

**UNIVERSIDADE FEDERAL DE MINAS GERAIS**  
**Escola de Veterinária**  
**Departamento de Tecnologia e Inspeção de Produtos de Origem Animal**

Daniela Cristina Solo de Zaldivar Ribeiro

**ESPECTROSCOPIA FTIR, HPLC e REDES NEURAIS ARTIFICIAIS PARA  
DETERMINAÇÃO ANALÍTICA DA LACTOSE RESIDUAL E OUTROS AÇÚCARES  
NO LEITE**

BELO HORIZONTE

2021

Daniela Cristina Solo de Zaldivar Ribeiro

**ESPECTROSCOPIA FTIR, HPLC e REDES NEURAS ARTIFICIAIS PARA  
DETERMINAÇÃO ANALÍTICA DA LACTOSE RESIDUAL E OUTROS AÇÚCARES  
NO LEITE**

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

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

Orientador: Prof. Dr. Leorges Moraes da Fonseca

Coorientadoras: Prof<sup>a</sup> Dra. Kelly Moura Keller e Prof<sup>a</sup> Dra. Débora Cristina Sampaio de Assis

Coorientador: Prof. Dr. Habib Asseiss Neto

Belo Horizonte

2021

R484e Ribeiro, Daniela Cristina Solo de Zaldivar, 1973 -  
Espectroscopia FTIR, HPLC e redes neurais artificiais para determinação analítica da lactose residual e outros açúcares no leite/ Daniela Solo de Zaldivar Ribeiro. - 2021.  
123 f.:il

Orientador: Leorges Moraes da Fonseca

Coorientadores: Kelly Moura Keller

Débora Cristina Sampaio de Assis

Habib Asseiss Neto.

Tese (Doutorado) apresentado à Escola de Veterinária da Universidade Federal de Minas Gerais, como requisito para obtenção do grau de Doutor em Ciência Animal.

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

Bibliografias: f.: 109 a 120.

1. Leite – Análise - Teses – 2. Leite – Qualidade – Teses - 3. Valor nutricional – Teses -  
I. Fonseca, Leorges Moraes da - II. Keller, Kelly Moura – III. Assis, Débora Cristina Sampaio –  
IV. Neto, Habib Asseiss – V. Universidade Federal de Minas Gerais, Escola de Veterinária -  
VI. Título.

CDD – 637

Bibliotecária responsável Cristiane Patrícia 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  
DANIELA CRISTINA SOLO DE ZALDIVAR RIBEIRO

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 27 de maio de 2021, pela banca constituída pelos membros:

Dr.(a). Leorges Moraes da Fonseca

Presidente – Orientador(a)

Dr.(a). Mônica de Oliveira Leite

Dr.(a). Sérgio Vale Aguiar Campos

Dr.(a). Marco Antônio Sloboda Cortez

Dr.(a). Leticia Goulart de Oliveira

Documento assinado eletronicamente por **Leorges Moraes da Fonseca, Professor do Magistério Superior**, em 27/05/2021, às 17:26, 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 **Leticia Goulart de Oliveira, Usuário Externo**, em 28/05/2021, às 20:18, 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 17:01, 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 **Monica de Oliveira Leite, Membro**, em 01/06/2021, às 11:35, 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?acao=documento\\_conferir&id\\_orgao\\_acesso\\_externo=0](https://sei.ufmg.br/sei/controlador_externo.php?acao=documento_conferir&id_orgao_acesso_externo=0), informando o código verificador **0734162** eo código CRC **5FFDCA8C**.

Referência: Processo nº 23072.222241/2021-01 SEI nº 0734162

## Agradecimentos

Meu longo caminho de doutoramento chega ao fim. Imbuída de muita gratidão, de sentimento de realização, de dever cumprido, despeço-me deste período que foi um tanto desafiador, um tanto instigante, excitante, exaustivo, fascinante e turbulento; o caminhar pela pesquisa científica teve vários percalços, mas não tiraram a beleza da ciência. Foi em meio a uma pandemia, que nos tirou a todos dos lugares comuns, e nos obrigou a pensar diferente e a agradecer ainda mais à ciência. Mas não caminhei só, nem poderia ser, uma pesquisa científica é resultado de muita colaboração, compartilhamento, encontros perfeitos, discussões, investigações, descobertas e constatações. Na companhia de pessoas maravilhosas, pude realizar meu estudo, estive com mentes brilhantes, com corações generosos, com mãos amigas e apoios sinceros. É hora de agradecer a todos e a cada um.

Agradeço ao professor Leorges, meu orientador, que me guiou nesta trajetória sinuosa, dividindo conhecimento e doando tempo e esforços. Às minhas coorientadoras, professoras Débora e Kelly, que participaram com generosidade e dedicação. Ao querido amigo, e coorientador, professor Habib, “cérebro” da ciência da computação, que mesmo de longe, possibilitou que esta tese existisse. Ao Daniel, por fazer com que a cromatografia permanecesse neste projeto. Às amigas, Juliana e Wanessa, companheiras de estudo e da vida; ao Ramon, pelos conselhos nas horas mais certas; aos colegas, servidores e professores do DTIPOA, Rommel, Aline, Marco Antônio, prof. Mônica Leite, prof. Cláudia, prof. Cléia, prof. Mônica Pinho, prof. Marcelo, prof. Ronon. Às colaboradoras do laboratório de leite, Marcinha, Rosemary e toda equipe, pela dedicação constante. Aos integrantes da banca examinadora e à Gabi, do Lamico, pelas horas doadas. À UFMG, instituição que tanto me orgulho, e à Fazenda Experimental. À FUNED, pela parceria. À Capes, pelo suporte financeiro (88882.348827/2010-01), FAPEMIG (APQ-02740-17), FINEP e MAPA.

Nesta jornada do conhecimento, pessoal e profissional, meus filhos, Lara e Luan, são meus maiores incentivos. Faço também por eles, ensino-lhes o amor pelo estudo, pela curiosidade e investigação. Divido as alegrias e os desafios de se fazer pesquisa no Brasil. De ser mulher e cientista. De ser mãe e profissional. Agradeço por eles estarem ao meu lado, de mãos dadas, e muitas vezes, me empurrando pra frente. Agradeço à minha família, minha mãe querida, à memória de meu pai, ao Carlinhos, às minhas irmãs, às amigas e amigos que torceram por mim. À Laurinha, pela alegria e amizade. São longos anos, muitas pessoas passaram, dividiram e compartilharam. Sintam-se agradecidos.

Agradeço a Deus, pelos animais e pelo leite, alimento perfeito, fontes de vida e inspiração.

Relembro quando tudo começou, quando meu orientador me convidou para estudarmos a qualidade do leite associando análises físico-químicas, equipamentos eletrônicos e inteligência artificial. Estudaríamos redes neurais artificiais. Confesso que quis dar meia-volta, naquela primeira reação quando levamos um susto; mas a pesquisa científica fascina e a descoberta de novos temas atrai os curiosos. Seguimos buscando sempre mais. E é com a certeza de que fomos felizes na escolha de nossa pesquisa, que fecho esta etapa de minha vida com o coração repleto de felicidade.

“Estou entre aqueles que acham que a ciência tem uma grande beleza.”

Marie Curie (1867-1934)

Aos cientistas que se debruçaram para desenvolver uma vacina, em tempo recorde, contra a COVID-19.

## Resumo

A lactose é o açúcar naturalmente presente no leite e possui grande importância nutricional, porém este sacarídeo não é bem absorvido pela maior parte da população mundial, em virtude da diminuição na produção da enzima lactase, que age no intestino delgado, hidrolisando a lactose ingerida em dois monossacarídeos, a glicose e a galactose. A hidrólise enzimática da lactose pelas indústrias oferece uma diversidade de produtos que contém baixa lactose. A metodologia utilizada para mensurar a concentração dos açúcares no leite após a hidrólise é, preferencialmente, a cromatografia líquida (HPLC) com detecção por índice de refração (RID), pela sua capacidade de distinguir os açúcares. A espectroscopia no infravermelho médio apresenta deficiência nesta distinção devido à absorção da luz pelos açúcares ocorrerem na mesma região do espectro. Neste trabalho foi proposta a associação da espectroscopia FTIR (*Fourier Transform Infrared*) às ferramentas da inteligência artificial por aprendizagem profunda para identificar e quantificar lactose residual e outros açúcares no leite. Foram construídas redes neurais convolucionais a partir dos espectros do infravermelho médio, sem pré-processamento dos dados, para resolver problemas de classificação de amostras de leite e prever qual a classe do açúcar presente, e para a predição dos valores da concentração dos açúcares no leite, um modelo de regressão foi proposto. Leite cru, pasteurizado e UHT foram adicionados de lactose, glicose e galactose em seis concentrações (0.1, 0.5, 1.0, 3.0, 5.0 e 7.0%) e, ao todo, 342 amostras foram submetidas às redes neurais artificiais para identificar e quantificar os açúcares no leite. Além disso, a hidrólise enzimática da lactose pela  $\beta$ -galactosidase foi realizada em 20 amostras de leite pasteurizado em quatro tratamentos diferentes (controle e três níveis de hidrólise) que resultou na presença simultânea dos três açúcares (lactose, glicose e galactose) para quantificação por HPLC-RID e comparação com os resultados previstos pelas redes neurais convolucionais. O resultado do desafio destas amostras aos algoritmos de classificação e regressão indicaram uma capacidade preditiva de 80% para classificação e 75% para quantificação dos açúcares nas amostras hidrolisadas. A associação dos espectros do FTIR às ferramentas de inteligência artificial mostrou ser uma metodologia promissora para o controle de qualidade do leite.

Palavras-chaves: Lactose reduzida. Açúcar. Leite. Cromatografia líquida. Redes neurais artificiais. Aprendizagem profunda.

## ***Abstract***

Lactose is the sugar naturally present in milk and has great nutritional importance, but this saccharide is not well absorbed by most of the world population, due to the decrease in the production of the lactase enzyme, which acts in the small intestine, hydrolyzing the lactose ingested in two monosaccharides, glucose and galactose. The enzymatic hydrolysis of lactose by industries offers a variety of products with low lactose. The methodology used to measure the concentration of sugars in milk after hydrolysis is preferably liquid chromatography (HPLC) with refractive index detection (RID), due to its ability to distinguish sugars. Mid-infrared spectroscopy lacks this distinction due to the absorption of light by sugars occurring in the same region of the spectrum. In this work, the association of the FTIR (Fourier Transform Infrared) spectroscopy with the tools of artificial intelligence for deep learning was proposed to identify and quantify residual lactose and other sugars in milk. Convolutional neural networks were built from the mid-infrared spectra, without preprocessing the data, to solve problems in the classification of milk samples, and to predict which sugar class was present, and to predict the sugar concentration values in milk, a regression model was proposed. Raw, pasteurized and UHT milk were added with lactose, glucose and galactose at six concentrations (0.1, 0.5, 1.0, 3.0, 5.0 and 7.0%) and, in total, 342 samples were submitted to artificial neural networks to identify and quantify the sugars in milk. In addition, the enzymatic hydrolysis of lactose by  $\beta$ -galactosidase was performed on 20 samples of pasteurized milk in four different treatments (control and three levels of hydrolysis) which resulted in the simultaneous presence of the three sugars (lactose, glucose and galactose) for quantification by HPLC-RID and comparison with results predicted by convolutional neural networks. The result of the challenge of these samples to classification and regression algorithms indicated a predictive capacity of 80% for classification and 75% for quantification of sugars in hydrolyzed samples. The association of FTIR spectra to artificial intelligence tools proved to be a promising methodology for milk quality control.

**Keywords:** Low-lactose. Sugar. Milk. Liquid Chromatography. Artificial neural network. Deep learning.

## Lista de figuras

2.1	Estrutura química da lactose.....	16
2.2	Representação esquemática da hidrólise enzimática da lactose por $\beta$ -galactosidase.....	18
2.3	Mapa de prevalência mundial da Hipolactasia Primária.....	21
2.4	Cromatograma de formulação láctea com determinação de açúcares por HPLC-RID.....	32
2.5	Cromatogramas de leite pasteurizado, com injeções de padrões de lactose, glicose e galactose.....	33
2.6	Espectros de absorção de leite cru com adição de sacarídeos, analisados pelo FTIR.....	38
2.7	Espectros FTIR de D-frutose, D-galactose, D-glicose e D-manose na região do C-O ( $1200-900\text{ cm}^{-1}$ ).....	39
2.8	Espectros de infravermelho por transformada de <i>Fourier</i> para leite puro de vaca, cabra e ovelha.....	42
2.9	Espectros de absorção de diferentes adulterantes em leite: (a)água, (b)amido, (c)citrato de sódio, (d)formaldeído, (e)sacarose.....	43
2.10	Espectros obtidos por FTIR da análise de leite de búfala com adição de leite de vaca em diferentes proporções.....	44
2.11	Etapas do processo KDD.....	46
2.12	Arquitetura de uma Rede Neural Artificial.....	50
2.13	Representação esquemática de uma rede neural convolucional para classificação de adulterantes em leite cru.....	53
2.14	Diagrama de validação cruzada de 5 subconjuntos ( <i>5-fold cross-validation</i> ).....	54
2.15	Representação de uma matriz de confusão num arranjo 2 x 2, para um problema genérico de classificação binária.....	57
2.16	Representação de curva ROC ( <i>receiver operator characteristic</i> ) para um modelo genérico de classificação que demonstra a relação entre a taxa de verdadeiros positivos (sensibilidade) e a taxa de falsos positivos ( $1 - \text{especificidade}$ ).....	58
3.1	Esquema de preparação das amostras de leite do Projeto Piloto.	60
3.2	Equipamento CombiScope™ FTIR 400, do Laboratório de Análise da Qualidade do Leite da Escola de Veterinária da Universidade Federal de Minas Gerais (LabUFMG).....	60
3.3	Esquema de preparação das amostras de leite analisadas por FTIR.....	62
3.4	Diagrama de validação cruzada de dez subconjuntos ( <i>10-fold cross-validation</i> ).....	68
3.5	Esquema de preparação das amostras de leite para hidrólise enzimática da lactose.....	71
3.6	Foto do cromatógrafo HPLC-RID LC 20-A Prominence, com amostrador automático (Shimadzu, Japão).....	73

4.1	Espectros de absorção pelo FTIR, de amostras de leite hidrolisado adicionado de lactose, glicose, galactose e sacarose na concentração de 50mg/mL, na região de 940 a 1300cm <sup>-1</sup> .....	78
4.2	Espectros de absorção pelo FTIR, de amostras de água deionizada fortificada com lactose, glicose, galactose e sacarose, na concentração de 50mg/mL, na região de 920 a 1300 cm <sup>-1</sup> .....	79
4.3	Espectro de uma amostra de leite cru, analisado por FTIR.....	80
4.4	Representação esquemática de uma Rede Neural Convolutacional para identificação de açúcares adicionados ao leite.....	82
4.5	Gráficos de acurácia e perda durante o treinamento e validação do modelo CNN para classificação de açúcares adicionados ao leite, treinada por 300 <i>epochs</i> .....	82
4.6	Representação esquemática de uma Rede Neural Convolutacional para identificar a presença de açúcar no leite.....	83
4.7	Gráficos de acurácia e perda durante o treinamento e validação do modelo CNN para classificação de açúcares presentes no leite, treinada por 300 <i>epochs</i> .....	84
4.8	Matriz de confusão do modelo de classificação de açúcares adicionados ao leite.....	84
4.9	Curvas ROC e AUC para as Redes Neurais Convolutacionais executadas para classificação de açúcares no leite.....	85
4.10	Saliências calculadas pela rede neural convolutacional com os espectros das amostras de leite adicionadas de 50 mg/mL de lactose, glicose e galactose.....	87
4.11	Representação esquemática de Rede Neural Convolutacional para quantificar açúcar em leite, com modelo de regressão.....	88
4.12	Gráfico de resíduos dos algoritmos de regressão para quantificar lactose, glicose e galactose.....	89
4.13	Gráficos do histórico de erros e do coeficiente de determinação ao longo das 250 <i>epochs</i> de treinamento da CNN para quantificação de lactose, glicose e galactose no leite.....	90
4.14	Relação e correlação da temperatura do leite (12°C, 30°C e 40°C) e os tempos para atingirem o índice crioscópico limite do estudo, na hidrólise enzimática da lactose com concentração da enzima 40 µL/100mL.....	93
4.15	Relação e correlação das concentrações de enzima β-galactosidase (40, 60 80 e 100 µL/100mL) e os tempos para atingirem o índice crioscópico limite do estudo, na hidrólise enzimática da lactose, a 7°C.....	95
4.16	Curva da lactose em solvente e na matriz, na faixa de trabalho de 0,3mg/mL a 3,0mg/mL, por HPLC-RID.....	96
4.17	Curvas da glicose e galactose em solvente e na matriz, na faixa de trabalho de 0,3mg/mL a 3,0mg/mL, por HPLC-RID.....	96
4.18	Curvas de linearidade da lactose, glicose e galactose, com a equação da regressão linear, o coeficiente de correlação(r) e coeficiente de determinação (R <sup>2</sup> ) nas concentrações de 0,3mg/mL a 3,0mg/mL.....	98

4.19	Cromatograma dos padrões dos analitos em solvente para observação dos tempos de retenção da lactose, glicose e galactose, por HPLC-RID.....	99
4.20	Cromatogramas superpostos da repetição 1, da amostra controle e das amostras com os três protocolos de hidrólise enzimática da lactose.....	100
4.21	Cromatogramas representando as cinco repetições de leite pasteurizado refrigerado, após hidrólise enzimática da lactose com três concentrações de $\beta$ -galactosidase 10 $\mu$ L/100mL, 20 $\mu$ L/100mL e 80 $\mu$ L/100mL, por 24 horas, nos protocolos 1, 2 e 3.....	101

## Lista de tabelas

2.1	Teor de lactose no leite de diferentes espécies.....	16
2.2	Prevalência de hipolactasia primária do adulto e intolerância à lactose em diferentes populações.....	20
2.3	Prevalência da hipolactasia primária e intolerância à lactose, por etnia, faixa etária e região do Brasil.....	22
2.4	Teor de lactose em alguns alimentos.....	23
3.1	Regiões do espectro de absorção do infravermelho por alguns carboidratos.....	69
4.1	Valores considerados como “lactose” pelo equipamento FTIR em diferentes amostras de leite cru sem hidrólise, leite cru hidrolisado e leite cru hidrolisado (LH) com adição de açúcares.....	77
4.2	Hiperparâmetros utilizados para controlar o treinamento da rede neural convolucional.....	81
4.3	Resultados das métricas de desempenho dos modelos de regressão com as redes neurais convolucionais em duas diferentes regiões espectrais.....	86
4.4	Resultados das métricas de desempenho dos modelos de regressão para quantificar lactose, glicose e galactose, após 10- <i>fold</i> validação cruzada.....	91
4.5	Influência da temperatura do leite no tempo da reação enzimática da lactose (40 µL/100mL de enzima), a partir do índice crioscópico (IC) e percentual de hidrólise .....	94
4.6	Concentração de lactose, glicose e galactose em leite pasteurizado, antes (C) e após 3 protocolos de hidrólise enzimática da lactose (H1, H2, H3), em 5 repetições (R1 a R5), quantificados por HPLC-RID).....	103
4.7	Resultados da previsão do algoritmo de classificação da rede neural convolucional com 4 neurônios na camada de saída.....	104
4.8	Comparação entre valores previstos pelo algoritmo e valores reais pela cromatografia para lactose, glicose e galactose em leite pasteurizado, antes (controle) e depois da hidrólise enzimática da lactose em três níveis (H1, H2 e H3), em cinco repetições.....	105

# Sumário

I-Introdução .....	12
1.1 Hipótese.....	13
1.2 Objetivos .....	14
II-Referencial Teórico .....	15
2.1 Lactose.....	15
2.2 $\beta$ -galactosidase .....	17
2.3 Intolerância à lactose .....	19
2.3.1 Intolerância à lactose e Alergia à Proteína do Leite de Vaca (APLV) .....	24
2.4 Hidrólise industrial da lactose.....	25
2.5 Determinação analítica da lactose .....	29
2.5.1 Quantificação da lactose e outros açúcares no leite por HPLC .....	30
2.6 Análise do leite por Espectroscopia FTIR .....	35
2.7 Mineração de dados ( <i>data mining</i> ).....	45
2.8 Aprendizado de Máquina .....	46
2.8.1 Regressão e Classificação.....	48
2.8.2 Redes Neurais Artificiais (RNA) .....	49
2.8.3 Redes Neurais Convolucionais .....	51
2.9 Separação de conjunto de dados.....	53
2.10 Validação Cruzada.....	54
2.11 Métricas de desempenho .....	55
III-Metodologia Aplicada.....	59
3.1 Projeto Piloto.....	59
3.2 Análises do leite pelo método FTIR .....	61
3.3 Aprendizado de Máquina .....	64
3.4 Otimização das condições de hidrólise enzimática da lactose.....	69
3.5 Quantificação de açúcares pelo método HPLC.....	72
3.6 Comparação entre métodos HPLC e FTIR/redes neurais.....	75
IV-Resultados e Discussão .....	76
4.1 Resultados preliminares do Projeto Piloto.....	76
4.2 Resultados do FTIR para construção das Redes Neurais Convolucionais .....	79
4.3 Resultados dos algoritmos.....	81
4.4 Resultados da otimização da hidrólise da lactose.....	93

4.5 Resultados da Cromatografia.....	95
4.6 Resultados da comparação entre métodos.....	104
V-Conclusões.....	107
VI-Referências.....	109
VII-Colaborações interinstitucionais para elaboração do projeto.....	121
VIII-Contribuições Acadêmicas .....	122

## Introdução

A lactose é o principal carboidrato do leite, representando em torno de 5% de sua composição química. É um dissacarídeo formado por dois monossacarídeos, a glicose e a galactose, unidos por uma ligação  $\beta$ -1,4-glicosídica. Possui grande importância nutricional, sendo a principal fonte de energia do leite, auxilia na absorção de vitaminas, cálcio, magnésio, tem baixo índice glicêmico, baixo risco cariogênico e possui potencial prebiótico no trato intestinal. A lactose tem várias aplicações na indústria alimentícia por fornecer boa textura e cor; e tem doçura menor do que a sacarose e a glicose. Porém, estima-se que mais da metade da população mundial sofra de intolerância à lactose, que é a incapacidade de digerir a lactose por deficiência ou ausência da enzima intestinal chamada  $\beta$ -galactosidase ou, simplesmente, lactase. Esta enzima possibilita a hidrólise da molécula da lactose em dois monossacarídeos, glicose e galactose, que serão absorvidos pelo intestino. O acúmulo da lactose no lúmen intestinal leva à fermentação pela microbiota presente, causando desconfortos como distensão, gases, cólica e diarreia. Assim, alternativas alimentares são importantes para pessoas com intolerância à lactose, pois um copo de leite que contém aproximadamente 10g de lactose é suficiente para desencadear os sintomas, embora pessoas que possuam alta sensibilidade possam apresentar os sintomas com doses muito baixas, como as encontradas em alguns medicamentos quando a lactose é utilizada como excipiente.

A hidrólise enzimática, pela ação da enzima  $\beta$ -galactosidase, é um método eficiente para redução dos teores de lactose no leite e nos produtos lácteos e sua eficiência depende da concentração da enzima e das condições do processamento, como temperatura e pH.

A ANVISA publicou as RDC's 135 e 136 em 08/02/2017 (Brasil, 2017), que dispõem sobre os alimentos para dietas com restrição de lactose e a declaração obrigatória na rotulagem da presença de lactose. Ficou estabelecido que os alimentos que contenham a lactose devam indicar sua presença nos rótulos e aqueles que tenham alterado o teor original de lactose, devam informar o teor de lactose remanescente. Com estas novas exigências, cresceu a necessidade por métodos laboratoriais mais rápidos e mais precisos para a determinação de lactose nos alimentos.

A espectroscopia no infravermelho com Transformada de Fourier (*Fourier Transform Infrared* - FTIR) permite análise simultânea dos componentes do leite, em uma única amostra, sem tratamento prévio, com rapidez analítica e possibilidade de automatização total do processo. A absorção das ligações O-H da lactose pela radiação infravermelha no comprimento de onda de 9,61 $\mu$ m permite a determinação quantitativa da lactose (AOAC 972.16). Os monossacarídeos do leite hidrolisado, a glicose e a galactose, interferem na leitura da lactose pelo FTIR, devido à sobreposição das ondas de absorção dos espectros dos açúcares (Ribeiro *et al.*, 2018), sendo necessário a utilização de novas ferramentas analíticas, a fim de se assegurar uma metodologia com precisão e seletividade para determinação da lactose residual de produtos hidrolisados.

Dessa forma, o presente trabalho visa utilizar os dados analíticos resultantes dos equipamentos de espectroscopia por infravermelho FTIR para a identificação e quantificação de açúcares presentes e residuais no leite, em associação ao uso de ferramentas de inteligência artificial.

## **1.1 Hipótese**

a) A mineração dos dados espectrais e o uso de redes neurais artificiais são capazes de distinguir os açúcares presente no leite hidrolisado, identificá-los e quantificá-los, sem interferência da sobreposição das ondas de absorção, quando analisado por FTIR.

## 1.2 Objetivos

- a) Identificar diferenças nos espectros de absorção pelos açúcares no leite analisado por FTIR;
- b) Associar a espectroscopia FTIR às redes neurais artificiais para identificar e quantificar lactose residual e outros açúcares no leite;
- c) Treinar modelos utilizando algoritmos de classificação e algoritmos de regressão para detecção de açúcares adicionados ao leite, por meio de pontos no espectro do infravermelho;
- d) Otimizar a hidrólise da lactose utilizando a enzima  $\beta$ -galactosidase em diferentes tempos, temperaturas e concentrações;
- e) Realizar a quantificação da lactose, glicose e galactose pelo método *High Performance Liquid Chromatography* (HPLC);
- f) Comparar os resultados da cromatografia líquida HPLC com os resultados da espectroscopia FTIR associada às redes neurais artificiais.



## Referencial Teórico

Este capítulo contém uma revisão bibliográfica da importância da lactose e outros sacarídeos em leite, a aplicação da hidrólise enzimática, sua contribuição na área da saúde, o impacto causado quando os indivíduos não conseguem digerir a lactose normalmente, as soluções apresentadas pela indústria, o posicionamento das agências reguladoras, a necessidade dos laboratórios de qualidade de leite em monitorar os produtos com lactose reduzida, as ferramentas disponíveis para a realização das análises laboratoriais e as novas possibilidades que se apresentam no campo da inteligência artificial.

### 2.1 Lactose

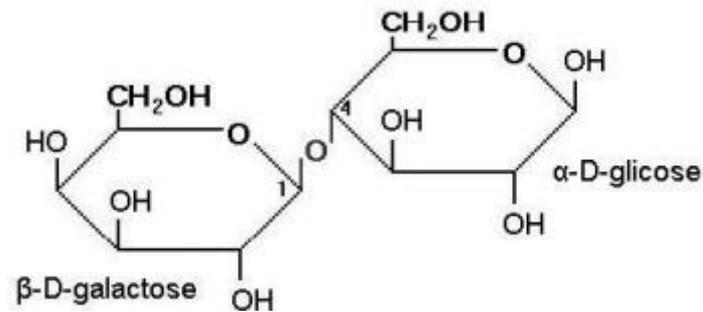
A lactose é o açúcar naturalmente presente no leite, sintetizada na glândula mamária, a partir da glicose sanguínea. Nos ruminantes, a glicose sanguínea é derivada principalmente da gliconeogênese no fígado que utiliza como substrato o ácido propiônico, um ácido graxo volátil absorvido do rúmen (Moraes, 2016). É o constituinte predominante e menos variável da matéria seca do leite e quantitativamente o mais importante dos sólidos não graxos.

Industrialmente, a fermentação da lactose por ação dos microrganismos que a transformam em ácido láctico ocupa lugar de grande destaque, sendo utilizada para obtenção de vários derivados lácteos como iogurtes, leites fermentados, queijos, requeijões, ácido láctico e outros (Pereira *et al.*, 2012).

A lactose é um dissacarídeo constituído por um radical  $\beta$ -D-galactose e um radical  $\alpha$ -D-glicose e é considerada um açúcar redutor, porque o grupo hidroxila no carbono

anomérico da porção glicose não está envolvido na ligação glicosídica, portanto, ela está livre para reagir com agentes oxidantes (Campbell, 2000). Na Figura 2.1 está representada a estrutura química da lactose.

Figura 2.1- Estrutura química da lactose



Fonte: Campbell (2000)

A lactose é encontrada apenas no leite, apresentando diferentes concentrações entre as espécies mamíferas. Em 100g de leite de vaca existem 4,9g de lactose e em 100g de leite humano, 7g de lactose, aproximadamente. A concentração de lactose no leite dos mamíferos apresenta uma considerável variabilidade entre as diferentes espécies (Tabela 2.1) e mantém uma relação inversamente proporcional com as concentrações das proteínas e gorduras (Fagundes-Neto, 2009; Mattar e Mazo (2010).

Tabela 2.1- Teor de lactose no leite de diferentes espécies

Fêmea da espécie	Lactose (%)
Mulher	7,0
Égua	6,1
Vaca	5,0
Ovelha	4,8
Cabra	4,3
Camela	3,3
Cadela	3,3
Rena	2,8
Rata	2,6
Coelha	2,0
Golfinho	1,1

Fonte: Fagundes-Neto (2009); Mattar e Mazo (2010); Fox *et al.* (2015); Moraes (2016) - adaptado

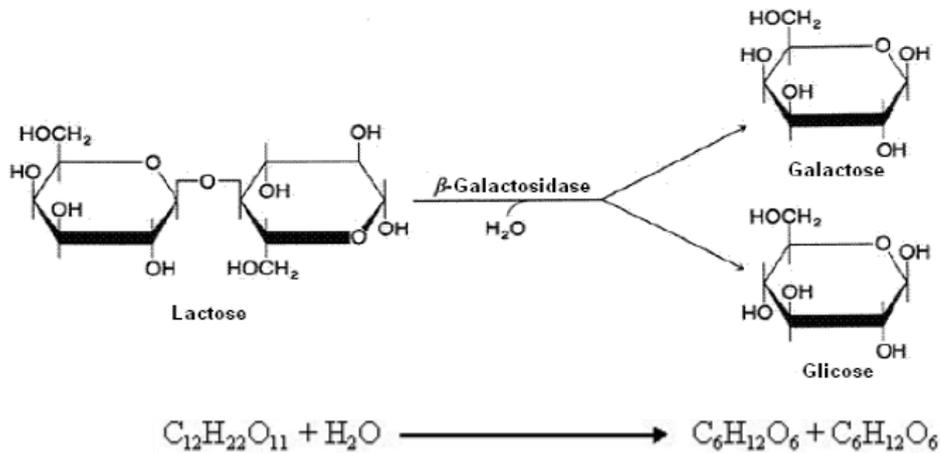
A lactose é encontrada no leite de todos os mamíferos, sendo que a maior concentração deste carboidrato é verificada no leite materno humano. Com exceção para o leite da foca, *Zalophus californianus*, que não possui carboidratos e a glicose encontrada no sangue dos filhotes durante o período da amamentação é derivada primariamente do glicerol da gordura do leite da foca (Fagundes-Neto, 2009).

A lactose é a principal fonte energética dos alimentos lácteos e é muito utilizada na indústria alimentícia, pois acentua os aromas e sabores dos alimentos. Participa da Reação de Maillard, uma reação de escurecimento não-enzimático responsável pela cor e aroma de produtos submetidos ao aquecimento ou a longos períodos de estocagem, onde a presença de um açúcar redutor é essencial para que ocorra a reação com o grupamento amino livre de aminoácidos, peptídeos ou proteínas (Francisquini *et al.*, 2017). Além disso, ela apresenta várias funcionalidades nos diversos setores químicos e industriais. Utilizada nas formulações infantis, a fim de aproximar o teor de lactose ao do leite materno; utilizada na indústria farmacêutica, como excipiente de medicamentos, na fabricação do lactitol e lactulose (laxantes e prebióticos); e como substrato para fermentação na produção de penicilinas (Brito, 2007).

## 2.2 $\beta$ -galactosidase

Lactase é o nome usualmente utilizado para denominar a enzima  $\beta$ -D-galactoside galactohidrolase ou  $\beta$ -galactosidase (Figura 2.2). É classificada como uma hidrolase e catalisa, entre outras, a reação de hidrólise da lactose em  $\beta$ -D-galactose e  $\alpha$ -D-glicose (Trevisan, 2008).

Figura 2.2- Representação esquemática da hidrólise enzimática da lactose por  $\beta$ -galactosidase



Fonte: Fischer (2010)

As  $\beta$ -galactosidasas podem ser encontradas na natureza, distribuídas entre vegetais, particularmente amêndoas, pêssego, damasco e maçã, e em órgãos de animais como intestino, cérebro, testículos, placenta. Também são produzidas por grande quantidade de microrganismos, tais como fungos filamentosos, bactérias e leveduras, sendo as leveduras e fungos as fontes preferidas para aplicações comerciais (Santiago *et al.*, 2004). As  $\beta$ -galactosidasas pertencem ao grupo de enzimas que têm despertado crescente interesse por parte da indústria de alimentos, pois a hidrólise da lactose permite o consumo de leite e outros produtos lácteos por pessoas intolerantes a este dissacarídeo, sem os efeitos adversos causados pela má absorção de lactose no intestino.

A legislação brasileira específica, por meio da RDC 53 de 07/10/2014/ANVISA (Brasil, 2014), que a enzima lactase, utilizada na indústria de alimentos, deva ser de origem microbiana, proveniente dos seguintes microrganismos: *Aspergillus niger*, *Aspergillus oryzae*, *Candida pseudotropicalis*, *Kluyveromyces fragilis*, *Kluyveromyces lactis*, *Kluyveromyces marxianus* e *Saccharomyces sp.* Tais espécies são classificadas como *Generally Recognized As Safe* (GRAS) pela *Food and Drug Administration* (FDA), sendo esse um importante critério para aplicações alimentícias (Klein, 2014).

A lactase está presente em todos os segmentos do intestino delgado, em todos os mamíferos, com exceção dos pinípedes (foca, p.ex.), cujo leite não tem lactose. Também em todos os mamíferos, e em aproximadamente 75% da população mundial, os níveis desta enzima caem significativamente após o desmame, em níveis próximos a 10% de sua atividade anterior. Essa condição leva algumas populações, nas quais a atividade lactásica é muito pequena, a desenvolver sintomatologia digestiva após a ingestão de leite (Fagundes-Neto, 2009; Lopez, 2009; Mattar *et al.*, 2012). A lactose que não for hidrolisada pela lactase não será absorvida e será fermentada pela microbiota intestinal, principalmente no cólon, levando à formação de gases e ácidos que causam os sintomas da intolerância à lactose (IL), como diarreia, distensão e dor abdominal (Mattar e Mazo, 2010).

### **2.3 Intolerância à lactose**

Os estudos epidemiológicos mostram que as populações que nos seus primórdios dependiam da pecuária mais do que da agricultura, e eram grandes consumidores de leite e laticínios em geral, apresentam menor prevalência de hipolactasia primária em relação àquelas que dependeram mais da agricultura para sobreviver (Montgomery *et al.*, 2007; Mattar e Mazo, 2010) (Tabela 2.2).

Alguns autores preferem utilizar a terminologia intolerância à lactose (IL) quando se refere aos sintomas ocasionados pela má digestão da lactose na ausência ou redução da enzima lactase no intestino delgado. E a hipolactasia primária se refere à condição genética de algumas populações e/ou indivíduos que têm a redução da enzima (“lactase não-persistente”) após a primeira infância (Fagundes-Neto, 2009; Mattar e Mazo, 2010; Faedo *et al.*, 2013). As duas condições nem sempre estão relacionadas (Wortmann, 2014; IDF, 2021).

Tabela 2.2- Prevalência de hipolactasia primária do adulto e intolerância à lactose em diferentes populações

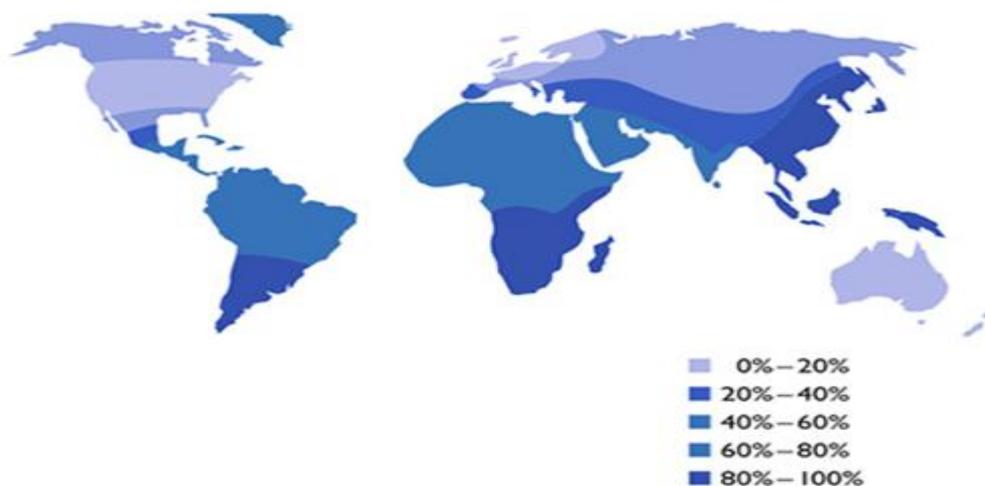
<b>Local</b>	<b>Prevalência (%)</b>	<b>Método</b>
Alemanha	14,8	Hidrogênio expirado
Áustria	20,1	Hidrogênio expirado
Brasil (brancos)	57	Genético
Brasil (crianças índias Terenas)	89,3 após 4 anos	Hidrogênio expirado
Brasil (japoneses)	100	Genético
Brasil (negros)	80	Genético
China (Norte)	87,3	Genético
Estônia (ligado aos finlandeses)	24,8	Genético
França	23,4	Hidrogênio expirado
Hungria	37	Genético
Índia (norte)	67,5	Genético
Índia (sul)	86,8	Genético
Itália	51(Norte) 71(Sicília)	Hidrogênio expirado
Japão (adultos)	89	Hidrogênio expirado
Jordânia (beduínos)	24	Hidrogênio expirado
Jordânia (oeste) e Palestina	75	Hidrogênio expirado
Rússia (nordeste)	35,6	Genético
Sibéria (oeste, nativos Khants)	94	Hidrogênio expirado
Somalis	76	Hidrogênio expirado
Sudão (tribo Béja, pecuaristas)	16,8	Hidrogênio expirado
Sudão (tribo Nilotes, agricultores)	74,5	Hidrogênio expirado
Suécia (crianças caucasianas)	10	Genético
Suécia (crianças não-caucasianas)	66	Genético
Suécia (idoso caucasiano)	5	Genético
Tuaregues	12,7	Hidrogênio expirado
Turquia	71,3	Hidrogênio expirado

Fonte: Mattar e Mazo (2010)-adaptado

Em um estudo realizado com 1088 indivíduos que apresentavam distúrbios gastrointestinais ou sintomatologia suspeita de intolerância à lactose, 44,11% apresentaram hipolactasia no teste de curva glicêmica (Pereira Filho e Furlan, 2004). Wortmann (2014), num estudo com 408 indivíduos, encontrou a condição genética para hipolactasia primária em 44,7% dos que apresentavam dispepsia funcional (sintomas digestivos) e em 50,7% dos assintomáticos.

A hipolactasia primária varia de acordo com a etnia e afeta mais de 75% da população mundial, dos quais aproximadamente 5% ocorrem no norte da Europa, e mais de 90% em alguns países da Ásia e da África (Moreira *et al.*, 2009; Mattar *et al.*, 2012 e Zychar e Oliveira, 2017). A distribuição populacional da intolerância à lactose no mundo está representada na Figura 2.3.

Figura 2.3- Mapa de prevalência mundial da Hipolactasia Primária



Fonte: <https://www.yamoo.org/patient-info/lactose-intolerance/183-how-common-is-lactose-intolerance-> adaptado.

No Brasil, estudos mostram prevalências de 5% a 45% na região Sul (Pretto *et al.*, 2002; Wortmann, 2014), com aumento significativo na região Nordeste (75%) (Sparvoli, 1989). Segundo Mattar e Mazo (2010), 43% da população brasileira apresenta o alelo da persistência da lactase, sendo a hipolactasia mais frequente entre os negros e japoneses (Tabela 2.3).

A intolerância à lactose pode ser classificada em três grupos, que representam o modo como se manifesta: congênita, primária e secundária. A intolerância congênita, manifestada em recém-nascidos, é uma condição extremamente grave e também muito rara, onde a enzima lactase está ausente ou modificada por condição genética (Longo, 2006; Mattar e Mazo, 2010).

Tabela 2.3- Prevalência de hipolactasia primária e intolerância à lactose, por etnia, faixa etária e região do Brasil

<b>Autor (Ano)</b>	<b>Método diagnóstico</b>	<b>Prevalência (%)</b>	<b>Etnia</b>	<b>N</b>	<b>Faixa etária (anos)</b>	<b>Região do Brasil</b>
<b>Sevá-Pereira (1982)</b>	Teste de tolerância à lactose	45	Caucasóides	80	20-52	Sudeste
		85	Negróides			
		100	Mongolóides			
<b>Sparvoli (1989)</b>	Teste de tolerância à lactose	75,7	Triíbridos (Caucasóides Negróides Mongolóides)	37	18-59	Nordeste
<b>Sparvoli (1990)</b>	Teste de tolerância à lactose	46	Caucasóides	70	19-59	Sul
		73	Negróides			
<b>Pretto (2002)</b>	Teste do hidrogênio expirado	5,2 15,5	Branca Não-branca	225	8-18	Sul
<b>Alves (2002)</b>	Teste do hidrogênio expirado	89,3	Indígenas	251	0 a 10	Centro-oeste
<b>Escoboza (2004)</b>	Imunohistoquímica de biópsias intestinais (endoscópicas)	53,2	Branca	115	5-60	Sudeste
		91,3	Não-branca			
<b>Mattar (2009)</b>	Teste genético (SNP – 13910C/T)	57	Branca	567	Idade média 42,1(±16,8) Limites não informados	Sudeste
		80	Negra			
		100	Orientais			

Fonte: Wortmann (2014)-adaptado

Na maioria dos mamíferos, a atividade da enzima lactase na parede intestinal diminui após o desmame, caracterizando a hipolactasia primária que provoca sintomas de intolerância à lactose. A intensidade dos sintomas de distensão, flatulência, dor abdominal e diarreia variam, dependendo da quantidade de lactose ingerida, e aumentam com o passar da idade. A hipolactasia é determinada geneticamente, porém uma mutação ocorreu e parte da humanidade tornou-se tolerante ao leite na idade adulta (lactase persistente). A hipolactasia também pode ser secundária a doenças que causem dano na borda em escova da mucosa do intestino delgado ou que aumentem significativamente o tempo de trânsito intestinal, como nas enterites infecciosas, giardíase, doença celíaca, doença inflamatória intestinal, enterites induzidas por drogas ou radiação, doença diverticular do cólon e anemia. Diferentemente da hipolactasia primária do adulto, a hipolactasia secundária é transitória e reversível (Mattar e Mazo, 2010; Pereira *et al.*, 2012).

A maioria dos adultos intolerantes à lactose pode consumir uma pequena quantidade da mesma (6 a 12g) sem os principais sintomas, especialmente quando consumida com refeições ou na forma de iogurte com culturas ativas, mas isso varia de uma pessoa a outra (Trevisan, 2008; IDF 2021). Indivíduos com alta sensibilidade podem apresentar queixas gastrointestinais após a ingestão de pequenas quantidades de lactose encontrada em medicamentos orais, utilizada como excipiente na fabricação dos mesmos (Nathan *et al.*, 2008). Enquanto alguns indivíduos podem tolerar em torno de 240 mL de leite, que contém aproximadamente 12g de lactose, em outros os sintomas podem aparecer com a ingestão de 1 a 2g de lactose, o que seria o consumo de uma porção de 100 gramas de queijo não maturado (Tabela 2.4). Como forma de reduzir os sintomas, os indivíduos evitam o consumo de produtos contendo lactose, ou fazem uso da enzima lactase junto com a ingestão dos produtos lácteos.

Tabela 2.4- Teor de lactose em alguns alimentos

<b>Alimento</b>	<b>% lactose</b>
Soro de leite em pó	69
Leite em pó integral	37,5
Leite condensado	12,3
Sorvete elaborado com leite	5,7
Leite integral / semi-desnatado / desnatado	5,0
Iogurte integral	4,0
Chantily	2,9
Ricota	2,0
Queijos	1,0
Manteiga	0,9
Margarina elaborada com leite	0,5
Queijos maturados	<0,5
Leite com lactose reduzida	0

Fonte: Furtado (2018?); IDF (2021) - adaptado

No entanto, o consumo de leite e seus derivados contribuem para a ingestão de numerosos nutrientes essenciais, entre eles a lactose, que age como um promotor na absorção e na retenção de cálcio no intestino e na absorção de magnésio e manganês. Também prolonga a ação da vitamina D, em caso de redução da radiação solar, e ajuda na prevenção do raquitismo e da osteomalácia. A redução ou

eliminação do leite e seus derivados da dieta de crianças intolerantes à lactose pode comprometer a absorção de proteína e riboflavina, além do cálcio. Portanto é recomendada a adição de cálcio nos produtos lácteos sem lactose ou com quantidade de lactose reduzida, uma vez que a absorção desse mineral no intestino é baixa na ausência de lactose (Longo, 2006; Mattar *et al.*, 2012; IDF, 2021).

Existem vários métodos para medir a tolerância à lactose. A técnica mais difundida nos laboratórios de análises clínicas é a curva glicêmica, que é realizada como um desafio, onde o paciente ingere de 25g a 50g de lactose e se avaliam os sintomas por duas a três horas. Em um estudo com 51 indivíduos, 48 deles apresentaram concordância nos resultados entre o exame de glicemia e o exame genético (Ridefelt e Håkansson, 2005).

O teste respiratório do hidrogênio expirado é considerado padrão-ouro para o diagnóstico de intolerância à lactose, seguindo o primeiro Consenso de Roma para doenças gastrintestinais. Com a validação do polimorfismo *LCT-13910C>T*, o exame genético passou a incorporar a rotina de exames de laboratório para diagnosticar hipolactasia. O diagnóstico que se baseia em medir a atividade da enzima lactase em biópsia intestinal com o kit Quick Lactase Test®, embora também tenha apresentado sensibilidade e especificidade de 100% comparado ao teste genético de hipolactasia, é um exame invasivo, dependente de endoscopia para a coleta da biópsia da segunda porção do duodeno, e a atividade da enzima lactase no duodeno é menor que no jejuno (Mattar e Mazo, 2010).

### **2.3.1 Intolerância à lactose e Alergia à Proteína do Leite de Vaca (APLV)**

É importante estabelecer a diferença entre APLV e intolerância à lactose (Pardo *et al.*, 2021). A alergia à proteína do leite de vaca é uma reação imunológica adversa às proteínas do leite, que se manifesta após a ingestão de leite ou derivados, e que pode provocar alterações gastrointestinais, na pele e no sistema respiratório (tosse e

bronquite, por exemplo) (Ferreira *et al.*, 2014). Entre os alérgenos, têm-se os subtipos da caseína (alfa, beta e kappa), que compõe cerca de 80% do total proteico do leite, e as proteínas do soro lácteo com destaque para beta-lactoglobulina, alfa-lactoalbumina, soroalbumina e imunoglobulinas. As reações alérgicas variam de sintomas leves até anafilaxia e acontecem até duas horas após exposição ao alérgeno. Estas reações podem ser mediadas por IgE ou não. Para diagnosticar APLV é necessária uma avaliação individualizada do paciente com sintomas para esta condição clínica (Araújo *et al.*, 2017; Silva *et al.*, 2021).

No Brasil, estudos demonstram que APLV é a alergia mais comum em crianças abaixo de quatro anos de idade, com prevalência de 5,4% (Araújo *et al.*, 2017; Silva *et al.*, 2021). Diferentemente do que ocorre com a intolerância à lactose, onde a redução da enzima lactase é mais grave na idade adulta (Branco *et al.*, 2017).

#### **2.4 Hidrólise industrial da lactose**

Hoje, muitos países, inclusive o Brasil, possuem grande variedade de produtos com baixo teor de lactose. A hidrólise da lactose é uma das principais tecnologias aplicadas para a produção de produtos lácteos com baixo teor de lactose, sendo um processo promissor para a indústria de alimentos, pois possibilita o desenvolvimento de novos produtos sem lactose em suas composições (EMBRAPA, 2020). Essa operação oferece certas vantagens tecnológicas, na medida em que diminui os riscos de cristalização nos derivados lácteos e aumenta o poder adoçante (Moreira, 2009).

Segundo o Anuário de leite, elaborado pela Embrapa, em 2019, houve uma grande mudança de hábitos revelada em pesquisa realizada no ano anterior em 11.300 lares, que identificou que 20% das famílias brasileiras já compraram algum produto sem lactose. Enquanto o leite longa vida apresentou uma redução nas suas taxas anuais de crescimento, a categoria de leites sem lactose cresceu 7,9% em volume e

12,6% em valor. Nos Estados Unidos, a tendência é semelhante, enquanto as vendas totais de volume de leite fluido no varejo caíram 2,4% em 2017, o volume de vendas de leite sem lactose aumentou 11,5% (EMBRAPA, 2019). Em 2020, o Anuário da Embrapa apresentou um cenário onde os lácteos representaram os alimentos funcionais mais consumidos no mercado mundial, deixando para trás cereais, carnes, peixes, ovos e outros. A procura por alimentos que promovam benefícios à saúde intestinal recai também ao consumo de produtos sem lactose, como leite, iogurte e manteiga (EMBRAPA, 2020).

Existem dois métodos principais para a hidrólise da lactose: o método ácido e o enzimático. O método ácido não é utilizado na indústria de alimentos, pois depende de altas temperaturas e utilização de ácidos fortes causando problemas tecnológicos como desnaturação de proteínas, alteração de cor e sabor nos alimentos. A hidrólise pelo método enzimático é o método preferencialmente utilizado em escala industrial. A reação é catalisada pela enzima lactase ( $\beta$ -galactosidase), que pode ser adicionada antes ou após o tratamento térmico. A vantagem deste método reside no fato de que a reação processa-se em condições amenas de pH e temperatura relativamente baixa, em uma faixa que pode variar de 4 a 40°C, sendo a temperatura ótima de 30 a 40°C, permitindo uma maior economia energética, além de não se formarem subprodutos indesejáveis (Longo, 2006; Fischer, 2010).

A crioscopia é um método de mensuração da eficiência da hidrólise da lactose. O ponto de congelamento tem alta correlação com o grau de hidrólise (Nijpels *et al.*, 1980; Longo, 2006). A redução no ponto de congelamento deve-se ao fato de que com a hidrólise da lactose ocorre um aumento dos açúcares redutores do leite, que passa a apresentar, além da lactose, glicose e galactose resultantes da reação. Com isto aumenta-se a concentração dos constituintes solúveis no leite, reduzindo o índice crioscópico, isto é, o ponto de congelamento. A crioscopia é um método indireto para mensurar o teor de lactose residual em produtos hidrolisados, desde que se conheça o valor inicial da lactose (Colinas *et al.*, 2006). Trevisan (2008), ao estudar a influência de duas marcas de enzimas sobre a hidrólise de leite pasteurizado, utilizou a crioscopia para mensurar a eficiência das mesmas e encontrou redução de 33,6% e 44,1% do ponto de congelamento, e uma hidrólise de

72,3% e 88,6%. O teor de lactose foi estimado em 1,4g/100g e 0,5g/100g, mensurando-se a glicose residual por espectrofotometria.

Produtos lácteos fermentados e leite com baixo teor de lactose têm sido recomendados para pessoas com intolerância à lactose, que podem usufruir dos nutrientes presentes no leite como as proteínas, o cálcio e a vitamina B, por exemplo (IDF, 2021). Entre os produtos fermentados, o iogurte apresenta uma tolerância atribuída à alta atividade da lactase pelos microrganismos usados na produção do iogurte. Porém, Longo (2006) encontrou uma redução de apenas 10,22% no teor de lactose em iogurte integral, em comparação ao iogurte com hidrólise enzimática que reduziu em 88,07% o teor de lactose. Os queijos possuem teor de lactose muito baixo, resultado do seu processamento tecnológico, onde a lactose se apresenta dissolvida no soro e não na massa do queijo. A lactose residual nos queijos está em torno de 1 a 2% dependendo da tecnologia empregada, do tipo de queijo, da lavagem da massa e do período de maturação (Furtado, 2018?). Queijos maturados podem conter traços de lactose (IDF, 2021).

Faedo *et al.* (2013) realizaram um estudo para obtenção de leite com baixo teor de lactose por processos de separação por membranas associados à hidrólise enzimática e utilizaram o método do ácido dinitrossalicílico (DNS) para mensurar a lactose. O ácido DNS reage com o açúcar redutor formando um composto reduzido e corado cuja absorção máxima de luz se dá a 540 nm (Santos *et al.*, 2017). O leite desnatado utilizado na pesquisa de Faedo *et al.* (2013) possuía um teor inicial de lactose de 5,86% e o leite elaborado apresentou 0,39% de lactose, com redução de 93,34%, superior à observada em leites com baixo teor de lactose encontrados nos mercados (90%).

A hidrólise da lactose apresenta outras vantagens tecnológicas como a prevenção da cristalização da lactose em produtos como sorvete, doce de leite e leite condensado, melhorando assim suas características sensoriais, como suavidade e cremosidade, pois a solubilidade da lactose é menor do que a de outros açúcares (Brito, 2007). Além disso, durante a hidrólise da lactose, pode haver a formação

simultânea de galactooligosacarídeos (GOS) que são carboidratos formados por duas a nove unidades de galactose e uma glicose terminal, e são considerados como um ingrediente prebiótico (Cardelle-Cobas *et al.*, 2011; Lisboa *et al.*, 2012). A aplicação de GOS em alimentos tem aumentado consideravelmente nos últimos anos devido às suas importantes propriedades funcionais, como por exemplo: melhoria da microbiota intestinal, alívio de transtornos gastrointestinais, prevenção de tumores, melhoria na absorção de minerais como o cálcio, e, além disso, a possibilidade de utilização como adoçantes naturais não-cariogênicos e de baixa caloria (Klein, 2014).

Um outro método pode remover a lactose e, conseqüentemente os açúcares residuais da hidrólise por separação cromatográfica. O leite sem lactose e sem os carboidratos resultantes da hidrólise não apresenta o gosto doce do leite hidrolisado, e possui redução de, aproximadamente, 60% no valor energético (Fischer, 2010). Porém, esta técnica não é comumente utilizada em escala industrial, ao contrário da hidrólise enzimática.

- **Legislação para lactose nos alimentos**

Até 2018, a Portaria 29, de 1998, do Ministério da Saúde (Brasil, 1998) era o documento legal que regulamentava os alimentos especialmente formulados para atender portadores de intolerância à ingestão de dissacarídeos como a lactose. Esses alimentos poderiam conter no máximo 500 mg do nutriente por 100 g ou 100 mL do produto final.

Em 2017, as RDC's 135 e 136 (ANVISA) definiram requisitos de composição e rotulagem para alimentos isentos e com baixo teor de lactose. Os alimentos classificados como "isentos de lactose" são aqueles em que a quantidade de lactose foi reduzida a um valor igual ou menor a 100 mg por 100 g ou 100 mL do alimento pronto para o consumo de acordo com as instruções de preparo do fabricante. Já os classificados como "baixo teor de lactose" são os que tiveram o teor deste açúcar

reduzido a uma quantidade maior que 100 mg e igual ou inferior a 1 g por 100 g ou 100 mL do alimento pronto para o consumo (Brasil, 2017). Em fevereiro de 2018, entrou em vigor a RDC 136, que obriga a declaração nos rótulos de produtos que contenham lactose (Brasil, 2017b). Comparativamente, na Alemanha, os produtos considerados “isentos de lactose” devem ter no máximo 10 mg de lactose por 100 g do alimento (GDCh, 2005).

## 2.5 Determinação analítica da lactose

Existem vários métodos para determinação de lactose em leite e derivados (Gambelli, 2017), tais como: glícidos redutores (método Lane-Eynon) (Brasil, 2006), crioscopia (Longo, 2006; Trevisan, 2008), *High Performance Liquid Chromatography* - HPLC (cromatografia líquida de alta eficiência) (AOAC 980.13) (Xinmin *et al.*, 2008; Erich *et al.*, 2012; Acquaro Jr *et al.*, 2013); espectroscopia no infravermelho médio (AOAC 972.16) (Ribeiro *et al.*, 2018; Hansen e Holroyd, 2019).

O método Lane-Eynon era o método oficial do Ministério da Agricultura Pecuária e Abastecimento-MAPA para mensurar glícidos redutores até setembro de 2018, pela IN 68/2006, revogada pela IN 30/2018 (Brasil, 2018a). Porém, este método não distingue os glícidos resultantes da hidrólise enzimática da lactose. Longo (2006) utilizou o método Lane-Eynon para mensurar lactose do leite pasteurizado, antes da hidrólise, utilizado como matéria-prima para o iogurte. Para analisar o teor de lactose residual no iogurte hidrolisado foi utilizado o método HPLC com detector por índice de refração.

O método HPLC tem sido citado como possuidor de alta sensibilidade e especificidade (Erich *et al.*, 2012; Gambelli, 2017), sendo a técnica mais amplamente utilizada para determinar açúcares em leite e derivados lácteos. Atualmente, é o método utilizado pela Rede de Laboratórios Federais de Defesa Agropecuária (Brasil, 2018c) e também pelo Laboratório Central de Saúde Pública

do Estado de Minas Gerais (Lacen-MG). Esse método é usado para separar e determinar os sacarídeos e, acoplado ao detector de índice refratométrico (RID) é uma poderosa técnica de quantificação de vários tipos de sacarídeos (ISO 22662/IDF198:2007; Xinmin *et al.*, 2008; Acquaro Jr *et al.*, 2013).

Como técnica de análise quantitativa para o controle de alimentos e produtos industrializados, a espectroscopia no infravermelho obteve crescimento após o advento da Transformação de Fourier e a utilização do interferômetro de Michelson, o que tornou o método mais rápido e robusto (Cienfuegos e Vaitsman, 2000; Saliba *et al.*, 2003). O uso do FTIR (*Fourier Transform Infrared*) permite análise simultânea dos componentes do leite, em uma única amostra, sem tratamento prévio, com rapidez analítica e possibilidade de automatização total do processo. A absorção da energia nas ligações dos grupos hidroxila da lactose pela radiação infravermelha no comprimento de onda de 9,61µm permite a determinação quantitativa da lactose (AOAC 972.16).

### **2.5.1 Quantificação da lactose e outros açúcares no leite por HPLC**

O método HPLC ocupa um lugar de destaque pela sua facilidade em efetuar a separação, identificação e quantificação das espécies químicas. É um método físico-químico de separação de componentes de uma mistura, realizada através da distribuição desses componentes entre duas fases. Uma das fases permanece estacionária, enquanto a outra se move através dela. Durante a passagem da fase móvel sobre a fase estacionária, os componentes da mistura são distribuídos entre as duas fases, de tal forma que cada um dos componentes seja seletivamente retido pela fase estacionária, resultando em migrações diferenciais (Collins *et al.*, 1997).

A diferenciação entre os carboidratos presentes no leite é realizada com precisão somente por HPLC, em comparação aos métodos polarimétrico, gravimétrico, enzimático, diferencial de pH e espectrofotometria no infravermelho, segundo Kouaouci (2007).

Os métodos 984.15 e 2006.06 recomendados pela AOAC (*Association of Official Analytical Chemists*) baseiam-se na hidrólise enzimática da lactose em glicose e galactose, seguida da oxidação da  $\beta$ -galactose em ácido galacturônico e redução do  $\text{NAD}^+$  (nicotinamida adenina dinucleotídeo). A determinação da lactose residual é calculada pela quantidade formada de NADH medido por espectrofotometria a 340nm (Silva K. *et al.*, 2020).

O método HPLC com diferentes tipos de detectores é considerado o método de escolha para mensurar lactose em produtos lácteos com lactose reduzida porque é direto, automatizado e capaz de diferenciação entre carboidratos (Erich *et al.*, 2012).

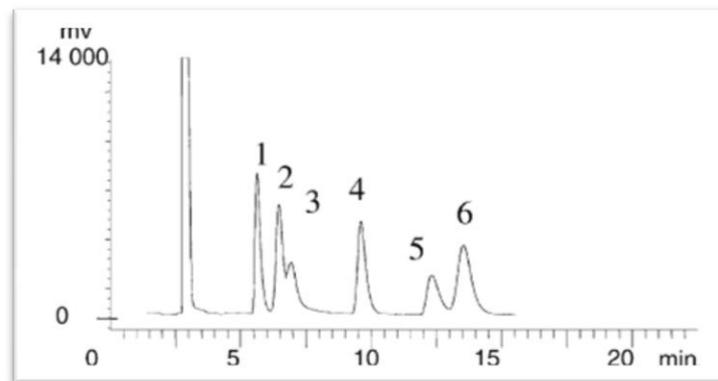
O método mais utilizado de HPLC obtém medições usando um detector de índice de refração (RID) com diferenciação entre vários monossacarídeos, dissacarídeos e oligossacarídeos do leite (AOAC 980.13; ISO 22662/IDF 198:2007; Xinmin *et al.*, 2008; Fusch *et al.*, 2011; Acquaro Jr *et al.*, 2013). O índice de refração requer alto controle da temperatura, pressão e pureza dos reagentes. Sem essas precauções a resposta fica susceptível à imprecisão com alguma mudança no índice de refração. Apesar de suas desvantagens, o detector de índice de refração é comumente usado para análise de mistura de açúcares, uma vez que não é necessário a derivatização (Sexton, 2004).

O detector por índice de refração mede as alterações na refração da luz à medida que flui através de uma célula de medição. Quando a luz de refração muda de direção ao atingir a célula, a quantidade de luz que chega ao refrator diminui, e essa diferença na incidência da luz na célula contendo a amostra e na célula contendo a referência, resulta na concentração da amostra. Este tipo de detector não é adequado para eluição por gradiente, pois o eluente é sempre comparado ao solvente puro em uma célula de referência (Talamona, 2005).

Chávez-Servin *et al.*, em 2004, desenvolveram e validaram um método para separar mono e dissacarídeos em fórmulas lácteas e extraí-los de componentes que possam interferir no sistema, como as proteínas. Foi utilizado o método de HPLC-RID com uma coluna para carboidratos Tracer (250 x 4,6mm, 5  $\mu\text{m}$ ) e uma pré-coluna  $\text{NH}_2$

(13 x 3 mm). Na extração foram utilizados os reagentes de Carrez I e II, etanol, acetonitrila e metanol; como fase móvel foi usado acetonitrila:água (75:25, v/v), em um fluxo de 1,8ml/min, a 25°C. Os resultados permitiram determinar frutose, glicose, galactose, sacarose, lactulose e lactose. Porém, os picos de glicose e galactose ficaram sobrepostos (*overlapped*) (Figura 2.4).

Figura 2.4- Cromatograma de formulação láctea com determinação de açúcares por HPLC-RID



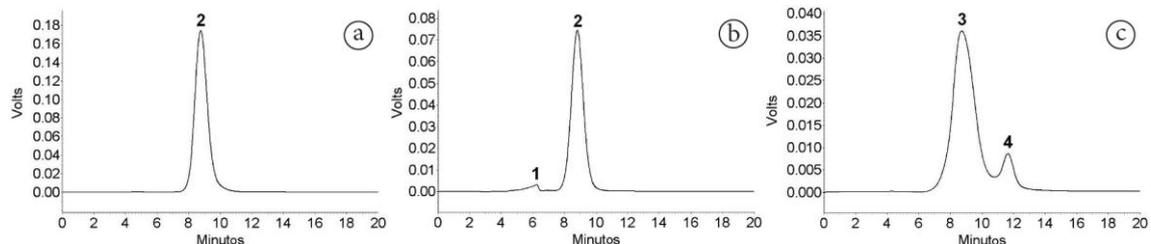
Legenda: picos (1) frutose; (2) glicose; (3) galactose; (4) sacarose; (5) lactulose; (6) lactose.

Fonte: Chávez-Servin *et al.* (2004)

Xinmin *et al.* (2008) estabeleceram e validaram um método analítico sensível, confiável e seletivo para a determinação de glucosamina e lactose em formulações lácteas por um método de HPLC-RID. A separação cromatográfica foi obtida com uma Coluna Shodex Asahipak NH2P-50 (250 x 4,6mm, 5  $\mu$ m). A separação foi realizada a 35°C utilizando água-acetonitrila (30/70, v/v) como fase móvel e com fluxo de 1mL/min. O volume de injeção foi de 20  $\mu$ L. O tempo de corrida foi de 10 minutos. Os picos foram identificados comparando tempo de retenção dos analitos com o tempo de retenção dos padrões dos açúcares. As áreas dos picos foram utilizadas para a análise quantitativa. As curvas de calibração para cada açúcar foram preparadas em seis níveis, de 0,2 a 10 mg/mL para glucosamina e 1 a 20 mg/mL para lactose, todos dissolvidos em água. Para a precipitação das proteínas vários reagentes foram testados, como ácido acético, acetato de zinco, ferrocianeto de potássio, acetato de chumbo, oxalato de sódio e hidrogenofosfato de sódio. O método proposto apresentou boa linearidade, precisão e acurácia, e mostrou sensibilidade apropriada para a determinação de glucosamina e lactose, simultaneamente, em fórmulas à base de leite.

Acquaro Jr *et al.* (2013) validaram um método para extração e quantificação de lactose em leite pasteurizado, pelo método HPLC-RID. O tempo de centrifugação, a intensidade da rotação e a proporção da diluição entre solvente/amostra foram analisadas estatisticamente como variáveis independentes. Com taxa de recuperação média de 105%, a proporção de etanol:leite foi de 4:1 (v/v), com rotação de 10.000rpm por 37 minutos. A corrida cromatográfica em HPLC-RID utilizou uma coluna de troca catiônica de Pb(II)-Aminex HPX-87P (300 x 7,8 mm, 9  $\mu$ m), a 80°C. A água, empregada como fase móvel, correu com fluxo de 0,8 mL/min. A validação foi avaliada pela seletividade, precisão, linearidade, faixa dinâmica de trabalho, LOD, LOQ e exatidão pelas taxas de recuperação. O método proposto apresentou boa resolução entre lactose e lactulose. Porém, quando injetados padrões de glicose e galactose, houve coeluição da lactose com galactose. E a presença de um pico aos 6,9 minutos deveu-se a presença de lactulose, composto resultante do aquecimento da pasteurização, formado a partir da isomerização alcalina da lactose, conforme demonstrado nos cromatogramas, na Figura 2.5.

Figura 2.5- Cromatogramas de leite pasteurizado, com injeções de padrões de lactose, glicose e galactose



Legenda: (a) Cromatograma do padrão de lactose 5,0% (m/v); (b) leite pasteurizado integral comercial com extração realizada utilizando leite: etanol (4: 1, v/v), em 10.000 rpm, por 37 minutos; (c) padrão de lactose 1,5%, galactose 0,25% e glicose 0,25% (m/v); picos: (1) lactulose, (2) lactose, (3) lactose e galactose, (4) glicose.

Fonte: Acquaro Jr *et al.* (2013)

A detecção UV-Visível é desejável, pois oferece a seletividade e a sensibilidade que faltam na detecção do índice de refração. No entanto, esta técnica requer a derivatização do açúcar, a fim de obter a absorção em comprimentos de onda mais altos, para evitar a interferência de outros compostos orgânicos, aumentando assim a seletividade da técnica. Sexton (2004) aplicou o método HPLC com detector UV-

Vis, derivatizou as amostras com PABA (ácido para-aminobenzoico) e determinou o teor de lactose com precisão em fórmulas infantis, sucos infantis e leite em pó desnatado.

O método HPLC acoplado ao detector UV-Vis ou fluorescência apresenta desvantagens por estes detectores precisarem de grandes quantidades de reagentes e de uma redução adicional com borohidreto de sódio. Além disso, a presença de grandes quantidades de glicose e galactose no leite hidrolisado também pode resultar em interferências (Garbalo-Rubio *et al.*, 2018).

Em geral, a detecção por fluorescência é mais sensível que o UV e tem um limite de detecção inferior. No entanto, sem derivatização, este método de detecção não é aplicável à maioria dos açúcares, porque eles não têm um fluoróforo. A detecção de fluorescência oferece alta sensibilidade e boa seletividade e muitas técnicas para a derivatização de carboidratos para detecção de fluorescência têm sido propostas. Uma vez que a derivatização é realizada, também pode tornar o carboidrato de interesse detectável por detectores de UV. Isto é vantajoso porque os detectores UV são de baixo custo e prontamente disponíveis para uso na maioria dos laboratórios analíticos (Sexton, 2004).

A técnica HPLC de fase reversa (RP-HPLC) consiste na utilização de uma fase estacionária de material apolar, geralmente hidrocarboneto, com uma fase móvel polar. Exemplos de fase estacionária apolar seriam as colunas C8 ou C18 que consistem em uma cadeia de oito ou dezoito hidrocarbonetos de carbono ligada a um suporte sólido, como a sílica. Solventes polares tais como metanol, água ou uma combinação deles podem ser utilizados como fase móvel. A HPLC de fase reversa é a técnica de cromatografia líquida mais popular, com ampla aplicabilidade a uma diversidade de grupos de compostos. Uma vez escolhida a fase estacionária, a fase móvel pode ser alterada para aumentar ou melhorar a separação (Sexton, 2004).

Erich *et al.* (2012) utilizaram o método de RP-HPLC acoplado ao detector UV-Vis e fluorescência, derivatização pré-coluna com PABA, para quantificar lactose em leite desnatado hidrolisado. O método foi validado pela avaliação dos parâmetros:

linearidade, taxa de recuperação, limite de detecção, limite de quantificação e reprodutibilidade. O método se mostrou adequado para a investigação de alimentos rotulados "sem lactose" que, no país em estudo, Alemanha, pode conter até 10mg por 100g de produto lácteo.

Um novo método para determinação de açúcares em produtos lácteos foi proposto em 2017 e aceito em 2020 pela AOAC (Sanders *et al.*, 2017; Brunt *et al.*, 2021). O método ISO 22184/IDF244:2021 usando cromatografia de troca aniônica de alta performance com detector de amperometria pulsada (HPAEC-PAD) detecta inclusive galactose e lactulose que até o momento não possuem protocolos padronizados, além da glicose, frutose, sacarose, lactose e maltose. No entanto, este método não é aplicável para quantificar lactose em produtos com lactose reduzida, com teores abaixo de 1mg/g (0,1%).

## **2.6 Análise do leite por Espectroscopia FTIR**

A região do infravermelho médio e as técnicas de refletância para análises em alimentos vêm sendo amplamente utilizadas desde o início dos anos noventa. Nesta década, inúmeros estudos revelaram as aplicações da espectroscopia na análise de leite, carne, óleos, gorduras e frutas, avaliando aspectos quantitativos e qualitativos, bem como, para obtenção de certificação de qualidade (Reeves e Zapf, 1998).

Um espectrômetro FTIR obtém espectros infravermelhos ao coletar o sinal do interferograma de uma amostra com um interferômetro, que mede todas as frequências infravermelhas simultaneamente. Desta forma, a técnica permite uma rápida triagem, quantificação de componentes e, portanto, um alto rendimento de amostras (Luykx e Van Ruth, 2008).

O pressuposto básico da aplicação das técnicas de espectroscopia baseia-se na geração da "impressão digital" dos alimentos. Um produto lácteo com uma dada composição química exposta a uma fonte de luz tem um espectro característico, que

é um resultado da absorção de vários componentes químicos. Nesta técnica, são detectadas as frequências específicas de energia absorvida por uma molécula dentro de um grupo funcional, uma vez que uma molécula ao receber a radiação no infravermelho médio (400 a 4000  $\text{cm}^{-1}$ ) atinge um estado vibracional excitado, passando para um nível de energia mais alto (Luykx e Van Ruth, 2008; Karoui *et al.*, 2010; Kamal e Karoui, 2015; Su *et al.*, 2018). O resultado final desta absorbância é um espectro de infravermelho altamente específico.

O princípio de todos os analisadores de infravermelho se baseia na capacidade de absorção de radiação dos grupos químicos específicos de alguns componentes do leite como gordura, proteína e lactose, em diferentes comprimentos de ondas. Os grupos CH em cadeias de ácidos graxos de moléculas de gordura absorvem radiação no comprimento de onda de 3,480  $\mu\text{m}$ , os grupos carbonila (C=O) das ligações éster dos triglicerídeos absorvem em 5,723  $\mu\text{m}$ , os grupos amina (CONH) das ligações peptídicas das proteínas em 6,465  $\mu\text{m}$  e os grupos hidroxila (OH) da lactose em 9,610  $\mu\text{m}$  (AOAC 972.16).

A técnica que revolucionou os laboratórios de análise de leite fluido foi a espectroscopia de infravermelho médio, também chamada de MIR (*mid-infrared*), que apresenta grandes vantagens em relação às técnicas anteriores, como a determinação simultânea de proteína, gordura e lactose em uma única amostra sem tratamento prévio, com rapidez analítica e possibilidade de automatização total do processo (Rodriguez-Otero *et al.*, 1997). A espectroscopia no infravermelho médio é um dos métodos mais utilizados para análise do leite e de produtos lácteos. Isto porque quase todas as substâncias químicas (exceto alguns sais e compostos químicos muito simples) têm o seu próprio espectro distinto; apenas as substâncias que ocorrem em concentrações muito baixas podem ser difíceis de se determinar (Luykx e Van Ruth, 2008; Karoui *et al.*, 2010).

A espectroscopia de infravermelho com transformada de Fourier (FTIR) vem ganhando espaço no controle de qualidade do leite cru, principalmente por ser um método capaz, potencialmente, de resolver muitos destes problemas de adulteração, pois é uma técnica rápida, de fácil execução, que exige o mínimo de preparação da

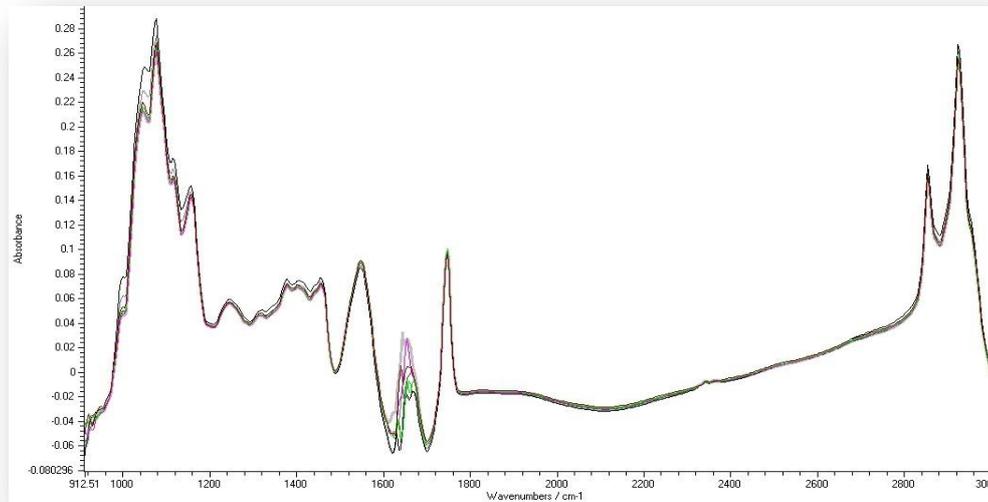
amostra, além de apresentar alta sensibilidade e baixo custo operacional (Pappas *et al.*, 2008; Nicolau *et al.*, 2010; Ribeiro *et al.*, 2018; Conceição *et al.*, 2019; Hansen e Holroyd, 2019; Tavares, 2019; Silva, L. *et al.*, 2020).

A Norma Internacional ISO 9622/IDF 141:2013 fornece diretrizes para a análise quantitativa da composição do leite e produtos lácteos líquidos, como leite cru, leite processado, creme e soro de leite, através da medição da absorção da radiação infravermelha média, aplicáveis à análise do leite de vaca. As diretrizes também são aplicáveis à análise de leite de outras espécies (cabra, ovelha, búfalo, etc.) e derivados de produtos lácteos líquidos, desde que sejam geradas calibrações adequadas para cada aplicação e procedimentos de controle adequados estejam em vigor. A aplicação é limitada a produtos de baixa viscosidade que podem ser bombeados através do sistema de fluxo do analisador e para analitos que não resultem em saturação óptica nos comprimentos de onda específicos que estão sendo utilizados (ISO 9622/IDF 141:2013).

A espectroscopia por infravermelho nem sempre pode eliminar a necessidade de análises laboratoriais mais detalhadas, mas pode ajudar a rastrear amostras que requerem mais investigação. Um cuidado especial é necessário com relação a possível interferência entre componentes que possuem regiões espectrais similares, que ocorrem com muita frequência quando produtos alimentares são analisados (Luykx e Van Ruth, 2008; Hansen e Holroyd, 2019).

O FTIR apresenta uma fragilidade em relação à presença de ingredientes extras, como açúcares, que levam à variação nas respostas do equipamento (Lactoscope FTA, 2010). Silva (2013) e Ribeiro *et al.* (2018) encontraram valores mais altos de lactose pelo FTIR quando o leite foi adicionado de sacarose, demonstrando a interferência dos açúcares na determinação do teor de lactose (Figura 2.6). Esta interferência na leitura da lactose por FTIR, na presença de sacarídeos no leite, deve-se à curva PLS (*Partial Least Squares*) pré-estabelecida no equipamento considerar os mesmos pontos de absorbância no espectro infravermelho (Ribeiro *et al.*, 2018; Conceição *et al.*, 2019).

Figura 2.6- Espectros de absorção de leite cru com adição de sacarídeos, analisados pelo FTIR



Fonte: Ribeiro *et al.* (2018)

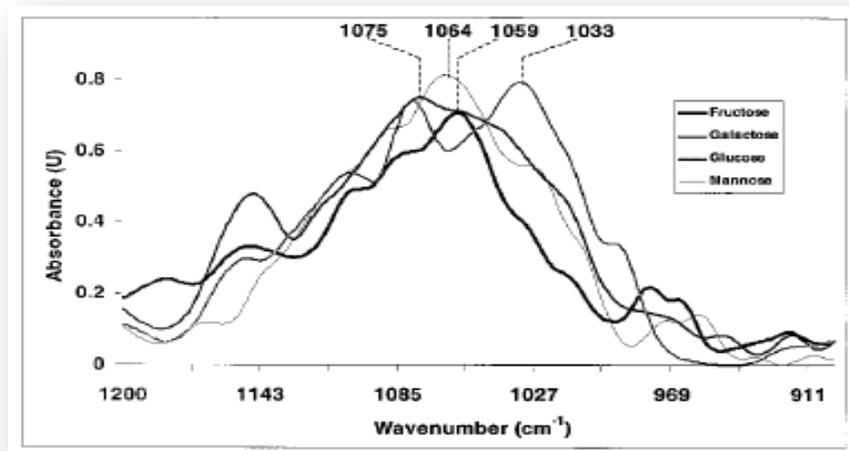
A adição de açúcares como reconstituintes da densidade ou do índice crioscópico do leite é uma prática que, associada à adição fraudulenta de água, visa recompor os parâmetros alterados do leite aos índices previstos na legislação brasileira (Brasil, 2018b). Os açúcares adicionados como sacarose ou amido, assim como os açúcares resultantes da hidrólise da lactose do leite, a glicose e a galactose, podem estar presentes no leite sem que o equipamento FTIR distinga a presença dos mesmos. Essa dificuldade na distinção destes açúcares presentes no leite se deve à similaridade espectral com a lactose (Hansen e Holroyd, 2019).

- **Espectros de absorbância do infravermelho pelos carboidratos**

Petibois *et al.* (1999) desenvolveram um método que permitiu a determinação da concentração de glicose em soro desidratado. Foram estudadas 32 amostras de soro analisadas por FTIR e comparadas ao método da glicose oxidase. Um pico dentro do espectro foi mais específico para glicose (997 a 1062  $\text{cm}^{-1}$ ). Sua integração de superfície mostrou forte relação com os dados de referência ( $r=0,998$ ;  $p < 0,001$ ). Análises FTIR de cinco soluções de glicose foram realizadas para

determinar sua absorção específica no mesmo pico. Desta forma, os espectros das concentrações de glicose séricas puderam ser medidos. Pela primeira vez, ao usar espectroscopia FTIR, nenhuma manipulação de espectros nem o uso de padrões internos foi necessário para obter resultados em alta concordância com a concentração de glicose medida por um método convencional (glicose-oxidase). Para discriminação espectral entre glicose e outras hexoses que poderiam estar presentes no sangue, foram realizadas as análises FTIR de soluções de 55 mmol L<sup>-1</sup> de D-frutose, D-galactose e D-manose (Figura 2.7). Segundo os autores, o pico localizado em 1033cm<sup>-1</sup> foi o mais específico para glicose.

Figura 2.7- Espectros do FTIR de D-frutose, D-galactose, D-glicose e D-manose na região do C-O (1200-900 cm<sup>-1</sup>)



Fonte: Petibois *et al.* (1999)

A técnica de espectroscopia no infravermelho, com acessório de refletância total atenuada (ATR) foi empregada por Morgano *et al.* (2003) para quantificar a concentração de sacarose em amostras de açúcar cristal comercial. A região do infravermelho médio, situada entre 1400 e 870 cm<sup>-1</sup>, foi utilizada para obtenção de um espectro e construção de um modelo de regressão pelo método PLS. Os resultados foram comparados com o método de referência da polarimetria. O modelo multivariado permitiu prever a concentração de sacarose com erros médios relativos de 0,34% e erro padrão de previsão de 0,021 g/100 mL.

Saliba *et al.* (2009) avaliaram a concentração de amido na raiz de nove genótipos de mandioca através do método convencional (enzimático), pela técnica espectroscópica no Infravermelho com Transformada de Fourier (FTIR) e análise particular por difração a laser (Partica) e estabeleceram padrões para a utilização dessas técnicas instrumentais na análise de alimentos. Procedeu-se a leitura da concentração linear na região espectral de 1644,284 a 859,88  $\text{cm}^{-1}$ . A comparação entre técnicas de determinação foi feita através de análise de covariância no delineamento de blocos ao acaso e utilizou-se o teste SNK a 5% de probabilidade de erro para comparação das médias. Os métodos instrumentais, por sua vez, apresentaram menores variações, maior rapidez, mínima manipulação e a vantagem de não necessitar do uso de reagentes.

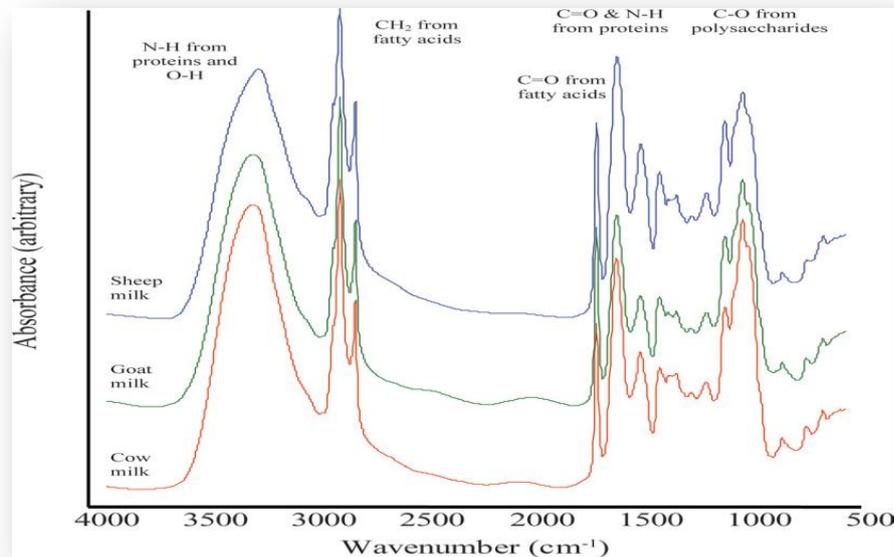
- **Espectros de absorvância do infravermelho pelos componentes do leite**

As regiões espectrais e os picos ou bandas geradas pela análise espectroscópica correspondem a possíveis compostos presentes nas amostras de leite e seus respectivos grupos funcionais, bem como as características das moléculas orgânicas. Normalmente, as ligações de ácidos graxos são identificadas em picos surgidos na faixa entre 3500 $\text{cm}^{-1}$  a 1747 $\text{cm}^{-1}$ , advindos de ligações CH<sub>2</sub> e C-O. Os grupos funcionais C=O, N-H e C-N originam os picos referentes às amidas I e II e III, respectivamente, entre as faixas de 1660  $\text{cm}^{-1}$  a 1241  $\text{cm}^{-1}$ . Os carboidratos, como a lactose, geram bandas envolvendo a presença de ligações C-O/C-C/C-O-O na faixa entre 1150  $\text{cm}^{-1}$  a 1000  $\text{cm}^{-1}$  de absorvância (Nicolau *et al.*, 2010).

Diferenças foram detectadas entre os leite de cabra e ovelha por Pappas *et al.* (2008), em que foi utilizado o FTIR para a obtenção de espectros com a finalidade de identificar e diferenciar os leites desnatados por meio da detecção de uma região espectral (impressão digital) de cada espécie. Dos onze picos gerados entre 1840-950  $\text{cm}^{-1}$  foram notadas pequenas diferenças entre as amostras de leite, sendo que a principal diferença foi localizada na banda 1745  $\text{cm}^{-1}$ , referente à vibração do carbonil do grupo éster, a qual foi mais intensa para o leite de cabra devido ao maior grau de esterificação que o leite de ovelha.

Nicolau *et al.* (2010) investigaram se a espectroscopia FTIR é uma técnica precisa e válida para a detecção e quantificação da adulteração de leite de cabra ou ovelha com leite de vaca em misturas binárias e terciárias. A faixa de comprimento de onda no IV médio de 4000 a 600  $\text{cm}^{-1}$  foi empregada e 64 varreduras foram coadicionadas e calculadas para melhorar a relação sinal-ruído; todos os espectros foram coletados com uma resolução de 4  $\text{cm}^{-1}$ . No total, 378 espectros foram coletados das diferentes combinações de leite. Os espectros representativos de FTIR coletados de leites puros de vacas, cabras e ovelhas são mostrados na Figura 2.8. Em combinação com métodos estatísticos multivariados, verificaram as diferenças existentes em suas composições com a sobreposição das bandas derivadas dos espectros dos leites, com destaque ao número de onda referente às gorduras, na faixa de absorção aproximadamente de 2927  $\text{cm}^{-1}$ , cuja intensidade se apresentou maior para o leite de ovelha. As outras diferenças visíveis entre os tipos de leite apareceram nas bandas de absorção relacionadas aos componentes restantes do leite, proteína (1654  $\text{cm}^{-1}$  e 1544  $\text{cm}^{-1}$  para amida I e II, respectivamente) e lactose (1159  $\text{cm}^{-1}$  e 1076  $\text{cm}^{-1}$ ). Estes pareciam estar presentes em maiores quantidades no leite de ovelha do que no leite de vaca e de cabra, com o primeiro exibindo maior absorção de IR para essas bandas. A integração das bandas derivadas dos espectros para cada tipo de leite confirma as diferenças acima na composição do leite de ovelha com maiores valores de absorção para as bandas de CH<sub>2</sub> e C-O.

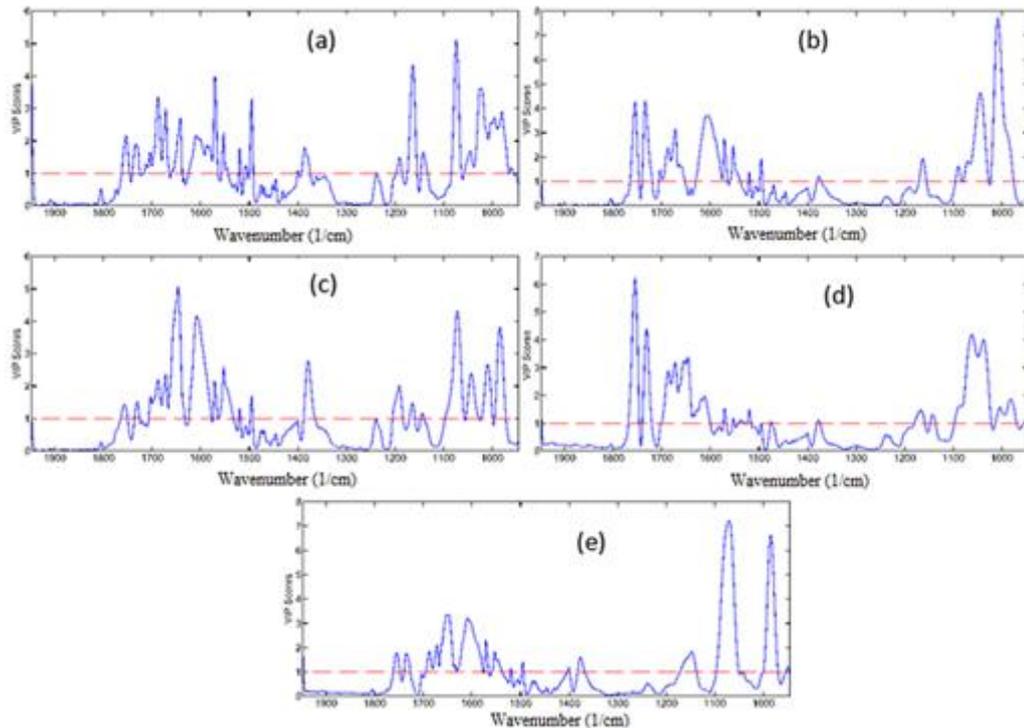
Figura 2.8- Espectros de infravermelho por transformada de Fourier para leite de vaca, cabra e ovelha



Fonte: Nicolau *et al.* (2010)

Botelho *et al.* (2015) desenvolveram um método qualitativo rápido e não destrutivo para a detecção simultânea de cinco adulterantes em leite cru (água, amido, citrato de sódio, formaldeído e sacarose), utilizando espectroscopia de infravermelho médio com reflexão total atenuada (ATR) e análise discriminante de mínimos quadrados parciais (PLS-DA). Os espectros foram coletados como uma média de 20 *scans*, de 4000 a 600  $\text{cm}^{-1}$ . A região entre 2000-900  $\text{cm}^{-1}$  apresentou o maior número de variações e foi considerado o mais discriminativo, contendo bandas características para quase todos os adulterantes avaliados. Como pode ser visto na Figura 2.9, cada adulterante apresentou picos em regiões específicas, como os dois picos entre 1080-1040 e 1015-980  $\text{cm}^{-1}$  na curva do citrato, a faixa larga com dois picos entre 1100-1000  $\text{cm}^{-1}$  na curva do formaldeído, e os dois picos maiores a 1080 e 980  $\text{cm}^{-1}$  na curva da sacarose.

Figura 2.9- Espectros de absorção de diferentes adulterantes em leite: (a) água, (b) amido, (c) citrato de sódio, (d) formaldeído, (e) sacarose



Fonte: Botelho *et al.* (2015)

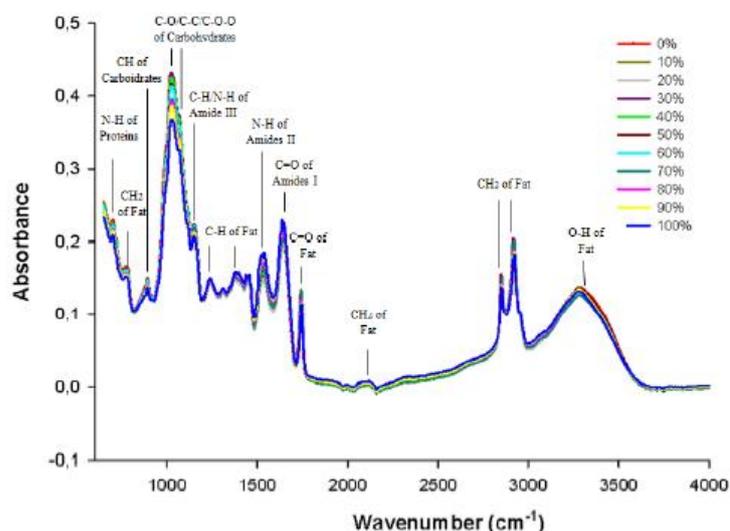
Conceição *et al.* (2019) identificaram adulteração com neutralizantes, conservantes e reconstituintes em amostras de leite por meio da análise de FTIR e quimiometria. E testaram esta técnica como ferramenta de triagem. Foram observadas as características dos espectros do leite e a sobreposição dos cinco níveis de concentração dos adulterantes. As principais frequências de absorção da ureia que estão relacionadas aos picos presentes no número de onda em 1680 e 1600  $\text{cm}^{-1}$ . A presença de picos entre 1200 e 1000  $\text{cm}^{-1}$  estão associados à presença de sacarose na amostra. O amido está associado a picos na região entre 1200-900  $\text{cm}^{-1}$ . Enquanto que a presença de picos na região 2500-2000  $\text{cm}^{-1}$  está associada à presença de peróxido de hidrogênio, picos na região 1250-1100  $\text{cm}^{-1}$  e na região 1200-1100  $\text{cm}^{-1}$  estão associados à presença de bicarbonato de sódio e hidróxido de sódio, respectivamente.

A detecção e identificação de adulterantes em leite cru foi realizada por Tavares (2019) com uso de Espectroscopia por FTIR associado à mineração de dados. Os espectros de infravermelho produzidos pelo FTIR foram 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 foram usados como entrada para o classificador da rede neural. As alterações nos espectros de absorção dos diferentes adulterantes (formol, bicarbonato de sódio, peróxido de hidrogênio, amido e sacarose) foram utilizados para o treinamento de diferentes algoritmos de aprendizado de máquina. De acordo com a arquitetura de rede neural proposta, obteve-se com mais de 99% de segurança que a amostra analisada estava ou não adulterada (método de triagem) e, em caso positivo, identificar qual o adulterante adicionado no modelo treinado.

Silva, L. *et al.* (2020) empregaram a espectroscopia no Infravermelho FTIR-ATR para realizar comparações entre os perfis espectroscópicos do leite de búfala e de vaca e amostras adulteradas com inclusão de diferentes percentuais de leite bovino em bubalino. Considerando que o leite de búfala apresenta maior teor dos seus principais constituintes em relação ao leite de vaca essas diferenças repercutiram na absorção dos seus respectivos grupos funcionais (Figura 2.10). Os picos referentes às vibrações das ligações das proteínas na região entre  $699\text{-}703\text{ cm}^{-1}$  e  $1240\text{-}1250\text{ cm}^{-1}$  também se apresentaram maiores para as amostras de leite bubalino, assim como os demais picos referentes aos carboidratos ( $1019\text{-}1133\text{ cm}^{-1}$  e  $890\text{-}893\text{ cm}^{-1}$ ) e às gorduras ( $768\text{-}780\text{ cm}^{-1}$ ), confirmando a relação entre o grau de absorção dos picos e o teor dos componentes de cada tipo de leite.

Figura 2.10- Espectros obtidos por FTIR da análise de leite de búfala com adição de leite de vaca em diferentes proporções



Fonte: Silva, L. *et al.* (2020)

## 2.7 Mineração de dados (*data mining*)

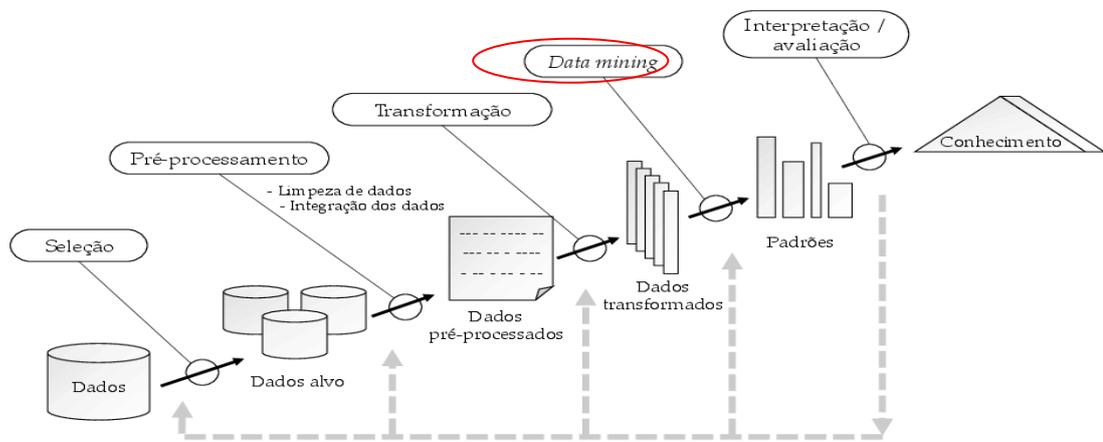
As análises de leite realizadas por instrumentos que utilizam um interferômetro para obter informações completas sobre o espectro dentro da região MIR, usando espectroscopia de infravermelho com Transformada de Fourier (FTIR), tornam possível obter uma extensa capacidade de computação e mineração de dados (Gambelli, 2017).

Bancos de dados geralmente contém informações desconhecidas que podem ser consideradas de grande importância. Com o objetivo de revelar tais informações, surge o processo chamado de Descoberta de Conhecimento em Banco de Dados (KDD - *Knowledge Discovery in Databases*), que se define como sendo um processo não trivial de extração de informações implícitas, previamente desconhecidas e potencialmente úteis, a partir dos dados armazenados em um banco de dados (Fayyad *et al.*, 1996).

KDD é uma análise exploratória, automática, que modela grandes repositórios de dados. É um processo que envolve a utilização de um banco de dados e de procedimentos que selecionam, pré-processam, realizam amostragem ou transformam os dados. *Data Mining* (Mineração de Dados) é o núcleo do processo KDD, envolvendo a inferência de algoritmos que exploram os dados, desenvolvem o modelo e descobrem padrões até então desconhecidos (Maimon e Rokach, 2010). As etapas principais do processo KDD são seleção, pré-processamento, transformação, *data mining* e interpretação e avaliação dos dados (Figura 2.11).

Mineração de dados (*data mining*) é o processo de encontrar anomalias, padrões e correlações em grandes conjuntos de dados para prever resultados. O termo "mineração" só foi usado nos anos 1990, mas sua base compreende três disciplinas científicas entrelaçadas e previamente existentes: estatística (o estudo numérico das relações entre dados), inteligência artificial (inteligência exibida por softwares e/ou máquinas, que se assemelha à humana) e *machine learning* (algoritmos que podem aprender com dados para realizar previsões) (Maimon e Rockach, 2010).

Figura 2.11- Etapas do processo KDD



Fonte: Fayyad *et al.* (1996)

## 2.8 Aprendizado de Máquina

Aprendizado de máquina (*machine learning*-ML) é o estudo computacional onde são desenvolvidos algoritmos baseados em métodos estatísticos que, a partir de conjuntos de dados de entrada, aprendem e reconhecem os padrões desses dados. Algoritmos de aprendizado de máquina vêm sendo amplamente utilizados devido à sua capacidade de contribuir na descoberta de conhecimentos de diversos temas. Técnicas de ML têm sido aplicadas em várias áreas, como reconhecimento de padrões, processamento de linguagem natural e aprendizagem computacional. Com estas técnicas, computadores são dotados com a capacidade de agir sem serem explicitamente programados, construindo algoritmos que podem aprender com os dados, e tomar decisões ou previsões baseadas neste aprendizado. Durante as últimas décadas, ML tem influenciado positivamente a vida cotidiana como, por exemplo, a pesquisa eficiente na Web, sistemas de autogestão, visão computacional, reconhecimento de voz e imagens, serviços financeiros, análise de sentimentos, diagnósticos médicos, entre outros (Kubat, 2017; Liu, W. *et al.*, 2017; Alpaydin, 2020).

Considerado uma subárea da inteligência artificial, o aprendizado de máquina é um sistema que aprende com dados, identifica padrões e prevê novos resultados com o mínimo de intervenção humana (Oliveira, 2018).

O aprendizado de máquina utiliza a estatística, principalmente para estimar problemas não-lineares. Os métodos estatísticos tradicionais, durante muito tempo, inferiram modelos lineares, com pequenas amostras de dados, até que os computadores se tornaram baratos e abundantes, a computação evoluiu e o estudo de métodos estatísticos mais complexos se tornou mais comum. Atualmente, é impraticável analisar a enorme quantidade de dados disponíveis somente de forma manual. Há um interesse crescente em modelos computacionais capazes de analisar e extrair informações dos dados. Estatística, inteligência artificial, mineração de dados, aprendizado de máquina, processamento de sinais e outras áreas se misturam em suas definições e abordagens (Alpaydin, 2020).

No aprendizado supervisionado, o objetivo é o aprendizado entre os dados de entrada e saída com base em valores corretos (rótulos) fornecidos por um supervisor, que são submetidos a um algoritmo que aprende regras e mapeia as instâncias dos rótulos fornecidos. O aprendizado supervisionado envolve a construção de um modelo estatístico com o objetivo de prever a resposta para observações futuras (predição) ou melhor entender o relacionamento entre a resposta e as entradas (inferência) (James *et al.*, 2017). São exemplos de métodos de aprendizado supervisionado *Support Vector Machines* (SVM), *K-Nearest Neighbor* (KNN), Árvores de Decisão, Redes Neurais Artificiais (RNA), entre outros.

O aprendizado não-supervisionado é um tipo de aprendizado que classifica os padrões similares sem utilizar pares (entrada-saída), ou seja, no treinamento da rede são usados apenas valores de entrada. A rede trabalha essas entradas como elementos de classificação e classifica a resposta mediante o critério da semelhança (Baptistella, 2005). O agrupamento, um modelo de classificação não supervisionada, visa ordenar diferentes objetos em categorias ou *clusters* com base na chamada medida de distância (Baker *et al.*, 2014).

Na etapa de treinamento do aprendizado supervisionado, os métodos tentam descobrir o relacionamento entre um ou mais atributos de entrada (variáveis independentes) e um atributo alvo (variável dependente). O relacionamento descoberto é representado em uma estrutura denominada *modelo*, que provém do treinamento e descreve e explica fenômenos que estão ocultos no conjunto de dados e que podem posteriormente ser usados para a predição de um novo valor (atributo alvo), a partir dos atributos de entrada (Maimon e Rockach, 2010).

As capacidades preditivas do modelo são avaliadas em um processo posterior, denominado teste, utilizando um conjunto de dados separado do treinamento.

### **2.8.1 Regressão e Classificação**

O aprendizado supervisionado pode ser dividido em modelos de regressão e modelos de classificação. Na regressão, os modelos mapeiam o espaço de entrada em um valor contínuo, caracterizando o conjunto de dados de forma quantitativa. A saída esperada é uma variável numérica (variável dependente) como uma função da entrada (variáveis independentes). A variável de saída é sempre uma variável contínua, de valor real, e, na prática, refere-se a quantidades, medidas e tamanhos a serem preditos pelos problemas (James *et al.*, 2017). Um problema de regressão requer a previsão de uma quantidade e as variáveis de entrada podem ser múltiplas, frequentemente chamado de problema de regressão múltipla (Brownlee, 2017).

Nos modelos de classificação, os modelos mapeiam o espaço de entrada em classes predefinidas, identificando a qual categoria um novo registro pertence. Este registro é determinado de forma qualitativa. A saída esperada é uma variável categórica (variável dependente) como uma função da entrada (variáveis independentes) (James *et al.*, 2017).

Modelos de classificação para determinação da presença de adulterantes em leite foram aplicados nos trabalhos de Asseiss Neto *et al.* (2019); Conceição *et al.* (2019);

Condé *et al.* (2020) e Silva, L. *et al.* (2020). Foram encontradas taxas de acerto de até 98% a depender do adulterante e do método de classificação utilizado.

Técnicas de classificação em mineração de dados são capazes de processar uma grande quantidade de dados e podem ser usados 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 atualmente disponíveis (Nikam, 2015).

De forma geral, é possível transformar os problemas de regressão em classificação, adequando a extensão da variável alvo contínua em conjuntos de intervalos que serão usados como classes discretas. Esse processo é chamado de discretização e a variável de saída resultante é uma classificação em que os rótulos têm um relacionamento ordenado (Brownlee, 2017). Em alguns casos, pode ocorrer o inverso, um problema de classificação pode ser convertido em um problema de regressão, quando um rótulo é convertido em um intervalo contínuo. Por exemplo: Classe 1 (valores de 0 a 49) e Classe 2 (valores de 50 a 100).

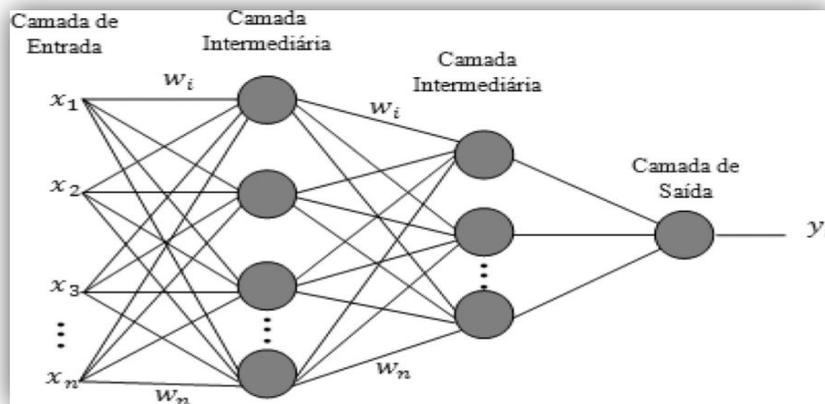
### **2.8.2 Redes Neurais Artificiais (RNA)**

Modelos de aprendizado supervisionado, as RNA são tipos de arquitetura computacional inspirada na biologia das redes neurais (sistema nervoso do cérebro) e são usados para aproximar funções que dependem de um grande número de entradas e são geralmente desconhecidas. Redes neurais artificiais 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).

As RNAs se assemelham ao cérebro humano por dois fatores: o conhecimento ( $\chi$ ) e os pesos ( $w$ ) (Figura 2.12). Sendo assim, o conhecimento é adquirido através da aprendizagem, e os pesos, que são a intensidade das conexões entre os neurônios,

são utilizados para armazenar estes conhecimentos até o neurônio de saída ( $y$ ) (Coutinho *et al.*, 2016). A interconexão entre os neurônios transmite o sinal, semelhante a uma sinapse biológica; e este sinal é representado por um número real e o processamento é a computação de uma função não linear da combinação de suas entradas (Alpaydin, 2020).

Figura 2.12- Arquitetura de uma Rede Neural Artificial



Fonte: Coutinho *et al.* (2016)

A estrutura de um neurônio, juntamente com suas entradas, pesos e saída é denominada *perceptron*. Em uma RNA, os perceptrons se organizam em camadas que, geralmente, se caracterizam em uma camada de entrada, uma camada de saída e uma ou mais camadas intermediárias ou ocultas. Este tipo de arquitetura é denominado *Multilayer Perceptron* (MLP). As redes multicamadas usam uma variedade de técnicas de aprendizado, sendo a mais popular a retropropagação (*backpropagation*), e os valores de saída são comparados com a resposta esperada para calcular o valor de alguma função de erro pré-definida. Por várias técnicas, o erro é então realimentado pela rede. Usando essas informações, o algoritmo ajusta os pesos de cada conexão para reduzir o valor da função de erro em um pequeno valor. Após repetir este processo por um número suficientemente grande de ciclos de treinamento, a rede geralmente converge para algum estado onde o erro dos cálculos é pequeno (Goyal e Goyal, 2012).

A aplicação das redes neurais na indústria de alimentos tem sua aplicabilidade estudada nas últimas décadas. A avaliação da qualidade dos produtos, da integridade, da autenticidade, da origem geográfica, do prazo de validade (*shelf-life*), todos estes parâmetros resultados da metodologia computacional das redes neurais traz um novo cenário às análises habituais, que são dispendiosas de tempo e custo (Goyal e Goyal, 2012). Impulsionada pelo lucro econômico, a substituição e a adulteração dos alimentos leva a potenciais riscos à saúde e à degradação de transações comerciais justas (Weng *et al.*, 2021).

As redes neurais têm sido amplamente utilizadas em quimiometria para substituir os métodos tradicionais de análise multivariada com base em modelos de regressão linear múltipla. A regressão pode ser considerada uma camada da rede neural, que é capaz de modelar problemas não lineares e mais complexos. Redes neurais já foram construídas com dados extraídos de análises espectroscópicas por FTIR e associadas à análise multivariada para identificação de adulteração em leite (Asseiss Neto *et al.*, 2019; Conceição *et al.*, 2019 e Silva, L. *et al.*, 2020).

Trabalhos utilizando as redes neurais para extrair informações de um grande número de dados, e fazê-lo de maneira mais rápida, mais eficiente e comparados aos métodos de estatística tradicionais, demonstraram que os resultados utilizando as redes neurais foram superiores na capacidade preditiva dos modelos propostos (Baptistella *et al.*, 2006; Goyal e Goyal, 2012; Coutinho *et al.*, 2016; Asseiss Neto *et al.*, 2019).

### **2.8.3 Redes Neurais Convolucionais**

As Redes Neurais Convolucionais (CNN) são arquiteturas de Redes Neurais que se tornaram uma ferramenta importante em várias áreas de pesquisa que abordam a visão computacional, o reconhecimento de padrões de imagens e vídeos, reconhecimento de fala, análises de textos, processamento de linguagem, entre outras. As CNN são inspiradas no processamento visual do cérebro humano e

possuem camadas adicionais em relação às redes neurais tradicionais (Liu, J. *et al.*, 2017).

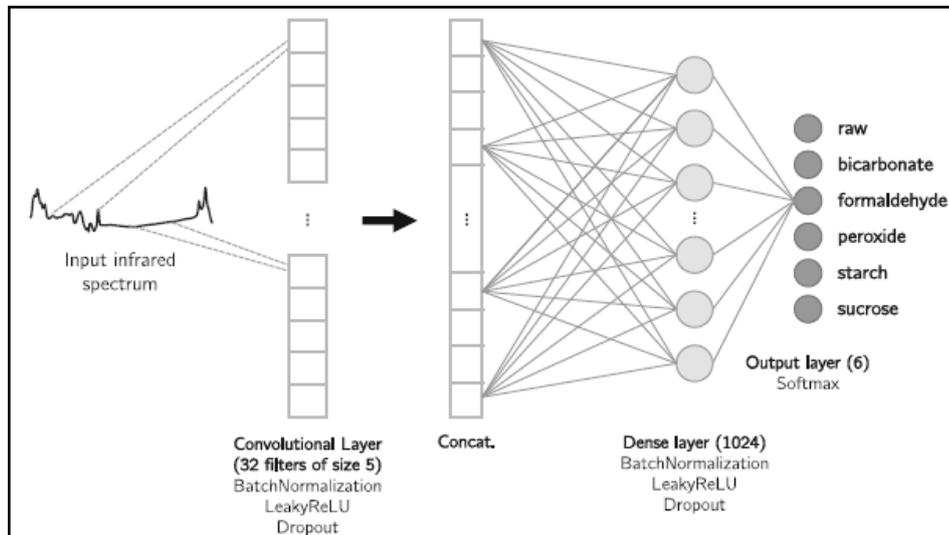
A definição da arquitetura de uma rede neural, por exemplo, requer um processo de decisões que denomina-se engenharia de arquitetura. Por meio deste processo, definições de hiperparâmetros, que são variáveis que controlam o próprio processo de treinamento, como o número de camadas ocultas utilizadas, quantos neurônios devem compor cada camada, quantos filtros convolucionais, entre outros, são tarefas que otimizam o desempenho do modelo. Esta engenharia permite que cientistas utilizem modelos de aprendizagem de máquina extremamente precisos, sem a necessidade de entender em detalhes a tecnologia por trás do seu funcionamento. A definição de uma arquitetura dos modelos geralmente é feita de forma manual, por meio de tentativa e erro, e usando o conhecimento de especialistas (Elsken *et al.*, 2019). Com a automação do processo, pela engenharia de arquitetura, o modelo torna-se mais adequado e com alto desempenho (Hutter *et al.*, 2019).

A arquitetura da CNN consiste em uma ou mais camadas convolucionais, que são um conjunto de filtros que aprendem, durante o treinamento, a detectar alguma característica importante dos dados de entrada; camadas de ativação não linear, como a *Rectified Linear Unit* (ReLU) que adicionam não linearidade ao modelo; camadas de agrupamento (*pooling*) que têm a função de reduzir a representação espacial dos dados e a complexidade da rede; e as camadas de normalização que ajustam a escala dos dados. Finalmente, a última camada convolucional passa para as camadas de *Multilayer Perceptron* e segue o funcionamento de uma rede neural tradicional (Witten *et al.*, 2016). As CNN são aplicadas a dados de imagem, em arranjos bidimensionais, que detectam bordas e outras regiões da imagem; no entanto, a mesma ideia é aplicada para dados unidimensionais, como dados espectrais (Liu, J. *et al.*, 2017).

Asseiss Neto *et al.* (2019), propuseram a construção de uma rede neural convolucional para classificação multiclasse, a partir do espectro de absorção do

FTIR, em estudo para identificar a presença de adulterantes em leite cru. Uma representação esquemática da CNN está na Figura 2.13.

Figura 2.13- Representação esquemática de uma rede neural convolucional para classificação de adulterantes em leite cru



Fonte: Asseiss Neto *et al.* (2019)

## 2.9 Separação de conjunto de dados

O desempenho preditivo dos métodos de aprendizado de máquina é avaliado utilizando uma separação de conjuntos de dados de forma padronizada: conjunto de treinamento, conjunto de validação e conjunto de teste. Se um modelo for avaliado com os mesmos dados do treinamento, pode-se ter uma estimativa incondizente com a realidade prática do modelo (Alpaydin, 2020).

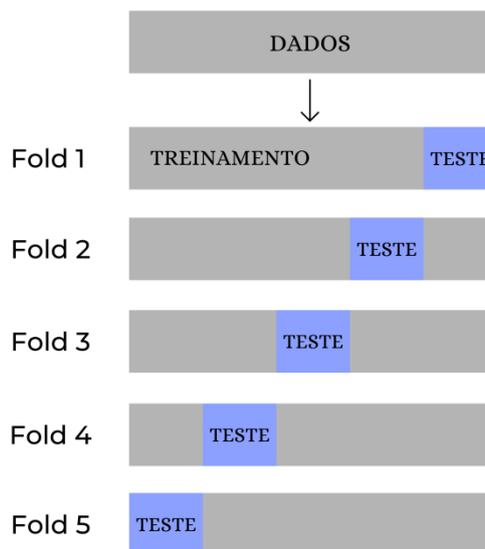
Uma abordagem trivial para se avaliar um modelo é dividir de maneira aleatória o conjunto de dados em três partes, numa proporção típica de 50% dos dados para treinamento, 25% para validação e 25% para teste. Esses valores variam com o autor, com o estudo e com os dados. As proporções dos conjuntos deve ser decidida pela quantidade de dados disponíveis (Asseiss Neto, 2020).

## 2.10 Validação Cruzada

As técnicas de validação cruzada (*cross-validation*) oferecem uma visão mais realista do poder preditivo dos modelos. Normalmente, um conjunto de validação específico não é abordado quando se trata de *cross-validation*, apenas conjuntos de treinamento e teste (Kuhn e Johnson, 2018).

O método de validação mais utilizado atualmente é o de validação cruzada *k-fold* (*k-fold cross-validation*), no qual o total de casos é dividido em  $k$  grupos aleatórios, de tamanhos aproximados. Um dos subconjuntos é usado para teste com os demais, sendo esse processo repetido com cada um dos  $k$  grupos. Obtêm-se, ao final, uma taxa de acerto/erro a partir das taxas de cada um dos subconjuntos (Faceli *et al.*, 2011; Dantas, 2017). Esta forma de validação cruzada, chamada de *k-fold cross-validation* (Figura 2.14), utiliza todas as instâncias tanto para treinamento quanto para teste, e cada instância é usada uma única vez. Os subconjuntos ( $k$ ) geralmente são do mesmo tamanho, e a escolha geralmente é de 5 ou 10  $k$ , mas não há regra formal (Kuhn e Johnson, 2018).

Figura 2.14- Diagrama de validação cruzada de 5 subconjuntos (*5-fold cross-validation*)



Fonte: autora (2021)

## 2.11 Métricas de desempenho

Há diversas métricas de desempenho para analisar a capacidade preditiva dos modelos de regressão e classificação. Para cada modelo, métricas podem e devem ser utilizadas. A acurácia, uma métrica de modelos de classificação, é a porcentagem de exemplos classificados corretamente dentre todas as previsões realizadas. Em um modelo de regressão não se usa a métrica de acurácia, como se prevê uma quantidade, a habilidade do modelo deve ser relatada como um erro nessas previsões (Brownlee, 2017).

Em um modelo de regressão, a variável resposta é uma variável numérica contínua e a métrica mais comum para caracterizar as capacidades preditivas do modelo é a raiz do erro quadrático médio (*root mean square error*, RMSE). O RMSE é uma função dos resíduos do modelo de regressão, que são as previsões do modelo subtraídas dos valores observados (Kuhn e Johnson, 2018). O RMSE é calculado pela raiz quadrada do Erro Quadrático Médio, representado na equação 1:

$$\text{RMSE} = \sqrt{\frac{1}{n} \sum_{i=1}^n (\hat{y}_i - y_i)^2} \quad (1)$$

onde  $n$  representa o número de amostras do conjunto,  $\hat{y}_i$  é o valor predito e  $y_i$  é o valor observado.

O Erro Quadrático Médio (*mean square error*, MSE) é a média do somatório dos quadrados dos erros de cada caso, tanto do conjunto de treinamento, quanto do conjunto de teste (Baptistella *et al.*, 2006). O MSE é calculado como a média do quadrado da diferença entre os valores preditos e os valores observados em um conjunto de dados, observado na equação 2:

$$\text{MSE} = \frac{1}{n} \sum_{i=1}^n (\hat{y}_i - y_i)^2 \quad (2)$$

onde  $n$  representa o número de amostras do conjunto,  $\hat{y}_i$  é o valor predito e  $y_i$  é o valor observado.

A média dos desvios, descrita pelo Erro Médio Absoluto (*mean absolute error*, MAE) possui a unidade resposta igual à unidade dos valores de entrada. Seu valor representa o desvio médio entre os valores preditos e reais (observados). Comparando o MSE e o MAE, o primeiro dá um peso maior para grandes desvios (pois são elevados ao quadrado), enquanto o MAE (Equação 3) dá um peso igual a todos os desvios (Willmott, 1981).

$$\text{MAE} = \frac{1}{n} \sum_{i=1}^n |\hat{y}_i - y_i| \quad (3)$$

onde  $n$  representa o número de amostras do conjunto,  $\hat{y}_i$  é o valor predito e  $y_i$  é o valor observado.

Outra métrica muito utilizada em modelos de regressão é o coeficiente de determinação, apresentado como  $R^2$ . Esse valor expressa a proporção de informação nos dados que é explicada pelo modelo. O valor de  $R^2$  varia de 0 a 1 e geralmente é representado em porcentagem. Por exemplo, um  $R^2$  igual a 0,85, significa que o modelo explica 85% da variação nos resultados (Kuhn e Johnson, 2018).

No modelo de classificação, a acurácia é a métrica que representa a precisão correta entre as classes observadas e preditas. Apesar de ser amplamente utilizada, é uma métrica que deve ser usada com cautela, pois não considera o tipo de erro que o modelo possa estar cometendo. Acurácia é a proporção de casos que foram corretamente previstos, sejam eles verdadeiro positivo ou verdadeiro negativo. É considerada a taxa de acerto (Kuhn e Johnson, 2018).

Para visualizar resultados verdadeiro positivo (VP), falso positivo (FP), verdadeiro negativo (VN) e falso negativo (FN), uma matriz de confusão é a melhor maneira (Figura 2.15). Os verdadeiros positivos são os previstos corretamente dentro dos realmente positivos. Os falsos positivos são erroneamente previstos como positivos. Os verdadeiros negativos são previstos corretamente como negativos e os falsos negativos são previstos erroneamente como negativos.

Figura 2.15- Representação de uma matriz de confusão num arranjo 2 x 2, para um problema genérico de classificação binária

		Previsto	
		positivo	negativo
Real	positivo	VP	FN
	negativo	FP	VN

Fonte: autora (2021)

A proporção de casos positivos que foram previstos corretamente (VP) entre todos os casos positivos é a sensibilidade, e a proporção de casos negativos que foram previstos corretamente (VN) entre todos os casos negativos é a especificidade, dadas pelas fórmulas:

$$\text{Sensibilidade} = \text{VP}/(\text{VP} + \text{FN})$$

$$\text{Especificidade} = \text{VN}/(\text{VN} + \text{FP})$$

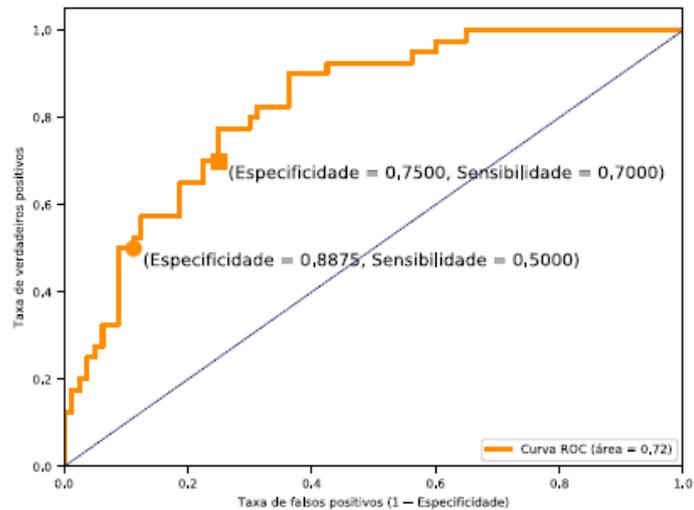
Asseiss Neto *et al.* (2019) realizaram classificações binária e multiclasse em estudos para identificar adulterantes em leite, considerando os mesmos conjuntos de dados. Foram gerados valores de até 99,34% de acurácia para a classificação binária (presença ou não de adulterantes), e de até 98,49% de acurácia para a classificação multiclasse (sacarose, amido, formol, bicarbonato e peróxido).

A Curva ROC (*Receiver Operator Characteristic*) e a AUC (*Area Under the ROC Curve*) são métricas que avaliam a relação entre a sensibilidade e a especificidade. A construção da curva ROC é a partir da relação da taxa de verdadeiros positivos (sensibilidade) e da taxa de falsos positivos (1 - especificidade). A área sob a curva ROC, ou AUC, é uma medida que sumariza o desempenho da ROC para um valor único que indica o quanto um modelo é capaz de distinguir entre as classes avaliadas. Na Figura 2.16, a área sob a curva é de 0,72. Quanto mais próximo de 1, mais eficiente é o modelo em distinguir as classes com 100% de sensibilidade e especificidade (Asseiss Neto, 2020).

Numa comparação de modelos, quanto mais alta e mais próxima do canto superior esquerdo a curva ROC estiver, melhor será o modelo. E pode-se considerar a maior AUC dos modelos para determinar aquele mais eficiente (Kuhn e Johnson, 2018).

As curvas ROC são definidas para problemas de classificação binária, mas existem extensões da técnica que consideram o problema da classificação multiclasse (Kuhn e Johnson, 2018).

Figura 2.16- Representação de curva ROC (*receiver operator characteristic*) para um modelo genérico de classificação que demonstra a relação entre a taxa de verdadeiros positivos (sensibilidade) e a taxa de falsos positivos (1 – especificidade)



Legenda: dois pontos são destacados e mostram a especificidade e a sensibilidade em dois limiares distintos. A área sob a curva (AUC) tem o valor de 0,72.

Fonte: Asseiss Neto (2020)



## Metodologia Aplicada

Este capítulo apresenta as técnicas aplicadas nas análises laboratoriais das amostras de leite, como FTIR, citometria de fluxo e HPLC e, posteriormente, o tratamento dos dados computacionais gerados pelos equipamentos eletrônicos. A associação de ferramentas de diferentes áreas, biológica e informática, faz da Bioinformática uma área promissora e de grande aplicabilidade na rotina laboratorial de monitoramento da qualidade do leite.

### 3.1 Projeto Piloto

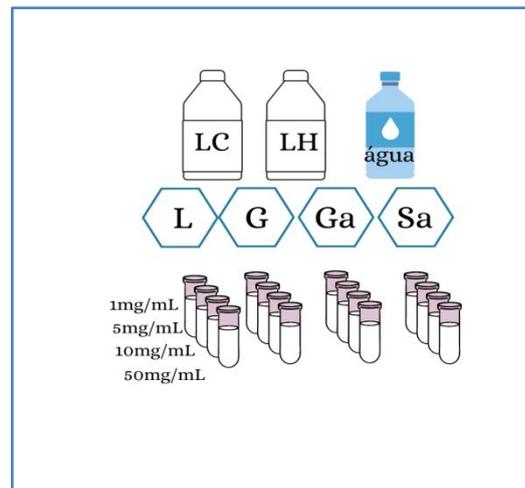
Em agosto de 2018, foram realizadas análises experimentais em um projeto piloto, para exploração de dados e direcionamento dos estudos.

O leite cru foi coletado no tanque de resfriamento da Fazenda Experimental Hélio Barbosa, da Escola de Veterinária da UFMG, localizada no município de Igarapé - MG. O leite foi armazenado em galão plástico e imediatamente transportado até o LabUFMG e mantido sob refrigeração até o preparo das amostras.

A hidrólise do leite foi realizada com enzima  $\beta$ -D-galactosidase comercial de *Kluyveromyces marxianus var. lactis* (Rica Nata®), 0,2mL/L de leite, por 24 horas, sob refrigeração, conforme recomendação do fabricante.

Foram utilizadas 33 amostras de leite cru (LC), 20 amostras de leite cru hidrolisado (LH), totalizando 53 amostras adicionadas de quatro açúcares P.A., lactose (L), glicose (G), galactose (Ga) e sacarose (Sa) nas concentrações de 1mg/mL; 5mg/mL; 10mg/mL e 50mg/mL, adicionadas em alíquotas de 40mL de leite (Figura 3.1).

Figura 3.1- Esquema de preparação das amostras de leite do Projeto Piloto



Fonte: autora (2021)

As amostras foram analisadas para composição e crioscopia pelo método de espectrofotometria no infravermelho por FTIR e Contagem de Células Somáticas (CCS) pela citometria de fluxo, no equipamento CombiScope™ FTIR 400 (Delta Instruments, Drachten, Holanda) (Figura 3.2).

Figura 3.2- Equipamento CombiScope™ FTIR 400, do Laboratório de Análise da Qualidade do Leite da Escola de Veterinária da Universidade Federal de Minas Gerais (LabUFMG)



Fonte: arquivo pessoal (2018)

Também foram realizadas análises em 40 amostras de água deionizada adicionadas de quatro açúcares P.A. (lactose, glicose, galactose e sacarose) nas concentrações de 1mg/mL; 5mg/mL; 10mg/mL; 30mg/mL e 50mg/mL, para exploração dos espectros de absorção pelos carboidratos, em alíquotas de 40mL de água (Figura 3.1).

## **3.2 Análises do leite pelo método FTIR**

### **Descrição**

Nesta etapa do trabalho foram realizadas análises por FTIR para criar um banco de dados para servir de entrada no aprendizado de máquina e construção das redes neurais convolucionais, por meio dos espectros de absorção das amostras de leite.

### **Local de realização do experimento**

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

### **Preparo das amostras para análise por FTIR**

As amostras de leite cru foram obtidas na Fazenda Experimental da Escola de Veterinária da UFMG, localizada no município de Igarapé – MG. O leite, proveniente do tanque de expansão da ordenha completa do rebanho, foi coletado diretamente em galão plástico, após agitação e imediatamente transportado até o LabUFMG e mantido sob refrigeração até o dia seguinte para o procedimento das análises. O leite pasteurizado e o leite UHT foram obtidos no mercado local, de diferentes marcas e lotes. Foram seis repetições para cada tipo de leite.

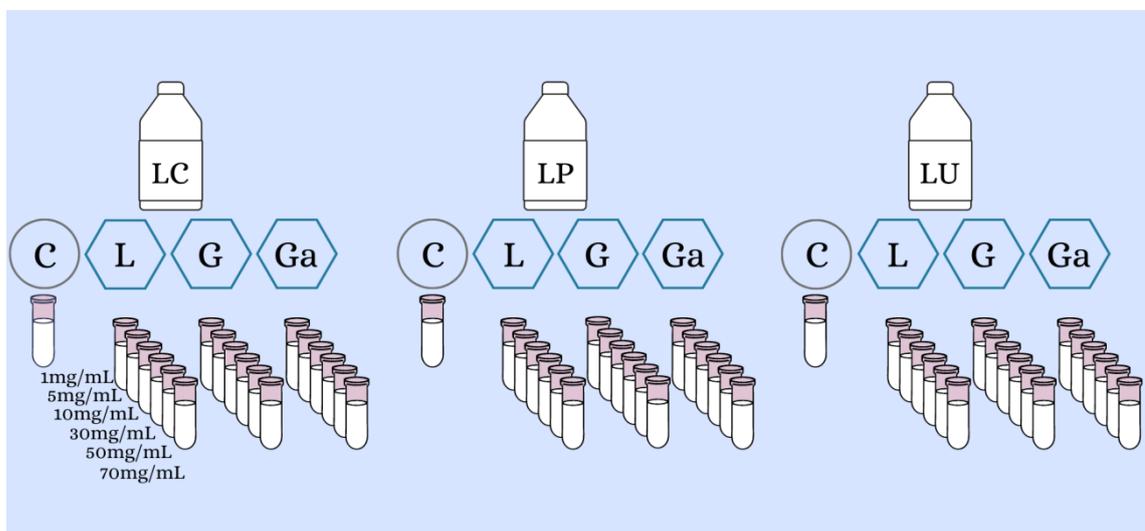
Para análise da qualidade das amostras foram realizadas análises de Contagem Padrão em Placas (CPP), CCS e composição antes de submetê-las ao esquema de preparo das amostras.

Alíquotas de 40mL de amostras de leite cru (LC), pasteurizado (LP) e UHT (LU) foram adicionadas com lactose (L), glicose (G) e galactose (Ga) (Sigma-Aldrich, USA), nas concentrações de 1mg/mL; 5mg/mL; 10mg/mL; 30mg/mL; 50mg/mL e 70mg/mL (Figura 3.3). O leite adicionado com açúcares foi armazenado em frascos contendo bronopol<sup>®</sup> (Advanced Instruments) para a conservação das amostras

analisadas para composição e CCS. Para análise de CPP as amostras foram armazenadas em frascos contendo azidiol<sup>1</sup> para conservação das mesmas.

O delineamento experimental incluiu 342 amostras ao todo, sendo 324 amostras adicionadas de açúcares (6 repetições x 3 açúcares x 6 concentrações x 3 tipos de leite) e 18 amostras sem adição para controle (C) (6 repetições x 3 tipos de leite).

Figura 3.3- Esquema de preparação das amostras de leite analisadas por FTIR



Fonte: autora (2021)

Foram realizadas análises de composição centesimal (gordura, proteína, lactose, sólidos totais, sólidos não gordurosos, caseína), nitrogênio ureico do leite, crioscopia, contagem de células somáticas e contagem padrão em placas, em equipamentos eletrônicos.

### Contagem Padrão em Placas

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, que permite análises precisas e instantâneas das bactérias no leite. A citometria de fluxo é uma técnica que conta e caracteriza partículas e células. Uma suspensão de células é corada e flui através de um tubo capilar com a separação dos aglomerados de bactérias. Durante o período de incubação, as bactérias são

<sup>1</sup> Composição do comprimido de azidiol: cloranfenicol e azida sódica.

coradas com um meio de coloração específico para o DNA (neste trabalho, brometo de etídio). 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, 2018).

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

### **Contagem de Células Somáticas**

A contagem de células somáticas (CCS) foi realizada no citômetro do equipamento CombiScope™ FTIR 400, por meio do SomaScope (Delta Instruments, Drachten, Holanda). Uma alíquota da amostra, pré-aquecida a 40° C, é aspirada para o interior do equipamento e conduzida a uma seringa contendo o corante tampão. O instrumento requer o uso de um corante fluorescente para corar o DNA das células. Em seguida, 50 µL da amostra são conduzidos por um fluido carreador para o citômetro de fluxo, onde recebem incidência de raio laser. A luz emitida passa por uma série de filtros ópticos e lentes focalizadas em comprimentos de onda adequados e é captada como pulso elétrico. Este pulso é ampliado, filtrado e convertido em contagem de células somáticas. O histograma resultante representa o número de células no eixo vertical e o tamanho das células no eixo horizontal (Delta, 2007).

A calibração de CCS foi feita utilizando-se três amostras padrão com CCS em diferentes níveis. As amostras padrão foram fornecidas pelo laboratório Valacta (Dairy Production Centre of Expertise, Quebec, Canadá).

### **Composição centesimal**

A análise composicional e crioscopia foram realizadas no equipamento CombiScope™ FTIR 400, por meio do LactoScope FTIR (Delta Instruments, Drachten, Holanda). A amostra de leite é pré-aquecida a 40°C. Após passar por um

homogeneizador, para reduzir o diâmetro dos glóbulos de gordura, a amostra recebe irradiação pelo feixe de luz infravermelha em uma cubeta. A diferença de energia absorvida entre a amostra a ser analisada e a amostra de referência é captada por um detector de infravermelho e em seguida, quantificada e transformada automaticamente em teores de componentes, tendo como referência a curva de calibração (Delta, 2007).

A calibração do instrumento refere-se ao ajuste do sinal instrumental de modo que, a cada nível do componente em quantificação, a média de resultados individuais dados pelo instrumento seja o mais próximo do valor verdadeiro da concentração do componente. 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 EST. Para ureia foram utilizadas 6 amostras e para caseína 8 amostras. A calibração foi feita utilizando-se o programa de calibração do próprio equipamento. As amostras padrão foram fornecidas pelo laboratório Valacta (Dairy Production Centre of Expertise, Quebec, Canadá).

A calibração para leitura da crioscopia foi feita com amostras padrão fornecidas pelo laboratório Eurofins/DQCI Services (Dairy Quality Control Incorporated, Estados Unidos), credenciado pela Secretaria de Agricultura do Estado de Mineápolis dos Estados Unidos.

### **3.3 Aprendizado de Máquina**

#### **Descrição**

Esta etapa do trabalho consistiu na execução de técnicas da inteligência artificial, desde o processo de mineração de dados (*data mining*) até a construção das redes neurais convolucionais, a partir dos resultados das análises das amostras de leite por FTIR.

### **Aquisição de dados**

Para cada amostra analisada, o equipamento FTIR utilizado gerou dois arquivos com dados distintos: um arquivo de espectro infravermelho, no formato SPC, que contém as coordenadas do espectro de absorção do infravermelho, e um arquivo de composição, no formato CSV, que contém variáveis numéricas e são *features* (atributos) de composição do leite. Estes atributos são calculados pelo equipamento a partir do espectro, que utiliza PLS (*Partial Least Squares*) como método de calibração para determinar os valores da composição centesimal (gordura, proteína, lactose, sólidos totais, sólidos não gordurosos, nitrogênio ureico do leite e caseína) (Delta, 2007).

Para este trabalho, utilizou-se as informações provenientes das coordenadas espectrais, que foram analisadas e condensadas em um único conjunto de dados no formato CSV. Os dados das coordenadas espectrais originais das amostras de leite sem pré-processamento serviram de entrada para a construção de uma rede neural convolucional (Tavares, 2019).

Os conjuntos de dados referentes à composição centesimal do leite não foram utilizados como entrada para treinamento do aprendizado de máquina. Apesar de mostrarem grande potencial de exploração e execução de redes neurais (Goyal e Goyal, 2012; Valente *et al.*, 2014; Asseiss Neto *et al.*, 2019 e Condé *et al.*, 2020), *features* de componentes apresentam uma grande variância e escalas diferentes que podem impactar na execução de modelos estatísticos e de aprendizado de máquina (Asseiss Neto, 2020). Essa variância na natureza dos dados da composição do leite faz necessário um pré-tratamento dos dados e análises estatísticas para escolha dos atributos de maior importância para o modelo.

Uma rede neural convolucional (CNN) tem capacidade de exploração sem necessidade de um pré-processamento dos dados, contrário do que ocorre em modelos de aprendizado como árvores de decisão, por exemplo. As CNN são capazes de aprender, no treinamento da rede, um número definido de filtros que destacam diferentes características. Os neurônios das camadas convolucionais respondem a estímulos em diferentes regiões espectrais fazendo com que os filtros

passem para as camadas profundas da rede somente os dados do espectro que a rede considera importante. Por isso, a CNN é capaz de receber os dados espectrais brutos sem o pré-processamento, realizando a extração de *features* sem quaisquer interações manuais (Schmidhuber, 2015).

As regiões mais significativas e representativas do espectro que servem de entrada para as camadas convolucionais da CNN, que aprenderam com os filtros, são denominadas mapas de saliências. Pelo mapa de saliências é possível visualizar quais regiões da imagem foram mais significativas para a rede (Simonyan *et al.*, 2014). A obtenção das saliências nos espectros do FTIR das amostras de leite possibilitam a extração das regiões espectrais que mais contribuíram para ativação de uma determinada saída na CNN.

### **Engenharia de Arquitetura**

A arquitetura da Rede Neural Convolucional executada possui uma camada convolucional unidimensional, 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 do FTIR. Esta camada também pode ser chamada de camada detectora de *features*. Os filtros são concatenados (agrupados) e seguidos por uma camada mais densa (totalmente conectada) composta de 2048 neurônios. São utilizadas camadas auxiliares que realizam a ativação dos neurônios como a função LeakyReLU (Maas *et al.*, 2013) que adicionam não linearidade ao modelo, função *MaxPooling1D* que reduz os dados fazendo agrupamentos, operações de *Batch Normalization* (Ioffe e Szegedy, 2015) que normalizam os dados das camadas e *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.

### **Execução da Rede Neural Convolutacional com espectros do FTIR**

As arquiteturas das CNN foram implementadas utilizando Keras (Chollet F., 2018) e TensorFlow (Abadi *et al.*, 2016) em Python. Todos os modelos utilizaram, em suas implantações, padrões da biblioteca Scikit-learn (Pedregosa *et al.*, 2011). Metodologia descrita por Asseiss Neto (2020) com modificações no número de filtros, tamanho de *kernel* convolutacional, número de neurônios na camada densa e número de neurônios na camada de saída.

As coordenadas dos espectros das amostras de leite analisadas pelo FTIR serviram de entrada para a construção das CNN propostas. Estas coordenadas foram compostas de 550 pontos e representam a região de absorção do espectro de 3100 a 900  $\text{cm}^{-1}$ , com resolução de 4  $\text{cm}^{-1}$ . A CNN interpreta o espectro da amostra de leite como uma “imagem”, assim como as camadas convolucionais de detecção de filtros que selecionam características importantes do espectro (Liu, J. *et al.*, 2017).

### **Algoritmo de Classificação**

Para o problema de classificação, uma CNN foi treinada com o conjunto de dados das amostras de leite sem adição e adicionadas de açúcares (lactose, glicose e galactose), descritos na seção 3.2. Uma vez que se conhece o açúcar adicionado e os valores dos resultados das análises de cada amostra, o modelo se torna capaz de reconhecer o espectro de absorção de cada sacarídeo. Os resultados dos modelos de classificação são variáveis respostas de natureza categórica, se a presença ou não de açúcar no leite foi identificada pelo modelo.

Para o modelo de classificação, a métrica de desempenho mais comum é a acurácia, que representa a precisão correta entre as classes observadas e preditas. É considerada como o percentual de acerto (Kuhn e Johnson, 2018).

### **Algoritmo de Regressão**

Alguns algoritmos têm a palavra “regressão” em seu nome, mas a distinção se faz importante entre os modelos, porque a regressão linear é um algoritmo (modelo) de regressão, enquanto a regressão logística é um algoritmo (modelo) de classificação (James *et al.*, 2017).

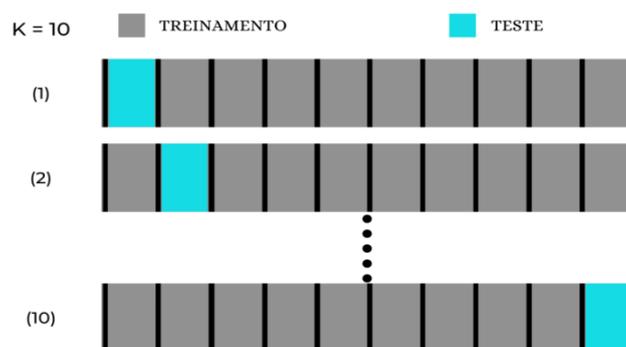
As CNN treinadas para modelos de regressão receberam como dados de entrada os resultados dos espectros das amostras de leite sem adição e adicionadas com os açúcares (lactose, glicose e galactose), em seis diferentes concentrações (1mg/mL; 5mg/mL; 10mg/mL; 30mg/mL; 50mg/mL e 70mg/mL). Os açúcares adicionados ao leite causam picos de absorvância que estão correlacionados à quantidade do componente no leite (Luykx e Van Ruth, 2008).

Os resultados dos modelos de regressão são variáveis respostas de natureza contínua, numérica, que correspondem à quantidade presente de cada açúcar no leite que foi identificada pelo modelo. A métrica mais comum para caracterizar as capacidades preditivas do modelo é a raiz do erro quadrático médio (*root mean square error*, RMSE). O RMSE é uma função dos resíduos do modelo de regressão, que são as diferenças entre os valores reais do conjunto de dados e os valores preditos, calculados com a equação de regressão (Kuhn e Johnson, 2018).

### **Cross- validation**

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 *k-fold 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. Neste trabalho, o conjunto de dados foi dividido em 10 partes (*10-fold*) ou subconjuntos (Figura 3.4).

Figura 3.4- Diagrama de validação cruzada de dez subconjuntos (*10-fold cross-validation*)



Fonte: autora (2021)

### Região do espectro na execução da CNN

A região do espectro de absorção do infravermelho pelo equipamento FTIR compreende a região do MIR (*mid-infrared*), e neste trabalho a região compreendida entre 902 e 3097  $\text{cm}^{-1}$  foi analisada e condensada em um arquivo CSV para execução da rede neural convolucional. Uma outra CNN compreendendo a região de 875 a 1481 $\text{cm}^{-1}$  também foi testada, a fim de explorar os dados na região de absorção dos açúcares (Tabela 3.1).

Tabela 3.1- Regiões do espectro de absorção do infravermelho por alguns carboidratos

Carboidrato	Região do espectro ( $\text{cm}^{-1}$ )	Referência
<b>Glicose (soro sanguíneo)</b>	997 a 1062	Petibois <i>et al.</i> , 1999
<b>Sacarose (açúcar cristal)</b>	870 a 1400	Morgano <i>et al.</i> , 2003
<b>Amido (mandioca)</b>	859 a 1644	Saliba <i>et al.</i> , 2009
<b>Lactose</b>	1076 a 1159	Nicolau <i>et al.</i> , 2010
<b>Sacarose (leite)</b>	980 a 1080	Botelho <i>et al.</i> , 2015
<b>Amido (leite)</b>	900 a 1200	Conceição <i>et al.</i> , 2019
<b>Sacarose (leite)</b>	1000 a 1200	Conceição <i>et al.</i> , 2019
<b>Lactose</b>	890 a 1133	Silva, L. <i>et al.</i> , 2020

Fonte: autora (2021)

### 3.4 Otimização das condições de hidrólise enzimática da lactose

#### Descrição

Nesta etapa do trabalho realizou-se um estudo de otimização da hidrólise enzimática da lactose para definir um protocolo de hidrólise em amostras de leite, resultando em diferentes concentrações finais de lactose, glicose e galactose, para serem analisadas por HPLC e por FTIR.

#### Local de realização do experimento

O experimento foi realizado no Laboratório de Análise da Qualidade do Leite da Escola de Veterinária da Universidade Federal de Minas Gerais (LabUFMG).

### Otimização das condições

Na etapa de otimização das condições de hidrólise enzimática da lactose, amostras de leite pasteurizado foram testadas em diferentes temperaturas (12°C, 30°C e 40°C), tempos (0, 30, 60, 90, 120, 150 e 180 minutos) e concentrações de enzima (10, 20, 40, 60, 80 e 100µL/100mL). A hidrólise enzimática foi realizada com enzima β-galactosidase de *Kluyveromyces lactis*, ≥ 2600 units/g (Sigma-Aldrich, USA, CAS 9031-11-2).

Para análise da qualidade das amostras foram realizadas análises de CPP, CCS e composição antes de submetê-las ao esquema de preparo das amostras. Também foi mensurado o índice crioscópico (IC) das amostras em Crioscópio Eletrônico LK-7000 (Laktron, PR, Brasil).

### Condições de hidrólise em diferentes temperaturas e tempos

Alíquotas de 100 mL de leite pasteurizado foram adicionadas de 40 µL de enzima β-galactosidase e mantidas às temperaturas de 12°C, 30°C e 40°C, em banho-maria. O grau de hidrólise foi acompanhado por meio da crioscopia inicial e a crioscopia medida de 30 em 30 minutos. O equipamento utilizado apresentou um limite de detecção do IC em -0,653°H. Ao alcançar esse IC, as medições cessavam e as interpretações eram feitas com os valores obtidos até este ponto.

O grau de hidrólise da lactose foi determinado aplicando-se a equação, segundo Longo (2006):

$$\% \text{ hidrólise alcançada} = 350,877 \times (\text{IC final}) - \frac{(\text{IC inicial})}{0,00285}$$

### Condições de hidrólise em diferentes concentrações da enzima

Alíquotas de 100 mL de leite pasteurizado foram adicionadas com 20, 40, 60, 80 e 100 µL de β-galactosidase, a 7°C. O grau de hidrólise foi acompanhado por meio da crioscopia inicial e a crioscopia medida em intervalos de 30 minutos.

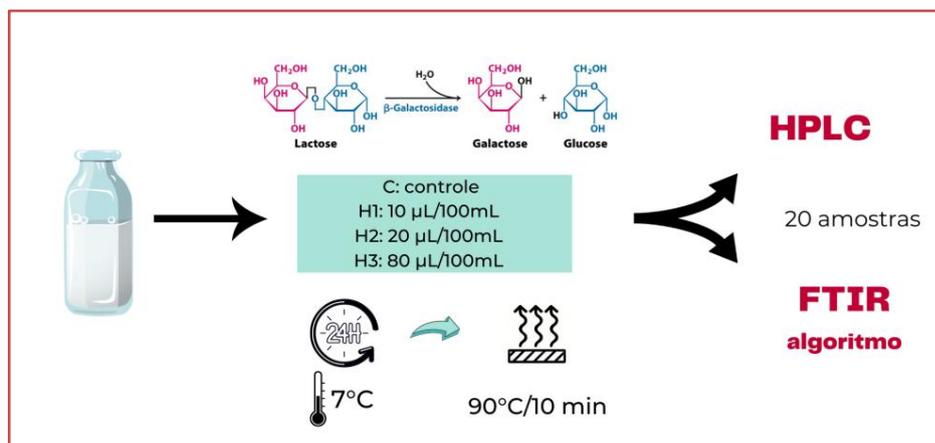
Outro teste realizado com quatro diferentes concentrações de enzima (10, 20, 40 e 60  $\mu\text{L}$ ) em 100mL de leite pasteurizado, conservadas a 7°C por 24 horas, demonstraram que as duas primeiras concentrações não atingiram o limite de IC do equipamento ao final do período testado.

### Preparo das amostras para hidrólise enzimática

A otimização das condições de hidrólise foram fundamentais para determinação dos quatro tratamentos nas amostras de leite pasteurizado. Uma alíquota de 100 mL foi destinada à amostra controle, sem adição de enzima. Para os três protocolos de hidrólise, alíquotas de 100 mL foram adicionadas com a enzima  $\beta$ -galactosidase, em três diferentes concentrações (10  $\mu\text{L}/100\text{mL}$ , 20  $\mu\text{L}/100\text{mL}$ , 80  $\mu\text{L}/100\text{mL}$ ). Após a adição da enzima, todas as amostras, inclusive as amostras controle, foram conservadas sob refrigeração, a 7°C, e mantidas por 24 horas até ser realizado o aquecimento, 90°C por 10 minutos, em banho-maria, com a finalidade de parar a hidrólise por inativação da enzima pela desnaturação (Churakova *et al.*, 2019). Após o aquecimento, as amostras foram rapidamente resfriadas e submetidas às análises por HPLC e FTIR.

No total foram 20 amostras de leite hidrolisado analisadas por HPLC e por FTIR, para estudo comparativo, conforme Figura 3.5 O leite pasteurizado foi obtido no mercado varejista local, de diferentes marcas e lotes. Foram realizadas cinco repetições.

Figura 3.5- Esquema de preparação das amostras de leite para hidrólise enzimática da lactose



Fonte: autora (2021)

### 3.5 Quantificação de açúcares pelo método HPLC

#### Descrição

Nesta etapa do trabalho foram realizadas as análises das amostras de leite com hidrólise enzimática da lactose, por HPLC-RID - considerado o método analítico de referência para a quantificação de açúcares no leite.

#### Local de realização do experimento

As análises por HPLC para determinação analítica de lactose, glicose e galactose foram realizadas na FUNED (Fundação Ezequiel Dias), no Laboratório de Análises de Microcomponentes.

#### Metodologia para quantificação de lactose, glicose e galactose

A metodologia utilizada foi descrita no método oficial 980.13 da AOAC (*Association of Official Analytical Chemists*), com modificações, e foi validada no laboratório em conformidade com os Procedimentos Operacionais Padronizados da FUNED (FUNED, 2018). As principais alterações do método foram a extração da amostra sem reagentes químicos e a utilização de água ultrapurificada como fase móvel.

#### Condições Cromatográficas

A análise foi realizada em um cromatógrafo líquido de alta eficiência (LC-20A Prominence – Shimadzu, Japão), com detecção por índice de refração (RID-10A - Shimadzu, Japão), equipado com o software LC-Solutions e amostrador automático (Figura 3.6). A separação cromatográfica foi obtida com uma coluna Shim-pack SCR101P 300 mm x 7,9 mm, 10 µm (Shimadzu Co., Kyoto, Japão) e pré-coluna Shim-pack SCR101P 50 mm x 4 mm, 10 µm (Shimadzu Co., Kyoto, Japão), temperatura do forno da coluna a 80°C e temperatura da célula do detector a 40°C. Água ultrapurificada foi utilizada como fase móvel em eluição isocrática, a um fluxo de 0,3mL/min, volume de injeção de 50 µL, em um tempo de corrida de 40 minutos.

A escolha da fase móvel foi uma modificação do método de referência, AOAC 980.13. A água ultrapurificada utilizada como fase móvel está citada no método ISO22662/IDF198:2007.

Figura 3.6- Foto do cromatógrafo HPLC-RID LC 20-A Prominence, com amostrador automático (Shimadzu, Japão)



Fonte: arquivo pessoal (2021)

### **Preparo das soluções padrão**

As soluções padrão dos açúcares, glicose, galactose e lactose, grau de pureza >99,9% (Sigma-Aldrich, USA), foram preparadas na concentração de 7,5mg/mL, a partir da pesagem de cada analito e diluição em 10 mL de água ultrapurificada.

### **Curvas de calibração**

Para a construção da curva de calibração de cinco pontos, foram pipetados 200 $\mu$ L, 400 $\mu$ L, 800 $\mu$ L, 1200 $\mu$ L e 2000 $\mu$ L das soluções padrão, em balão volumétrico de 5mL, correspondendo aos cinco níveis de concentrações: 0,3mg/mL; 0,6mg/mL; 1,2mg/mL; 1,8mg/mL e 3,0mg/mL.

Foram construídas as curvas no solvente e nos extratos das matrizes, para cada açúcar de interesse (glicose, galactose e lactose). Foram utilizadas como matrizes brancas o leite pasteurizado integral e o leite pasteurizado sem lactose, obtido no comércio varejista local. O uso das duas matrizes visaram reduzir a interferência da presença dos açúcares em quantidades desconhecidas na matriz, para se aproximar de uma matriz isenta do analito de interesse.

### **Extração da amostra**

Uma alíquota de 2 mL de cada amostra foi diluída em 25 mL de água ultrapurificada e, depois, homogeneizada em banho-maria a 40°C por 30 minutos. Após agitação no ultrassom por mais 30 minutos, o conteúdo foi transferido para um balão volumétrico de 50 mL e completado com água ultrapurificada. Transferiu-se 25 mL para ultracentrifugação a 19000 rpm, a 10°C, por 60 minutos. O resultado do centrifugado foi pipetado e filtrado em membrana de 0,45µm para vials. As corridas cromatográficas ocorreram imediatamente à extração.

Foram construídos gráficos correlacionando área do pico versus concentração e, por regressão linear, foram definidas as equações das curvas, o coeficiente de correlação ( $r$ ) e coeficiente de determinação ( $R^2$ ). A linearidade foi avaliada por meio do teste  $F$  (*Fischer-Snedecor*) na análise da variância da regressão, pelo teste de *Cochran* para verificação da homogeneidade dos resíduos e pela dispersão aleatória dos resíduos (INMETRO, 2020).

O efeito matriz foi avaliado pela comparação entre os coeficientes angulares das curvas de calibração construídas com os padrões dos analitos em solvente e no extrato da matriz fortificada. Para calcular o efeito matriz da lactose, foi utilizada como matriz “livre do analito” o leite pasteurizado sem lactose. Para o efeito matriz da glicose e da galactose, foi utilizado como matriz “livre do analito” o leite pasteurizado integral.

A comparação estatística entre as curvas construídas no solvente e nas duas matrizes foi realizada pelo teste  $F$  (*F-Snedecor*) para análise de homogeneidade de variâncias e teste  $t$  (*Student*) para comparação de médias, por nível de concentração, com nível de significância de 5% (INMETRO, 2020).

### 3.6 Comparação entre métodos HPLC e FTIR/redes neurais

#### Descrição

Nesta etapa do trabalho realizou-se a comparação dos resultados das concentrações de glicose, galactose e lactose por HPLC-RID nas amostras de leite hidrolisado com os resultados previstos pelos algoritmos de classificação e regressão.

#### Verificação dos algoritmos

Os espectros resultantes das análises por FTIR foram testados pelos algoritmos de classificação e regressão desenvolvidos na seção 3.3. Esta etapa de verificação do algoritmo desenvolvido, compreende um importante passo em relação a trabalhos anteriores que não realizaram esta etapa nos processos de construção de redes neurais com amostras de leite (Asseiss Neto *et al.*, 2019; Silva, L. *et al.*, 2020; Condé *et al.*, 2020).

A avaliação da qualidade preditiva dos modelos pela técnica do *cross-validation* e pelas métricas de desempenho, como acurácia, taxa de erros, taxa de perdas, coeficiente de determinação, entre outros, são fundamentais para validação do desempenho do método. Porém, realizar uma etapa de verificação com novas amostras, desconhecidas, submete o modelo a uma espécie de contra-prova do seu poder preditivo.

As respostas da execução dos algoritmos de classificação e regressão foram comparadas às respostas das amostras submetidas à identificação e quantificação dos açúcares por HPLC-RID, pelo teste de igualdade de variâncias e teste *t*, para comparação das médias, utilizando *software* estatístico Stata® 14 (StataCorp, USA).

# IV

## Resultados e Discussão

Este capítulo apresenta os resultados de toda a fase experimental desenvolvida. Resultados do Projeto Piloto para identificar diferenças nos espectros de absorção entre os açúcares no leite; resultados da espectroscopia por FTIR das amostras de leite adicionadas de açúcares para construção de redes neurais convolucionais; resultados dos algoritmos de classificação e regressão que identificaram e quantificaram os açúcares presentes no leite; resultados da otimização da hidrólise enzimática da lactose; resultados da cromatografia líquida para determinação de açúcares residuais em leite hidrolisado; resultados da comparação entre as concentrações de açúcares por HPLC e pelos algoritmos.

### 4.1 Resultados preliminares do Projeto Piloto

Os resultados das análises por FTIR são fornecidos em planilhas (arquivo .csv) com os valores centesimais de composição (lactose, proteína, sólidos totais, sólidos não gordurosos, caseína), nitrogênio ureico do leite, crioscopia e CCS.

Os resultados demonstraram que a leitura da lactose por FTIR sofre interferência dos outros açúcares, conforme observado por Ribeiro *et al.*, em 2018. Mesmo o leite hidrolisado, ou seja, sem lactose, apresenta leitura dentro dos parâmetros normais da legislação brasileira, mínimo de 4,3g/100mL (Brasil, 2018b), na coluna referente à lactose. E as amostras de leite sem lactose, com adição de diferentes açúcares, apresenta resultados como “lactose” pelo FTIR (Tabela 4.1).

Tabela 4.1- Valores considerados como “lactose” pelo equipamento FTIR em diferentes amostras de leite cru sem hidrólise, leite cru hidrolisado e leite cru hidrolisado (LH) com adição de açúcares

<b>Amostra</b>	<b>Lactose no FTIR</b>
leite cru sem hidrólise	4.69
leite cru hidrolisado	4.82
LH com lactose adicionada 0,1%	4.89
LH com lactose adicionada 0,5%	5.26
LH com lactose adicionada 1,0%	5.71
LH com lactose adicionada 5,0%	9.25
LH com glicose adicionada 0,1%	4.95
LH com glicose adicionada 0,5%	5.30
LH com glicose adicionada 1,0%	5.78
LH com glicose adicionada 5,0%	9.68
LH com galactose adicionada 0,1%	4.91
LH com galactose adicionada 0,5%	5.22
LH com galactose adicionada 1,0%	5.67
LH com galactose adicionada 5,0%	9.08
LH com sacarose adicionada 0,1%	4.90
LH com sacarose adicionada 0,5%	5.10
LH com sacarose adicionada 1,0%	5.36
LH com sacarose adicionada 5,0%	7.51

Fonte: dados da tese (2018)

O software do equipamento CombiScope™ Delta FTIR também gera dados espectrais das curvas resultantes da absorção do infravermelho (arquivo .spc, Thermo Scientific, Galactic Grams, SPC file format). Esses dados fornecem informações com grande potencial de interpretação e mineração.

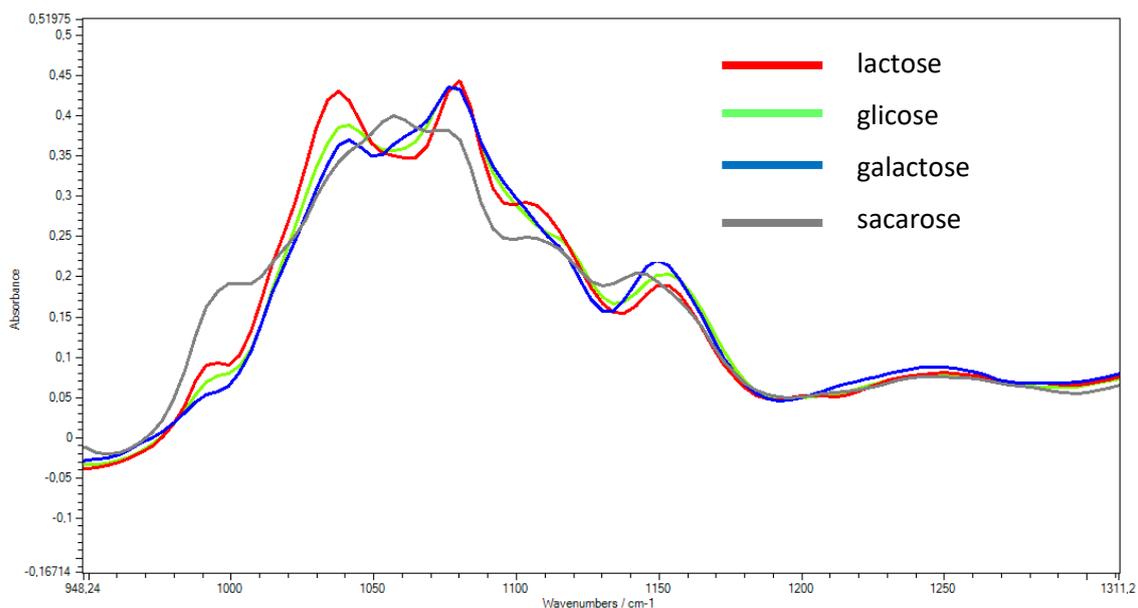
Com o programa Spectrum Viewer (Protea©, Reino Unido), os dados espectrais foram visualizados e processados, explorando as interações entre os açúcares adicionados e os outros componentes constituintes do leite, com o objetivo de identificar os perfis das curvas das substâncias e posterior mineração dos dados.

A absorção das ligações hidroxilas da lactose pela radiação infravermelha no comprimento de onda de 9,61µm permite a determinação quantitativa da lactose (AOAC 972.16). Este comprimento de onda corresponde a 1040 cm<sup>-1</sup>. Os açúcares adicionados ao leite causam picos de absorbância na mesma região da lactose, o que explica a interferência nos resultados de composição do leite, pois a

absorbância de luz infravermelha está correlacionada à quantidade do componente no leite (Luykx e Van Ruth, 2008).

A exploração das curvas e dos dados espectrais (Software Spectrum Viewer) demonstrou que a região do espectro de absorção dos açúcares compreende a faixa de 875 a 1481 $\text{cm}^{-1}$ . Apesar da similaridade espectral entre lactose, glicose e galactose, foi possível observar diferenças nos picos de absorbância pelos diferentes açúcares (Figura 4.1), na região que compreende a absorção do infravermelho pelos carboidratos do leite.

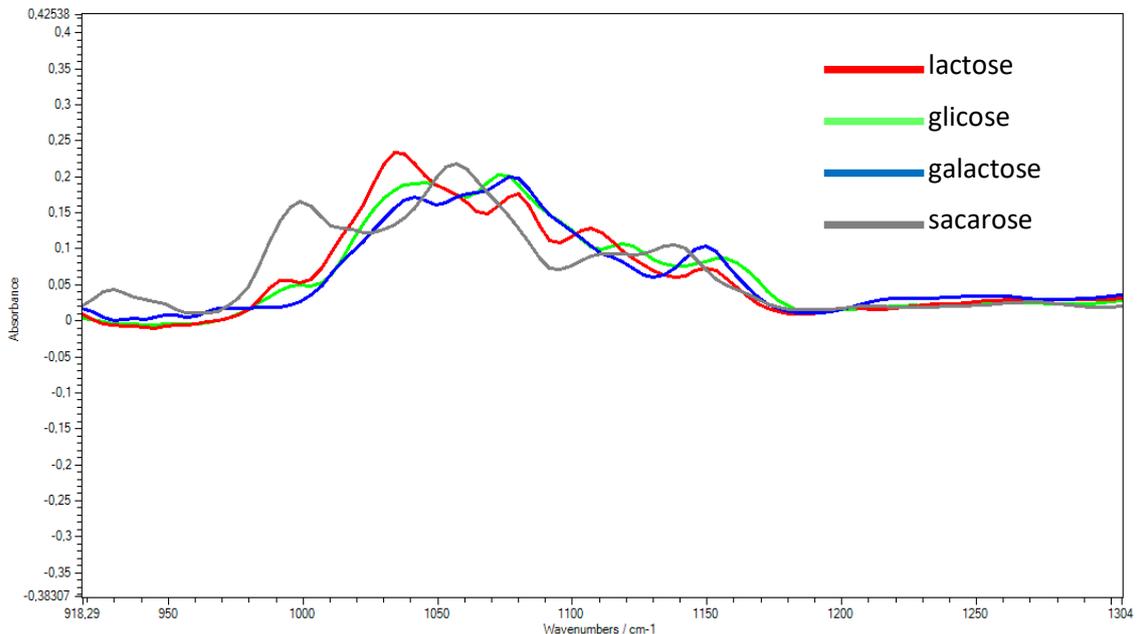
Figura 4.1- Espectros de absorção pelo FTIR, de amostras de leite hidrolisado adicionado de lactose, glicose, galactose e sacarose na concentração de 50mg/mL, na região de 940 a 1300 $\text{cm}^{-1}$



Fonte: autora (2018)

Pelas curvas espectrais de amostras de água adicionada de açúcares, na região do espectro que compreende a absorbância do infravermelho pelos carboidratos, observou-se que a absorbância (eixo y) foi menor nas curvas de água em comparação às curvas das amostras de leite, ambas com adição de açúcares (Figura 4.2), fato que decorre da intensidade de absorção da luz infravermelha estar diretamente relacionada à concentração da molécula na qual incide a luz (Luykx e Van Ruth, 2008).

Figura 4.2- Espectros de absorção pelo FTIR, de amostras de água deionizada fortificada com lactose, glicose, galactose e sacarose, na concentração de 50mg/mL, na região de 920 a 1300  $\text{cm}^{-1}$



Fonte: autora (2018)

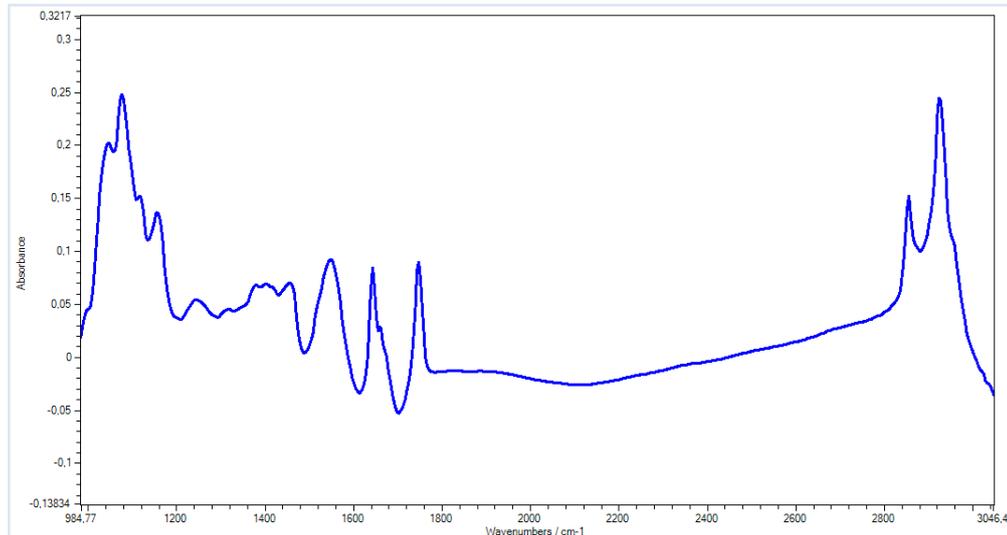
Os resultados obtidos no Projeto Piloto confirmaram que a leitura da lactose pelo FTIR sofre interferência dos monossacarídeos presentes no leite hidrolisado, glicose e galactose, devido à sobreposição das ondas de absorção dos espectros dos açúcares que levam à variação nas respostas do equipamento. Porém, a partir da exploração dos dados espectrais foi possível observar os perfis das curvas de absorção do infravermelho e identificar a distinção nos espectros de cada açúcar presente no leite hidrolisado.

#### 4.2 Resultados do FTIR para construção das Redes Neurais Convolucionais

Neste trabalho, as variáveis usadas como entrada para a Rede Neural Convolucional (CNN) foram todas as coordenadas do espectro resultante do FTIR. O resultado do espectro é uma sequência de valores resultantes da absorção de luz do equipamento na região do infravermelho médio, entre 400 e 4000  $\text{cm}^{-1}$ . Para a mineração dos dados, foi utilizada a região de 900 a 3100  $\text{cm}^{-1}$ , em intervalos de 4  $\text{cm}^{-1}$ , num total de 550 pontos do espectro (Figura 4.3). Uma rede neural

convolucional (CNN) tem capacidade de exploração sem necessidade de um pré-processamento dos dados.

Figura 4.3- Espectro de uma amostras de leite cru, analisado por FTIR



Fonte: autora (2021)

### Rede Neural Convolucional

A engenharia de arquitetura da CNN executada neste trabalho, considerou os hiperparâmetros que estão listados na Tabela 4.2. Por meio deste processo, foram definidas as variáveis para controlar o treinamento.

O treinamento da CNN foi feito pelo otimizador *Adam (Adaptive Moment Estimation) Optimizer* com 300 iterações (*epochs*) para as classificações e 250 iterações para as regressões. À medida que o treinamento ocorreu, parte do conjunto foi submetido à uma validação interna à rede. Na CNN executada, 20% do conjunto de treinamento foi considerado como conjunto de validação; e para cada *epoch* da rede, durante o treinamento e validação, históricos de acurácia e perda foram armazenados. Com o aumento das *epochs*, o aprendizado da rede evoluiu (acurácia) e os erros diminuíram (perdas).

Tabela 4.2- Hiperparâmetros utilizados para controlar o treinamento da rede neural convolucional

Hiperparâmetro	Descrição	Valores
Camadas convolucionais	Número de camadas convolucionais	1
Filtros convolucionais	Número de filtros convolucionais	60
Tamanho de <i>kernel</i>	Comprimento da janela de convolução	20
Camadas densas	Número de camadas totalmente conectadas inseridas após as camadas convolucionais	1
Tamanho de camada densa	Número de neurônios em cada camada densa	2048
Taxa de <i>dropout</i>	Taxa de neurônios ignorados aleatoriamente para prevenir <i>overfitting</i> na rede	0.1
Tamanho de <i>max-pooling</i>	Reduz a resolução da entrada	4
Ativação	Função que define a saída de um neurônio	LeakyReLU

Fonte: dados da tese (2021)

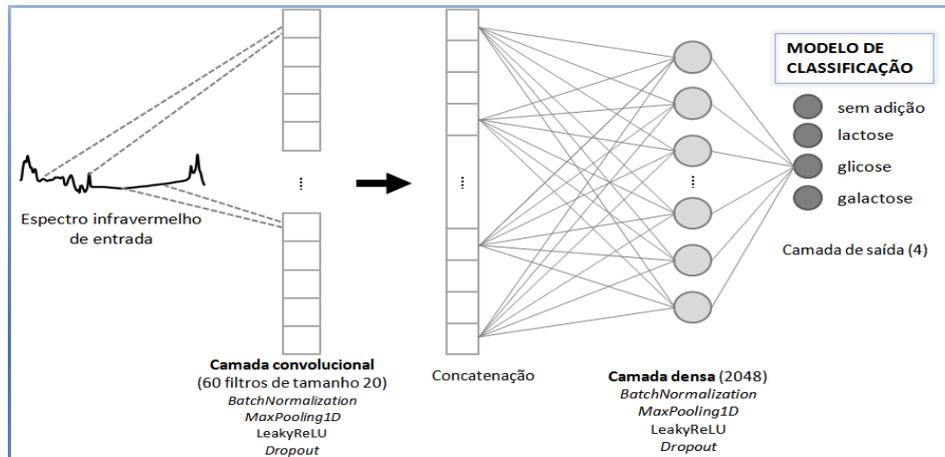
### 4.3 Resultados dos algoritmos

#### CNN para classificação

Nos modelos de classificação, considerando um problema multiclasse, o número de neurônios de saída corresponde ao número de respostas possíveis. Foram testados dois modelos de classificação multiclasse, um para diferenciar o leite normal (sem adição) do leite adicionado com algum dos açúcares adicionados; e no outro, o modelo identificou qual açúcar foi adicionado ao leite.

A primeira CNN para classificação utilizou como dados de entrada os resultados dos espectros das amostras de leite normal (sem adição de açúcar – amostra branca) e das amostras de leite com adição de açúcares - glicose, galactose e lactose, num total de 311 amostras. O resultado foi uma rede com quatro neurônios na saída (Figura 4.4), que pode ser interpretado como um modelo para identificação de açúcar adicionado ao leite.

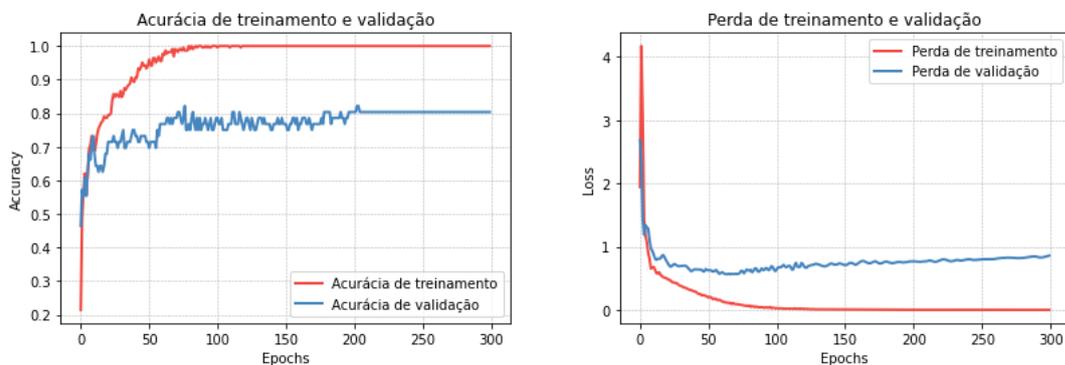
Figura 4.4- Representação esquemática de uma Rede Neural Convolucional para identificação de açúcares adicionados ao leite



Fonte: autora (2021), modificado de Asseiss Neto *et al.* (2019)

O modelo apresentou ótimos resultados no treinamento, observados nas plotagens da acurácia e da perda durante as *epochs* (iterações) do treinamento do algoritmo (Figura 4.5). A acurácia foi calculada comparando a previsão correta entre a classe predita e a classe real, e a perda foi calculada pelo erro de previsão entre a classe predita e a classe real. Para este problema multiclasse, a classe predita é a classe com maior probabilidade entre as quatro classes possíveis. E após o método de validação cruzada (*10-fold cross validation*), o teste apresentou acurácia de 80,07%.

Figura 4.5- Gráficos de acurácia e perda durante o treinamento e validação do modelo CNN para classificação de açúcares adicionados ao leite, treinada por 300 *epochs*

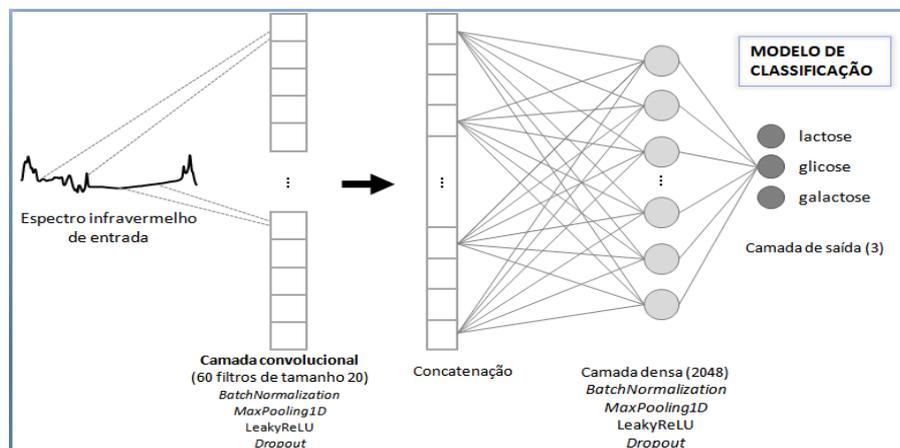


Legenda: Divisão do conjunto de dados do treinamento de 80%/20%.

Fonte: autora (2021)

Considerando que toda amostra de leite normal tem lactose naturalmente presente em torno de 5%, então, uma outra construção de algoritmo foi proposta. Nesta outra execução de CNN para classificação, os dados de entrada foram somente as amostras adicionadas de lactose, glicose e galactose, num total de 225 amostras. Para este modelo não foram utilizadas amostras de leite normal (sem adição de açúcar). O resultado foi uma rede com três neurônios na saída (Figura 4.6), que podem ser interpretados como um modelo para identificar a presença do açúcar no leite.

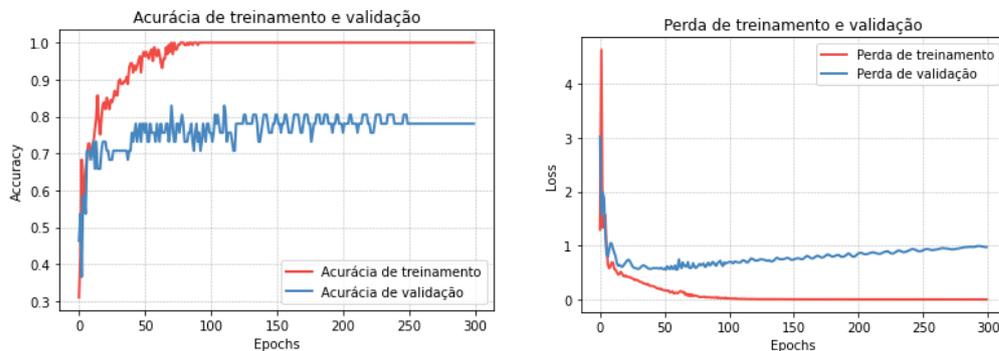
Figura 4.6- Representação esquemática de uma Rede Neural Convolutacional para identificar a presença de açúcar no leite



Fonte: autora (2021), modificado de Asseiss Neto *et al.* (2019).

Na Figura 4.7 observamos as plotagens da acurácia e da perda durante as *epochs* do treinamento do algoritmo. Para este problema multiclasse, a classe predita é a classe com maior probabilidade entre as três classes possíveis. Este modelo também apresentou ótimos resultados no treinamento. E após o método de validação cruzada (*10-fold cross validation*), o teste apresentou acurácia de 75,99%.

Figura 4.7- Gráficos de acurácia e perda durante o treinamento e validação do modelo CNN para classificação de açúcares presentes no leite, treinada por 300 *epochs*



Legenda: Divisão do conjunto de dados do treinamento de 80%/20%.

Fonte: autora (2021)

A acurácia é a proporção de casos que foram corretamente previstos, sejam eles verdadeiro positivo ou verdadeiro negativo. É considerada a taxa de acerto. Na Figura 4.8, uma matriz de confusão foi construída para melhor visualização dos verdadeiros positivos (VP), falsos positivos (FP), verdadeiros negativos (VN) e falsos negativos (FN).

Figura 4.8- Matriz de confusão do modelo de classificação de açúcares adicionados ao leite

		Valor Previsto			
		lactose	glicose	galactose	total
Valor Real	lactose	55	11	11	<b>77</b>
	glicose	5	53	14	<b>72</b>
	galactose	6	11	59	<b>76</b>
	total	66	75	84	<b>225</b>

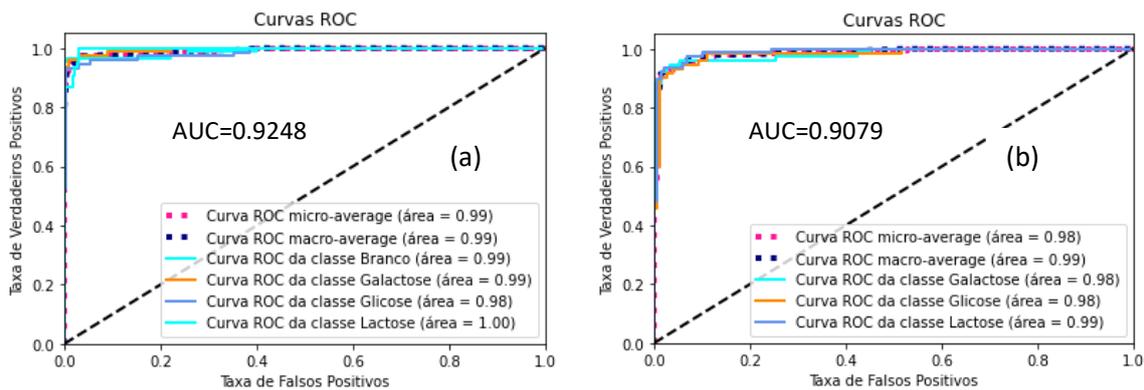
Acurácia por classe de açúcar: Lactose = 0.7143; Glicose = 0.7361; Galactose = 0.7763

Fonte: autora (2021)

A curva ROC (*Receiver Operating Characteristics*) e AUC (*Area Under the Curve*) são duas métricas muito utilizadas para modelos de classificação, que avaliam a relação entre a sensibilidade e a especificidade. Elas foram avaliadas para demonstrar mais uma medida de desempenho dos modelos. A Figura 4.9 apresenta a curva ROC para cada açúcar identificado.

A área sob a curva ROC, ou AUC, é uma medida que sumariza o desempenho da ROC. Neste trabalho os valores de AUC das duas CNN para classificação foram de 0.9248 e 0.9079, e significa que os modelos têm alta capacidade de distinguir entre as classes avaliadas.

Figura 4.9- Curvas ROC e AUC para as Redes Neurais Convolucionais executadas para classificação de açúcares no leite



Legenda: a estratégia *micro-average* agrega os valores de todas as classes e a *macro-average* computa os valores de cada classe. (a) CNN com amostras brancas; (b) CNN sem amostras brancas. Fonte: autora (2021)

### CNN para regressão

A partir do conhecimento de que cada açúcar tem um espectro próprio, o modelo aprendeu a reconhecer qual açúcar está presente no leite e em qual quantidade. Foram executadas cinco diferentes redes neurais. Todas utilizaram como dados de entrada os espectros resultantes das amostras adicionadas com os três açúcares em seis diferentes concentrações, numa faixa de 1mg/mL a 70mg/mL.

Os dados foram considerados de forma distinta para execução de cada CNN. Uma CNN tomou todas as amostras de leite com adição de açúcares como um único grupo, um único rótulo. Neste modelo, o resultado foi a capacidade preditiva de quantificar açúcar em leite, sem distinção por sacarídeo.

A segunda CNN baseou-se na mesma informação do modelo anterior, porém, esta rede utilizou os espectros de uma região menor, específica de absorção do

infravermelho pelos açúcares, de 875 a 1481  $\text{cm}^{-1}$ . Os resultados foram um pouco inferiores aos da primeira CNN, conforme demonstrado na Tabela 4.3, pelas métricas de desempenho para modelos de regressão, como MSE (*mean square error*), RMSE (*root mean square error*), MAE (*mean absolute error*) e  $R^2$  (coeficiente de determinação). O coeficiente de determinação de 0.9538 significa que o modelo explica mais de 95% da variação nos valores esperados (Kuhn e Johnson, 2018). Estas métricas de capacidade preditiva foram determinadas após o método de validação cruzada (*10-fold cross validation*).

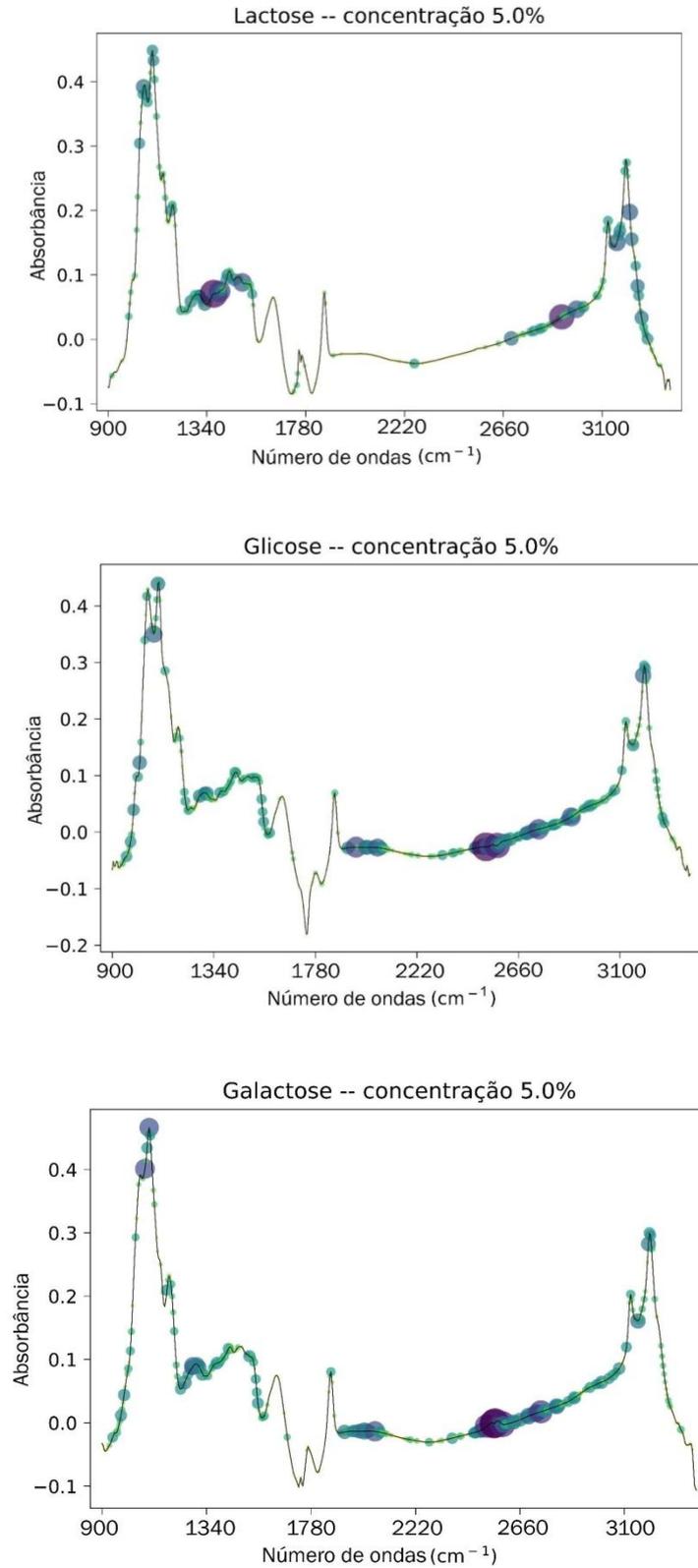
Tabela 4.3- Resultados das métricas de desempenho dos modelos de regressão com as redes neurais convolucionais em duas diferentes regiões espectrais

Métricas de desempenho	Região 900 a 3100 $\text{cm}^{-1}$	Região 875 a 1481 $\text{cm}^{-1}$
RMSE	0.4828	0.5861
MSE	0.2331	0.3435
MAE	0.3273	0.4324
$R^2$	0.9538	0.9336

Fonte: dados da tese (2021)

Utilizar todas as coordenadas do espectro resultante do FTIR mostrou melhores resultados que a utilização de uma região mais restrita, específica à absorção de luz pelos carboidratos. Uma rede neural convolucional, que aprende com os dados do espectro semelhante à uma imagem unidimensional, extrai todas as possíveis informações quando utiliza todo o espectro. Com a construção dos mapas de saliências, observa-se que todo o espectro contribuiu para a decisão da rede. Na Figura 4.10, as regiões com destaques maiores e mais escuros representam a intensidade com que cada coordenada contribuiu para a ativação de cada saída (açúcar) na CNN.

Figura 4.10- Saliências calculadas pela rede neural convolucional com os espectros das amostras de leite adicionadas de 50 mg/mL de lactose, glicose e galactose

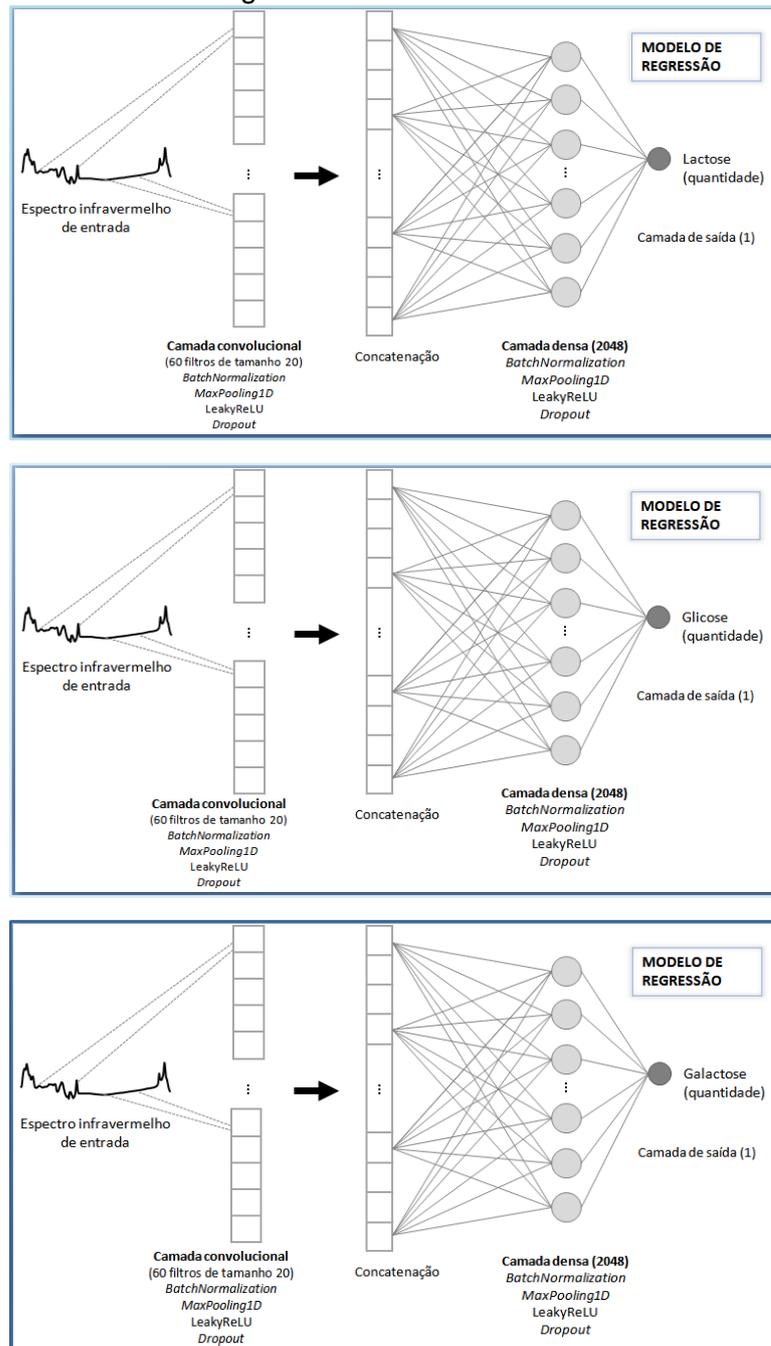


Fonte: autora (2021)

## Capacidade preditiva da quantificação de açúcares no leite

Para quantificação de lactose, glicose e galactose foram executadas três diferentes CNN (Figura 4.11). Cada rede neural utilizou como dados de entrada os espectros próprios de cada açúcar. E o neurônio de saída correspondeu à quantidade predita do açúcar.

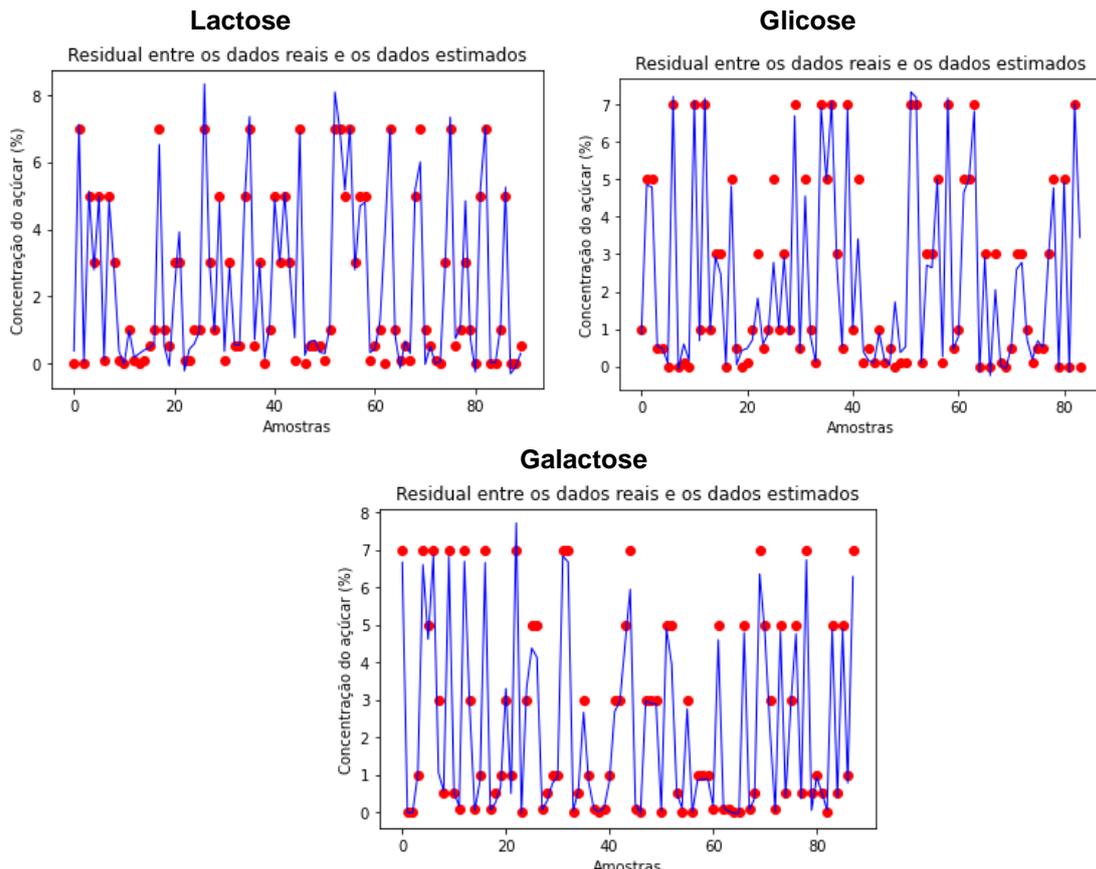
Figura 4.11- Representação esquemática de Rede Neural Convolutacional para quantificar açúcar em leite, com modelo de regressão



Fonte: autora (2021), modificado de Asseiss Neto *et al.*, 2019.

Na Figura 4.12, a adequação dos modelos pode ser observada pela construção do gráfico de resíduos, que são as diferenças entre os valores preditos e os valores reais.

Figura 4.12- Gráfico de resíduos dos algoritmos de regressão para quantificar lactose, glicose e galactose

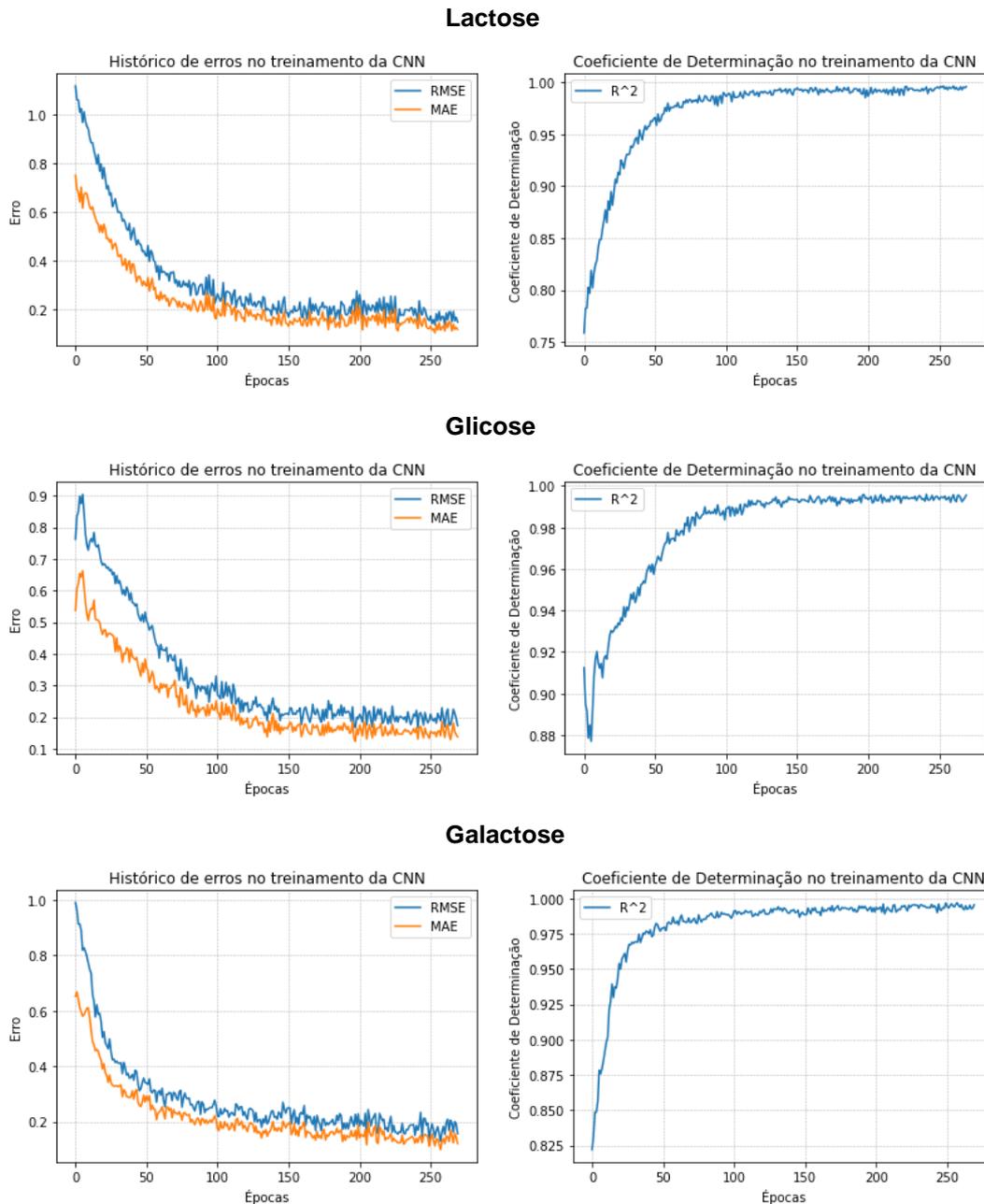


Legenda: a linha azul - valores preditos; pontos vermelhos - valores reais das concentrações das amostras.

Fonte: autora (2021)

Durante o treinamento, a rede foi aprendendo e diminuindo a taxa de erros gerados pelo modelo. Na Figura 4.13, a diminuição no histórico de erros e o aumento do coeficiente de determinação, na fase de treinamento da rede, demonstra que o modelo foi aumentando a sua capacidade preditiva ao longo das 250 *epochs*, com os algoritmos de regressão para lactose, glicose e galactose.

Figura 4.13- Gráficos do histórico de erros e do coeficiente de determinação ao longo das 250 *epochs* de treinamento da CNN para quantificação de lactose, glicose e galactose no leite



Fonte: autora (2021)

A validação cruzada utilizada, *10-fold cross-validation*, nos modelos de regressão resultaram em métricas de desempenho que estão apresentadas na Tabela 4.4. Os coeficientes de determinação estão próximos de 1, e representam a proporção de variância nos valores preditos que podem ser explicados pelo modelo. O RMSE é uma métrica independente de escala, quanto menor o valor, melhor.

Tabela 4.4- Resultados das métricas de desempenho dos modelos de regressão para quantificar lactose, glicose e galactose, após 10-fold cross-validation

Métricas de desempenho	Lactose	Glicose	Galactose
RMSE	0.8646	0.7981	0.6110
MSE	0.7475	0.6369	0.3733
MAE	0.5997	0.5564	0.4748
R <sup>2</sup>	0.8106	0.8552	0.9177

Fonte: dados da tese (2021)

Para modelos que preveem um resultado numérico, existem maneiras diferentes de medir a precisão, cada uma com sua própria nuance. Para entender os pontos fortes e fracos de um modelo específico, não se deve contar apenas com uma única métrica. Além do RMSE, MSE, R<sup>2</sup>, as visualizações do ajuste do modelo, particularmente pelo gráfico de resíduos (Figura 4.12), são essenciais para entender se o modelo é adequado para o propósito (Kuhn e Johnson, 2018).

Um problema de classificação binária tende a ser mais simples e pode obter melhores resultados que um problema de classificação multiclasse (Asseiss Neto *et al.*, 2019). Neste trabalho, o algoritmo de classificação foi executado para um problema multiclasse, em que as classes foram açúcares (lactose, glicose e galactose) com a mesma região do espectro de absorção da luz infravermelho.

Apesar da similaridade espectral entre os açúcares, as diferenças observadas nos picos dos espectros de cada um foram reconhecidas pelas redes neurais convolucionais e tornaram possíveis identificar e quantificar os açúcares presentes no leite.

Amido e sacarose também absorvem a luz infravermelho na mesma região do espectro (Ribeiro *et al.*, 2018; Conceição *et al.*, 2019; Hansen e Holroyd, 2019), porém os picos de absorção no espectro da sacarose são ainda mais distintos da lactose, glicose e lactose, conforme observado na Figura 4.1.

Conceição *et al.* (2019) utilizaram RNA para identificação de adulterantes em leite. No modelo de classificação proposto, encontraram um coeficiente de determinação ( $R^2$ ) de 0,58 para identificação de sacarose e 0,53 para amido.

Silva, L. *et al.* (2020) propuseram um modelo de RNA para identificar autenticidade de leite de búfala. Amostras de leite de búfala foram adicionadas de leite de vaca, elaborando um problema de classificação com cinco possíveis resultados. O resultado da melhor rede neural artificial foi escolhido com base nas métricas de RMSE e  $R^2$ , e obteve uma alta taxa de classificação somente quando a adulteração foi superior a 30%.

A CNN utilizada neste trabalho foi capaz de identificar a presença de açúcares no leite a partir da adição de 0,1%.

Condé *et al.* (2020) executaram um modelo de RNA para quantificar fraude de leite por adição de soro de leite, utilizando como dados de entrada *features* de composição do leite (gordura, proteína, etc.), porém os valores preditos pelo modelo desviaram dos valores reais, e 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.

A construção de rede neural convolucional para problemas de regressão difere do trabalho de Asseiss Neto *et al.* (2019) que utilizaram CNN para classificação de amostras de leite com cinco adulterantes e não utilizaram a técnica de *K-fold cross-validation*.

A técnica *K-fold cross-validation* em comparação às técnicas de divisão do conjunto de dados em proporções para treinamento, teste e validação, oferece uma visão mais realista do poder preditivo dos modelos. Na *cross-validation* todas as instâncias são treinadas e testadas, evitando incorrer em uma possível distribuição desigual dos dados nos subconjuntos. Trabalhos anteriores, empregando redes neurais em leite, também não utilizaram *cross-validation* para validação do método. Conceição *et al.* (2019) e Silva, L. *et al.* (2020), utilizaram grupos de treinamento e validação, na

proporção 70/30 e 80/20, respectivamente. E Condé *et al.* (2020) dividiram os dados em 60/20/20, para treinamento, teste e validação.

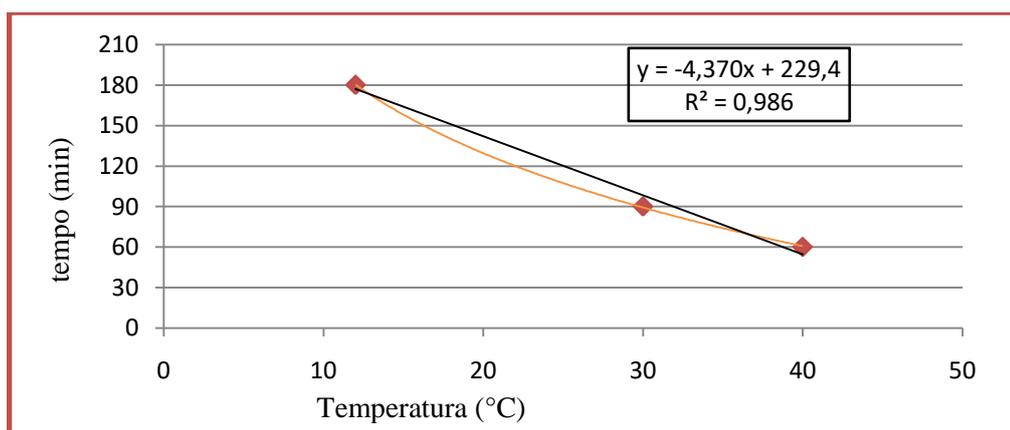
#### 4.4 Resultados da otimização da hidrólise da lactose

##### Resultados das condições de hidrólise em diferentes temperaturas e tempos

Os resultados indicaram que o limite de IC (-0,653°H) do equipamento foi alcançado após 3 horas a 12°C, após 90 minutos a 30°C e 60 minutos a 40°C. Houve correlação inversa da temperatura do leite em relação ao tempo da reação enzimática. Quanto mais alta a temperatura da amostra durante a hidrólise, respeitando o limite para que não ocorresse a desnaturação da enzima, menor foi o tempo da reação, conforme observado por Longo (2006) e Trevisan (2008).

Foi construída uma curva correlacionando a velocidade da hidrólise da lactose em função da temperatura da amostra, apresentada na Figura 4.14. Observa-se um alto grau de correlação ( $R^2 = 0,986$ ) entre as variáveis.

Figura 4.14- Relação e correlação da temperatura do leite (12°C, 30°C e 40°C) e os tempos para atingirem o índice crioscópico limite do estudo, na hidrólise enzimática da lactose com concentração da enzima 40  $\mu\text{L}/100\text{mL}$



Fonte: autora (2021)

Devido ao limite de mensuração do IC do equipamento, o máximo de hidrólise, estimada pela equação citada, foi de 37,19%. A Tabela 4.5 apresenta o tempo de

hidrólise da lactose em função da temperatura, considerando o IC e o percentual de hidrólise medidos em intervalos de 30 minutos.

Tabela 4.5- Influência da temperatura do leite no tempo da reação enzimática da lactose (40µL/100mL de enzima), a partir do índice crioscópico (IC) e percentual de hidrólise

12°C			30°C			40°C		
tempo (min)	Hidrólise (%)	IC (°H)	tempo (min)	Hidrólise (%)	IC (°H)	tempo (min)	Hidrólise (%)	IC (°H)
0	0	-0,547	0	0	-0,547	0	0	-0,547
30	12,98	-0,584	30	14,03	-0,587	30	35,09	-0,647
60	21,05	-0,607	60	30,88	-0,635	60	37,19	-0,653
90	27,02	-0,624	90	37,19	-0,653	90	..	..
120	34,74	-0,646	120	..	..	120	..	..
150	36,49	-0,651	150	..	..	150	..	..
180	37,19	-0,653	180	..	..	180	..	..

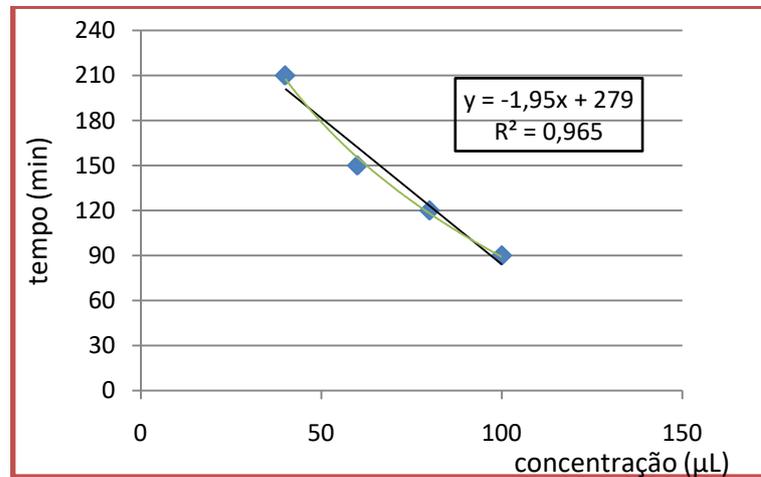
Fonte: dados da tese (2021)

### Resultados das condições de hidrólise em diferentes concentrações da enzima

Após 6 horas de hidrólise, a 7°C, a amostra com 20 µL/100mL não atingiu o limite de IC do equipamento, ao contrário do observado nas amostras com as concentrações maiores. Na concentração de 40 µL/100mL o IC limite foi atingido aos 210 minutos, com 60 µL/100mL foi aos 150 minutos, com 80 µL/100mL foi aos 120 minutos e com 100 µL/100mL foi aos 90 minutos.

Foi construída uma curva correlacionando o tempo que a hidrólise alcançou o IC limite estudado em função das concentrações de enzima testadas, conforme apresentada na Figura 4.15. Observa-se um alto grau de correlação ( $R^2= 0,965$ ) entre as variáveis. A correlação inversa pode ser observada pelo sinal negativo no eixo x da equação, ou seja, quanto mais alta a concentração de enzima, menor o tempo de reação enzimática.

Figura 4.15- Relação e correlação das concentrações de enzima  $\beta$ -galactosidase (40, 60 80 e 100  $\mu$ L/100mL) e os tempos para atingirem o índice crioscópico limite do estudo, na hidrólise enzimática da lactose, a 7°C



Fonte: autora (2021)

## 4.5 Resultados da Cromatografia

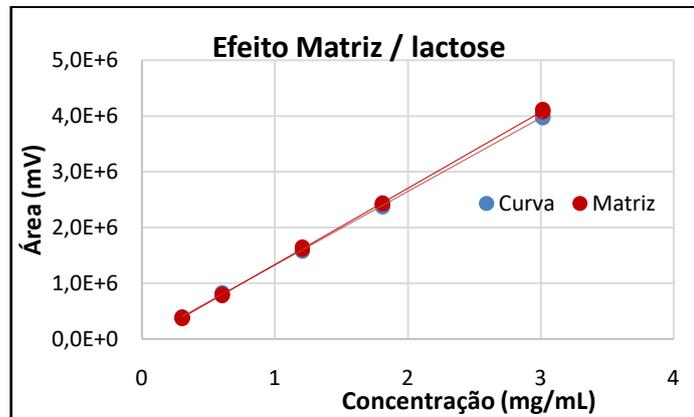
### Linearidade, Efeito Matriz e Seletividade

Linearidade é a habilidade de um método analítico em produzir resultados que sejam diretamente proporcionais à concentração do analito na amostra, em uma dada faixa de concentração (INMETRO, 2020). A faixa de trabalho compreende as concentrações em que a linearidade for atingida, definindo-se, assim, a curva de calibração correspondente.

No método proposto, a faixa de trabalho foi de 0,3mg/mL a 3,0mg/mL, definida pelo próprio método normalizado (FUNED, 2020).

Para calcular o efeito matriz duas curvas foram construídas para cada analito, para comparar os coeficientes angulares das curvas na presença da matriz e sem matriz (somente padrão em solvente). Para avaliar o efeito matriz da lactose, utilizou-se, como matriz “livre do analito”, o leite pasteurizado com teor reduzido de lactose (“zero lactose”). A curva foi construída nas concentrações de 0,3mg/mL; 0,6mg/mL; 1,2mg/mL; 1,8mg/mL e 3,0mg/mL, em três repetições (Figura 4.16).

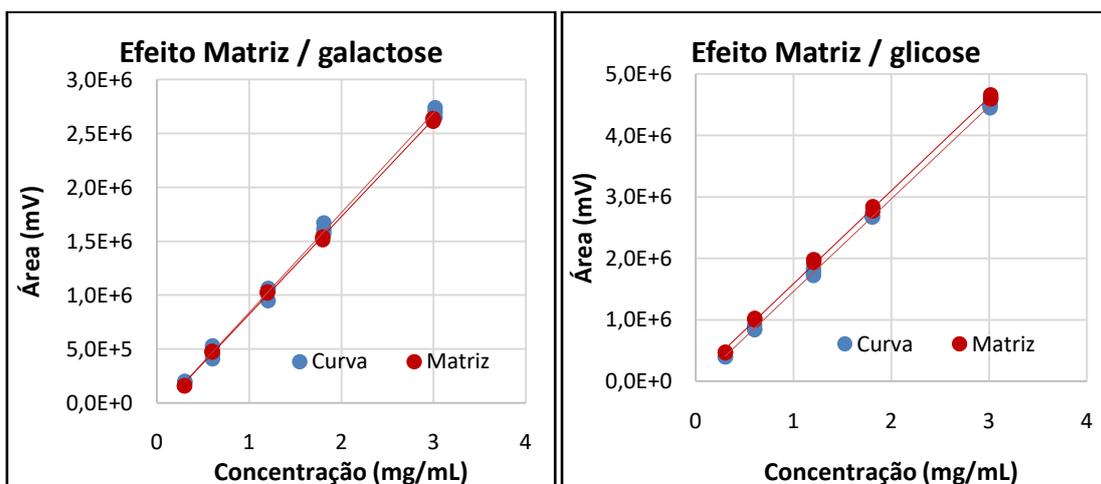
Figura 4.16- Curva da lactose em solvente e na matriz, na faixa de trabalho de 0,3mg/mL a 3,0mg/mL, por HPLC-RID



Fonte: autora (2021)

Para avaliar o efeito matriz da glicose e da galactose, utilizou-se, como matriz “livre do analito”, o leite pasteurizado integral normal, que não possui esses analitos na sua composição. As duas curvas foram construídas, para cada analito, na presença da matriz e no solvente (Figura 4.17), nas concentrações de 0,3mg/mL; 0,6mg/mL; 1,2mg/mL; 1,8mg/mL e 3,0mg/mL, em três repetições.

Figura 4.17- Curvas da glicose e galactose em solvente e na matriz, na faixa de trabalho de 0,3mg/mL a 3,0mg/mL, por HPLC-RID



Fonte: autora (2021)

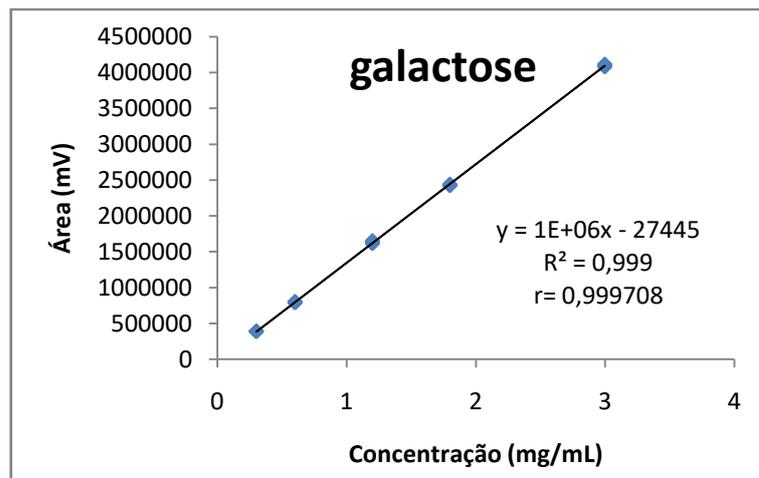
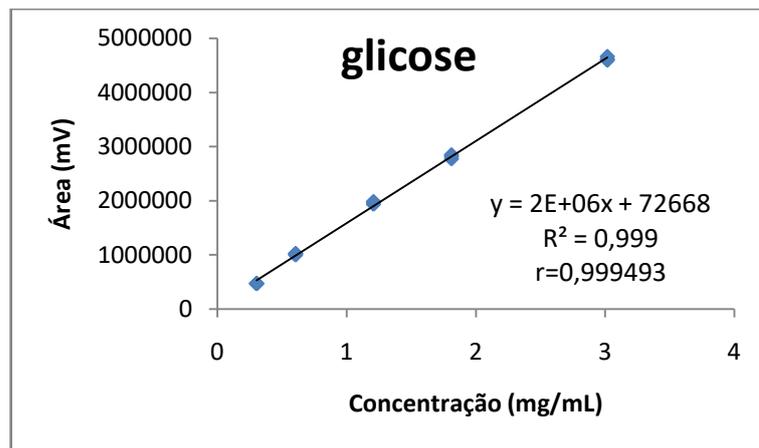
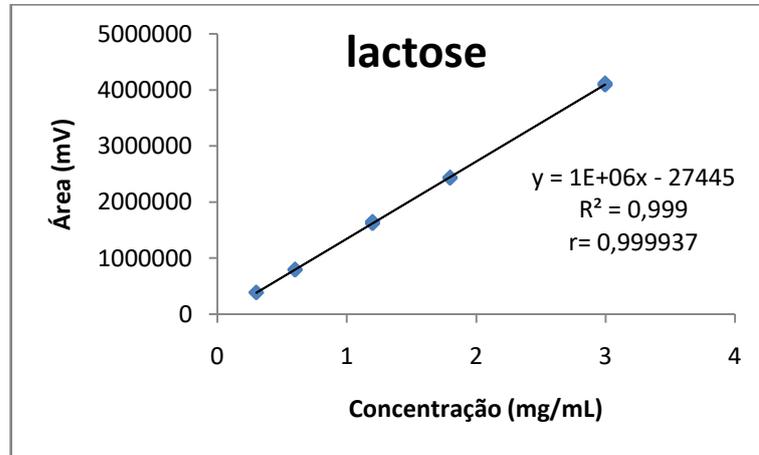
Apesar do paralelismo das retas ser indicativo de ausência de interferência dos constituintes da matriz, a avaliação do efeito matriz foi realizada por meio de avaliação estatística adequada (Brasil, 2017c). Para comparar estatisticamente as curvas, foram realizados, para cada analito, o Teste  $F$  ( $F$ -Snedecor) de homogeneidade de variâncias, por nível de concentração e o teste  $t$  ( $Student$ ) de comparação de médias, com um nível de significância de  $\alpha = 0,05$  (5%). O teste  $F$  não foi significativo, logo, a matriz não teve efeito importante sobre a precisão do método nessa faixa de fortificação considerada. Já o teste  $t$  foi significativo, portanto, a matriz causou efeito sobre o resultado, então, as inclinações das curvas para os três analitos foram consideradas estatisticamente diferentes, ou seja, houve efeito matriz (INMETRO, 2020).

Confirmado o efeito matriz nos três analitos de interesse, os gráficos foram construídos na curva matrizada. Estes dados foram usados para estudo da linearidade e seletividade.

A linearidade foi avaliada pela equação linear, pelo coeficiente de correlação ( $r$ ) e coeficiente de determinação ( $R^2$ ). As curvas foram construídas nas matrizes fortificadas com cada açúcar de interesse (glicose, galactose e lactose), na faixa de trabalho descrita, em triplicata (Figura 4.18).

A variância dos resíduos foi homogênea, ou seja, homocedástica, avaliada pelo teste de *Cochran*, e o teste  $F$  não foi significativo, ou seja, as curvas dos três açúcares se apresentaram lineares aos testes.

Figura 4.18- Curvas de linearidade da lactose, glicose e galactose, com a equação da regressão linear, o coeficiente de correlação (r) e coeficiente de determinação (R<sup>2</sup>) nas concentrações de 0,3mg/mL a 3,0mg/mL



Fonte: autora (2021)

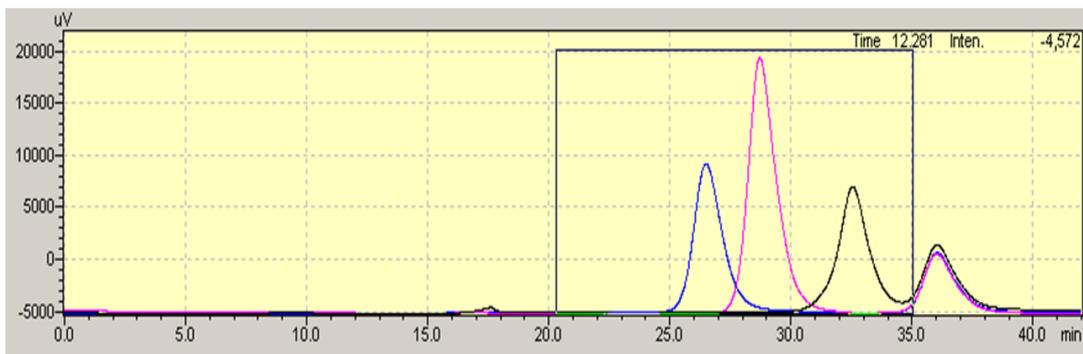
O coeficiente de correlação linear ( $r$ ) é um bom indicativo do quanto a reta pode ser considerada adequada como modelo matemático. Os coeficientes de correlação ( $r$ ) obtidos foram acima de 0,99. Este parâmetro permite uma estimativa da qualidade da curva obtida, pois quanto mais próximo de 1 menor a dispersão dos dados e menor a incerteza da equação da regressão estimada. O coeficiente de determinação ( $R^2$ ) também se apresentou adequado, sendo superior a 0,99, indicando um excelente ajuste dos dados para a linha de regressão.

### Avaliação dos Cromatogramas

O tempo de retenção da lactose, glicose e galactose foram observados no cromatograma após corrida dos padrões dos analitos em solvente e os resultados estão demonstrados na Figura 4.19. O tempo de retenção ( $t_r$ ) da lactose foi 26,5 minutos, o  $t_r$  da glicose foi 28,7 minutos e da galactose foi 32,6 minutos. O tempo total da corrida foi de 40 minutos. No  $t_r$  36,2 minutos observa-se a presença de um contaminante, que não interferiu na análise por não se encontrar no mesmo tempo de retenção dos analitos de interesse.

A identificação dos açúcares foi realizada através da comparação dos tempos de retenção obtidos pela injeção de soluções padrões individuais de cada um dos açúcares estudados, com os sinais obtidos das amostras e das matrizes fortificadas com as soluções padrões.

Figura 4.19- Cromatograma dos padrões dos analitos em solvente para observação dos tempos de retenção da lactose, glicose e galactose, por HPLC-RID



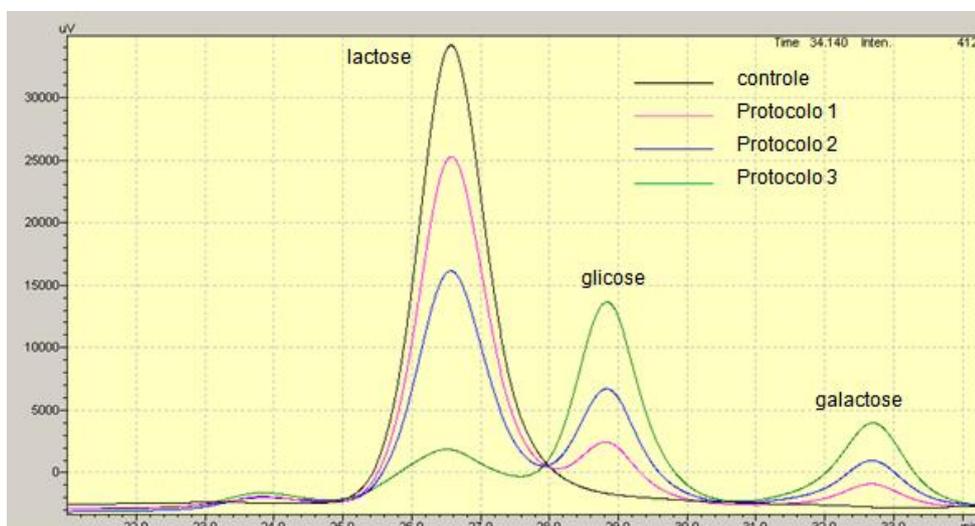
Fonte: autora (2021)

A água ultrapurificada empregada como fase móvel, apresentou excelente resolução entre lactose, glicose e galactose. Porém, Acquaro Jr e colaboradores (2013), que também utilizaram água como fase móvel, obtiveram coeluição da lactose com galactose.

Chávez-Servin *et al.*, em 2004, também utilizaram HPLC-RID para separação de sacarídeos, porém os picos de glicose e galactose ficaram sobrepostos.

Os cromatogramas resultantes dos três protocolos (1, 2 e 3) de hidrólise enzimática da lactose descritos na seção 3.4 estão demonstrados na Figura 4.20. Observe que os três níveis de hidrólise (baixo, médio e alto) foram alcançados, conforme o esperado, a fim de se obter um número de resultados significativamente diferentes para serem comparados, a posteriori, com os resultados das redes neurais.

Figura 4.20- Cromatogramas superpostos da repetição 1, da amostra controle e das amostras com os três protocolos de hidrólise enzimática da lactose

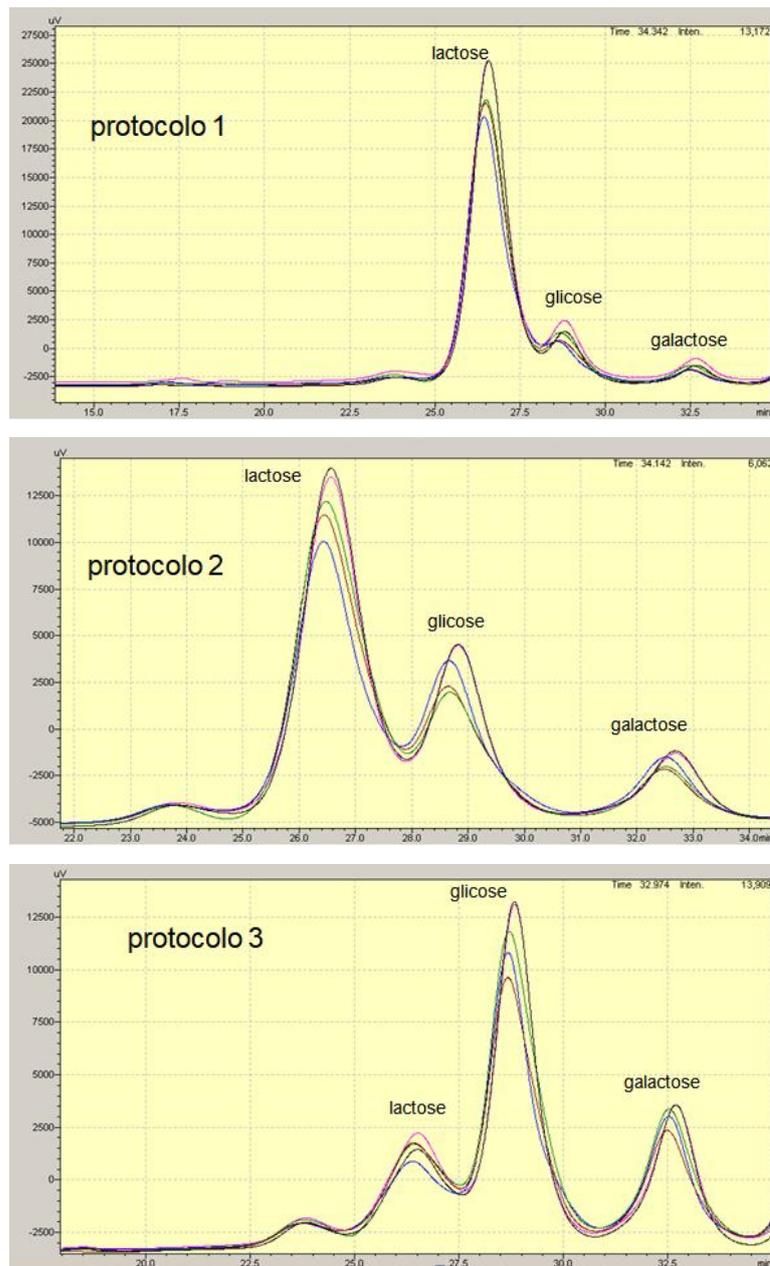


Legenda: Controle: sem hidrólise; Protocolo 1: 10  $\mu\text{L}/100\text{mL}$  de leite a  $7^\circ\text{C}$ ; Protocolo 2: 20  $\mu\text{L}/100\text{mL}$  de leite a  $7^\circ\text{C}$ ; Protocolo 3: 80  $\mu\text{L}/100\text{mL}$  de leite a  $7^\circ\text{C}$ . Tempo de hidrólise: 24 horas. Enzima:  $\beta$ -galactosidase de *Kluyveromyces lactis*,  $\geq 2600$  units/g (Sigma-Aldrich, USA, CAS 9031-11-2).

Fonte: autora (2021)

Foram cinco repetições de amostras de leite pasteurizado, conservadas a 7°C, e submetidas aos três protocolos de hidrólise. Na Figura 4.21, os cromatogramas superpostos das cinco repetições, submetidas aos protocolos de hidrólise enzimática da lactose estão representados.

Figura 4.21- Cromatogramas representando as cinco repetições de leite pasteurizado refrigerado, após hidrólise enzimática da lactose com três concentrações de  $\beta$ -galactosidase 10  $\mu$ L/100mL, 20  $\mu$ L/100mL e 80  $\mu$ L/100mL, por 24 horas, nos protocolos 1, 2 e 3



Fonte: autora (2021)

A concentração final dos açúcares nas amostras foi calculada pela fórmula abaixo, extraída da documentação do procedimento (FUNED, 2020):

$$C_{(g/Pg \text{ ou } mL)} = \frac{(y_{(mV)} - b) \cdot V_b \text{ (mL)} \cdot V_f \text{ (mL)} \cdot P_{(g \text{ ou } mL)}}{a \cdot V_i \text{ (\mu L)} \cdot V_a \text{ (mL)} \cdot A_{(g \text{ ou } mL)} \cdot 1000}, \text{ em que:}$$

C = concentração do analito

y = área do pico

b = intercepto

V<sub>b</sub> = volume do balão utilizado

V<sub>f</sub> = volume final antes da injeção

P = porção declarada na rotulagem

a = coeficiente angular

V<sub>i</sub> = volume de injeção

V<sub>a</sub> = volume da tomada de amostra

A = volume ou massa da amostra

Fonte: FUNED (2020)

Os resultados obtidos das análises cromatográficas dos açúcares em amostras de leite, antes e após a hidrólise enzimática da lactose, estão apresentados na Tabela 4.6. Os percentuais de hidrólise alcançados, a partir dos três protocolos, demonstraram que o objetivo de se obter três níveis de hidrólise, baixo (H1), médio (H2) e alto (H3), e um número de resultados significativamente diferentes, foi alcançado.

Os valores de glicose e galactose não apresentaram diferença significativa entre as cinco repetições e entre os três níveis de hidrólise, pelo teste F e teste *t* ( $\alpha=5\%$ ), utilizando *software* estatístico Stata® 14 (StataCorp, USA). Os valores de lactose foram diferentes significativamente ( $p<0,05$ ) da glicose e galactose, pelos mesmos testes. Os quatro tratamentos (controle, H1, H2 e H3) levaram a diferenças significativas ( $p<0,05$ ) entre si, para cada açúcar.

Tabela 4.6- Concentração de lactose, glicose e galactose em leite pasteurizado, antes (C) e após 3 protocolos de hidrólise enzimática da lactose (H1, H2, H3), em 5 repetições (R1 a R5), quantificados por HPLC-RID

	Lactose <sup>a</sup> (g/100mL)	Glicose <sup>b</sup> (g/100mL)	Galactose <sup>b</sup> (g/100mL)	Hidrólise da lactose(%)*
R1 C	5,16	0,00	0,00	-
R1 H1	3,79	0,46	0,55	26,55
R1 H2	2,62	0,96	0,92	49,22
R1 H3	0,89	1,92	1,80	82,75
R2 C	4,91	0,00	0,00	-
R2 H1	3,84	0,40	0,51	21,79
R2 H2	2,60	1,00	0,98	47,05
R2 H3	0,97	1,83	1,73	80,24
R3 C	4,71	0,00	0,00	-
R3 H1	3,32	0,26	0,41	29,51
R3 H2	2,30	1,08	1,07	51,17
R3 H3	0,80	1,80	1,73	83,01
R4 C	4,64	0,00	0,00	-
R4 H1	3,52	0,34	0,49	24,14
R4 H2	2,38	0,77	0,75	48,71
R4 H3	0,95	1,75	1,67	79,52
R5 C	4,91	0,00	0,00	-
R5 H1	3,60	0,43	0,54	26,68
R5 H2	2,59	0,76	0,78	47,25
R5 H3	0,91	2,04	1,90	81,47

Fonte: dados da tese (2021)

Notas: H1: 10 µL/100mL de leite; H2: 20 µL/100mL de leite; H3: 80 µL/100mL de leite. Temperatura da reação: 7°C. Tempo de hidrólise: 24 horas. Enzima: β-galactosidase de *Kluyveromyces lactis*, ≥ 2600 units/g (Sigma-Aldrich, USA, CAS 9031-11-2). C: controle (antes da hidrólise).

(\*) %hidrólise = 1- (concentração de lactose final/concentração de lactose inicial)

a,b: letras diferentes, entre colunas, diferem entre si pelo teste t, a 5% de probabilidade.

Este método HPLC-RID utilizado para separar, identificar e quantificar lactose, glicose e galactose nas amostras de leite hidrolisado, utilizando água ultrapurificada tanto como fase móvel como para extração das amostras, demonstrou ser capaz de reduzir a utilização de grandes quantidades de reagentes químicos.

A cromatografia é o método eletivo para identificação e quantificação de açúcares em leite. Embora o mais novo método oficializado pela AOAC (ISO 22184/IDF244:2021) não se aplique para leite com lactose reduzida, inferior a 0,1%, a quantificação dos açúcares residuais após hidrólise enzimática da lactose é

realizada por cromatografia, utilizando métodos oficiais modificados com validação intra-laboratorial.

#### 4.6 Resultados da comparação entre métodos

O algoritmo de classificação foi executado nas 20 amostras de leite com hidrólise enzimática da lactose e seus resultados previstos estão demonstrados na Tabela 4.7. Considerando que os resultados da concentração de glicose e galactose não foram significativamente diferentes na análise por HPLC ( $\alpha=5\%$ ), a previsão do algoritmo para um desses açúcares foi considerada correta no nível 3 da hidrólise (H3).

Tabela 4.7- Resultados da previsão do algoritmo de classificação da rede neural convolucional com 4 neurônios na camada de saída

	Lactose (g/100mL)	Glicose (g/100mL)	Galactose (g/100mL)	Açúcar previsto
R1 C	5,16 <sup>b</sup>	0 <sup>a</sup>	0 <sup>a</sup>	lactose
R1 H1	3,79 <sup>b</sup>	0,46 <sup>a</sup>	0,55 <sup>a</sup>	lactose
R1 H2	2,62 <sup>b</sup>	0,96 <sup>a</sup>	0,92 <sup>a</sup>	lactose
R1 H3	0,89 <sup>b</sup>	1,92 <sup>a</sup>	1,80 <sup>a</sup>	galactose
R2 C	4,91 <sup>b</sup>	0,00 <sup>a</sup>	0,00 <sup>a</sup>	lactose
R2 H1	3,84 <sup>b</sup>	0,40 <sup>a</sup>	0,51 <sup>a</sup>	lactose
R2 H2	2,60 <sup>b</sup>	1,00 <sup>a</sup>	0,98 <sup>a</sup>	galactose
R2 H3	0,97 <sup>b</sup>	1,83 <sup>a</sup>	1,73 <sup>a</sup>	glicose
R3 C	4,71 <sup>b</sup>	0,00 <sup>a</sup>	0,00 <sup>a</sup>	lactose
R3 H1	3,32 <sup>b</sup>	0,26 <sup>a</sup>	0,41 <sup>a</sup>	lactose
R3 H2	2,30 <sup>b</sup>	1,08 <sup>a</sup>	1,07 <sup>a</sup>	galactose
R3 H3	0,80 <sup>b</sup>	1,80 <sup>a</sup>	1,73 <sup>a</sup>	galactose
R4 C	4,64 <sup>b</sup>	0,00 <sup>a</sup>	0,00 <sup>a</sup>	lactose
R4 H1	3,52 <sup>b</sup>	0,34 <sup>a</sup>	0,49 <sup>a</sup>	lactose
R4 H2	2,38 <sup>b</sup>	0,77 <sup>a</sup>	0,75 <sup>a</sup>	glicose
R4 H3	0,95 <sup>b</sup>	1,75 <sup>a</sup>	1,67 <sup>a</sup>	galactose
R5 C	4,91 <sup>b</sup>	0,00 <sup>a</sup>	0,00 <sup>a</sup>	lactose
R5 H1	3,60 <sup>b</sup>	0,43 <sup>a</sup>	0,54 <sup>a</sup>	lactose
R5 H2	2,59 <sup>b</sup>	0,76 <sup>a</sup>	0,78 <sup>a</sup>	galactose
R5 H3	0,91 <sup>b</sup>	2,04 <sup>a</sup>	1,90 <sup>a</sup>	galactose

Fonte: dados da tese (2021)

Nota: letras iguais, entre colunas, não diferem entre si, pelo teste t, a 5% de probabilidade.

Das 20 amostras analisadas, quatro apresentaram resultados diferentes na previsão pelo algoritmo para o açúcar que estava em maior concentração (valor real, quantificado pela cromatografia), ou seja, erro de 20%. Esse achado concorda com a acurácia alcançada no modelo de CNN proposta para classificação com quatro neurônios na camada de saída, que foi de 80%.

A CNN de regressão também foi executada nas 20 amostras de leite que foram analisadas por cromatografia. Os valores previstos pelos algoritmos para quantificar lactose, glicose e galactose foram comparados aos valores “reais” medidos pela cromatografia. Na Tabela 4.8, os resultados dos açúcares foram comparados entre tratamentos (controle e níveis 1, 2 e 3 da hidrólise enzimática da lactose), nas cinco repetições.

Tabela 4.8- Comparação entre valores previstos pelo algoritmo e valores reais pela cromatografia para lactose, glicose e galactose em leite pasteurizado, antes (controle) e depois da hidrólise enzimática da lactose em três níveis (H1, H2 e H3), em cinco repetições

Açúcar (g/100mL)		Média (DP)		
		VR	VP	p-value
Lactose	C	4,86 (0,20)	2,98 (3,04)	0,238
	H1	3,61 (0,21)	2,68 (1,94)	0,345
	H2	2,49 (0,15)	1,12 (1,20)	0,061
	H3	0,90 (0,07)	-0,16 (1,08)	0,105
Glucose	C	0 (0)	-0,45 (0,60)	0,169
	H1	0,38 (0,08)	0,63 (1,07)	0,631
	H2	0,91 (0,14)	2,37 (1,17)	0,048
	H3	1,87 (0,12)	5,46 (1,61)	0,007
Galactose	C	0 (0)	-0,02 (1,24)	0,969
	H1	0,50 (0,05)	0,67 (0,70)	0,610
	H2	0,90 (0,13)	2,28 (1,47)	0,102
	H3	1,77 (0,09)	5,72 (1,36)	0,003

Fonte: dados da tese (2021)

Notas: C – tratamento controle; H1 – hidrólise nível 1; H2 - hidrólise nível 2; H3 - hidrólise nível 3; VP- valor previsto pelo algoritmo ; VR - valor real pela cromatografia; DP – desvio padrão; p-value >0,05 - não há diferença entre as médias.

Os resultados da Tabela 4.8 demonstraram que a predição para quantificação da lactose nos quatro tratamentos de hidrólise não apresentaram diferença estatística pelo teste F e teste  $t$  ( $\alpha=5\%$ ), enquanto as previsões para glicose foram diferentes nos tratamentos 2 e 3, e para galactose as previsões foram diferentes somente no tratamento 3 ( $p<0,05$ ).

Os resultados comparativos entre os valores previstos pelo algoritmo de regressão e os valores reais quantificados pela cromatografia apontam para uma semelhança em 75% das previsões.

O desafio de verificar os algoritmos com novas amostras, desconhecidas, diferentes do conjunto de dados inicial, foi uma maneira de validar o algoritmo, de trazê-lo para “vida real”. Desafio maior devido à presença simultânea dos três açúcares nas mesmas amostras. A detecção simultânea por redes neurais não foi executada em trabalhos anteriores, onde cada substância era adicionada às amostras separadamente, e uma parte do conjunto de dados era submetido à validação do método.

# V

## Conclusões

Técnicas promissoras associando a espectroscopia do infravermelho ao uso da inteligência artificial, por meio de algoritmos, para garantir a qualidade do leite, têm sido executadas nos últimos anos. Neste trabalho, as redes neurais convolucionais utilizando os dados espectrais como informações de entrada foram capazes de identificar a distinção entre os espectros dos sacarídeos presentes no leite, sem interferências da sobreposição de ondas, quando analisado pelo FTIR.

A proposta de submeter amostras de leite com teor reduzido de lactose aos algoritmos de classificação e regressão se mostraram possíveis. Açúcares foram identificados e quantificados quando a rede neural convolucional foi aplicada aos dados espectrais de cada amostra. Os algoritmos de regressão quantificaram os açúcares em amostras de leite com lactose, glicose e galactose presentes simultaneamente na mesma amostra.

A concentração de lactose, glicose e galactose em amostras de leite com hidrólise por  $\beta$ -galactosidase, em condições otimizadas de tempo, temperatura e concentração, foram obtidas por HPLC-RID.

O desenvolvimento de técnicas cromatográficas com menor utilização de reagentes químicos é uma preocupação dos dias atuais. O método realizado neste trabalho mostrou a possibilidade de executar a extração da amostra de leite somente com água, assim como a utilização de água ultrapurificada como fase móvel, para detecção e quantificação de açúcares em leite. A redução no uso de reagentes químicos vai de encontro aos preceitos da Química Verde, uma preocupação ambiental fundamentada por princípios que norteiam a pesquisa.

A comparação dos resultados do HPLC-RID com os resultados dos algoritmos de classificação e regressão indicaram uma capacidade preditiva de 80% para classificação e 75% para quantificação dos açúcares nas amostras de leite com hidrólise enzimática da lactose.

### **Considerações Finais**

A procura por produtos com teor reduzido de lactose é uma realidade crescente, que atende às necessidades fisiológicas dos indivíduos com hipolactasia e sintomatologia de intolerância à lactose, assim como agrega valor aos produtos lácteos e promove o crescimento do setor. O comércio de leite fluido, vem recuando ao longo dos últimos anos, no Brasil e em outros países, em troca de produtos com propostas inovadoras, como o leite sem lactose, que promove saúde e bem-estar intestinal para quem precisa.

Os métodos que existem para quantificar o teor de lactose e outros açúcares em leite são de alto custo e laboriosos. As indústrias necessitam de métodos rápidos, baratos e eficientes para quantificar o teor dos açúcares nos lácteos a fim de atender à legislação e assegurar o controle de qualidade dos produtos. Métodos de cromatografia líquida são os que conseguem fazer a distinção entre os açúcares presentes no leite e derivados, porém incorrem nas desvantagens apontadas. Ainda não há método oficial, aprovado pela AOAC, que quantifique açúcares residuais em leite com lactose reduzida abaixo de 0,1%.

Há uma grande variedade de modelos de redes neurais artificiais propostas para análises de leite e também uma diversidade na demonstração de métricas de desempenho. A não uniformidade dos modelos traz resultados muito distintos para problemas com certa semelhança, o que justifica o contínuo aprimoramento das construções de algoritmos para resolverem problemas de autenticidade, qualidade e composição do leite, quando aliados às técnicas de espectroscopia por infravermelho. A engenharia de arquitetura das redes neurais é um campo vasto e promissor para utilização nas análises de leite, como vêm demonstrando os trabalhos realizados até hoje.

# VI

## Referências

ABADI, M. *et al.* TensorFlow: A System for Large-scale Machine Learning. In: Google Brain. arXiv:1605.08695v2, 2016.

ABNT - Associação Brasileira de Normas Técnicas. ISO - *International Standard Organization*. ABNT NBR ISO/IEC 17025. Requisitos gerais para a competência de laboratório de ensaio e calibração. Rio de Janeiro: ABNT, 2017.

ACQUARO JR, V.R. *et al.* Desenvolvimento e validação de método para extração e quantificação através de HPLC com índice de refração para lactose em leite pasteurizado. *Scientia Chromatographica*, v.5, n.2, p.137-145, 2013.

ALPAYDIN, E. *Introduction to Machine Learning*. 4ª ed. Cambridge: MIT Press. 2020.

AOAC *Official Methods of AOAC INTERNATIONAL* (2005) 18ª ed., Gaithersburg, MD, USA, Official Method 972.16.

AOAC *Official Methods of AOAC INTERNATIONAL* (2005) 18ª ed., Gaithersburg, MD, USA, Official Method 980.13.

ARAÚJO, E.C.S.; FARIAS, A.C.M.; COSTA ONE, G.M. Intolerância a lactose e alergia à proteína do leite: o papel do nutricionista. *Revista Campo do Saber*, v.3, n.3, p.10, 2017.

ASSEISS NETO, H. Metodologia de aprendizado autoML baseado em informações de complexidade de instâncias. 2020. Tese (Doutorado em Ciências da Computação), Universidade Federal de Minas Gerais, Minas Gerais, 2020.

ASSEISS NETO, H. *et al.* On the utilization of deep and ensemble learning to detect milk adulteration. *Bio Data Mining*, n.13, 2019.

BAPTISTELLA, M. *O uso de redes neurais e regressão linear múltipla na engenharia de avaliações: determinação dos valores venais de imóveis urbanos*. 2005. Dissertação (Mestrado em Métodos Numéricos em Engenharia), Universidade Federal do Paraná, Paraná, 2005.

BAKER, M.J.; TREVISAN, J.; BASSAN, P. *et al.* Using Fourier Transform IR spectroscopy to analyze biological materials. *Nature protocols*. v.9, n.8, p.1771-1791, 2014.

BOTELHO, B. G. *et al.* 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.

BRANCO, M.S.C. *et al.* Classification of lactose intolerance: An overview of causes and treatments. *Revista Ciências Médicas*, v.26, n.3, p.117-125, 2017.

BRASIL. Ministério da agricultura Pecuária e Abastecimento. Manual de métodos oficiais para análises de alimentos de origem animal. Brasília. 2018c.

BRASIL. Ministério da Agricultura, Pecuária e Abastecimento (MAPA). Instrução Normativa Nº30, de 26 de junho de 2018. Ficam estabelecidos como oficiais os métodos constantes do Manual de Métodos Oficiais para Análise de Alimentos de Origem Animal. *Diário Oficial da União*, Brasília, 13 de jul. de 2018a.

BRASIL. Ministério da Agricultura, Pecuária e Abastecimento (MAPA). Instrução Normativa Nº76, de 26 de novembro de 2018. Ficam aprovados os Regulamentos Técnicos que fixam a identidade e as características de qualidade que devem apresentar o leite cru refrigerado, o leite pasteurizado e o leite pasteurizado tipo A. *Diário Oficial d União*, Brasília, 30 de nov. de 2018b.

BRASIL. Ministério da Agricultura, Pecuária e abastecimento. Instrução Normativa Nº68, de 12 de dezembro de 2006. Oficializa os métodos analíticos oficiais físico-químicos, para controle de leite e produtos lácteos, em conformidade com o anexo desta Instrução Normativa, determinando que sejam utilizados nos Laboratórios Nacionais Agropecuários. *Diário Oficial da União*, Brasília, 14 de dez. de 2006.

BRASIL. Ministério da Saúde. Portaria Nº29, de 13 de janeiro de 1998. Aprova o Regulamento Técnico referente a Alimentos para Fins Especiais. *Diário Oficial da União*, Brasília, 30 de mar. de 1998.

BRASIL. Ministério da Saúde. RDC Nº53, de 07 de outubro de 2014. Dispõe sobre a lista de enzimas, aditivos alimentares e veículos autorizados em preparações enzimáticas para uso na produção de alimentos em geral. *Diário Oficial da União*, Brasília, 08 de out. de 2014.

BRASIL. Ministério da Saúde. RDC Nº135, de 08 de fevereiro de 2017. Altera a Portaria SVS/MS Nº 29, de 13 de janeiro de 1998, que aprova o regulamento técnico referente a alimentos para fins especiais, para dispor sobre os alimentos para dietas com restrição de lactose. *Diário Oficial da União*, Brasília, 09 de fev. de 2017.

BRASIL. Ministério da Saúde. RDC Nº136, de 8 de fevereiro de 2017. Estabelece os requisitos para declaração obrigatória da presença de lactose nos rótulos dos alimentos. *Diário Oficial da União*, Brasília, 09 de fev. de 2017b.

BRASIL. Ministério da Saúde. RDC Nº 166, DE 24 de julho de 2017. Dispõe sobre a validação de métodos analíticos e dá outras providências. *Diário Oficial da União*, Brasília, 25 de jul. de 2017c.

BRITO, A.B.N. *Estudo da cristalização de lactose em diferentes solventes*. 2007. Tese (Doutorado em Engenharia Química), Universidade Federal de São Carlos, São Paulo, 2007.

BROWNLEE, J. Difference between classification and regression in Machine Learning. *Machine Learning Mastery*. 2017. Disponível em: <https://machinelearningmastery.com/classification-versus-regression-in-machine-learning/>. Acesso em: 25 de jan. 2021.

BRUNT, K. *et al.* Results Multi-Laboratory Trial ISO/CD 22184 – IDF/WD 244: Milk and milk products – Determination of the sugar contents – High performance anion exchange chromatography method with pulsed amperometric detection (HPAEC-PAD). *Journal of AOAC International*. 2021.

CAMPBELL, M.K. *Bioquímica*. 3ªed. Porto Alegre: Artmed Editora, 2000.

CARDELLE-COBAS, A. *et al.* Galactooligosaccharides derived from lactose and lactulose: Influence of structure on *Lactobacillus*, *Streptococcus* and *Bifidobacterium* growth. *International Journal of Food Microbiology*, n.149, p.81-87, 2011.

CHÁVEZ-SERVÍN, J.L. *et al.* Analysis of mono- and disaccharides in milk-based formulae by high-performance liquid chromatography with refractive index detection. *Journal of Chromatography A*, n.1043, p.211-215, 2004.

CHURAKOVA E. *et al.* Accurate analysis of residual lactose in low-lactose milk: Comparing a variety of analytical techniques. *International Dairy Journal*, n.96 p.126-131, 2019.

CIENFUEGOS, F.; VAITSMAN, D. *Análise Instrumental*. 1ªed. Rio de Janeiro: Interciência, 2000.

CHOLLET, F. Introduction to keras. March 9th, Latvijas Universitāte, 2018.

COLLINS, C.H.; BRAGA, G.; BONATO, P.B. *Introdução a métodos cromatográficos*. 7ªed. Campinas: Editora Unicamp, 1997.

CONCEIÇÃO, D. G. *et al.* 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.

CONDÉ, V.A.; VALENTE, G.F.S.; MINIGHIN, E.C. Milk fraud by the addition of whey using an artificial neural network. *Ciência Rural*, v.50, n.7, e20190312, 2020.

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.

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

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

EMBRAPA. Anuário Leite 2019. Disponível em: <https://ainfo.cnptia.embrapa.br/digital/bitstream/item/198698/1/Anuario-LEITE-2019.pdf>. Acesso em: 08 jan. 2021.

EMBRAPA. Anuário Leite 2020. Disponível em: <https://www.embrapa.br/en/busca-de-publicacoes/-/publicacao/1124722/anuario-leite-2020-leite-de-vacas-felizes>. Acesso em: 30 maio 2021.

ELSKEN, T.; METZEN, J.H.; HUTTER, F. Neural architecture search: A survey. *Journal of Machine Learning Research*, v.20, n.55, p.1-21, 2019.

ERICH, S.; ANZMANN, T.; FISCHER, L. Quantification of lactose using íon-pair RP-HPLC during enzymatic lactose hydrolysis of skim milk. *Food Chemistry*., n.135, p.2393-2396, 2012.

FACELI, K. *et al.* Inteligência artificial: Uma abordagem de aprendizado de máquina. LTC: Rio de Janeiro, 2011.

FAEDO, R. *et al.* Obtenção de leite com baixo teor de lactose por processos de separação por membranas associados à hidrólise enzimática. *Revista CIATEC – UPF*, v.3, n.1, p.44-54, 2013.

FAGUNDES-NETO, U. Intolerância à lactose: Mitos e Realidade (1). 2009. Disponível em: <http://gastropedinutri.blogspot.com.br/2009/05/intolerancia-lactose-mitos-e-realidade.html>. Acesso em: 22 abr. 2017.

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.

FERRÃO, M.F. *et al.* Determinação simultânea dos teores de cinza e proteína em farinha de trigo empregando NIR-PLS e DRIFT-PLS. *Ciências Tecnologia de Alimentos*. Campinas, v.24, n.3, p.333-340, 2004.

FERREIRA, S. *et al.* Alergia às proteínas do leite de vaca com manifestações gastrointestinais. *NASCER E CRESCER revista de pediatria do centro hospitalar do Porto*, v.23, n.2, 2014.

FISCHER, J. *Hidrólise da Lactose por  $\beta$ -galactosidase de Aspergillus oryzae Imobilizada em Reator de Leito Fixo*. 2010. Dissertação (Mestrado em Engenharia Química), Universidade Federal de Uberlândia, Minas Gerais, 2010.

FOSS. Foss in raw milk testing. Analytics beyond measure. Out. 2018. Disponível em: [https://www.fossanalytics.com/pt-br/products/bactoscan#Papers\\_and\\_Brochures\\_BrochureSpot](https://www.fossanalytics.com/pt-br/products/bactoscan#Papers_and_Brochures_BrochureSpot). Acesso em: 27 dez. 2019.

FOX, P.F. *et al.* *Dairy Chemistry and Biochemistry*. 2ªed. Springer, 2015.

FRANCISQUINI, J.A. *et al.* Reação de Maillard: uma revisão. *Revista do Instituto de Laticínio Cândido Tostes*, Juiz de Fora, v.72, n.1, p.48-57, 2017.

FUNED - FUNDAÇÃO EZEQUIEL DIAS. Unidade de Gestão do Sistema da Qualidade. Elaboração de Procedimento Operacional Padrão – GA / MET (UGSQ-GA-0001). Revisão 06. Belo Horizonte, MG. 2018.

FUNED - FUNDAÇÃO EZEQUIEL DIAS. Diretoria do Instituto Otávio Magalhães. Determinação de açúcares em alimentos contendo polióis (DIOM-DIVISA-SQE-LAM-MET-0017). Revisão 02. Belo Horizonte, MG. 2020.

FURTADO, M.M. Fermentações nos queijos: Interações com a lactose. Disponível em: <http://www.abiq.com.br/nutricao/4%20-%20Fermenta%C3%A7%C3%A3o%20no%20Queijo%20e%20em%20Prod.%20Lacteos%20-%20Mucio%20M.%20Furtado.pdf>. Acesso em: 16 mai. 2018.

FUSCH, G. *et al.* Quantification of lactose content in human and cow's milk using UPLC–tandem mass spectrometry. *Journal of Chromatography B*, v.879 , p.3759-3762, 2011.

GAMBELLI, L. Milk and Its Sugar-Lactose: A Picture of Evaluation Methodologies. *Beverages*, v.3, n.35, 2017.

GARBALLO-RUBIO, A. *et al.* Determination of residual lactose in lactose-free cow milk by hydrophilic interaction liquid chromatography (HILIC) coupled to tandem mass spectrometry. *Journal of Food Composition and Analysis*, v.66, p.39-45, 2018.

GDCh – Gesellschaft Deutscher Chemiker - Positionspapier der Lebensmittelchemischen Gesellschaft zu den Angaben "laktosefrei" und "laktosearm". *Lebensmittelchemie*, v.59, n.45, 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.

HANSEN, P.W.; HOLROYD, S.E. Development and application of Fourier transform infrared spectroscopy for detection of milk adulteration in practice. *International Journal of Dairy Technology*, v.72, n.3, 2019.

HUTTER, F.; KOTTHOFF, L.; VANSCHOREN, J. Automated Machine Learning. 1ªed. Springer: Cham, 2019.

IDF – Lactose, an important nutrient: Advocating a revised policy approach for dairy and its intrinsic sugar. *Bulletin of the International Dairy Federation*. IDF: Brussels, Belgium. 2021.

INMETRO - Instituto Nacional de Metrologia, Normalização e Qualidade Industrial. DOQ-CGCRE-008. *Orientação sobre validação de métodos analíticos*. INMETRO, Revisão 09. JUN/2020.

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.

ISO 22662/IDF 198:2007. Milk and milk products - Determination of lactose content by high-performance liquid chromatography (Reference method). *International Organization for Standardization*: Geneva, Switzerland. 2007.

ISO 9622/IDF 141:2013. Milk and liquid milk products – Guidelines for the application of mid-infrared spectrometry. *International Organization for Standardization*: Geneva, Switzerland. 2013.

ISO 22184/IDF 244:2021. Milk and milk products — Determination of the sugar contents — High performance anion exchange chromatography with pulsed amperometric detection method (HPAEC-PAD). Geneva, Switzerland. 2021.

JAMES, G. *et al.* An Introduction to Statistical Learning. 8ª ed. New York: Springer. 2017.

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 and Technology*, n.46, v.1, p.27–48, 2015.

KAROUI, R.; DOWNEY, G.; BLECKER, C. Mid-Infrared spectroscopy coupled with chemometrics: a tool for the analysis of intact food and the exploration of their

molecular structure – quality relationships – a review. *Chemical Reviews*, n.110, p. 6144-6168, 2010.

KLEIN, M.P. *Hidrólise da lactose e síntese de galactooligossacarídeos utilizando  $\beta$ -galactosidase imobilizada em suportes à base de quitosana*. 2014. Tese (Doutorado em Engenharia de Alimentos). Universidade Federal de Santa Catarina, Florianópolis, Santa Catarina, 2014.

KOUAOUCI, R. Analytical methods for lactose quantification in milk and milk products. IDF International Symposium – Lactose and its Derivatives. Valacta - Quebec, Canadá. 2007.

KUBAT, M.. An introduction to machine learning. 2<sup>a</sup>ed. Switzerland: Springer. 2017.

KUHN, M.; JOHNSON, K. *Applied Predictive Modeling*. 2<sup>a</sup>ed. New York: Springer. 2018.

LACTOSCOPE FTA. The determination of residual lactose in lactose-reduced milk by Lactoscope FTIR advanced – calibration, validation and application to milk and milk products. *Residual lactose by Lactoscope FTA in lactose reduced milk*. Version 1.3. p.1-10. 2010.

LA ROCA, M.F. *et al.* Desenvolvimento e validação de método analítico: passo importante na produção de medicamentos. *Revista Brasileira de Farmácia*, v.88, n.4, p.177-180, 2007.

LISBOA, C.R. *et al.* Síntese de galacto-oligossacarídeos a partir de lactose usando  $\beta$ -galactosidase comercial de *Kluyveromyces lactis*. *Brazilian Journal of Food Technology*, Campinas, v.15, n.1, p.30-40. 2012.

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

LIU, W. *et al.* A survey of deep neural network architectures and their applications. *Neurocomputing*, v.234, p.11-26, 2017.

LONGO, G. *Influência da adição de lactase na produção de iogurte*. 2006. Dissertação (Mestrado em Tecnologia de Alimentos) – Setor de Tecnologia, Universidade Federal do Paraná, Curitiba, 2006.

LOPEZ, F. A. Os dilemas da intolerância à lactose. *Ades atualiza*, ano 1, n.2, 2009.

LUYKX, D.M.A.M.; VAN RUTH, S.M. An overview of analytical methods for determining the geographical origin of food products. *Food Chemistry*, n.107, p.897-911, 2008.

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.

MAIMON, O.; ROKACH, L. *Data Mining and Knowledge Discovery Handbook*. Springer: Boston, MA. 2ª ed. 2010.

MATTAR, R.; MAZO, D. F. C. Intolerância à lactose: mudança de paradigmas com a biologia molecular. *Revista da Associação Médica Brasileira*, São Paulo, n.56, v.2, p.230-236. 2010.

MATTAR, R.; MAZO, D.F.C.; CARRILHO, F.J. Lactose intolerance: Diagnosis, genetic, and clinical factors. *Clinical and Experimental Gastroenterology*, v.5, p.113, 2012.

MONTGOMERY, R.K. *et al.* Lactose and Lactase-Who Is Lactose Intolerant and Why? *Journal of Pediatric Gastroenterology and Nutrition*, n.45, p.131-137, 2007.

MORAES, I.A. Fisiologia da glândula mamária. 2016. Disponível em: <http://www.uff.br/fisiovet/lactacao.pdf>. Acesso em: 22 abr. 2017.

MOREIRA, K.M.M. *et al.* Produção de doce de leite com teor reduzido de lactose por  $\beta$ -galactosidase. *Revista Acadêmica Ciências Agrárias e Ambientais*. Curitiba, v.7, n.4, p.375-382, 2009.

MORGANO, M. A.; MORIYA, C.; FERREIRA, M. M. C. Determinação quantitativa do teor de sacarose em açúcar cristal por espectroscopia FTIR/ATR e regressão multivariada. *Brazilian Journal of Food Technology*, v.6, n.1, p.77-83, 2003.

NATHAN, J.P.; SCHILIT, S.; ROSENBERG, J.M. Can lactose-intolerant patients use lactose-containing medications? *Drug Topics*. 2008. Disponível em: <https://www.drugtopics.com/view/clinical-q-can-lactose-intolerant-patients-use-lactose-containing-medications>. Acesso em: 28 mar. 2017.

NICOLAU, N.; XU, Y.; GOODACRE, R. Fourier transform infrared spectroscopy and multivariate analysis for the detection and quantification of different milk species. *Journal of Dairy Science*, v.93, n.12, p.5651-5660, 2010.

NIJPELS, H.H. *et al.* Application of cryoscopy for the measurement of enzymatic hydrolysis of lactose. *Journal Food Science*, v.45, p.1864-1687, 1980.

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.

OLIVEIRA, R.F. Inteligência Artificial. Editora: Londrina, PR, 2018.

PAPPAS, C.S. *et al.* Identification and differentiation of goat and sheep milk based on diffuse reflectance infrared Fourier transform spectroscopy (DRIFTS) using cluster analysis. *Food Chemistry*, v.106, n.3, p.1271-1277, 2008.

PARDO, R.B. *et al.* Intolerance, allergy and the ambiguous concepts of milk and dairy products technology: does the consumer know how to choose the milk wants to drink? *Brazilian Journal of Animal and Environmental Research*, v.4, n.1, p.684-695, 2021.

PEDREGOSA, F. *et al.* Scikit-learn: Machine Learning in Python. *Journal of Machine Learning Research*, n.12, p.2825-2830, 2011.

PEREIRA FILHO, D.; FURLAN, S.A. Prevalência da intolerância à lactose em função da faixa etária e do sexo: experiência do laboratório Dona Francisca, Joinville, SC. *Revista Saúde e Meio Ambiente*, v.5, n.1, p.24-30, 2004.

PEREIRA, M.C.S. *et al.* Lácteos com baixo teor de lactose: uma necessidade para portadores de má digestão da lactose e um nicho de mercado. *Revista do Instituto de Laticínios Cândido Tostes*, v.67, n.389, p.57-65, 2012.

PETIBOIS, C. *et al.* Determination of glucose in dried serum samples by Fourier-Transform Infrared spectroscopy. *Clinical Chemistry*, v.45, n.9, p.1530-1535, 1999.

PRETTO, F.M. *et al.* Má absorção de lactose em crianças e adolescentes: diagnóstico através do teste do hidrogênio expirado com o leite de vaca como substrato. *Journal of Pediatric*, v.78, n.3, p.213-218, 2002.

REEVES, J.B.; ZAPF, C.M. Mid-Infrared Diffuse Reflectance Spectroscopy for Discriminant Analysis of Food Ingredients. *Journal of Agricultural Food Chemistry*, v.46, p.3614-3622, 1998.

RIDEFELT, P.; HÁKANSSON, L.D. Lactose intolerance: lactose tolerance test versus genotyping. *Scandinavian Journal of Gastroenterology*, v.40, n.7, p.822-826, 2005.

RIBEIRO, D.C.S.Z. *et al.* Adulterants interference on Fourier Transform Infrared analysis of raw milk. *Arquivo Brasileiro de Medicina Veterinária e Zootecnia.*, v.70, n.5, p.1649-1654, 2018.

RODRIGUEZ-OTERO, J.L.; HERMIDA, M.; CENTENO, J. Analysis of dairy products by near-infrared spectroscopy: a review. *Journal Agricultural Food Chemical*, n.45, v.8, p.2815-2819, 1997.

SALIBA, E.O.S. *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.

SAMPAIO, I. B. M. Estatística aplicada à experimentação animal. 2ªed. Belo Horizonte: FEP-MVZ. , 2002.

SANDERS, P. *et al.* The determination of sugars in dairy products: development of a new standard method of the International Dairy Federation and the Internal Organization of Standardization. *Journal of AOAC International*, n.100, v.5, p.1577-1581, 2017.

SANTIAGO, P. *et al.* Estudo da produção de  $\beta$ -galactosidase por fermentação de soro de queijo com *Kluyveromyces marxianus*. *Ciência e Tecnologia de Alimentos.*, v. 24, n. 4, p.567-572, 2004.

SANTOS, A.A. *et al.* Dosagem de açúcares redutores com o reativo DNS em microplaca. *Brazilian Journal of Food Technology*, Campinas, v. 20, p.1-9, 2017.

SCHMIDHUBER, J. Deep learning in neural networks: An overview. *Neural Networks*, v.61, p.85-117, 2015.

SEXTON, D.L. *Determination of Lactose by Reversed-Phase High Performance Liquid Chromatography*. 2004. Tese (Master of Science in Chemistry), East Tennessee State University. Tennessee – USA, 2004.

SILVA, K *et al.* Lactose determination in relation to contemporary methodologies. *Revista do Instituto de Laticínios Cândido Tostes*, Juiz de Fora, v.75, n.1, p. 59-71, 2020.

SILVA, L.C.C. *Capacidade de detecção de adulterações e suficiência das provas oficiais para assegurar a qualidade do leite pasteurizado*. 2013. Tese (Doutorado em Ciência Animal). Universidade Estadual de Londrina. Londrina, Paraná, 2013.

SILVA, L.K.R. *et al.* Spectroscopic method (FTIR-ATR) and chemometric tools to detect cow's milk addition to buffalo's milk. *Revista Mexicana de Ingeniería Química*, v.19, n.1, p.11-20, 2020.

SILVA, N.F.; FILHA, A.A.D.A.L.; ANDRADE, A.C.M.; DE SOUZA, J.D.; DE SÁ, M.D.S.V.S. Aspectos imunológicos da alergia à proteína do leite de vaca mediada por IgE. *Revista Multidisciplinar em Saúde*, v.2, n.2, p.41, 2021.

SIMONYAN, K.; VEDALDI, A.; ZISSERMAN, A. Deep inside Convolutional Networks: Visualizing Image Classification Models and Saliency Maps. *Cornell University*. Computer vision and pattern recognition: arXiv- prints, p.arXiv:1312.6034v2, 2014.

SPARVOLI, A.C. *Malabsorção de lactose do adulto em uma população nordestina*. 1989. Dissertação (Mestrado em Ciências Médicas). Universidade Estadual de Campinas. Campinas, São Paulo, 1989.

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.

SU, W-H.; ARVANITOYANNIS, I.S.; SUN, D-W. Trends in Food Authentication. *In: Modern Techniques for Food Authentication*. Cambridge: Academic Press. 1ªed., chapter 18, p.731-758, 2018.

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*. 2019. Tese (Doutorado em Ciência Animal). Universidade Federal de Minas Gerais, Minas Gerais, 2019.

TALAMONA, A. *Laboratory Chromatography Guide*. Flawil, Switzerland: BUCHI Labortechnik AG. 2005.

THOMPSON, M.; ELLISON, S. L. R.; WOOD, R. Harmonized guidelines for single-laboratory validation of methods of analysis. *Pure and Applied Chemistry*, v.74, p.835-855, 2002.

TREVISAN, A.P. *Influência de diferentes concentrações de enzimas lactase e temperaturas sobre a hidrólise da lactose em leite pasteurizado*. 2008. Dissertação (Mestrado em Ciência e Tecnologia de Alimentos). Universidade Federal de Santa Maria, Rio Grande do Sul.

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 do Instituto de Laticínios Cândido Tostes*, v.69, n.6, p.425-432, 2014.

WENG, S. *et al.* Feasibility of authenticating mutton geographical origin and breed via hyperspectral imaging with effective variables of multiple features. *Food Analytical Methods*, v.14, p.834-844, 2021.

WILLMOTT, C. J. On the validation of models. *Physical Geographic*, v.2, 184-194, 1981.

WITTEN, I. *et al.* Data Mining: Practical machine learning tools and techniques. Cambridge: Morgan Kaufmann, 4ªed, 2016.

WORTMANN, A.C. *Análise molecular dos polimorfismos associados à hipolactasia primária do tipo adulto em dispépticos funcionais e controles assintomáticos*. Tese

(Doutorado em Ciências em Gastroenterologia e Hepatologia). Universidade Federal do Rio Grande do Sul, Porto Alegre, Rio Grande do Sul, 2014.

XINMIN, W. *et al.* Determination of glucosamine and lactose in milk-based formulae by high-performance liquid chromatography. *Journal of Food Composition and Analysis*. v.21, p.255-258, 2008.

ZYCHAR, B.C.; OLIVEIRA, B.A. Fatores desencadeantes da intolerância á lactose: metabolismo enzimático, diagnóstico e tratamento. *Atas de Ciências da Saúde*, São Paulo, v.5, n.1, p.35-46, jan.-mar. 2017.

# VII

## **Colaborações interinstitucionais para elaboração do projeto**

- Departamento de Tecnologia e Inspeção de Produtos de Origem Animal/ EV-UFMG

- Prof<sup>a</sup> Débora Cristina Sampaio

- Departamento de Medicina Veterinária Preventiva/ EV-UFMG

- Prof<sup>a</sup> Kelly Moura Keller

- 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)

# VIII

## Contribuições Acadêmicas

Artigo 1: publicado em co-autoria

ASSEISS NETO, H., Wanessa L.F. Tavares, Daniela C.S.Z. Ribeiro, Ronnie C.O. Alves, Leorges M. Fonseca and Sérgio V.A. Campos. On the utilization of deep and ensemble learning to detect milk adulteration. *Bio Data Mining*, n. 13. 2019.

Artigo 2: submetido ao periódico *Food Analytical Methods*

Determination of the lactose content in low-lactose milk by Spectroscopy FTIR, HPLC-RID and RNA.

Artigo 3: elaborado para submissão ao periódico *Journal Dairy Science*

Identification of fraud with addition of sugars in milk by Spectroscopy FTIR, Artificial Neural Network and KNN.

Resumo 1: Simultaneous analysis of three adulterants in raw milk using Fourier-transform infrared spectroscopy. D. C. S. Z. Ribeiro; W. L. F. Tavares ; J. S. Lima; M. O. Leite; M. M. O. P. Cerqueira; L. F. Ferreira ; J. P. A. Haddad; J. P. M. Heringer and L. M. Fonseca. Veterinary School, Universidade Federal de Minas Gerais, Belo Horizonte, MG, Brazil. *Journal of Dairy Science*, v.101, Suppl.2, p.55, 2018.

Resumo 2: Use of infrared spectroscopy to estimate the lactose content in hydrolyzed milk. D. C. S. Z. Ribeiro; W. L. F. Tavares; J. S. Lima; 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, 2019.

Resumo 3: Detection of adulteration in milk using infrared spectroscopy and machine learning. H. Asseiss Neto; W. L. F.Tavares; D. C. S.Z. Ribeiro; J. S. Lima; 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; Departament of Computer Science, Universidade Federal de Minas Gerais, Belo Horizonte, MG, Brazil. *Journal of Dairy Science*, v.102, Suppl.1, p.38, 2019.

Resumo 4: 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; Departament of Computer Science, Universidade Federal de Minas Gerais, Belo Horizonte, MG, Brazil. *Journal of Dairy Science*, v.102, Suppl.1, p.38-39, 2019.