permite o processamento simultâneo de diversas amostras e dispensa a presença do analista em algumas de suas etapas. No entanto, os analistas devem ser treinados e o custo desta análise é considerado alto frente à técnica do ASL (Aquino, 2013). Este método permite detectar adulterações de, pelo menos, 0,5% de soro líquido ou em pó adicionado ao leite pasteurizado ou ao leite em pó (Silva, 1997). Métodos que utilizam eletroforese capilar têm sido usados com sucesso na identificação e quantificação do caseinomacropéptido em leites (Magalhães, 2008).

Recio *et al.* (2000) estudaram a aplicação da eletroforese capilar para detecção de sólidos de soro de queijo em amostras de leite UAT genuínas e adulteradas, bem como amostras comerciais. Além disso, dois lotes de leite, com e sem soro de queijo, foram adicionados ao leite UAT, sob condições controladas de armazenamento e temperaturas diferentes, para avaliar a influência da atividade proteolítica durante o armazenamento. O estudo indicou que a eletroforese capilar foi eficiente na detecção de sólidos de soro

de queijo adicionado ao leite e que interferências, devido apenas à proteólise intensa, ocorrem após longo período de armazenamento ou a temperaturas elevadas.

Outros diversos métodos têm sido estudados e propostos para a detecção da adulteração do leite com soro de queijo, como mostra a Tabela 2.1. Em geral, esses métodos exigem excessivo trabalho laboratorial, longo tempo de análise e apresentam baixa sensibilidade (Carvalho *et al.*, 2007).

**Tabela 2.1** - Métodos relatados na literatura para detecção de adição de soro de queijo ao leite

PRINCÍPIOS DO MÉTODO	AUTOR
Proporção cisteína/cistina	Hill e Leary (1968)
Proporção cisteína/cistina	Koning e Rooijen (1971)
Microscopia eletrônica	Kalab (1980)
Teor de sulfidrila	Hill <i>et al.</i> (1988)
Comparação da composição de aminoácidos	Greenberg e Domer (1988)
Comparação da composição de aminoácidos	Walsh e Brown (1988)
Crioscopia	Castaneda <i>et al.</i> (1987)
Determinação da relação fósforo caseínico/nitrogênio proteico	Furtado e Wolfschoon-Pombo (1988)

FONTE: CARVALHO *et al.* (2007).

Kasemsuran *et al.* (2007) investigaram a viabilidade da utilização de uma técnica baseada na espectroscopia de infravermelho próximo (NIR), para detecção e quantificação de água e de soro de leite como adulterantes em leite. Os resultados obtidos por eles demonstram que espectroscopia NIR é viável para detectar adulterantes tais como água e soro de leite, e para determinar o seu conteúdo no leite.

Bremer *et al.* (2008) testaram um ensaio imunoenzimático (ELISA) para a detecção de soro de queijo em leite em pó. Este teste foi criado com objetivo de ser um método de triagem alternativo, de fácil utilização, barato e específico. O ELISA utiliza um anticorpo monoclonal específico desenvolvido contra o CMP com limite de detecção de 1% (m/m). Para avaliar a adequação do ELISA, 60 amostras comerciais de leite em pó foram analisadas pela técnica do ELISA, CLAE-FG (Filtração em gel) e CLAE-FR

(Fase reversa). O método ELISA foi considerado de fácil execução e tempo de ensaio curto, quando comparado com a CLAE-FG, método oficial pela UE (União Europeia). A sensibilidade alcançada pelo teste foi adequada para utilização como triagem de acordo com a legislação da UE. Não foram verificados resultados falsos negativos pelo ELISA e o número de resultados falsos positivos foi inferior ou comparável aos resultados da CLAE-FG.

Oancea (2009) analisou amostras de leite e bebidas lácteas obtidas a partir de fontes comerciais, pelo teste imunocromatográfico com anticorpos monoclonais altamente específicos. O CMP foi detectado em 70% das amostras selecionadas. Concluiu-se que o procedimento descrito é uma ferramenta útil para a detecção de rotina de adição fraudulenta de soro a leite e produtos lácteos.

Mendonça *et al.* (2017) utilizaram um Kit rápido, STICK c-GMP®, que é um teste imunocromatográfico para detecção de glicomacropéptido de caseína no leite (c-GMP). Este teste é muito sensível, rápido e capaz de detectar níveis de GMP que podem indicar adulteração de 4% de uso de soro de queijo em leite ou ainda apontar falhas nas condições de ordenha, conservação, transporte e processamento do leite, podendo gerar entre 2% até 1% de c-GMP na amostra. O teste é realizado por meio de tiras que são imersas na amostra, previamente preparada, e então é realizada a leitura a partir da formação ou não de linhas nas tiras. O c-GMP, quando presente em amostras de leite, reage com as partículas recobertas com anticorpos monocromáticos específicos frente ao glicomacropéptido. Este complexo de partículas passa por um processo cromatográfico e esta reação origina a formação de uma linha vermelha na tira de teste. Das 15 amostras analisadas, três amostras de leite estavam fora da especificação permitida para GMP, sendo que duas destas não deveriam estar sendo comercializadas para alimentação humana pois afeta a qualidade nutricional do leite.

## **2.5 Fundamentos da Espectrofotometria no Infravermelho**

A espectroscopia compreende o estudo da interação da radiação eletromagnética com a matéria, cujo principal objetivo é a determinação dos níveis de energia e transições de espécies atômicas e moleculares. A energia de certo estado corresponde ao movimento

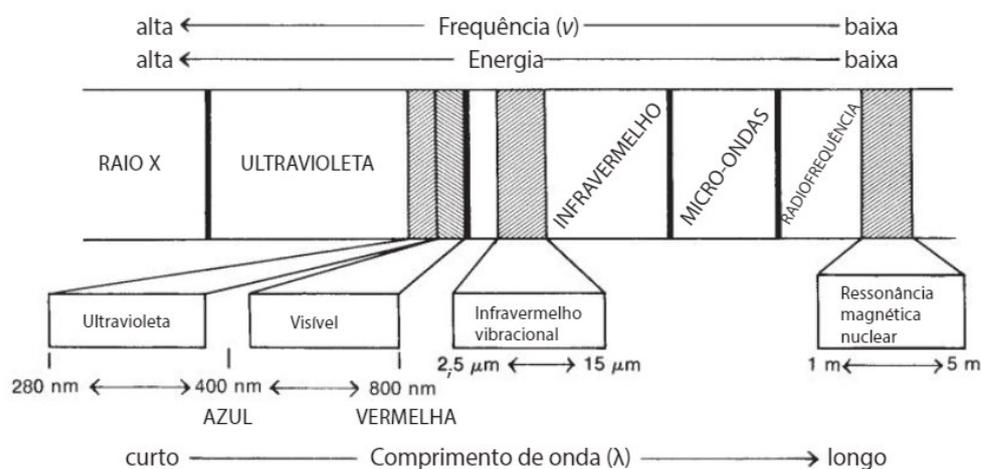
dos elétrons em torno do núcleo. Assim, os vários estágios são chamados de estados eletrônicos. Neste caso, as transições eletrônicas estão associadas a mudanças de orbital atômico, spin eletrônico e movimento angular total. Já para as moléculas, onde existem as ligações químicas, as transições eletrônicas envolvem mudança de energia dos elétrons de valência, ou seja, variações nas populações eletrônicas de orbitais moleculares. Em razão da existência das ligações químicas, as moléculas também possuem energia vibracional e rotacional (Skoog e Leary, 1992).

William Herschel, um astrônomo, descobriu a radiação infravermelha em 1800. Sabendo que a luz solar continha todas as cores do espectro e que era também uma fonte de calor, Herschel queria descobrir qual, ou quais, eram as cores responsáveis pelo aquecimento dos objetos. Idealizou um experimento utilizando um prisma e termômetros com bulbos pretos onde mediu as temperaturas de diferentes cores. Ele observou um aumento de temperatura à medida que movia o termômetro da cor violeta para o vermelho no espectro criado pela luz do sol atravessando o prisma. Descobriu que a temperatura mais quente ocorria além da luz vermelha. A radiação que causou esse aquecimento não era visível e foi denominado por Herschel como radiação invisível de infravermelho (Biggs *et al.*, 1987; Silveira *et al.*, 2004).

A espectrofotometria no infravermelho fundamenta-se nas mudanças de estados energéticos vibracionais causados pela interação da radiação eletromagnética com a matéria na região do infravermelho, que apresenta comprimentos de onda que vão de aproximadamente 0,78 a 1000  $\mu\text{m}$ . Essa faixa é dividida em três áreas: a radiação no infravermelho próximo (NIR-*Near Infrared*), que compreende o intervalo entre 0,78 a 2,5  $\mu\text{m}$  ( $\sim 14.000$  e  $4.000\text{ cm}^{-1}$ ); a radiação no infravermelho médio (MIR-*Middle Infrared*), que vai de 2,5 a 50  $\mu\text{m}$  ( $\sim 4.000$  e  $400\text{ cm}^{-1}$ ); e a radiação no infravermelho distante (FIR-*Far Infrared*), que abrange a faixa de 50 a 1000  $\mu\text{m}$  ( $\sim 400$  e  $10\text{ cm}^{-1}$ ) (HSU, 1997). A absorção de radiação infravermelha, em diferentes comprimentos de ondas, permite obter informações químicas qualitativas, quantitativas e físicas, as quais são intrínsecas de cada átomo e molécula (Stuart, 2004). A Figura 2.2 ilustra a relação da região do infravermelho com outras contidas no espectro eletromagnético. A maior parte dos químicos refere-se à radiação na região do infravermelho por meio de uma unidade chamada número de onda, em vez de comprimento de onda. Números de onda

são expressos em centímetros recíprocos ( $\text{cm}^{-1}$ ), o infravermelho vibracional médio vai de  $4.000$  a  $400 \text{ cm}^{-1}$  (Pavia *et al.*, 2010). Em uma leitura de um espectro, cada molécula apresentará o seu próprio espectro na região do infravermelho, tornando o método viável para identificar diferentes tipos de amostras. Os picos presentes no gráfico do espectro correspondem às frequências de vibrações entre os átomos que compõem cada amostra (Kamal e Karoui, 2015).

**Figura 2.2** - Uma parte do espectro eletromagnético que mostra a relação do infravermelho vibracional com outros tipos de radiação



FONTE: PAVIA *et al.* (2010).

Os padrões de absorção no infravermelho, ou espectros infravermelhos, de duas moléculas diferentes jamais serão idênticos. Assim, o espectro infravermelho pode servir para moléculas da mesma forma que impressões digitais servem para seres humanos. Quando se comparam os espectros infravermelhos de duas substâncias que se acredita serem idênticas, pode-se descobrir se elas são, de fato, idênticas (Pavia *et al.*, 2010).

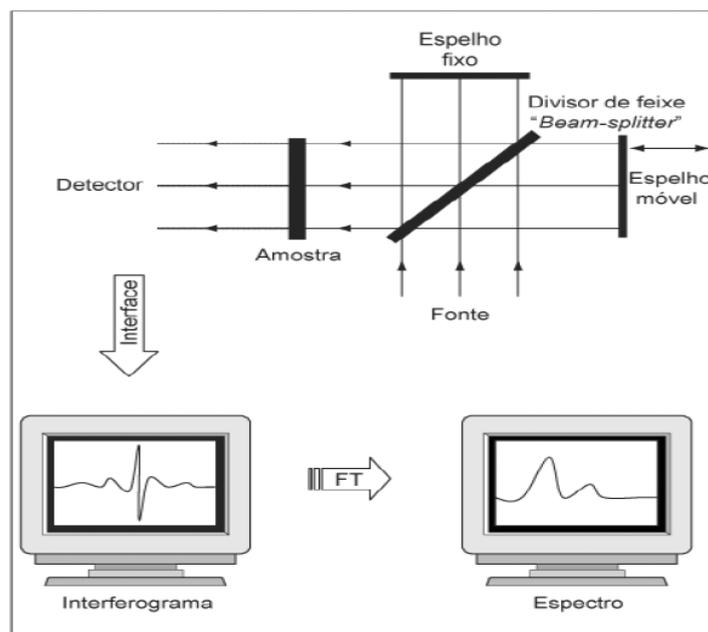
O instrumento que obtém o espectro de absorção no infravermelho de um composto é chamado de espectrômetro ou espectrofotômetro de infravermelho, que pode ser de dois tipos, dispersivo ou de transformada de Fourier (FTIR). Ambos produzem espectros praticamente idênticos, porém espectrômetros de infravermelho de transformada de Fourier têm maior velocidade e maior sensibilidade do que os dispersivos (Pavia *et al.*, 2010).

As bases da FTIR, foram estabelecidas no fim do século XX. O cientista Albert Abraham Michelson construiu um interferômetro e o descreveu em suas publicações, em 1891 e 1892. No ano 1892, Lord Rayleigh confirmou que o interferômetro tinha relação com um espectro por uma operação matemática denominada Transformada de Fourier. No entanto, a técnica foi pouco usada por mais de cinquenta anos, esperando o avanço dos computadores que deram grande impulso nesse ramo (Perkins, 1986).

A espectroscopia por infravermelho ganhou impulso com o surgimento da espectroscopia com transformada de Fourier e a utilização do interferômetro de Michelson, que tornou a técnica ainda mais rápida (Saliba *et al.*, 2003).

O interferômetro de Michelson consiste basicamente em dois espelhos (um fixo e um móvel) e um divisor de feixes (espelho semitransparente). A radiação proveniente da fonte incide sobre o divisor de feixes, que transmite metade da radiação para o espelho móvel e reflete a outra metade para o espelho fixo. Os espelhos refletem os dois feixes para o divisor e recombinam-se. Mas as diferenças de caminhos dos dois feixes causam interferências que podem ser tanto construtivas como destrutivas. Se os dois espelhos se encontram equidistantes do divisor, combinam-se construtivamente, porém se o espelho móvel se mover a determinadas distâncias do divisor, as amplitudes combinam-se destrutivamente. O feixe resultante passa pela amostra e em seguida, é redirecionado para o detector. Para a radiação no infravermelho (policromática), a soma de todas as interações construtivas e destrutivas para cada componente resulta num sinal complexo chamado de interferograma. Como o interferograma não pode ser interpretado, é empregada a técnica matemática chamada de Transformada de Fourier que converte os dados obtidos no interferômetro em um espectro que relaciona a intensidade *versus* frequência (número de onda), conforme pode ser observado na Figura 2.3 (Stuart, 2004; Helfer *et al.*, 2006).

**Figura 2.3** - Esquema ilustrativo para o interferômetro de michelson e do resultante da aplicação da transformada de fourier

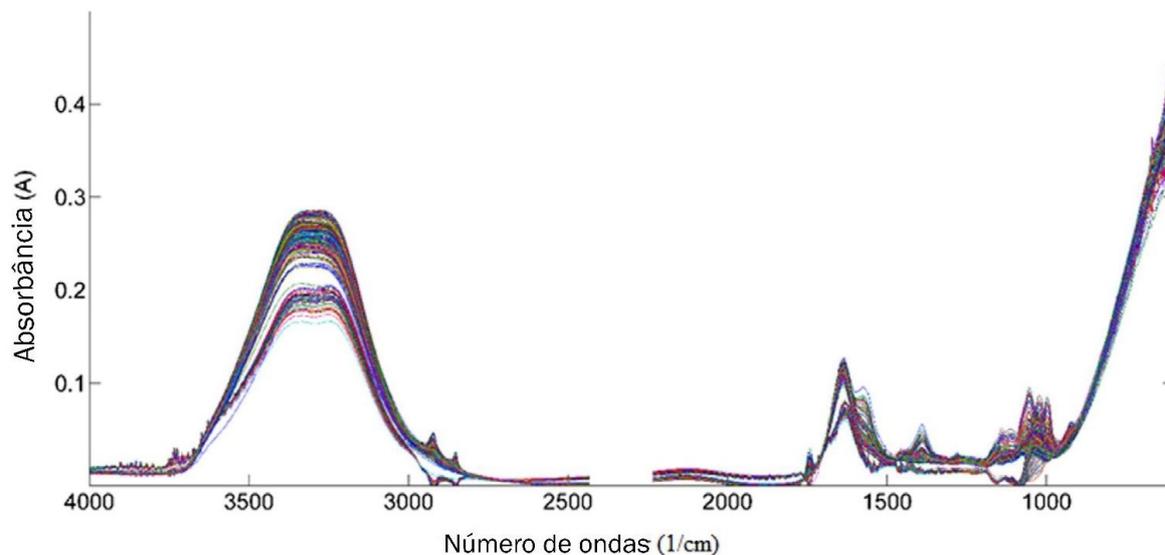


FONTE: HELFER *et al.* (2006).

A radiação no infravermelho atravessa a amostra a ser analisada, a radiação transmitida é comparada com aquela transmitida na ausência de amostra. O espectrômetro registra o resultado na forma de uma banda de absorção, fornecendo evidências da presença de vários grupos funcionais na estrutura orgânica devido à interação das moléculas ou átomos com a radiação eletromagnética em um processo de vibração molecular. A radiação no infravermelho faz com que átomos e grupos de átomos de compostos orgânicos vibrem com amplitude aumentada ao redor das ligações covalentes que os ligam (Silverstein, 2000).

Segundo Pavia *et al.* (2010), as frequências de radiação no infravermelho absorvidas equivalem às frequências vibracionais naturais de uma molécula e a energia absorvida eleva a amplitude dos movimentos vibracionais das suas ligações, ou seja, quando a radiação interage com moléculas orgânicas são gerados movimentos vibracionais e rotacionais que podem interagir com o campo elétrico da radiação e causar variações na amplitude de um de seus movimentos o que ocasiona na absorção da radiação no infravermelho e formação das bandas/picos. O espectro formado relaciona a intensidade *versus* frequência (número de onda), como ilustrado na Figura 2.4 (Botelho *et al.*, 2015).

**Figura 2.4** - Espectros de amostras de leite obtidos por FTIR



FONTE: BOTELHO *et al.* (2015).

Dois picos são facilmente observados. O primeiro, localizado aproximadamente entre 3700 e 2800  $\text{cm}^{-1}$ , corresponde à região de estiramento das ligações entre átomos de hidrogênio e átomos de oxigênio e carbono, ou seja, grupos OH e CH, respectivamente. A região onde se observa o segundo pico, entre 1700 e 1500  $\text{cm}^{-1}$ , é atribuída a ligações duplas, como ligações entre átomos de carbono e oxigênio e também a interações de proteínas e lipídios. Esses dois picos estão associados aos componentes naturais do leite, incluindo a água (Kohler *et al.*, 2010; Botelho *et al.*, 2015).

Como uma “impressão digital”, cada molécula apresentará o seu próprio espectro na região do infravermelho, tornando o método viável para identificar diferentes tipos de amostras (análise qualitativa). Os picos presentes no gráfico do espectro correspondem às frequências de vibrações entre os átomos que compõem cada amostra. A altura desses picos corresponde à quantidade de determinada amostra (análise quantitativa) (Morgano *et al.*, 2005).

As vantagens da técnica de FTIR sobre outros métodos incluem a rapidez na aquisição dos espectros, melhor relação sinal-ruído, alta resolução espectral, ser uma técnica não destrutível e apresentar boa reprodutibilidade. Baseia-se no fato de que as ligações químicas das substâncias possuem frequências de vibrações específicas, as quais correspondem aos níveis de energia da molécula (chamados de níveis vibracionais).

Tais frequências dependem da forma da superfície de energia potencial da molécula, da geometria molecular, das massas dos átomos e eventualmente do acoplamento vibrônico (Smith, 1999).

### **2.5.1 Espectrofotometria FTIR para análise de leite**

Para a análise de leite, os métodos baseados na turbidez e corante de ligação para a determinação de gordura e proteína, respectivamente, foram utilizados no passado, mas a técnica que revolucionou de fato os laboratórios de análise de leite fluido foi a espectrofotometria de infravermelho, através da metodologia FTIR (*Fourier Transform Infrared*). Tal metodologia apresenta grandes vantagens em relação àquelas de referência: rapidez e precisão, necessidade de pouca mão de obra, pouca ou nenhuma preparação da amostra para análise, assim como o baixo custo relativo à multiplicidade de análises em um único equipamento, tornando desnecessária a aquisição de vários equipamentos para análises distintas (Oliveira, 2011).

Os principais componentes do leite lidos no equipamento são sólidos totais, gordura, proteína, caseína, nitrogênio ureico e lactose. A sua determinação por meio de métodos convencionais é bastante demorada e trabalhosa, por isso, vários trabalhos foram realizados para comparação da espectrofotometria de infravermelho com as metodologias de referência; para evidenciar a aplicabilidade desta tecnologia como uma alternativa mais rápida para a análise de leite cru. Nos laboratórios oficiais da RBQL, a análise da composição centesimal dos constituintes do leite cru refrigerado é realizada em equipamentos eletrônicos com grande capacidade de análise, podendo realizar de 400 a 700 análises por hora. Esses equipamentos realizam a análise por absorção infravermelha, sendo os mais recentes baseados em tecnologia de Infravermelho por Transformada de Fourier - FTIR (Oliveira *et al.*, 2012).

A espectrofotometria no infravermelho próximo e médio é muito utilizada para determinar constituintes do leite (Rodriguez-Otero *et al.*, 1997; Stuart, 2004; Rodriguez-Saona *et al.*, 2006). O leite corresponde a uma emulsão composta por água, proteínas, lactose, lipídeos e sais minerais. Cada um desses componentes possui diferentes estruturas químicas que se apresentam livres ou sob algum tipo de interação. A gordura

absorve radiação no comprimento de onda de 5,73  $\mu\text{m}$  através dos grupos carbonila (C=O) das ligações éster dos triglicerídeos; as proteínas 6,46  $\mu\text{m}$ , através dos grupos amida (CONH) das ligações peptídicas; e a lactose 9,53  $\mu\text{m}$ , através dos grupos hidroxila (OH) (Oliveira, 2011). Os sólidos totais podem ser obtidos pelo somatório do conteúdo de proteína, gordura e lactose, acrescidos de uma constante de minerais, ou por meio da absorção da radiação no comprimento de onda 4,3  $\mu\text{m}$  dos grupos hidroxilas da água (Biggs *et al.*, 1987).

Os equipamentos de infravermelho não fazem leitura direta do ponto de congelamento do leite, mas calculam os componentes que contribuem para esse parâmetro e possuem uma sonda. A concentração de gordura, proteína e lactose podem ser medidas por meio de infravermelho, mas os sais e ácidos dissolvidos não podem ser determinados dessa forma. Então é utilizada a condutividade elétrica para determinar principalmente os sais, levando em consideração que os outros componentes e a água também influenciam nessa medida. Combinando-se a condutividade com os teores de gordura, proteína e lactose, é possível calcular a contribuição de cada componente no ponto de congelamento do leite. Utilizando-se a Regressão Linear Múltipla entre o ponto de congelamento, que é a variável dependente, e os conteúdos de gordura, proteína e lactose em conjunto com a condutividade, que são as variáveis independentes, é possível calcular coeficientes que atuam no ponto de congelamento (Arnvidarson *et al.*, 1998; DELTA INSTRUMENTS, 2007, 2009).

Silveira *et al.* (2004) avaliaram a metodologia de determinação da composição do leite por meio de analisador de infravermelho comparando com a metodologia de referência, utilizando-se 48 amostras individuais de leite de vacas da raça Holandesa e cinco amostras de leite de conjunto. A média dos resultados para gordura, proteína, lactose e sólidos totais das amostras individuais foram, respectivamente, 3,8%, 3,3%, 4,6% e 12,4%, usando-se metodologia de referência e 3,8%, 3,3%, 4,5% e 12,4%, usando-se equipamento eletrônico FTIR. Para amostras de tanque, as médias foram 3,9% e 4,0% para gordura, 3,2% e 3,2% para proteína, 4,7% e 4,6% para lactose e 12,6% e 12,6% para sólidos totais, respectivamente, para análises feitas por métodos de referência e por analisador de infravermelho. Não houve diferença ( $p>0,05$ ) entre os valores obtidos pelos diferentes métodos avaliados.

Em artigo sobre uso do FTIR para rápida autenticação e detecção de adulterações em alimentos, Rodriguez-Saona e Allendorf (2011) apresentaram uma série de estudos relacionados à utilização do MIR e do NIR. Segundo os autores, o FTIR-MIR é comumente utilizado para identificação estrutural de compostos orgânicos, devido às vibrações fundamentais promovidas nos grupos funcionais que proporcionam boa absorção e bandas mais bem definidas que no FTIR-NIR. Apesar das diferenças nos processos de absorção nas regiões do MIR e do NIR, ambos têm sido utilizados para determinação de padrões de alimentos de origem vegetal, sucos de frutas industrializados, produtos agroindustriais, óleos refinados, gorduras e em vários outros produtos alimentícios. Enfatizaram a aplicação do FTIR na detecção de grande número de adulterações, a partir de análise que demanda pequena quantidade de amostra, que dispensa preparo prévio da mesma e que proporciona resultados rápidos. Concluíram que os métodos baseados em análises de espectros vibracionais estão emergindo com rapidez, apresentando técnicas modernas e eficientes para monitoramento de adulterações e contaminações.

Oliveira *et al.* (2012) realizaram um estudo para avaliar o teor de nitrogênio ureico no leite (NUL) cru produzido no Brasil através do equipamento CombiScope™ FTIR (Delta Instruments). A repetibilidade e reprodutibilidade do CombiScope™ FTIR (Delta Instruments) e a comparação com um método enzimático automatizado (ChemSpec® 150; Bentley Instruments) foram testados para a medição do nitrogênio ureico no leite (NUL) cru. Houve alta correlação entre os métodos FTIR e enzimático automatizado para a análise de ureia, sem diferença significativa entre ambos ( $p > 0,05$ ). Os resultados obtidos evidenciam que o equipamento CombiScope™ FTIR é um método confiável para a análise do teor de ureia no leite cru. Entretanto, a detecção de ureia adicionada de forma fraudulenta ao leite cru não foi linearmente proporcional, com diferença significativa para adição de ureia em níveis de 40mg/dL.

O uso de infravermelho médio com reflectância atenuada foi avaliado em pesquisa desenvolvida por Santos *et al.* (2013). Os autores realizaram adulterações artificiais com soro de queijo, ureia, urina, peróxido de hidrogênio em leites comerciais. Dois modelos quimiométricos foram aplicados aos dados espectrais: Pattern Recognition Analysis by Soft Independent Modeling of Class Analogy (SIMCA) e Partial Least Squares

Regression (PLSR). Constataram que os dois modelos possibilitaram a identificação das adulterações artificiais promovidas no leite. A pesquisa comprovou que a técnica espectroscópica no FTIR-MIR dotado de reflectância total atenuada permite a detecção e a quantificação de adulterantes em leite em níveis satisfatórios.

## 2.6 Aprendizado de máquina

Nos últimos anos, com a globalização e a evolução da tecnologia, a quantidade de dados armazenados aumentou consideravelmente. Bancos de dados de grandes empresas recebem diariamente milhares de novas informações e essas são, na maioria das vezes, “jogadas” no repositório de dados e “esquecidas” lá, quando deveriam ser estudadas, analisadas para que esses registros não representem apenas dados, mas, sim, conhecimento. Bancos de dados podem conter informações desconhecidas que podem ser consideradas importantíssimas, visto que, na atualidade, um dos maiores tesouros é a informação (Asseiss, 2017).

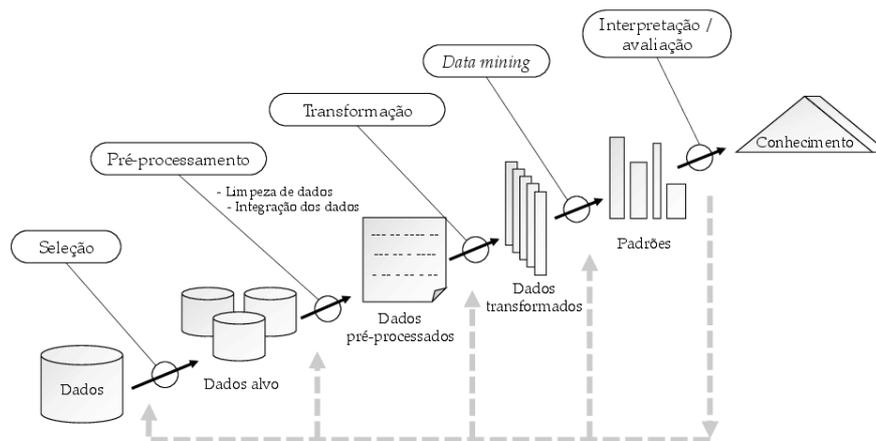
À busca por informações em banco de dados é dado o nome de Processo de Descoberta de Conhecimentos em Banco de Dados (do inglês, *Knowledge Discovery in Database - KDD*). O processo de KDD é composto por cinco etapas, sendo elas: etapa de seleção de dados, etapa de pré-processamento dos dados, etapa de transformação de dados, etapa de *Data Mining* e Aprendizado de Máquina e etapa de interpretação e avaliação dos dados. Todas as etapas são importantes, uma vez que preparam os dados e fornecem um procedimento organizado para a descoberta de novos conhecimentos, mas destaca-se a importância da etapa de *Data Mining* e Aprendizado de Máquina, onde algoritmos específicos são aplicados aos conjuntos de dados, permitindo a detecção de padrões e relacionamentos imprescindíveis para a descoberta de conhecimento (Asseiss, 2017). O KDD refere-se ao processo global de descobrir conhecimento útil a partir de dados, e a mineração de dados refere-se a um determinado passo nesse processo, que é a aplicação de algoritmos específicos para extrair padrões de dados (Fayyad *et al.*, 1996).

A descoberta de conhecimento de bancos de dados foi apresentada no primeiro workshop de KDD em 1989, na qual foi enfatizado que o conhecimento é o produto

final de uma descoberta orientada por dados. Sendo então, popularizados os campos de Inteligência Artificial e aprendizado de máquina (Fayyad *et al.*, 1996).

Pode-se dizer que o processo de KDD compreende, em outras palavras, todo o ciclo que o dado percorre até que este se transforme em um conhecimento (Cereda e José, 2014). As principais etapas do processo são apresentadas, na forma de uma visão geral, na Figura 2.5.

**Figura 2.5** - Etapas do processo de descoberta de conhecimentos em banco de dados (KDD)



FONTE: Adaptado de Fayyad *et al.* (1996).

O uso de um banco de dados é essencial para armazenar tanta informação de forma organizada e fácil de ser recuperada, pois computacionalmente a recuperação de dados é significativamente mais rápida (Asseiss, 2017). A etapa de seleção de dados requer um tempo considerável do processo, pois é necessário conhecer e entender o banco de dados em questão. (Boente *et al.*, 2008).

Na etapa de pré-processamento é feita uma limpeza nos dados, ou seja, pode ser retirado algum atributo considerado sem relevância para o domínio em questão, ficando somente atributos que realmente integram o contexto (Boente *et al.*, 2008), além de remover dados duplicados e corrompidos (Goldschmidt e Passos, 2005).

A etapa de transformação permite otimizar o tempo de processamento do algoritmo, uma vez que reduz o número e a complexidade dos atributos disponíveis para análise (Jiang *et al.*, 1999). Nesta etapa, é feita a execução de uma série de funções e a aplicação de conjuntos de regras aos dados extraídos para convertê-los em um formato

padrão para atender aos requisitos de esquema do banco de dados de destino (Han e Kamber, 2006).

*Data Mining* (em português, Mineração de Dados) é o processo de encontrar anomalias, padrões e correlações em grandes conjuntos de dados para prever resultados. A mineração de dados extrai informações de um conjunto de dados e o transforma em uma estrutura compreensível. É o processo computacional que envolve métodos na interseção de inteligência artificial, aprendizado de máquina, estatística e sistemas de banco de dados. A tarefa de mineração de dados real é a análise automática ou semiautomática de grandes quantidades de dados para extrair padrões interessantes previamente desconhecidos (Kesavaraj e Sukumanan, 2013). As diferentes tarefas de *data mining* podem ser definidas como sumarização, classificação, agrupamento, associação e regressão (Fayyad *et al.*, 1996; Chen *et al.*, 1996). A classificação que é bastante utilizada é a aprendizagem de regras ou funções que determinam a classe de um objeto de acordo com seus atributos (Chen *et al.*, 1996).

As principais áreas de aplicação do KDD incluem marketing, finanças (especialmente investimento), detecção de fraude, produção, telecomunicações e agentes da Internet (Fayyad *et al.*, 1996).

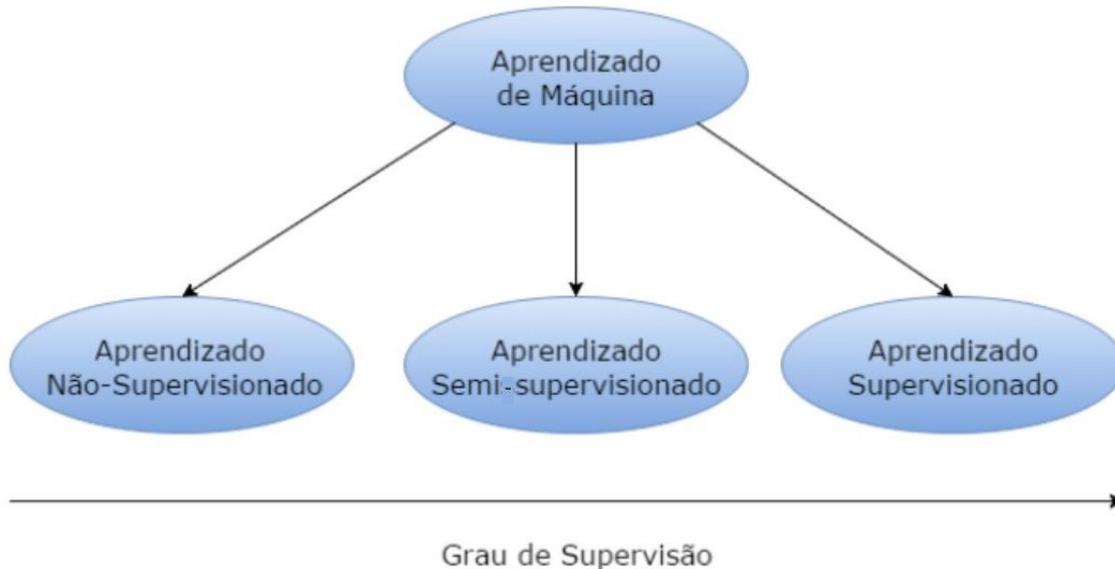
### **2.6.1 Técnicas de Aprendizado de Máquina**

O aprendizado de máquina é um conceito relacionado à área de Inteligência Artificial (IA), cujo objetivo é desenvolver métodos computacionais capazes de tomar decisões com base em conhecimentos já existentes (Michalski *et al.*, 2013). Em outras palavras, a aprendizagem de máquina trata da construção de técnicas automatizadas que possibilitem que um computador, a partir da aquisição de novos conhecimentos, melhore seu desempenho preditivo e habilidades automaticamente de acordo com a experiência prévia. Desta forma, o aprendizado de máquina possibilita a aquisição de conhecimento por meios computacionais (Ferreira, 2016).

Conforme ilustrado na figura 2.6, o aprendizado pode ser dividido em três categorias: aprendizado supervisionado, aprendizado não supervisionado e aprendizado semi-

supervisionado (por reforço) (Monard e Baranauskas, 2003). Esta classificação está relacionada com os métodos adotados no processo de generalização do conhecimento.

**Figura 2.6** - Hierarquia do aprendizado de acordo com o grau de supervisão



FONTE: Monard e Baranauskas (2003).

Na aprendizagem não supervisionada considera-se o conjunto de exemplos não rotulados, e, portanto, estão disponíveis somente os exemplos de entradas. Neste caso, o algoritmo indutor não possui a informação sobre a classe à qual pertence cada exemplo. O objetivo na aprendizagem não supervisionada é encontrar similaridades ou diferenças no conjunto de valores dos atributos dos exemplos que permitam a criação de agrupamentos ou *clusters*. Desta forma, exemplos agrupados podem ser atribuídos a uma mesma classe, enquanto exemplos dispersos estão mais propensos a pertencer a classes distintas (Russell e Norvig, 2003).

O aprendizado semi-supervisionado combina ambos os métodos supervisionado e não supervisionado. Esta abordagem utiliza um pequeno conjunto de exemplos rotulados, e um conjunto maior de exemplos não rotulados. O objetivo neste caso é observar ambos os conjuntos de exemplos para encontrar uma hipótese capaz de classificar novas observações entre as classes já existentes (Ferreira, 2016).

No aprendizado supervisionado são aplicadas técnicas onde algoritmos indutores realizam inferências a partir de exemplos rotulados, ou seja, cada exemplo observado é

descrito por um conjunto de atributos e pelo valor da classe, também denominada como atributo meta, à qual o exemplo pertence (Russell e Norvig, 2003). Neste caso, o objetivo do algoritmo é determinar um modelo capaz de classificar exemplos não rotulados recebidos como entrada. Em outras palavras, o processo de aprendizado supervisionado pode ser comparado com o processo de aprendizagem de um aluno em um ambiente onde há um professor. Ao resolver determinada tarefa, o professor avalia a resposta informando qual a resposta correta. Após a repetição contínua deste processo, o aluno saberá responder às novas questões corretamente sem o auxílio do professor. Em uma técnica de aprendizagem supervisionada a tarefa de aprendizado pode ser descrita como classificação ou regressão, de acordo com o tipo a qual pertence o atributo classe (Ferreira, 2016). São exemplos de métodos de aprendizado supervisionado *Support Vector Machines* (SVM), *K-NearestNeighbor* (KNN), Árvores de Decisão, Redes Neurais Artificiais (RNA), entre outros.

Exemplificando o aprendizado supervisionado, um agente externo (professor) apresenta à rede neural alguns conjuntos de padrões de entrada e seus correspondentes padrões de saída. Portanto, é necessário ter um conhecimento prévio do comportamento que se deseja ou se espera da rede. Para cada entrada, o professor indica explicitamente se a resposta calculada é boa ou ruim. A resposta fornecida pela rede neural é comparada à resposta esperada. O erro verificado é informado à rede para que sejam feitos ajustes a fim de melhorar suas futuras respostas (Ferneda, 2006).

No contexto das tarefas de aprendizado de máquina, o processo de classificação se dá por meio de um algoritmo indutor, cujo objetivo é predizer o rótulo de novas entradas com base nos exemplos de entrada já rotulados. Neste caso, o primeiro passo para a classificação é mapear e modelar o conjunto de dados do domínio ao qual pretende-se aplicar o algoritmo, preparando os exemplos para o processamento. Desta forma, o conjunto de entrada é composto por uma coleção de exemplos, sendo que cada exemplo é constituído por uma lista de atributos independentes, bem como um atributo classe, que corresponde ao rótulo. Em vista disso, o atributo classe representa o alvo no processo de classificação (Ferreira, 2016).

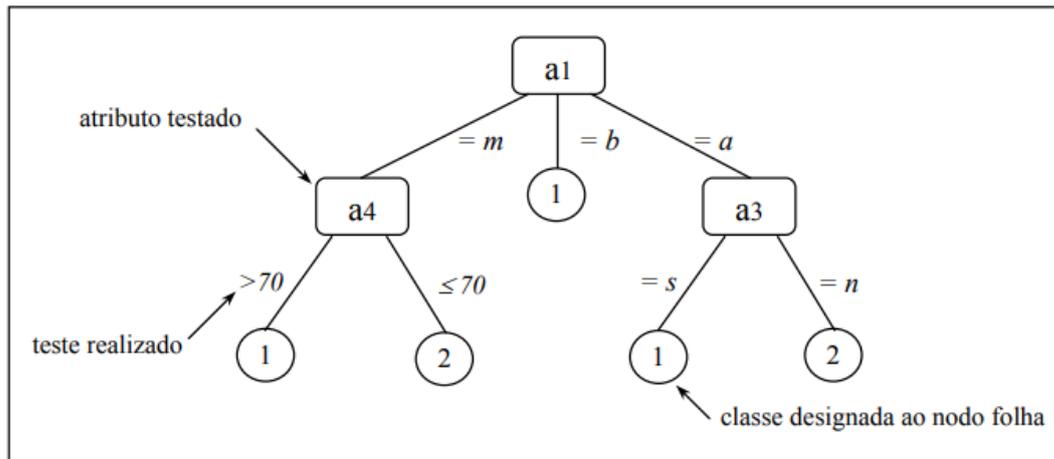
A classificação é uma das técnicas da mineração de dados utilizada principalmente para analisar um determinado conjunto de dados, de modo que o erro de classificação seja menor. É usada para extrair modelos que definem com precisão classes de dados importantes dentro do conjunto de dados determinado, e constitui em um processo de duas etapas. Durante a primeira etapa, o modelo é criado aplicando o algoritmo de classificação no conjunto de dados de treinamento e, na segunda etapa, o modelo extraído é testado em relação a um conjunto de dados de teste predefinido para medir o desempenho e a precisão do modelo (Nikam, 2015). O método de classificação pode ser utilizado em aplicações que incluem diagnósticos médicos, avaliação de risco em empréstimos, detecção de fraudes etc. permitindo obter padrões através da construção de classificadores (Garcia, 2003).

Os tipos de classificadores mais aplicados baseiam-se na utilização de árvores de decisão e de redes neurais artificiais como uma forma clara de representar o conhecimento contido no conjunto de dados analisado (Garcia, 2003).

### **2.6.1.1 Técnica de árvore de decisão**

Árvore de decisão é uma técnica de aprendizado de máquina que utiliza a estratégia de dividir para resolver um problema de decisão, dividindo um problema complexo em problemas mais simples, aos quais recursivamente é aplicada a mesma estratégia, ou seja, divide-se um conjunto de casos em classes. A primeira separação é o principal atributo de classificação, chamada de nó, sendo os casos divididos com base no positivo ou negativo para esse nó. O processo segue sucessivamente para cada subgrupo criado, terminando quando todos os exemplos em um dado subgrupo são da mesma classe, ou quando o resultado da divisão não pode ser melhorado. O resultado é um gráfico tipo árvore, onde, além dos nós citados, representando um atributo, cada ramo que nasce do nó significa um valor para ele, e os nós terminais (ou folhas) correspondem ao resultado final da classificação, ou seja, a qual classe a amostra foi associada, como exemplificado na Figura 2.7 (Dantas, 2017).

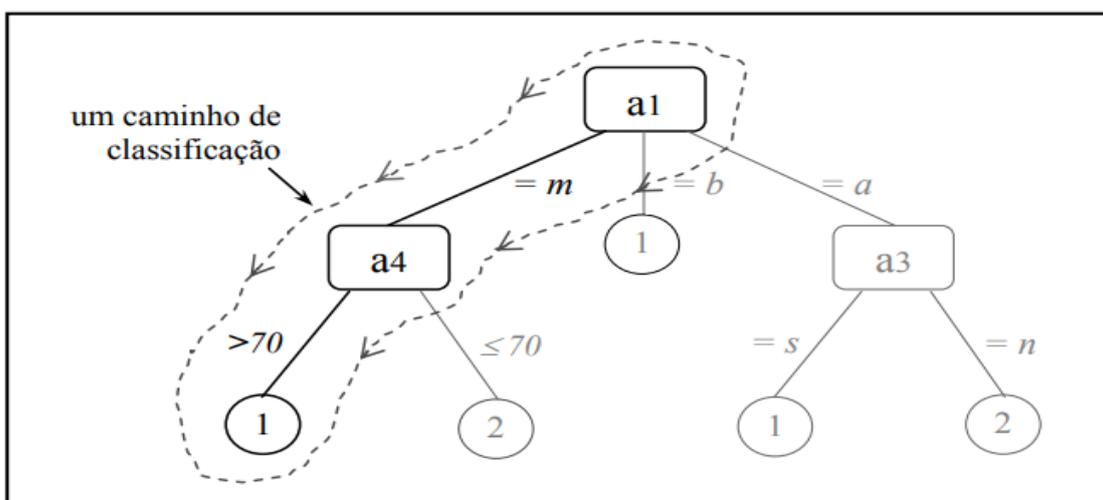
**Figura 2.7** - Exemplo de um classificador utilizando árvore de decisão



FONTE: GARCIA (2003).

Ao classificar um registro do conjunto de dados, percorre-se um caminho na árvore que se inicia a partir do nó raiz, onde é realizada a aplicação da condição de teste do nó, seguindo, posteriormente, para a ramificação apropriada baseado no resultado do teste. O próximo nó pode ser um nó interno, onde uma nova condição de teste é aplicada, ou um nó folha. Neste caso, o rótulo deste nó é atribuído ao registro que, por sua vez, é classificado (Tan *et al.*, 2009). A classificação da árvore ocorre ao se percorrer o caminho que se inicia no nodo raiz (a1) e se estende até as folhas. A situação é mostrada na Figura 2.8.

**Figura 2.8** - Exemplo de um caminho de classificação



FONTE: GARCIA (2003).

Os algoritmos se baseiam em uma estratégia que expande uma árvore de decisão escolhendo localmente quais atributos usar para particionar os dados (Asseiss, 2017).

### 2.6.1.2 Redes Neurais Artificiais (RNA)

A busca por um modelo computacional que simule o funcionamento das células do cérebro data dos anos 40, com o trabalho de McCulloch e Pitts. O entusiasmo pela pesquisa neste campo cresceu durante os anos 50 e 60, à medida que os computadores se tornaram mais avançados. Nesse período, Rosenblatt propôs, em 1958, um método inovador de aprendizagem para as redes neurais artificiais denominado *perceptron*. No final dos anos 60, Minsky e Papert publicaram um livro no qual apresentam importantes limitações do *perceptron*. As dificuldades metodológicas e tecnológicas, juntamente com os ataques extremamente pessimistas de Papert e Minsky, fizeram com que as pesquisas não evoluíssem nos anos seguintes. Durante os anos 70, esta linha de pesquisa contava com um número reduzido de cientistas, o que resultou numa redução drástica de grande parte do financiamento em pesquisas. Porém, durante os anos 80, o entusiasmo ressurgiu graças a avanços metodológicos importantes e ao aumento dos recursos computacionais disponíveis. Mozer, em 1984, foi o pioneiro na utilização de técnicas de redes neurais na recuperação de informação. Ele utilizou uma arquitetura bastante simples que não empregava uma das principais características das redes neurais, que é a capacidade de aprender (Ferneda, 2006).

Em meados dos anos 2000 o termo “aprendizagem profunda” começa a ganhar popularidade após um artigo de Geoffrey Hinton e Ruslan Salakhutdinov mostrar como uma rede neural de várias camadas poderia ser pré-treinada uma camada por vez. Em 2009 aconteceu o NIPS Workshop sobre Aprendizagem Profunda para Reconhecimento de Voz e descobriu-se que com um conjunto de dados suficientemente grande, as redes neurais não precisam de pré-treinamento e as taxas de erro caíram significativamente. Em 2012, algoritmos de reconhecimento de padrões artificiais alcançam desempenho em nível humano em determinadas tarefas. Em 2015, o Facebook coloca a tecnologia de aprendizado profundo – chamada DeepFace – em operação para marcar e identificar automaticamente usuários do Facebook em fotografias. Algoritmos executam tarefas superiores de reconhecimento facial usando redes profundas que levam em conta 120 milhões de parâmetros. Em 2017, ocorre a adoção em massa do *Deep Learning* em diversas aplicações corporativas e celulares, além do avanço em pesquisas.

As RNAs podem funcionar como modelos preditivos que descrevem a relação funcional entre as variáveis de entrada e variáveis de saída de um sistema (Valente *et al.*, 2014). São apresentadas como sistemas de “neurônios” interconectados que podem calcular valores de entradas e são capazes de realizar aprendizado de máquina, bem como reconhecimento de padrões devido à sua natureza adaptativa (Nikam, 2015). O conhecimento é adquirido por meio da aprendizagem, e os pesos, que são a intensidade das conexões entre os neurônios, são utilizados para armazenar estes conhecimentos (Coutinho *et al.*, 2016).

Uma rede neural artificial opera criando conexões entre muitos diferentes elementos de processamento, cada um correspondendo a um único neurônio em um cérebro biológico. Esses neurônios podem ser construídos ou simulados por um sistema digital de computador. Cada neurônio recebe muitos sinais de entrada, em seguida, com base em uma ponderação interna, produz um único sinal de saída, que é enviado como entrada para outro neurônio. Os neurônios estão fortemente interconectados e organizados em diferentes camadas. A camada de entrada recebe os sinais iniciais, e a camada de saída produz o resultado final. Em geral, uma ou mais camadas ocultas são inseridas entre as duas, tornando impossível prever ou conhecer o fluxo exato de dados (Nikam, 2015).

O principal objetivo da computação baseada em RNA (neurocomputação) é desenvolver algoritmos matemáticos que permitam que as RNAs aprendam mimetizando o processamento de informações e a aquisição de conhecimento, como no cérebro humano. A analogia grosseira entre o neurônio artificial e o neurônio biológico se baseia na representação das conexões entre os nós pelos axônios e dendritos, enquanto os pesos de conexão representam as sinapses, e o limiar se relaciona à atividade na soma de sinais. Como um neurônio tem um grande número de dendritos/sinapses, ele pode receber e transferir muitos sinais simultaneamente, que podem excitar ou inibir o disparo do neurônio. Esse mecanismo simplificado de transferência de sinal constituiu a etapa fundamental do desenvolvimento precoce da neurocomputação e da operação da unidade de construção das RNAs (Basheer e Hajmeer, 2000).

Como parte do aprendizado supervisionado, as RNAs “aprendem” regras considerando os exemplos fornecidos como parte do treinamento. Uma rede neural é formada por nós conectados, denominados neurônios artificiais, ou unidades, e cada conexão transmite o sinal de um neurônio ao outro (semelhante à sinapse biológica). O neurônio possui pesos que são utilizados para ponderar as entradas recebidas e produzir uma saída. Um neurônio recebe um sinal, processa-o e então transmite uma saída para outros neurônios conectados a ele. Em geral, o sinal em um neurônio é representado por um número real e o processamento é a computação de uma função não linear da combinação de suas entradas (Chollet *et al.*, 2015; Alpaydin, 2014).

A estrutura de neurônio, juntamente com suas entradas, pesos e saída é denominada *perceptron*. Um *perceptron*, em si, é um classificador binário linear, que decide o valor de uma saída binária de acordo com uma combinação linear baseada nos pesos das entradas. Em uma rede neural artificial, os *perceptrons* se organizam em camadas, e elas se caracterizam, geralmente, em uma camada de entrada, uma camada de saída e uma ou mais camadas internas (também chamadas de camadas ocultas). Este tipo de arquitetura de rede neural é denominado *Perceptron Multicamadas* (MLP) (Alpaydin, 2014).

Em uma rede neural, as conexões entre *perceptrons* de diferentes camadas são ponderadas pelos pesos associados, que representam a força de conexão entre as unidades. As conexões são os resultados das funções de ativação dos *perceptrons* de uma camada, que são usados como entrada para outra camada. As funções de ativação podem ser a função sigmoide, ReLU (*Rectified Linear Unit*), *Softmax*, entre outras. Quando utilizadas funções não lineares, a rede neural é capaz de resolver problemas não lineares (Chollet *et al.*, 2015). No Quadro 2.1 encontram-se as principais informações sobre os tipos de neurônios.

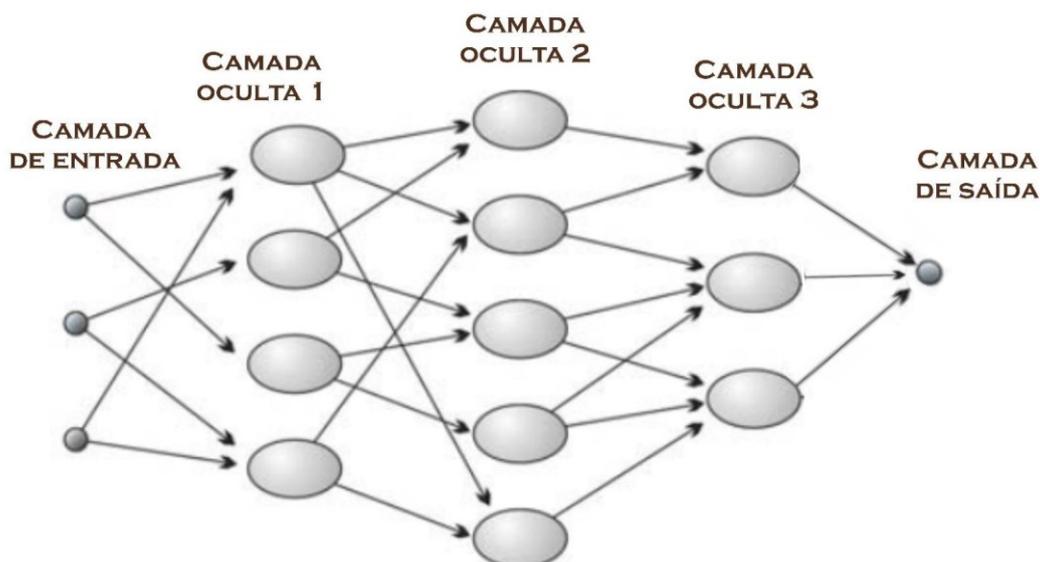
Quadro 2.1 - Principais informações dos tipos de neurônios

Tipo de Neurônio	Principais Características	Desvantagens	Vantagens
Perceptron	Usa variáveis binárias de entrada	Não possui camada oculta	Sistema de base para demais tipos de neurônios
	Introduz pesos nas variáveis de entrada		
	Produz uma única saída binária	Aprendizado de recursos pouco complexos	Utilização para dados separados linearmente
Perceptron Multicamadas	Possui camadas ocultas empilhadas	Grande consumo de memória para aprendizado	Camadas ocultas com pesos distintos permite o aprendizado de recursos mais complexos
		Função de ativação Step (linear e não suave)	
Sigmóide	Utiliza backpropagation	Redes de neurônios mais complexos	Ajuste repetidos dos pesos minimiza a diferença entre a saída real e a saída desejada
	Função de ativação sigmóide		Função de ativação estável e suave
			Permite classificação de limites ou padrões de decisões não lineares

FONTE: ALVES (2020).

As redes neurais devem ser treinadas com conjuntos de dados rotulados a fim de realizar predições. O treinamento envolve a execução das entradas do conjunto de treinamento na rede e, para cada entrada, sua saída é observada. Cada saída é comparada com o rótulo original e os valores de erro são utilizados em uma função de perda (*loss*) utilizada para atualizar os pesos dos neurônios da rede, em uma fase de retro propagação. Este processo é conhecido como *backpropagation* (Witten *et al.*, 2016). Os valores de saída são comparados com a resposta correta para calcular o valor de alguma função de erro pré-definida. Após retornar o erro para a rede, o algoritmo ajusta os pesos de cada conexão para reduzir o valor da função de erro. Depois de repetir este processo por vários ciclos de treinamento seguidos, a rede converge para um estado no qual o erro dos cálculos se torna pequeno (Goyal e Goyal, 2012). O padrão de treinamento de entrada e saída da RNA é ilustrado na Figura 2.9.

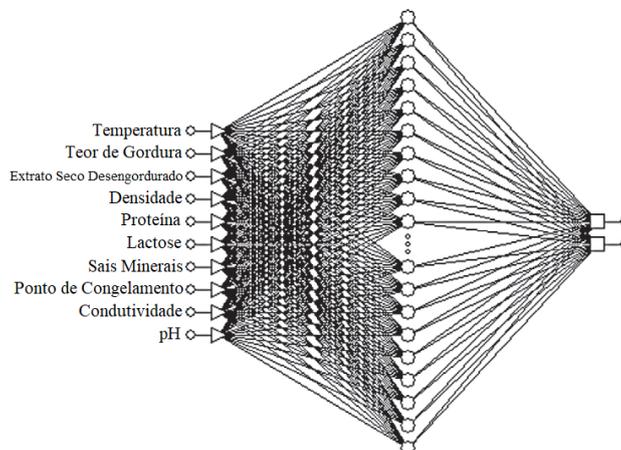
**Figura 2.9** - Representação simplificada de uma rede neural artificial mostrando camada de entrada, oculta e saída



FONTE: Adaptado de FERNEDA (2006).

Valente *et al.* (2014) encontraram resultados que demonstram que as Redes Neurais Artificiais (RNA) apresentam grande potencial de uso como ferramenta de avaliação de possíveis adulterações de leite com soro de queijo. As análises de rotina do leite de um laticínio e de amostras adulteradas foram realizadas por meio de um analisador de leite por ultrassom (*MASTER COMPLETE-AKSO®*). Para determinar a configuração da rede neural artificial foi usado o *software* SNN®. Essa técnica apresenta uma grande possibilidade de aplicação reduzindo o número de amostras que iriam ser analisadas pela metodologia oficial. Com a rede neural artificial (Fig. 2.10) foi possível classificar as amostras em leite normal ou adulterado pela adição de soro de queijo.

**Figura 2.10** - Estrutura da RNA usada na detecção de adição de soro de queijo no leite



FONTE: VALENTE *et al.* (2014).

No estudo de Alves da Rocha *et al.* (2015), foi utilizada a microscopia confocal Raman associada à rede neural artificial para avaliar e quantificar a adulteração de leite fluido pela adição de soro de queijo.

Na pesquisa realizada por Conceição *et al.* (2019), foram utilizados os métodos de regressão linear múltipla e redes neurais artificiais nos dados obtidos por FTIR – ATR para identificar adulterações em leite cru.

Já no estudo de Hernández-Ramos *et al.* (2019), as redes neurais artificiais foram empregadas para prever a contagem de células somáticas do leite de ovelha utilizado na produção de queijo.

Uma grande vantagem da rede neural artificial é que pode ser aprimorada, pois quanto maior o número de dados disponíveis melhor será a predição da rede (Valente *et al.*, 2014).

- **Redes Neurais Convolucionais**

A Rede Neural Convolucional (ConvNet, Convolutional Neural Network ou CNN) é um algoritmo de aprendizagem profunda capaz de captar dados de entrada, atribuir importância, pesos e vieses que podem ser aprendidos, e diferenciá-los. O pré-processamento exigido pela CNN é muito menor em relação a outros algoritmos de classificação. Enquanto nos métodos anteriores os filtros são construídos manualmente, as CNN são capazes de aprendê-los com o treinamento suficiente (Tavares, 2019).

As CNN são comumente aplicadas a dados de imagens, interpretados como arranjos bidimensionais. Nesse caso, as camadas convolucionais detectam bordas e outras características que representam regiões importantes da imagem. A mesma ideia pode ser aplicada para dados unidimensionais, como séries temporais e dados espectrais (Liu *et al.*, 2017).

A CNN consiste em múltiplas partes com funções diferentes, na qual é comum aplicar camadas de convolução sobre o dado de entrada. Cada camada de convolução é

composta por diversos neurônios, responsáveis pela aplicação de filtros e pesos. A combinação das entradas de um neurônio, utilizando os pesos respectivos de cada uma de suas conexões, produz uma saída passada para a camada seguinte. Os pesos atribuídos às conexões de um neurônio podem ser interpretados como uma matriz, que representa o filtro de uma convolução no domínio espacial, também conhecido como kernel ou máscara (Vargas *et al.*, 2016).

- **Validação Cruzada**

O método de validação mais utilizado atualmente é o de validação cruzada (*cross-validation*), no qual o total de casos é dividido em  $n$  grupos aleatórios, de tamanhos aproximados. Um dos subconjuntos é usado para teste com os demais, sendo esse processo repetido com cada um dos  $n$  grupos. Obtêm-se, ao final, uma taxa de acerto/erro a partir das taxas de cada um dos subconjuntos (Dantas, 2017).

Esse método é simples e amplamente utilizado para estimar erros de previsão, que funciona dividindo os dados de treinamento aleatoriamente em  $k$  partes iguais (*k-fold cross-validation*). O método de aprendizado é adequado na parte dos dados, e o erro de predição é computado na parte restante. Isso é feito para cada parte dos dados por vez, calculando-se a média das  $k$  estimativas de erro de previsão (Hastie *et al.*, 2017).

A validação cruzada é aplicada ao conjunto de treinamento, uma vez que selecionar o parâmetro de contração é parte deste processo. O objetivo do conjunto de testes é julgar o desempenho do modelo selecionado (Hastie *et al.*, 2017). Em geral, a escolha de  $k$  é 5 ou 10, mas não há regra formal (Kuhn e Johnson, 2013).

Na *k-fold cross-validation*, o conjunto de dados original é dividido aleatoriamente em  $k$  subconjuntos de mesmos tamanhos aproximados. Um modelo é inferido (treinado) usando todas as instâncias exceto as da primeira partição ( $k = 1$ ). As instâncias separadas são usadas para teste e estimativas de desempenho do modelo. O procedimento se repete com a seleção das instâncias da segunda partição ( $k = 2$ ) para o teste, enquanto o modelo é treinado com todas as instâncias remanescentes, e assim por

diante. As estimativas dos  $k$  testes são sumarizadas utilizando-se, normalmente, a média e o desvio padrão de cada execução (Kuhn e Johnson, 2013; Hastie *et al.*, 2017).

- **Matriz de Confusão**

Uma maneira comum de se descrever a predição de classes de um modelo de classificação é através da matriz de confusão. Esta matriz é uma tabela simples que indica os valores observados e os valores preditos pelo modelo. O Quadro 2.2 mostra um exemplo quando o valor esperado tem duas classes, a ocorrência e a não ocorrência de um determinado evento. A primeira linha da matriz corresponde a amostras preditas como ocorrência do evento, algumas são preditas corretamente (verdadeiros positivos, VP) e algumas são classificadas incorretamente (falsos positivos, FP). De forma análoga, a segunda linha da matriz contém os valores de predições de não ocorrência do evento, algumas predições corretas (verdadeiros negativos, VN) e outras predições incorretas (falsos negativos, FN). As células na diagonal principal denotam os casos na qual as classes são corretamente preditas, enquanto as células fora da diagonal mostram o número de erros para cada caso possível (Asseiss Neto, 2020).

**Quadro 2.2** - Matriz de confusão para um problema genérico de classificação de duas classes (ocorrência e não ocorrência de um evento)

Predição	Observação	
	Ocorrência	Não Ocorrência
Ocorrência	VP	FP
Não Ocorrência	FN	VN

FONTE: ASSEISS NETO, 2020.

Na classificação, a acurácia é a métrica que representa a precisão correta entre as classes observadas e preditas. Apesar de ser amplamente utilizada, a acurácia deve ser usada com cautela, pois é uma métrica que não leva em conta o tipo de erro que o modelo está fazendo. É considerada a taxa de acerto (Kuhn e Johnson, 2018).

No caso de uma classificação binária (quando há duas classes possíveis), algumas métricas adicionais podem ser relevantes: sensibilidade e especificidade. A sensibilidade de um modelo é a proporção em que um evento de interesse é predito corretamente para todas as amostras em que esse evento ocorre. Já a especificidade é

definida como a proporção das amostras em que há a não ocorrência de um evento para todas as amostras em que esse evento não ocorre (Asseiss Neto, 2020).

- **Curva ROC (*Receiver Operating Characteristic*)**

Análise ROC é um método gráfico para avaliação, organização e seleção de sistemas de diagnóstico e/ou predição. Gráficos ROC foram originalmente utilizados em detecção de sinais, para se avaliar a quantidade de transmissão de um sinal em um canal com ruído (Prati, 2008).

A Curva ROC é uma alternativa de avaliação do classificador, que consiste na utilização de um gráfico bidimensional, onde o eixo vertical representa a taxa de verdadeiros positivos e o eixo horizontal representa a taxa de falsos positivos obtida pelo classificador (Prati, 2008).

A curva ROC pode ser usada para uma avaliação quantitativa de modelos. A área sob a curva ROC, representada pela sigla AUC (Area Under the ROC Curve), oferece uma medida condensada de desempenho relacionada a todos os possíveis limiares de classificação. Essa medida é interessante pois pode sumarizar o desempenho da ROC para um valor escalar único que permite comparar diferentes classificadores. Em outras palavras, a AUC indica o quanto um modelo é capaz de distinguir entre as classes avaliadas pelo modelo (Fawcett, 2006).

Curvas ROC utilizam a probabilidade de classificação de cada classe e consideram um intervalo de limiares de decisão sobre essas probabilidades. A curva ROC é calculada, no intervalo crescente dos limiares, através da plotagem dos valores da taxa de verdadeiros positivos (sensibilidade) e da taxa de falsos positivos (especificidade) (Fawcett, 2006).

## CAPÍTULO 3

### ESPECTROFOTOMETRIA POR INFRAVERMELHO FTIR E CROMATOGRAFIA LIQUIDA DE ALTA EFICIÊNCIA (CLAE) NA IDENTIFICAÇÃO DE FRAUDES POR ADIÇÃO DE SORO DE QUEIJO AO LEITE.

#### 3.1 - Introdução

A sociedade demanda por alimentos seguros, saudáveis e nutritivos, que tenham qualidade em seus aspectos físico-químicos, sensoriais e microbiológicos. Neste contexto faz-se necessário combater as fraudes do leite, que prejudica toda a cadeia produtiva. Para combatê-las é preciso intensificar a fiscalização e punição, aprimorar os sistemas de fiscalização e conscientizar produtores, transportadores, empresas e consumidores em relação à importância da qualidade do leite. Paralelamente a isso é importante escolher os métodos de análise mais adequados (Mafud *et al.*, 2007).

A adulteração do leite cru com soro de queijo é um problema sério para a indústria de laticínios, especialmente em países em desenvolvimento. Esta prática fraudulenta é motivo de preocupação para agências de inspeção de alimentos e consumidores por causa da redução do valor nutricional do leite e alguns derivados, e até questões de segurança (Brandão *et al.*, 2010, Robim *et al.*, 2012, Tibola *et al.*, 2018).

A análise de dados é uma tarefa importante, pois permite a obtenção de informações úteis. Uma das técnicas utilizadas para analisar a composição do leite é denominada Espectrofotometria do infravermelho por Transformada de Fourier (FTIR), que permite extrair informações por meio da emissão da luz infravermelha nas amostras.

A FTIR é uma técnica utilizada para avaliar propriedades dos componentes alimentares e apresenta vantagens em relação às demais técnicas, como fácil execução, rapidez e custos reduzidos, devido a não necessidade do uso de reagentes e precisão na obtenção dos resultados, especialmente, quando associada à análise multivariada.

A legislação brasileira utiliza como critério para a avaliação de qualidade do leite quanto à presença de soro, a determinação quantitativa do caseinomacropeptídeo (CMP) e considera inadequado ao consumo humano o leite com concentração de CMP superiores a 75 mg/L (Brasil, 2006). O método oficial aprovado (Brasil, 2018a) para detectar e quantificar o CMP é a Cromatografia Líquida de Alta Eficiência (CLAE), que embora apresente boa sensibilidade, possui custo elevado, requer o preparo das amostras e demanda longos períodos na condução das análises, sendo muito custosa na rotina de análise de leite, além de poder apresentar resultados falso-positivos pela ação de bactérias psicrotóxicas.

O objetivo deste capítulo foi avaliar os resultados encontrados nas análises FTIR de amostras fraudadas por soro de queijo e analisar os resultados obtidos na CLAE para avaliar a presença de CMP nas amostras fraudadas com soro de queijo e verificar a correlação dos níveis de CMP com o aumento da Contagem Bacteriana.

### **3.2 Material e Métodos**

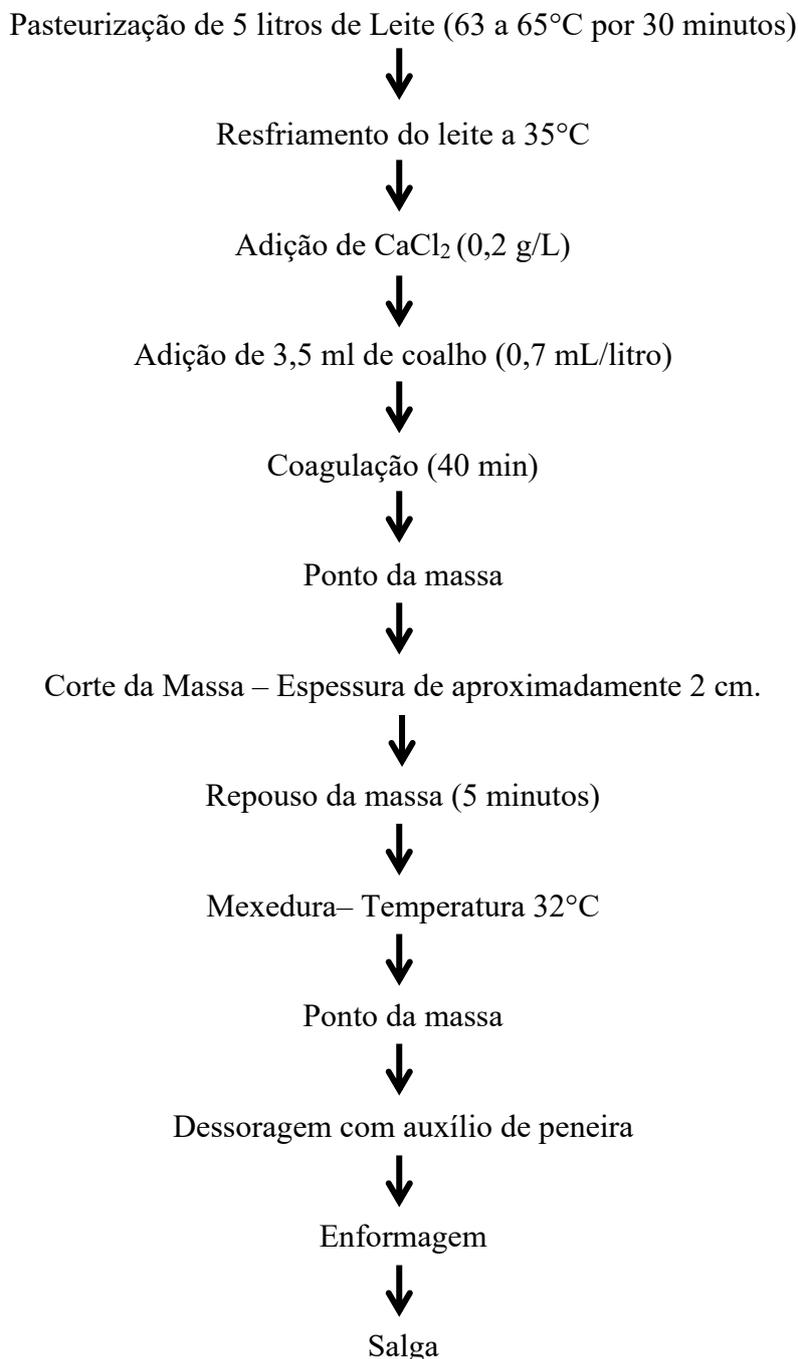
- **Local de realização do experimento**

O experimento foi realizado no Laboratório de Análise da Qualidade do Leite da Escola de Veterinária da UFMG (LabUFMG), credenciado pelo Ministério da Agricultura, Pecuária e Abastecimento (MAPA) e certificado pela norma ABNT NBR ISO/IEC 17025:2017 e no Laboratório Piloto de Laticínios (Escola de Veterinária da Universidade Federal de Minas).

- **Preparo das amostras**

Foram realizadas cinco coletas, de 15 litros de leite cru em cada, na Fazenda Experimental Professor Hélio Barbosa da Escola de Veterinária da UFMG, localizada no município de Igarapé – MG, nos meses de agosto e setembro de 2018. O leite foi coletado, após agitação, diretamente do tanque de resfriamento em galão plástico higienizado e imediatamente transportado até o LabUFMG, onde foi colocado em câmara fria de 4 a 7°C, por aproximadamente 1,5 horas até o momento da utilização.

A primeira etapa foi a fabricação do queijo Minas Frescal no Laboratório Piloto para a separação do soro do queijo, seguindo o esquema:

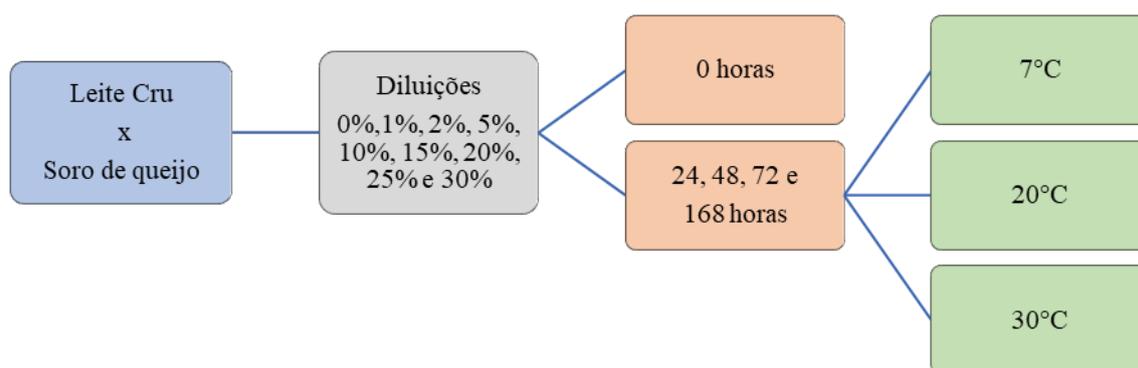


O soro de queijo produzido foi submetido ao tratamento térmico de 72°C a 75°C por 10 minutos para inativação enzimática da renina (quimosina) e impedir o processo de coagulação.

Foram feitas as diluições (0%, 1%, 2%, 5%, 10%, 15%, 20%, 25% e 30%), conforme Fig. 3.1. O leite adicionado de soro foi armazenado em frascos contendo bronopol (Advanced Instruments) para análise de composição e CCS (Contagem de Células Somáticas), e em frascos contendo azidiol<sup>1</sup> para conservação das amostras analisadas para CPP (Contagem Padrão em Placas). Para cada repetição foram 117 análises de composição, 117 para CPP e 16 para cromatografia (que foram congeladas para posterior análise) (Fig. 3.2), em cada repetição.

As racks foram montadas com as amostras e com uma amostra controle (branco) de leite sem nenhuma adição, no início de cada rack. Todos os frascos foram devidamente identificados com etiquetas à prova de água.

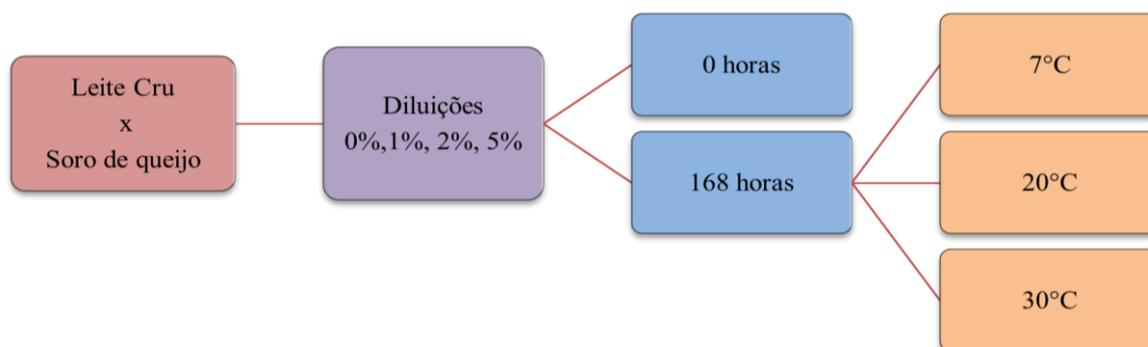
**Figura 3.1.** Esquema da preparação das amostras com diluições (0%, 1%, 2%, 5%, 10%, 15%, 20%, 25% e 30%) armazenadas a 7, 20 e 30°C nos tempos 0h, 24h, 48h, 72h e 168h.



FONTE: A autora (2021)

<sup>1</sup> Composto de Azida sódica e cloranfenicol.

**Figura 3.2** - Esquema da preparação das amostras para cromatografia com diluições (0%, 1%, 2% e 5%), armazenadas a 7, 20 e 30°C nos tempos 0h e 168h



FONTE: A autora (2021)

As amostras foram mantidas em estufa a 20°C, estufa a 30°C e refrigeradas a 7°C ±2 e analisadas nos tempos 0, 24, 48, 72 e 168 horas para composição, CCS e CPP.

Foram realizadas 5 repetições do esquema anterior (Fig. 3.1), gerando 585 amostras, já do esquema da Fig. 3.2 foram geradas 80 amostras.

As análises de Contagem Padrão em Placas foram utilizadas para avaliação das amostras, como controle de qualidade e para comparação com os resultados de Cromatografia.

- **Métodos analíticos**

- **FTIR e Citometria de Fluxo**

As análises de composição centesimal (gordura, proteína, lactose, sólidos totais, sólidos não gordurosos e caseína), nitrogênio ureico do leite e crioscopia foram realizadas no equipamento CombiScope FTIR 400®, por meio do LactoScope FTIR (Delta Instruments, Drachten, Holanda). A contagem de células somáticas (CCS) foi feita no citômetro do equipamento CombiScope FTIR 400®, por meio do SomaScope (Delta Instruments, Drachten, Holanda).

Para a realização das análises de composição, CCS e crioscopia no equipamento Delta FTIR, cada rack previamente preparada foi colocada no banho-maria à temperatura de

40°C, pelo período de 20 minutos, e homogeneizada por inversão 10 vezes com o auxílio de régua metálica.

A calibração do equipamento CombiScope™ FTIR (Delta Instruments) foi realizada por meio de 14 amostras padrão de leite cru, com diferentes faixas de composição para gordura, proteína, lactose e sólidos totais, obtidas do laboratório VALACTA® (Dairy Production Centre of Expertise, Quebec, Canadá), juntamente com os laudos contendo os resultados das análises realizadas no laboratório de origem.

Um total de 10 variáveis foram obtidas a partir de calibrações no equipamento que, a partir de um modelo de regressão linear múltipla (*Multiple Linear Regression*, MLR) interno ao equipamento, considera a absorção da luz por regiões pré-determinadas. O modelo de regressão é capaz de estimar numericamente as seguintes *features*: gordura (*fat*), proteína (*protein*), lactose, sólidos totais (*total solids*), sólidos não-gordurosos (*solids non-fat*, SNF), caseína (*casein*) e nitrogênio ureico do leite (*milk urea nitrogen*, MUN). Outros três valores também obtidos incluíram a contagem de células somáticas (*Cells*) por citometria de fluxo, o valor do ponto de congelamento (*FrzPoint*) e um valor de controle de qualidade da amostra (*QValue*). É a partir do espectro em si que o equipamento executa um método de regressão para consolidar os dez atributos dos componentes descritos.

- **Contagem Padrão em Placas (CPP)**

A Contagem Padrão em Placas foi efetuada em equipamento eletrônico BactoScan FC® (Foss Electric A/S, Hillerod, Dinamarca) que tem por princípio a citometria de fluxo. O BactoScan™ FC é um equipamento que conta e caracteriza partículas e células, permitindo análises precisas e instantâneas das bactérias no leite. Uma suspensão de células é corada e flui através de um tubo capilar. Os aglomerados de bactérias são separados em bactérias individuais. Durante o período de incubação, as bactérias são coradas com um meio de coloração específico para o DNA. Para a mensuração, o material da amostra é exposto a um feixe de luz gerado por um laser, que resulta na emissão da fluorescência. O dispositivo eletrônico conta os pulsos e os exibe em um

diagrama de análise de altura de pulso no monitor do computador (Foss, 2016).

O equipamento foi calibrado de acordo com o método de Contagem Padrão em Placa de referência, definido pela norma Internacional (IDF 100B: 1991), permitindo a conversão da contagem bacteriana individual em UFC/mL, em atendimento às Instruções Normativas (Brasil, 2018b).

#### ▪ **Cromatografia Líquida de Alta Eficiência (CLAE)**

A metodologia foi realizada conforme o Manual de Métodos Oficiais para Análises de Alimentos de Origem Animal (Brasil, 2018a).

As amostras foram descongeladas em temperatura ambiente e, logo após, 10 mL de cada amostra foi pipetada em um béquer, e adicionada de 5 mL de ácido tricloroacético a 24%, gota a gota e sob agitação constante. Após repouso de 60 min em temperatura ambiente, as amostras foram filtradas em papel qualitativo (WHATMAN®, nº 5).

As amostras filtradas foram injetadas no cromatógrafo (Shimadzu), composto dos seguintes módulos: bomba LC 10AT, injetor automático SCL-10A, detector-UV-visível SPD 10AV. A coluna utilizada foi a Zorbax GF-250 da Agilent, AG884973-901, 250 mm de comprimento, 9,4 mm de diâmetro interno e tamanho de partícula de 4 µm.

A fase móvel foi uma solução tampão fosfato pH 6,0 (1,74g de hidrogenofosfato de potássio; 12,37g de dihidrogenofosfato de potássio e 21,41g de sulfato de sódio em, aproximadamente, 700 mL de água deionizada destilada), com fluxo de 1,5 mL/min. O pH da solução foi ajustado quando necessário para 6,0 utilizando solução de ácido fosfórico a 3 mol/L e solução de hidróxido de potássio a 3 mol/L. A solução foi transferida para um balão volumétrico de 1000 mL e o volume completado com água deionizada destilada e filtrada em membrana de 0,45µm. Antes do uso, a solução foi degaseificada em banho de ultrassom durante 30 min.

Antes de iniciar as análises das amostras, em balões de 25 mL foram colocados respectivamente 150, 300, 450, 600, 750 e 900 microlitros de CMP liofilizado padrão

reconstituído com água deionizada para ser preparada a curva padrão. Posteriormente foi adicionado 5 mL de ácido tricloroacético para, ao final, obter padrões conhecidos de 0, 15, 30, 45, 60, 75 e 90 mg/L. Dessa forma, foi construído um gráfico de concentração (mg/L) versus a altura efetiva do pico e cálculo da área e assim calculada a regressão linear da curva aceitando valores de  $R > 0,95$ . Após a preparação da curva padrão, foi feita a identificação dos picos com o mesmo tempo de retenção do CMP padrão e as concentrações presentes nas amostras foram calculadas em mg/L, com base na curva de calibração. Esse cálculo é feito automaticamente pelo equipamento.

- **Análise Estatística**

Os resultados das análises de composição, gordura, proteína, lactose, sólidos totais, sólidos não gordurosos (SNF), nitrogênio ureico do leite (MUN), caseína e Ponto de Congelamento (PC) realizadas no equipamento FTIR, foram comparados pelo teste de Student-Newman Keuls (SNK) ao nível significância de 5% empregando-se o software SPSS (Versão 22; IBM). Os resultados de Índice de CMP e Contagem Bacteriana Total foram comparados pelo teste de Tukey ao nível significância de 5% empregando-se o software SPSS (Versão 22; IBM).

### **3.3 - Resultados e Discussão**

Durante o processo de leitura do espectro infravermelho, o equipamento FTIR executa uma série de cálculos gerados como valores numéricos para diferentes componentes do leite. De acordo com a documentação do equipamento, os cálculos são baseados em um modelo de Regressão Linear Múltipla (*Multiple Linear Regression - MLR*) que considera a absorvância da energia da luz pela amostra para regiões de comprimento de onda específicas. A informação extraída depende das calibrações do equipamento para a concentração dos componentes do leite (leitura do percentual de gordura, proteína, lactose, sólidos totais, sólidos não gordurosos (Solids Non Fat - SNF), caseína e nitrogênio ureico do leite (Milk Urea Nitrogen - MUN), conforme ilustrado na Figura 3.3.

**Figura 3.3** Componentes numéricos gerados pelo equipamento FTIR em formato .CSV de amostras aleatórias

Cells	Q-Value	Fat	Protein	Lactose	Solids	FFA	Citrate	FPD	SNF	NPN-CU	Casein	Porcentagem soro
209	98.00	4.13	3.33	4.65	13.08	50	1339	0.548	9.23	14.0	2.64	
191	90.31	4.01	3.45	4.62	13.04	25>	1224>	0.542	9.36	12.0	2.77	0%
224	92.09	3.99	3.46	4.66	13.08	28>	1225>	0.549	9.42	12.2	2.76	1%
241	90.91	3.96	3.45	4.68	13.05	24>	1218>	0.550	9.42	11.7	2.76	2%
224	88.99	3.88	3.38	4.70	12.92	27>	1227>	0.549	9.37	12.0	2.70	5%
227	88.27	3.75	3.27	4.73	12.71	24>	1240>	0.549	9.28	13.6	2.60	10%
239	91.45	3.59	3.17	4.79	12.50	26>	1267>	0.550	9.22	11.9	2.50	15%
239	92.02	3.45	3.05	4.83	12.28	25>	1269>	0.550	9.13	11.6	2.41	20%
248	90.89	3.32	2.95	4.87	12.08	26>	1295>	0.550	9.05	12.8	2.31	25%
243	87.22	3.18	2.85	4.96	11.92	26>	1308>	0.554	9.04	11.1	2.23	30%

FONTE: A autora (2021).

O quadro 3.1 mostra os limites de composição dos componentes do leite de acordo com a Instrução Normativa nº 76 de 26 de novembro de 2018 do Ministério da Agricultura, Pecuária e Abastecimento.

**Quadro 3.1** – Requisitos físico-químicos para o leite cru refrigerado de acordo com a Instrução Normativa nº 76 de 2018

Requisitos	Limites
Teor de Gordura	Mínimo 3,0g/100g
Proteína Total	Mínimo 2,9g/100g
Lactose	Mínimo 4,3g/100g
Sólidos não gordurosos	Mínimo 8,4g/100g
Sólidos totais	Mínimo 11,4g/100g
Índice crioscópico	Entre -0,530°H e -0,555°H

FONTE: BRASIL (2018b).

A adição de soro de queijo ao leite em diferentes concentrações, não causou mudanças significativas na composição do leite quando se observa as médias, observando os limites estabelecidos pela legislação. Os valores médios e coeficientes de variação (CV) dos parâmetros analisados nas amostras adicionadas de soro estão apresentados na Tabela 3.1. Na Figura 3.4 são apresentadas as médias em relação a concentração de soro adicionado.

**Tabela 3.1** – Resultados das análises de composição por FTIR das amostras de leite adicionadas de soro nas concentrações 0, 1, 2, 5, 10, 1, 10, 25 e 30% (n = 585)

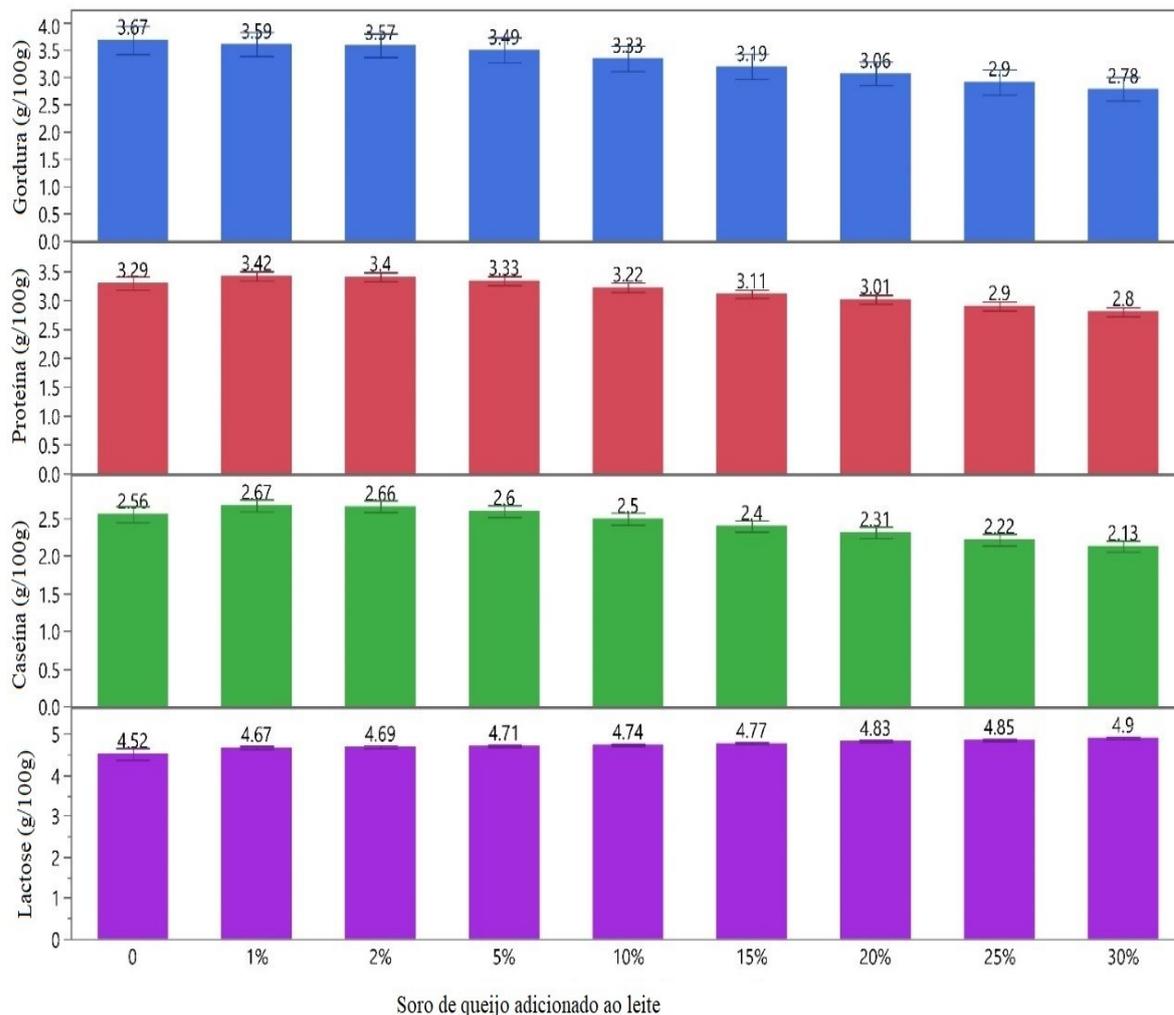
<b>Parâmetros</b>	<b>Média</b>	<b>Mínimo</b>	<b>Máximo</b>	<b>Desv. Padrão</b>	<b>CV (%)</b>
<b>Gordura (g/100g)</b>	3,28	2,36	4,09	0,371	13,8
<b>Proteína (g/100g)</b>	3,18	2,65	3,52	0,238	5,7
<b>Lactose (g/100g)</b>	4,75	4,60	4,97	0,085	0,7
<b>Sólidos não gordurosos (g/100g)</b>	9,14	8,67	9,48	0,196	3,8
<b>Sól. Totais (g/100g)</b>	12,18	11,03	13,21	0,520	27,1
<b>Ponto de Congelamento (°H)</b>	-0,544	-0,528	-0,555	0,003	0
<b>Nitrogênio não Protéico (mg/dL)</b>	11,314	6,2	19	2,02	4,09
<b>Caseína (g/100g)</b>	2,46	1,99	2,80	0,21	0,045

FONTE: A autora (2021).

O coeficiente de variação (CV) é usado para analisar a dispersão em termos relativos do valor médio de uma variável. Dos componentes do leite, a Gordura e os Sólidos Totais apresentaram maior CV, seguidos da proteína, sólidos não gordurosos e lactose. Segundo Linn (1988) o teor de gordura varia muito no leite, pois é fortemente influenciado por fatores genéticos, fisiológicos e ambientais.

A adição de soro de queijo ao leite cru provocou a redução da concentração dos componentes, exceto lactose, com o aumento da concentração correlacionado com a quantidade crescente de soro de queijo adicionado ( $R^2 = 0,60$ ;  $p < 0,001$ ). Nenhuma diferença foi encontrada para índice crioscópico nos diferentes tratamentos.

**Figura 3.4** – Médias das concentrações dos componentes do leite em relação a concentração de soro de queijo adicionado ao leite



FONTE: A autora (2021).

O teste de Student-Newman-Keuls (SNK) é um teste de comparações múltiplas usado para detectar diferenças significativas entre as médias. O teste consiste em declarar, toda vez que o valor absoluto da diferença de duas médias que abrangem as médias ordenadas for igual ou maior que a diferença crítica, apresentada pelo valor de  $\alpha$ . Neste caso,  $\alpha$  foi considerado como sendo 0,05. A Tabela 3.2 mostra os resultados do teste SNK para as variáveis de composição do leite, apresentando as médias por concentração de soro adicionado.

**Tabela 3.2** – Valores médios\* dos componentes do leite em relação a concentração de soro de queijo quando aplicado o Teste de Student-Newman-Keuls (Valor de  $\alpha = 0,05$ )

Soro (%)	Gordura (g/100g)	Proteína (g/100g)	Lactose (g/100g)	SNG (g/100g)	Sólidos Totais (g/100g)	Ponto de congelamento (°H)	Caseína (g/100g)
0	3,62 <sup>a</sup>	3,40 <sup>a</sup>	4,67 <sup>a</sup>	9,34 <sup>a</sup>	12,69 <sup>a</sup>	- 0,544 <sup>abc</sup>	2,69 <sup>a</sup>
1	3,59 <sup>a</sup>	3,41 <sup>a</sup>	4,67 <sup>a</sup>	9,32 <sup>a</sup>	12,65 <sup>a</sup>	- 0,544 <sup>abc</sup>	2,67 <sup>a</sup>
2	3,57 <sup>ab</sup>	3,40 <sup>a</sup>	4,69 <sup>b</sup>	9,32 <sup>a</sup>	12,63 <sup>a</sup>	- 0,545 <sup>c</sup>	2,65 <sup>a</sup>
5	3,49 <sup>b</sup>	3,33 <sup>b</sup>	4,70 <sup>c</sup>	9,27 <sup>b</sup>	12,49 <sup>b</sup>	- 0,545 <sup>bc</sup>	2,59 <sup>b</sup>
10	3,33 <sup>c</sup>	3,22 <sup>c</sup>	4,73 <sup>d</sup>	9,16 <sup>c</sup>	12,25 <sup>c</sup>	- 0,544 <sup>abc</sup>	2,49 <sup>c</sup>
15	3,19 <sup>d</sup>	3,10 <sup>d</sup>	4,77 <sup>e</sup>	9,07 <sup>d</sup>	12,02 <sup>d</sup>	- 0,543 <sup>a</sup>	2,39 <sup>d</sup>
20	3,06 <sup>e</sup>	3,01 <sup>e</sup>	4,82 <sup>f</sup>	9,02 <sup>e</sup>	11,85 <sup>e</sup>	- 0,544 <sup>abc</sup>	2,31 <sup>e</sup>
25	2,91 <sup>f</sup>	2,89 <sup>f</sup>	4,85 <sup>g</sup>	8,92 <sup>f</sup>	11,60 <sup>f</sup>	- 0,543 <sup>ab</sup>	2,21 <sup>f</sup>
30	2,77 <sup>g</sup>	2,80 <sup>g</sup>	4,90 <sup>h</sup>	8,86 <sup>g</sup>	11,42 <sup>g</sup>	- 0,544 <sup>abc</sup>	2,13 <sup>g</sup>

\* Valores médios para 5 repetições. Legenda: Médias seguidas de letras distintas diferem entre si pelo teste de SNK, em nível de significância de 5%.

FONTE: A autora (2021).

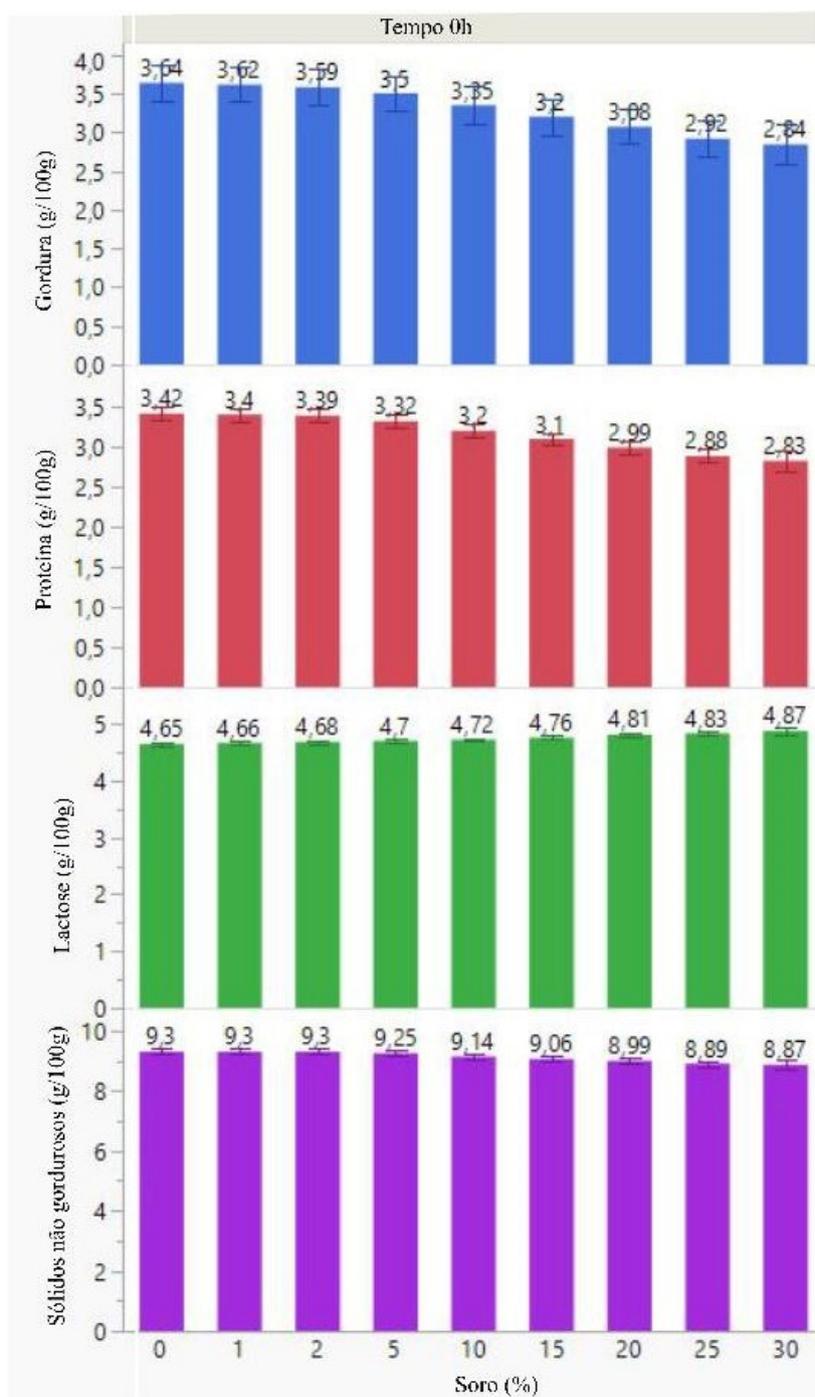
A adição de soro de queijo ao leite diminui os níveis de alguns componentes, como a gordura e proteína, mas dependendo dos teores iniciais dos componentes e da quantidade de soro adicionada, os níveis dos constituintes podem estar dentro dos padrões legais (Conde *et al.*, 2020).

A adição de soro até 2% não alterou os resultados da composição significativamente, como pode ser observado nas variáveis gordura, proteína, sólidos não gordurosos, lactose e sólidos totais.

No trabalho realizado por Conde *et al.* (2020) foi identificado que as variáveis mais afetadas pela adição de soro de queijo ao leite foram a quantidade de gordura e a densidade. Eles relataram que, embora o conteúdo de gordura tenha sido o mais significativo, a importância relativa de cada variável foi muito próxima quando o soro é adicionado em baixas proporções. Isso dificulta a identificação da fraude de soro de queijo adicionado ao leite.

A Figura 3.5 mostra a amplitude dos resultados das variáveis Gordura, Proteína, Lactose e Sólidos Não Gordurosos nas concentrações 0%, 1%, 2%, 5%, 10%, 15%, 20%, 25% e 30% no tempo 0h. As barras apresentadas no gráfico correspondem ao desvio padrão.

**Figura 3.5** – Teores (g/100g) de gordura, proteína, lactose e sólidos não gordurosos após adição de soro obtido da fabricação de queijo ao leite nas concentrações de 0%, 1%, 2%, 5%, 10%, 15%, 20%, 25% e 30%



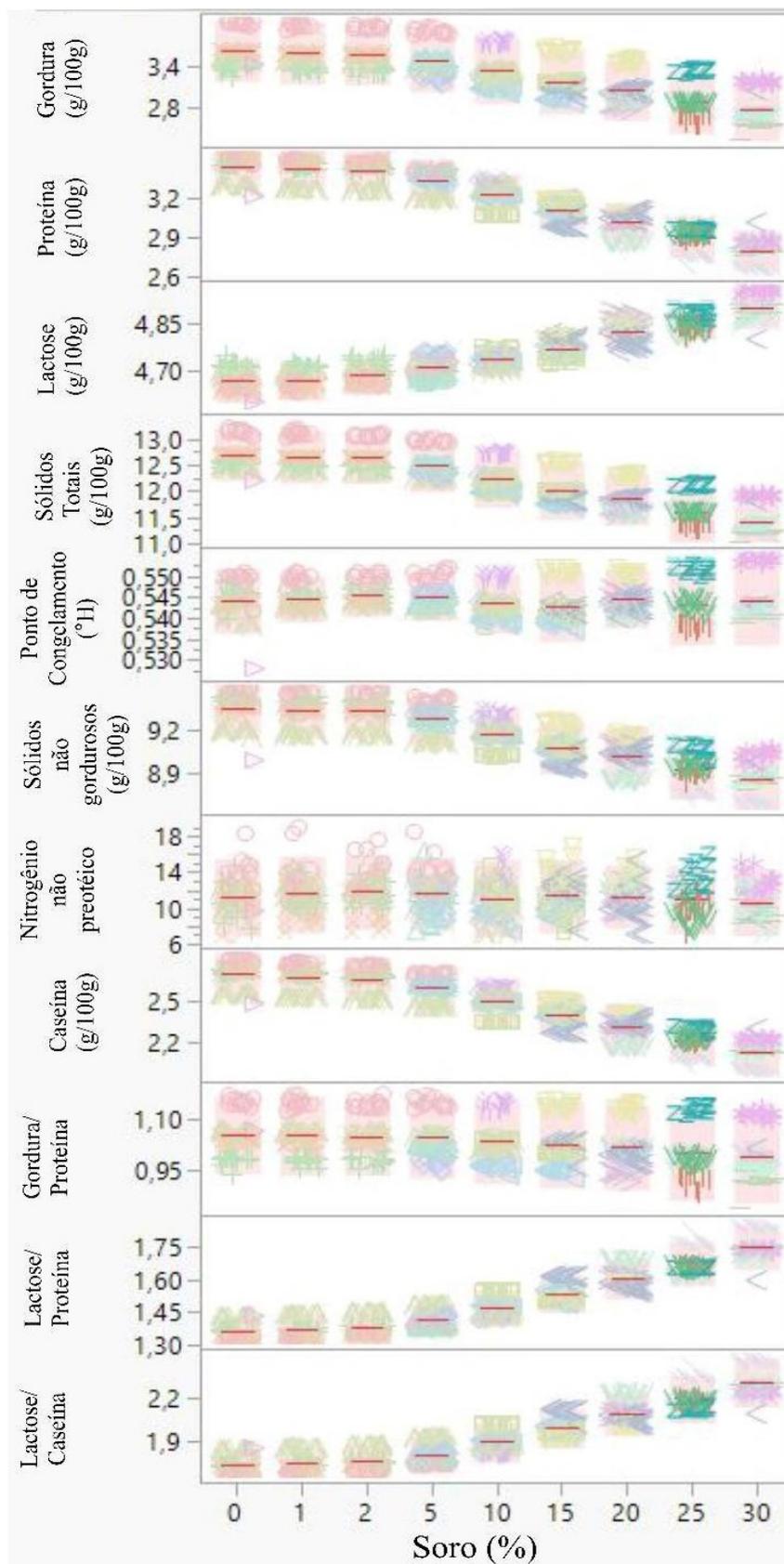
FONTE: A autora (2021).

Os teores de gordura e proteína se mantiveram dentro dos padrões legais exigidos pela legislação brasileira até a adição de 20% de soro, ocorrendo uma leve queda na adição de 25% e 30% de soro. O teor de lactose aumentou e de sólidos não gordurosos diminuiu com o aumento da concentração de adição de soro ao leite, porém permaneceram dentro dos padrões da legislação. Esse fato destaca a importância de análises complementares, para detectar possíveis fraudes, visto que o leite pode ser fraudado com soro e ainda apresentar resultados de composição dentro dos limites estabelecidos.

As variáveis mais afetadas quando o soro é adicionado ao leite são proteína e gordura. Embora o conteúdo de gordura seja o mais significativo no leite, a importância relativa de cada variável é muito próxima quando o soro é adicionado. Isso frequentemente evita a fraude de soro de leite sendo identificada. O soro de leite bovino tem aproximadamente 0,5% de gordura, justificando a queda acentuada dos teores quando o soro é adicionado ao leite (Antunes, 2003). O soro contém água, lactose e sais, apresentando densidade e índice crioscópico muito próximos ao leite, tornando a detecção mais laboriosa e complexa (Tronco, 2008). No entanto, a mudança no teor de gordura nem sempre permite que a fraude pela adição de soro de queijo seja identificada; em vez disso, as variações que ocorrem em todas as variáveis são mais reveladoras.

A Figura 3.6 ilustra o Gráfico de Dispersão para as variáveis de composição do leite. Os dados foram plotados separadamente, assim a relação se torna mais clara. Esse gráfico permite enxergar anormalidades na composição, sendo possível analisar se a variável independente “concentração de soro” tem impacto real na variável dependente “composição”.

**Figura 3.6** – Gráfico de dispersão das variáveis de composição do leite versus a concentração de soro



FONTE: A autora (2021).

O Gráfico de dispersão trata-se de uma representação gráfica que analisa a relação entre duas variáveis quantitativas — uma de causa e uma de efeito. Esse tipo de diagrama traz números simultâneos das duas variáveis, deixando visível se o que acontece em uma variável causou interferência na outra. Nesse modelo pode se observar a correlação positiva entre a concentração de soro e quantidade de componentes do leite, quanto mais adiciona-se soro, maior é a alteração na composição, observando que a maior variação foi encontrada quando se adiciona 25% e 30% de soro em todas as variáveis.

Em estudo realizado por Leite (2006) concluiu-se que amostras com conservantes armazenadas a 30°C podem ser analisadas, quanto à contagem de células somáticas e composição centesimal até o quinto e quarto dia, respectivamente, após a coleta e se mantidas sob refrigeração até dez dias.

Em testes realizados por Cortez *et al.* (2010), foi analisado que os valores de proteína para a adição de 30% de soro de queijo, apesar de apresentarem uma tendência à redução, permaneceram dentro do preconizado pela legislação, tendo o valor mínimo de 3,15%. Para a concentração de lactose, foi observado a mesma redução, porém de forma menos marcante. Já o valor de gordura, foi de 2,7% na adição de 15% de soro de queijo, sendo considerado fora dos padrões da legislação.

A quantificação do CMP é o indicador atualmente mais recomendável para se detectar a adição de soro de queijo ao leite. De acordo com a IN n°69 (Brasil, 2006), leites que apresentam índices de CMP até 30 mg/L podem ser destinados ao consumo direto, enquanto leites com índices entre 30 até 75 mg/L devem ser destinados à produção de derivados lácteos. Leite com valores de CMP acima de 75 mg/L devem ser destinados à alimentação animal, à indústria química em geral ou a outro destino a ser avaliado tecnicamente, pelo Departamento de Inspeção de Produtos de Origem Animal (DIPOA).

Leites que apresentam concentração de CMP acima de 30 mg/L já podem ser considerado leite fraudado ou de má qualidade. Níveis de CMP acima de 75mg/L sugerem uma adição fraudulenta de soro de queijo ou extrema deterioração do produto causada por micro-organismos psicrotróficos (Friedrich *et al.*, 2007).

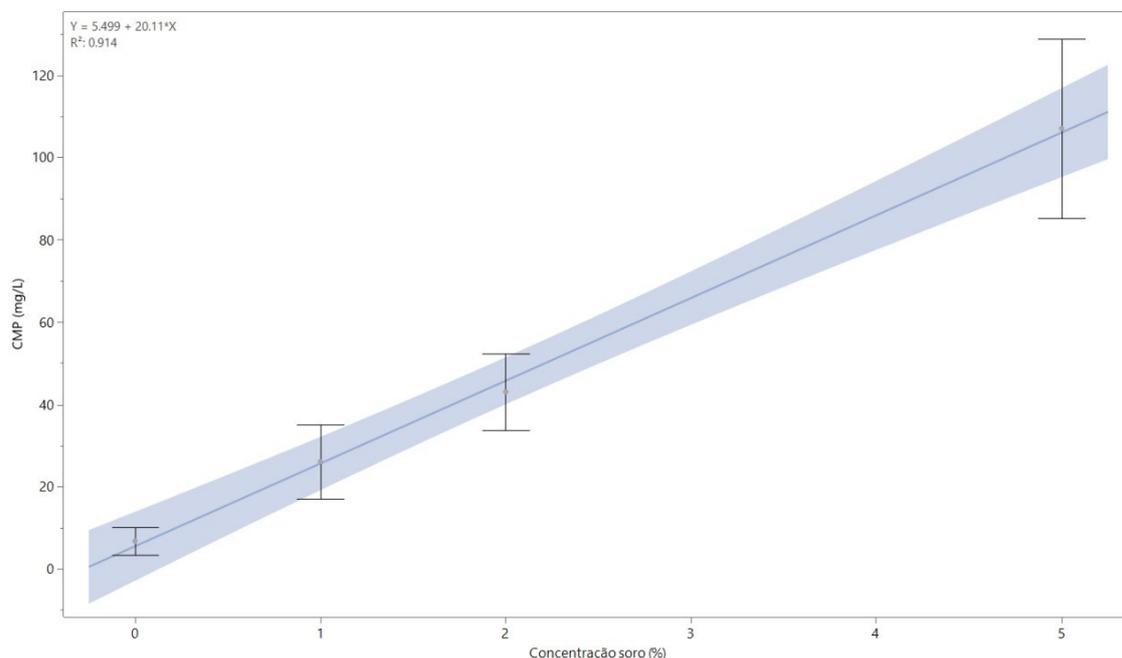
Ao adicionar percentuais crescentes de soro de queijo ao leite e analisar no tempo 0h, obteve-se concentrações também crescentes de CMP, conforme apresentado na Tabela 3.3, o que também foi observado por Oliveira *et al.* (2009) em seus estudos. Quando ao leite foram adicionadas diferentes concentrações de soro de queijo, observou-se que houve aumento do CMP e da área do pico característico do CMP, à medida que foi aumentada a concentração de soro na amostra. O aumento do tempo de incubação das amostras de leite foi significativo no resultado.

**Tabela 3.3** - Resultados médios de CMP (mg/L), obtidos por CLAE em amostras de leite adicionadas de soro com diferença de concentração, tempo e temperatura

Tratamentos	Concentração de CMP (mg/L)			
	Porcentagem de Soro adicionado			
	0%	1%	2%	5%
<b>Tempo 0h</b>	6,76	26,06	43,04	107,06
<b>Tempo 168 h e temperatura 5°C</b>	311,01	702,11	813,26	769,13
<b>Tempo 168 h e temperatura 20°C</b>	103,26	111,79	93,66	122,77
<b>Tempo 168 h e temperatura 30°C</b>	99,48	138,71	100,86	133,08

FONTE: A autora (2021).

Houve correlação direta entre a concentração de soro de queijo e o índice de CMP. Quanto maior a concentração de soro de queijo, maior o índice de CMP. Foi construída uma curva correlacionando o Índice de CMP em função da concentração de soro, apresentada na Figura 3.7. Observa-se um alto grau de correlação ( $R^2 = 0,914$ ) entre as variáveis.

**Figura 3.7** Relação e correlação entre a concentração de soro (0, 1%, 2%, 5%) e o Índice de CMP

FONTE: A autora (2021).

Considerando somente o tratamento 168 horas, a contagem bacteriana (UFClog/mL) e o CMP não foram afetados pela concentração de soro, conforme observado na tabela 3.4.

**Tabela 3.4** – Valores médios da Contagem Bacteriana e do índice de CMP em relação a concentração de soro de queijo (0, 1%, 2%, 5%) quando aplicado o Teste de Tukey (Valor de  $\alpha = 0,05$ )

Soro (%)	N	Contagem Bacteriana (UFClog/mL)	CMP (mg/L)
0	14	5,91010 <sup>a</sup>	183,48250 <sup>a</sup>
1	15	6,03970 <sup>a</sup>	317,53820 <sup>a</sup>
2	15	5,93895 <sup>a</sup>	335,92767 <sup>a</sup>
5	15	5,95624 <sup>a</sup>	341,66027 <sup>a</sup>

Legenda: Médias seguidas de letras iguais não diferem entre si pelo teste de Tukey HSD, em nível de significância de 5%.

FONTE: A autora (2021).

Entretanto, o Índice de CMP e a Contagem Bacteriana sofreram efeito da temperatura de incubação, conforme pode ser observado na tabela 3.5

**Tabela 3.5** – Valores médios da Contagem Bacteriana e do índice de CMP em relação a temperatura de incubação (7°C, 20°C e 30°C) quando aplicado o Teste de Tukey (Valor de  $\alpha = 0,05$ )

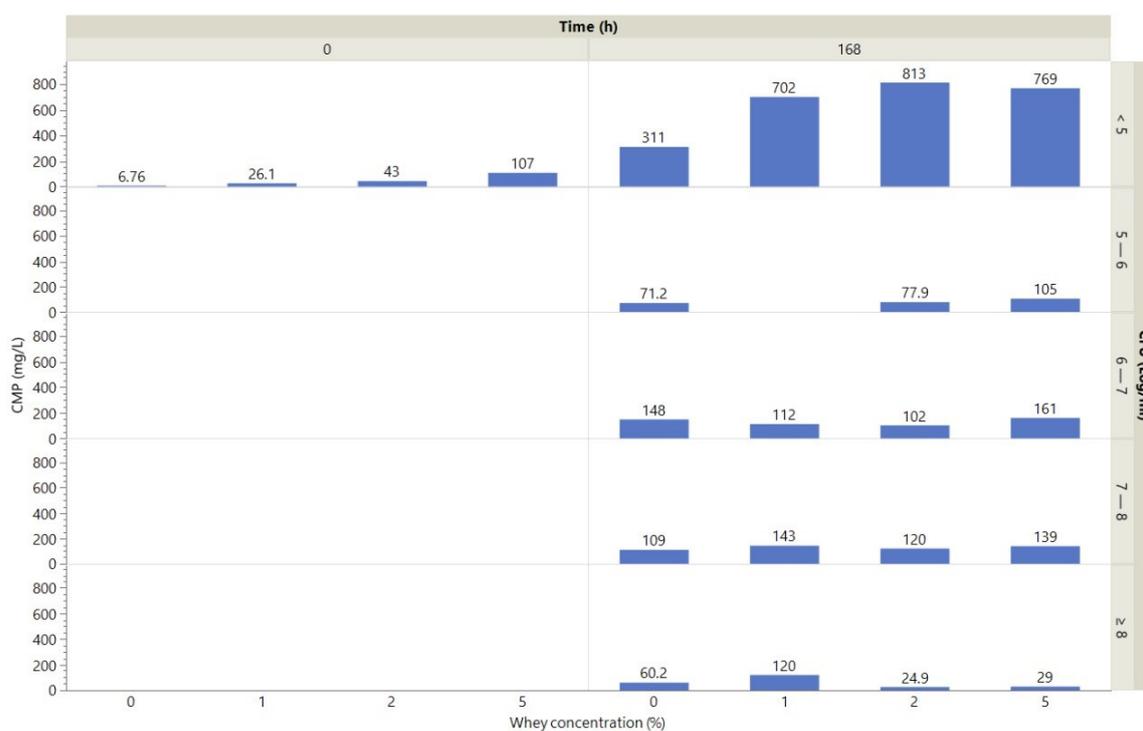
Temperatura (°C)	N	Contagem Bacteriana (LogUFC)	CMP (mg/L)
7	20	4,18563 <sup>a</sup>	648,87745 <sup>a</sup>
20	19	6,26469 <sup>b</sup>	113,54826 <sup>b</sup>
30	20	7,45129 <sup>c</sup>	118,03405 <sup>b</sup>

Legenda: Médias seguidas de letras distintas diferem entre si pelo teste de Tukey HSD, em nível de significância de 5%.

FONTE: A autora (2021).

A proteólise causada por micro-organismos no leite, principalmente *Pseudomonas sp.*, proporciona a degradação parcial da caseína liberando peptídeos contendo ácido siálico, que influenciam nos métodos de detecção de adição de soro ao leite (Firmino *et al.*, 2010). Espera-se que quanto maior o tempo e a temperatura de armazenamento mais evidente fica a proteólise. A Figura 3.8 mostra relação entre os níveis de CMP, concentração de soro (0, 1%, 2% e 5%) e Contagem Padrão em Placas definida por UFC (log/ml) para o tempo 0h e 168h nas temperaturas 7°C, 20°C e 30°C.

**Figura 3.8** Relação entre os níveis de CMP, concentração de soro (0, 1%, 2% e 5%) e Contagem Padrão em Placas definida por UFC (log/ml)



FONTE: A autora (2021).

O índice de CMP foi alto com temperatura de 7°C armazenado por 168h. Isso ocorreu muito provavelmente pela proteólise, mas com menos proporção e essa proteólise nesse caso específico foi por causa da liberação de CMP a partir da micela de caseína além do CMP do soro que já existia, ou seja, pseudo-CMP.

Às temperaturas de armazenamento de 20°C e 30°C, as baixas concentrações de CMP, ocorreram possivelmente por proteólise extensa que leva a quebra de polipeptídios, inclusive o CMP. A diminuição significativa das concentrações de CMP do tempo 168h, pode ser atribuída à alta concentração e/ou atividade das enzimas produzidas durante este período e que, devido à baixa especificidade, passaram, ao longo do tempo, a degradar o próprio CMP, o que vai de encontro a resultados obtidos por outros autores (Oliveira, 2009; Firmino *et al.*, 2010).

Houve proteólise microbiana nas amostras armazenadas a 7°C, e em temperaturas maiores com crescimento microbiano, nesse ponto já tem uma extensa proteólise inclusive de CMP formado, ou seja, o valor de CMP pode ser afetado em armazenamento prolongado.

Lasmar (2007) trabalhou com porcentagem de soro adicionado em leite de 0, 2, 5, 10 e 20% utilizando conservante bronopol e amostras congeladas e não encontrou diferenças significativas entre os tratamentos realizados, indicando que amostras adicionadas de bronopol e armazenadas a 7°C e 30°C por períodos de até 8 dias, podem ser analisadas por CLAE para detecção de soro de queijo em leite sem que ocorra alteração dos resultados esperados. Estes resultados indicam que no estudo citado provavelmente não houve crescimento bacteriano nem produção de enzimas proteolíticas, o que causaria um aumento nos valores encontrados, gerando resultados falso-positivos, ou seja, resultados que poderiam ser negativos, mas estão dando positivo pela ação de enzimas proteolíticas e não pela adição de soro.

Em estudos realizados por Magalhães (2008), foi possível observar que não houve aumento da área do pico relativo ao CMP em função do tempo de incubação até dois dias de incubação. No período entre dois dias e quatro dias de incubação pode-se observar um aumento rápido da área do pico relativo ao CMP. A área

relativa ao CMP continuou aumentando até os 6 dias de incubação, porém com menor intensidade. Após 6 dias de incubação a área relativa ao pico do CMP tendeu a decrescer, possivelmente devido à extensa proteólise que estava ocorrendo. Observa-se, portanto, que a concentração de CMP no leite pode determinar a extensão de sua proteólise.

Prata (2009) relatou que o teor de caseinomacropéptido em leite e derivados tem assumido características de importante marcador das ações proteolíticas sofridas, seja na condição de matéria-prima ou mesmo no monitoramento de etapas importantes do processo tecnológico do leite.

Andrade *et al.* (2010) analisaram amostras de bebidas lácteas fermentadas com adição de soro de queijo na proporção de (0, 10, 20 e 40%) respectivamente. Foi observada uma alta correlação ( $p < 0,0001$ ) entre os resultados de porcentagem de soro e índices de CMP, ou seja, o índice de CMP acompanha proporcionalmente a porcentagem de soro lácteo adicionado.

Villanoeva *et al.* (2014), avaliaram a interferência da temperatura (21°C, 6°C, e -12°C) e do tempo de armazenamento (dia da amostragem, 30, 60, 90 e 120) do leite UHT em relação ao índice de CMP. Um aumento progressivo do índice CMP em leite UHT, medido por meio de CLAE, foi observado com o aumento da temperatura e tempo de armazenamento. As amostras armazenadas a -12°C apresentaram menos alterações nos níveis de CMP, confirmando que, se o armazenamento for necessário, o congelamento de amostras de leite UHT para uma análise posterior é a melhor opção, quando não for possível a análise imediatamente após produção.

Esses resultados demonstram a importância das condições de estocagem do leite cru exigidas pela IN 76/2018 (Brasil, 2018b), que é de no máximo 7°C admitindo-se, excepcionalmente, até 9°C no recebimento e de 4°C até a pasteurização. Leite em temperatura mais elevada, ou estocado por tempo maior, e com presença de bactérias psicrotróficas, apresenta proteólise elevada, caracterizado pela presença de CMP. Dessa forma, o resultado do Índice de CMP, como previsto na IN 69/2006 (Brasil, 2006), é um

indicador da intensidade de proteólise das proteínas do leite, ou então da adição fraudulenta do leite com soro de queijo.

## CAPÍTULO 4

### APRENDIZADO DE MÁQUINA PARA DETECÇÃO DE FRAUDE POR ADIÇÃO DE SORO OBTIDO DA FABRICAÇÃO DE QUEIJO AO LEITE.

#### 4.1 Introdução

A adulteração do leite cru com o soro do queijo é um problema sério para a indústria de laticínios, especialmente nos países em desenvolvimento. Essa prática fraudulenta é motivo de preocupação para agências de fiscalização de alimentos e consumidores devido à redução do valor nutricional do leite e alguns derivados, e até mesmo por questões de segurança (Brandão *et al.* 2010; Robim *et al.* 2012; Tibola *et al.* 2018).

No entanto, métodos analíticos adequados para investigar essa fraude geralmente são caros, demorados, com baixo rendimento analítico com precisão e exatidão limitadas (De La Fuente e Juárez 2005; Carvalho *et al.* 2015; Tibola *et al.* 2018). Um dos métodos mais conhecidos para detecção de soro de queijo, baseado na quantificação do caseinomacropéptido (CMP) por cromatografia líquida de alta eficiência (CLAE) é um assunto controverso por causa de questões relacionadas a resultados falso-positivos e precisão (Lenardon *et al.* 2017; Alves *et al.* 2018; Raymundo *et al.* 2018; Lobato *et al.* 2020).

Os avanços atuais com o uso de aprendizado de máquina em métodos analíticos podem ser uma resposta a este problema, uma vez que técnicas novas e inovadoras poderiam ser criadas e incorporadas a uma rotina de laboratório (De La Fuente e Juárez 2005).

As técnicas de Espectroscopia no infravermelho (FTIR) tem capacidade de produzir um grande volume de dados. A associação desta tecnologia com algoritmos de aprendizado de máquina pode ser uma ferramenta de otimização para processar esses dados e triagem de amostras para integridade.

O Aprendizado de Máquina, vem para auxiliar nessa tarefa de análise, oferecendo maneiras eficientes de produzir resultados e interpretações importantes a partir dos dados obtidos da técnica FTIR.

A base da aplicação das técnicas espectroscópicas depende da geração da "impressão digital" do analito. Um produto lácteo específico como o leite, que apresenta uma determinada composição química, quando exposto a uma fonte de luz, apresentará um espectro característico, que é resultado da absorção por vários constituintes químicos diferentes. Como uma "impressão digital", cada molécula apresentará o seu próprio espectro na região do infravermelho, tornando o método viável para identificar diferentes tipos de amostras. Devido às pequenas variações de composição exata, é necessário comparar o espectro do analito a uma biblioteca de espectros representativos, a fim de estabelecer sua qualidade ou autenticidade (Karoui e Baerdemaerker, 2007).

Redes neurais artificiais, especialmente as redes neurais convolucionais, vêm ganhando espaço crescente na área de aprendizado de máquina, com propostas de arquiteturas específicas para diferentes habilidades, como reconhecimento de imagens, processamento de linguagem natural, previsões de dados temporais como de mercados financeiros, entre outros (Witten *et al.*, 2016).

Uma das propriedades mais importantes de uma rede neural artificial é a capacidade de aprender por intermédio de exemplos e fazer inferências sobre o que aprendeu, melhorando gradativamente o seu desempenho. As redes neurais utilizam um algoritmo de aprendizagem cuja tarefa é ajustar os pesos de suas conexões (Braga; Carvalho; Ludemir, 2000).

O objetivo deste trabalho foi utilizar metodologias computacionais capazes de identificar a adulteração do leite pela adição de soro de queijo em amostras analisadas por equipamento de Espectroscopia de Infravermelho (FTIR), uma vez que por meio desta técnica é gerado um banco de dados que pode conter informações desconhecidas e muito importantes que podem ser utilizadas por meio do Aprendizado de Máquina podendo trazer resultados reveladores.

Além disso, é relatado em diversos estudos a possibilidade de resultados falso positivos no método oficial (Brasil, 2018a) para identificação de fraude por soro de queijo que demonstraram problemas de precisão e desempenho deste método devido a vários fatores, como acidez do soro e condições de armazenamento. Portanto, um método alternativo, não suscetível a tal fatores, é uma grande necessidade para a indústria de laticínios.

## 4.2 Material e Métodos

- **Local de realização do experimento**

O experimento foi realizado no Laboratório de Análise da Qualidade do Leite da Escola de Veterinária da UFMG (LabUFMG), credenciado pelo Ministério da Agricultura, Pecuária e Abastecimento (MAPA) e certificado pela norma ABNT NBR ISO/IEC 17025:2017 e no Laboratório Piloto de Laticínios (Escola de Veterinária da Universidade Federal de Minas).

- **Preparo das amostras e análise laboratorial**

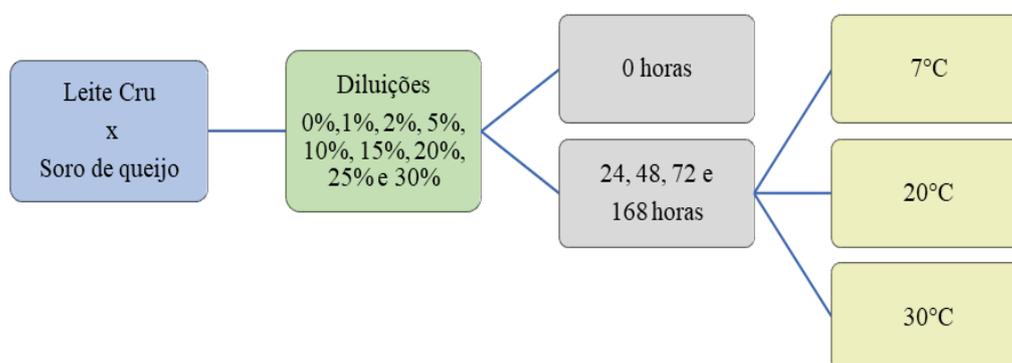
Foram realizadas cinco coletas, de 15 litros de leite cru em cada, na Fazenda Experimental Hélio Barbosa da Escola de Veterinária da UFMG, localizada no município de Igarapé – MG, nos meses de agosto e setembro de 2018. O leite foi coletado, após agitação, diretamente do tanque de resfriamento em galão plástico higienizado e imediatamente transportado até o LabUFMG, onde foi colocado em câmara fria de 4 a 7°C, por aproximadamente 1,5 horas até o momento da utilização.

A primeira etapa foi a fabricação do queijo Minas Frescal no Laboratório Piloto para a separação do soro do queijo. Foi feita a Pasteurização de 5 litros de leite (63 a 65°C por 30 minutos), após isso foi feito o resfriamento a 35°C e foram adicionados CaCl<sub>2</sub> (0,2 g/L) e 3,5 ml de coalho bovino líquido (HA-LA). Após agitação, o leite com o coalho foi deixado em repouso por 40 minutos para coagulação. Depois o coágulo foi cortado em cubos de 2 cm de lado, e agitado lentamente. A massa foi então deixada em repouso por 5 minutos. A massa foi novamente agitada lentamente, até que o soro estivesse

límpido, sem a presença de matéria leitosa. O soro de queijo produzido foi submetido ao tratamento térmico de 72°C a 75°C por 10 minutos para inativação enzimática da renina (quimosina) e impedir o processo de coagulação.

Foram feitas as diluições em leite resfriado (0%, 1%, 2%, 5%, 10%, 15%, 20%, 25% e 30%), conforme Fig. 4.1. O leite adicionado de soro foi armazenado em frascos contendo bronopol (Advanced Instruments). Para cada repetição foram 117 análises de composição.

**Figura 4.1** - Esquema da preparação das amostras com diluições (0%, 1%, 2%, 5%, 10%, 15%, 20%, 25% e 30%), armazenadas a 7, 20 e 30°C nos tempos 0h, 24h, 48h, 72h e 168h



FONTE: A autora (2021).

As racks foram montadas com as amostras e com uma amostra controle (branco) de leite sem nenhuma adição, no início de cada rack. Todos os frascos foram devidamente identificados com etiquetas à prova de água.

As amostras foram mantidas em estufa a 20°C, estufa a 30°C e refrigeradas a 7°C ±2 e analisadas nos tempos 0, 24, 48, 72 e 168 horas para composição. Foram realizadas 5 repetições, gerando 585 amostras, para fins de mineração computacional de dados.

As análises de composição centesimal (gordura, proteína, lactose, sólidos totais, sólidos não gordurosos e caseína), nitrogênio ureico do leite e crioscopia foram realizadas no equipamento CombiScope FTIR 400®, por meio do LactoScope FTIR (Delta Instruments, Drachten, Holanda).

Para a realização das análises de composição e crioscopia no equipamento Delta FTIR, cada rack previamente preparada foi colocada no banho-maria à temperatura de 40 ( $\pm 2$ ) °C, pelo período de 20 minutos, e homogeneizada por inversão 10 vezes com o auxílio de régua metálica.

A calibração do equipamento CombiScope™ FTIR (Delta Instruments) foi realizada por meio de 14 amostras padrão de leite cru, com diferentes faixas de composição para gordura, proteína, lactose e sólidos totais, obtidas do laboratório VALACTA® (Dairy Production Centre of Expertise, Quebec, Canadá), juntamente com os laudos contendo os resultados das análises realizadas no laboratório de origem.

Um total de 10 variáveis são obtidas e dependem de calibrações no equipamento que, a partir de um modelo de regressão linear múltipla (*Multiple Linear Regression*, MLR) interno ao equipamento, considera a absorção da luz por regiões pré-determinadas. O modelo de regressão é capaz de estimar numericamente as seguintes *features*: gordura (fat), proteína (*protein*), lactose, sólidos totais (*total solids*), sólidos não-gordurosos (*solids non-fat*, SNF), caseína (*casein*) e nitrogênio ureico do leite (*milk urea nitrogen*, MUN). Outros valores também obtidos incluíam o valor do ponto de congelamento (*FrzPoint*) e um valor de controle de qualidade da amostra (*QValue*). É a partir do espectro em si que o equipamento executa um método de regressão para consolidar os dez atributos dos componentes descritos.

- **Aquisição dos dados**

O processo teve início com a preparação dos dados, quando os registros resultantes das análises do equipamento FTIR foram organizados de forma a servir como entrada para o algoritmo de mineração de dados. Esses registros consistiram nos resultados de composição e nos espectros obtidos pela absorção dos raios infravermelhos em cada amostra.

Com o conjunto de dados preparado, deu-se início à aplicação do Aprendizado de Máquina através do problema classificação. Neste método, os dados obtidos foram utilizados para gerar um modelo estatístico com base em um treinamento realizado com

as amostras. Utilizou-se a Classificação binária, em que as respostas possíveis para uma classificação de amostra são a presença ou a ausência de soro de queijo.

A espectroscopia FTIR foi aplicada a todas as amostras de leite coletadas para obter espectros de infravermelho e resultados de composição. O equipamento FTIR produz duas informações para cada amostra analisada: um arquivo com dados espectrais (formato SPC, Thermo Scientific, Galactic Grams), que contém coordenadas para o espectro infravermelho; e um arquivo de componentes (formato CSV), que contém variáveis numéricas, calculadas por meio de fórmulas de análises multivariadas (Partial Least Squares, PLS) pelo equipamento a partir do espectro. A Classificação foi feita baseada nos resultados da composição e nos resultados dos espectros, utilizando aprendizagem profunda (RNA) e conjunto de árvores de decisão.

Para cada amostra no conjunto de dados, tem-se as coordenadas do espectro lidas durante a análise pelo equipamento. Essas coordenadas são compostas de 498 pontos e representam a absorvância de cada número de onda (ou região espectral). As coordenadas espectrais partem de  $2.997\text{ cm}^{-1}$  até  $1.002\text{ cm}^{-1}$ , medidos em número de onda.

As estruturas dos dados relativas aos atributos componentes (10 *features*) e as coordenadas espectrais (498 pontos) exigem diferentes abordagens na aplicação de métodos de aprendizado de máquina. Nos atributos componentes, cada *feature* representa fortemente uma característica da composição do leite, enquanto as coordenadas espectrais representam curvas do espectro infravermelho, onde cada valor é um ponto da curva.

- Árvore de Decisão

Foram utilizadas 520 amostras com soro de queijo adicionado ao leite cru em diferentes concentrações (1%, 2%, 5%, 10%, 15%, 20% e 30%) e 585 de leite cru (amostras de controle) coletados do banco de dados do laboratório que foram analisadas durante 2019/2020, gerando um total de 1105 amostras.

A Árvore de Decisão CART® (Minitab 19.2020) foi usada como método de classificação usando o grupo contendo as amostras para Treinamento e Teste. O algoritmo preditivo resultante foi aplicado ao conjunto de dados de validação com o objetivo de detectar soro de queijo adicionado ao leite.

Para a Árvore de Decisão os parâmetros incluídos foram: probabilidades correspondentes às frequências da amostra, validação do modelo em uma proporção de 70:30 (treinamento: teste), dividido aleatoriamente de 895 amostras (80% do número total de amostras); Método de divisão de Gini, com um erro padrão de custo mínimo de classificação incorreta e 20% do amostras foram usadas para validação.

- Rede Neural pela composição

Após coleta, os dados gerados foram normalizados e inseridos no *software* SPSS 22 (IBM). A rede escolhida para realização do presente trabalho foi do tipo Perceptron de Múltiplas Camadas (MLP).

Foram utilizadas 520 amostras com soro de queijo adicionado ao leite cru em diferentes concentrações (1%, 2%, 5%, 10%, 15%, 20% e 30%) e 585 de leite cru (amostras de controle) coletados do banco de dados do laboratório que foram analisadas durante 2019/2020, gerando um total de 1105 amostras.

O procedimento para Rede Neural Perceptron Multicamadas (MLP; SPSS 22, IBM) incluiu os seguintes parâmetros: As amostras de treinamento, teste e validação foram divididas aleatoriamente a uma taxa de aproximadamente 60:25:15; camada de entrada com os recursos selecionados como covariáveis; otimização de algoritmo baseada em treinamentos; tempo máximo de treinamento de 15 minutos; épocas de treinamento calculadas automaticamente.

As métricas de desempenho foram avaliadas com base na acurácia, precisão, sensibilidade e especificidade, onde:

$$\text{Acurácia} = \frac{\text{amostras previstas corretamente}}{\text{todas as amostras}}$$

$$\text{Precisão} = \frac{\text{verdadeiro positivo}}{(\text{verdadeiro positivo} + \text{falso positivo})}$$

$$\text{Sensibilidade} = \frac{\text{verdadeiro positivo}}{(\text{verdadeiro positivo} + \text{falso negativo})}$$

$$\text{Especificidade} = \frac{\text{verdadeiro negativo}}{(\text{verdadeiro negativo} + \text{falso positivo})}$$

A sensibilidade de um modelo é a proporção em que um evento de interesse é predito corretamente para todas as amostras em que esse evento ocorre. Em outras palavras, a sensibilidade é a taxa de verdadeiros positivos. Já a especificidade é definida como a proporção das amostras em que há a não ocorrência de um evento para todas as amostras em que esse evento não ocorre. A especificidade é a taxa de verdadeiros negativos.

Uma vez que o método de referência de Índice de caseinomacropéptido usando CLAE detecta níveis com certeza acima de 1% de soro de queijo adicionado, os tratamentos com baixos níveis de adição de soro de leite (1%) foram testados adicionalmente como soro de queijo não detectável em ambos os métodos, CART® e MLP. Estas análises de Rede Neural foram baseadas na composição.

- Rede Neural pelos Espectros

Os dados espectrais são compostos por coordenadas espectrais completas, que podem ser interpretadas como imagens para o reconhecimento das redes neurais. Neste caso, foram aplicadas as Redes Neurais Convolucionais (*Convolutional Neural Networks* - CNNs), que são capazes de detectar características específicas do espectro, sem a necessidade de qualquer pré-processamento. Esses dados espectrais, geralmente são repletos de informações a serem exploradas, e são adequados às redes neurais

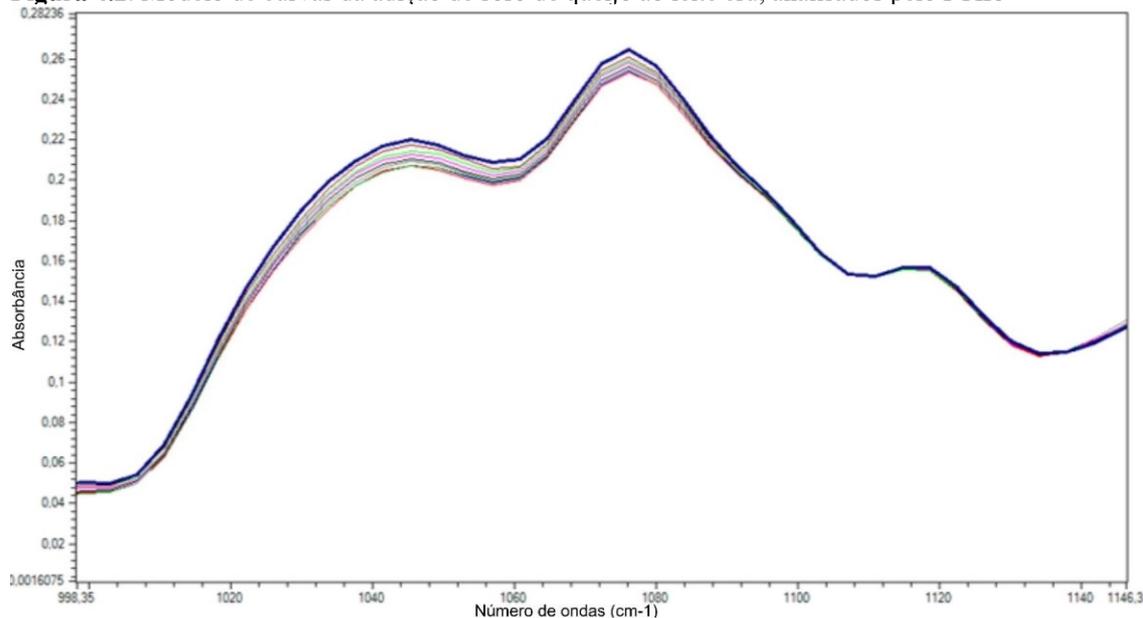
convolucionais devido ao seu formato semelhante a imagens. Essa característica, portanto, motivou a utilização desse tipo de rede neural na metodologia proposta.

Para a avaliação do método de aprendizado de máquina proposto, foi utilizada a classificação binária. Nessa classificação foi considerado a presença ou ausência de soro de queijo, com valor leite cru para amostras puras e soro para amostras adicionadas de soro.

Do total de 585 amostras, foram excluídas as amostras com 0% de concentração de soro, gerando 520 amostras. Além dessas, utilizou-se 585 amostras livres de qualquer adulteração que eram utilizadas como amostras controle do laboratório. O conjunto de dados utilizado para a classificação por Redes Neurais foi composto, então, por 1105 amostras de leite.

Os dados espectrais das curvas resultantes da absorção do infravermelho (arquivo.spc) foram visualizados por meio do software Spectrum Viewer (Protea, Reino Unido). Os dados espectrais foram processados, explorando as interações entre a adição de soro de queijo e os outros componentes constituintes do leite, com o objetivo de identificar os perfis das curvas e posterior mineração dos dados. A Figura 4.2 representa os espectros sobrepostos de amostras aleatórias com concentrações diferentes de soro.

**Figura 4.2.** Modelo de curvas da adição de soro de queijo ao leite cru, analisados pelo FTIR

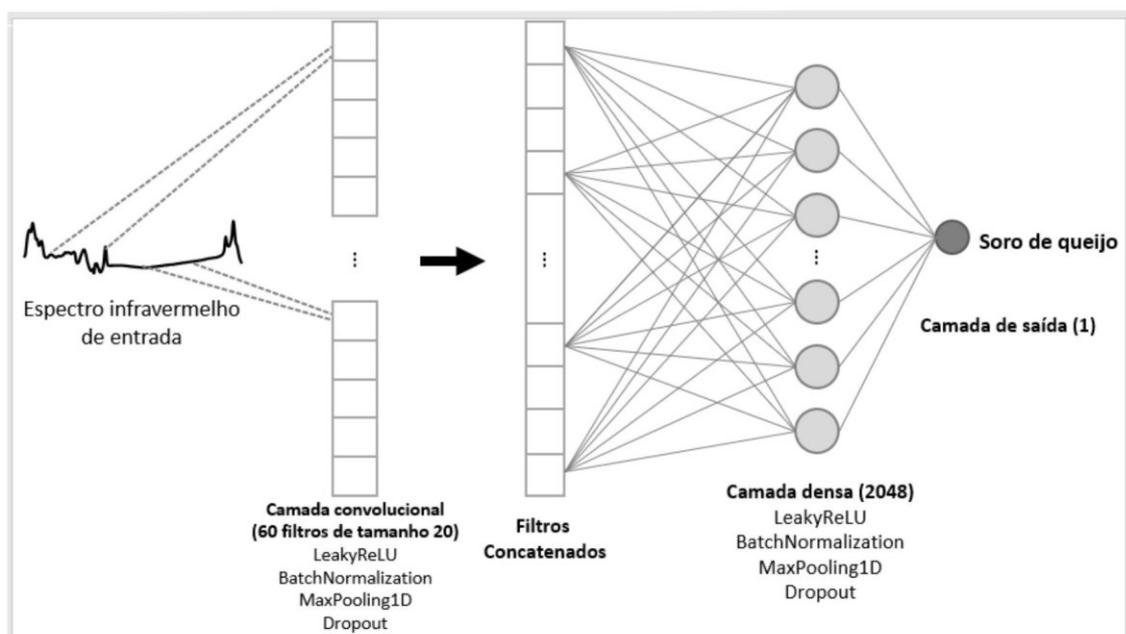


FONTE: A autora (2021).

Com os resultados fornecidos pelas curvas foi utilizada a técnica de Redes Neurais Artificiais para análises dos dados. A propriedade mais importante dessa rede é a habilidade de aprender e com isso melhorar o seu desempenho. Isso é feito através de um processo iterativo de ajustes dos valores de seus neurônios, de acordo com a inserção de amostras, pelo treinamento. O aprendizado ocorre quando a rede neural atinge uma solução generalizada para uma classe de dados (Oliveira, 2000).

Para a execução do método, uma arquitetura de Rede Neural Convolutiva foi proposta. Esta arquitetura possui uma camada convolutiva de uma dimensão, para o recebimento dos dados espectrais, que aprende 60 filtros com tamanho de *kernel* 20. Isso significa que a rede irá passar uma janela deslizante pelos dados 60 vezes e que cada janela tem tamanho 20 para os cálculos de convolução (tamanho de *kernel*). A camada pode, portanto, detectar 60 *features* diferentes diretamente a partir do espectro infravermelho. Essa camada também pode ser chamada de camada detectora de *features*. A arquitetura de CNN proposta foi baseada no trabalho de Liu *et al.* (2017) e Asseiss Neto (2020), porém com uma estrutura diferente, necessitando de menos camadas, filtros e neurônios. A Figura 4.3 apresenta os detalhes da arquitetura de CNN proposta.

**Figura 4.3** - Arquitetura da rede neural convolutiva para classificação binária de espectros infravermelhos puros



FONTE: Modificado de ASSEISS NETO (2020).

A saída da camada convolucional é concatenada, seguindo uma sequência lógica, e passada para uma camada densa, mais conhecida por camada totalmente conectada, composta de 2048 neurônios. São utilizadas camadas auxiliares que realizam a ativação dos neurônios com a função LeakyReLU (Maas *et al.*, 2013), que adicionam não linearidade ao modelo; função *BatchNormalization* (Ioffe e Szegedy, 2015), que organiza os dados, gerando lotes; função *MaxPooling1D* que reduz os dados, fazendo agrupamentos e a função de *Dropout* (Srivastava *et al.*, 2014), que ignoram um determinado número de neurônios aleatoriamente a fim de oferecer um poder maior de generalização do modelo, evitando o efeito de *overfitting* com os dados de treinamento, que pode se tornar ineficaz para prever novos resultados. O modelo binário classifica as amostras entre a presença ou a ausência de soro.

O treinamento da CNN foi feito usando o otimizador Adam (Kingma e Ba, 2015) com a execução de 300 *epochs* (interações). Durante o treinamento e a validação do modelo, para cada épocas da rede, históricos de acurácia e perda são armazenados. A cada épocas, a acurácia é calculada comparando-se a classe predita com a classe real, e a perda é calculada pelo valor de entropia cruzada entre a classe predita e a classe real. No cálculo de acurácia, a ideia é que as probabilidades de classe próximas a 0 são consideradas como classe whey (com soro) e valores próximos a 1 são considerados pure (sem adição de soro).

A metodologia empregada permite uma grande flexibilização de execuções, com filtragem de determinados dados de forma a se analisar características específicas e a diversificação do treinamento e dos testes com o modelo através da técnica denominada *cross-validation*. Nesta técnica, todas as partes do conjunto de dados são utilizadas para treinamento e teste, oferecendo uma visão realista da capacidade de predição do modelo.

A exploração de saliências é um aspecto importante para interpretar modelos de aprendizado de máquina. Utilizando os conceitos para a determinação do mapa de saliências em imagens, foi aplicada esta metodologia nas coordenadas espectrais obtidas das amostras de leite. Os métodos foram readequados para a diferença de estrutura de entrada, pois, ao contrário de imagens, os espectros obtidos são coordenadas

unidimensionais. A obtenção das saliências nos espectros obtidos, realizam, portanto, a extração das regiões espectrais que mais contribuíram para a ativação de uma determinada saída na CNN.

Foi utilizada a Linguagem Python para implementar as arquiteturas da CNN. Para acelerar o desenvolvimento foram utilizadas as bibliotecas Keras (Chollet *et al.*, 2015), Scikit-learn (Pedregosa *et al.*, 2011) e TensorFlow (Abadi *et al.*, 2016) que possuem domínio público.

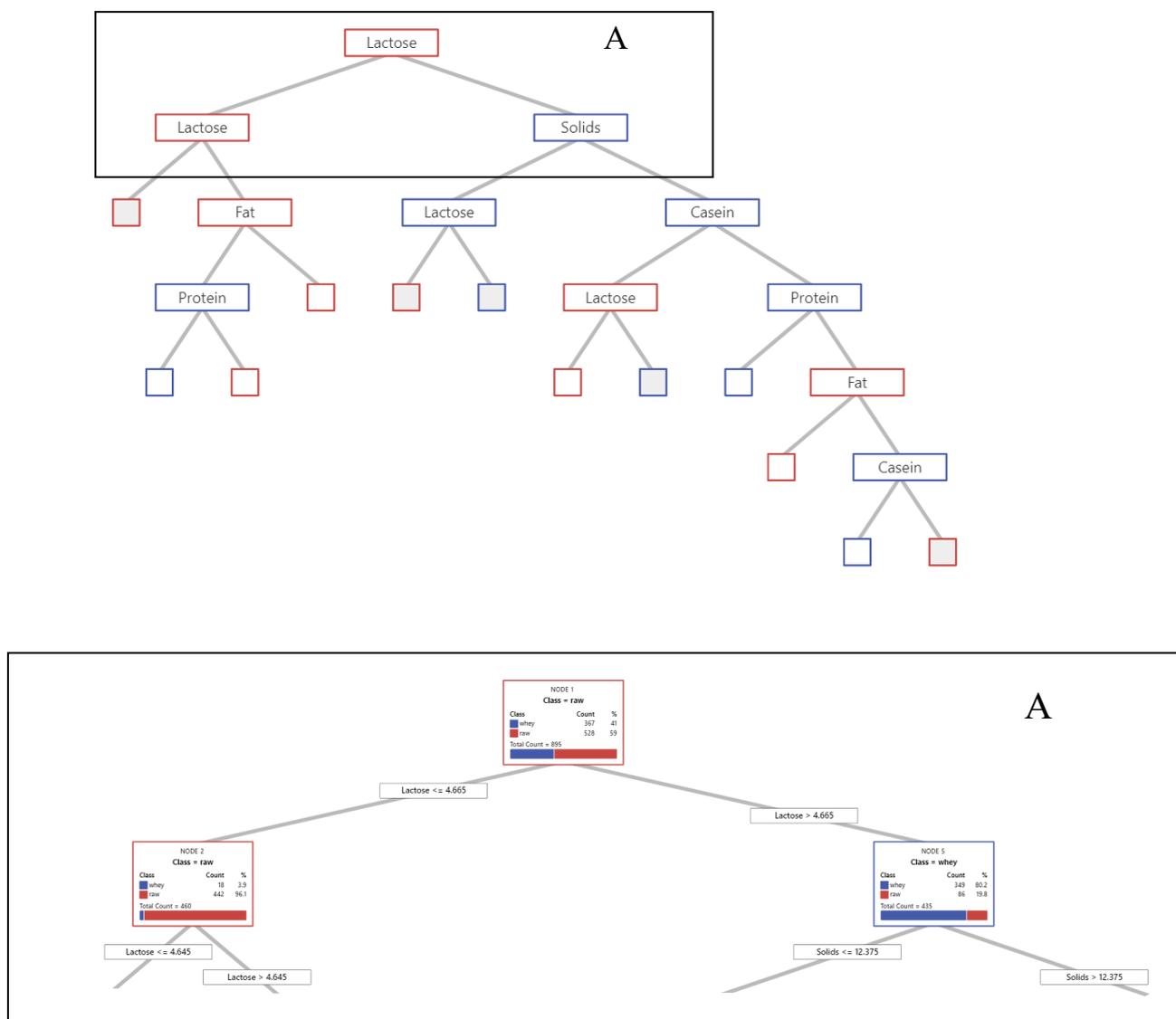
### **4. 3 Resultados e Discussão**

O equipamento FTIR gera além de dados numéricos de composição, as imagens dos espectros dos resultados que foram utilizados para alimentar o Aprendizado de Máquina de diferentes formas.

- **Árvore de decisão**

As variáveis foram analisadas utilizando aprendizagem em conjuntos de árvores de decisão. Esse modelo foi treinado em amostras de leite adicionadas de soro de queijo, a fim de reconhecer os padrões que identificam as características da adulteração e quais são os componentes mais afetados. A classificação binária foi realizada considerando amostras com soro (Whey) e sem soro (no whey). A figura 4.4 ilustra a árvore de decisão utilizada no modelo. Através da árvore de decisão, utilizando os resultados de composição foi possível observar a formação de 5 nós terminais (em cinza), que apresentaram maior importância para identificar as amostras com soro.

**Figura 4.4.** Árvore de decisão das variáveis de composição gerada pelo algoritmo de classificação. A visualização do nó está completa e parte é ampliada abaixo. Os nós cinzas apresentaram uma forte influência



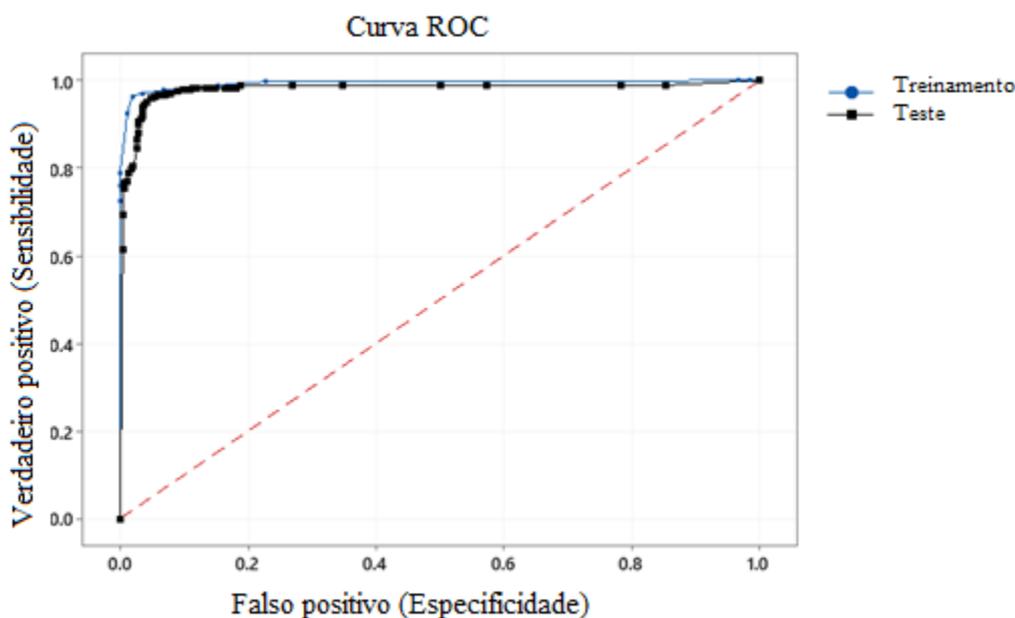
FONTE: A autora (2021).

O resultado mais importante para este modelo foi a concentração de lactose com uma importância relativa de 100% em comparação com a proteína que teve uma importância relativa de 60,8%. Isso significa que a proteína tem uma parte de importância próxima à metade da lactose nesta árvore de classificação. A variável com a menor importância relativa foi nitrogênio ureico do leite (MUN) com 18,8%. A quantidade de classificação incorreta para esta simulação foi de 0,027 para as amostras de treinamento e 0,054 para as amostras de teste.

Existe normalmente uma relação de compromisso entre a sensibilidade e a especificidade. Intuitivamente, o aumento da sensibilidade de um modelo pode provocar perda de especificidade, já que mais amostras estão sendo preditas como ocorrência de evento. Essa troca entre sensibilidade e especificidade pode ser interessante quando há penalidades diferentes associadas com cada tipo de erro. Uma técnica que avalia a troca e a relação de compromisso é denominada *Curva Receiver Operating Characteristic* (ROC) (Kuhn e Johnson, 2013)

A área sobre a curva ROC, denominada *AUC* (*Area Under Curve*) foi avaliada a fim de se obter mais uma medida de desempenho do modelo de classificação. A Figura 4.5 apresenta a curva ROC calculada para o método.

**Figura 4.5** - Curva ROC do modelo apresentado mostrando a relação entre a taxa de verdadeiros positivos e a taxa de falsos positivos para um intervalo de limiares de probabilidades de classificação. A área sob a curva (AUC) possui o valor de 0,99 no treinamento e 0,98 para o teste



FONTE: A autora (2021).

A curva plotada indica um bom desempenho do modelo apresentando um valor de AUC médio para o teste próximo de 1 (0,9799). Isso significa que o modelo está realizando praticamente todas as predições corretas para cada classe. Se a curva ROC de um modelo se aproxima bastante do ponto (0,0, 1,0), na parte superior esquerda do gráfico, esse modelo é bom. Uma outra forma de visualizar isso é quanto mais longe a curva

ROC estiver da linha pontilhada, que é a curva ROC de um modelo aleatório, melhor. Os valores da AUC são interpretados como: 0,5-0,6 (péssimo), 0,6-0,7 (ruim), 0,7-0,8 (pobre), 0,8-0,9 (bom), > 0,9 (excelente) (Metz, 1978).

A curva ROC do teste possibilita acompanhar a variação da sensibilidade e da especificidade à medida que evoluem os pontos. Quanto mais a curva ROC se aproxima do canto superior esquerdo, melhor é a qualidade do teste. A AUC é o resultado da integração de todos os pontos durante o trajeto da curva, e computa simultaneamente a sensibilidade e a especificidade, sendo um estimador do comportamento da acurácia global do teste. A AUC fornece uma estimativa da probabilidade de classificação correta (acurácia do teste); uma AUC de 0,98 reflete uma chance de classificação correta de 98%.

No campo da aprendizagem de máquina e especificamente na classificação estatística, uma matriz de confusão, também conhecida como matriz de erro, é um layout de tabela específico que permite a visualização do desempenho de um algoritmo, tipicamente um algoritmo de classificação. Através dela é possível avaliar não somente se o sistema estava recuperando as informações corretas, mas também verificar o quanto o sistema está abstraindo as informações erradas. Ao exibir a quantidade e os percentuais em cada célula desta matriz, é possível consultar rapidamente com que frequência o modelo é previsto com precisão. A Tabela 4.1 apresenta a Matriz de confusão desse modelo.

**Tabela 4.1** - Matriz de confusão para uma árvore de classificação CART® com gordura, proteína, caseína, lactose, sólidos totais, sólidos não gordurosos e ponto de congelamento como recursos de entrada e saída binomial para leite cru e leite cru adicionado com soro de queijo\* (2%, 5%, 10%, 15%, 20%, 25% e 30%).

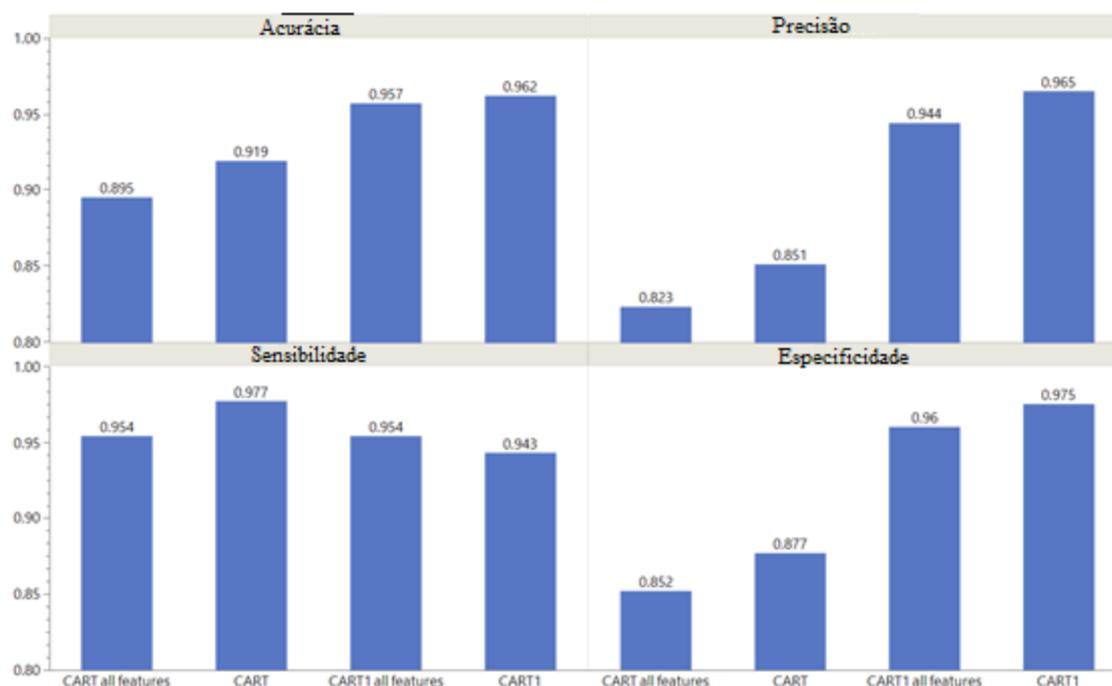
Amostra	Valor observado	Valor Predito		
		Leite cru	Leite com Soro	Porcentagem
Treinamento	Leite cru	365	10	97.3%
	Leite com Soro	10	250	96.2%
	Porcentagem Geral			96.9%
Teste	Leite cru	141	12	92.2%
	Leite com Soro	3	104	97.2%
	Porcentagem Geral			94.2%

\* Para otimização, o leite adicionado com 1% de soro de queijo foi considerado não detectável neste modelo de árvore de classificação

Através da matriz de classificação foi possível observar altas taxas de amostras com predição correta e baixas taxas de erros de classificação, tanto no treinamento quanto no conjunto de teste. Previsões corretas para detectar amostras com adição de soro foram de 96,2% no conjunto de treinamento e 97,2% no conjunto de amostras de teste.

Os algoritmos do melhor modelo e modelos adicionais foram usados na validação de conjunto de amostras e estão ilustrados na Figura 4.6.

**Figura 4.6** Desempenho do método de árvore de classificação CART® para detectar soro de queijo adicionado ao leite cru com base em dados de composição obtidos com o método do infravermelho com transformada de Fourier (FTIR)



FONTE: A autora (2021).

A primeira análise (CART all features), foi utilizado os fatores proteína, caseína, lactose, sólidos não gordurosos, sólidos totais, gordura, ponto de congelamento e nitrogênio ureico do leite (MUN) como recursos de entrada e saída binomial leite cru e leite cru adicionado com soro de queijo (1%, 2%, 5%, 10%, 15%, 20%, 25% e 30%) sendo observado uma acurácia de quase 90%, precisão de 82%, sensibilidade de 95% e especificidade de 85%.

No segundo tratamento (CART), foram utilizados proteína, caseína, lactose e sólidos não gordurosos como recurso de entrada e saída binomial leite cru e leite cru adicionado

com soro de queijo (1%, 2%, 5%, 10%, 15%, 20%, 25% e 30%) sendo observado uma acurácia de 92%, precisão de 85%, sensibilidade 98% e especificidade de 88%.

Já no terceiro tratamento (CART1 all features), foi utilizado os fatores proteína, caseína, lactose, sólidos não gordurosos, sólidos totais, gordura, ponto de congelamento e nitrogênio ureico do leite como recursos de entrada e saída binomial como leite cru e leite cru adicionado com soro de queijo (2%, 5%, 10%, 15%, 20%, 25% e 30%), sendo observado uma acurácia de quase 96%, precisão de 94%, sensibilidade de quase 95% e especificidade de 96%.

O teste de validação com alta taxa de desempenho foi o CART1, sendo utilizados os fatores de entrada proteína, caseína, lactose e sólidos não gordurosos e saída binomial como leite cru e leite cru adicionado com soro de queijo (2%, 5%, 10%, 15%, 20%, 25% e 30%). Foi observado uma acurácia de 96%, precisão de 97%, sensibilidade de quase 94% e especificidade de 98%.

No estudo de Coitinho *et al.* (2017), foram adulteradas 1650 amostras com amido, bicarbonato de sódio, citrato de sódio, formaldeído, sacarose, soro e água para a calibração de um equipamento compacto de espectroscopia FTIR, utilizando a análise de componentes principais (Principal Component Analysis – PCA). Para a avaliação das 12 calibrações obtidas, foram determinadas a sensibilidade e a especificidade de cada uma delas. Os melhores resultados demonstraram 84% de sensibilidade para todos os adulterantes simultaneamente.

- **Análise dos componentes numéricos por RNA**

Os dados de composição produzidos a partir do FTIR foram utilizados para a análise de Rede Neural (Perceptron Multicamada). A Tabela 4.2 mostra os resultados médios da composição do leite fraudado com soro de queijo. As variáveis de entrada foram as variáveis obtidas na análise do FTIR: Gordura, Proteína, Lactose, Sólidos não gordurosos, Sólidos totais, Caseína, Nitrogênio ureico do leite e Ponto de Congelamento. A camada de saída considerada foi a presença de soro.

**Tabela 4.2** - Características das amostras usadas para configurar a RNA

Parâmetros	Média	Mínimo	Máximo	Desv. Padrão	CV (%)
<b>Gordura (g/100g)</b>	3,28	2,36	4,09	0,371	13,8
<b>Proteína (g/100g)</b>	3,18	2,65	3,52	0,238	5,7
<b>Lactose (g/100g)</b>	4,75	4,60	4,97	0,085	0,7
<b>Sólidos não gordurosos (g/100g)</b>	9,14	8,67	9,48	0,196	3,8
<b>Sól. Totais (g/100g)</b>	12,18	11,03	13,21	0,520	27,1
<b>Ponto de Congelamento (°H)</b>	-0,544	-0,528	-0,555	0,003	0
<b>Caseína (g/100g)</b>	2,46	1,99	2,80	0,21	0,045

FONTE: A autora (2021).

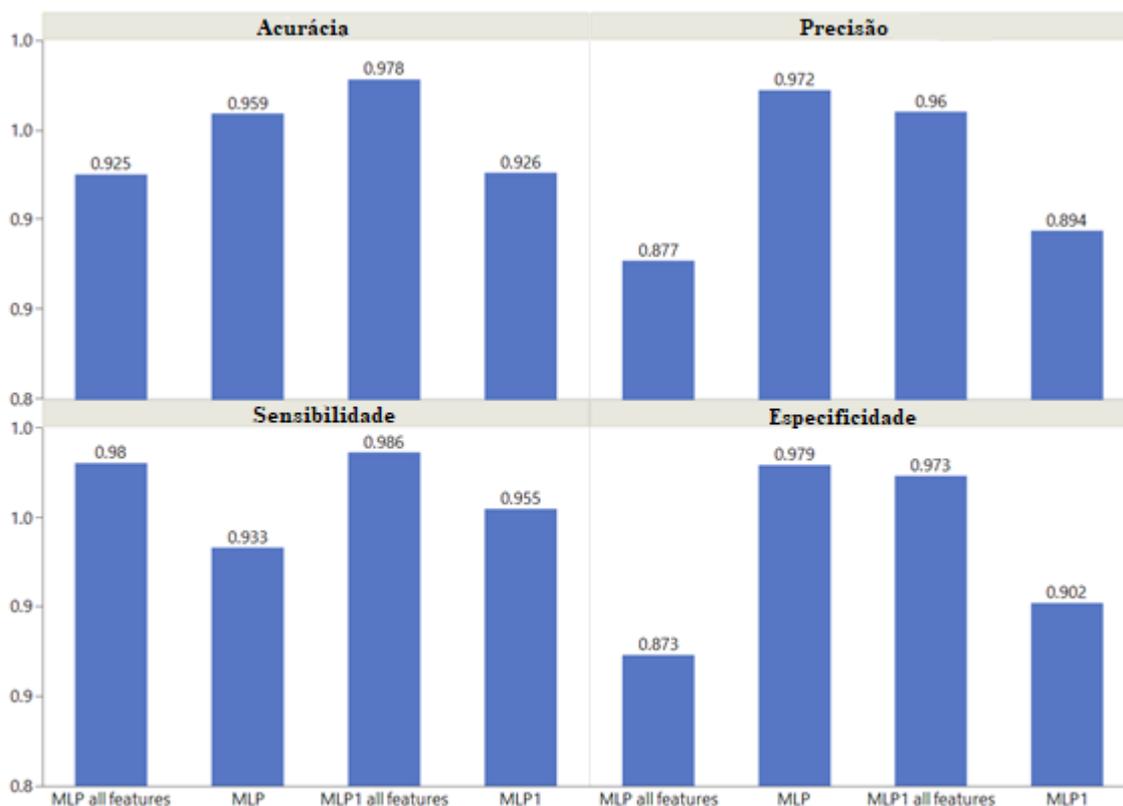
O conjunto de treinamento resultou em 1,8% de predições incorretas, enquanto para o conjunto de teste, predições incorretas foram 1,5%, conforme mostrado na Tabela 4.3 que apresenta a matriz de confusão desse modelo. Esses resultados indicam que esta rede tem uma excelente previsão, com grande potencial para testar amostras reais.

**Tabela 4.3** - Matriz de confusão para uma rede perceptron multicamadas com gordura, proteína, caseína, lactose, sólidos totais, sólidos não gordurosos e ponto de congelamento como recursos de entrada e saída binomial para leite cru e leite cru adicionado com soro de queijo (2%, 5%, 10%, 15%, 20%, 25% e 30%).

Amostra	Valor Observado	Valor predito		
		Leite cru	Leite com soro	Porcentagem
Treinamento	Leite Cru	367	4	98.9%
	Leite com soro	8	274	97.2%
	Porcentagem Geral			98.2%
Teste	Leite Cru	165	2	98.8%
	Leite com soro	2	99	98.0%
	Porcentagem Geral			98.5%

Os resultados de casos testados podem ser observados na Figura 4.7. Utilizou-se os seguintes cofatores: Proteína, Lactose, Sólidos não gordurosos, Caseína, Gordura, Sólidos totais e Ponto de Congelamento. A melhor rede é aquela que apresenta melhor acurácia.

**Figura 4.7** - Resultados de acurácia, precisão, sensibilidade e especificidade para diferentes tratamentos de Redes Neurais Artificiais, em modelo binário para identificação de soro de queijo em leite



FONTE: A autora (2021).

A primeira análise (MLP all features), foi utilizado como recursos de entrada os fatores proteína, caseína, lactose, sólidos não gordurosos, sólidos totais, gordura, ponto de congelamento e nitrogênio ureico do leite (MUN) e saída binomial leite cru e leite cru adicionado com soro de queijo (1%, 2%, 5%, 10%, 15%, 20%, 25% e 30%) foi observado uma acurácia de 93%, precisão de 88%, sensibilidade de 98% e especificidade de 87%.

No segundo tratamento (MLP), foram utilizados os fatores de entrada proteína, caseína, lactose e sólidos não gordurosos e saída binomial leite cru e leite cru adicionado com soro de queijo (1%, 2%, 5%, 10%, 15%, 20%, 25% e 30%) sendo observado uma acurácia de 96%, precisão de 97%, sensibilidade 94% e especificidade de 98%.

Já no terceiro tratamento (MLP1 all features), foi utilizado como recursos de entrada os fatores proteína, caseína, lactose, sólidos não gordurosos, sólidos totais, gordura, ponto de congelamento e nitrogênio ureico do leite e saída binomial como leite cru e leite cru

adicionado com soro de queijo (2%, 5%, 10%, 15%, 20%, 25% e 30%), sendo observado uma acurácia de 98%, precisão de 97%, sensibilidade de quase 99% e especificidade de 97%.

A quarta análise (MLP1), foram utilizados os fatores de entrada proteína, caseína, lactose e sólidos não gordurosos e saída binomial como leite cru e leite cru adicionado com soro de queijo (2%, 5%, 10%, 15%, 20%, 25% e 30%), sendo observado uma acurácia de 93%, precisão de 89%, sensibilidade de quase 96% e especificidade de 90%.

Valente *et al.* (2014) utilizaram amostras preparadas através da mistura do leite com diferentes concentrações de soro (0, 1, 5, 10 e 20%) e analisaram através de um analisador de leite por ultrassom quanto a temperatura, teor de gordura, extrato seco desengordurado, densidade, proteína, lactose, sais minerais, ponto de congelamento, condutividade e pH. Esses resultados serviram para alimentar a Rede Neural Artificial. Foram testadas as redes de função radial e *Multilayer Perceptron*, sendo que a melhor rede de classificação foi a rede neural de função de base radial. Os resultados desse artigo demonstraram erro de classificação menor que 5% aplicando a rede neural artificial.

No geral, o tratamento “MLP1 all features” obteve melhores resultados. No modelo de classificação trabalhado foi aplicado ao conjunto de validação de amostra considerando 1% de soro adicionado ao leite como não detectável, apresentando erro de pouco mais de 2%. Esses resultados mostram que as Redes Neurais Artificiais apresentam grande potencial de uso como ferramenta de avaliação de possíveis adulterações de leite com soro de queijo. Essa técnica apresenta uma grande possibilidade de aplicação reduzindo o número de amostras que iriam ser analisadas pela metodologia oficial.

A utilização de métodos de aprendizado de máquina para detectar a adição fraudulenta de soro de queijo ao leite é escasso. O uso de redes neurais artificiais foi relatado em outro trabalho, usando os resultados de composição de análises de rotina em amostras de leite como variáveis de entrada. Conde e colaboradores (2020) realizaram um trabalho para quantificar o teor de soro em amostras de leite adulterado por meio de redes neurais artificiais, usando como variáveis de entrada os resultados de composição

de análises rotineiras em amostras de leite. As análises foram realizadas com diferentes concentrações em relação à adição de soro de queijo (0, 5, 10 e 20%), e as amostras foram analisadas quanto à gordura, sólidos não gordurosos, densidade, proteína, lactose, minerais e ponto de congelamento, totalizando 164 ensaios, dos quais 60% foram utilizados para treinamento em rede, 20% para validação de rede e 20% para teste de rede neural. Embora os autores terem afirmado que o uso das redes neurais provou ser eficiente, os autores sugeriram a utilização de um método mais preciso para confirmação da fraude. Apesar de ser um modelo para quantificação, não foram apresentadas métricas de desempenho do modelo. Entre as variáveis de maior relevância estavam o teor de gordura e a densidade.

- **Rede Neural Convolutacional - Análise dos espectros**

Os dados espectrais produzidos a partir da composição das amostras de leite foram utilizados para o treinamento do algoritmo de aprendizado de máquina. O método proposto é utilizado para prever a presença de soro de queijo em um problema de classificação binária. Foi proposta uma arquitetura de Rede Neural Convolutacional (*Convolutional Neural Network - CNN*) que dispensa o pré-processamento de dados espectrais. A Figura 4.8 mostra os dados convertidos em arquivo .CSV, em que os espectros são apresentados como coordenadas.

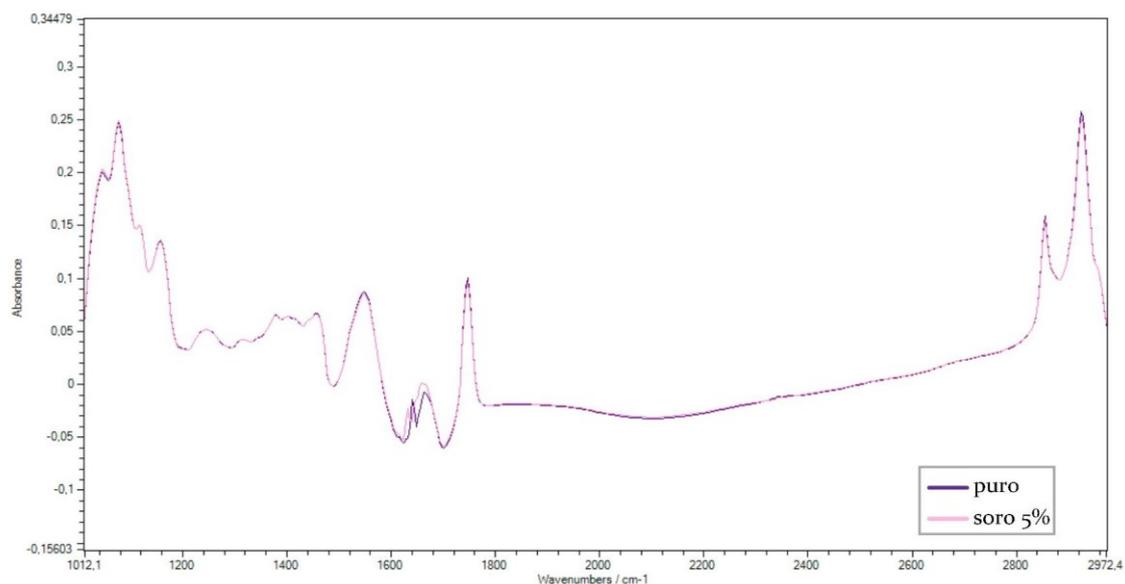
**Figura 4.8** - Exemplo de aquisição de dados convertidos em arquivo formato .CSV com todos os componentes numéricos obtidos de amostras aleatórias

1	Week	Nr	Temp	Temp	Ingredis	Concer	IsRaw	M	SPC 1002	SPC 1006	SPC 1010	...	SPC 2985	SPC 2989	SPC 2993	SPC 2997
2	1	1	0		raw	0.0	yes		0.04043000	0.04345983	0.05661538243293762		0.01531188189	0.0077411383	0.000716209	-0.005273386836051941
3	1	2	0		whey	0.0	no		0.04051002	0.04435856	0.05766892433166504		0.0177055895	0.009675458	0.00282457	-0.002519264817237854
4	1	3	0		whey	0.0	no		0.04091134	0.04494839	0.05843093991279603		0.0168894678	0.009124428	0.00283153	-0.002293705940246582
5	1	4	0		whey	0.0	no		0.04116952	0.04545474	0.0583987683057785		0.0187065154	0.0110307484	0.00367450	-0.002246350049972534
6	1	5	0		whey	0.0	no		0.04286008	0.04651050	0.0593220210075378		0.0190013796	0.0115191364	0.00443609	-0.0006889253854751588
7	1	6	0		whey	0.0	no		0.04303368	0.04679444	0.0607554018497467		0.0177536606	0.0109145641	0.00480204	-0.0003943443298339844
8	1	7	0		whey	0.0	no		0.04339522	0.04739806	0.061492532096328735		0.0184671282	0.0112087728	0.00544996	0.0005766302347183228
9	1	8	0		whey	0.0	no		0.04557239	0.04962255	0.06325101852416992		0.0193098485	0.0125457644	0.00584444	0.0034010261297225948
10	1	9	0		whey	0.0	no		0.04572753	0.04953588	0.06387586891651154		0.0191809982	0.012479022	0.00643856	0.0017436593770980835
11	1	1	24	20.0	raw	0.0	yes		0.03708292	0.040741652	0.0541117787361145		0.0123582329	0.0046571195	-0.001904119	-0.0071630775928497314
12	1	2	24	20.0	whey	0.0	no		0.03734974	0.041324677	0.0543319433927536		0.0144782960	0.0067109167	0.003401026	-0.006031587199172965
13	1	3	24	20.0	whey	0.0	no		0.03677168	0.04045791	0.0537058413028717		0.0138082057	0.0060167753	-0.00042919	-0.006050929427469312
14	1	4	24	20.0	whey	0.0	no		0.03804363	0.041692942	0.05498751997947693		0.0132555961	0.0055746138	-0.00069834	-0.006112903356552124
15	1	5	24	20.0	whey	0.0	no		0.03784912	0.04135408	0.055109575390815735		0.0153825432	0.007832303	0.001681163	-0.004205867848124696
16	1	6	24	20.0	whey	0.0	no		0.04024049	0.043911874	0.05700670182704925		0.0149334073	0.007607877	0.001865506	-0.002842053771018982
17	1	7	24	20.0	whey	0.0	no		0.04070678	0.04550555	0.05812628567218781		0.0174604207	0.0106280448	0.00585899	0.0007185488939285277
18	1	8	24	20.0	whey	0.0	no		0.04050943	0.044491978	0.05867445468902587		0.0165207833	0.010005105	0.00420872	-0.0006808459150283812
19	1	9	24	20.0	whey	0.0	no		0.04254601	0.04675630	0.06057447195053101		0.0163646489	0.010737776	0.00566478	0.0006057173013687134
20	1	1	24	30.0	raw	0.0	yes		0.03693024	0.041257068	0.05442772805690765		0.0126262903	0.004576832	-0.00230728	-0.00726340711167908
21	1	2	24	30.0	whey	0.0	no		0.03657174	0.04072953	0.05390141904354096		0.0127599537	0.004779472	-0.00178587	-0.0077959299087524405
22	1	3	24	30.0	whey	0.0	no		0.03707188	0.041074708	0.05425982177257538		0.0115785747	0.003824159	-0.00239552	-0.007973146378326416
23	1	4	24	30.0	whey	0.0	no		0.03829878	0.04237747	0.055992543697357185		0.0142293423	0.0062541612	2.37E+11	-0.005123421549797058
24	1	5	24	30.0	whey	0.0	no		0.03942388	0.04304409	0.05621084570884705		0.0147764533	0.007266804	0.001256378	-0.003466516731695557
25	1	6	24	30.0	whey	0.0	no		0.03976170	0.04369695	0.05702109634876251		0.0138908326	0.006556496	0.00053686	-0.004212096333503722
26	1	7	24	30.0	whey	0.0	no		0.04100543	0.04510666	0.05897006392478942		0.0155492424	0.0081330839	0.00222378	-0.00287945568561554
27	1	8	24	30.0	whey	0.0	no		0.04152828	0.045129522	0.05857571959495544		0.0142831951	0.007203578	0.001416236	-0.003102719783782959
28	1	9	24	30.0	whey	0.0	no		0.04218088	0.046558021	0.060399755835533135		0.0157931894	0.0098119378	0.004778444	-0.0001055449247360229

FONTE: A autora (2021)

Os espectros de infravermelho produzidos pela técnica de FTIR são formados por 518 pontos ou varreduras, com resolução de  $4\text{ cm}^{-1}$ , medidos em números de onda variando de  $3000\text{ cm}^{-1}$  a  $1000\text{ cm}^{-1}$ , que são usados como entrada para o classificador da CNN. Cada espectro é seguido pelo rótulo de classe (possuir soro) no conjunto de dados, o que permite o processo de treinamento da rede. No modelo proposto foi feita a varredura nos números de onda variando de  $2997\text{ cm}^{-1}$  a  $1002\text{ cm}^{-1}$ , o que pode ser observado na Figura 4.8. Segundo estudos realizados por Vinciguerra (2019), as bandas de absorção de amida que poderiam estar associadas à presença de CMP devem estar na região  $1200\text{-}1000\text{ cm}^{-1}$  e  $1690\text{-}1670\text{ cm}^{-1}$ . Na Figura 4.9, é demonstrado o espectro da amostra com adição de 5% de soro de queijo e o espectro do leite normal (puro), revelando as alterações provocadas no mesmo.

**Figura 4.9** Alterações provocadas no espectro de amostra com soro (5%), em relação ao espectro do leite puro



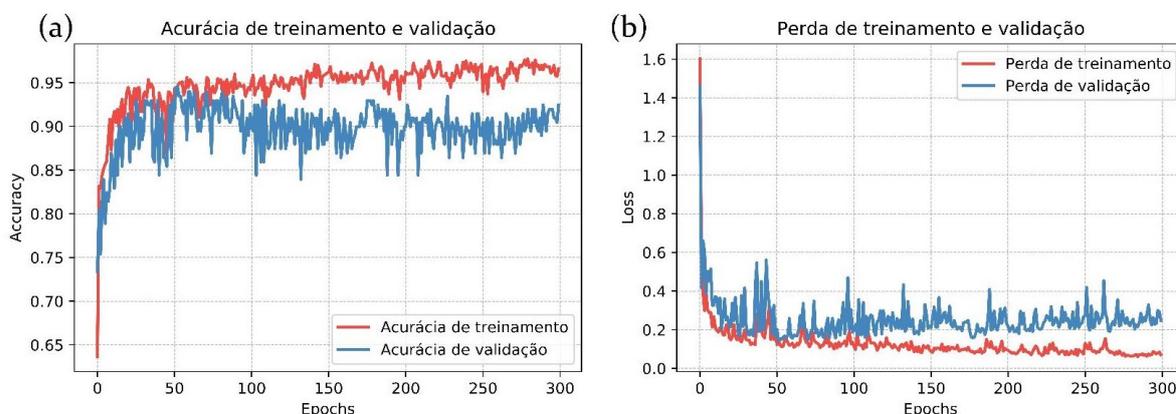
FONTE: A autora (2021).

Foram observadas as características dos espectros do leite e a sobreposição do espectro de leite com 5% de soro, sendo possível observar que não existiu diferença visual entre os espectros de leite e leite com soro.

Botelho *et al.* (2015) adulteraram amostras de leite com peróxido de hidrogênio, amido e sacarose variando a concentração de 0,5% a 10% identificaram pela visualização do espectro esses adulterantes com uma concentração de 10%.

O treinamento da CNN foi feito com 300 *epochs* (épocas ou interações) para a classificação binária. Cada execução da CNN considerou 20% do conjunto de treinamento como o conjunto de validação interno à rede. A Figura 4.10 mostra os gráficos da precisão e da perda do modelo em conjuntos de treinamento e validação, revelando o aprendizado da rede neural. À medida que se aumentam as épocas, o aprendizado da rede evolui (acurácia) e diminuem-se os erros (perdas).

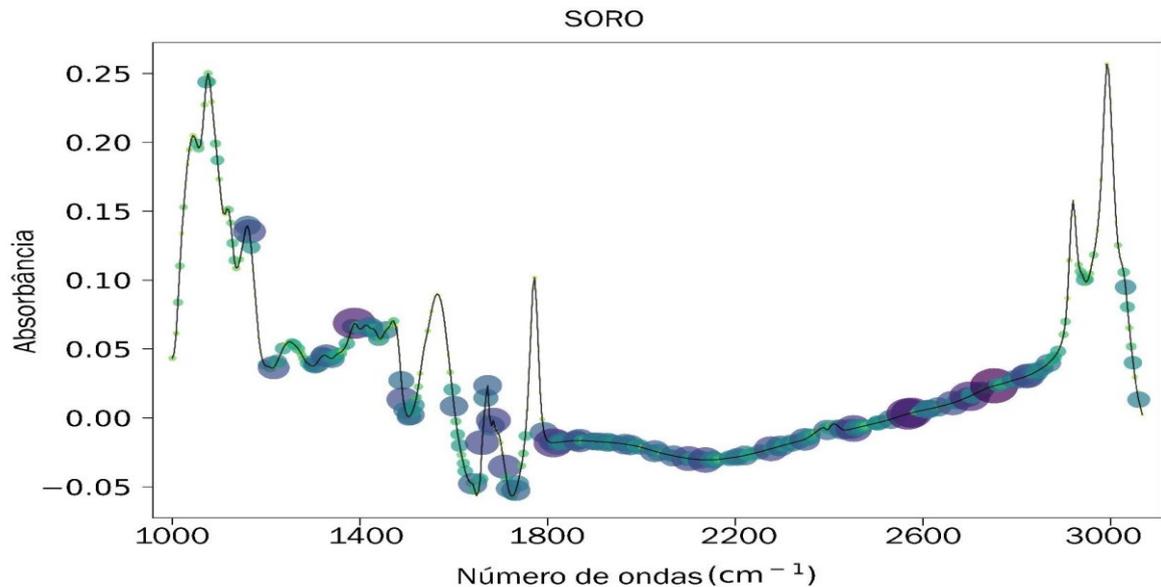
**Figura 4.10** - Gráfico da precisão e da perda do modelo da CNN nas etapas de treinamento e validação. O modelo foi treinado por 300 épocas. (a) precisão do treinamento e validação. (b) perda de treinamento e validação



FONTE: A autora (2021).

A Figura 4.11 apresenta as saliências obtidas com o espectro de uma amostra com adição de soro de queijo, isto é, as regiões que ativaram o nó final “com soro” na CNN utilizada. Então, pode-se utilizar os pontos de saliência que ocorrem com mais frequência considerando todas as amostras do conjunto de treinamento para formar um conjunto de regiões espectrais. Essas regiões compreendem os pontos que a rede considerou mais importantes, utilizando todas as amostras de treinamento. Selecionam-se assim, os pontos de saliência mais frequentes, formando-se, portanto, um conjunto de *features* que podem ser utilizadas para substituir o processamento realizado pelo equipamento FTIR. Esse conjunto de *features* é extraído automaticamente, sem conhecimento prévio das amostras.

**Figura 4.11** - Saliências calculadas pela CNN com o espectro de uma amostra com soro de queijo.  
Regiões mais escuras compreendem os pontos onde a rede considerou mais importante



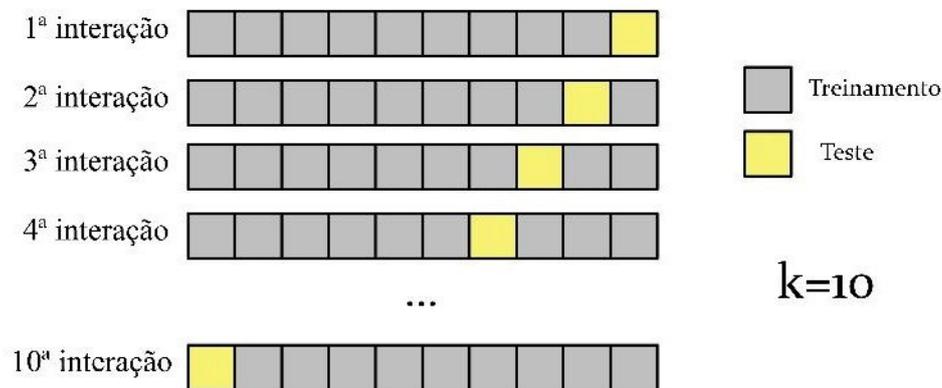
FONTE: A autora (2021).

As regiões com os destaques maiores e mais escuros representam a intensidade com que cada coordenada contribuiu para a decisão da rede, isto é, as regiões que ativaram o nó final “soro”. Visualmente, a única região possível de ver uma pequena diferença, de acordo com a figura 4.9, que apresentou uma curva do leite e do leite com adição de soro, foi na região próxima a 1600/1700 número de ondas e com a Rede Neural Convolutiva, foram destacados inúmeros pontos.

Apesar de os resultados obtidos com a extração de *features* descrita anteriormente não representar ganhos de desempenho dos modelos, os resultados são, em média, equivalentes ao uso das *features* do equipamento FTIR. Além disso, a exploração de saliências é um aspecto importante da interpretação de modelos de aprendizado de máquina.

Para a avaliação da qualidade dos classificadores, a técnica de validação cruzada utilizada foi a *k-fold*. A saída foi uma matriz com 10 valores que corresponde a cada execução do *cross-validation*, já que o número de *folds* foi definido como 10, conforme Figura 4.12. Nesse modelo a acurácia média é considerada como a acurácia do modelo.

**Figura 4.12** - Técnica de validação cruzada *k-fold*, com treinamento e teste com  $k=10$



FONTE: A autora (2021).

A arquitetura de rede neural convolucional proposta produziu valores de acurácia entre 89,19% a 95,50%, sendo que a acurácia média foi de 93,84%. Pode-se prever, com > 93% de segurança que a amostra analisada está ou não com adição de soro. O valor da acurácia indica a taxa de acertos, uma medida da variabilidade estatística. Quanto maior a acurácia, melhor é a exatidão e a precisão do experimento.

Pode-se definir a precisão de um modelo como a proporção de previsões corretas de uma categoria em relação a todas as previsões feitas dessa categoria. As previsões corretas da categoria alvo são chamadas de Verdadeiros Positivos (*true positive — TP*), e as previsões incorretas para a categoria alvo são chamadas de Falsos Positivos (*false positive — FP*). Nesse estudo, um verdadeiro positivo é uma situação na qual o modelo previu que há adição de soro no leite cru e ele realmente tinha, enquanto um falso positivo é uma situação na qual o modelo previu que há adição de soro, mas não tinha. Pelos resultados, o modelo proposto tem uma precisão média de 91,81%, em um intervalo de 84,21% até 98,18%, ou seja, em todas as situações em que o modelo considerou a presença de soro de queijo, ele esteve certo, em média, 92% vezes.

No algoritmo construído observou-se uma matriz de confusão conforme a figura 4.13.

Figura 4.13 - Medidas de desempenho da RNA calculadas a partir da matriz de confusão

		Valor previsto		
		Somente Leite	Presença de Soro	Total
Valor Esperado	Somente Leite	555 Verdadeiro positivo	30 Falso negativo	585
	Presença de Soro	19 Falso positivo	501 Verdadeiro negativo	520
	Total	574	531	1105

FONTE: A autora (2021).

Através dessa Matriz é possível observar que a primeira célula de resultado, contém o valor 555, indicando o número de *verdadeiros positivos* para obter o valor somente leite. Dessa forma, essa estatística indica que o modelo previu o valor correto para “somente leite” em 555 casos. A célula logo abaixo dessa, que contém o valor 19, indica o número de *falsos positivos* ou quantas amostras deveriam ser “somente leite”, mas foram classificadas como contendo soro.

A célula que contém o valor 30 indica o número de *falsos negativos* para a presença de soro. Essa estatística indica que em 30 casos, o modelo previu que há soro, mas na verdade não tem. Já em 501 casos, o modelo previu corretamente que as amostras contêm soro de queijo.

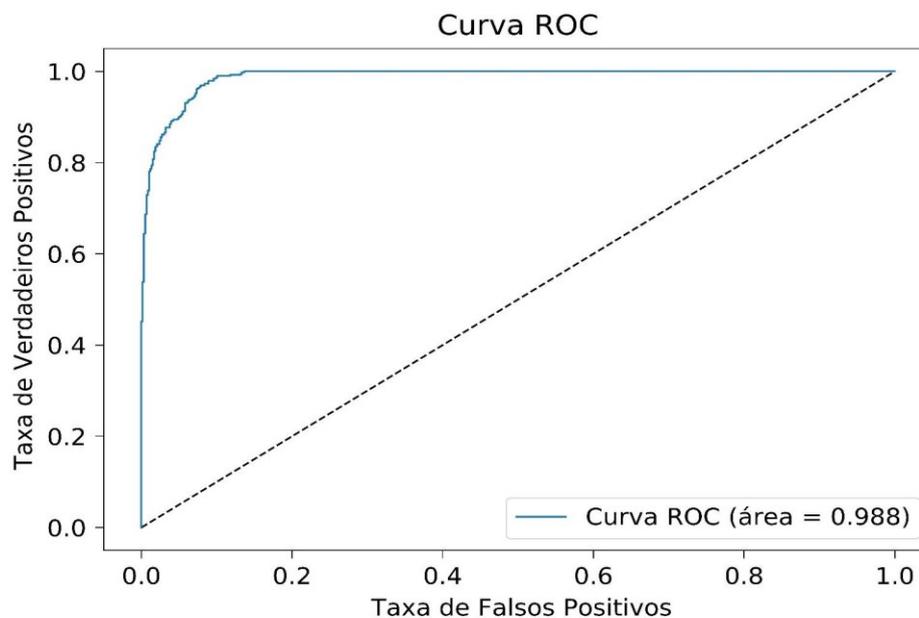
Somando os valores das células que são diagonalmente adjacentes, é possível determinar a exatidão geral do modelo. Uma diagonal mostra o número de previsões corretas, e a outra diagonal mostra o número de previsões incorretas. Num total de 1105 casos, 49 (19+30) foram consideradas incorretas, observando uma taxa de erro de aproximadamente 4,43%. Para métodos de triagem, um erro de 5% é um número aceitável. Os resultados obtidos indicam um bom método para auxílio na detecção de fraude por soro de queijo em leite.

Pela Matriz de confusão foi possível determinar algumas métricas adicionais relevantes: A acurácia foi determinada através da fórmula:  $VP + VN/Total$  mostrando um valor de 0,955. Já a precisão do modelo foi determinada através da fórmula:  $VP/(VP + FP)$  mostrando um valor de 0,966.

A sensibilidade do modelo foi determinada através da fórmula:  $VP/(VP + FN)$  mostrando um valor de 0,948. A especificidade foi determinada através da fórmula:  $VN/(VN+ FP)$  mostrando um valor de 0,963.

A área sobre a curva ROC foi avaliada a fim de se obter mais uma medida de desempenho do modelo de classificação. O cálculo foi obtido a partir do *score* AUC de dez repetições. A Figura 4.14 apresenta a curva ROC calculada para o método. A curva plotada indica um bom desempenho do modelo apresentando um o valor de AUC médio próximo de 1 (0.988). Isso significa que o modelo é capaz de fazer previsões com alto grau de confiança.

**Figura 4.14** - Curva ROC do modelo apresentado mostrando a relação entre a taxa de verdadeiros positivos e a taxa de falsos positivos para um intervalo de limiares de probabilidades de classificação. A área sob a curva (AUC) possui o valor de 0,988



FONTE: A autora (2021).

Uma vez que se conhece a quantidade de soro e os valores dos resultados das análises de cada amostra, o modelo treinado se tornou capaz de reconhecer as características específicas afetadas pela adição de soro. Dessa forma, novas amostras em que se desconhece uma possível fraude por soro podem ser inseridas no modelo e este, por sua vez, classificará cada amostra predizendo se a fraude está presente, com uma certa taxa de acerto (acurácia). Uma alta taxa de acerto significa que o modelo foi treinado com um número representativo de amostras e que ele tem a capacidade de predição satisfatória.

Silva *et al.* (2020) utilizaram diferentes métodos quimiométricos para detectar a adição de leite de vaca no leite de búfala, através da técnica Espectroscopia Infravermelho por Transformada de Fourier associada à refletância total atenuada (FTIR-ATR). Foi feita a adição de leite de vaca ao leite de búfala, variando de 0% a 100%, totalizando 495 amostras. Foram obtidos 15 picos na região espectral entre  $4000\text{ cm}^{-1}$  e  $600\text{ cm}^{-1}$  através dos grupos FTIR correspondentes as variáveis gorduras, proteínas e lactose, cujas absorvâncias foram submetidas à Análise de Componentes Principais (ACP) e Rede Neural Artificial (RNA). O ACP separou as amostras à medida que o leite de vaca era adicionado ao leite de búfala. A RNA mostrou ser mais eficiente, exibindo taxas de classificação de 100% para as amostras de referência e para a maioria dos grupos com a mistura de leite de vaca e de búfala, taxa de classificação total de 95,5% e boa previsão com correlação e erro satisfatórios. Os resultados mostraram a importância do método espectroscópico FTIR-ATR associado à análise quimiométrica para a detecção da autenticidade do leite de búfala.

Nos trabalhos realizados por Tavares (2019) e Asseiss Neto (2020) os resultados dos experimentos mostraram que arquiteturas de rede neural convolucional apresentam desempenho superior em todos os casos, quando comparados a métodos estatísticos tradicionalmente empregados nesse tipo de análise. Em Tavares (2019) avaliou-se amostras de leite adicionadas de amido, sacarose, bicarbonato de sódio, peróxido de hidrogênio e formaldeído, em concentrações, temperaturas e tempos de armazenamento distintos. De acordo com a arquitetura CNN proposta, pode-se prever, com mais de 99% de acurácia, que a amostra analisada estava ou não adulterada (método de triagem).

O estudo de Conceição e colaboradores (2019) teve como objetivo utilizar os dados espectrais de FTIR – ATR (Attenuated Total Reflection) combinados à análise multivariada para identificar adulterações no leite cru. Um total de 620 amostras foram adulteradas com bicarbonato de sódio, hidróxido de sódio, peróxido de hidrogênio, amido, sacarose e ureia. Foram empregados os métodos de análise de regressão linear múltipla e rede neural artificial, que apresentaram média de correlação de apenas 76% para detecção dos adulterantes.

Alves da Rocha *et al.* (2015) usaram a rede neural artificial para identificar soro de queijo em leite fluido com concentrações variando de 0% a 100%, diferentes configurações de rede foram testadas e a melhor configuração obteve  $R^2 > 99,99\%$  usando análises com Raman. Eles concluíram que o uso de microscopia Raman confocal em conjunto com a Rede Neural Artificial é um excelente método para quantificar a adulteração de leite por soro de queijo por causa de sua alta velocidade, simplicidade e eficiência.

Em nenhum dos estudos anteriores pesquisados, que avaliaram a identificação de fraude por soro de queijo no leite cru por espectroscopia FTIR, foi utilizada metodologia de aprendizado de máquina similar com a utilização dos espectros, ou obtidos resultados superiores aos deste trabalho.

## **CONCLUSÕES**

A adição de soro de queijo ao leite altera os níveis de alguns componentes, mas dependendo das quantidades adicionadas, os níveis dos constituintes continuam a atender aos padrões da legislação.

Foram utilizadas técnicas de aprendizado de máquina para identificar fraude de leite com soro de queijo que mostrou uma taxa de acerto de até 97,8%, revelando ser um bom método para detectar esse tipo de fraude, principalmente para triagem analítica.

Os resultados obtidos neste estudo confirmam que a espectroscopia por infravermelho, aliada aos métodos de Aprendizado de Máquina têm grande potencial e podem ser poderosas técnicas analíticas a serem empregadas na garantia de qualidade e autenticidade das indústrias de laticínios em geral.

## **CONSIDERAÇÕES FINAIS**

Os métodos utilizados para analisar a presença de soro de queijo no leite são de alto custo e laboriosos, entre outras desvantagens. A análise por Cromatografia Líquida de Alta Eficiência é o método oficial para detectar esse tipo de fraude, porém os níveis de CMP, podem sofrer variações por ação de bactérias psicrotóxicas, dependendo do tempo e temperatura de estocagem do leite antes da análise.

## REFERÊNCIAS BIBLIOGRÁFICAS

ABADI, M. *et al.* **TensorFlow: A System for Large-scale Machine Learning**. In: Proceedings of the 12th USENIX Conference on Operating Systems Design and Implementation. OSDI'16, p. 265–283. USENIX Association, Savannah, GA, USA, 2016.

ABRANTES, M. R.; CAMPÊLO, C.S; SILVA, J.B.A. Fraude em leite: Métodos de detecção e implicações para o consumidor. **Revista Instituto Adolfo Lutz**. v.73, n.3, p.244-251, 2014.

ABREU, L. R. **Leite e derivados – caracterização físico-química, qualidade e legislação**. Textos acadêmicos. Universidade Federal de Lavras. Ed. Centro de editoração/FAEPE. 2005. 151p.

ALPAYDIN, E. (2014). **Introduction to Machine Learning**. The MIT Press, Cambridge, 3ª edição.

ALVES DA ROCHA, R. da. PAIVA, I. M. ANJOS, V. FURTADO, M. A. M. BELL, M. J. V. Quantification of whey in fluid milk using confocal Raman microscopy and artificial neural network. **Journal of Dairy Science**. v.98, n.6, 2015.

ALVES, PRISCILA MELLO. **Inteligência Artificial e Redes Neurais**. Publicado em 11/06/2020. Disponível em: <https://www.ipea.gov.br/cts/pt/central-de-conteudo/artigos/artigos/106-inteligencia-artificial-e-redes-neurais>. Acessado em: 8 de março de 2021.

ALVIM, T.C. **Efeito da qualidade do leite na detecção de soro lácteo por Cromatografia Líquida de Alto Desempenho – Filtração Gélida (GF- HPLC)**. 69f. Dissertação (Mestrado em Ciência e Tecnologia de Alimentos) – Universidade Federal de Viçosa, Viçosa-MG, 1992.

ANDRADE, E. H. P.; LEITE, M. O.; SOUZA, M. R; FONSECA, L. M. *et al.* Use of caseinomacropetide by high performance liquid chromatography to estimate cheese whey addition in fermented milk beverages. **Journal of Dairy Science**, v. 93, p. 606-607, 2010.

ANTUNES, A. J. Funcionalidade de proteínas do soro de leite bovino. São Paulo: Ed. Manole, p.142, 2003.

AQUINO, L. F.M.C. **Estudo da fraude com soro de queijo em leite através das metodologias de ácido siálico livre, eletroforese SDS-PAGE e análise sensorial**. 70f. Dissertação (Mestrado em Medicina Veterinária) – Universidade Federal Fluminense, Niterói – RJ, 2013.

ARNVIDARSON, B., NYGAARD, L., HANSEN, P. W. Determination of extraneous water in milk samples, or the freezing point depression of milk samples. Foss Electric A/S. Hillerod, Denmark. **United States Patent** N° 5.739.034. Apr, 1998.

ASSEISS, M. S. G. **Aplicação do processo de descoberta de conhecimento em banco de dados acadêmico utilizando as tarefas de agrupamento e classificação**. 115f. Dissertação (Mestrado em Engenharia Elétrica) - UNESP. Ilha Solteira, 2017.

ASSEISS NETO, H. **Metodologia de aprendizado AUTOML baseado em informações de complexidade de instâncias**. 113f. Tese (doutorado em Ciência da Computação - UFMG – Belo Horizonte, 2020.

AYALA, A. R.; LEÓN, S. V.; FLORES, G. P. *et al.* Aplicación de três métodos analíticos para la detección de suero de quesería em leche UHT comercializada em la ciudad de México. **Interciencia**, v. 34, n. 6, p. 406-412, 2009.

BENÍTEZ, E; PONCE, P; NOA, M. Detección de suero de quesería em leche em polvo por HPLC de filtración por gel (GFC-HPLC). **Revista de Salud Animal**, v.23, n. 1, p. 27-31, 2001.

BIGGS, D. A.; JOHNSON, G.; SJAUNJA, L. O. Analysis of fat, protein, lactose and total solids by infrared absorption. In: Monograph on rapid indirect methods for measurement of the major components of milk. **Bulletin of the International Dairy Federation**, n.2008, p.21-29, 1987.

BOENTE, A. N. P.; GOLDSCHMIDT, R. R.; ESTRELA, V. V. **Uma metodologia de suporte ao processo de descoberta de conhecimento em bases de dados**. In: SIMPÓSIO DE EXCELÊNCIA EM GESTÃO E TECNOLOGIA, 5., 2008, Rio de Janeiro. Anais... Rio de Janeiro: Associação Educacional Dom Bosco, 2008. p. 4–5.

BORGES, P.F.Z., SGARBIERI, V.C., DIAS, N.F.G., JACOBUCCI, H.B., PACHECO, M.T.B., BALDINI, V.L.S. Produção piloto de concentrados de proteínas de leite bovino: composição e valor nutritivo. **Brazilian Journal of Food Technology**, Campinas, v. 4, p.1-8, 2001.

BOTELHO, B. G.; REIS, N.; OLIVEIRA, L. S.; SENA, M. M. Development and analytical validation of a screening method for simultaneous detection of five adulterants in raw milk using mid-infrared spectroscopy and PLS-DA. **Food Chemistry**, v.181, p.31-37, 2015.

BRAGA, A. P.; CARVALHO, A. C. P. L. F.; LUDEMIR, T. B. Redes neurais artificiais: teoria e aplicações. **Livros Técnicos e Científicos**, Rio de Janeiro: LTC, 2000.

BRASIL. MINISTÉRIO DA AGRICULTURA, PECUÁRIA E ABASTECIMENTO. Portaria nº 352, de 04 de setembro de 1997. Regulamento Técnico Para Fixação de Identidade e Qualidade do Queijo Minas Frescal. Diário Oficial da União. Brasília, 1997.

BRASIL. Ministério da Agricultura, Pecuária e Abastecimento. Instrução Normativa nº 69, de 13 de dez. 2006. **Institui Critério de Avaliação da Qualidade do Leite in natura, Concentrado e em Pó, Reconstituídos, com base no Método Analítico Oficial Físico-Químico denominado “Índice CMP”**, de que trata a Instrução

Normativa nº68, de 12 de dezembro de 2006. *Diário Oficial da União*, Brasília, Seção 1, p. 67. 15 dez. 2006.

BRASIL. Ministério da Agricultura Pecuária e Abastecimento. Decreto Nº 9.013 de 29 de março de 2017. **Regulamento da Inspeção Industrial e Sanitária de Produtos de Origem Animal**. Brasília, 108p, 2017.

BRASIL. Ministério da Agricultura, Pecuária e Abastecimento. **Manual de métodos oficiais para análise de alimentos de origem animal** / Ministério da Agricultura, Pecuária e Abastecimento. Secretaria de Defesa Agropecuária. – Brasília: MAPA, 2018a.

BRASIL. Ministério da Agricultura, Pecuária e Abastecimento. Instrução Normativa nº 76, de 26 de nov. 2018. **Regulamento técnico de identidade e qualidade de leite cru refrigerado**. *Diário Oficial da União*, Brasília, Edição 230, Seção 1, p. 9, 2018b.

BREMER, M. C. E. G.; KEMMERS-VONCKEN, A. E.M.; BOERS, E. A. M. *et al.* Enzyme-linked immunosorbent assay for the detection of bovine rennet whey powder in milk powder and buttermilk powder. **International Dairy Journal**, v. 18, p. 294-302, 2008.

CARVALHO, B.M.A.; CARVALHO, L.M.; ALCÂNTRA, L.A.P.; BONOMO, R.C.F. Métodos de detecção de fraude em leite por adição de soro de queijo. **Revista Electrónica de Veterinária**, v. 8, n. 6, p. 1-7, 2007.

CEREDA, P. R. M.; JOSE, J. Adaptive data mining: preliminary studies. **IEEE Latin America Transactions**, Piscataway, v. 12, n. 7, p. 1258–1270, Oct 2014.

CHEN, M.S.; HAN, J.; YU, P. S. Data mining: an overview from a database perspective. *IEEE Transactions on Knowledge and data Engineering*, **Los Alamitos**, v. 8, n. 6, p. 866–883, 1996.

CHOLLET, F. *et al.* (2015). **Keras**. Disponível em: <https://keras.io> Acessado em: 8 de março de 2021.

COITINHO, T. B.; CASSOLI, L. D.; CERQUEIRA, P. H. R.; SILVA, H. K.; COITINHO, J. B.; MACHADO, P. F. Adulteration identification in raw milk using fourier transform infrared spectroscopy. **Journal of Food Science and Technology**, v.54, n.8, p.2394-2402, 2017.

CONCEIÇÃO, D. G.; GONÇALVES, B. R. F.; da HORA, F. F.; FALEIRO, A. S.; SANTOS, L. S.; FERRÃO, S. P. B. Use of FTIR-ATR spectroscopy combined with multivariate analysis as a screening tool to identify adulterants in raw milk. **Journal of the Brazilian Chemical Society**, v.30, n.4, p.780-785, 2019.

CONDE, V.A; VALENTE, G.F.S; MINIGHIN, E.C. Milk fraud by the addition of whey using an artificial neural network. **Ciência Rural**, Santa Maria, v.50, n.7, 2020.

CORTEZ, M.A.S; DIAS, V.G; MAIA, RG; COSTA, C.C.A. Características físico-químicas e análise sensorial do leite pasteurizado adicionado de água, soro de queijo, soro fisiológico e soro glicosado. **Revista Instituto Laticínios Cândido Tostes**, n.376, v. 65, p.18-25, 2010.

COUTINHO, E.R.; SILVA, R.M.; DELGADO, A.R.S. Utilização de Técnicas de Inteligência Computacional na Predição de Dados Meteorológicos. **Revista Brasileira de Meteorologia**, v.31, n.1, P.24-36, 2016.

CUNHA, M.F.; BRANDÃO, S.C.C. A coleta a granel pode aumentar os riscos com as bactérias psicrotróficas. **Indústria de Laticínios**, p.71-73. Jul/ago, 2000.

DANTAS, P.B. **Utilização da aprendizagem de máquina e seleção de atributos para o diagnóstico de ceratocone a partir de parâmetros biomecânicos**. 66f. Dissertação (Mestrado em Modelagem Computacional do Conhecimento). Instituto de Computação. Universidade Federal de Alagoas. Maceió, 2017.

DEGANI, A.L.; CASS, Q. L.; VIEIRA, P.C. Cromatografia um breve ensaio. **Química nova na escola**, São Paulo, n. 7, p. 21- 25, 1998.

DELTA INSTRUMENTS, *An Advanced Instruments Company*. **LactoScope FTIR user's guide**. *Milk and Liquid Dairy Product Analyzer*. Netherlands: Delta Instruments, 2007. 138p.

DELTA INSTRUMENTS, *An Advanced Instruments Company*. **SomaScope MKII user's manual**. Netherlands: Delta Instruments, 2009. 44p.

FAWCETT, T. An introduction to ROC analysis. **Pattern Recognition Letters**, v. 27, n. 8, p. 861 – 874, 2006.

FAYYAD, U.; PIATETSKY-SHAPIRO, G.; SMYTH, P. From data mining to knowledge discovery in databases. **AI Magazine**, Palo Alto, v. 17, n. 3, p. 37–53, 1996.

FENNEMA, O.R. **Química de los alimentos**. Espanha: Editora Acribia S.A., 2000, 1280p.

FERNEDA, E. Redes neurais e sua aplicação em sistemas de recuperação de informação. **Ciência da Informação**, Brasília, v. 35, n. 1, p. 25-30, jan./abr. 2006.

FERREIRA, L. D. **Técnicas de aprendizado de máquina aplicadas à identificação de perfis de aprendizado em um ambiente real de ensino**. 73f. Qualificação (Mestrado em Ciências da computação e Matemática Computacional) - USP. São Carlos, 2016.

FIRMINO, F.C; TALMA, S.V; MARTINS, M.L; LEITE, M.O; MARTINS, A.D.O. Detecção de fraudes em leite cru dos tanques de expansão da região de Rio Pomba, Minas Gerais. **Revista Instituto Laticínios Cândido Tostes**, v.65, n. 376, p. 5-11, 2010.

FLORENCIO, ISANNA M.; ELIANE R. FLORENTINO FLÁVIO L. H. DA SILVA, RICARDO S. MARTINS, MÔNICA T. CAVALCANTI; JOSIVANDA P. GOMES. Produção de etanol a partir de lactossoro industrial. **Revista brasileira de engenharia agrícola e ambiental**, Campina Grande, v.17 n.10, 2013.

FOGAÇA, G. N. **Utilização de eletroforese microfluídica na detecção da adição de soro de queijo em leite cru, pasteurizado, UHT e em pó**. 71 f. Dissertação (Mestrado em Ciência e Tecnologia do Leite e Derivados) – Universidade Federal de Juiz de Fora, Juiz de Fora, 2017.

FRIEDRICH, M. T.; FRANKEN, R. B. C.; AZEVEDO, M. S. *et al.* Avaliação da estabilidade do leite in natura e UHT quanto ao índice de CMP. **Revista CIATEC UPF**, Passo Fundo, v. 2, n. 1, p. 21-27, 2010.

FUENTE, M. A.; JUAREZ, M. Authenticity Assessment of Dairy Products. **Critical Reviews in Food Science and Nutrition**, v.45, p.563–585, 2005.

FUKUDA, S. P; ROIG, S. M; PRATA, L. F. Correlation between acidic ninhydrin and HPLC methods to evaluate fraudulent addition of whey in milk. **Le Lait**, v. 84, p. 501-512, 2004.

HOFFMANN, F. L.; SILVA, J. V.; VINTURIM, T. M. Qualidade microbiológica de queijos tipo “minas frescal”, vendidos em feiras livres na região de São José do Rio Preto, SP. **Higiene Alimentar**. São José do Rio Preto, v.16, n.96, p.69-76, maio de 2002.

GABARDO, S.; RECH, R.; AYUB, M. A. Z. Determination of lactose and ethanol diffusion coefficients in calcium alginate gel spheres: Predicting values to be used in immobilized bioreactors. **Journal of Chemical and Engineering Data**, v.56, p. 2305-2309, 2011.

GARCIA, S.C. **O Uso de Árvores de Decisão na Descoberta de Conhecimento na Área da Saúde**. 87f. Dissertação (Mestrado em Ciência da Computação) – Universidade Federal do Rio Grande do Sul. Porto Alegre, 2003.

GOLDSCHMIDT, R. R.; PASSOS, E. P. L. **Data mining: um guia prático - conceitos, técnicas, ferramentas, orientações e aplicações**. Rio de Janeiro: Editora Campus, 2005.

GOYAL, S.; GOYAL, G. K. Artificial neural networks for dairy industry: A review. **Journal of Advanced Computer Science and Technology**, v.1, n.3, p.101-115, 2012.

GUINEE. T. P.; WILKINSON, M. G. Rennet coagulation and coagulants in cheese manufacture. **Journal of the Society of Dairy Technology**, v.45, n.4, p.94-104, 1992.

HAN, J.; KAMBER, M. **Data mining: concepts and techniques**. 3. ed. Boston: Elsevier Science, 2006.

HASTIE, T.; TIBSHIRANI, R.; FRIEDMAN, J. **The Elements of Statistical Learning: Data Mining, Inference, and Prediction**. Springer, 2nd ed., 745p., New York, 2017.

HELPER, G. A.; FERRÃO, M. F.; FERREIRA, C. V. *et al.* Aplicação de métodos de análise multivariada no controle qualitativo de essências alimentícias empregando espectroscopia no infravermelho médio. **Ciência e Tecnologia de Alimentos**, v.26, n. 4, p. 779-786, 2006.

HERNÁNDEZ-RAMOS, P. A.; VIVAR-QUINTANA, A. M.; REVILLA, I. Estimation of somatic cell count levels of hard cheeses using physicochemical composition and artificial neural networks. **Journal of Dairy Science**, v.102, n.2, p.1014-1024, 2019.

HSU, C P.S. **Infrared spectroscopy**. In: SETTLE, F. (Ed). Handbook of Instrumental techniques for analytical chemistry. Cap. 15, p. 243-283. 1997.

JIANG, M.-F.; TSENG, S.-S.; LIAO, S.-Y. **Data types generalization for data mining algorithms**. In: IEEE INTERNATIONAL CONFERENCE ON SYSTEMS, MAN, AND CYBERNETICS- SMC, 12., Tokyo, Anais, v. 5, p. 928–933, 1999.

KAMAL, M.; KAROUI, R. Analytical methods coupled with chemometric tools for determining the authenticity and detecting the adulteration of dairy products: A review. **Trends in Food Science & Technology**, v.46, p.27-48, 2015.

KAROUI, R.; BAERDEMAEKER, J. A review of the analytical methods coupled with chemometric tools for the determination of the quality and identity of dairy products. **Food Chemistry**, v.102, p.621-640, 2007.

KASEMSURAN, S.; THANAPASE, W.; KIATSOONTHON, A. Feasibility of Near-Infrared Spectroscopy to Detect and to Quantify Adulterants in Cow Milk. **Analytical Sciences**, v.23, p.907-910, 2007.

KESAVARAJ, G.; SUKUMARAN, S. A Study On Classification Techniques in Data Mining. IEEE International Conference on Computational Intelligence and Computing Research. **Tiruchengode**, India, july, 2013.

KINGMA, D. P; Ba, J. L. Adam: A method for stochastic optimization. **Conference paper at the 3rd International Conference for Learning Representations**, San Diego, v.1, 2015.

KOSIKOWSKI, F.V. Whey utilization and whey products. **Journal of Dairy Science**. n.62, p.1149-1160, 1979.

KUHN, M. E JOHNSON, K. **Applied Predictive Modeling**. Springer-Verlag, New York, 1ª edição, 2013.

IOFFE, S; SZEGEDY, C. Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift. **32nd International Conference on Machine Learning**, Lille, France, p. 448--456, 2015.

LASMAR, M. M. **Detecção de soro lácteo por cromatografia líquida de alta eficiência em amostras de leite cru conservadas em bronopol**. 38f. Dissertação (Mestrado em Ciência Animal) – Escola de Veterinária, Universidade Federal de Minas Gerais, Belo Horizonte, 2007.

LEITE, M.O. **Fatores interferentes na análise eletrônica da qualidade do leite cru conservado com azidiol líquido, azidiol comprimido e bronopol**. 62f. Tese (Doutorado em Ciência Animal) – Escola de Veterinária, Universidade Federal de Minas Gerais, Belo Horizonte, 2006.

LINN, J. G. *Factors Affecting the Composition of Milk from Dairy Cows*. National Research Council. Washington, 1988.

LIU, J. et al. Deep convolutional neural networks for Raman spectrum recognition: a unified solution. **Analyst**, v.142, p.4067-4074, 2017.

LÓPEZ-FANDIÑO, R. RAMOS, M. Revisión: El caseinomacropéptido bovino. II. Detección de la presencia de suero de quesería en productos lácteos. **Revista Espanhola Ciência e Tecnologia de Alimentos.**, v.33, p.1-12, 1992.

MAAS, A. L.; HANNUN, A. Y. & NG, A. Y. Rectifier Nonlinearities Improve Neural Network Acoustic Models. In **Proceedings of the 30th International Conference on Machine Learning**, Atlanta, USA, 2013.

MADRONA, G. S.; TERRA, C. O.; PENHA, C. B. Efeito da substituição do açúcar por oligofrutose em bebida láctea achocolatada. **Revista Brasileira de Tecnologia Agroindustrial**, v.3, p.29- 37, 2009.

MAFUD, M. D.; ROSSI, R. M.; CAMPOS, E. M.; NEVES, M. F.; SCARE, R. F. Não conformidade na cadeia produtiva do leite: problemas institucionais. In: **XLV CONGRESSO DA SOCIEDADE BRASILEIRA DE ECONOMIA, ADMINISTRAÇÃO E SOCIOLOGIA RURAL**, Londrina: SOBER, p. 1-15, 2007.

MAGALHÃES, M. A. **Determinação de fraude de leite com soro de leite pela análise de CMP e PSEUDO-CMP por cromatografia líquida de alta eficiência em fase reversa com detecção por espectrometria de massa**. 57 f. Dissertação (Mestrado em Ciência e Tecnologia de Alimentos) – Universidade Federal de Viçosa, Viçosa, 2008.

MENDONÇA, L.C.O; MENDONÇA, A.S; SANTANA, A.C.M. Detecção de soro de leite como adição fraudulenta em leites comercializados no município de Itabuna, BA. **Higiene Alimentar** - Vol.31 - nº 266/267 - Março/Abril de 2017.

MEREZE, J; RODRIGUES, L; MARIOTO, M; GONZAGA, N; DANIEL,G.C; TAMANINI, R; BELOTI,V. Detecção de adulterações do leite pasteurizado por meio

de provas oficiais. **Semina: Ciências Biológicas e da Saúde**, Londrina, v. 36, n. 1, supl, p. 283-290, ago. 2015.

METZ, CE. Basic principles of ROC analysis. **Semin Nucl Med.** n. 8, v. 4, p. 283-98, 1978.

MICHALSKI, R. S.; CARBONELL, J. G.; MITCHELL, T. M. **Machine learning: An artificial intelligence approach**. Springer Science and Business Media, 2013.

MOLLÉ, D.; LÉONIL, J.; Heterogeneity of the bovine K-casein caseinomacropéptide, resolved by liquid chromatography on-line with electrospray ionization mass spectrometry. **Journal of Chromatography Analyse**, v.708, p.223- 230, 1995.

MOLLEA, C.; MARMO, L.; BOSCO, F. Valorisation of Cheese Whey, a by-product from the dairy industry. In: **Food Industry**. [s.l.] InTech, 2013. Disponível em: <https://www.intechopen.com/books/food-industry/valorisation-of-cheese-whey-a-by-product-from-the-dairy-industry> Acessado em: 20 de abril de 2021.

MOORE, J. C.; SPINK, J.; LIPP, M. Development and application of a database of food ingredient fraud and economically motivated adulteration from 1980 to 2010. **Journal of Food Science**, Chicago, v. 77, n. 4, p. 118-126, 2012.

MONARD, M. C.; BARANAUSKAS, J. A. Conceitos sobre aprendizado de máquina. **Sistemas Inteligentes-Fundamentos e Aplicações**, v. 1, n. 1, 2003.

MONTÁNEZ, C. D. A.; RAMÍREZ, J. R.; ARANGO, C. J. J.; BETANCOURT, S. D. P. Detección de glucomacropéptido (GMP) como indicador de adulteración com suero de quesería em leche deshidratada. **Veterinaria Mexico**, v.31, n.3, 2000.

MORGANO, M. A.; FARIA, C. G.; FERRÃO, M. F. *et al.* Determination of protein in raw coffee for NIR spectroscopy and regression PLS. **Ciência e Tecnologia de Alimentos**, v.25, n.1, p.25-31, 2005.

NIKAM, S.S. A comparative study of classification techniques in data mining algorithms. **Oriental Journal of Computer Science & Technology**, v.8, n.1, p.13-19, 2015.

OANCEA, S. Identification of glycomacropéptide as indicator of milk and dairy drinks adulteration with whey by immunochromatographic assay. **Romanian Biotechnological Letters**, v. 14, n.1, p. 4.146-4.151, 2009.

OLIVEIRA, K. P. S. **Aplicação das técnicas de Redes Neurais e de análise de componentes principais na modelagem de uma Lagoa Aerada da RIPASA S/ A**. Dissertação (Mestrado em Engenharia Química). Universidade Estadual de Campinas, Campinas. 2000.

OLIVEIRA, G.B; GATTI, M.D.S; VALADÃO, R.C; MARTINS, J.F.P; LUCHESE, R.H. Detecção da adição fraudulenta de soro de queijo em leite: interferência da

atividade de proteases bacterianas. **Revista Instituto Laticínios Cândido Tostes**, n. 367/368, v. 64, p.56-65, Mar/Jun, 2009.

OLIVEIRA, M.C.P.P. **Espectrofotometria no infravermelho por metodologia FTIR (Fourier Transform Infrared): validação da análise do teor de uréia e de outros parâmetros de qualidade do leite**. Dissertação (Mestrado em veterinária). Universidade Federal de Minas Gerais. Belo Horizonte. 2011.

OLIVEIRA, M.C.P.P.; SILVA, N.M.A.; BASTOS, L.P.F. *et al.* Fourier Transform Infrared Spectroscopy (FTIR) for MUN analysis in normal and adulterated milk. **Arquivo Brasileiro de Medicina Veterinária e Zootecnia**, v.64, n.5, p.1360-1366, 2012.

PAVIA, D. L.; LAMPMAN, G. M.; KRIZ, G. S. **Introdução à espectroscopia**. 1.ed. São Paulo: Cengage Learning, 2010. 716 p.

PEDREGOSA, F.; VAROQUAUX, G.; GRAMFORT, A.; MICHEL, V.; THIRION, B.; GRISEL, O.; BLONDEL, M.; PRETTENHOFER, P.; WEISS, R.; DUBOURG, V.; VANDERPLAS, J.; PASSOS, A.; COURNAPEAU, D.; BRUCHER, M.; PERROT, M.; DUCHESNAY, E. Scikit-learn: Machine Learning in Python. **Journal of Machine Learning Research**, v.12, n.2825–2830, 2011.

PERKINS, W. D. Fourier transform-infrared spectroscopy: Part I. Instrumentation. **J. Chem. Educ**, v. 63, n. 1, p. A5-A10, 1986.

PRATA, L.F. **Distribuição de GMP\* livre, em leites crus e pasteurizados, na definição da detecção de fraudes por adição de soro (\*Glicomacropéptídeo)**. Tese de Livre docência apresentada à FCAVJ / Unesp. Jaboticabal, SP. 105p. 2002.

PRATA, L. F. Determinação de GMP (glicomacropéptídeo) no leite por meio da quantificação do ácido siálico - esclarecimentos para adoção e implantação do método. **Revista Indústria de Laticínios**, n. 81, p. 75-80, jul/ago, 2009.

PRATI, R. C.; BATISTA, G. E. A. P. A.; MONARDO, M. C. Curvas ROC para avaliação de classificadores. **Revista IEEE América Latina**, n.6, v. 2, 215-222, 2008.

RECIO, I; FRUTOS de, M; OLANO, A; RAMOS, M. Protein changes in stored ltrahigh-temperature-treated milks studied by capillary electrophoresis and high-performance liquid chromatography. **Journal Agricultural of Food Chemistry**,v.44, n.12, p.3955-3959, 1996.

RECIO, I; GARCÍA- RISCO, M. R; LÓPEZ-FANDIÑO, R; OLANO, A.; RAMOS, M; Detection of rennet whey solids in UAT milk by capillary electrophoresis; **International Dairy Journal** 10, p. 333-338, 2000.

RICHARDS, N.S.P.S. Uso racional de soro lácteo. **Revista Indústria de Laticínios**, v.2, n.9, p. 67-69, 1997.

ROBIM, M.S; CORTEZ, M.A. S; SILVA, A.C.O; FILHO, R.A; GEMAL, N.H; NOGUEIRA, E. B. Pesquisa de fraude no leite UAT integral comercializado no estado do Rio de Janeiro e comparação entre os métodos de análises físico químicas Oficiais e o método de ultrassom. **Revista Instituto Laticínios Cândido Tostes**, n. 389, v. 67, p. 43-50, Nov/Dez, 2012.

RODRIGUEZ-OTERO, J. L; HERMIDA, M; CENTENO, J. Analysis of dairy products by nearinfrared spectroscopy: A review. **Journal of Agricultural and Food Chemistry**, v.45, n.8 p. 2815–2818, 1997.

RODRIGUEZ-SAONA, L. E; KOCA, N; HARPER, W. J. *et al.* Rapid determination of Swiss cheese composition by Fourier transform infrared/attenuated total reflectance spectroscopy. **Journal of Dairy Science**, v.89, n.5, p.1407-1412, 2006.

RODRIGUEZ-SAONA, L. E.; ALLENDORF, M. E. Use of FTIR for Rapid Authentication and Detection of Adulteration of Food. **Annual Review of Food Science and Technology**. v. 2, p. 467–83. 2011.

RUSSELL, S.; NORVIG, P. **Artificial intelligence – a modern approach**. Prentice Hall, v. 2, n. 1, 2003.

SALIBA, E.O.S.; GONTIJO NETO, M.M; RODRIGUES, N.M. *et al.* Predição da composição química do sorgo pela técnica de espectroscopia de refletância no infravermelho próximo. **Arquivo Brasileiro de Medicina Veterinária e Zootecnia**, v.55, n.3, 2003.

SANTOS, J. P. V.; FERREIRA, C. L. L. F. Alternativas para o aproveitamento de soro de queijo nos pequenos e médios laticínios. **Revista do Instituto de Laticínios**, v.56, n.321, p.44-50, 2001.

SANTOS, P.M; PREREIRA-FILHO, E.R; RODRIGUEZ-SOANA, L.E. Rapid detection and quantification of milk adulteration using infrared microspectroscopy and chemometrics analysis. **Food Chemistry**, v. 138, p. 19-24, 2013.

SGARBIERI, V.C. Revisão: Propriedades estruturais e físico-químicas das proteínas do leite. **Brazilian Journal of Food Technology**, v.8, n.1, p.43-56, jan/mar., 2005.

SKOOG, D. A.; LEARY, J.J. **Principles of Instrumental Analysis**, 4.ed., Saunders: London, 1992.

SILVA, P. H. F. **Físico Química do Leite e Derivados – Métodos Analíticos**. Juiz de Fora, Fevereiro 1997, 190p.

SILVA, C. A.; HERMAN-GOMEZ, R. C. Qualidade proteica do soro de leite fermentado pela levedura *Kluyveromyces fragilis*. **Ciência Rural**, v.30, p.515-520, 2000.

SILVA, N.; JUNQUEIRA, V. C. A.; SILVEIRA, N. F. A. **Manual de métodos de análise microbiológica de alimentos**. São Paulo: Livraria Varela, 2001. 105 p.

SILVA, L.K.R; GONÇALVES, B.R.F; DA HORA, F.F; SANTOS, L.S; FERRÃO, S.P.B. Spectroscopic method (FTIR-ATR) and chemometric tools to detect cow's milk addition to buffal's milk. **Revista Mexicana de Ingeniería Química**, v. 19, n. 1, p. 11-20, 2020.

SILVEIRA, T.M.L et al. Comparação entre métodos de referência e análise eletrônica na determinação da composição do leite bovino. **Arquivo Brasileiro de Medicina Veterinária e Zootecnia**, v.56, n.6, p.782-787, 2004.

SILVERSTEIN, R.M.; BASSLER, G.C.; MORRILL, T.C. **Identificação espectrométrica de compostos orgânicos**. 6. ed. Rio de Janeiro: Guanabara, 460pp, 2000.

SMITH, B. **Infrared Spectral Interpretation: A Systematic Approach**. New York: CRC Press. 200pp, 1999.

SOUZA, E.M.T; ARRUDA, S.F; BRANDÃO, P.O; SIQUEIRA, E.M.A. Eletrophoretic analysis to detect and quantify additional whey in milk and dairy beverages. **Ciência e Tecnologia de Alimentos**, v. 20, n. 3, p. 314-317, 2000.

SOUZA, C.M; NICOLAU, E.S; LAGE, M.E; MESQUITA A.J; COELHO K.O. Comparação entre três métodos analíticos para determinação de soro em leite cru refrigerado. **Arquivo Brasileiro de Medicina Veterinária e Zootecnia**, v.70, n.2, p.606-614, 2018.

SRIVASTAVA, N.; HINTON, G.; KRIZHEVSKY, A.; SUTSKEVER, I. & SALAKHUTDINOV, R. Dropout: A Simple Way to Prevent Neural Networks from Overfitting. **Journal of Machine Learning Research**, v. 15, p.1929-1958, 2014.

STUART, B. **Infrared Spectroscopy: Fundamentals and Applications**. John Wiley & Sons. West Sussex. England. 2004, 223p.

TAN, P.; STEINBACH, M.; KUMAR, V. Introdução ao *data mining*. Rio de Janeiro: **Ciência Moderna**, 2009.

TAVARES, W.L.F. **Uso de espectrofotometria FTIR (Fourier Transform Infrared) e mineração de dados para a detecção e identificação de adulterantes no leite cru**. 98f. Tese (Doutorado em Ciência Animal). Universidade Federal de Minas Gerais, Minas Gerais, 2019.

TIBOLA, C. S; DA SILVA, S. A; DOSSA, A. A; PATRÍCIO, D. I. Economically Motivated Food Fraud and Adulteration in Brazil: Incidents and Alternatives to Minimize Occurrence. **Journal of Food Science**, v.83, n.8, p.2028-2038, 2018.

TRONCO, V.M. **Manual para inspeção da qualidade do leite**. 4ª ed. Porto Alegre: UFSM, 2008. 166 p.

TULLIO, L.T. **Isolamento e caracterização do glicomacropéptido do soro de leite.** 97f. Dissertação (Mestrado em Tecnologia de Alimentos) - Universidade Federal do Paraná, Curitiba, 2007.

URBÁN, G; PÉREZ, N; PÉREZ, J; FRESÁN, C; GONZÁLEZ, C; VEJA, S; GUTIÉRREZ, T; DÍAZ, G. Detección de adulteración com suero de quèsería em leches fluidas mexicanas mediante electroforesis em gel de poliacrilamida (SDS-PAGE). **Revista Salud Animal**, v.24, n.1, p. 60-64, 2002.

VALDUGA, E.; PAVIANI, L. C.; MAZUR, S. P.; FINZER, J. R. D. Aplicação do soro de leite em pó na panificação. **Alimentos e Nutrição**, v.17, p.393-400, 2006.

VALENTE, G.F.S; *et al.* Aplicação de redes neurais artificiais como teste de detecção de fraude de leite por adição de soro de queijo. **Revista Instituto Laticínios Cândido Tostes**, Juiz de Fora, v. 69, n. 6, p. 425-432, nov/dez, 2014.

VARGAS, A. C. G.; CARVALHO, A. M. P.; VASCONCELOS, C. N. Um estudo sobre Redes Neurais Convolucionais e sua aplicação em detecção de pedestres. **In: Proceedings of the XXIX Conference on Graphics, Patterns and Images**, 4p., 2016.

VINCIGUERRA, L.L; MARCELO, C.A; MOTTA, T.M; MENEGHINI, L.Z, BERGOLD, A.M; FERÃO, M.F. Chemometric tools and ftir-atr spectroscopy applied in milk adulterated with cheese whey. **Quimica Nova**, v. 42, n. 3, p. 249-254, 2019.

VILLANOEVA, C. N. B. C.; ANDRADE, E. H. P.; BAFFA JUNIOR, J. C. *et al.* Caseinomacropéptide index in UHT whole milk stored under different conditions of temperature and time. **Arquivo Brasileiro de Medicina Veterinária**, v.66, n.1, p.289-296,2014.

WITTEN, I.; FRANK, E.; HALL, M. & PAL, C. Data Mining: **Practical Machine Learning Tools and Techniques**. Morgan Kaufmann, Cambridge, 4ª edição, 2016.

## **COLABORAÇÕES INTERINSTITUCIONAIS PARA ELABORAÇÃO DO PROJETO**

- Departamento de Ciências da Computação/ Instituto de Ciências Exatas – UFMG

- Prof. Sérgio Vale Aguiar Campos (Departamento de Ciências da Computação, UFMG)
- Prof. Habib Asseiss Neto (Instituto Federal de Educação, Ciência e Tecnologia do Mato Grosso do Sul)

## CONTRIBUIÇÕES ACADÊMICAS

### RESUMO - ADSA 2019

**Machine learning applied to Fourier-transform infrared spectroscopy for detection of cheese whey addition to raw milk.** J.S. Lima; D. C. S.Z. Ribeiro; W. L.F.Tavares; H. Asseiss Neto; S. V. A. Campos and L. M. Fonseca. Instituto Federal do Mato Grosso do Sul, Três Lagoas, MS, Brazil; Veterinary School, Universidade Federal de Minas Gerais, Belo Horizonte, MG, Brazil; Department of Computer Science, Universidade Federal de Minas Gerais, Belo Horizonte, MG, Brazil. *Journal of Dairy Science*, v.102, Suppl.1, p.38-39, 2019.

### PRÊMIOS - 2020

- ✓ Grande Prêmio Profa. Conceição Ribeiro da Silva Machado, PRPq/UFMG.
- ✓ Relevância Acadêmica - Semana do Conhecimento, PRPq/UFMG.
- ✓ Destaque - Semana do Conhecimento, PRPq/UFMG.

### APRESENTAÇÃO DE TRABALHO - 2020

CARVALHO, B. P. M; LIMA, J. S. Uso da espectrofotometria FTIR e técnicas de aprendizado de máquina para detecção de fraude por adição de soro de queijo ao leite cru.

**ARTIGO** – Submissão Periódico *Journal of Dairy Science*

LIMA, J. S; RIBEIRO, D.C.S.Z; ASSEISS NETO, H; CAMPOS, S.V.A; LEITE, M.O; FORTINI, M.E.R; FONSECA, L.M. **Classification Tree and Multilayer Perceptron Neural Network for detection of cheese whey illegally added to milk.**