

UNIVERSIDADE FEDERAL DE MINAS GERAIS
ESCOLA DE VETERINÁRIA
DEPARTAMENTO DE TECNOLOGIA E INSPEÇÃO DE
PRODUTOS DE ORIGEM ANIMAL

Claudia Ferreira Viana

**Deteção de adulteração por adição de leite bovino ao leite bubalino utilizando
redes neurais artificiais e outras técnicas de mineração de dados**

Belo Horizonte

2022

Claudia Ferreira Viana

Detecção de adulteração por adição de leite bovino ao leite bubalino utilizando redes neurais artificiais e outras técnicas de mineração de dados

Tese apresentada ao Colegiado de Pós-Graduação em Ciência Animal da Escola de Veterinária da Universidade Federal de Minas Gerais –UFMG como requisito parcial à obtenção do grau de Doutora em Ciência Animal

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

Prof. Orientador: Leorges Moraes da Fonseca

Prof. Co-orientadoras: Mônica de Oliveira Leite e Elisa Helena Paz Andrade

Belo Horizonte

2022

V614d

Viana, Claudia Ferreira ,1992 -

Detecção de adulteração por adição de leite bovino ao leite bubalino utilizando redes neurais artificiais e outras técnicas de mineração de dados/ Claudia Ferreira Viana. – 2022.

149 f.il

Orientador: Leorges Moraes da Fonseca

Coorientadoras: Mônica de Oliveira Leite

Elisa Helena Paz Andrade

Tese (Doutorado) apresentado à Escola de Veterinária da Universidade Federal de Minas Gerais para obtenção do título de Doutora.

Bibliografias: f. 130 a 149.

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

CLAUDIA FERREIRA VIANA

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

Aprovado(a) em 26 de outubro de 2022, pela banca constituída pelos membros:

Dr.(a). Leorges Moraes da Fonseca - Presidente - Orientador(a)

Dr.(a). Sérgio Vale Aguiar Campos

Dr.(a). Daniela Cristina Solo de Zaldivar Ribeiro

Dr.(a). Débora Cristina Sampaio de Assis

Dr.(a). Bruna Maria Salotti de Souza



Documento assinado eletronicamente por **Leorges Moraes da Fonseca, Professor do Magistério Superior**, em 27/10/2022, às 08:23, 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 **Bruna Maria Salotti de Souza, Professora do Magistério Superior**, em 27/10/2022, às 09:52, conforme horário oficial de Brasília, com fundamento no art. 5º do [Decreto nº 10.543, de 13 de novembro de 2020](#).



Documento assinado eletronicamente por **Debora Cristina Sampaio de Assis, Professora do Magistério Superior**, em 27/10/2022, às 21:35, 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 28/10/2022, às 09:54, 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 **Daniela Cristina Solo de Zaldivar Ribeiro, Usuário Externo**, em 28/10/2022, às 15:48, 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, informando o código verificador **1844161** e o código CRC **430A03E1**.

AGRADECIMENTOS

À Deus, por sempre me proporcionar tantas oportunidades boas e alegrias.

Aos meus pais, Claudio e Sônia, pelo amor e apoio incondicional em todas as etapas da minha vida acadêmica. Se eu cheguei até aqui foi porque vocês sempre acreditaram que eu conseguiria mesmo quando eu pensei em desistir.

Ao meu irmão, Caio, que sempre foi meu parceiro de conversas e desabafos acadêmicos.

Ao Marsel, por todo apoio e incentivo durante as etapas mais difíceis, e por acreditar sempre que eu era capaz de finalizar essa etapa.

Ao meu orientador Leorges, por me acompanhar desde o Mestrado e confiar no meu trabalho.

Prof. Sérgio, por toda colaboração para enriquecimento do trabalho.

Às minhas Co-orientadoras Mônica e Elisa pelas correções e sugestões na melhoria do trabalho.

À Ana Paula, por ter sido o anjo que Deus colocou na minha vida no meu momento de maior dúvida com as redes neurais. Sem sua ajuda não conseguiríamos ter o resultado que tivemos em relação à execução do trabalho.

À Isabella por toda a ajuda e companheirismo durante o experimento no laboratório.

À Sarah pelas incontáveis horas de conversa durante toda a pós-graduação, sempre com uma palavra de incentivo.

A todas as minhas “flats” por todo companheirismo ao longo dos anos.

Aos amigos de turma Ranier e Gustavo pelo apoio durante as aulas.

Aos funcionários do Laboratório de Qualidade do Leite, que sempre foram prestativos e ajudaram na execução dos experimentos.

Aos professores do DTIPOA pelo carinho, suporte e todo aprendizado que foi passado durante toda a pós-graduação.

Às amigas Mariana, Carol, Nathália, Úrsula e Paulinha por todo companheirismo mesmo à distância.

Aos membros da banca por toda contribuição e sugestão para melhorarmos a tese.

À CAPES pela concessão da bolsa (Processo número: 88882.348895/2019-01)

Ao Laboratório de Análise da Qualidade do Leite da Escola de Veterinária da UFMG (Projeto 882, FEPE-UFMG). Aos órgãos financiadores e de apoio ao Laboratório de Análise da Qualidade do Leite da Escola de Veterinária da UFMG (FINEPE, CNPq, FAPEMIG, Ministério da Agricultura Pecuária e Abastecimento).

À FAPEMIG pelo apoio no projeto APQ-02740-17.

RESUMO

A crescente demanda por produtos derivados de leite de búfala no mercado brasileiro nos últimos anos tem incentivado cada vez mais produtores a investir nessa área de produção. O alto valor nutricional do leite bubalino comparado principalmente ao leite bovino tem contribuído para esse aumento de consumo, o que acarreta em dificuldades de aquisição de leite no período de entressafra. A legislação brasileira permite a comercialização de leite com mistura de diferentes espécies, desde que seja devidamente identificado no rótulo da embalagem. Porém, muitas vezes essa prática acontece sem que o consumidor seja informado ou antes mesmo da chegada do leite na indústria. Visando aprimorar a detecção das fraudes, objetivou-se nesse estudo a realização de análise físico-química dos leites bovino e bubalino a partir de espectroscopia no infravermelho a fim de se realizar a comparação entre a composição do leite das duas espécies e o desenvolvimento de modelos de Redes Neurais Artificiais e outras técnicas de mineração de dados para detectar a adição de leite bovino ao leite bubalino. Para a análise físico-química de caracterização do leite bubalino em diferentes estações do ano foram coletadas amostras de tanques de refrigeração durante 24 meses, totalizando 837 amostras. Foram feitas análises dos teores de composição gordura, lactose, proteína, sólidos totais e sólidos não gordurosos, encontrando maiores teores de gordura (6,19%) e proteína (4,25%) nas estações de primavera e verão respectivamente. Em relação aos sólidos totais os maiores valores foram encontrados nas estações de primavera (16,14%) e os sólidos não gordurosos tiveram maiores valores no verão (10,03%), coincidindo com período de chuvas da região. Para os estudos envolvendo a comparação da composição do leite bovino com o leite bubalino e inteligência artificial, foram coletadas 300 amostras de leite bubalino e 300 amostras de leite bovino no período de outubro/2021 a março/2022 e preparadas misturas com nove percentuais de adição de leite bovino ao leite bubalino, simulando adulteração (1%, 2%, 5%, 10%, 20%, 30%, 40%, 50%, 75%), além de leite bubalino sem adição (0%) e leite bovino sem adição (100%). As amostras foram analisadas por infravermelho com transformada de Fourier (FITR) e os resultados obtidos foram tabelados e comparados estatisticamente quanto à composição tendo como resultado uma superioridade do leite bubalino em todos os parâmetros avaliados, além da utilização do banco de dados na montagem das arquiteturas de redes neurais juntamente com dados do histórico da rotina do laboratório. Foram testadas *Redes Multilayer Perceptron* de uma e duas camadas ocultas e Rede Neural de Função de Base Radial através do software IBM SPSS Statistics®. O mesmo banco de dados foi empregado no teste das técnicas de

mineração de dados utilizando a plataforma RapidMiner®, com os testes *Naive Bayes*, *Generalized Linear Model*, *Logistic Regression*, *Fast Large Margin*, *Deep Learning*, *Decision Trees*, *Random Forest*, *Gradient Boosted Trees* e *Support Vector Machine*. As redes *Multilayer Perceptron* obtiveram resultados de classificação de 97,4% e 97,0% para uma e duas camadas respectivamente, enquanto as de função de base radial obtiveram 97,3% de acerto. Os modelos de mineração de dados com melhores resultados foram *Random Forest*, *Support Vector Machine* e *Decision Trees*. Os resultados demonstraram boa acurácia, sensibilidade, especificidade e precisão tanto para os testes de Redes Neurais quanto para os outros testes de mineração de dados, evidenciando que são boas ferramentas de análise e predição de fraudes a serem usados como triagem pela indústria.

Palavras chave: Fraude, Leite de búfala, Rapidminer, Redes Neurais

ABSTRACT

The growing demand for buffalo milk products in the Brazilian market in recent years has encouraged more producers to invest in this production area. The high nutritional value of buffalo milk compared to bovine milk has contributed to this increase in consumption, which leads to difficulties in acquiring milk during the off-season. Brazilian legislation allows the marketing of milk with a mixture of different species, as long as it is properly identified on the package label. However, this practice often happens without consumers awareness, or before the milk even reaches the industry. In order to improve fraud detection, this study aimed to perform physical-chemical analysis of cow and buffalo milk using infrared spectroscopy to compare the milk composition the two species and to develop models of Artificial Neural Networks and other data mining techniques from compositional data to detect the addition of cow milk to buffalo milk. For the physical-chemical analysis of buffalo, samples were collected from refrigeration tanks over 24 months, totalling 837 samples. Analyses were performed for the composition, fat, lactose, protein, total solids, and solids non-fat contents. Higher fat contents (6.19%) and protein (4.25%) were find in the spring and summer seasons, respectively. Regarding total solids, the highest values were found in the spring seasons (16.14%) while solids non-fat had higher values in the summer (10.03%), which is the rainy period of the region. For the studies involving the comparison of bovine milk composition with buffalo milk and artificial intelligence, 300 samples of buffalo milk samples and 300 samples of cow milk were collected during the months of October/2021 to March/2022, and mixtures with nine levels of bovine milk addition to buffalo milk were prepared, simulating adulteration (1%, 2%, 5%, 10%, 20%, 30%, 40%, 50%, 75%) and two levels without mixing (0% - buffalo milk without addition and 100% - cow milk without addition). The samples were analyzed by FITR and the results obtained were tabulated and statistically compared with higher components concentration for buffalo milk in all evaluated parameters, these results were used to set up the neural network architectures along with laboratory routine data. Multilayer Perceptron Networks with one and two hidden layers and Radial Base Function Neural Networks were tested using IBM SPSS Statistics® software. The same database was employed to test data mining techniques using the Rapidminer® software. Naive Bayes, Generalized Linear Model, Logistic Regression, Fast Large Margin, Deep Learning, Decision Trees, Random Forest, Gradient Boosted Trees and Support Vector Machine tests were used. The Multilayer Perceptron networks achieved classification results of 97.4% and 97.0% for one and two layers respectively, while the radial basis function obtained 97.3% of accuracy. The data mining

models with the best results were Random Forest, Support Vector Machine and Decision Trees. The results showed good accuracy, sensitivity, specificity and precision for both the Neural Network tests and the other data mining tests, demonstrating that they are good analysis and adulteration prediction tools to be used as a screening test by the industry.

Keywords: Fraud, Buffalo Milk, Rapidminer, Artificial Neural Networks

Lista de Figuras:

Figura 0-1: Etapas do processo de KDD (Knowledge Discovery in Databases) ou Descoberta de Conhecimento em Banco de Dados, desde as etapas de limpeza e integração até a etapa da avaliação do conhecimento a que se deseja estudar.	33
Figura 0-2: Representação esquemática de um neurônio biológico humano com dendritos, núcleo, soma, axônio e terminais sinápticos.....	36
Figura 0-3: Representação de um neurônio artificial com as entradas, pesos, somador, função de ativação e camada de saída.	37
Figura 0-4: Funções de ativação comumente utilizadas em Redes Neurais Artificiais. (a) Função de Limiar, (b) Função de ativação linear por partes, (c) Função Sigmoide, (d) Função Tangente Hiperbólica, (e) Função Softmax.....	39
Figura 0-5: Arquitetura de uma rede neural de múltiplas camadas com camada de entrada, duas camadas ocultas e uma camada de saída.	41
Figura 0-6: Representação de uma rede de função de bases radiais (RBF) com camada de entrada, camada intermediária e camada de saída.	43
Figura 1-1: Teores de gordura, lactose e proteína de leite de búfala coletado mensalmente de tanques refrigeradores durante o período de julho/2020 a junho/2022 em Minas Gerais. Fonte: A autora (2022).....	55
Figura 1-2: Variação sazonal dos teores de sólidos totais e extrato seco desengordurado do leite de búfalas coletado mensalmente de tanques refrigeradores durante o período de julho/2020 a junho/2022.....	58
Figura 1-3: Variação sazonal das contagens de células somáticas (médias aritméticas mensais e geométricas trimestrais) (mil CS/ml) e das contagens padrão em placas (médias aritméticas mensais e geométricas trimestrais) (mil UFC/ml) de leite de búfala coletadas mensalmente de tanques refrigeradores durante o período de julho/2020 a junho/2022.	60
Figura 2-1: Demonstração da adulteração experimental das amostras de leite de acordo com as porcentagens de leite bovino adicionadas ao leite bubalino, utilizando amostras de leite bovino e bubalino sem adição obtidas de tanques refrigeradores de unidades produtoras do estado de Minas Gerais no período de outubro/2021 a março/2022. A amostra 0% corresponde ao leite bubalino sem adição e a 100% leite bovino sem adição. As demais porcentagens se referem ao percentual de leite bovino na amostra.....	66
Figura 2-2: Composição do banco de dados a partir dos resultados das análises FTIR realizadas em amostras de leite bovino e bubalino obtidas no período de 2020-2022 em tanques refrigeradores de unidades produtoras no estado de Minas Gerais, e mistura de leite bovino em leite bubalino em diferentes porcentagens simulando fraude.....	68
Figura 3-1: Esquema da redução do desbalanceamento das amostras utilizadas para formação do banco de dados utilizando dados obtidos de análises FTIR em leite bovino, leite bubalino e adição de leite bovino ao leite bubalino em diferentes porcentagens simulando fraude, em amostras obtidas no período de 2020-2022 em tanques de refrigeração de unidades produtoras no estado de Minas Gerais.....	81
Figura 3-2: Gráfico Boxplot comparativo dos diferentes tipos de Rede Neural Multilayer Perceptron e Função de Base Radial em relação à acurácia e a diferentes porcentagens de adição de leite bovino em leite bubalino simulando fraude.	92
Figura 3-3: Gráfico Boxplot comparativo dos diferentes tipos de Rede Neural Multilayer Perceptron e Função de Base Radial em relação à precisão e a diferentes porcentagens de adição de leite bovino em leite bubalino simulando fraude.	93

Figura 3-4: Gráfico Boxplot comparativo dos diferentes tipos de Rede Neural Multilayer Perceptron e Função de Base Radial em relação à especificidade e à diferentes porcentagens de adição de leite bovino em leite bubalino simulando fraude.....	94
Figura 3-5: Gráfico Boxplot comparativo dos diferentes tipos de Rede Neural Multilayer Perceptron e Função de Base Radial em relação à sensibilidade e à diferentes porcentagens de adição de leite bovino em leite bubalino simulando fraude.	95
Figura 4-1: Esquema representativo do funcionamento de uma Decision Tree em relação às tomadas de decisão desde o nó raiz até os nós folhas em diferentes tipos de resposta.	102
Figura 4-2: Curva ROC relativa aos testes de mineração de dados feita pela plataforma RapidMiner® em amostras com 1% de adição de leite bovino em leite bubalino simulando fraude.	107
Figura 4-3: Area Under the Curve relativa aos testes de mineração de dados feitos pelo Rapidminer® utilizando dados FTIR relativos à composição do leite em amostras com 1% de adição de leite bovino em leite bubalino simulando fraude. Em ordem: Naive Bayes, Generalized Linear Model, Logistic Regression, Fast Large Margin, Deep Learning, Decision Trees, Random Forest, Gradient Boosted Trees e Support Vector Machine	107
Figura 4-4: Curva ROC relativa aos testes de mineração de dados feita pela plataforma RapidMiner® em amostras com 2% de adição de leite bovino em leite bubalino simulando fraude.	109
Figura 4-5: Area Under the Curve relativa aos testes de mineração de dados feitos pelo Rapidminer® utilizando dados FTIR relativos à composição do leite em amostras com 2% de adição de leite bovino em leite bubalino simulando fraude. Em ordem: Naive Bayes, Generalized Linear Model, Logistic Regression, Fast Large Margin, Deep Learning, Decision Trees, Random Forest, Gradient Boosted Trees e Support Vector Machine	109
Figura 4-6: Curva ROC relativa aos testes de mineração de dados feita pela plataforma RapidMiner® em amostras com 5% de adição de leite bovino em leite bubalino simulando fraude.	111
Figura 4-7: Area Under the Curve relativa aos testes de mineração de dados feitos pelo Rapidminer® utilizando dados FTIR relativos à composição do leite em amostras com 5% de adição de leite bovino em leite bubalino simulando fraude. Em ordem: Naive Bayes, Generalized Linear Model, Logistic Regression, Fast Large Margin, Deep Learning, Decision Trees, Random Forest, Gradient Boosted Trees e Support Vector Machine.	111
Figura 4-8: Curva ROC relativa aos testes de mineração de dados feita pela plataforma RapidMiner® em amostras com 10% de adição de leite bovino em leite bubalino simulando fraude.	113
Figura 4-9 Area Under the Curve relativa aos testes de mineração de dados feitos pelo Rapidminer® utilizando dados FTIR relativos à composição do leite em amostras com 10% de adição de leite bovino em leite bubalino simulando fraude. Em ordem: Naive Bayes, Generalized Linear Model, Logistic Regression, Fast Large Margin, Deep Learning, Decision Trees, Random Forest, Gradient Boosted Trees e Support Vector Machine.	113
Figura 4-10: Curva ROC relativa aos testes de mineração de dados feita pela plataforma RapidMiner® em amostras com 20% de adição de leite bovino em leite bubalino simulando fraude.	115
Figura 4-11: Area Under the Curve relativa aos testes de mineração de dados feitos pelo Rapidminer® utilizando dados FTIR relativos à composição do leite em amostras com 20% de adição de leite bovino em leite bubalino simulando fraude. Em ordem: Naive Bayes,	

Generalized Linear Model, Logistic Regression, Fast Large Margin, Deep Learning, Decision Trees, Random Forest, Gradient Boosted Trees e Support Vector Machine.	115
Figura 4-12: Curva ROC relativa aos testes de mineração de dados feita pela plataforma RapidMiner® em amostras com 30% de adição de leite bovino em leite bubalino simulando fraude.	117
Figura 4-13: Area Under the Curve relativa aos testes de mineração de dados feitos pelo Rapidminer® utilizando dados FTIR relativos à composição do leite em amostras com 30% de adição de leite bovino em leite bubalino simulando fraude. Em ordem: Naive Bayes, Generalized Linear Model, Logistic Regression, Fast Large Margin, Deep Learning, Decision Trees, Random Forest, Gradient Boosted Trees e Support Vector Machine.	117
Figura 4-14: Curva ROC relativa aos testes de mineração de dados feita pela plataforma RapidMiner® em amostras com 40% de adição de leite bovino em leite bubalino simulando fraude.	119
Figura 4-15: Area Under the Curve relativa aos testes de mineração de dados feitos pelo Rapidminer® utilizando dados FTIR relativos à composição do leite em amostras com 40% de adição de leite bovino em leite bubalino simulando fraude. Em ordem: Naive Bayes, Generalized Linear Model, Logistic Regression, Fast Large Margin, Deep Learning, Decision Trees, Random Forest, Gradient Boosted Trees e Support Vector Machine.	119
Figura 4-16: Curva ROC relativa aos testes de mineração de dados feita pela plataforma RapidMiner® em amostras com 50% de adição de leite bovino em leite bubalino simulando fraude.	121
Figura 4-17: Area Under the Curve relativa aos testes de mineração de dados feitos pelo Rapidminer® utilizando dados FTIR relativos à composição do leite em amostras com 50% de adição de leite bovino em leite bubalino simulando fraude. Em ordem: Naive Bayes, Generalized Linear Model, Logistic Regression, Fast Large Margin, Deep Learning, Decision Trees, Random Forest, Gradient Boosted Trees e Support Vector Machine.	121
Figura 4-18: Curva ROC relativa aos testes de mineração de dados feita pela plataforma RapidMiner® em amostras com 75% de adição de leite bovino em leite bubalino simulando fraude.	123
Figura 4-19: Area Under the Curve relativa aos testes de mineração de dados feitos pelo Rapidminer® utilizando dados FTIR relativos à composição do leite em amostras com 75% de adição de leite bovino em leite bubalino simulando fraude. Em ordem: Naive Bayes, Generalized Linear Model, Logistic Regression, Fast Large Margin, Deep Learning, Decision Trees, Random Forest, Gradient Boosted Trees e Support Vector Machine.	123
Figura 4-20: Comparativo da acurácia entre os diferentes testes de mineração de dados realizados pelo Rapidminer® em relação aos diferentes níveis de adição de leite bovino ao leite bubalino, utilizando dados obtidos de análise FTIR realizados em amostras obtidas de 2020-2022.	124

Lista de Tabelas:

Tabela 0-1: Comparativo da composição entre os leites de búfala e vaca	24
Tabela 0-2: Parâmetros Físico-Químicos para o Leite de Búfala determinada pela Resolução SAA-03 de 10 de janeiro de 2008 do estado de São Paulo.	26
Tabela 1-1: Teores médios e desvios padrão (DP) sazonais de gordura, lactose, proteína, sólidos totais (ST) e extrato seco desengordurado (ESD) de leite de búfala coletado mensalmente de tanques refrigeradores durante o período de julho/2020 a junho/2022 em Minas Gerais (n=837).....	55
Tabela 1-2: Valores médios e desvios padrão (DP) sazonais de contagens de células somáticas (CCS), CCS médias geométricas trimestrais, contagens padrão em placas (CPP) e CPP médias geométricas trimestrais de leite de búfala coletado mensalmente de tanques refrigeradores durante o período de julho/2020 a junho/2022 em Minas Gerais (N=837).....	59
Tabela 2-1: Estatística descritiva da composição do leite de vacas e de búfalas obtidos de tanques de refrigeração em unidades produtoras de MG durante o período de 2020-2022 e analisadas por FTIR no Laboratório de Análise da Qualidade do Leite da Escola de Veterinária da UFMG.	69
Tabela 2-2: Estatística descritiva de componentes do leite em cada um dos níveis de mistura de leite e em amostras de leite bovino e bubalino sem adição, obtidos de tanques de refrigeração de fazendas no estado de MG no período de 2020-2022 e analisadas por FTIR no Laboratório de Análise da Qualidade do Leite da Escola de Veterinária da UFMG.	72
Tabela 2-3: Quadro de ANOVA relativo à comparação entre as análises de leite bovino e bubalino obtidos de tanques de refrigeração de fazendas no estado de MG no período de 2020-2022 e analisadas por FTIR no Laboratório de Análise da Qualidade do Leite da Escola de Veterinária da UFMG.	74
Tabela 2-4: Comparação entre amostras de leites sem adição bovino e bubalino obtidos de tanques de refrigeração de fazendas no estado de MG no período de 2020-2022 e misturas de leite em diferentes porcentagens de adição de leite bovino em leite bubalino, em relação às variáveis de composição do leite obtidas por análise FTIR no Laboratório de Análise da Qualidade do Leite da Escola de Veterinária da UFMG.	74
Tabela 3-1: Arquitetura da Rede Neural Multilayer Perceptron de uma camada (MLP1) utilizando dados obtidos de análises FTIR feitas no Laboratório de Análise da Qualidade do Leite da Escola de Veterinária da UFMG, utilizando amostras coletadas durante os anos de 2020-2022.....	83
Tabela 3-2: Percentual médio de acerto e classificação das redes Multilayer Perceptron de uma camada em relação aos níveis de adição de leite bovino em leite bubalino simulando fraude experimental.	84
Tabela 3-3: Arquitetura de Rede Neural Multilayer Perceptron de duas camadas (MLP2) utilizando dados obtidos de análises FTIR feitas no Laboratório de Análise da Qualidade do Leite da Escola de Veterinária da UFMG, utilizando amostras coletadas durante os anos de 2020-2022.....	85
Tabela 3-4: Percentual médio de acerto e classificação das redes Multilayer Perceptron de duas camadas em relação aos níveis de adição de leite bovino em leite bubalino simulando fraude experimental.	86
Tabela 3-5: Arquitetura de Rede Neural de Função de Base Radial (RBF) utilizando dados obtidos de análises FTIR feitas no Laboratório de Análise da Qualidade do Leite da Escola de Veterinária da UFMG, utilizando amostras coletadas durante os anos de 2020-2022.....	87

Tabela 3-6: Percentual médio de acerto e classificação das redes de Função de Base Radial em relação aos níveis de adição de leite bovino em leite bubalino simulando fraude experimental	88
Tabela 3-7: Comparação múltipla entre as métricas de qualidade dos testes (acurácia, precisão, sensibilidade e especificidade) e os diferentes modelos de Redes Neurais Artificiais utilizados segundo o teste de Tukey.	90
Tabela 4-1: Comparativo dos métodos de mineração de dados analisados pelo RapidMiner [®] em relação à acurácia, erro, precisão, sensibilidade, especificidade e tempo de análise em amostras de leite com adição de 1% de leite bovino em leite bubalino simulando fraude 1%	106
Tabela 4-2: Comparativo dos métodos de mineração de dados analisados pelo RapidMiner [®] em relação à acurácia, erro, precisão, sensibilidade, especificidade e tempo de análise em amostras de leite com adição de 2% de leite bovino em leite bubalino simulando fraude. ...	108
Tabela 4-3: Comparativo dos métodos de mineração de dados analisados pelo RapidMiner [®] em relação à acurácia, erro, precisão, sensibilidade, especificidade e tempo de análise em amostras de leite com adição de 5% de leite bovino em leite bubalino simulando fraude	110
Tabela 4-4: Comparativo dos métodos de mineração de dados analisados pelo RapidMiner [®] em relação à acurácia, erro, precisão, sensibilidade, especificidade e tempo de análise em amostras de leite com adição de 10% de leite bovino em leite bubalino simulando fraude ..	112
Tabela 4-5: Comparativo dos métodos de mineração de dados analisados pelo RapidMiner [®] em relação à acurácia, erro, precisão, sensibilidade, especificidade e tempo de análise em amostras de leite com adição de 20% de leite bovino em leite bubalino simulando fraude. .	114
Tabela 4-6: Comparativo dos métodos de mineração de dados analisados pelo RapidMiner [®] em relação à acurácia, erro, precisão, sensibilidade, especificidade e tempo de análise em amostras de leite com adição de 30% de leite bovino em leite bubalino simulando fraude ..	116
Tabela 4-7: Comparativo dos métodos de mineração de dados analisados pelo RapidMiner [®] em relação à acurácia, erro, precisão, sensibilidade, especificidade e tempo de análise em amostras de leite com adição de 40% de leite bovino em leite bubalino simulando fraude ..	118
Tabela 4-8: Comparativo dos métodos de mineração de dados analisados pelo RapidMiner [®] em relação à acurácia, erro, precisão, sensibilidade, especificidade e tempo de análise em amostras de leite com adição de 50% de leite bovino em leite bubalino simulando fraude ..	120
Tabela 4-9: Comparativo dos métodos de mineração de dados analisados pelo RapidMiner [®] em relação à acurácia, erro, precisão, sensibilidade, especificidade e tempo de análise em amostras de leite com adição de 75% de leite bovino em leite bubalino simulando fraude ..	122

SUMÁRIO

1. INTRODUÇÃO GERAL	19
2. REVISÃO DE LITERATURA	22
1. Bubalinocultura brasileira.....	22
2. Características do leite de vaca e leite de búfala.....	24
3. Fraude em leite de búfala.....	26
4. Métodos convencionais para detecção de fraudes	29
5. Espectroscopia no infravermelho.....	31
6. Infravermelho com Transformada de Fourier.....	31
7. Uso da técnica FTIR na análise de adulteração em leite e derivados	32
8. Mineração de dados	33
9. Aprendizado de máquina	34
10. Redes Neurais Artificiais (RNA).....	35
10.1. O neurônio	35
10.2. O neurônio artificial.....	37
10.3. Função de Ativação	38
10.4. Arquiteturas de Redes Neurais Artificiais	39
10.4.1. Redes Perceptron Multicamadas.....	41
10.4.2. Redes de Função Base Radial	42
10.4.3. Comparação entre Redes de Função de Base Radial e Perceptron de Múltiplas Camadas	43
10.5. Utilização de Redes Neurais Artificiais na análise de Alimentos	44
10.6. Treinamento das Redes Neurais Artificiais	46
CAPÍTULO 1: QUALIDADE DO LEITE DE BÚFALA EM DIFERENTES ESTAÇÕES DO ANO	48
1.1. INTRODUÇÃO	48
1.2. CARACTERÍSTICAS DO LEITE DE BÚFALA	48
1.3. MATERIAL E MÉTODOS	52
1.3.1. Obtenção das amostras	52
1.3.2. Local de realização do experimento	53
1.3.3. Análise de Contagem de Células Somáticas.....	53
1.3.4. Análise de Composição do leite	53
1.3.5. Análise de Contagem Padrão em Placas (CPP).....	54
1.3.6. Análise Estatística	54
1.4. RESULTADOS E DISCUSSÃO	55
1.5. CONCLUSÃO	62

CAPÍTULO 2: ANÁLISE FTIR DO LEITE BUBALINO ACRESCIDO DE DIFERENTES NÍVEIS DE LEITE BOVINO E ANÁLISE ESTATÍSTICA DOS RESULTADOS	63
2.1. INTRODUÇÃO.....	63
2.2. MATERIAL E MÉTODOS.....	64
2.2.1. Local de realização do experimento.....	64
2.2.2. Obtenção de amostras	64
2.2.3. Análise de composição centesimal do leite	64
2.2.4. Análise da contagem de células somáticas.....	65
2.2.5. Pools de leite com misturas de diferentes espécies.....	66
2.2.7. Análise estatística.....	69
2.3. RESULTADOS E DISCUSSÃO	69
2.4. CONCLUSÃO	76
CAPÍTULO 3: CONSTRUÇÃO DAS REDES NEURAIS A PARTIR DOS RESULTADOS DOS COMPONENTES OBTIDOS POR FTIR	77
3.1 INTRODUÇÃO.....	77
3.2 MATERIAL E MÉTODOS	78
3.2.1 <i>Multilayer Perceptron (MLP)</i>	78
3.2.2 <i>Função de Base Radial</i>	79
3.2.3 Análise estatística.....	79
3.3 RESULTADOS E DISCUSSÃO	80
3.3.1 Arquiteturas de Redes.....	82
3.3.2 Testes de comparação múltipla das Redes Neurais	89
3.3.3 Gráficos Boxplot	92
3.4 CONCLUSÃO	95
CAPÍTULO 4: APRENDIZADO DE MÁQUINA E APRENDIZADO PROFUNDO A PARTIR DOS RESULTADOS DOS COMPONENTES OBTIDOS POR FTIR.....	97
4.1 INTRODUÇÃO.....	97
4.2 RAPIDMINER E MODELOS DE CLASSIFICAÇÃO	97
4.2.1 <i>Naive Bayes:</i>	98
4.2.2 <i>Generalized Linear Model:</i>	99
4.2.3 <i>Logistic Regression:</i>	99
4.2.4 <i>Fast Large Margin:</i>	100
4.2.5 <i>Deep Learning:</i>	100
4.2.6 <i>Decision Tree:</i>	101
4.2.7 <i>Random Forest:</i>	102

4.2.8	<i>Gradient Boosted Trees:</i>	103
4.2.9	<i>Support Vector Machine:</i>	103
4.3	<i>Curva ROC e AUC:</i>	103
4.4	MATERIAL E MÉTODOS	104
4.4.1	Análise estatística	104
4.5	RESULTADOS E DISCUSSÃO:	105
4.5.1	Fraude 1%	105
4.5.2	Fraude 2%	108
4.5.3	Fraude 5%:.....	110
4.5.4	Fraude 10%	112
4.5.5	Fraude 20%	114
4.5.6	Fraude 30%	115
4.5.7	Fraude 40%	118
4.5.8	Fraude 50%	119
4.5.9	Fraude 75%	121
4.6	CONCLUSÃO	127
3.	CONCLUSÃO GERAL	128
	REFERÊNCIAS	130

DETECÇÃO DE ADULTERAÇÃO POR ADIÇÃO DE LEITE BOVINO AO LEITE BUBALINO UTILIZANDO REDES NEURAIAS ARTIFICIAIS

1. INTRODUÇÃO GERAL

Segundo a legislação brasileira, entende-se por leite, sem outra especificação, o produto oriundo da ordenha completa, ininterrupta, em condições de higiene, de vacas sadias, bem alimentadas e descansadas. O leite de outros animais deve denominar-se segundo a espécie de que proceda, sendo permitida a mistura de leite de espécies animais diferentes, desde que conste na denominação de venda do produto e seja informada na rotulagem a porcentagem do leite de cada espécie (Brasil, 2017).

Dotada de uma produção de 34,84 bilhões de litros de leite, em 2019, a atividade leiteira no Brasil se distribui por quase todo o país. A produção total de leite no ano de 2019 segundo os últimos dados disponíveis, aumentou cerca de 2,7% em relação ao ano anterior. O principal estado responsável pela produção leiteira de forma geral é Minas Gerais que detém uma participação de 27,11% no total produzido (IBGE, 2021).

No leite bubalino, por se tratar de um produto disponível em pequenas quantidades em vários países do mundo, fraudes e adulterações por adição de matérias-primas com maior disponibilidade e/ ou custo mais baixo, são frequentemente descritas (Drummond, 2013). Esta prática infringe a Legislação brasileira disposta no Regulamento de Inspeção Industrial e Sanitária de Produtos de Origem Animal (RIISPOA) que define no Art. 504 as definições de fraude e falsificação, e o Código de Defesa do Consumidor no Art. 6º que assegura a informação clara acerca de quantidades, características, composição, qualidade e riscos dos produtos destinados ao consumo (Brasil, 1997; Brasil, 2017). Pode também comprometer a saúde do consumidor pela ingestão de compostos não relatados no rótulo do produto (Dias, 2009) como a beta-caseína A1, por exemplo. Sendo assim, a detecção da adulteração de forma mais rápida e eficiente é fundamental para que se garanta a autenticidade, evitando fraudes e dúvidas quanto aos produtos comercializados.

Dessa forma, a constante busca por soluções de problemas complexos, tem levado pesquisadores a se inspirarem em fontes da própria natureza para criar novos algoritmos baseados em inteligência de enxame, sistemas biológicos, físicos e químicos. Essas soluções baseadas em modelos da natureza, inspiram otimizações verificadas no comportamento coletivo de formigas, abelhas, pássaros e derivando de metodologias para obtenção de resultados com

melhor desempenho e menor caminho possível (Fister, *et al.*, 2013). Dentro desse cenário, a inteligência artificial relacionada à capacidade das máquinas de realizarem funções envolvendo aprendizagem e solução de problemas se assemelha ao comportamento “cognitivo” exclusivo dos seres humanos (Sage, 1990).

As Redes Neurais Artificiais (RNA) surgiram, portanto, como modelos de aprendizagem supervisionado que apresentam arquitetura computacional inspiradas nas redes neurais (sistema nervoso) e são usadas para aproximar funções que dependem de um grande número de entradas que geralmente são desconhecidas. Esse sistema é apresentado então como um sistema de “neurônios” interconectados que são capazes de calcular valores de entrada e realizar aprendizado de máquina, bem como reconhecimento de padrões devido à sua natureza adaptativa (Nikam, 2015). Sendo assim, as RNAs podem ser utilizadas como um modelo de predição de amostras anormais de leite, indicativas de fraude por mistura de leites de diferentes espécies.

1.1. Objetivos:

O presente estudo objetivou realizar a caracterização do leite bubalino quanto à composição e a contagem de células somáticas e realizar a predição de adulteração por adição de leite bovino ao leite bubalino, utilizando análises FTIR com associação a redes neurais artificiais e outras técnicas de mineração de dados.

1.1.1. Objetivos Específicos

- Realizar a análise FTIR em amostras de leite bovino e bubalino sem adição e em amostras com mistura de leite bovino e bubalino em diferentes porcentagens quanto à contagem de células somáticas e composição;
- Utilizar metodologias computacionais capazes de identificar adulteração no leite por misturas de leite de diferentes espécies (bovino e bubalino) em amostras analisadas por espectroscopia FTIR;
- Utilizar um banco de dados criado por resultados obtidos da espectroscopia FTIR a partir de amostras adulteradas experimentalmente com misturas de leite bovino e bubalino e amostras de leite sem adição e treinar modelos de aprendizado de máquina;
- Propor arquiteturas de rede neural *Multilayer Perceptron* e Função de Base Radial capazes de reconhecer e extrair informações a partir dos resultados obtidas por FTIR;

- Treinar modelos com algoritmos de classificação para detecção de adulteração de adição de leite bovino ao leite bubalino utilizando os resultados da análise de composição como dados de entrada.

2. REVISÃO DE LITERATURA

1. Bubalinocultura brasileira

O búfalo doméstico, *Bubalus bubalis*, é uma espécie leiteira com crescente importância no Brasil, que possui atualmente o maior rebanho de bubalinos do ocidente, com 1.502.482 cabeças (IBGE, 2021; Damé, 2006). Originários da Índia, são animais de fácil criação, rústicos e mais resistentes a doenças. Esses animais chegaram ao Brasil na década de 1880 no Estado do Pará e entre os anos de 1919 e 1920 ocorreram as importações de búfalos no Estado de Minas Gerais. A partir das reproduções desses animais surgiram grande parte da população bubalina do centro-sul e sudeste do Brasil, principalmente de Minas Gerais e São Paulo (Zava, 2011).

A maior resistência dos búfalos a algumas patologias permite aos produtores reduzirem a utilização de pesticidas e medicamentos, o que resulta na possibilidade de produzir leite e carne com redução de resíduo de medicamentos com maior facilidade em relação aos bovinos (Neto e Bastianetto, 2009).

Atualmente existem quatro raças de búfalos no Brasil: Mediterrâneo, Murrah e Jafarabadi, com dupla aptidão para a produção de leite e carne, e Carabao com aptidão para produção de carne e trabalho de tração. Os bubalinos da raça Murrah são considerados os de melhor aptidão leiteira em razão de sua conformação e temperamento dócil sendo a predominante na produção leiteira no Brasil. Possuem como padrão racial altura de 1,35m e peso corporal médio de 550kg e 750kg para fêmeas e machos respectivamente. Possuem corpo curto, reto e profundo, simétrico e equilibrado, com conformação média e compacta. O pescoço tem comprimento médio, forte no macho e descarnado na fêmea. O dorso é largo e um pouco selado, as costelas são arqueadas, as ancas salientes e a garupa larga. Possuem úbere volumoso, e tetos longos, sendo os anteriores mais curtos e bem separados (Marques *et al.*, 2011).

Os animais da raça Jafarabadi possuem aptidão para leite e carne, temperamento dócil e, juntamente com a raça Mediterrâneo, apresentam população genética mais homogênea (Marques *et al.*, 2011). Caracteriza-se pela forma peculiar da cabeça com os chifres longos, caídos e voltados para cima. A pelagem é preta e bem definida. No Brasil há duas variedades bem distintas: a Gir búfalo, mais delicada e de ossatura leve; e a Palitana, que possui ossatura mais pesada e grande carapaça na região frontal (Cavalli e Pereira, 2020).

Os bubalinos da raça Mediterrâneo foram selecionados na Itália para a produção de leite, porém pelo seu porte, possuem aptidão mista para leite e carne. Morfologicamente apresentam características das raças Murrah e Jafarabadi. É comumente conhecido no Brasil como búfalo “preto” ou “italiano”. Apresenta características corporais semelhantes à raça Murrah, porém com porte mais alto e chifres mais longos, fortes e grossos, de seção oval ou triangular, dirigidos para trás, para fora e para o alto. As linhagens com maior aptidão leiteira mostram corpo mais longo e menos musculoso (Cavalli e Pereira, 2020).

A produção de leite é a aptidão mais importante do bubalino, portanto o conhecimento acerca do controle da produção e do produto, são fundamentais para o sucesso econômico da atividade leiteira. A espécie possui uma capacidade intrínseca de conversão de forragens de baixa qualidade e resíduos de culturas de áreas marginais em carne e leite de alta qualidade. Além disso, possuem longa vida produtiva, podendo chegar a nove ou dez parições e produzindo de 7 a 11 litros de leite por dia.

Os búfalos têm um comportamento reprodutivo influenciado pela diminuição de horas de luz do dia (Zicarelli, 2010). Pode-se dizer que os búfalos são poliéstricos estacionais de dias curtos, semelhantemente aos ovinos e caprinos. Na região tropical quente e úmida, a sazonalidade reprodutiva dos búfalos é caracterizada pela concentração de partos de janeiro a maio, apesar de ocorrerem partos durante todo o ano (Pereira *et al.*, 2007). Apesar da concentração de partos facilitar o trabalho e o manejo em propriedades leiteiras, é fundamental que tenham partos o ano inteiro para a manutenção do setor (Cavalli e Pereira, 2020). A sazonalidade reprodutiva da espécie é um problema para a produção leiteira, pois interfere negativamente na oferta dos produtos lácteos à indústria, havendo uma queda de, aproximadamente, 60% da produção no período da entressafra. Portanto, enquanto a demanda por derivados se mantém, a disponibilidade desse produto diminui pela escassez de matéria-prima.

Devido a essas flutuações sazonais para disponibilidade do leite de búfala, aliado ao preço mais elevado comparado com o leite de vaca, a adição de leite de outras espécies no momento da chegada do leite ao laticínio, bem como, durante o preparo de produtos lácteos de búfala torna-se uma fraude ou falsificação frequente. Por isso, atualmente, são aplicados diversos métodos laboratoriais a fim de se detectar tal adulteração (Filho *et al.*, 2014).

2. Características do leite de vaca e leite de búfala

O leite de búfala apresenta características que o diferenciam de qualquer outro tipo de leite. Quando comparado ao leite bovino apresenta vantagens nutricionais, como maiores teores em sólidos totais, gordura, proteína, cálcio e fósforo (Tabela 0-1) (Oliveira *et al.*, 2009; Silva e Ribeiro, 2021). Os valores de lipídios, proteínas, lactose, sólidos totais e resíduo mineral do leite de búfala são de grande importância nutricional.

Tabela 2-1: Comparativo da composição entre os leites de búfala e vaca

Parâmetro	Espécie	
	Búfala	Vaca
Gordura (%)	8,16	3,68
Proteína (%)	4,50	3,70
Cinzas (%)	0,70	0,70
Sólidos Totais (%)	17,00	12,00
Cálcio (mg)	180,00	107,00
Ferro (mg)	0,12	0,07
Fósforo (mg)	120,00	82,00
Vitamina A (U.I)	204,27	185,49
Calorias por 100 mL	104,29	62,83

Fonte: Adaptado de Verruma e Salgado (1994)

A vitamina A é a que apresenta maior destaque, conferindo uma das características mais marcantes ao leite de búfala, que é uma coloração acentuadamente branca e opaca devido à ausência dos pigmentos carotenóides (provitamina A e β -caroteno), responsável pela cor amarelada dos produtos derivados do leite bovino (Amaral *et al.*, 2005). É mais concentrado que o leite bovino, apresentando assim menos água e mais matéria seca. Outra característica é possuir um sabor mais adocicado quando comparado ao leite de vaca, apesar de não possuir mais lactose que o leite bovino (Macedo *et al.*, 2001).

Seu alto teor de cálcio faz com que seja recomendado com a finalidade de minimizar a ocorrência de osteoporose. É importante aliado no processo de recuperação de pacientes, em luta contra doenças de todos os tipos, que precisam ingerir alimentos de qualidade em pequenas doses ao longo do dia devido ao seu alto valor nutricional (Dubey, 1997; Macedo *et al.*, 2001).

O leite de búfala é quase duas vezes mais rico em gordura quando comparado ao leite de vaca. Esta é a fração mais importante responsável por seu alto valor energético e nutritivo. Apesar dos valores relatados por Verruma e Salgado (1994), trabalhos realizados com leite de

búfala encontraram valores diferentes no teor de gordura. Varricchio *et al.* (2007) relataram um valor médio de 8,3%, Amaral *et al.*, (2004) encontraram média geral de 6,83%, Santos (2022) obtiveram média de 6,25% com amostras variando entre 3,04% e 10,78%. Os níveis de colesterol, total e livre, na gordura do leite de búfala são mais baixos do que no leite de vaca. A concentração total de colesterol no leite de búfala é menor que a encontrada no leite de vaca (275 mg versus 330 mg por 100 g de gordura), e é 1,5 a 1,9 vezes mais calórico que o leite de vacas (Zicarelli, 2010).

A maior vantagem do leite bubalino, em relação ao leite de outras espécies, é a qualidade nutricional, principalmente pelos valores de sólidos totais, por possuir teores de proteínas, gorduras e minerais que superam consideravelmente os do leite da vaca e, portanto, o seu aproveitamento industrial é maior, chegando comparativamente a sobrepujar o rendimento do leite bovino em mais de 40%. No Brasil são produzidos queijos que tradicionalmente são feitos com o leite de búfala, apesar de terem sofrido adaptações no país não se tendo uma definição na legislação acerca do tipo de leite a ser utilizado. O principal queijo bubalino é um queijo fresco de massa filada com aroma e sabor característicos, denominado como muçarela, originalmente produzido na Itália. Além desse, são produzidos queijos nacionais como o Marajoara, branco e macio, o provolone, a ricota e o mascarpone utilizando o leite de búfala como matéria-prima (Carvalho e Lourenço Júnior, 2001).

O rendimento de produção a partir da utilização do leite bubalino para a fabricação de derivados lácteos pode ser observado na produção de muçarela. Apenas cinco litros de leite bubalino são necessários para produzir 1 kg de muçarela de búfala, em contrapartida, são necessários 10 litros de leite bovino para produzir 1 kg de muçarela. Esse fator possibilita que a indústria pague mais pelo leite de búfala em relação ao leite bovino (Cavalli e Pereira, 2020).

Desse modo, esse produto tem grande importância tanto *in natura* quanto para a elaboração de derivados lácteos, principalmente os queijos (Drummond, *et al.* 2013). É importante lembrar que a legislação federal brasileira não estabelece parâmetros de identidade e qualidade para o leite bubalino. Somente o Estado de São Paulo possui uma legislação para alguns parâmetros de qualidade do leite bubalino (São Paulo, 2008), que estabelece valores mínimos de 4,5% para o teor de gordura, 8,57% para o extrato seco desengordurado e aspectos sensoriais, não fazendo referências para lactose, proteína, e outros requisitos importantes para esse leite (Tabela 0-2).

Tabela 2-2: Parâmetros Físico-Químicos para o Leite de Búfala determinada pela Resolução SAA-03 de 10 de janeiro de 2008 do estado de São Paulo.

Análise	Padrão
Aspecto e Cor	Normais para a espécie
Odor	Característico
Temperatura de Recebimento	7°C
Gordura	Mínimo de 4,5g/100g (4,5%)
Extrato seco desengordurado	Mínimo de 8,57g/100g
Acidez Titulável	0,14% a 0,23% em gramas de ácido láctico
Densidade a 15°C	Entre 1,028 e 1,034 g/mL
Índice Crioscópico	Entre -0,520 e -0,570°C.
Substâncias estranhas à sua composição ou corpos estranhos de qualquer natureza	Não deve apresentar

Fonte: Resolução SAA-03 de 10 de janeiro de 2008 - SP

Apesar do maior valor nutritivo e rendimento industrial do leite de búfalas quando comparados com o leite de vacas e do crescimento de sua exploração no país, pouco se tem feito para regulamentação de normas de padrão de identidade e qualidade do leite bubalino, o que dificulta a realização de medidas de controle e fiscalização aliada à falta de padrões a serem seguidos (Amaral *et al.*, 2005).

Diversos fatores contribuem para a composição do leite. Da mesma forma que a alimentação e nutrição, outros aspectos como o fator racial, estágio da lactação, temperatura ambiental, condições de estresse do animal e a perda excessiva de condição corporal podem ter influência direta. Além desses fatores, a estação do ano, a contagem de células somáticas, a mastite e a saúde geral, a manifestação de cio, a frequência e a técnica de ordenha e o melhoramento genético, exercem um maior ou menor efeito sobre a composição do leite (Mühlbach, 2009).

3. Fraude em leite de búfala

Na espécie bubalina ocorre a concentração dos partos entre os meses de janeiro a maio. No Sudeste do Brasil a concentração da produção de leite vai até o mês de setembro, período no qual a búfala produz cerca de 82% de leite, havendo ausência da matéria prima nos meses de verão (Bastianetto *et al.*, 2005; Czerwenka *et al.*, 2010; Andrighetto, 2011). A escassez, o

alto valor agregado, a maior oferta do leite de vaca, aliados à dificuldade da identificação desse tipo de fraude ou falsificação devido às semelhanças visuais e sensoriais entre esses produtos são alguns dos possíveis motivos que levam à fraude (Rodrigues *et al.*, 2012).

Segundo o Art. 504, parágrafo 2º, do Decreto nº 9.013, de 29 de março de 2017, que dispõe sobre o Regulamento da Inspeção Industrial e Sanitária de Produtos de Origem Animal (RIISPOA), consideram-se fraudados as matérias-primas ou produtos de origem animal: (1) Que tenham sido privados parcial ou totalmente de seus componentes característicos em razão da substituição por outros inertes ou estranhos e não atendem ao disposto na legislação específica; (2) Com adição de ingredientes, de aditivos, de coadjuvantes de tecnologia ou de substâncias com o objetivo de dissimular ou de ocultar alterações, deficiências de qualidade da matéria-prima ou defeitos na elaboração do produto; (3) Elaborados com adição de ingredientes, de aditivos, de coadjuvantes de tecnologia ou de substâncias com o objetivo de aumentar o volume ou o peso do produto; (4) ou elaborados ou comercializados em desacordo com a tecnologia ou o processo de fabricação estabelecido em normas complementares ou em desacordo com o processo de fabricação registrado, mediante supressão, abreviação ou substituição de etapas essenciais para qualidade ou identidade do produto (Brasil, 2017).

Além da fraude, o RIISPOA no mesmo Art. 504 define como falsificados: (1) as matérias-primas e os produtos que tenham sido elaborados, fracionados ou reembalados, expostos ou não ao consumo, com a aparência e as características gerais de outro produto registrado junto ao Departamento de Inspeção de Produtos de Origem Animal da Secretaria de Defesa Agropecuária do Ministério da Agricultura, Pecuária e Abastecimento e que se denominem como este, sem que o seja; (2) as matérias-primas e os produtos que tenham sido elaborados de espécie diferente da declarada no rótulo ou divergente da indicada no registro do produto; (3) as matérias-primas e os produtos que não atendam às especificações referentes à natureza ou à origem indicadas na rotulagem (Brasil, 2017).

Dentre as adulterações de leite bubalino mais comumente relatadas, a adição de leite bovino é a mais frequente. Segundo o Art. 235, parágrafo 2º, do RIISPOA, é permitida a mistura de leite de espécies animais diferentes, desde que conste na denominação de venda do produto e seja informada na rotulagem a porcentagem do leite de cada espécie (Brasil, 2017). Tratando-se de adulteração, essa mistura não é informada ao consumidor e por esta razão, além da avaliação da qualidade sensorial, microbiológica e físico-química do leite, se torna necessário que métodos eficazes sejam empregados para detecção desses processos (Mendes *et al.*, 2010).

Essas práticas podem lesar os consumidores em relação aos valores agregados ao produto (Dias, *et al.*, 2012) e acarretar riscos indiretos a saúde pública, como reações alérgicas devido à ocorrência de alergias específicas à proteína do leite de vaca. É de suma importância a identificação do leite utilizado na fabricação do queijo, especialmente aqueles feitos a partir de uma única espécie, tais como queijos de ovelhas, cabras ou búfalas. Para evitar a possível substituição fraudulenta de leite destas espécies pelo de vaca, que possui valor mais acessível, é necessário desenvolver procedimentos analíticos capazes de detectar tais adulterações e proteger os consumidores de rotulagem enganosa (López-Calleja *et al.*, 2007). Além da mistura de leite de diferentes espécies que compreende a principal fraude do leite bubalino, podem ocorrer adição de água, além da presença de resíduos de antibióticos.

3.1. Mistura de leites de diferentes espécies

Devido à maior disponibilidade do leite de vaca, principalmente devido ao fato de ser uma matéria-prima mais barata em relação ao leite de búfala, muitos produtores acreditam ser atrativo financeiramente realizar a mistura dos leites para obter uma maior rentabilidade. Essa mistura é realizada principalmente quando o leite é utilizado para a produção de derivados como a muçarela. Desta forma, o incremento de leite de vaca em qualquer quantidade, em queijos comercializados como bubalinos sem especificação em rótulo pode vir a lesar o consumidor, que acredita estar adquirindo um produto derivado exclusivamente do leite de búfala (Silva *et al.* 2015).

Silva *et al.* (2015) analisando amostras de queijos bubalinos comercializados no Estado do Pará encontraram 13% das amostras com presença de leite bovino. Em um estudo semelhante, Buzi *et al.* (2009) confirmaram a presença de leite de vaca em queijo de búfala em 22% das amostras obtidas no comércio varejista em São Paulo. Em todos os produtos cujos resultados foram positivos para leite de vaca, havia a indicação no rótulo de que o leite de búfala era a única matéria-prima utilizada.

Darwish *et al.* (2009) analisaram 21 amostras de leite fluido de búfala disponíveis no mercado e detectaram fraude em 38% das amostras coletadas. Do mesmo modo, Dias (2009) avaliando muçarela de búfala disponíveis no varejo do Rio de Janeiro, encontrou 23% de amostras positivas para mistura com leite bovino acima de 50% em relação ao leite bubalino sem que essa informação constasse no rótulo do produto. Esses dados evidenciam que a ocorrência de mistura de leite de diferentes espécies é uma prática comum e que podem acarretar prejuízos sérios aos consumidores.

4. Métodos convencionais para detecção de fraudes

Para evitar a fraude de mistura de leites de diferentes espécies, principalmente nos casos da adição de leite bovino ao leite bubalino, muitos métodos foram desenvolvidos, entre eles: eletroforese bidimensional (Chianese *et al.*, 1990), focalização isoelétrica (Moio *et al.*, 1990), eletroforese capilar (Recio *et al.*, 2004; Trimboli *et al.*, 2019), Cromatografia Líquida de Alta Eficiência (Mayer *et al.*, 1997; Mayer, 2005), ELISA (Hurley, 2004; Asensio *et al.*, 2008; Dias, 2009), PCR (Herman, 2001), PCR *real-time* (Drummond *et al.*, 2013), PCR multiplex (Silva *et al.*, 2015) e técnicas de cromatografia (Chou *et al.*, 2007).

Os métodos baseados em DNA, devido à sua extrema sensibilidade, reprodutibilidade e simplicidade, tornaram-se mais populares nas últimas décadas, substituindo os métodos a base de proteínas, em função de algumas limitações, como por exemplo, a sensibilidade dessas moléculas ao tratamento térmico (Mafra, 2007). Particularmente, a Reação em Cadeia pela Polimerase (PCR) tem sido a técnica mais utilizada para a detecção específica de origem animal em produtos lácteos e queijos. As células somáticas do leite, principalmente representadas por leucócitos, ainda persistem durante os processos de fabricação de queijo e podem ser usadas como uma fonte de DNA amplificável (Chavéz, 2008). Esse teste possibilita que determinada região do genoma de qualquer organismo seja multiplicada em milhões de cópias, o que facilita a análise genética e permite o desenvolvimento de técnicas de diagnóstico mais sensíveis e mais específicas do que as tradicionalmente utilizadas (Oliveira *et al.*, 2007). Quantidades mínimas de DNA podem ser detectadas usando primers espécie-específicos (Mafra *et al.*, 2007).

A técnica do PCR *Multiplex* foi desenvolvida com a finalidade de, com um único teste altamente específico, promover a diferenciação entre várias espécies ou entre vários gêneros simultaneamente (Oliveira, 2007).

Cottenet *et al.* (2011) estudaram o uso do PCR *multiplex real-time* com amplificação do DNA mitocondrial para análise de leite fluido de diferentes marcas comercializadas na China, Índia e Paquistão em busca de contaminação e fraudes. A técnica apresentou resultados satisfatórios de identificação das amostras que continham DNA de mais de uma espécie. Dentre essas alterações foram identificadas amostras com a presença de mistura de leite de diferentes espécies sem que constasse na rotulagem do produto. Foram detectados DNA de búfala em 21% das amostras de leite bovino analisadas no total. No leite bovino indiano, 20% das amostras

continham leite de búfala em pequenas quantidades (<5%) e 50% possuíam porcentagem de leite de búfala entre 5 e 50%. No mesmo trabalho investigaram a presença do leite de vaca no leite de búfala provenientes da Índia e Paquistão, com presença de DNA bovino no leite bubalino apenas nos leites indianos, 70% do leite com presença de leite bovino.

Lopez-Calleja *et al.*, (2007), relataram o uso da PCR como alternativa simples e rápida para detecção de DNA específico de vaca em amostras de leite e queijo tipo muçarela de búfala adquiridas na Espanha. Do mesmo modo Darwish *et al.* (2009) analisaram 21 amostras de leite fluido de búfala disponíveis no mercado e detectaram fraude em oito das amostras coletadas. Silva *et al.* (2015) utilizaram a técnica do PCR *Multiplex* para a diferenciação do DNA de espécies diferentes em queijos fabricados com leite de búfala fraudados experimentalmente e de 44 amostra de queijos obtidas no comércio varejista do Pará. Tanto nas amostras fraudadas experimentalmente quanto nas amostras coletadas em pontos de venda para o consumidor, o PCR foi capaz de identificar se houve ou não mistura de leites de diferentes espécies. Foram detectadas fraudes em 13,63% das amostras comerciais analisadas, a maioria proveniente do mercado informal e de feiras livres. Esses estudos evidenciaram que o método é rápido e eficaz para garantir a autenticidade do produto.

Apesar da importância do leite de búfala para o mercado, a legislação brasileira ainda não possui um método oficial validado para detecção de fraude em leite bubalino, bem como, de outras espécies com adição de leite bovino. Embora a autenticidade do leite e de produtos lácteos sempre tenha sido determinada por métodos analíticos tradicionais e sejam considerados como referência, eles requerem análises laboriosas, equipamentos sofisticados, de alto custo, que exigem operadores altamente qualificados e treinados, além de consistirem em processos demorados, que necessitam da aquisição de reagentes químicos, os quais, na maioria dos casos oferecem risco à saúde humana (Lin, 2009; Kamal e Karoui, 2015)

Apesar de serem métodos com grande eficácia e resultados promissores, as análises tradicionais demandam tempo, pessoal especializado e laboratórios equipados com instrumentos específicos. O uso de tecnologias mais recentes como análise de dados e uso de redes neurais artificiais podem ser úteis no processo de triagem de amostras anormais, evitando análises complementares desnecessárias e direcionando para provas definitivas, os casos suspeitos de adulteração.

5. Espectroscopia no infravermelho

A espectroscopia pode ser definida como a interação de qualquer tipo de radiação magnética com a matéria (Oliveira, 2001). A espectroscopia estuda a interação de radiações eletromagnéticas com as moléculas, tendo como objetivo a determinação dos níveis de energia de átomos. As diferenças de energia entre os níveis são determinadas pelo espectro, e, a partir destas medidas, determinam-se as posições relativas dos níveis energéticos. A ligação entre dois átomos envolve diferentes tipos de energia, como a energia translacional, vibracional e a eletrônica. As transições eletrônicas estão situadas na região do ultravioleta, as vibracionais na região do infravermelho e as rotacionais na região de micro-ondas (Sala, 2008)

O infravermelho corresponde à região do espectro eletromagnético situada na faixa de números de onda entre 14.290 e 10 cm^{-1} . De acordo com a faixa numérica, recebe uma denominação específica. Entre $400-10\text{ cm}^{-1}$ é denominada infravermelho distante. A região que compreende número de ondas entre $4000 - 400\text{ cm}^{-1}$ é a mais comumente utilizada pela química orgânica, sendo denominada infravermelho médio. A região chamada infravermelho próximo, 14.290 a 4000 cm^{-1} , tem recebido, recentemente, muita atenção, em particular com relação às análises quantitativas de amostras com matrizes complexas (Barbosa, 2007). Em uma leitura de um espectro, cada molécula será absorvida em diferentes comprimentos de onda e apresentará o seu próprio espectro na região do infravermelho. Isso permite obter informações químicas qualitativas, quantitativas e físicas, as quais são intrínsecas de cada átomo e molécula, tornando o método viável para identificar diferentes tipos de amostras (Stuart, 2004; Kamal e Karoui, 2015)

6. Infravermelho com Transformada de Fourier

A espectroscopia por infravermelho teve crescimento com o surgimento da espectroscopia com transformada de Fourier e a utilização do interferômetro de Michelson, que possibilitou que a técnica se tornasse ainda mais rápida (Saliba *et al.*, 2003). O interferômetro de Michelson consiste basicamente em dois espelhos (um fixo e um móvel) e um divisor de feixes (espelho semitransparente). As reflexões vão gerar interações construtivas e destrutivas que combinadas formarão um interferograma. Como o interferograma não pode ser interpretado, é empregada a técnica matemática chamada de Transformada de Fourier que converte os dados obtidos no interferômetro em um espectro que relaciona a intensidade *versus*

frequência (número de onda) (Stuart, 2004; Helfer *et al.*, 2006). A operação matemática transformada de Fourier do infravermelho (FT-IR) é aplicada, quer por um microprocessador acoplado ao espectrofotômetro, quer por um computador associado externamente ao aparelho, distribuindo assim a energia absorvida pela amostra por cada componente de onda, apresentando assim os valores de cada frequência, resultando numa compreensão e leitura dos resultados (Almeida, 2009).

O FTIR é uma tecnologia atraente especialmente para a indústria de alimentos, pois permite medições simples, rápidas e não destrutivas de componentes físicos e químicos. Os avanços em sua instrumentação juntamente com o desenvolvimento de métodos de análise de dados tornam esta tecnologia ideal para grandes volumes de amostras, triagem rápida e caracterização até dos menores componentes dos alimentos. As técnicas de FTIR já são conhecidas pelas indústrias de alimentos, no entanto, mais recentemente tem-se expandido seu potencial para o monitoramento de adulterações em alimentos (Cassoli *et al.*, 2011; Kamal e Karoui, 2015; Lohumi *et al.*, 2015).

7. Uso da técnica FTIR na análise de adulteração em leite e derivados

O uso das técnicas de FTIR vem sendo estudado principalmente nas últimas décadas em associação com as técnicas de tecnologia avançada e *data mining* para a avaliação da qualidade de produtos lácteos e sua aplicação na substituição de métodos de referência (Etzion *et al.* 2004; Nicolaou *et al.*, 2010; Cassoli *et al.*, 2011; Botelho *et al.*, 2015; Jaiswal *et al.*, 2015).

Muitos trabalhos foram realizados usando os gráficos de espectros. Pappas *et al.* (2008) encontraram diferenças entre leite de cabra e ovelha utilizando o FTIR para gerar espectros e diferenciar leites desnatados através da detecção da região espectral da que forma uma "impressão digital" de cada espécie. Dos onze picos gerados entre 1840 cm^{-1} e 950 cm^{-1} , pequenas diferenças foram notadas entre as amostras de leite, sendo a principal diferença na banda de 1745 cm^{-1} , referente à vibração da carbonila do grupo éster referente a açúcares, que foi mais intenso para o leite de cabra do que para o leite de ovelha devido ao maior grau de esterificação. Outro estudo realizado por Nicolaou *et al.* (2010) avaliou as diferenças dos leites das espécies vaca, cabra e ovelha por meio de combinação com métodos estatísticos multivariados e verificaram as diferenças existentes em suas composições com a sobreposição

das bandas derivadas dos espectros, com destaque ao comprimento de onda referente às gorduras (2.927 cm^{-1}) cuja intensidade foi maior no leite de ovelha.

Dessa forma, o FTIR tem sido utilizado para a diferenciação de leites oriundos de distintas espécies animais e também de diferentes tipos de queijos por meio da determinação da sua composição, diferenciação quanto à textura, aspectos sensoriais, vida de prateleira, variações sazonais e identificação das regiões onde foram produzidos por meio da análise da maturação, acidez, grau de deformação, gordura, dentre outros aspectos (Coroian *et al.*, 2012)

8. Mineração de dados

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).

Muitas pessoas consideram o termo Mineração de Dados como sinônimo de *Knowledge Discovery in Databases* (KDD) ou Descoberta de Conhecimento em Banco de Dados. Na verdade, KDD é um processo mais amplo consistindo das seguintes etapas, como ilustrado na Figura 0-1.

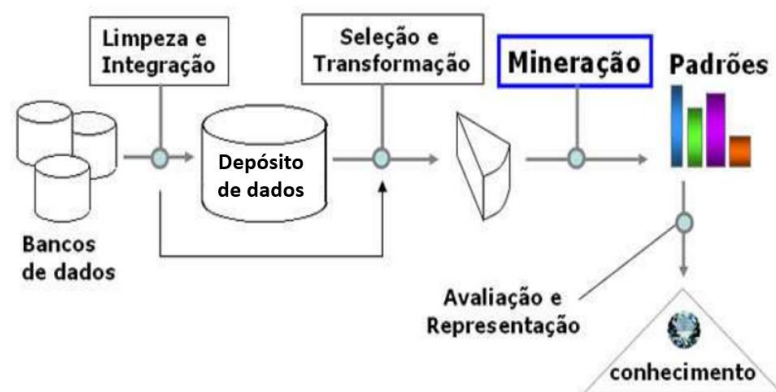


Figura 2-1: Etapas do processo de KDD (Knowledge Discovery in Databases) ou Descoberta de Conhecimento em Banco de Dados, desde as etapas de limpeza e integração até a etapa da avaliação do conhecimento a que se deseja estudar.

Fonte: De Amo, 2004.

Como demonstrado na Figura 0-1, ocorrem diferentes etapas até se obter o resultado pretendido pela mineração de dados. Essas etapas podem ser definidas como: (1) *Limpeza dos dados*: etapa onde são eliminados ruídos e dados inconsistentes. (2) *Integração dos dados*: etapa onde diversas fontes de dados podem ser combinadas produzindo um único repositório de dados. (3) *Seleção*: etapa onde são selecionados os atributos que interessam ao usuário de acordo com o problema a qual se quer solucionar. (4) *Transformação dos dados*: etapa onde os dados são transformados num formato apropriado para aplicação de algoritmos de mineração. (5) *Mineração*: etapa essencial do processo consistindo na aplicação de técnicas inteligentes a fim de se extrair os padrões de interesse. (6) *Avaliação ou Pós-processamento*: etapa onde são identificados os padrões interessantes de acordo com algum critério do usuário. (7) *Visualização dos Resultados*: etapa onde são utilizadas técnicas de representação de conhecimento a fim de apresentar ao usuário o conhecimento minerado (de Amo, 2004).

Os repositórios usados normalmente possuem muitos de registros. Neste contexto, o uso de todos os registros do repositório para a construção do modelo de Mineração de Dados é inviável. Assim, utiliza-se uma amostra com maior representatividade que é dividida em três conjuntos: (1) Conjunto de Treinamento (Training Set): conjunto de registros usados no qual o modelo é desenvolvido; (2) Conjunto de Validação (Validation Set): conjunto de registros usados para validar o modelo construído; (3) Conjunto de Testes (Test Set): conjunto de registros usados para testar o modelo construído;

Essa divisão em grupos é necessária para que o modelo não fique dependente de um conjunto de dados específico e, ao ser submetido a outros conjuntos (com valores diferentes dos usados na construção e validação do modelo), apresente resultados insatisfatórios. Este efeito é chamado de efeito Bias (Camilo e Silva, 2009).

9. Aprendizado de máquina

O aprendizado de máquina é realizado com o objetivo de construir e desenvolver programas que melhorem seu desempenho por meio de exemplos (Mitchell, 1997). Para isso, é preciso uma grande quantidade de exemplos a fim de gerar o conhecimento do computador, que consiste em hipóteses geradas a partir dos dados (Laudermir, 2021).

As técnicas de aprendizado de máquina são orientadas a dados, ou seja, aprendem automaticamente a partir de grandes volumes de dados. Dessa forma, os algoritmos formados, geram hipóteses a partir dos dados. Existem três tipos principais de aprendizado de máquina: Supervisionado, o Não Supervisionado e por Reforço.

No caso de Aprendizado Supervisionado, para cada exemplo apresentado ao algoritmo de aprendizado é necessário que se apresente a resposta desejada, como se fossem rótulos que classificam em classes. O objetivo do algoritmo é, portanto, construir um classificador que consiga determinar corretamente a classe dos novos exemplos ainda não rotulados. Para rótulos de classe discretos, o problema é chamado de classificação e para valores contínuos, regressão. Esse tipo de aprendizado é o mais comumente utilizado (Ludermir, 2021).

No aprendizado Não Supervisionado, os exemplos são apresentados ao algoritmo sem rótulos. O próprio algoritmo irá agrupar os exemplos pelas similaridades dos atributos, analisando se podem ser agrupados de alguma maneira formando os chamados *clusters*. Após o agrupamento, é necessária uma verificação para determinar o que cada agrupamento significa no contexto do problema (Ludermir, 2021).

Já o Aprendizado por Reforço utiliza o algoritmo para classificar os exemplos, mas não fornece as respostas corretas, apenas um sinal de reforço, de recompensa ou punição. Dessa forma a hipótese baseada nos exemplos determina se a resposta foi boa ou ruim (Ludermir, 2021).

10. Redes Neurais Artificiais (RNA)

O modelo mais comum de aprendizado supervisionado são as redes neurais artificiais que foram baseadas no sistema neural do cérebro (Nikam, 2015).

10.1. O neurônio

O cérebro humano é composto por bilhões de neurônios interconectados que permitem receber, processar e transmitir informações por meio de reações bioquímicas. Esse complexo sistema é denominado rede neural. Cada neurônio é uma simples unidade de microprocessamento que recebe e combina os sinais de muitos outros neurônios. O resultado

disso são tarefas bem-sucedidas, como reconhecimento de rostos, fala e movimentação corporal das pessoas (Baughman,1995; Silva *et al*, 2010; Abraham, 2005).

Os neurônios são constituídos por três seções principais: os dendritos, o corpo celular (ou soma) e o axônio (Figura 0-2).

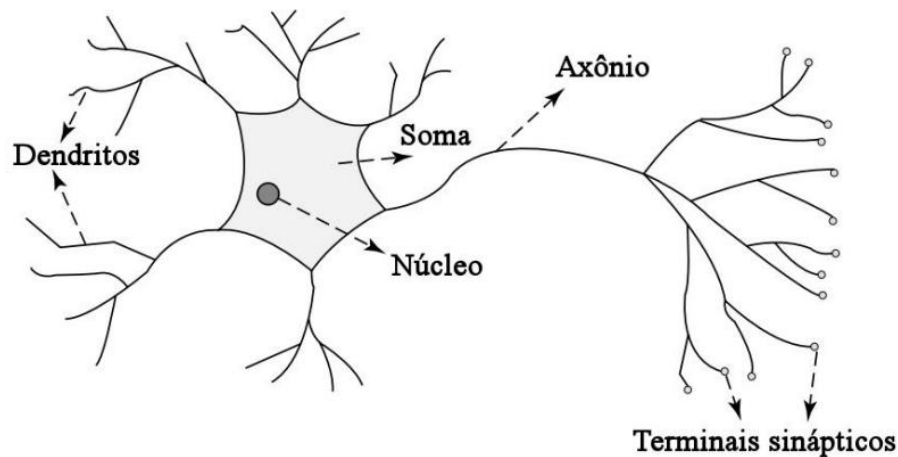


Figura 2-2: Representação esquemática de um neurônio biológico humano com dendritos, núcleo, soma, axônio e terminais sinápticos.

Fonte: Traduzido de Abraham (2005)

No interior do corpo celular, localiza-se um núcleo responsável por funções lógicas. Externamente estende-se ao axônio, fibra comprida e fina com terminais sinápticos, bem como os dendritos, mais curtos, semelhantes à galhos de árvores, formando uma rede. Os axônios de um neurônio se conectam aos terminais sinápticos de outro neurônio, transmitindo sinais complexos quimicamente liberados pelo efeito de aumentar ou diminuir o potencial elétrico dentro do corpo da célula. A comunicação entre neurônios se dá a partir da recepção dos sinais ou impulsos nervosos, oriundos de outros neurônios. Quando recebidos pelos dendritos e a soma dos impulsos ultrapassa o limiar da excitação da célula, ocorre a ativação do neurônio, sendo esse disparado e produzindo um sinal-resposta no corpo celular. Esse sinal percorre o axônio da célula até os dendritos dos próximos neurônios (Braga *et al.*, 2000; Silva *et al.*,2010, Abraham, 2005).

Este contato entre a terminação do axônio de uma célula com o dendrito de outro neurônio é conhecido como sinapse. Estas funcionam como válvulas que são capazes de controlar a transmissão dos impulsos entre os nodos da rede neural. Isso ocorre pela conversão de sinais elétricos pré-sinápticos em sinais químicos e novamente em sinal elétrico pós-

sináptico. A contribuição de todos os nodos pré e pós-sinápticos irá determinar a geração ou não de um impulso nervoso, possibilitando a maioria das funções realizada pelo cérebro, como memorização e o aprendizado (Braga *et al.*, 2000; Haykin, 2001).

As Redes Neurais Artificiais são baseadas na estrutura individual dos nodos, nas suas conexões e no comportamento conjunto e dinâmico desses nodos naturais a estímulos externos. Esta ferramenta pode aprender a relação entre entradas e saídas com base nos dados de treinamento, sendo, portanto, amplamente utilizada para modelagem, predição e classificação, por possibilitar a identificação de relações presentes em fontes de dados limitados, incompletos, com ruídos, dinâmicos e não lineares. Dessa forma, a habilidade da RNA de aprender e generalizar o comportamento de processos complexos e não lineares, a torna uma poderosa ferramenta de modelagem (Youssefi *et al.*, 2009; Aghbashlo *et al.*, 2012; Mohanraj *et al.*, 2012).

10.2. O neurônio artificial

Os neurônios artificiais (Figura 0-3) são elementos de processamento interconectado que fazem parte das RNAs. Esses neurônios baseiam-se nos neurônios humanos mediante generalizações de sistemas matemáticos, em que os efeitos das sinapses são estabelecidos como pesos que modulam os sinais de entradas. Uma função de ativação transforma a soma ponderada dos sinais de entrada e resultam no impulso do neurônio. De acordo com o algoritmo utilizado na RNA se estabelece a capacidade de aprendizado de um neurônio artificial (Abraham, 2005).

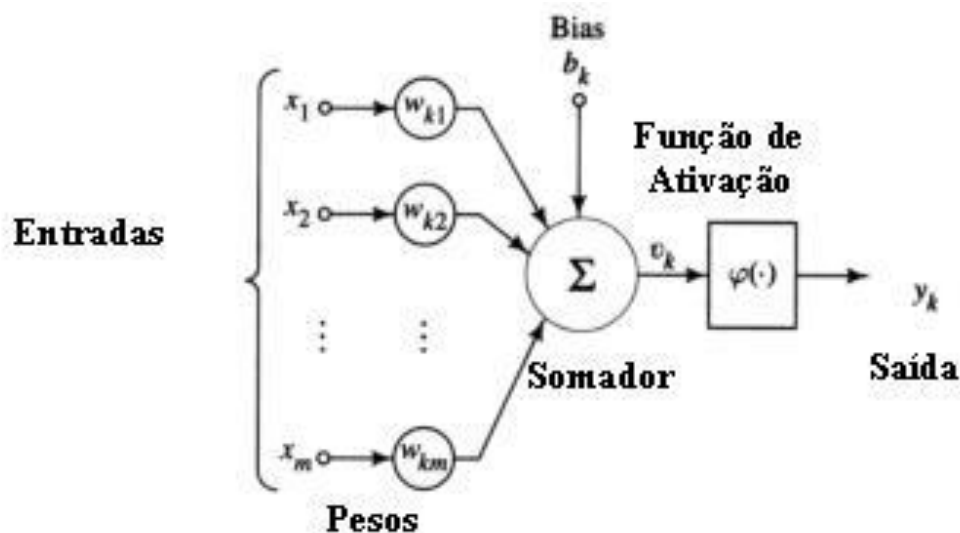


Figura 2-3: Representação de um neurônio artificial com as entradas, pesos, somador, função de ativação e camada de saída.

Fonte: Adaptado de Haykin (2001)

Na Figura 0-3 ficam evidenciados os três elementos básicos de um neurônio artificial:

1. Conjunto de sinapses ou elos de conexão caracterizados por um peso ou força própria. Um sinal x_i na entrada da sinapse j conectada ao neurônio k é multiplicado pelo peso sináptico w_{kj} .

Os valores de entrada x_i ($i=1$ até n) são sinais ou medidas advindas do meio externo que representam os valores assumidos pelas variáveis de uma determinada aplicação. Cada entrada é multiplicada por um peso de conexão (w_{ij}), cujos valores podem ser tanto positivos quanto negativos, caso as sinapses sejam excitatórias ou inibitórias respectivamente. Isso permite quantificar as suas relevâncias em relação à funcionalidade do respectivo neurônio (Baughman e Liu, 1995; Braga *et al*, 2000; Haykin, 2001; Silva *et al*, 2010).

2. Um *somador* para somar os sinais de entrada, ponderados pelas respectivas sinapses do neurônio; as operações descritas aqui constituem um *combinador linear*
3. Uma função de ativação para restringir a amplitude de saída de um neurônio. Essa função também é descrita como função restritiva uma vez que restringe (limita) o intervalo permissível de amplitude de sinal de saída a um valor finito. Tipicamente esse intervalo é escrito como o intervalo unitário fechado $[0, 1]$ ou alternativamente $[-1, 1]$. (Haykin, 2001)

10.3. *Função de Ativação*

A função de ativação (ou função de transferência) é usada nas redes neurais artificiais para transformar um sinal de entrada em um sinal de saída, que, por sua vez, pode alimentar outro sinal de entrada. Se uma rede neural não utiliza função de ativação, o sinal de saída se torna uma função linear simples, sendo incapaz de aprender e tampouco executar tarefas mais complexas (Sharma, 2020). Existem diversos tipos de função de ativação e as principais estão representadas na Figura 0-4. A função de limiar (ou degrau, ou passo) binário é mais simples, utilizada para classificadores binários, dada pela Equação (a). A função de ativação linear por partes é caracterizada em forma de rampa, com suas condições na Equação (b). A função sigmoide, representada na Equação (c), muito utilizada, torna a curva entre 0 e 1 mais suave (Haykin, 2001). A função tangente hiperbólica é similar à sigmoide, diferindo em poder assumir valores entre -1 e 1 (ao invés de 0 a 1), resultando em diferentes sinais de saídas de camadas

anteriores que podem alimentar as próximas camadas da rede. Está representada na Expressão (d).

$$\varphi(v) = \begin{cases} 1, & \text{se } v \geq 0 \\ 0, & \text{se } v < 0 \end{cases} \quad (\text{a})$$

$$\varphi(v) = \begin{cases} 1, & \text{se } v \geq +\frac{1}{2} \\ v, & +\frac{1}{2} > v > -\frac{1}{2} \\ 0, & \text{se } v \leq -\frac{1}{2} \end{cases} \quad (\text{b})$$

$$\varphi(v) = \frac{1}{1 + e^{-av}} \quad (\text{c})$$

$$\varphi(v) = \tanh(v) \quad (\text{d})$$

$$Pr(m - \text{ésima categ.}) = \frac{e^{W_m \cdot A}}{\sum_{L=1}^N e^{W_L \cdot A}} \quad (\text{e})$$

Figura 2-4: Funções de ativação comumente utilizadas em Redes Neurais Artificiais. (a) Função de Limiar, (b) Função de ativação linear por partes, (c) Função Sigmoides, (d) Função Tangente Hiperbólica, (e) Função Softmax.

Fonte: Haykin (2001)

A função *softmax* é uma combinação de múltiplas sigmóides, capaz de lidar com mais de dois tipos de classes, transformando a saída delas para valores entre 0 e 1 e dividindo pela soma das saídas. A saída do m-ésimo neurônio na camada da função *softmax* da teoria probabilística está demonstrada na Equação (e), sendo que: N é a quantidade de classes, W e A são vetores com j-ésimo elementos (Yuan, 2016)

10.4. Arquiteturas de Redes Neurais Artificiais

A topologia de uma RNA refere-se a como os neurônios dessa rede se interconectam. Por outro lado, a arquitetura define a forma como os seus diversos neurônios estão arranjos em relação uns aos outros (Baughman e Liu, 1995; Silva *et al.*, 2010). A maneira pela qual os neurônios de uma rede neural estão estruturados está intimamente ligada com o algoritmo de aprendizagem usado para treinar a rede (Haykin, 2001). O planejamento da arquitetura e a

definição da topologia irão refletir o tipo de problema que se deseja solucionar com a rede neural (Braga *et al.*, 2000).

As redes neurais podem ser divididas em três partes: camada de entrada, camadas ocultas e camada de saída (Khanlari *et al.*, 2012). A camada de entrada recebe as informações (dados) provenientes do meio externo. As camadas ocultas ou invisíveis possuem neurônios que irão extrair as características relacionadas ao processo ou sistema em análise. A camada de saída produz e apresenta os resultados finais da rede, os quais são resultantes dos processamentos realizados pelos neurônios de camadas anteriores (Silva *et al.*, 2000).

As principais arquiteturas de redes neurais artificiais, podem ser divididas considerando a disposição e interconexão dos seus neurônios, além da constituição de suas camadas. Elas podem ser, portanto, rede *feedforward* (alimentação à frente ou progressivas) de camada simples, rede *feedforward* de camadas múltiplas e redes recorrentes (Haykin, 2001).

As redes *feedforward* têm fluxo de sinal em apenas uma direção, são estáticas e podem possuir uma ou múltiplas camadas de neurônios (também conhecidas como camadas ocultas, escondidas ou intermediárias). As redes com múltiplas camadas são as mais comuns, e também conhecidas como *perceptron* multicamada (Jain *et al.*, 1996). As redes *feedforward* de múltiplas camadas são usadas tipicamente na resolução de vários problemas complexos, tais como aqueles relacionados à classificação de padrões, identificação de sistemas, robótica e controle de processos (Rai *et al.*, 2005). Neste caso, ao contrário das redes de camada simples, há uma ou mais camadas ocultas presentes (Figura 0-5). A determinação do número de camadas escondidas e o número de neurônios que as constituem dependem, sobretudo, do tipo e da complexidade do problema a ser solucionado pela rede, assim como da quantidade e da qualidade dos dados disponíveis a respeito do problema (Silva *et al.*, 2010).

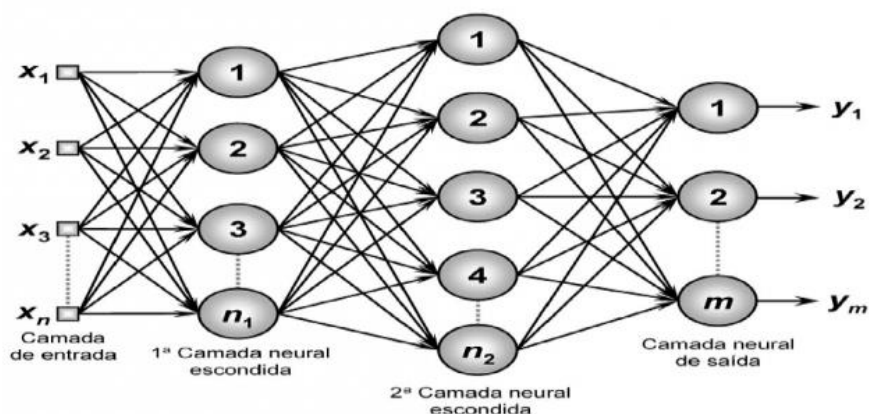


Figura 2-5: Arquitetura de uma rede neural de múltiplas camadas com camada de entrada, duas camadas ocultas e uma camada de saída.

Fonte: Silva *et al.* (2010)

10.4.1. *Redes Perceptron Multicamadas*

As redes neurais do tipo *Multilayer Perceptron* (MLP) consistem tipicamente por um conjunto de unidades sensoriais (nós de fonte) que constituem uma *camada de entrada*, uma ou mais *camadas ocultas* de nós computacionais e uma *camada de saída*. As camadas ocultas são responsáveis pela separação dos padrões de fronteiras de decisão, contendo uma quantidade de neurônios a ser definida para cada problema. A camada de saída constrói combinações lineares das fronteiras de decisão formadas pelos neurônios ocultos com um único neurônio responsável pela predição do valor a ser estudado. Em cada neurônio da rede, é realizado um somatório ponderado pelos pesos sinápticos dos sinais provenientes dos neurônios da camada anterior (Haykin, 2001; Marini, 2009).

A Figura 0-5 representa a arquitetura de uma rede *perceptron* de múltiplas camadas com duas camadas ocultas e uma camada de saída. Nessa arquitetura o fluxo do sinal progride através da rede para a frente, da esquerda para a direita e de camada por camada (Haykin, 2001).

Esse modelo de rede tem sido utilizado com sucesso para resolver diversos problemas difíceis, através do treinamento de forma supervisionada com um algoritmo de retro propagação de erro. Esse algoritmo é baseado na regra de aprendizagem por correção de erro que consiste em dois passos através das diferentes camadas da rede: um passo para frente (propagação) e um passo para trás (retro propagação). No passo para frente, um padrão de atividade é aplicado aos nós sensoriais da rede e seu efeito se propaga camada por camada. Durante esse passo, todos

os pesos da rede são fixos. Finalmente, um conjunto de saídas é produzido como resposta real da rede. Essa resposta real é subtraída de uma resposta desejada para produzir um sinal de erro. Esse sinal então é propagado para trás, contra a direção das conexões sinápticas, dando origem ao nome de “retropropagação de erro”. Nessa etapa os pesos são todos ajustados de acordo com uma regra de correção de erro, a fim de fazer com que a resposta real da rede se mova para mais perto da resposta desejada em um sentido estatístico (Haykin, 2001).

10.4.2. *Redes de Função Base Radial*

As redes neurais do tipo função de base radial consistem em uma camada de entrada, uma de saída e uma única camada oculta, as quais apresentam funções completamente diferentes. A arquitetura desse tipo de rede é mais simples quando comparada à *Multilayer Perceptron* (MLP), consistindo em duas camadas de pesos, e da mesma forma que as MLP multicamadas alimentadas adiante (*feedforward*) (Figura 0-6). A primeira camada possui os parâmetros das funções de base radial enquanto a segunda cria combinações lineares das funções a fim de gerar uma saída. A camada de entrada é formada por nós de fonte, que são unidades sensoriais com a função de conectar a rede ao seu ambiente e não executam nenhuma função sobre os dados de entrada. A segunda camada é a única camada oculta da rede, que aplica uma transformação não linear do espaço de entrada para o espaço oculto. Na maioria das aplicações, esse espaço oculto é de alta dimensionalidade e essas unidades ocultas fornecem um conjunto de funções radiais que constituem uma base arbitrária para os padrões de entrada. A camada de saída faz uma combinação linear de bases radiais, fornecendo a resposta da rede ao padrão de ativação aplicado à camada de entrada. Essa camada procura classificar os padrões recebidos da camada anterior (Haykin, 2001).

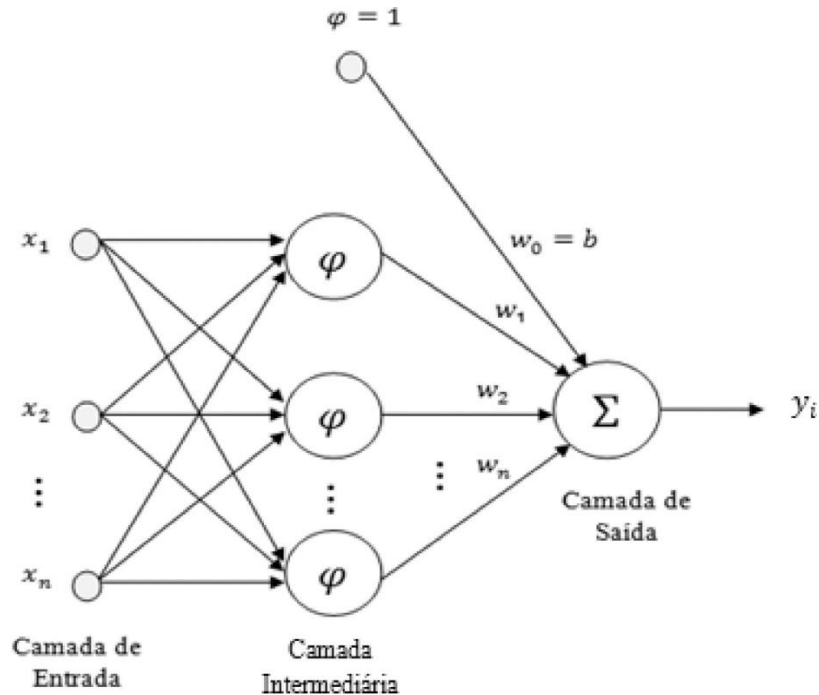


Figura 2-6: Representação de uma rede de função de bases radiais (RBF) com camada de entrada, camada intermediária e camada de saída.

Fonte: Haykin (2001)

Esse tipo de arquitetura é usualmente utilizado em estudos juntamente com as redes MLP para que se sejam comparados os resultados e utilizados os que apresentarem melhores valores com menor erro. Esse modelo de estudo foi utilizado com dados obtidos por espectroscopia FTIR para a quantificação do percentual de café arábica em misturas de café (Carvalho, 2014; Sabino, 2015). Em ambos os estudos, para as redes MLP foram utilizados de 6 a 13 neurônios na primeira camada, 0 a 8 na segunda camada. Para as redes RBF, a quantidade de bases radiais utilizadas foram de 5 a 80. Ambas as redes foram alimentadas com espectros sem adição e apresentaram erro menor que 5%. As redes obtiveram resultados similares, porém a rede MLP apresentou melhor desempenho pois obteve, de maneira geral, menores resultados de erro absoluto médio e maiores valores de coeficiente de correlação quando comparados à rede RBF.

10.4.3. Comparação entre Redes de Função de Base Radial e Perceptron de Múltiplas Camadas

As redes Perceptron de Múltiplas Camadas e as redes de Função de Base Radial são as mais utilizadas quando o sistema de alimentação da rede provém de dados obtidos de espectros do infravermelho, dessa forma é comum a constatação de que sempre existe uma rede RBF

capaz de imitar precisamente uma MLP específica ou vice-versa. Apesar de ambas serem exemplos de redes em camadas alimentadas adiante, não lineares e aproximadores universais, essas redes diferem entre si em alguns aspectos importantes.

- Uma rede RBF (na sua forma mais básica) possui apenas uma única camada oculta, enquanto as MLP podem apresentar uma ou mais camadas ocultas.
- Tipicamente, os nós computacionais de uma MLP, localizados na camada oculta ou em uma camada de saída, vão compartilhar um modelo neuronal comum. Em contrapartida, em uma RBF os nós da camada oculta são bastante diferentes e servem a um propósito diferente daqueles da camada de saída da rede.
- A camada oculta de uma rede RBF é não linear, enquanto a camada de saída é linear. Por outro lado, as camadas ocultas e de saída de uma MLP usadas como classificador de padrões, são normalmente todas não lineares. Dessa forma, quando o MLP é usado para solucionar problemas de regressão não linear, uma camada linear para a saída é normalmente a escolha preferida.
- As MLPs constroem aproximações globais de um mapeamento de saída não linear. Por outro lado, as redes RBF utilizando não linearidades localizadas com decaimento exponencial, constroem aproximações locais para mapeamentos de entrada-saída não lineares.

Portanto, isso demonstra que para a aproximação de um mapeamento de entrada-saída não-linear, os MLP requerem um número menor de parâmetros que a rede RBF para o mesmo grau de precisão. As características lineares da camada de saída da rede RBF indicam que essa rede está mais proximamente relacionada com o *perceptron de Rosenblatt* (tipo mais simples de rede neural artificial) do que com o perceptron de multicamadas. No entanto, RBF difere do perceptron pela capacidade de implementar transformações não lineares arbitrárias do espaço de entrada. Sendo assim, é capaz de resolver problemas que não conseguem ser resolvidos por nenhum perceptron (Haykin, 2001).

10.5. Utilização de Redes Neurais Artificiais na análise de Alimentos

Essa arquitetura na indústria de alimentos já foi bastante estudada e utilizada com sucesso. Khanmohammadi *et al.* (2009) propuseram o uso de redes neurais com camadas ocultas para a

determinação quantitativa do teor de proteína em iogurtes. Com base na absorbância característica da proteína na região espectral de 1800-1500 cm^{-1} por meio de espectroscopia FTIR foi possível a determinação do teor de proteína das amostras. O procedimento de seleção do comprimento de onda do algoritmo de projeção sucessiva juntamente com o modelo de retro propagação, foram as técnicas utilizadas que obtiveram bons resultados. Considerando a complexidade da amostra, o modelo de rede neural de múltiplas camadas com retro propagação se mostrou confiável por ser um método rápido e simples e sem necessidade de preparação de amostra.

Alguns outros trabalhos foram realizados com uso de rede neural MLP para fazer predição de fraudes em leite. Monteiro (2015) utilizou RNAs para a predição de fraude em leite fluido com adição de soro de leite. Os dados de entrada foram obtidos através de análise de espectroscopia no infravermelho (FTIR) enquanto os dados da camada de saída foram obtidos através da análise do caseínomacropéptídeo (CMP) por cromatografia líquida de alta eficiência. A utilização da RNA permitiu a quantificação de fraude por adição de soro ao leite, demonstrando um grande potencial de aplicação para amostras adulteradas. Contudo, o modelo utilizado no estudo tem pouca aplicabilidade prática uma vez que as análises usadas como camada de saída são complexas e caras.

Outro uso de RNAs para predição de adulterantes em leite cru foi realizado por Conceição (2018). No estudo foram testados adulterantes como bicarbonato de sódio, hidróxido de sódio, peróxido de hidrogênio, amido, ureia e sacarose. Todas as redes testadas obtiveram quinze neurônios na camada de entrada referente às absorbâncias obtidas nos espectros FTIR e dois neurônios na camada de saída (leite e leite adulterado). As RNAs apresentaram boas capacidades de predição para os adulterantes estudados, principalmente quando comparados à outras técnicas quimiométricas empregadas no trabalho. Utilizando amostras provenientes de produtores, foram capazes de realizar a predição de adulteração em 2,4% das amostras mesmo em baixas concentrações. Dessa forma foi possível afirmar que o FTIR aliado à análises de RNAs puderam obter resultados satisfatórios na identificação de adulteração a partir de 0,1%.

Alves *et al.* (2022) também analisaram o uso de redes MLP com fase de treinamento supervisionado e algoritmo de retropropagação, a fim de quantificar a fraude em leite em pó integral com adição de soro lácteo em pó. Os autores objetivaram determinar a melhor topologia de RNA para a predição da fraude utilizando como dados as análises físico-químicas de rotina por meio de espectroscopia FTIR. Foram testados três modelos. O primeiro utilizando dados de gordura, crioscopia, sólidos totais, extrato seco desengordurado, lactose, proteína e caseína. O

segundo modelo excluiu resultados de caseína devido ao alto custo e complexidade. O terceiro utilizou apenas resultados considerados de rotina em produtos lácteos e baixo custo (gordura, extrato seco desengordurado e sólidos totais). Como camada de saída todos os modelos utilizaram as porcentagens de fraude presente nas amostras. O experimento demonstrou grande potencial de uso para a indústria de alimentos, para controle de qualidade de leite em pó , principalmente, onde as possibilidades de fraude utilizando soro lácteo em pó podem estar presentes.

10.6. Treinamento das Redes Neurais Artificiais

Para que a RNA atenda aos objetivos propostos, ela deve ser treinada. Esse termo treinamento da rede consiste em fazê-la adotar valores de pesos e limiares baseados em suas amostras de entrada, de modo que qualquer amostra futura apresentada a ela seja corretamente classificada. Ou seja, serão apresentadas as amostras ou dados na entrada da rede e estes serão direcionados a uma resposta desejada, fazendo com que a rede seja obrigada a modificar os seus pesos e limiares. Dessa forma, qualquer outro valor apresentado a ela posteriormente ao final do treinamento, produzirá uma saída ou resultado igual ou próximo do desejado (Silva *et al.*; 2010).

Usualmente para treinar uma rede e verificar se os pesos estão adequados, é necessário dividir as amostras existentes em treinamento e validação da rede. Sugere-se que a divisão seja 60% a 90% amostras para treinamento e 40% a 10% para testes. A “época de treinamento” será, portanto, cada vez que os dados forem apresentados para ajuste dos pesos sinápticos e limiares da rede (Silva *et al.*, 2010).

Os tipos de treinamento podem ser divididos em:

1. **Supervisionado:** no qual deve-se dispor das amostras e das saídas desejadas para que os pesos e limiares sejam ajustados pelo algoritmo de aprendizagem.
2. **Não supervisionado:** nesse treinamento os dados não são rotulados, de forma que cabe ao algoritmo identificar particularidades dos conjuntos de amostras e os subconjuntos similares, os organizando na camada de saída. Os pesos e limiares são ajustados pelo algoritmo de aprendizagem de modo a refletir essas particularidades.

3. **Com reforço:** Similar ao supervisionado, porém seu algoritmo visa ajustar os pesos e limiares baseando-se em informações procedentes da interação com o sistema mapeado, visando reforçar respostas satisfatórias.

CAPÍTULO 1: QUALIDADE DO LEITE DE BÚFALA EM DIFERENTES ESTAÇÕES DO ANO

1.1. INTRODUÇÃO

A espécie bubalina ocupa um relevante papel na produção de alimentos nos países em desenvolvimento, localizados em sua maioria nas áreas tropicais. Assumem também um relevante papel no desenvolvimento social e econômico em muitos outros países. No Brasil, a produção de leite de búfala e seus derivados vêm ganhando importância (Gregory *et al.*, 2014), mesmo sem possuir, ainda, uma legislação federal específica que regularize o padrão de identidade e qualidade desse leite e seus derivados.

Todas as peculiaridades características do leite de búfala contribuíram fortemente para aumentar o interesse dos produtores pela bubalinocultura (Pereira, 2007). Dessa forma, o conhecimento da composição do leite e seu espectro de variação são essenciais para a determinação de sua qualidade, pois definem diversas características sensoriais e industriais (Dürr, 2004).

Dessa forma, objetivou-se avaliar a influência das estações do ano na composição e qualidade higiênico-sanitária do leite de búfalas exploradas no estado de Minas Gerais.

1.2. CARACTERÍSTICAS DO LEITE DE BÚFALA

Os búfalos são animais poliéstricos estacionais de dias curtos, dessa forma, apresentam maior atividade reprodutiva no período do outono e inverno. Em regiões próximas a linha do equador onde a diferença de fotoperíodo é discreta, o fator que mais influencia a reprodução é a disponibilidade de alimentos, que está diretamente relacionada ao período chuvoso do ano. Quando as búfalas são criadas em locais distantes da região equatorial, mesmo em boas condições de alimentação, apresentam comportamento reprodutivo influenciado positivamente pela diminuição de horas de luz do dia. Sendo assim, os partos concentram-se no primeiro semestre, sendo a entressafra da produção de leite na primavera e verão (Andrighetto, 2011).

Mesmo com sua facilidade de adaptação ao clima tropical, caracterizado por temperatura e umidade elevadas, a espécie necessita de condições favoráveis para expressar seu potencial produtivo, dispondo, portanto de condições básicas como sombra e água, a fim de

amenizar os efeitos causados pelo estresse calórico. O mesmo ocorre em regiões de baixas temperaturas, visto que a energia gasta para produção de leite é desviada para os mecanismos de termorregulação (Bernardes, 2007).

O leite de búfala apresenta alto valor nutricional, altos níveis de gordura, proteínas e minerais (cerca de 25% de cálcio), podendo ser utilizado tanto para o consumo *in natura* como matéria-prima para elaboração de produtos lácteos (Silva e Ribeiro, 2021). Jorge *et al.* (2011) relataram uma economia de 20 a 40% da matéria-prima na elaboração de produtos como iogurte, queijos, doce de leite e manteiga ao substituir o leite bovino pelo bubalino. Esta característica deve-se ao elevado teor de extrato seco total, além de características físicas como densidade e disposição coloidal dos constituintes do leite (Andrade *et al.*, 2011).

O leite de búfala apresenta níveis mais elevados de proteína total, ácidos graxos de cadeia média, teores de retinol e tocoferóis e ácido linoleico conjugado (CLA) do que os níveis do leite de vaca (Ahmad *et al.*, 2013). O CLA é um componente característico da gordura presente no tecido adiposo de ruminantes, o qual é formado no rúmen como intermediário da biohidrogenação do ácido linoléico isomerase. Estudos realizados com CLA demonstraram propriedades benéficas à saúde, como a redução da gordura corporal, efeitos anticancerígenos, como por exemplo a inibição do crescimento de células de câncer de colón e apoptose de células cancerígenas. Além disso, a suplementação com CLA melhora o índice de massa e peso corporal, redução da gordura abdominal e aumento de massa magra. (Gutiérrez-Tolentino *et al.*, 2022; Nörnberg *et al.*, 2022). Verruma e Salgado (1994) analisando a composição química do leite de búfala e comparando com o de bovinos verificaram que o leite bubalino apresentou maiores níveis dos principais constituintes, como gordura, proteína e sólidos totais.

Em relação à gordura, o leite de búfala é quase duas vezes mais rico quando comparado ao leite de vaca. O teor de gordura no leite de búfala normalmente se mantém acima de 5,5%, sendo essa uma característica particular da espécie. Esta é a fração mais importante, responsável por seu alto valor energético e nutritivo. Sob o ponto de vista nutricional, os lipídeos apresentam níveis apreciáveis de ácidos graxos essenciais ao organismo. A proporção de ácido graxo saturado/ácido graxo poli-insaturado é considerada nutricionalmente adequada. O leite de búfala possui o dobro de ácido linoléico conjugado (CLA) do que o leite de vaca. O CLA é um componente característico da gordura presente no tecido adiposo de ruminantes, o qual é formado no rúmen como intermediário da biohidrogenação do ácido linoléico isomerase (Gutiérrez-Tolentino *et al.*, 2022).

No que se refere à proteína, o conteúdo proteico do leite de búfala é maior que o de vaca (Verruma e Salgado, 1994; Ahmad *et al.*, 2008). Dentre as pesquisas que retratam os valores médios de proteína no leite de búfalas, os teores variam entre 3,8 e 4,5% (Duarte, 2001; Soares *et al.*, 2013). As micelas de caseína do leite de búfalas são maiores que as encontradas no leite de vaca fazendo com que a coalhada elaborada com leite de búfalas retenha menos água em relação a de leite de vacas, durante a ação do coalho (Ganguli, 1979). Na maioria das espécies mamíferas quatro diferentes proteínas compõem o grupo das caseínas, sendo elas α S1-caseína, α S2-caseína, β -caseína e κ -caseína. No leite de vaca, a β -caseína representa aproximadamente 30% do total das proteínas e, de acordo com a genética do animal, serão expressas no leite as variantes β -caseína A1 e/ou A2, originando a denominação leite A1 (havendo apenas β -caseína do tipo A1 ou uma mistura de β -caseína A1 e A2) e leite A2 (havendo apenas a β -caseína do tipo A2). O que diferencia essas duas variantes genéticas da β -caseína é a substituição de apenas um aminoácido na posição 67 dos 209 aminoácidos que compõem esta proteína. Por outro lado, o leite bubalino possui apenas a beta caseína A2 podendo ser consumido por pessoas com sensibilidade à beta-caseína A1 (Brooke-Taylor, 2017).

As proteínas do soro e as proteínas menores possuem níveis mais elevados no colostro que no leite de búfalas em estágios mais avançados de lactação. A fração de proteína sérica do leite contém quatro principais proteínas, β -lactoglobulina (50%), α -lactoglobulina (20%), soroalbumina (10%) e imunoglobulinas (10%), como igG1 (principalmente), IgG2, IgA e IgM (Fox *et al.*, 2000).

Apesar de não possuir teor maior de lactose em relação ao leite bovino, o leite de búfala apresenta sabor adocicado. O leite é um alimento que possui uma composição diversificada de nutrientes e muitos fatores influenciam na variação de suas características. Essas variações podem ser de origem genética, entre os animais, e diferenças não genéticas, que podem ser causadas por fatores ligados ao meio físico ou por funções de natureza fisiológica, recebendo ambas, a denominação geral de fatores de meio ou ambientais (Amaral *et al.*, 2005).

Muitos fatores podem influenciar a produção e a qualidade do leite de búfalas. Dentre esses, a mastite é o mais importante para a bubalinocultura leiteira. A mastite é uma inflamação da glândula mamária de origem multifatorial (Viguiet *et al.*, 2009), principalmente por infecção bacteriana por *Staphylococcus* spp., *Streptococcus* spp., *Corynebacterium* spp., e *Escherichia coli* (Patel *et al.*, 2019, Melo *et al.*, 2022). Vários fatores podem estar associados à ocorrência da mastite em búfalas como as más condições de manejo, e o tipo de micro-organismos no ambiente (Ali *et al.*, 2011). Portanto, o conhecimento desses fatores é essencial para melhorias

do controle dessa doença na espécie bubalina (Ali *et al.*, 2014). Estudos mostram que as búfalas têm maior predisposição à mastite subclínica do que à mastite clínica, podendo variar de 5,46% a 70,0%, e a incidência da doença está associada à vários fatores de risco como os efeitos ambientais e fisiológicos dos animais (Moura *et al.*, 2017; Melo *et al.*, 2022).

A contagem de células somáticas (CCS) é um instrumento preciso de avaliação da saúde da glândula mamária dos animais, tanto individualmente como de um rebanho. Altas contagens de células somáticas resultam em perdas na produção e alterações nos componentes individuais do leite em decorrência da mastite, com comprometimento sobre o rendimento, a qualidade de seus derivados e vida de prateleira (Santos e Fonseca, 2007). Por outro lado, raças, paridade, idade de parto, estágio de lactação, estação do ano, estresse, intervalo de ordenha e manejo afetam a CCS em leite de búfala (Muggli, 1995; Singh e Ludri, 2001).

A mastite pode contribuir para a alteração do percentual de sólidos totais, gordura, caseína e lactose, além disso, pode aumentar componentes inadequados para a fabricação de lácteos como glicogênio, pH e cloretos (Radostits *et al.*, 2006). Cerón-Muños *et al.* (2002) verificaram que animais com alta CCS tiveram redução na porcentagem de lactose. A redução da lactose no leite pode decorrer da passagem da lactose do leite para o sangue e pelo aumento da permeabilidade do tecido mamário e dos vasos sanguíneos, aumentando o extravasamento de hemocomponentes para o úbere e, conseqüentemente, alterando a composição do leite (Sharif *et al.*, 2007).

Outro importante fator a ser considerado são as condições de higiene as quais os animais são submetidos. As condições de higiene inadequadas no ambiente determinam a existência de animais mais sujos no rebanho. A pele do úbere e do teto no período entre as ordenhas fica exposta a lama e fezes, situação que acarreta maior risco de contaminação microbiana do leite. Os locais onde os animais são criados podem abrigar elevadas cargas microbianas, mesmo tendo aparência seca e limpa. O efeito final desta contaminação, ou seja, na contagem padrão em placas (CPP) do leite, vai depender do conjunto de procedimentos de limpeza e desinfecção realizados antes da ordenha (Santos e Fonseca, 2007).

Elevados valores para CPP no leite de búfalas são reflexo da higiene durante a ordenha, do armazenamento do leite e das deficiências no transporte como frequência de coleta, atrasos, exposição ao sol e elevadas temperaturas (Bailone *et al.*, 2017). O leite submetido a condições inadequadas de higiene, que favorecem a contaminação, e mesmo temperatura de armazenamento elevadas, passa por um processo de fermentação da lactose, causado pelos

micro-organismos, que resulta na sua acidificação devido à produção de ácido láctico. A indústria láctea utiliza a acidez do leite como indicativo da qualidade microbiológica para efetuar a compra da matéria-prima (leite) e garantir qualidade dos seus produtos derivados (Santos e Fonseca, 2007).

Sendo assim, as principais estratégias adotadas para diminuir a contaminação do leite estão relacionadas ao manejo adequado do ambiente, à ordenha e à limpeza e manutenção dos equipamentos de ordenha (Santos e Fonseca, 2007).

1.3. MATERIAL E MÉTODOS

1.3.1. Obtenção das amostras

Foram utilizados dados provenientes do banco de dados do Laboratório de Análise da Qualidade do Leite da Escola de Veterinária da Universidade Federal de Minas Gerais (LabUFMG). As amostras de leite bubalino analisadas foram coletadas mensalmente em propriedades produtoras diretamente do tanque de refrigeração. Foram coletados no total dados de 837 amostras de leite de búfala provenientes de rebanhos do estado de Minas Gerais entre os meses de junho/2020 a julho/2022 totalizando 24 meses.

O efeito das estações foi definido como: Primavera (23 de setembro a 21 de dezembro), Verão (21 de dezembro a 21 de março), Outono (21 de março a 21 de junho) e Inverno (21 de junho a 23 de setembro).

As amostras foram acondicionadas em frascos estéreis de 50mL. Para a análise de composição foram utilizados frascos com conservante Bronopol (2-bromo-2-nitro-1,3-propanodiol, natamicina) a serem analisados por espectroscopia FTIR. Para as análises de contagem bacteriana padrão em placas, foram utilizados frascos com conservante Azidiol (azida sódica e cloranfenicol) encaminhados à citometria de fluxo. Após a coleta, as amostras foram acondicionadas em caixas isotérmicas com gelo a fim de manter as amostras entre 2 e 10°C como determina a IN 76/2018 do Ministério da Agricultura pecuária e Abastecimento (MAPA) (Brasil, 2018) e enviadas para o laboratório.

1.3.2. Local de realização do experimento

As análises foram realizadas 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 de acordo com a norma ABNT NBR ISO/IEC 17025:2017.

1.3.3. Análise de Contagem de Células Somáticas

A contagem de células somáticas (CCS) do leite foi realizada no citômetro de fluxo do equipamento CombiScope™ FTIR, por meio do SomaScope (Delta Instruments, Drachten, Holanda) (ISO 13366/IDF 148 Part 2, 2006). Nessa análise, 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 do equipamento para essa análise foi feita utilizando-se três amostras padrão com CCS em diferentes níveis.

1.3.4. Análise de Composição do leite

A análise de composição foi realizada no equipamento CombiScope™ FTIR, por meio do LactoScope FTIR (Delta Instruments, Drachten, Holanda) (ISO 9622: 2013). As amostras foram aquecidas a 40°C antes da análise. Antes da entrada no aparelho, a amostra passa por homogeneização a fim de reduzir o diâmetro dos glóbulos de gordura e posteriormente recebe irradiação 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 equipamento é feita através do 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. Esse procedimento é realizado por meio de 14 amostras padrão de leite cru com diferentes faixas de composição para gordura, proteína, lactose e estrato seco total (EST) utilizando-se o programa de calibração do próprio equipamento. As amostras padrão foram fornecidas pelo laboratório Valacta (Dairy Production Centre of Expertise, Quebec, Canadá). Não existe no Brasil um padrão específico para calibração com leite bubalino, porém o padrão fornecido para a calibração do equipamento fornece uma variação ampla em valores de gordura e proteína que abrange os teores de componentes do leite de búfala.

1.3.5. Análise de Contagem Padrão em Placas (CPP)

As amostras para contagem bacteriana foram analisadas em duplicata no equipamento BactoScan FC® (Foss Analytics). Este equipamento realiza a contagem individual de bactérias como células únicas pelo método de citometria de fluxo, porém não realiza a diferenciação dos micro-organismos. O equipamento foi calibrado de acordo com o método de contagem padrão em placa de referência, definido pela norma Internacional (ISO 21187: 2021 [IDF 196:2021] ISO 16297: 2020[IDF 161:2020]), permitindo a conversão da contagem bacteriana total em contagem padrão, em atendimento à Instrução Normativa nº 77/2018 (Brasil, 2018). Antes de se passar a *rack* no equipamento, as amostras sofrem agitação garantindo que haverá homogeneidade na amostra a ser analisada.

1.3.6. Análise Estatística

As variáveis com distribuição normal dos dados foram analisadas pelo teste de Tukey e os resultados originais obtidos de CCS (contagem de células somáticas) e de CPP (unidades formadoras de colônias/mL) por equipamento eletrônico ou por contagem em placas analisados pelo teste de Kruskal Wallis. A análise estatística foi realizada com os softwares IBM SPSS Statistics (v. 22) e JMP (16.0; SAS Institute).

1.4. RESULTADOS E DISCUSSÃO

Os resultados das análises feitas por método de FTIR, geram arquivos de planilha em formato CSV com composição centesimal do leite (lactose, proteína, sólidos totais, sólidos não gordurosos, caseína), nitrogênio ureico do leite, crioscopia e CCS. Os resultados das análises FTIR e de citometria de fluxo foram computados e estão dispostos conforme as Tabela 1-1 e Figura 1-1.

Tabela 0-1: Teores médios e desvios padrão (DP) sazonais de gordura, lactose, proteína, sólidos totais (ST) e extrato seco desengordurado (ESD) de leite de búfala coletado mensalmente de tanques refrigeradores durante o período de julho/2020 a junho/2022 em Minas Gerais (n=837)

	Estação							
	Verão		Outono		Inverno		Primavera	
	Média	DP	Média	DP	Média	DP	Média	DP
Gordura (g/100g)	5.95 ^b	0.83	5.57 ^c	0.70	5.58 ^c	0.72	6.19 ^a	0.72
Proteína (g/100g)	4.24 ^a	0.31	3.87 ^c	0.32	3.74 ^d	0.34	4.16 ^b	0.32
Lactose (g/100g)	4.78 ^c	0.21	4.88 ^b	0.20	4.94 ^a	0.18	4.80 ^c	0.17
ST (g/100g)	15.98 ^a	0.88	15.27 ^b	0.85	15.23 ^b	0.90	16.14 ^a	0.84
ESD (g/100g)	10.03 ^a	0.32	9.70 ^b	0.34	9.65 ^b	0.35	9.96 ^a	0.28

a, b, c, d Médias dentro da mesma linha com diferentes letras, diferem estatisticamente ($p < 0,05$).

Fonte: A autora (2022)

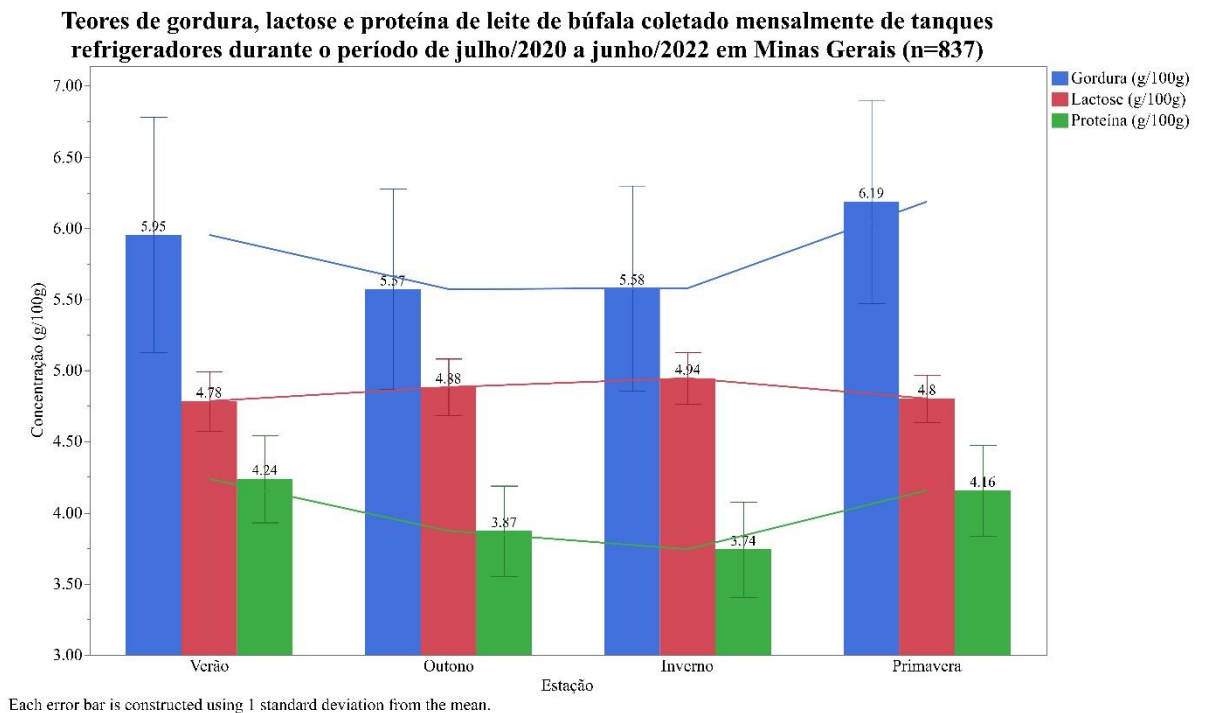


Figura 0-1: Teores de gordura, lactose e proteína de leite de búfala coletado mensalmente de tanques refrigeradores durante o período de julho/2020 a junho/2022 em Minas Gerais. Fonte: A autora (2022).

As médias de gordura encontradas no período avaliado demonstraram variação estatística entre as estações do ano. A maior média encontrada foi na estação da primavera com teor de 6,19%, seguida pelo verão com 5,95% e outono e inverno sem diferença estatística entre si com 5,57% e 5,58%, respectivamente. Esses resultados encontrados corroboram com os resultados encontrados por Amaral *et al.* (2004), que ao estudarem a variação sazonal da composição do leite bubalino em rebanhos de Minas Gerais, encontraram maior teor de gordura na primavera com teores de 7,59%, apesar de médias superiores ao encontrado no presente trabalho.

Por outro lado, Costa Filho *et al.* (2014) encontraram valores mais parecidos aos encontrados no presente estudo, com maior percentual no verão (6,00%), porém sem diferença estatística em relação aos valores encontrados na primavera (5,68%) e inverno (5,84%) em rebanhos do nordeste brasileiro. Vale ressaltar que a gordura se apresenta como o constituinte mais variável e ao ponto que assume valores acima de 5,5% está assumindo a característica particular da espécie bubalina (Costa Filho *et al.*, 2014).

Os teores de proteína variaram estatisticamente em todas as estações do ano. As maiores médias foram obtidas no verão (4,24%) seguidas por primavera (4,16%), outono (3,87%) e inverno (3,74%). Os resultados encontrados são semelhantes aos encontrados por Costa Filho *et al.* (2014) que apresentaram valores de proteína com maior teor no verão (4,28%), seguido pelo outono (4,22%), primavera (4,20%) e inverno (4,03%), somente apresentando variação estatística no inverno. Amaral *et al.* (2004) também encontraram maior teor de proteína na estação do verão (4,47%) com diferença estatística entre todas as estações.

O maior teor de proteína no verão pode ser devido ao estágio de início de lactação em que se encontram os animais (Cerón-Muñoz *et al.*, 2002; Amaral, 2005), uma vez que trabalhos anteriores já apontaram um aumento do teor de proteína no primeiro mês de lactação, com subsequente queda e aumento gradativo à medida que avança a lactação, além da provável melhor disponibilidade de alimento nesta estação (Patwardhan *et al.*, 1986).

A lactose apresentou maior teor percentual no inverno, com 4,94%, seguido pelo outono com 4,88%, primavera 4,80% e verão (4,78%). Esses valores corroboram aos encontrados por Amaral *et al.* (2005) que encontraram maiores porcentagens de lactose no inverno (4,97%) e menores na primavera (4,88%). A maior porcentagem de lactose no inverno pode estar vinculada a uma maior produção de leite nesse período ou maior disponibilidade de precursores do açúcar do leite.

A lactose do leite é proveniente da neoglicogênese via propionato (Martin e Sauvant, 2007) e animais suplementados com dietas ricas em sacarose, como a cana-de-açúcar, apresentam maior teor de propionato no fígado onde é metabolizado, em parte, a lactose. Dessa forma, fêmeas que são suplementadas com dietas ricas em sacarose apresentam maior teor de lactose no leite (Costa Filho *et al.*, 2014). O menor teor de lactose observado no período da primavera pode ser explicado pela menor produtividade de leite no final da lactação, uma vez que a lactose relaciona-se diretamente com o volume de leite produzido (Cerón-Muñoz *et al.*, 2002; Amaral, 2005).

Os teores de sólidos totais (ST) foram equivalentes durante os períodos de primavera e verão com maiores médias (16,14% e 15,98% respectivamente) e durante os períodos de outono e inverno com menores médias (15,27% e 15,23% respectivamente). Esses valores se encontram dentro da faixa descrita por Amaral *et al.* (2005) que indica a média entre 15% a 17% para o leite de búfalas. Os valores encontrados são semelhantes aos encontrados por Costa Filho *et al.* (2014) que relataram maiores médias no verão, com 16,03% e menores médias no outono (9,79%), com valores mais baixos em relação aos encontrados no presente estudo.

O aumento no teor de sólidos totais também pode ser vinculado ao efeito da estação na disponibilidade do pasto. O verão é um período chuvoso que pode influenciar diretamente o pasto em relação a qualidade nutricional refletindo na produção de leite das búfalas. Por outro lado, o outono e o inverno, abrange o período de seca prejudicando a qualidade e a quantidade de pasto disponível para as búfalas. Dessa forma, a variação no volume de leite produzido é o fator primário no comportamento do teor de ST do leite (Bezerra Júnior *et al.*, 2014).

Os sólidos totais têm especial importância no leite bubalino por se configurar em um dos principais parâmetros devido sua importância para indústria de lácteos, pois sua proporção no leite indica rendimento industrial no fabrico de queijos, iogurtes, e outros derivados lácteos (Araújo *et al.*, 2011). Pode-se inferir que médias baixas de sólidos totais no leite estão mais ligadas ao efeito de diluição, constatado com o elevado índice de produção.

É importante evidenciar também que o elevado percentual de gordura e proteína do leite de búfala o confere alto valor nutritivo, enquanto o elevado teor de sólidos favorece o alto rendimento industrial (Macedo *et al.*, 2001; Campanile *et al.*, 2007), e distingue decisivamente as qualidades físico-químicas do leite de bubalino, em relação ao leite bovino. A relação entre os sólidos totais e extrato seco desengordurado está evidenciada na Figura 1-2.

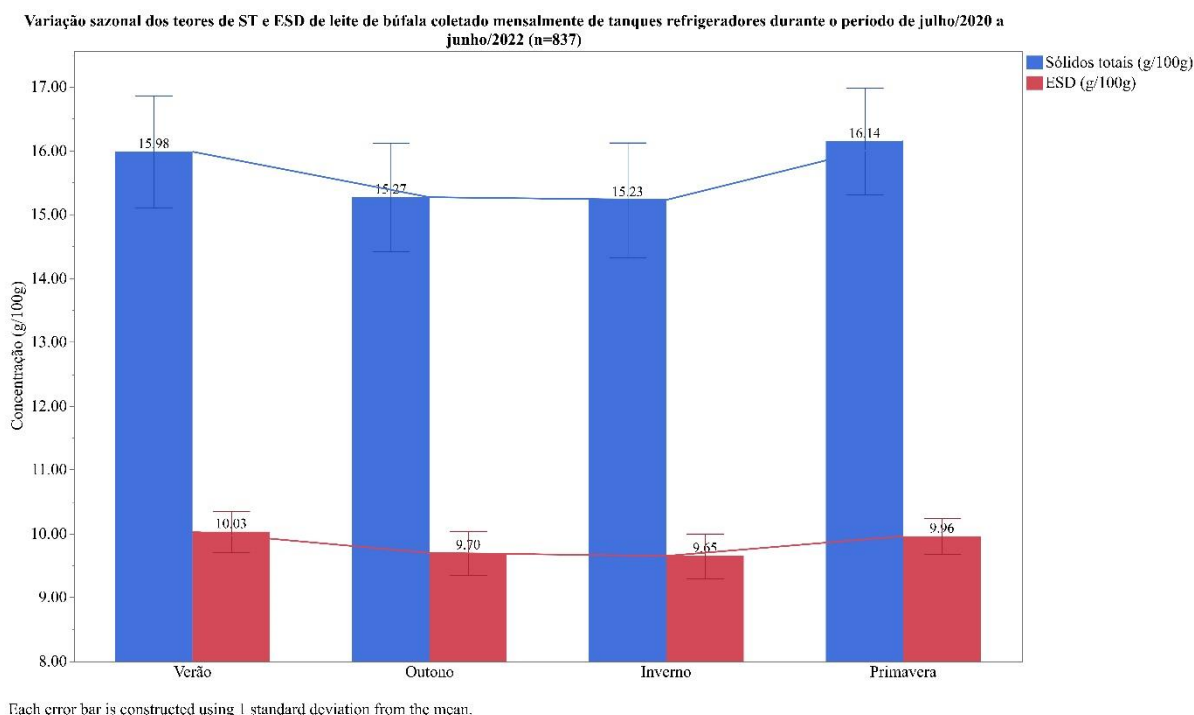


Figura 0-2: Variação sazonal dos teores de sólidos totais e extrato seco desengordurado do leite de búfalas coletado mensalmente de tanques refrigeradores durante o período de julho/2020 a junho/2022.

Fonte: A autora (2022)

De acordo com a Figura 1-2, o maior responsável pelo teor de sólidos totais é a gordura, que apresenta maiores porcentagens na primavera e no verão e menores no outono e no inverno. A maior concentração de gordura, ST e ESD no leite de búfalas durante a época seca pode ser atribuída a concentração desses componentes na glândula mamária devido a menor produção de leite dos animais durante a época seca do ano.

Os teores de extrato seco desengordurado foram mais elevados no verão e na primavera (10,03% e 9,96% respectivamente) e mais baixos no outono e no inverno (9,70% e 9,65% respectivamente). Esses valores corroboram aos achados por Costa Filho *et al.* (2014) que encontraram teores mais elevados no verão (10,02%) e mais baixos no inverno (9,79%). Os teores de lactose e, principalmente, os teores de proteína são parâmetros menos responsivos a variação nas condições ambientais e nutricionais (Figueiredo *et al.*, 2010). Provavelmente, maiores valores de ESD no verão estão relacionados com menor volume de leite produzido pelas búfalas, aumentando a concentração desse componente no leite (Araújo *et al.*, 2011).

O fato de as búfalas apresentarem o comportamento reprodutivo sazonal (Baruselli e Carvalho, 2002) com estação de monta concentrada no outono e consequente estação de parição

concentrada no verão pode explicar algumas alterações da composição do leite em função da fase da lactação em que se encontra o animal e a respectiva estação do ano (Amaral *et al.*, 2005).

Os resultados obtidos neste trabalho permitem inferir que os componentes do leite de búfalas variaram em função das estações do ano e isto, provavelmente se deve ao estágio da lactação em que se encontram os animais, uma vez que a espécie bubalina apresenta fator sazonal de reprodução, concentrando os partos no verão, além do fator nutricional relacionado com a disponibilidade de alimento em cada estação. Além disso, o fato de as búfalas serem suplementadas durante o período seco com cana-de-açúcar corrigido com ureia e, durante todo o ano, ser ofertado concentrado às búfalas leiteiras, pode trazer interferentes ao efeito da estação do ano, uma vez que a maior parte da variação na composição do leite entre as estações é de origem nutricional (Amaral *et al.*, 2004; Bastianetto, 2005; Lopes, 2009).

A contagem de células somáticas (CCS) e a contagem padrão em placas (CPP) durante o período está evidenciada na Tabela 1-2 e na Figura 1-3.

Tabela 0-2: Valores médios e desvios padrão (DP) sazonais de contagens de células somáticas (CCS), CCS médias geométricas trimestrais, contagens padrão em placas (CPP) e CPP médias geométricas trimestrais de leite de búfala coletado mensalmente de tanques refrigeradores durante o período de julho/2020 a junho/2022 em Minas Gerais (N=837)

	Estação							
	Verão		Outono		Inverno		Primavera	
	Média	DP	Média	DP	Média	DP	Média	DP
CCS (CS/mL*1000)	224 ^c	154	288 ^b	220	250 ^{bc}	204	341 ^a	252
CCS méd.geom.trim (CS/mL*1000)	250 ^b	141	273 ^{ab}	183	225 ^c	158	318 ^a	205
CPP (UFC/mL*1000)	346 ^a	1156	240 ^a	934	177 ^b	659	198 ^a	765
CPP méd.geom.trim (UFC/mL*1000)	113 ^a	144	115 ^a	159	69 ^c	143	77 ^b	81

Fonte: A autora (2022)

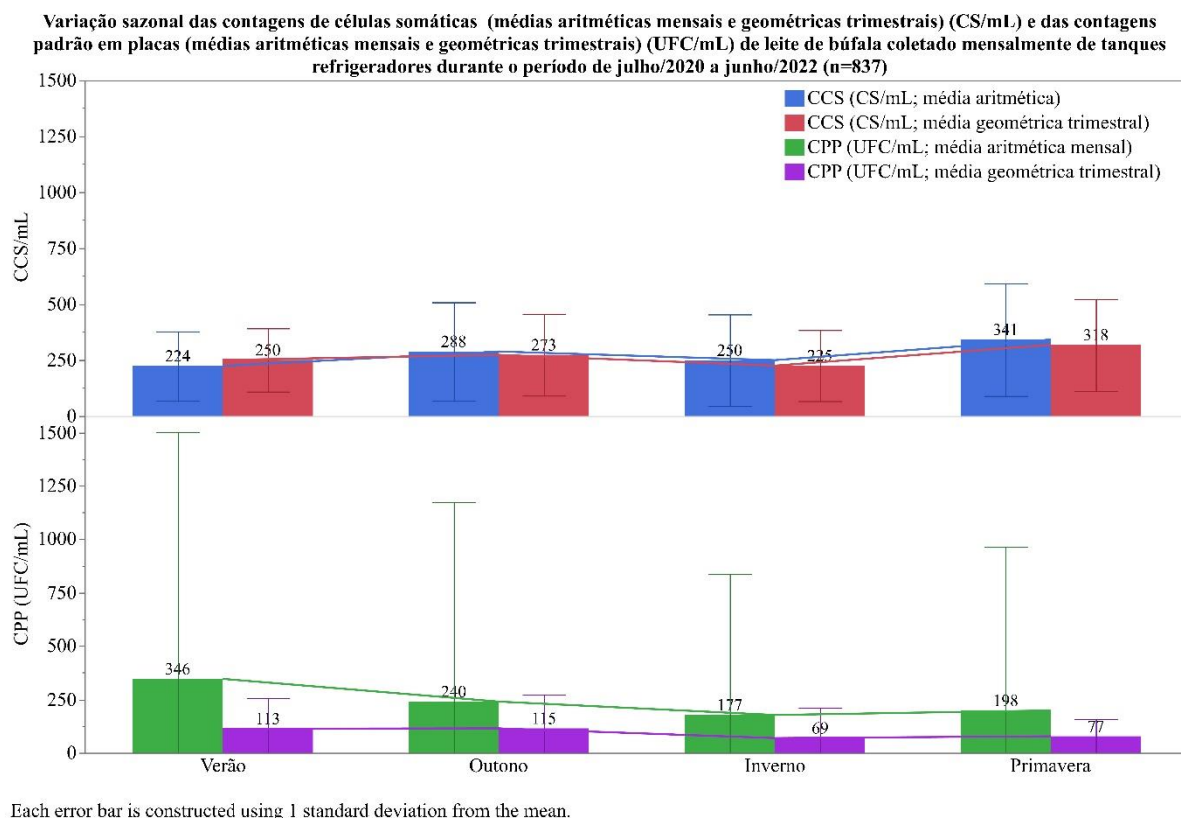


Figura 0-3: Variação sazonal das contagens de células somáticas (médias aritméticas mensais e geométricas trimestrais) (mil CS/ml) e das contagens padrão em placas (médias aritméticas mensais e geométricas trimestrais) (mil UFC/ml) de leite de búfala coletadas mensalmente de tanques refrigeradores durante o período de julho/2020 a junho/2022.

Fonte: A autora (2022)

A CCS apresentou-se elevada durante todas as estações do ano. As maiores médias aritméticas e geométricas foram registradas na estação da primavera, com 341 mil CS/mL e 318 mil CS/mL respectivamente. Em seguida as estações do outono e inverno sem diferença estatística entre si, e o verão com as menores contagens (288 mil CS/mL, 224 mil CS/mL e 250 mil CS/mL). Costa Filho *et al.* (2014) encontraram alta contagem de células somáticas durante todas as estações do ano e maior média no outono (350 mil CS/mL), período onde se concentra o maior volume de chuvas no agreste do Rio Grande do Norte. Por outro lado, Amaral *et al.* (2004) verificaram maior contagem de células no verão.

Na região sudeste do Brasil, principalmente no estado de Minas Gerais, o período de chuvas inicia-se na primeira quinzena de outubro com duração até o mês de março, o que compreende as estações da primavera e do verão. O excesso de umidade e alta temperatura ambiente criam condições favoráveis para maior infecção e prevalência de mastite nos rebanhos (Amaral *et al.*, 2005; Costa Filho *et al.*, 2014). O efeito estacional não deve ser considerado

como causa principal da variação da CCS, pois, o que ocorre é resultante do aumento da contaminação bacteriana dos tetos durante períodos em que as condições de crescimento microbiano são mais favoráveis e circunstâncias em que fatores contaminantes não são evitados por boas práticas de manejo (Amaral *et al.*, 2005).

A legislação brasileira não padroniza a CCS para o leite de búfala, sendo assim muitos trabalhos consideram outras referências como base para a classificação da contagem. Segundo Piccinini *et al.* (2006), contagens de até 200.000 CS/mL indicam glândulas mamárias sadias, entre 200.000 e 400.000 é considerado moderado e a partir de 400.000 CS/mL de leite, considera-se que o animal esteja acometido por mastite subclínica. Silva e Silva (1994), descreveram o intervalo de 50.000 a 375.000 CS/mL como sendo o esperado para o leite de búfalas sadias e em todas as estações do ano as médias encontradas estiveram dentro do intervalo esperado.

O uso da média geométrica tem sido recomendado em bovinos para a análise tanto da CCS de amostras analisadas individualmente quanto da CCS do leite do tanque de rebanhos leiteiros, sendo aprovada em diversos países em termos de legislação. Alguns países como os EUA, a média geométrica da CCS do tanque de três meses consecutivos é usada como medida regulatória para o limite máximo de CCS do leite a ser comercializado (Santos, 2002).

Quanto aos valores de CPP permitidos atualmente, a legislação para leite bovino (IN 76 do MAPA), determina médias geométricas trimestrais máximas de 300.000 UFC/mL (Brasil, 2018), sendo que contagens acima do permitido pela legislação podem acarretar inclusive na suspensão da coleta do leite da propriedade até que a assistência técnica consiga normalizar os valores. É importante ressaltar a ausência de padrões regulatórios específicos para o leite de búfala, principalmente em relação aos parâmetros de qualidade higiênico-sanitária e padrão de identidade e qualidade.

Apesar das altas médias aritméticas de CCS na estação da primavera, essa estação apresentou as segundas menores médias de contagem padrão em placas (CPP), com média geométrica de 198 UFC/mL, não possuindo diferença estatística com as estações de verão (346 UFC/mL) e outono (240 UFC/mL). As médias geométricas trimestrais de CPP são utilizadas como parâmetro de qualidade higiênico-sanitárias pelos laticínios. Devido à ausência de parâmetros específicos para o leite de búfala, os padrões para leite bovino são utilizados para a comparação. Todas as estações apresentaram média geométrica abaixo de 300.000 UFC/mL (Brasil, 2018).

Costa Filho *et al.* (2014) encontraram valores de CPP com menor contagem no verão (135mil UFC/mL) e maior contagem no outono (449 mil UFC/mL) onde se concentram grande volume de chuvas. Esses resultados corroboram com a descrição de Cunha Neto (2003) que cita que a presença de micro-organismos no leite de búfala *in natura*, de acordo com a estação do ano, pode resultar em valores entre $5,0 \times 10^4$ a $1,3 \times 10^3$ UFC/mL no inverno, e $1,5 \times 10^5$ a $3,2 \times 10^7$ UFC/mL no verão.

Apesar de não estarem relacionados com a qualidade intrínseca do leite, o volume e a sazonalidade de produção são critérios considerados para o pagamento do produto. Interessa aos laticínios captar leite junto aos produtores que forneçam grandes volumes diários de leite e que apresentem pequena variação sazonal da produção. Isso representa uma diluição nos custos operacionais e de transporte, além de uma melhor logística para recolhimento do produto. A pequena variação sazonal proporciona um melhor planejamento por parte da indústria e a minimização da ociosidade do parque industrial em determinadas épocas do ano (Fonseca, 2001)

1.5. CONCLUSÃO

O leite de búfala sofre influência da estação do ano principalmente no teor de sólidos totais, gordura e extrato seco desengordurado tendo aumento no período de primavera e verão onde se concentram a estação de chuva no estado de Minas Gerais. A contagem de células somáticas foi maior na primavera e a contagem padrão em placas apresentou maiores índices no verão, estações onde são concentrados os maiores volumes pluviométricos possibilitando maior chance de infecções mamárias.

CAPÍTULO 2: ANÁLISE FTIR DO LEITE BUBALINO ACRESCIDO DE DIFERENTES NÍVEIS DE LEITE BOVINO E ANÁLISE ESTATÍSTICA DOS RESULTADOS

2.1. INTRODUÇÃO

A demanda nacional por grandes volumes de leite de boa qualidade faz com que a pecuária leiteira seja bastante competitiva. Assim, é necessário quantificar e conhecer os parâmetros que podem interferir nos resultados obtidos na análise da qualidade do leite, diagnosticando os pontos que devem ser corrigidos, a fim de gerar ganhos efetivos na atividade (Dantas, *et al.*, 2020).

No Brasil, diversos laticínios têm focado na produção dos derivados de leite de búfala. Isto deve-se, principalmente, aos estudos realizados quanto às potencialidades produtivas encontradas nesses animais, as quais favoreceram a expansão de sua atividade em várias regiões do país (Santini *et al.*, 2013).

Devido a sazonalidade reprodutiva, os produtores de leite bubalino podem encontrar dificuldades no fornecimento à indústria e em alguns casos ocorrem fraude por adição de leite bovino ao bubalino visando o aumento do volume. Essa prática leva à prejuízos tanto ao consumidor quanto à indústria e sua identificação deve ser feita da forma rápida e eficaz.

Por outro, o mercado consumidor tem priorizado, cada vez mais, a certificação de segurança e qualidade nutricional dos alimentos adquiridos e consumidos em suas residências e avaliando, principalmente, se este produto pode trazer mais benefícios à saúde quando comparado com outras mercadorias presentes no setor comercial. Dessa forma, a certificação garante não somente a qualidade nutricional, mas também a autenticidade do produto, dando a certeza ao consumidor que o produto adquirido não foi adulterado (Dantas, *et al.*, 2020).

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

O objetivo desse capítulo foi analisar as amostras de leite bovino e bubalino sem adição e as amostras com misturas de leite bovino e bubalino em diferentes porcentagens quanto à composição e à contagem de células somáticas para realizar a comparação dos resultados nas diferentes espécies e a interferência da mistura nos resultados.

2.2. MATERIAL E MÉTODOS

2.2.1. 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 de acordo com a norma ABNT NBR ISO/IEC 17025:2017.

2.2.2. Obtenção de amostras

Foram utilizadas cerca de 600 amostras de leite de rebanhos bovinos (300 amostras) e bubalinos (300 amostras) de fazendas no estado de Minas Gerais que foram encaminhadas para a análise de rotina no LabUFMG entre os meses de outubro/2021 e março/2022. Após a coleta, as amostras foram acondicionadas em frascos estéreis de 50mL em caixas isotérmicas com gelo a fim de manter as amostras entre 2 e 10°C como determina a IN 76/2018 do MAPA (Brasil, 2018) e enviadas para o laboratório.

As amostras de leite bubalino foram obtidas de tanques de refrigeração de fazendas no estado de Minas Gerais, em frascos com conservante Bronopol (2-bromo-2-nitro-1,3-propanodiol, natamicina) para análise de composição por espectroscopia FTIR. As amostras foram analisadas individualmente e posteriormente foram formados os *pools* para a realização das adulterações. As amostras de leite bovino foram obtidas seguindo o mesmo critério de conservação.

Após a chegada no laboratório, as amostras foram encaminhadas à análise de composição e contagem de células somáticas.

2.2.3. Análise de composição centesimal do leite

A análise de composição foi realizada no equipamento CombiScope™ FTIR, por meio do LactoScope FTIR (Delta Instruments, Drachten, Holanda) (ISO 9622: 2013). As amostras foram aquecidas a 40°C antes da análise. Antes da entrada no aparelho, as amostras passaram

por homogeneização afim de reduzir o diâmetro dos glóbulos de gordura e posteriormente receberam irradiação de luz infravermelha em uma cubeta. A diferença de energia absorvida entre a amostra analisada e a amostra de referência foi 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 equipamento foi feita através do ajuste do sinal instrumental de modo que, a cada nível do componente em quantificação, a média de resultados individuais foram dados pelo instrumento o mais próximo do valor verdadeiro da concentração do componente. Esse procedimento foi realizado por meio de 14 amostras padrão de leite cru com diferentes faixas de composição para gordura, proteína, lactose e extrato seco total (EST) utilizando-se o programa de calibração do próprio equipamento. As amostras padrão foram fornecidas pelo laboratório Valacta (Dairy Production Centre of Expertise, Quebec, Canadá). Não existe no Brasil um padrão específico para calibração com leite bubalino, porém o padrão fornecido para a calibração do equipamento fornece uma variação ampla em valores de gordura e proteína que abrange os teores de componentes do leite de búfala.

2.2.4. Análise da contagem de células somáticas

A contagem de células somáticas (CCS) do leite foi realizada no citômetro de fluxo do equipamento CombiScope™ FTIR, por meio do SomaScope (Delta Instruments, Drachten, Holanda) (ISO 13366/IDF 148 Part 2, 2006). Nessa análise uma alíquota da amostra, pré-aquecida a 40° C, foi 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 foram conduzidos por um fluido carreador para o citômetro de fluxo, onde receberam incidência de raio laser. A luz emitida passou 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 representou o número de células no eixo vertical e o tamanho das células no eixo horizontal (Delta, 2007).

A calibração do equipamento para essa análise foi feita utilizando-se três amostras padrão com CCS em diferentes níveis.

Após a análise individual das amostras, os frascos foram utilizados para a formação das adulterações.

2.2.5. Pools de leite com misturas de diferentes espécies.

Após a análise de composição das amostras individuais, os *pools* de leite de búfala foram formados com leite de aproximadamente 10 fazendas diferentes cada um, totalizando 500mL de amostra. Paralelamente, foram formados *pools* com leite bovino (500 mL) com as amostras da rotina que foram analisadas no Laboratório de Qualidade do Leite, escolhidas aleatoriamente. Estas amostras eram amostras de composição normais, sem nenhum tipo de adulteração.

A partir desses pools foram feitas as adulterações do leite bubalino com adição de leite bovino seguindo as porcentagens de fraude, ou seja, quantidade de leite bovino adicionado: 0%, 1%, 2%, 5%, 10%, 20%, 30%, 40%, 50%, 75%, 100%, onde 0% correspondia a leite bubalino sem adulteração e 100% leite bovino sem adulteração. Posteriormente as amostras foram novamente analisadas por espectroscopia FTIR como demonstrado na Figura 2-1.

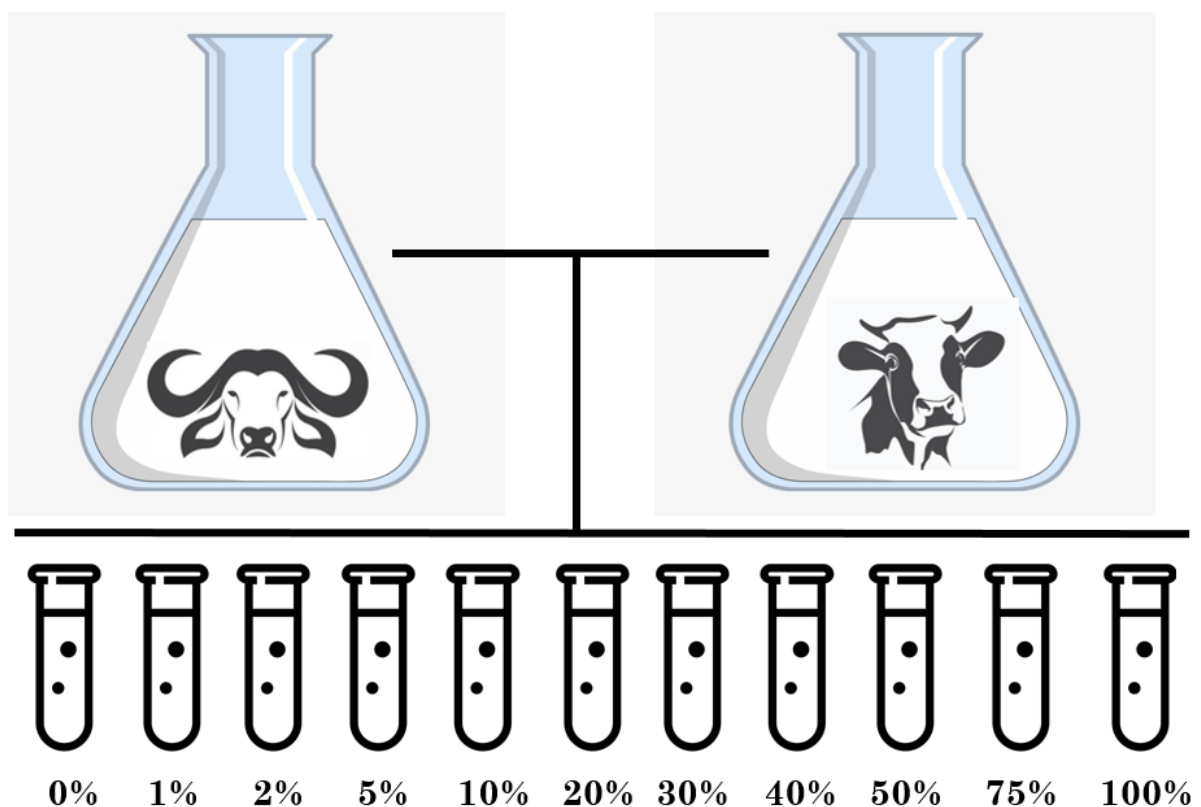


Figura 0-1: Demonstração da adulteração experimental das amostras de leite de acordo com as porcentagens de leite bovino adicionadas ao leite búfalo, utilizando amostras de leite bovino e búfalo sem adição obtidas de tanques refrigeradores de unidades produtoras do estado de Minas Gerais no período de outubro/2021 a março/2022. A amostra 0% corresponde ao leite

bubalino sem adição e a 100% leite bovino sem adição. As demais porcentagens se referem ao percentual de leite bovino na amostra.

Fonte: A autora (2022)

As amostras provenientes da adulteração foram novamente encaminhadas à análise de contagem de células somáticas e composição e os resultados registrados para a análise estatística.

2.2.6. Formação do banco de dados

Para cada amostra analisada, o equipamento FTIR gerou dois arquivos com dados distintos. Um arquivo foi referente ao espectro infravermelho, no formato SPC, com as coordenadas do espectro de absorção do infravermelho lido e outro arquivo continha a composição no formato CSV, com variáveis numéricas que compõe os *features* (atributos) de composição do leite. Os *features* foram calculados utilizando o espectro, a partir da PLS (*Partial Least Squares*) como método de calibração, para determinação dos valores da composição centesimal do leite (gordura, proteína, lactose, sólidos totais, sólidos não gordurosos, nitrogênio ureico do leite e caseína) (Delta, 2007).

Para este trabalho, foram utilizadas as informações provenientes das coordenadas espectrais, que foram analisadas e condensadas em um único conjunto de dados no formato CSV. Os dados utilizados da composição centesimal do leite foram os referentes aos teores de gordura, proteína, lactose, sólidos totais e sólidos não gordurosos.

Os dados foram compilados em um banco de dados geral, utilizando informações de amostras puras de leite bovino e bubalino proveniente do banco de dados do Laboratório de Qualidade do Leite no período de julho/2020 a junho/2022 e as amostras fraudadas experimentalmente no período de novembro/2021 a março/2022, totalizando 13.018 amostras. A demonstração do banco de dados está representada na Figura 2-2.

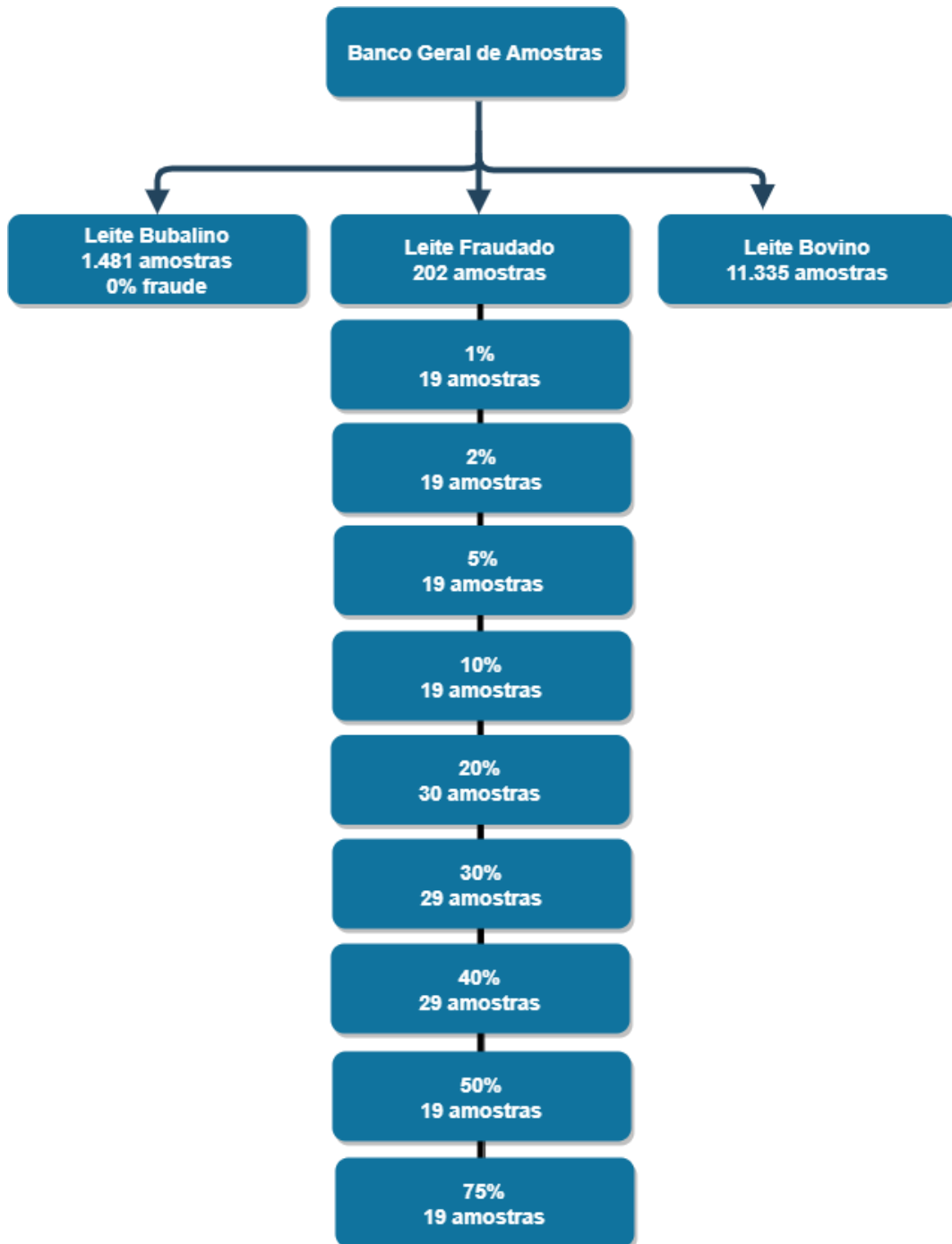


Figura 0-2: Composição do banco de dados a partir dos resultados das análises FTIR realizadas em amostras de leite bovino e bubalino obtidas no período de 2020-2022 em tanques refrigeradores de unidades produtoras no estado de Minas Gerais, e mistura de leite bovino em leite bubalino em diferentes porcentagens simulando fraude

2.2.7. Análise estatística

As análises estatísticas complementares incluindo a estatística descritiva e análises multivariadas foram realizadas com o software IBM SPSS Statistics (v. 22) para a diferenciação das amostras de leite de búfala das amostras fraudadas. O teste de Tukey foi utilizado como comparação post-hoc nos tratamentos com o erro tipo 1 controlado no nível de significância 5%.

2.3. RESULTADOS E DISCUSSÃO

Os resultados das análises feitas por método de FTIR, geram arquivos de planilha em formato CSV com composição centesimal do leite (lactose, proteína, sólidos totais, sólidos não gordurosos, caseína), nitrogênio ureico do leite e CCS. Os resultados foram utilizados para comparação entre os leites de vaca e búfala, a fim de se encontrar as diferenças de composição do leite das duas espécies e os efeitos de diluição que ocorrem nos casos de fraude por mistura de leite. Os resultados obtidos no banco de dados completo foram analisados e tabelados conforme a Tabela 2-1.

Tabela 0-1: Estatística descritiva da composição do leite de vacas e de búfalas obtidos de tanques de refrigeração em unidades produtoras de MG durante o período de 2020-2022 e analisadas por FTIR no Laboratório de Análise da Qualidade do Leite da Escola de Veterinária da UFMG.

Leite		CCS (1000 cels/ml)	Gordura (g/100g)	Proteína (g/100g)	Lactose (g/100g)	Sólidos Totais (g/100g)	Sólido Não- Gordurosos (g/100g)
Vaca	Média	652	3,75	3,31	4,47	12,52	8,78
	Desvio Padrão	516	0,37	0,18	0,13	0,478	0,226
	Mediana	518	3,76	3,30	4,48	12,52	8,78
	N	11.334	11.336	11.336	11.336	11.336	11.336
Búfala	Média	245	5,82	4,03	4,84	15,67	9,96
	Desvio Padrão	204	0,78	0,37	0,21	0,92	0,49
	Mediana	188,	5,77	4,05	4,86	15,65	9,90
	N	1.480	1.481	1.481	1.481	1.481	1.481

Pelas análises de composição observou-se que o leite bubalino apresenta médias de gordura, proteína, lactose, sólidos totais e sólidos não gordurosos superiores ao leite bovino.

Figueiredo *et al.* (2010) analisaram o leite bubalino em diferentes propriedades do estado do Pará em relação às variáveis físico-químicas. Os valores encontrados na análise de gordura foram de 8,14%, teor superior aos encontrados no presente estudo nos rebanhos de Minas Gerais (5,81%). Outros trabalhos encontraram valores mais baixos em relação aos rebanhos bubalinos estudados, como Neves (2002), Macedo *et al.* (2001) e Santos (2022) que encontraram valores de 6,0% e 6,59% e 6,25% respectivamente. Alguns trabalhos encontraram valores mais próximos aos encontrados, como Costa Filho *et al.* (2014) de 5,56% e Pignata *et al.* (2014) de 4,26%. Essas diferenças podem ocorrer por influências de fatores ambientais, como estação do ano e nutrição, além de efeitos intrínsecos ao animal, como raça, idade e estágio de lactação (Amaral *et al.* 2005)

Em relação à variável proteína, a média encontrada no leite de búfala (4,03%) foi semelhante aos achados em trabalhos realizados em rebanhos bubalinos por Zen *et al.* (2006), de 4,40%, por Neves (2002), de 4,26%, por Macedo *et al.* (2001), de 4,23% e Santos (2022), de 4,49%. Essa variável também sofre influência direta de fatores ambientais e nutricionais, como alimentação com diferentes fontes de nitrogênio e disponibilidade de energia.

A lactose apresentou menor variação em relação à comparação entre o leite bovino e o leite bubalino com 4,47% e 4,84% respectivamente. Figueiredo *et al.* (2010), Neves (2002) encontraram porcentagens mais baixas de lactose em leite bubalino com 4,55% e 4,33% respectivamente. Santos (2022) encontrou 4,96% na análise de lactose em leite bubalino, se assemelhando aos resultados encontrados no presente estudo.

Os sólidos totais especialmente no leite bubalino têm grande importância principalmente por se configurar um dos principais parâmetros para a indústria de laticínios. A proporção dessa variável indica o maior rendimento industrial na fabricação de derivados como queijos, iogurtes e outros produtos. O leite de búfala é mais concentrado em relação ao leite bovino, com maiores teores de sólidos e menor teor de umidade (Araújo *et al.* 2011; Santos, 2022). Os resultados encontrados nas análises dos rebanhos em Minas Gerais foram 12,52% em leite bovino e 15,56% em leite bubalino, mostrando o maior teor do leite de búfala. Esses valores em leite bubalino são semelhantes aos encontrados por Santos (2022) que encontrou 16,81% em rebanhos no norte do Brasil e aos encontrados por Macedo *et al.* (2001), 17,01% em rebanhos do Sudeste.

O extrato seco desengordurado é um parâmetro menos variável comumente usado para nivelar o teor de sólidos do leite de rebanhos em diferentes condições de manejo por desconsiderar a variação de gordura no leite (Araújo *et al.* 2011). Os valores encontrados para leite bovino e bubalino foram 8,77% e 9,95%, respectivamente, mostrando que para esse parâmetro o leite de búfala também apresentou maiores porcentagens. Outros trabalhos também encontraram valores semelhantes para essa variável. Figueiredo *et al.* (2010) encontraram 9,87%, Costa Filho *et al.* (2014) 9,87% e Santos (2022) 10,55%.

Em relação à contagem de células somáticas (CCS), o leite bubalino apresentou menores quantidades em relação ao leite bovino. As búfalas, possuem características anatômicas do úbere e tetos e sistema de defesa atuante na glândula mamária que conferem maior resistência contra a mastite, quando comparados com bovinos (Sollecito *et al.*, 2011). Apesar disso, as mesmas técnicas de manejo e criação indicadas para os bovinos devem ser aplicadas aos rebanhos bubalinos.

O valor médio encontrado para CCS no leite bovino foi 652 mil/mL enquanto para o leite bubalino a média foi menor, com 245 mil/mL, sendo compatível com o preconizado para um leite de qualidade. Os baixos valores encontrados indicam um estado adequado de saúde e higidez da glândula mamaria das búfalas estudadas. Os resultados obtidos são superiores aos encontrados por Santos (2022) que obteve média de 121,1 mil CS/mL, porém está de acordo com o descrito por Silva e Silva (1994) que destacaram o intervalo de 50.000 e 375.000 CS/mL como sendo o esperado para o leite de búfalas sadias.

Quando analisadas as amostras fraudadas com mistura dos leites de diferentes espécies, o resultado observado demonstrou efeito de diluição nos componentes. À medida que se aumenta a porcentagem de leite bovino adicionado à amostra, ocorre uma redução no teor dos componentes quando comparado à amostra de leite bubalino sem adição. Os resultados da análise FTIR das amostras fraudadas foram analisados e estão dispostos na Tabela 2-2.

Tabela 0-2: Estatística descritiva de componentes do leite em cada um dos níveis de mistura de leite e em amostras de leite bovino e bubalino sem adição, obtidos de tanques de refrigeração de fazendas no estado de MG no período de 2020-2022 e analisadas por FTIR no Laboratório de Análise da Qualidade do Leite da Escola de Veterinária da UFMG.

Parâmetro		Média	Desvio	N
Gordura	Vaca	3,75	0,37	11,335
	25% Buf+75% Vaca	4,04	0,30	19
	50% Buf+50% Vaca	4,65	0,26	19
	60% Buf+40% Vaca	4,89	0,25	19
	70% Buf+30% Vaca	5,12	0,24	19
	80% Buf+20% Vaca	5,29	0,37	30
	90% Buf+10% Vaca	5,57	0,22	29
	95% Buf+5% Vaca	5,67	0,24	29
	98% Buf+2% Vaca	5,79	0,27	19
	99% Buf+1% Vaca	5,82	0,28	19
	Búfala	5,81	0,78	1,481
	Total	4,01	0,80	13,018
Proteína	Vaca	3,30	0,17	11,335
	25% Buf+75% Vaca	3,51	0,06	19
	50% Buf+50% Vaca	3,75	0,10	19
	60% Buf+40% Vaca	3,85	0,11	19
	70% Buf+30% Vaca	3,95	0,13	19
	80% Buf+20% Vaca	3,98	0,21	30
	90% Buf+10% Vaca	4,10	0,16	29
	95% Buf+5% Vaca	4,16	0,16	29
	98% Buf+2% Vaca	4,21	0,19	19
	99% Buf+1% Vaca	4,22	0,20	19
	Búfala	4,03	0,36	1,481
	Total	3,39	0,32	13,018
Lactose	Vaca	4,47	0,13	11,335
	25% Buf+75% Vaca	4,58	0,05	19
	50% Buf+50% Vaca	4,65	0,05	19
	60% Buf+40% Vaca	4,68	0,05	19
	70% Buf+30% Vaca	4,71	0,06	19
	80% Buf+20% Vaca	4,75	0,08	30
	90% Buf+10% Vaca	4,80	0,08	29
	95% Buf+5% Vaca	4,83	0,09	29
	98% Buf+2% Vaca	4,79	0,09	19
	99% Buf+1% Vaca	4,79	0,09	19
	Búfala	4,84	0,20	1,481
	Total	4,51	0,18	13,018
Sólidos Totais	Vaca	12,54	0,47	11,335
	25% Buf+75% Vaca	13,14	0,35	19
	50% Buf+50% Vaca	14,05	0,32	19
	60% Buf+40% Vaca	14,41	0,32	19

Parâmetro		Média	Desvio	N
	70% Buf+30% Vaca	14,77	0,32	19
	80% Buf+20% Vaca	15,05	0,61	30
	90% Buf+10% Vaca	15,48	0,29	29
	95% Buf+5% Vaca	15,66	0,30	29
	98% Buf+2% Vaca	15,76	0,38	19
	99% Buf+1% Vaca	15,80	0,38	19
	Búfala	15,66	0,91	1,481
	Total	12,92	116,92	13,018
Sólidos Totais	Vaca	8,77	0,22	11,335
não gordurosos	25% Buf+75% Vaca	9,75	0,17	19
	50% Buf+50% Vaca	10,12	0,17	19
	60% Buf+40% Vaca	10,27	0,17	19
	70% Buf+30% Vaca	10,43	0,17	19
	80% Buf+20% Vaca	10,49	0,24	30
	90% Buf+10% Vaca	10,67	0,17	29
	95% Buf+5% Vaca	10,77	0,16	29
	98% Buf+2% Vaca	10,83	0,19	19
	99% Buf+1% Vaca	10,85	0,19	19
	Búfala	9,95	0,48	1,481
	Total	89,39	0,50	13,018

Para efeito de comparação entre as diluições, foi feito o quadro de análise de variância como demonstrado na Tabela 2-3.

Tabela 0-3: Quadro de ANOVA relativo à comparação entre as análises de leite bovino e bubalino obtidos de tanques de refrigeração de fazendas no estado de MG no período de 2020-2022 e analisadas por FTIR no Laboratório de Análise da Qualidade do Leite da Escola de Veterinária da UFMG.

			Soma dos Quadrados	df	Quadrado Médio	F	Sig.
SCC * fraude	Entre Grupos	(Combinado)	216,675,566.24	1	216,675,566.24	899.538	.000
	Dentro de Grupos		3,086,082,221.55	12,812	240,874.35		
	Total		3,302,757,787.79	12,813			
Gordura * fraude	Entre Grupos	(Combinado)	5,576.17	1	5,576.17	29,037.53	.000
	Dentro de Grupos		2,460.90	12,815	.192		
	Total		8,037.08	12,816			
Proteína * fraude	Entre Grupos	(Combinado)	69.444	1	694.443	16,084.67	.000
	Dentro de Grupos		55.327	12,815	.043		
	Total		1,247.72	12,816			
Lactose * fraude	Entre Grupos	(Combinado)	184.897	1	184.897	9,179.50	.000
	Dentro de Grupos		258.125	12,815	.020		
	Total		443.022	12,816			
Sólidos Totais * fraude	Entre Grupos	(Combinado)	12,931.929	1	12,931.92	43,286.13	.000
	Dentro de Grupos		3,828.539	12,815	.299		
	Total		16,760.469	12,816			
Sólidos totais não gordurosos * fraude	Entre Grupos	(Combinado)	1,822.040	1	1,822.04	25,115.57	.000
	Dentro de Grupos		929.680	12,815	.073		
	Total		2,751.719	12,816			

df: *degrees of freedom* (graus de liberdade); Sig (valor de p)

Fonte: A autora (2022)

Segundo a análise de ANOVA, o valor $p < 0,001$ indica que nas médias entre os grupos analisados, pelo menos duas médias vão apresentar valores diferentes entre si. Para comparação do conjunto de médias foi realizado um teste de comparações múltiplas (post-hoc) de Tukey (Tabela 2-4).

Tabela 0-4: Comparação entre amostras de leites sem adição bovino e bubalino obtidos de tanques de refrigeração de fazendas no estado de MG no período de 2020-2022 e misturas de leite em diferentes porcentagens de adição de leite bovino em leite bubalino, em relação às variáveis de composição do leite obtidas por análise FTIR no Laboratório de Análise da Qualidade do Leite da Escola de Veterinária da UFMG.

Leite	Número de amostras	Gordura (g/100g)	Proteína (g/100g)	Lactose (g/100g)	Sólidos Totais (g/100g)	ESD (g/100g)
		Média (DP)	Média (DP)	Média (DP)	Média (DP)	Média (DP)
Vaca	11.335	3.75 ^{ad} (0.37)	3.31 ^{ad} (0.18)	4.47 ^{ad} (0.13)	12.52 ^{ad} (0.48)	8.78 ^{ad} (0.23)
25% Buf+75% Vaca	19	4.04 ^{bd} (0.30)	3.52 ^{bd} (0.07)	4.58 ^{bd} (0.05)	13.14 ^{bd} (0.36)	9.76 ^{bc} (0.18)
50% Buf+50% Vaca	19	4.65 ^{bd} (0.26)	3.76 ^{bd} (0.10)	4.66 ^{bd} (0.05)	14.06 ^{bc} (0.32)	10.13 ^{bc} (0.17)
60% Buf+40% Vaca	19	4.90 ^{bd} (0.26)	3.85 ^{bd} (0.12)	4.68 ^{bd} (0.06)	14.42 ^{bd} (0.32)	10.27 ^{bc} (0.17)
70% Buf+30% Vaca	19	5.13 ^{bd} (0.24)	3.95 ^{bc} (0.14)	4.72 ^{bd} (0.06)	14.78 ^{bd} (0.32)	10.43 ^{bd} (0.17)

Leite	Número de amostras	Gordura (g/100g)	Proteína (g/100g)	Lactose (g/100g)	Sólidos Totais (g/100g)	ESD (g/100g)
		Média (DP)	Média (DP)	Média (DP)	Média (DP)	Média (DP)
80% Buf+20% Vaca	30	5.30 ^{bd} (0.37)	3.99 ^{bc} (0.22)	4.76 ^{bd} (0.09)	15.06 ^{bd} (0.61)	10.50 ^{bd} (0.25)
90% Buf+10% Vaca	29	5.57 ^{bc} (0.22)	4.10 ^{bc} (0.17)	4.80 ^{bc} (0.08)	15.49 ^{bd} (0.30)	10.67 ^{bd} (0.18)
95% Buf+5% Vaca	29	5.67 ^{bc} (0.24)	4.17 ^{bd} (0.16)	4.83 ^{bc} (0.10)	15.67 ^{bd} (0.31)	10.77 ^{bd} (0.17)
98% Buf+2% Vaca	19	5.80 ^{bc} (0.28)	4.22 ^{bd} (0.20)	4.79 ^{bc} (0.10)	15.77 ^{bd} (0.38)	10.84 ^{bd} (0.20)
99% Buf+1% Vaca	19	5.82 ^{bc} (0.28)	4.22 ^{bd} (0.20)	4.80 ^{bc} (0.10)	15.81 ^{bd} (0.39)	10.85 ^{bd} (0.19)
Búfala	1.481	5.82 ^{bc} (0.78)	4.03 ^{bc} (0.37)	4.85 ^{bc} (0.21)	15.67 ^{bc} (0.92)	9.96 ^{bc} (0.49)

^{a, b, c, d} Médias dentro da mesma coluna com diferentes letras, diferem estatisticamente ($p < 0,05$).

Fonte: A autora (2022)

Para a variável gordura, foi observado que o leite de vaca difere dos tratamentos entre 1% e 50%, não possuindo diferença com tratamento de maior porcentagem de leite bovino adicionado (75%). Por outro lado, o leite bubalino possui diferença com tratamentos acima de 20% de adição de leite bovino e leite bovino sem adição, não havendo diferença quando a adição é baixa (1% a 10%)

Para a variável proteína, amostras de leite bovino diferem de todos os tratamentos, enquanto o leite bubalino sem adição somente difere em amostras de fraude acima de 40% e menores que 5% de adição de leite bovino.

Em relação à lactose, o leite bovino diferiu em todos os tratamentos, em contrapartida, o leite bubalino só diferiu em fraudes acima de 20%.

Para a variável sólidos totais, a diferença ocorre em todos os níveis em relação ao leite bovino. Para o leite de búfala, apenas amostras com 50% de adição de leite bubalino não apresentaram diferença estatística.

O leite bovino difere de todos os tratamentos quanto ao extrato seco desengordurado. O leite de búfala difere do leite bovino sem adição e de amostras com até 30% de adição de leite bovino, não havendo diferença entre 40 a 75% de fraude.

Silva (2017) estudando a adição de leite bovino ao leite bubalino, encontrou resultados semelhantes ao encontrado no presente trabalho. Amostras com adição de até 30% de leite de vaca apresentavam maior semelhança com o leite bubalino, sendo detectada a mudança de comportamento em amostras acima de 40% de adição. A partir de 60% de adição de leite bovino na amostra, o leite passava a ser mais semelhante ao leite de vaca segundo a análise de

componentes principais. No presente trabalho foi-se estudado cada componente de forma isolada, de forma que cada um foi detectada a diferença estatística em uma porcentagem de adulteração.

2.4. CONCLUSÃO

O leite bubalino apresentou maiores teores de gordura, proteína, lactose, sólidos totais e sólidos totais não gordurosos em relação ao leite bovino, como esperado, resultado que reforça a qualidade nutricional do produto além do maior rendimento industrial pelo alto teor de sólidos encontrado.

. As misturas de leite em diferentes níveis evidenciaram que quanto maior a porcentagem de leite bovino adicionada à amostra de leite bubalino, ocorre um efeito de diluição distanciando os valores dos padrões de leite de búfala e aproximando aos padrões de leite de vaca.

De maneira geral o leite de vaca apresentou diferença significativa em relação ao leite de búfala e às diluições, porém a diferença do leite de búfala em relação às diluições foi significativa em diluições a partir de 20%. Sendo assim, torna-se necessária a importância de métodos de detecção de adulterações cada vez mais eficazes, visto que apenas acima de 20% de adição de leite bubalino se tem alteração significativa.

CAPÍTULO 3: CONSTRUÇÃO DAS REDES NEURAIAS A PARTIR DOS RESULTADOS DOS COMPONENTES OBTIDOS POR FTIR

3.1 INTRODUÇÃO

A crescente demanda por derivados de leite de búfalas nos últimos anos tem destacado a espécie bubalina na pecuária brasileira, encorajando muitos produtores a investir nessa área de produção. A superioridade e os benefícios à saúde humana dos componentes do leite de búfala contribuem com o aumento do consumo de seus derivados. Essa crescente demanda aliada ao o fato de a produção ser sazonal e possuir maior valor agregado, pode levar a um aumento da incidência de práticas ilícitas, como adição de quantidade variáveis do leite de vaca durante a fabricação, caracterizando fraude por adulteração de produto (Azevedo, 2021).

A adição de leite de vaca em leite e derivados lácteos de búfala, além de ser uma prática fraudulenta, essas adulterações podem representar riscos à saúde humana, como em indivíduos que possuem distúrbios metabólicos ou alergias específicas ao leite de origem bovina (Azevedo, 2021).

Dentre os métodos tradicionais para a identificação da mistura de leite de diferentes espécies, as técnicas de Reação em Cadeia pela Polimerase (PCR) são as mais utilizadas. Esses procedimentos são extremamente sensíveis com excelentes resultados, porém demandam tempo de execução além de reagentes específicos. Com o aprimoramento das fraudes, as indústrias e órgãos de fiscalização desenvolveram diversas técnicas para identificação de adulterações no leite. A grande maioria das inspeções são caras e lentas, por se tratarem de análises físico-químicas laboratoriais (Drummond *et al.*, 2013).

As técnicas de Espectroscopia no infravermelho (FTIR) tem capacidade de produzir um grande volume de dados. A associação desta tecnologia com algoritmos de aprendizado de máquina pode ser uma ferramenta de otimização para processar esses dados e triagem de amostras para integridade (Lima, 2021).

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

As redes neurais artificiais são apresentadas como sistemas de neurônios interconectados que podem calcular valores de entradas e são capazes de realizar aprendizado de máquina por intermédio de exemplos e fazer inferências sobre o que aprendeu, melhorando gradativamente o seu desempenho, bem como reconhecimento de padrões devido à sua natureza adaptativa (Braga, *et al.* 2000; Nikam, 2015)

O objetivo desse capítulo foi utilizar os dados obtidos a partir da análise FTIR para a montagem das redes neurais *Multilayer Perceptron* de uma e duas camadas e redes de Função de Base Radial para a predição de fraudes por adição de leite bovino em leite bubalino.

3.2 MATERIAL E MÉTODOS

As redes neurais foram utilizadas buscando verificar os desempenhos individuais com relação à identificação de amostras fraudadas e a comparação entre as opções de forma a determinar a que mais se adequa de forma simples e objetiva. Para a formação das redes, os dados obtidos da análise de composição foram utilizados na alimentação na camada de entrada (Gordura, Proteína, Sólidos Totais, Lactose e Extrato Seco Desengordurado) e na camada de saída foram dispostas as porcentagens de fraude (1%, 2%, 5%, 10%, 20%, 30%, 40%, 50%, 75% e 100%). O software utilizado para a análise das redes foi o SPSS® que possui a função de modelagem de dados não linear através de análises de redes neurais MLP e RBF, e está mais disponível em relação a softwares mais complexos. No software SPSS® as arquiteturas das redes MLP e RBF permitem separar os dados de forma automática em percentuais de treinamento, validação e teste. Foram efetuados ensaios com 60% dos dados destinados ao treinamento, 20% à validação e 20% ao teste, utilizando os mesmos conjuntos para a comparação dos três tipos de redes. Para todos os modelos os dados foram distribuídos de forma aleatória. Foram executadas as seguintes ferramentas no *software*: Rede *Perceptron* Multicamadas *Feed-Forward*, que consiste em uma das RNAs mais utilizadas, tanto para aproximação de funções, como para classificação de padrões e Redes de Função de Base Radial, comumente aplicadas principalmente para modelagem de sistemas de predição e classificações de dados.

3.2.1 *Multilayer Perceptron (MLP)*

Para a execução das redes MLP's, foram definidos o banco de dados, as entradas e os alvos. As arquiteturas desse modelo de rede possuem uma camada de entrada para recebimento dos dados

espectrais da análise FTIR. Após essas definições foram feitas as seleções dos parâmetros de ensaio e números de camadas ocultas, definição dos dados de treinamento e teste e, enfim, a execução da rede. Após a execução de todos os ciclos foi feito o cálculo das médias e preparação dos resultados que foram posteriormente armazenados.

3.2.2 Função de Base Radial

O funcionamento das RBFs é semelhante às MLPs. As redes perceptron multicamadas diferem das de função de base radial requerendo um número menor de parâmetros para se obter um mesmo grau de precisão, mas são mais lentas (Haykin, 2001)

Foram realizados dez ensaios para cada rede neural e os resultados armazenados para comparação de desempenhos. Para isso, foram feitos cálculos da quantidade de acertos das redes e os percentuais de acurácia, precisão, sensibilidade e especificidade.

3.2.3 Análise estatística

As métricas de performance das redes foram avaliadas segundo acurácia, precisão, sensibilidade e especificidade, onde:

$$\text{Acurácia} = \frac{\text{amostras classificadas corretamente}}{\text{total de amostras}}$$

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

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

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

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

As análises estatísticas complementares incluindo a estatística descritiva e análises multivariadas foram realizadas com o software IBM SPSS Statistics (v. 22) para a diferenciação das amostras de leite de búfala das amostras fraudadas. O teste de Tukey foi utilizado como comparação post-hoc nos tratamentos com o erro tipo 1 controlado no nível de significância 5%.

3.3 RESULTADOS E DISCUSSÃO

Os dados obtidos da análise FTIR foram utilizados como camada de entrada na montagem das arquiteturas de redes neurais *Multilayer Perceptron* e Função de Base Radial utilizando o software SPSS®.

Para a formação das redes de forma que os resultados sejam os mais assertivos possíveis, é preciso que a rede possua um banco de dados com subgrupos balanceados. Tecnicamente qualquer conjunto de dados que apresenta uma desigualdade da distribuição entre as classes pode ser considerado desbalanceado (He, 2008).

O banco de dados geral foi composto por 13.018 amostras, compreendendo tanto amostras puras quanto adulteradas. Para minimizar o erro proveniente do desbalanceamento dos grupos amostrais na montagem das redes neurais, uma segunda amostragem foi realizada de modo a homogeneizar o banco de dados com grupos em quantidades proporcionais escolhidas ao acaso, conforme representado na Figura 3-1. Dessa forma, foram criados 3 grupos de amostras aleatórias de leite bovino sem adição e bubalino sem adição que foram selecionadas e divididas aleatoriamente em 3 grupos com número aproximadamente similar de amostras. As amostras fraudadas foram repetidas em todos os grupos. Após essa divisão, cada grupo de amostras continha quantidades similares de amostras gerando uma redução significativa do desbalanceamento dos dados.

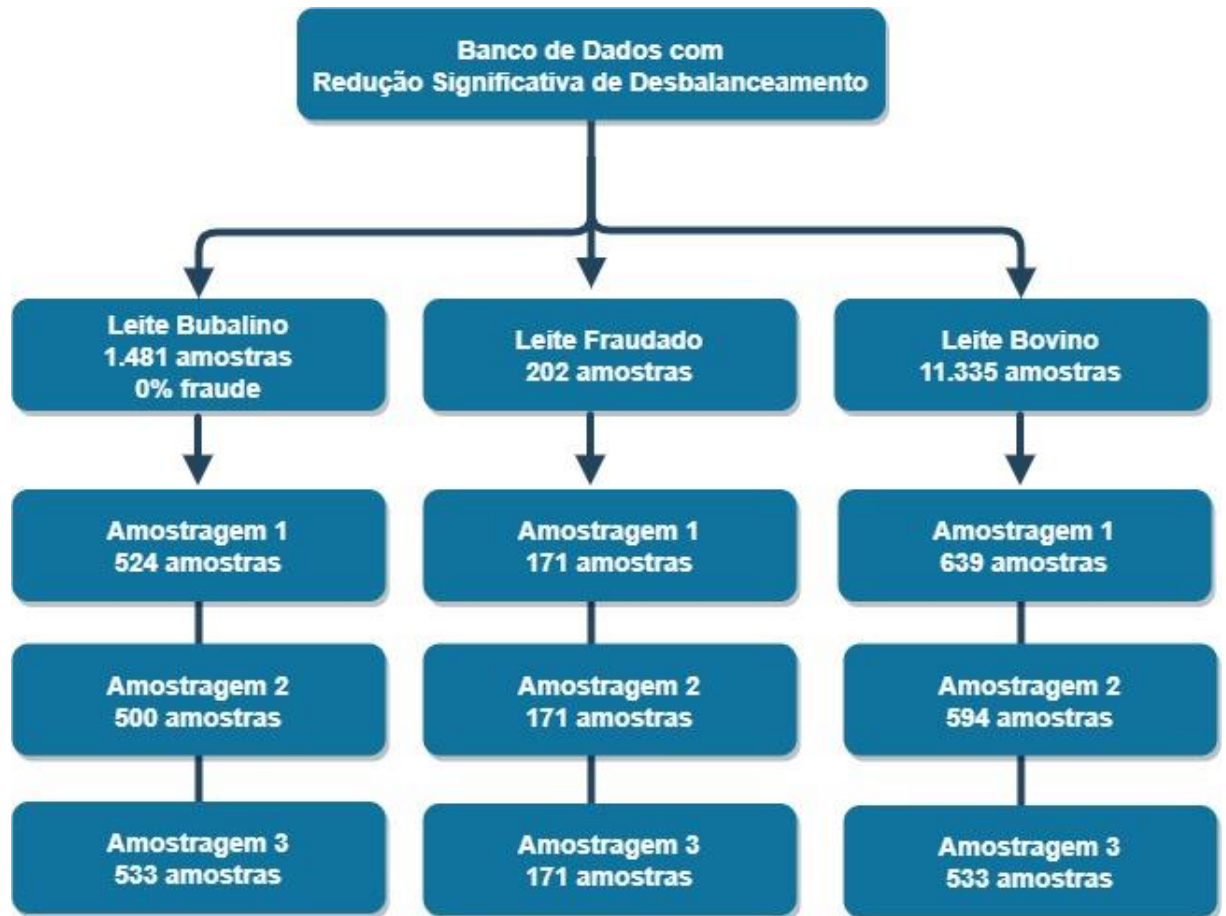


Figura 0-1: Esquema da redução do desbalanceamento das amostras utilizadas para formação do banco de dados utilizando dados obtidos de análises FTIR em leite bovino, leite bupalino e adição de leite bovino ao leite bupalino em diferentes porcentagens simulando fraude, em amostras obtidas no período de 2020-2022 em tanques de refrigeração de unidades produtoras no estado de Minas Gerais.

Fonte: A autora (2022)

No primeiro grupo amostral, chamado de banco geral de amostras, onde o número de amostras de leite bovino foi muito superior ao número de amostras de leite bupalino sem adição, as duas respostas possíveis da rede seriam de fraude e não fraude. O conjunto possuía, portanto, 13.018 amostras, sendo 11.537 amostras com leite bovino (consideradas fraudadas por não ser leite de búfala) e 1481 amostras de leite bupalino (não fraudadas). Preferencialmente, espera-se obter um classificador que forneça um grau equilibrado de acurácia preditiva (idealmente 100 por cento) para ambas as classes (fraudadas e não fraudadas) do conjunto de dados. De fato, os classificadores tendem a proporcionar um grave desequilíbrio no grau de precisão, com o grupo fraudado com acurácia perto de 100 por cento e o grupo não fraudado tendo acurácia de 0-10 por cento, por exemplo (Chawla, 2002). A consequência disso é a classificação errônea

de amostras não fraudadas sendo classificadas como fraudadas devido ao desbalanceamento com elevado número de amostras de leite bovino. No campo prático, onde o propósito é melhorar a detecção de amostras anormais na indústria láctea, classificar erroneamente uma amostra poderia trazer danos ao produtor e maior gasto com testes específicos desnecessários para detecção de fraudes em amostras puras. Dessa forma, para evitar resultados falso-positivos, foi realizado um balanceamento do grupo amostral de forma que os subgrupos tivessem dados proporcionais, gerando dessa forma um resultado mais confiável. Foram criados, portanto, 3 grupos amostrais: amostragem 1 com 1.334 amostras, amostragem 2 com 1.265 amostras e amostragem 3 com 1.237 amostras.

Uma das técnicas mais comuns para mitigar o problema dos grupos desbalanceados consiste na modificação do conjunto de dados originais, como por exemplo, com a reamostragem aleatória de dados. Estudos têm mostrado que para vários algoritmos de classificação, um conjunto balanceado é capaz de proporcionar uma melhoria significativa no desempenho das classificações comparadas com um conjunto de dados desbalanceados (He, 2008). Estes resultados justificam o uso de métodos de reamostragem para aprendizado desbalanceado (Provost, 2001). A reamostragem foi realizada de forma a reduzir o desbalanceamento existente no banco de dados.

3.3.1 Arquiteturas de Redes

3.3.1.1 Redes Multilayer Perceptron (MLP)

Os dados em formato CSV do banco de dados balanceado foram utilizados na camada de entrada das redes neurais *Multilayer Perceptron*. Foram rodadas cinco redes neurais de cada arquitetura para cada grupo de amostras, totalizando quinze redes para cada tipo de rede neural. Os resultados finais foram obtidos a partir da média dos resultados obtidos para cada arquitetura.

Foram criados dois modelos de redes neurais de múltiplas camadas, com uma e duas camadas ocultas. As informações das arquiteturas da rede *Multilayer Perceptron* de uma camada estão dispostas na Tabela 3-1.

Tabela 0-1: Arquitetura da Rede Neural Multilayer Perceptron de uma camada (MLP1) utilizando dados obtidos de análises FTIR feitas no Laboratório de Análise da Qualidade do Leite da Escola de Veterinária da UFMG, utilizando amostras coletadas durante os anos de 2020-2022.

Arquitetura de Rede Multilayer Perceptron de uma camada			
Camada interna	Covariáveis	1	Gordura
		2	Proteína
		3	Lactose
		4	Sólidos Totais
		5	Extrato Seco Desengordurado
	Número de unidades ^a		5
	Método de redimensionamento para covariáveis		Padronizado
Camada Oculta	Número de Camadas Ocultas		1
	Número de unidades na primeira camada		13
	Função de Ativação		Tangente Hiperbólica
Camada de Saída	Variáveis Dependentes	1	Fraude 1%
		2	Fraude 2%
		3	Fraude 5%
		4	Fraude 10%
		5	Fraude 20%
		6	Fraude 30%
		7	Fraude 40%
		8	Fraude 50%
		9	Fraude 75%
		10	Fraude 100%
	Número de unidades		20
	Função de Ativação		Softmax
	Função de Erro		Cross-entropy

a = Excluindo Bias

Fonte: A autora (2022)

As redes *Multilayer Perceptron* de uma camada oculta apresentaram média geral de acertos para a fase de validação 97,61% e teste com 97,42%.

Esse resultado corrobora com o encontrado por Silva (2020) utilizaram diferentes métodos quimiométricos para detectar a adição de leite de vaca no leite de búfala, por meio da técnica Espectroscopia de Infravermelho por Transformada de Fourier associada à refletância total atenuada (FTIR-ATR). As redes neurais se mostraram mais satisfatórias na identificação de leite bovino em leite bubalino, com percentual de acerto geral do teste de 95,55%. As outras análises realizadas foram as análises dos componentes principais, com diferenciação das

amostras somente a partir de 40% de adição de leite de vaca; análise de agrupamento, com resultado satisfatório apenas em amostras puras; e análise discriminante, com baixa classificação no treinamento e na validação para todos os grupos de amostras. Os resultados mostraram a eficácia do método espectroscópico FTIR-ATR associado à análise quimiométrica para a detecção da autenticidade do leite de búfala.

Em relação a taxa de classificação da rede por classe de fraude, os resultados estão dispostos na Tabela 3-2.

Tabela 0-2: Percentual médio de acerto e classificação das redes Multilayer Perceptron de uma camada em relação aos níveis de adição de leite bovino em leite bubalino simulando fraude experimental.

% de Fraude	% de Acerto
1%	95,70%
2%	95,72%
5%	96,54%
10%	97,38%
20%	98,08%
30%	98,80%
40%	99,04%
50%	98,90%
75%	99,46%
Vaca	99,32%

Fonte: A autora (2022)

Os resultados indicam que as redes tendem a ir aumentando a porcentagem de acerto à medida que aumenta a quantidade de leite bovino adicionada à amostra. Mesmo em pequenas adições, as redes *Multilayer Perceptron* de uma camada são capazes de fazer boa previsão de fraude com mais de 95% dos resultados corretos, ou seja, classificando amostras bubalinas como sendo puras e amostras fraudadas como sendo fraude.

Balestrieri *et al* (2001) realizaram na Itália a classificação de amostras de leite de cabra, ovelha e búfala por meio de redes neurais de acordo com a composição de cada espécie determinada pelo método oficial de análise italiana. Foram utilizados 21 neurônios de entrada, 4 neurônios de saída e 1 camada oculta com 10 neurônios tanto para treinamento quanto para validação e apresentou 96% de classificação correta, porcentagem semelhante ao encontrado no presente estudo.

Outro estudo feito por Nazário *et al.* (2009), projetou uma rede neural para detectar adulteração de amostras de leite por aguçagem na faixa de 1 a 60% através de equipamentos de ultrassom. Na detecção da adulteração foram obtidos resultados com boa capacidade de

classificação, sendo que as configurações contendo 40 e 52 neurônios obtiveram o maior potencial de classificação, obtendo 95,7% e 96,3% respectivamente.

Alves *et al.* (2022) avaliaram o uso de redes neurais Multilayer Perceptron de uma camada para a detecção de fraudes em leite em pó integral por adição de soro em pó. Os estudos concluíram que mesmo para amostras de leite em pó o uso das redes neurais é capaz de fazer boa predição de fraudes, sendo uma ótima alternativa de uso para a indústria de lácteos.

Lima (2021) estudou o uso de redes neurais Multilayer Perceptron de uma camada para identificação de amostras de leite cru fraudadas com soro de leite, utilizando dados FTIR. Como resultado, obteve 98% de acurácia no teste, mostrando a boa capacidade de detecção de fraudes pelo método.

A fim de comparar os resultados das redes neurais Multilayer Perceptron, foi executada a rede com duas camadas, como demonstrado na Tabela 3-3.

Tabela 0-3: Arquitetura de Rede Neural Multilayer Perceptron de duas camadas (MLP2) utilizando dados obtidos de análises FTIR feitas no Laboratório de Análise da Qualidade do Leite da Escola de Veterinária da UFMG, utilizando amostras coletadas durante os anos de 2020-2022

Arquitetura de Rede Multilayer Perceptron de duas camadas			
Camada interna	Covariáveis	1	Gordura
		2	Proteína
		3	Lactose
		4	Sólidos Totais
		5	Extrato Seco Desengordurado
		Número de unidades ^a	
	Método de redimensionamento para covariáveis		Padronizado
Camada Oculta	Número de Camadas Ocultas		2
	Número de unidades na primeira camada		13
	Número de unidades na secundacamada		10
	Função de Ativação		Tangente Hiperbólica
Camada de Saída	Variáveis Dependentes	1	Fraude 1%
		2	Fraude 2%
		3	Fraude 5%
		4	Fraude 10%
		5	Fraude 20%
		6	Fraude 30%
		7	Fraude 40%
		8	Fraude 50%
		9	Fraude 75%
		10	Fraude 100%
	Número de unidades		20
	Função de Ativação		Identity
	Função de Erro		Soma dos Quadrados

a = Excluindo Bias

Fonte: A autora (2022)

As porcentagens de acerto na classificação geral para as redes *Multilayer Perceptron* de duas camadas foram 97,62% na etapa de validação e 96,98% na de teste. Em relação à classificação considerando os níveis de fraude, os resultados estão dispostos na Tabela 3-4.

Tabela 0-4: Percentual médio de acerto e classificação das redes *Multilayer Perceptron* de duas camadas em relação aos níveis de adição de leite bovino em leite bubalino simulando fraude experimental.

% de Fraude	% de Acerto
1%	94,44%
2%	94,08%
5%	95,14%
10%	96,60%
20%	96,88%
30%	97,66%
40%	98,16%
50%	98,00%
75%	98,66%
Vaca	97,94%

Fonte: A autora (2022)

Do mesmo modo que as redes *Multilayer Perceptron* de uma camada, as redes *Multilayer Perceptron* de duas camadas tiveram aumento na porcentagem de acerto da classificação à medida que a porcentagem de adição de leite bovino à amostra se tornava maior, porém o uso de duas camadas não resultou em incremento quando comparado a *Multilayer Perceptron* de uma camada.

Conceição *et al.* (2019), avaliando o uso das redes neurais para a predição de adulterantes no leite cru, encontraram que todas as redes testadas apresentaram boa capacidade de predição para os adulterantes avaliados. Para todas as classes de adulterantes, o percentual de classificação encontrado foi superior a 76%, indicando boa capacidade de predição de fraudes mesmo em baixas porcentagens de adição. Os resultados encontrados no presente estudo foram superiores aos encontrados por Conceição *et al.* (2019) mesmo em baixas porcentagens de leite bovino adicionadas às amostras.

3.3.1.2 Redes de Função de Base Radial (RBF)

A fim de obter dados comparativos das efetividades das redes na detecção de fraudes, foram utilizadas as redes de função de base radial como dispostos na Tabela 3-5.

Tabela 0-5: Arquitetura de Rede Neural de Função de Base Radial (RBF) utilizando dados obtidos de análises FTIR feitas no Laboratório de Análise da Qualidade do Leite da Escola de Veterinária da UFMG, utilizando amostras coletadas durante os anos de 2020-2022.

Arquitetura de Rede de Função de Base Radial			
Camada interna	Covariáveis	1	Gordura
		2	Proteína
		3	Lactose
		4	Sólidos Totais
		5	Extrato Seco Desengordurado
	Número de unidades		5
	Método de redimensionamento para covariáveis		Padronizado
Camada Oculta	Número de unidades na primeira camada		a
	Função de Ativação		Softmax
Camada de Saída	Variáveis Dependentes	1	Fraude 1%
		2	Fraude 2%
		3	Fraude 5%
		4	Fraude 10%
		5	Fraude 20%
		6	Fraude 30%
		7	Fraude 40%
		8	Fraude 50%
		9	Fraude 75%
		10	Fraude 100%
	Número de unidades		20
	Função de Ativação		Identity
	Função de Erro		Soma dos Quadrados

a = Determinado pelo critério de dados de teste: O "melhor" número de unidades ocultas é aquele que produz o menor erro nos dados de teste.

Fonte: A autora (2022)

Os resultados de classificação geral para as redes de Função de Base Radial foram: na fase de validação 97,72% e na fase de teste 97,33%. Os resultados gerais foram similares em relação aos três modelos de redes neurais utilizados na comparação.

Em relação aos acertos de acordo com os níveis de fraude, os resultados estão dispostos na Tabela 3-6.

Tabela 0-6: Percentual médio de acerto e classificação das redes de Função de Base Radial em relação aos níveis de adição de leite bovino em leite bubalino simulando fraude experimental

% de Fraude	% de Acerto
1%	96,00%
2%	95,90%
5%	96,72%
10%	96,98%
20%	98,46%
30%	98,78%
40%	98,86%
50%	98,54%
75%	99,04%
100%	98,68%

Fonte: A autora (2022)

Da mesma forma que as redes *Multilayer Perceptron*, as redes Função de Base Radial também tiveram maior percentual de acerto à medida que aumenta a quantidade de leite bovino na amostra. Mesmo em pequenas porcentagens de fraude, as redes se mostraram muito eficazes na detecção das fraudes com ótimo percentual de respostas corretas.

Na área de alimentos, as redes neurais de Função de Base Radial e *Multilayer Perceptron* foram utilizadas na quantificação do percentual de café arábica em misturas de café, com metodologia semelhante à realizada na detecção das fraudes em leite (Carvalho, 2015; Sabino, 2015). Nos dois estudos foram comparadas as métricas obtidas nas redes neurais e os resultados apresentaram percentual de acerto acima de 95% em ambas. A conclusão do estudo após a comparação das redes, foi que de maneira geral, as redes *Multilayer Perceptron* apresentam melhor desempenho por gerarem menor erro absoluto médio.

Por outro lado, Valente *et al.* (2014) buscaram analisar amostras fraudadas com soro de leite em diferentes concentrações e compararam as redes *Multilayer Perceptron* com as redes de função de base radial. Como resultado, as redes de função de base radial obtiveram melhores índices de classificação em relação às redes *Multilayer Perceptron*, com erro de classificação menor que 5%.

Além desses modelos, outros modelos de redes neurais já foram estudados para a identificação de amostras adulteradas. Outra arquitetura estudada foi a de Redes Neurais Convolucionais que em estudo realizado por Tavares (2019) apresentou desempenhos melhores aos comparados a modelos tradicionais. Tavares (2019) avaliou a identificação de amostras de leite com adulterantes utilizando dados obtidos de análises FTIR, redes neurais convolucionais

e classificadores *Gradient Boosting Machine* (GBM) e *Random Forests* (RF). As redes convolucionais obtiveram resultados de 99,34% de classificação correta, demonstrando a boa capacidade dessa tecnologia na identificação de fraudes.

Ainda na área de lácteos, Ribeiro (2021) objetivando identificar e quantificar a lactose residual e outros açúcares de leite por associação de análises FTIR e redes neurais artificiais convolucionais, encontrou capacidade preditiva das redes de 80% para classificação e 75% para quantificação dos açúcares em amostras de leite após a hidrólise da lactose.

Esses resultados reforçam que a inteligência artificial tem ganhado cada vez mais espaço nas diferentes áreas com resultados satisfatórios principalmente na detecção de fraudes.

3.3.2 Testes de comparação múltipla das Redes Neurais

Foi realizado o teste de Tukey para comparar a variação em amostras multivariadas. Nesse teste, duas ou mais matrizes de covariância são testadas a fim de analisar se são iguais (homogêneas) ou não. O valor de p encontrado foi $p < 0,001$, indicando que as matrizes são diferentes estatisticamente.

Foram realizados testes de comparação múltipla dos métodos a fim de se comparar as métricas de eficiência das redes neurais. Os métodos foram comparados segundo acurácia, precisão, sensibilidade e especificidade. Os resultados estão dispostos na Tabela 3-7.

Tabela 0-7: Comparação múltipla entre as métricas de qualidade dos testes (acurácia, precisão, sensibilidade e especificidade) e os diferentes modelos de Redes Neurais Artificiais utilizados segundo o teste de Tukey.

		Comparação Múltipla do Método					95% Intervalo de Confiança	
Variável Dependente		(I)	(J)	Diferença de	Erro Padrão	p	Limite Inferior	Limite Superior
		Método(MLP1, MLP2, RBF)	Método(MLP1, MLP2, RBF)	Médias (I-J)				
Acurácia	Tukey	MLP1	MLP2	.00456^b	.001111	.000	.00194	.00717
			RBF	.00118 ^a	.001111	.537	-	.00380
	HSD	MLP2	MLP1	-.00456^b	.001111	.000	.00717	-.00194
			RBF	-.00338^b	.001111	.007	.00599	-.00076
Precisão	Tukey	MLP1	MLP2	.00338^b	.001111	.007	.00076	.00599
			RBF	.00794^b	.001587	.000	.00421	.01168
			MLP2	-.00573^b	.001587	.001	.00947	-.00200
			RBF	-.00794^b	.001587	.000	.01168	-.00421
	HSD	MLP2	MLP1	-.01368^b	.001587	.000	.01741	-.00994
			RBF	.00573^b	.001587	.001	.00200	.00947
			MLP2	.01368^b	.001587	.000	.00994	.01741
			RBF	.00573^b	.001587	.001	.00200	.00947
Sensibilidade	Tukey	MLP1	MLP2	-.00162^a	.001795	.639	-.00585	.00261
			RBF	.00509^b	.001795	.013	.00087	.00932
	HSD	MLP2	MLP1	.00162^a	.001795	.639	-.00261	.00585
			RBF	.00671^b	.001795	.001	.00249	.01094
Especificidade	Tukey	MLP1	RBF	-.00509^b	.001795	.013	.00932	-.00087
			MLP2	-.00671^b	.001795	.001	.01094	-.00249
			MLP2	.01052^b	.001701	.000	.00652	.01453
			RBF	-.00393^a	.001701	.056	-.00793	.00007
	HSD	MLP2	MLP1	-.01052^b	.001701	.000	.01453	-.00652
			RBF	-.01445^b	.001701	.000	.01846	-.01045
			RBF	.00393^a	.001701	.056	-.00007	.00793
			MLP2	.01445^b	.001701	.000	.01045	.01846

^{a,b} Letras diferentes na mesma coluna diferem estatisticamente entre si. Tratamentos que foram diferentes entre si estatisticamente foram destacados em negrito e com linhas vermelhas entre as colunas.

Fonte: A autora (2022)

Para as métricas de acurácia, sensibilidade e precisão, ambos os testes apresentaram os mesmos resultados em relação à diferença estatística. A especificidade foi a única métrica a apresentar resultados divergentes entre os testes. Para a acurácia as comparações entre as redes MLP1 x MLP2 e MLP2 x RBF apresentaram diferença estatística, enquanto as redes MLP1 x RBF não apresentaram diferenças. As redes MLP1 e RBF apresentaram melhores métricas, sendo as opções com melhor desempenho.

Em relação à sensibilidade ocorreram diferenças estatísticas entre as redes MLP1xRBF e MLP2xRBF, enquanto as redes MLP1xMLP2 não apresentaram diferenças o que indica que a adição de uma camada oculta não traz ganho aos resultados. Esse teste indica os verdadeiros positivos, ou seja, amostras fraudadas quando se tem fraude, e para esse teste, as redes com melhor desempenho foram as redes RBF.

Para a métrica de precisão, todas as comparações apresentaram diferenças estatísticas entre si nos dois testes avaliados.

Quanto a especificidade, na comparação multivariada utilizando o teste de Tukey foram observadas diferenças significativas entre as comparações de redes MLP1 x MLP2 e MLP2 e RBF. Não ocorreram diferenças estatísticas entre as redes MLP1 e RBF.

Os resultados evidenciam que os três modelos de redes, apesar das diferenças entre eles, possuem boa capacidade de detecção das fraudes e das amostras puras, evidenciando que são eficazes como ferramenta de triagem na detecção de amostras anormais.

Como todo teste, as redes neurais também apresentam limitações do método, como a necessidade de grande quantidade de dados para a alimentação das redes, dificuldade de lidar com os dados desbalanceados gerando resultados tendenciosos e falso-positivos ou falso-negativos, além da dificuldade de lidar com dados com alta dimensão, que aumenta o tempo de treinamento e pode ocasionar um pior desempenho.

3.3.3 Gráficos Boxplot

Os gráficos boxplot são utilizados para comparar a distribuição de pontuações em variáveis. Esses gráficos foram utilizados para uma melhor inspeção do padrão de pontuação nos grupos permitindo a observação das diferenças dentro de cada parâmetro de métrica avaliado como demonstrado nas Figuras 3-2 a 3-5.

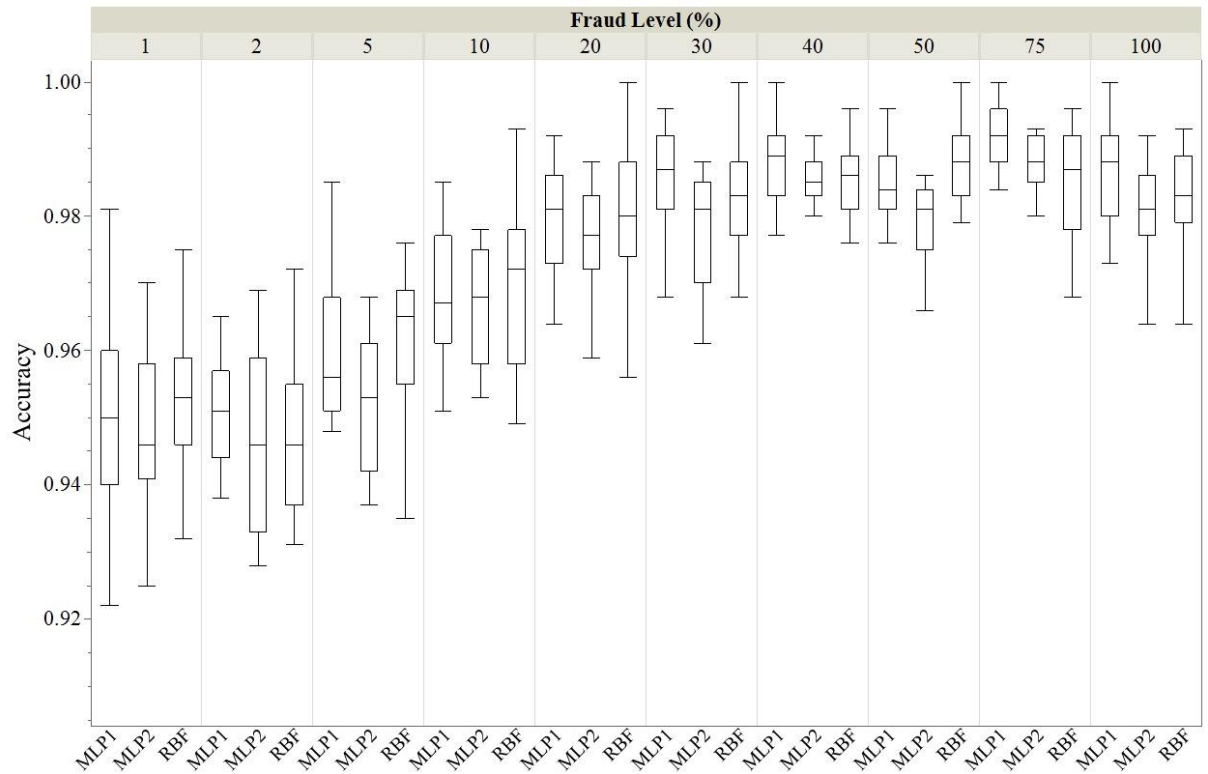


Figura 0-2: Gráfico Boxplot comparativo dos diferentes tipos de Rede Neural Multilayer Perceptron e Função de Base Radial em relação à acurácia e a diferentes porcentagens de adição de leite bovino em leite bubalino simulando fraude.

Fonte: A autora (2022)

Cada distribuição de pontuação é representada por uma caixa e linhas adjacentes. O comprimento da caixa é o intervalo interquartil da variável e contém 50% dos casos. A linha horizontal no interior da caixa representa o valor mediano e as linhas saindo da caixa correspondem aos valores menores e maiores da variável (Pallant, 2010).

Em relação à acurácia, as redes tiveram bom desempenho com resultados acima de 92%, o que indica uma boa capacidade de predição de fraude mesmo em porcentagens baixas de adição de leite bovino às amostras. A distribuição dos resultados foi bastante homogênea demonstrando que para todos os níveis de fraude, as redes tiveram uma boa porcentagem de

acerto nos três modelos avaliados. Isso é evidenciado pela baixa amplitude dos resultados entre a maior e a menor detecção. Quanto menor a porcentagem de fraude avaliada, maior foi a variação da distribuição dos resultados observados. À medida que o percentual de fraude aumenta os valores se tornam mais homogêneos e com menos variação.

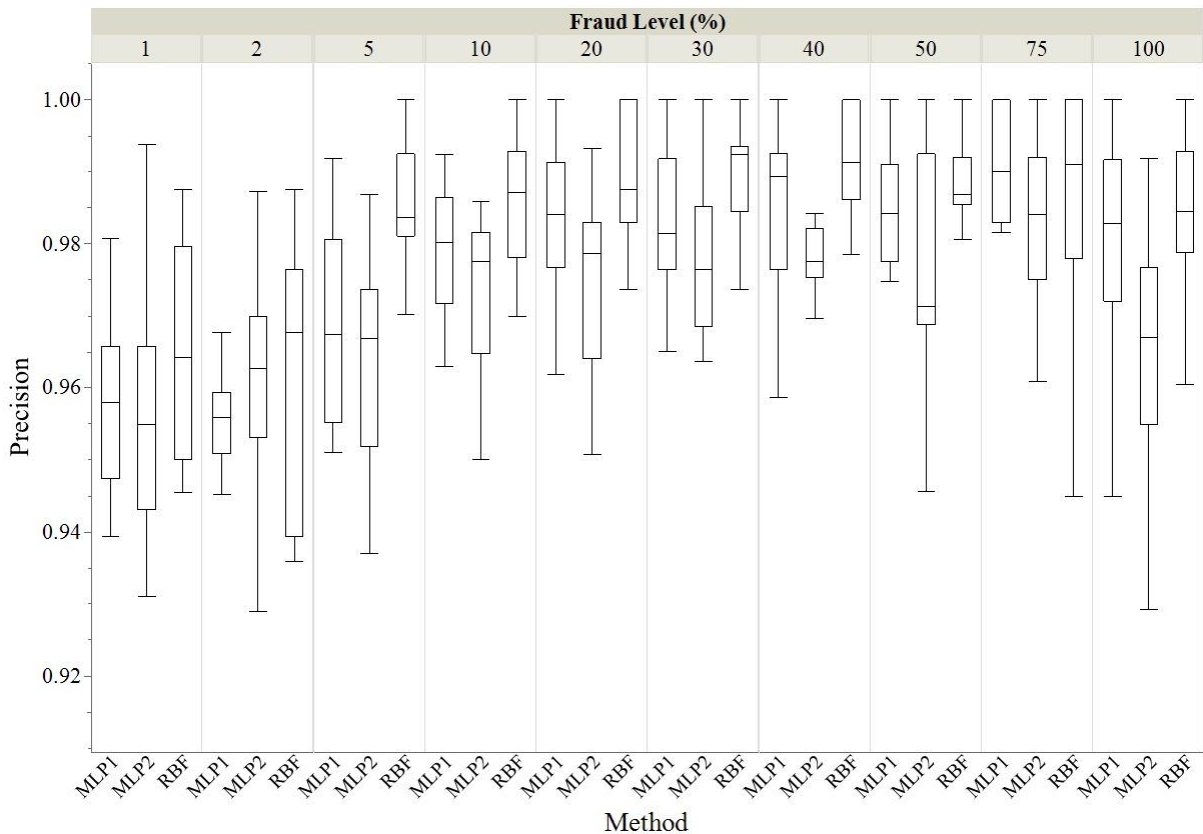


Figura 0-3: Gráfico Boxplot comparativo dos diferentes tipos de Rede Neural Multilayer Perceptron e Função de Base Radial em relação à precisão e a diferentes porcentagens de adição de leite bovino em leite bubalino simulando fraude.

Fonte: A autora (2022)

A análise de precisão mostrou bons resultados em todas as redes avaliadas, com índices acima de 0,93 em todos os níveis de porcentagem de fraude. Essa métrica de qualidade indica uma boa repetibilidade do teste, com pouca variação de erro. À medida que a porcentagem de fraude aumenta, aumenta a quantidade de resultados mais próximos a 1 evidenciados pela linha do limite superior, em todos os modelos de redes testadas e pela proporcionalidade das respostas distribuídas no boxplot.

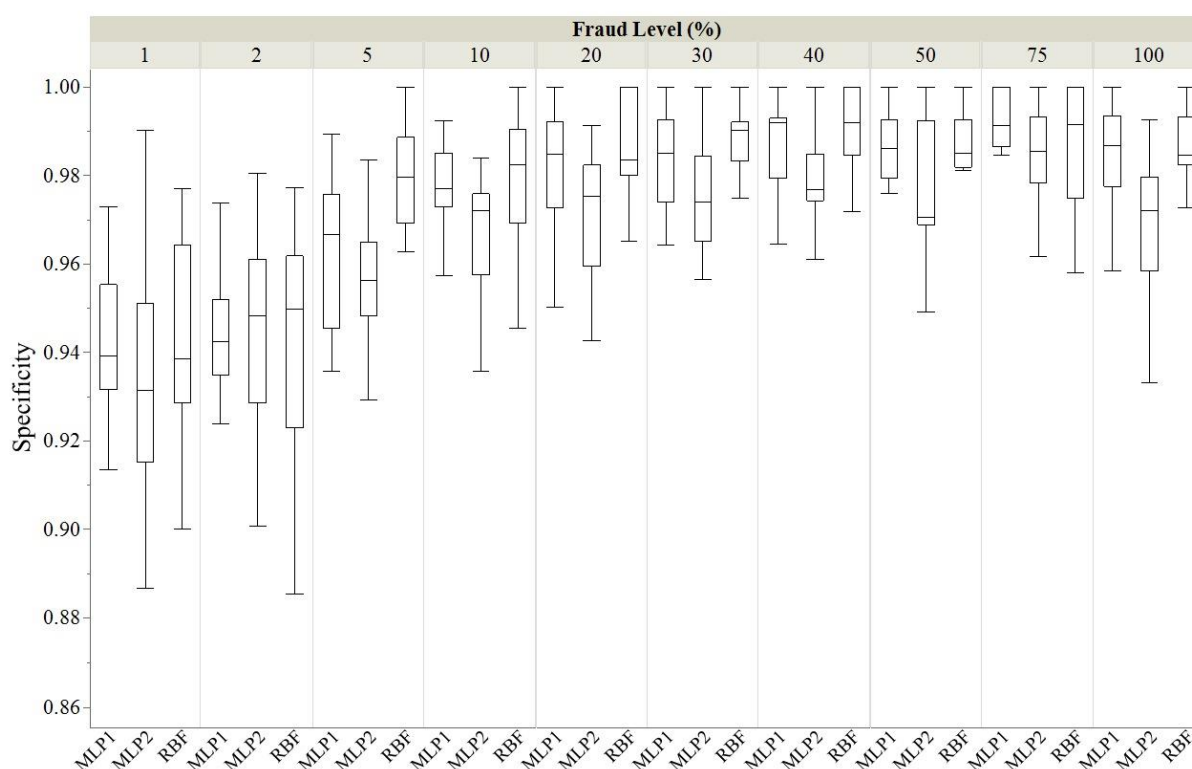


Figura 0-4: Gráfico Boxplot comparativo dos diferentes tipos de Rede Neural Multilayer Perceptron e Função de Base Radial em relação à especificidade e à diferentes porcentagens de adição de leite bovino em leite bubalino simulando fraude.

Fonte: A autora (2022)

Em relação à especificidade, como observado na Figura 3-4, todos os modelos são capazes de identificar a quase totalidade das amostras negativas como sendo negativas, ou seja, identificam as amostras de leite bubalino sem adição dentre as amostras de fraude. A especificidade para todas as redes aumenta à medida que a porcentagem de fraude é maior, mostrando maior diferenciação entre o leite bubalino e às amostras que contém alguma porcentagem de leite bovino adicionada. Para todos os modelos, os valores de especificidade foram acima de 0,88. Baixas porcentagens de leite bovino adicionadas à amostra (1 e 2%) apresentaram maior amplitude de resultados, variando de 0,88 a 0,98 de especificidade e mediana entre 0,93 e 0,95 em todos os três modelos de redes. A 5% de fraude, o melhor resultado foi da rede RBF com menor amplitude de resultados e valores mais próximos a 1. À 10% de fraude a rede MLP1 apresentou menor amplitude e resultados entre 0,96 e 0,99 com mediana a 0,98. A RBF foi a única a apresentar resultados de especificidade 1, porém com maior amplitude de resultados. A partir de 20% de fraude todas as porcentagens apresentaram pelo menos duas redes com resultados de especificidade máxima. À medida que se aumenta a

porcentagem de fraude, a amplitude dos resultados diminui, reduzindo inclusive os quadrantes do boxplot.

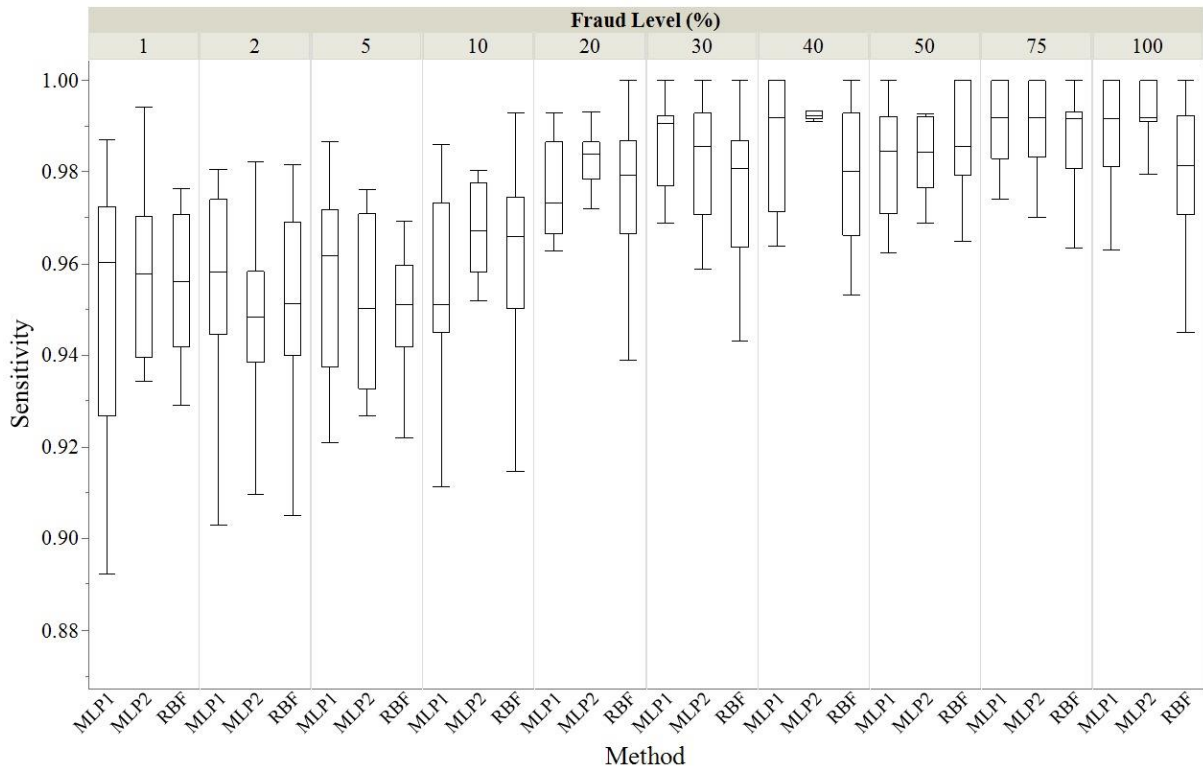


Figura 0-5: Gráfico Boxplot comparativo dos diferentes tipos de Rede Neural Multilayer Perceptron e Função de Base Radial em relação à sensibilidade e à diferentes porcentagens de adição de leite bovino em leite bubalino simulando fraude.

Fonte: A autora (2022)

Em relação à sensibilidade, todas as redes apresentaram resultados acima de 0,89, ou seja, as redes possuem boa capacidade de identificar como fraudadas as amostras que realmente contém fraude. Mesmo quando as porcentagens de fraude são baixas (1, 2 e 5%) a capacidade de detecção da rede é satisfatória, apesar da grande amplitude dos resultados. Amostras a partir de 20% alcançaram sensibilidade máxima em pelo menos uma das redes avaliadas. A partir 40% de fraude, os resultados estão mais concentrados mostrando uma menor amplitude nos três modelos de redes. Nas fraudes 50% e 75% as três redes apresentaram pouca variabilidade de resultados, contendo pelo menos duas redes com resultados de sensibilidade máxima.

3.4 CONCLUSÃO

Todas as redes neurais avaliadas no presente trabalho apresentaram ótimo desempenho na classificação e predição das amostras fraudadas, se mostrando uma excelente ferramenta a

ser utilizada por laticínios na identificação de triagem de possíveis infratores. Além da vantagem da ótima acurácia dos três modelos, o tempo gasto na montagem e análise dos dados pela rede é relativamente curto, o que permite uma análise rápida e eficiente dos resultados, tornando viável a aplicação.

Apesar dos resultados satisfatórios de forma geral, a adição de uma camada oculta nas redes MLP não trouxe ganho significativo. Quando comparadas às redes MLP1 e RBF, os resultados da rede MLP2 teve redução no desempenho, de forma que não justifica a sua utilização. Para as análises de adulteração de leite de búfala com adição de leite bovino, as redes neurais de uma camada oculta possuem resultados satisfatórios.

CAPÍTULO 4: APRENDIZADO DE MÁQUINA E APRENDIZADO PROFUNDO A PARTIR DOS RESULTADOS DOS COMPONENTES OBTIDOS POR FTIR

4.1 INTRODUÇÃO

A adulteração de alimentos com a matéria prima de outras espécies que possui menor custo e/ou maior disponibilidade é observada no mundo todo. Algumas práticas como a adição de leite bovino em produtos lácteos de ovelhas, cabras e laticínios de búfalo são as mais citadas (Herman, 2001).

A substituição não declarada de leite de búfala por leite de vaca que possui menor valor comercial é um problema comum, devido a diferenças de preço e disponibilidade sazonal do leite da espécie bubalina o que acaba levando o produtor a essa prática (Di Domenico *et al.* 2017).

Para a detecção cada vez mais precoce desse tipo de adulteração, as indústrias e pesquisadores tem investido em tecnologias avançadas. Técnicas que envolvem aprendizado de máquina demonstram ser extremamente eficazes para enfrentar esse desafio. Uma técnica que vem ganhando relevância no contexto de aprendizado de máquina é conhecida como Aprendizado Profundo (*Deep Learning*, em inglês). Essas técnicas foram inicialmente utilizadas em reconhecimento automático de fala, reconhecimento de imagem e processamento de linguagem natural e mais recentemente na detecção de fraudes em transações financeiras como mostrado por Roy *et al.* (2018). Buscando complementar as análises de redes neurais e oferecer outras alternativas dentro do campo da inteligência artificial, as técnicas de aprendizado profundo foram testadas como ferramentas de predição de fraude.

O objetivo desse capítulo é abordar outras técnicas de aprendizado de máquina e aprendizado profundo, além das redes neurais como técnicas para predição de fraude de adição de leite bovino em leite bubalino. Para isso, foi utilizada a plataforma RapidMiner®.

4.2 RAPIDMINER E MODELOS DE CLASSIFICAÇÃO

RapidMiner é uma plataforma de preparação de dados projetada para acelerar esse processo com facilidade de uso por analistas e pesquisadores sem necessidade de qualquer base extremamente técnica. Essa ferramenta fornece funções que incluem transformação, limpeza,

geração, mesclagem e permite que os usuários façam alterações passo a passo, criando em segundo plano um processo que pode ser aplicado posteriormente a conjuntos de dados semelhantes.

O software possui uma ferramenta de preparação de dados multifuncional que ajuda a fazer qualquer tipo de preparação de dados, seja limpeza ou transformação, além de ferramentas de visualização com uma ampla variedade de diagramas. O pré-processamento envolve a formatação dos dados inseridos de acordo com seu tipo, bem como a limpeza dos dados, tratamento de dados ausentes, seleção de atributos ou seleção de instâncias. Esta primeira fase é crucial porque a escolha dos descritores e o conhecimento preciso das variáveis a serem processadas dependerão do desenvolvimento dos modelos de previsão. As informações necessárias para construir um bom modelo de previsão podem estar disponíveis nos dados, mas a escolha errada de variáveis ou amostras de aprendizado pode fazer com que a operação falhe (Bjaoui *et al.* 2020).

Essa ferramenta fornece um ambiente integrado de aprendizado de máquina e análise preditiva que podem ser utilizados em diversas áreas com finalidades distintas, sendo amplamente difundidas inclusive na predição de fraudes, principalmente financeiras. Dentre as ferramentas disponíveis, o Auto model funciona de forma automática, onde a partir do fornecimento dos dados do conjunto de entrada, são escolhidas opções de maior relevância. A partir disso são criadas comparações das abordagens utilizando diversos tipos de algoritmo (Rapidminer, 2022).

Para o estudo foram avaliadas *Naive Bayes*, *Generalized Linear Model*, *Logistic Regression*, *Fast Large Margin*, *Deep Learning*, *Decision Tree*, *Random Forest*, *Gradient Boosted Trees* e *Support Vector Machine*.

4.2.1 Naive Bayes:

Este algoritmo é um caso especial baseado na classificação baseada no teorema de Thomas Bayes segundo uma técnica estatística (probabilidade condicional) (Zhang, 2004). Segundo esse teorema, é possível encontrar a probabilidade de um certo evento ocorrer, dada a probabilidade de um outro evento que já ocorreu. Devido à sua simplicidade e o alto poder preditivo, é um dos algoritmos mais utilizados. Estudos comparativos mostram que os *Naive Bayes* obtiveram resultados compatíveis com os modelos de árvore de decisão e redes neurais.

Essa técnica parte do pressuposto que não exista relação de dependência entre os atributos, entretanto nem sempre isso é possível. Nesses casos em particular, é utilizada uma variação conhecida como *Bayesian Belief Networks*, ou *Bayesian Networks* que possibilita a classificação dessas variáveis (Niedermayer, 2008). O *Naive Bayes* é, portanto, um modelo fácil para construção do classificador e adequado para grandes bancos de dados, pois a ideia básica é a suposição independente das características. Possui as vantagens de ser um algoritmo relativamente fácil de se construir e ter boa performance na classificação multi classes. Por outro lado, em casos de frequência zero é necessária uma técnica de suavização (Jeyaraj *et al.*, 2021).

4.2.2 Generalized Linear Model:

O *Generalized Linear Model* (GLM) permite generalizar a regressão linear simples para variáveis de resposta que têm padrões de distribuição de erros diferentes da distribuição normal, como uma distribuição gaussiana. O modelo linear generalizado assume que cada resultado das variáveis dependentes é formado a partir de uma distribuição conhecida em uma família exponencial, uma grande distribuição de probabilidade que inclui as distribuições usual, binomial, Poisson e gama (Alisherov *et al.*, 2021). Este modelo é composto por 3 componentes: família exponencial de distribuições de probabilidade, preditivo linear e função de conexão.

Possui as vantagens de flexibilidade na modelagem com modelo ajustado para as propriedades ótimas do estimador. Por outro lado, a função linear fornecerá apenas um estimador e as respostas devem ser independentes (Jeyaraj *et al.*, 2021).

4.2.3 Logistic Regression:

Logistic regression (LR) ou regressão logística é uma parte da análise de regressão que é usada quando a variável dependente (resposta) é uma variável dicotômica. As variáveis dicotômicas geralmente consistem apenas em dois valores que representam a ocorrência ou ausência de um evento que geralmente recebe um número 0 ou 1 (Saputra e Widodo, 2014). A regressão logística fornece uma razão de chances e um intervalo de confiança para cada preditor no modelo, ambos comumente usados e fáceis de interpretar. No entanto, as previsões baseadas em LR têm suas restrições. Primeiro, relacionamentos não lineares não podem ser incluídos

diretamente em modelos logísticos; isso restringe muitas variáveis que apresentam tais relações e exige a conversão dessas variáveis para uma escala linear, o que pode complicar as interpretações. Diante da *big data*, ou seja, dados com maior variedade que chegam a volumes crescentes as previsões de LR podem não ser tão precisas quanto as abordagens de mineração de dados (Alharthi, 2018).

4.2.4 *Fast Large Margin:*

O conceito de uma *Large Margin* foi identificado como um princípio para classificar dados com base em uma margem de classificação (ou seja, um parâmetro de escala) ao invés de um erro de treinamento bruto. A margem da classificação é determinada principalmente pela localização da função de decisão longe de quaisquer pontos de dado. As abordagens das *Fast Large Margin* estão procurando alcançar soluções de decisão resolvendo um problema de otimização quadrática restrita e impulsionando com técnicas de parada antecipada (Smola, *et al.* 2000).

O classificador de margem grande rápida considera as técnicas de otimização e convergência acima para reduzir erros generalizados e maximizar a margem de separação de hiperplanos. Tais algoritmos podem economizar tempo e recursos ao otimizar o treinamento. Esse algoritmo tem muitas vantagens, primeiro, ele converge em um número finito de atualizações. Em segundo lugar, a solução é otimizada para fornecer a margem máxima possível. Resultados experimentais usando um algoritmo *fast large margin* mostram maior desempenho e convergência mais rápida em comparação com *Support Vector Machines* e outros sistemas de última geração (Smola, *et al.* 2000).

4.2.5 *Deep Learning:*

Deep Learning ("aprendizado profundo") faz parte de uma família de métodos de aprendizado de máquina baseados em redes neurais artificiais. Essa técnica é usada principalmente nas seguintes áreas: bioinformática, indústria farmacêutica, diagnóstico médico por imagem, validação de materiais e desenvolvimento de programas e recomendações para jogos (Schmidhuber, 2015). *Deep Learning* é um algoritmo de aprendizado de máquina que usa

várias camadas para extrair, passo a passo, dependências de nível superior de dados brutos. O modelo “recomendações baseadas em conteúdo” também utiliza o *Deep Learning* para criar seus recursos. O modelo utiliza uma abordagem híbrida e orientada ao conteúdo e serve para melhorar as recomendações em diversas tarefas (Alisherov, 2021).

4.2.6 *Decision Tree:*

Decision Tree ou Árvores de decisão são modelos estatísticos que utilizam treinamento supervisionado para classificação e predição dos dados. Dessa forma, no conjunto de treinamento as variáveis preditivas Y são conhecidas. Uma árvore de decisão possui uma estrutura de árvore, onde cada nó interno (não-folha), pode ser entendido como um atributo de teste, e cada nó-folha (nó-terminal) possui um rótulo de classe (Figura 4-1) (Han e Kamber 2000).

As ligações entre os nós representam os valores possíveis do teste do nó superior, e as folhas indicam a classe (categoria) a qual o registro pertence. Após a árvore de decisão montada, para classificarmos um novo registro, basta seguir o fluxo na árvore (mediante os testes nos nós não-folhas) começando no nó raiz até chegar a uma folha. Pela estrutura que formam, as árvores de decisões podem ser convertidas em Regras de Classificação. Apesar de ser uma técnica extremamente poderosa, é necessária uma análise detalhada dos dados que serão usados para garantir bons resultados (Camilo e Silva, 2009). Uma das vantagens do método da árvore de decisão é que ele ordena o valor mais significativo e depende do algoritmo (Alisherov, 2021).

O tamanho da árvore impacta na classificação final e uma árvore muito grande geralmente possui grande acurácia, mas pequena generalização, dessa forma se torna muito particular para um subconjunto de problemas. Esse problema é conhecido como *overfitting*, onde o domínio de resolução do algoritmo de árvore se torna restrito a uma subclasse de problemas. Para a resolução desses problemas existem algoritmos de poda (*pruning*) que reduzem o tamanho da árvore e, conseqüentemente, aprimoram a capacidade de generalização do algoritmo (Gomes, 2019).

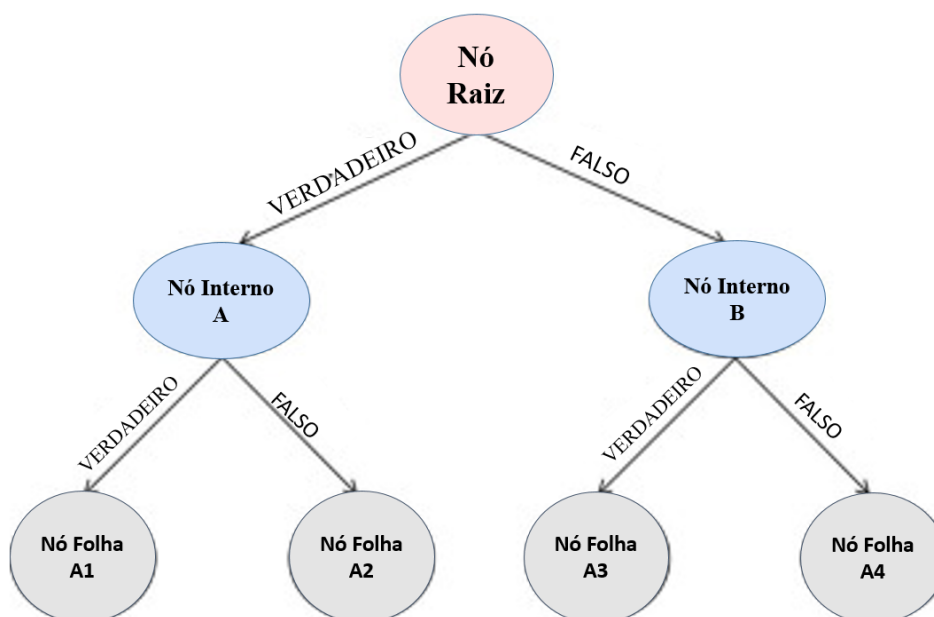


Figura 0-1: Esquema representativo do funcionamento de uma *Decision Tree* em relação às tomadas de decisão desde o nó raiz até os nós folhas em diferentes tipos de resposta.

Fonte: Adaptado de Sá, *et al.* (2016)

4.2.7 *Random Forest*:

Random Forest (RF) ou *Random Resolution Forest* é um método de aprendizado que usa algoritmos mistos para alcançar os melhores resultados para classificação, regressão e outros problemas. Este método é comumente usado durante o treinamento para construir muitas árvores de decisão e busca combater o problema de enviesamento que pode ocorrer nas árvores de decisão (Alisherov, 2021). Este algoritmo busca diluir o viés de classificação das árvores de decisão através de uma técnica de votação e distribuição de pesos para as árvores que o compõe. Além disso, outro problema que este algoritmo busca combater é o *overfitting*. Através da adição de um componente de aleatoriedade na divisão e na seleção dos subconjuntos de características, o algoritmo RF conseguirá ajustar melhor a divisão das árvores de decisão, evitando assim o crescimento exagerado das árvores e, conseqüentemente, o *overfitting* (Gomes, 2019).

4.2.8 Gradient Boosted Trees:

Este método é formalizado no algoritmo “AdaBoost” de Freund e Schapire, originalmente aplicado a problemas de classificação. O reforço funciona aplicando sequencialmente ao modelo que se quer ensinar versões repetidamente ponderadas dos dados de treinamento. Após cada interação de reforço, os exemplos classificados incorretamente têm seus pesos aumentados e, corretamente, os exemplos classificados têm seus pesos diminuídos. Assim, cada classificador sucessivo se concentra em exemplos que foram difíceis de classificar nas etapas anteriores (Alisherov, 2021).

4.2.9 Support Vector Machine:

Support Vector Machine (SVM) é um algoritmo de aprendizado desenvolvido com base na *Statistical Learning Theory* (STL) estabelecida por Vapnik, *et al.* (1995). Baseia-se na teoria da dimensão e no princípio da minimização do risco estrutural. Com base nas informações limitadas da amostra, se busca o melhor compromisso entre a complexidade do modelo e a capacidade de aprendizado, o que supera em muito os problemas dimensionais e de aprendizado de máquina tradicional, de modo a obter melhor capacidade de generalização. O SVM é usado para resolver problemas de classificação, encontrando o hiperplano com o maior espaço de separação (delimitado por suas margens) entre as categorias da variável escolhida, tendo sua dimensão definida pela dimensão dos dados. O hiperplano é a equação que separa as duas classes, enquanto os vetores de suporte são os pontos de dados mais próximos dos limites das margens (Kamilov *et al.*, 2019) SVMs têm uma ótima capacidade de generalização e são particularmente eficazes quando o número de características é alto e o número de pontos de dados é baixo. Esse método é, por natureza, um tipo de classificador binário, entretanto, pode ser utilizado em conjunto para realizar classificações multi-classes (Alisherov, 2021).

4.3 Curva ROC e AUC:

A curva ROC (*Receiver Operating Characteristics*) e AUC (*Area Under the Curve*) são duas métricas utilizadas em modelos de classificação, para avaliar a relação entre a sensibilidade e a especificidade.

A Curva ROC é uma alternativa de avaliação do classificador, que consiste na utilização de um gráfico bidimensional, onde o eixo vertical representa a taxa de verdadeiros positivos e o eixo horizontal representa a taxa de falsos positivos obtida pelo classificador (Prati, 2008). Essa curva é calculada no intervalo crescente dos limiares, através da plotagem dos valores da taxa de verdadeiros positivos (sensibilidade) e da taxa de falsos positivos (especificidade) (Fawcett, 2006)

A área sob a curva ROC, representada pela sigla AUC (*Area Under the ROC Curve*), é uma maneira eficaz de resumir a precisão diagnóstica geral do teste, por oferecer uma medida condensada de desempenho relacionada a todos os possíveis limiares de classificação. Ele assume valores de 0 a 1, onde um valor de 0 indica um teste perfeitamente impreciso e um valor de 1 reflete um teste perfeitamente preciso. De forma geral, uma AUC de 0,5 não sugere discriminação (ou seja, capacidade de identificar ou não a fraude com base no teste), 0,7 a 0,8 é considerado aceitável, 0,8 a 0,9 é considerado ótimo, e mais de 0,9 é considerado excelente (Hosmer e Lemeshow, 2000). Nessas curvas, quanto mais próximas ao canto superior esquerdo a curva se encontra, melhor é o resultado do teste. Essas curvas foram utilizadas na comparação entre os algoritmos e testes utilizados pela plataforma RapidMiner.

4.4 MATERIAL E MÉTODOS

Foram utilizados os dados obtidos a partir análise FTIR das amostras puras e fraudadas como descrito no Capítulo 1 para a alimentação da plataforma RapidMiner. A fim de minimizar os erros experimentais, foi utilizado o banco de dados balanceado composto por três grupos de amostragem descrito no Capítulo 2. A partir da inserção dos dados na plataforma, os testes foram feitos no modo *Auto model*, uma configuração da plataforma que permite que muitos testes sejam rodados simultaneamente utilizando a base de dados inserida na plataforma. Nessa configuração foram testados nove modelos de classificação: *Naive Bayes*, *Generalized Linear Model*, *Logistic Regression*, *Fast Large Margin*, *Deep Learning*, *Decision Tree*, *Random Forest*, *Gradient Boosted Trees* e *Support Vector Machine*, com divisão dos dados 90% para treinamento e 10% para teste.

4.4.1 Análise estatística

As métricas de performance dos métodos foram avaliadas segundo acurácia, precisão, sensibilidade e especificidade, onde:

$$\text{Acurácia} = \frac{\text{amostras classificadas corretamente}}{\text{total de amostras}}$$

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

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

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

4.5 RESULTADOS E DISCUSSÃO:

Os dados obtidos na análise FTIR foram utilizados para análise pela plataforma RapidMiner. Foram utilizados na alimentação do programa os dados do grupo de amostras balanceado para redução dos possíveis erros de amostragem e indução dos resultados. Os resultados obtidos na análise pelo software foram computados e tabelados e serão discutidos de acordo com as porcentagens de fraude. Para os cálculos de performance, é feito no programa o cálculo do desempenho de sete subconjuntos de dados amostrados aleatoriamente de acordo com os dados inseridos na plataforma. O maior e o menor resultado são desconsiderados e é realizada uma média com os cinco desempenhos restantes gerando o resultado final da análise.

4.5.1 Fraude 1%

Os nove métodos realizados tiveram os resultados dispostos na Tabela 4-1:

Tabela 0-1: Comparativo dos métodos de mineração de dados analisados pelo RapidMiner® em relação à acurácia, erro, precisão, sensibilidade, especificidade e tempo de análise em amostras de leite com adição de 1% de leite bovino em leite bubalino simulando fraude 1%

Modelo	Acurácia	Erro	Precisão	Sensibilidade	Especificidade	Tempo
Naive Bayes	94,2%	5,8%	95,7%	94,9%	93%	3min 42s
Generalized Linear Model	92,7%	7,3%	90,6%	98,3%	84,2%	6min 48s
Logistic Regression	94,5%	5,5%	93,4%	97,9%	89,4%	6min 44s
Fast Large Margin	93,2%	6,8%	91,9%	97,4%	86,7%	6min 56s
Deep Learning	94,0%	6,0%	93,7%	96,5%	90,1%	7min 21s
Decision Tree	94,5%	5,5%	95,3%	95,7%	92,7%	6min 46s
Random Forest	95,6%	4,4%	96,2%	96,6%	94,1%	7min 48s
Gradient Boosted Trees	95,5%	5,0%	95,8%	96,2%	93,3%	7min 34s
Support Vector Machine	90,8%	9,2%	100,0%	85,0%	100,0%	7min 00s

Fonte: A autora (2022)

De acordo com as análises realizadas, todos os testes tiveram bom desempenho com acurácia acima de 90%, indicando que todos são capazes de realizar boas predições de fraude mesmo com uma baixa porcentagem de leite bovino adicionado às amostras. A melhor acurácia foi do teste *Random Forest*, que apresentou 95,6% de acurácia, representando a melhor performance dos testes comparados. Consequentemente, esse teste também mostrou a menor taxa de erro com 4,4%.

Em relação à análise de precisão, nota-se que os valores atingidos por todos os testes foram superiores a 90% indicando pouca variabilidade dos resultados. Dentre os testes, a melhor performance é do *Support Vector Machine* com 100% seguido pelo *Random Forest* com 96,2%. Quanto ao tempo de análise, o teste com resultado mais rápido foi o *Naive Bayes*.

Para o teste de sensibilidade apenas um teste atingiu valores menores de 90%, o teste *Support Vector Machine* com 85%. O melhor resultado foi do *Generalized Linear Model* com 98,3%, indicando a boa capacidade desse teste de identificar as amostras fraudadas mesmo com baixas porcentagens de adição de leite bovino.

O teste com melhor capacidade de identificação de leite bubalino sem adição foi o *Support Vector Machine* com 100% de especificidade. Por outro lado, o *Generalized Linear Model* apresentou a pior performance dos testes comparados, com 84,2%.

A comparação geral entre os testes foi realizada com a curva ROC e a AUC, como demonstrado nas Figuras 4-2 e 4-3.

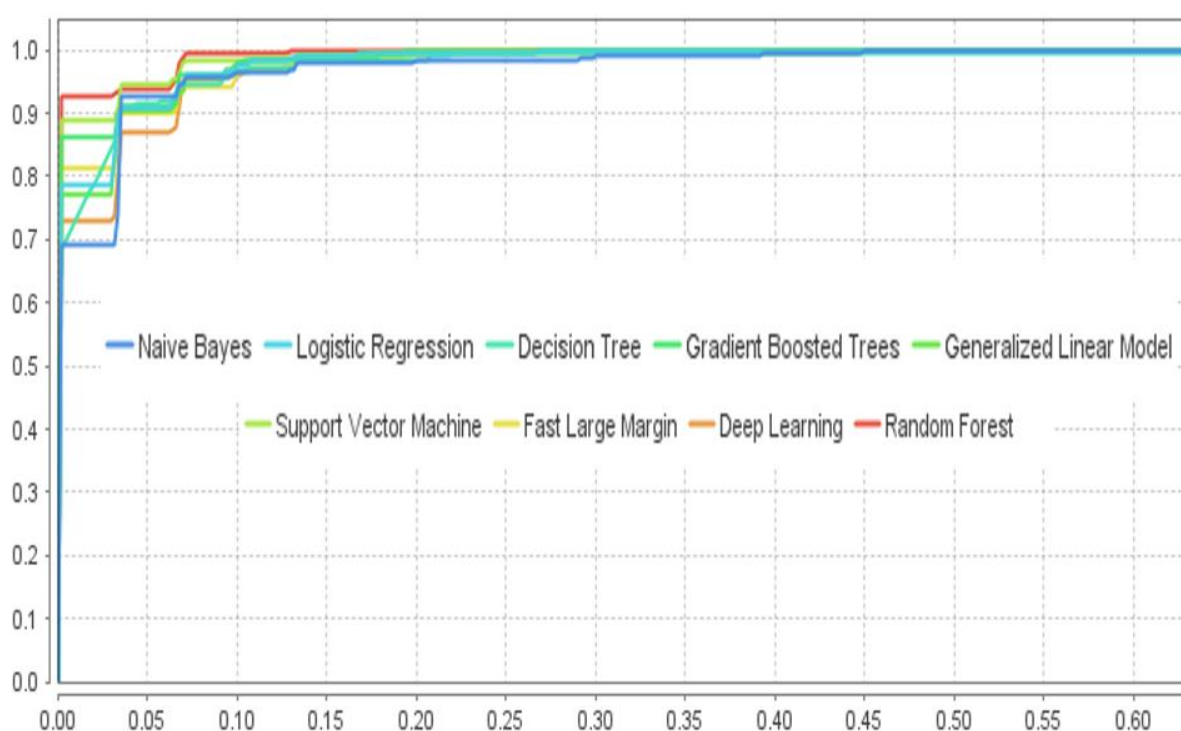


Figura 0-2: Curva ROC relativa aos testes de mineração de dados feita pela plataforma RapidMiner® em amostras com 1% de adição de leite bovino em leite bubalino simulando fraude.

Fonte: A autora (2022)

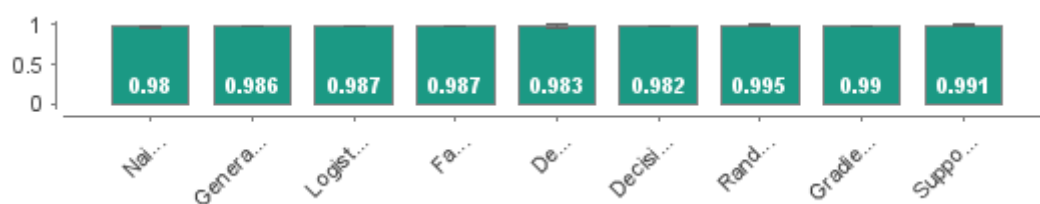


Figura 0-3: Área Under the Curve relativa aos testes de mineração de dados feitos pelo Rapidminer® utilizando dados FTIR relativos à composição do leite em amostras com 1% de adição de leite bovino em leite bubalino simulando fraude. Em ordem: Naive Bayes, Generalized Linear Model, Logistic Regression, Fast Large Margin, Deep Learning, Decision Trees, Random Forest, Gradient Boosted Trees e Support Vector Machine

Fonte: A autora (2022)

Comparando os testes através das curvas e das áreas sob a curva, os melhores resultados são evidenciados pelo *Random Forest*, que apresentou a curva mais alta e a maior AUC. A pior performance foi do teste *Naive Bayes* com a curva mais baixa e a menor AUC.

4.5.2 Fraude 2%

Os nove modelos de métodos realizados tiveram os resultados dispostos na Tabela 4-2.

Tabela 0-2: Comparativo dos métodos de mineração de dados analisados pelo RapidMiner® em relação à acurácia, erro, precisão, sensibilidade, especificidade e tempo de análise em amostras de leite com adição de 2% de leite bovino em leite bubalino simulando fraude.

Modelo	Acurácia	Erro	Precisão	Sensibilidade	Especificidade	Tempo
Naive Bayes	90,8%	9,2%	99,0%	85,5%	99%	6min 43s
Generalized Linear Model	93,4%	6,6%	95,6%	93,4%	93,5%	7min 09s
Logistic Regression	93,2%	6,8%	95,5%	92,9%	93,5%	7min 12s
Fast Large Margin	94,0%	6,0%	94,0%	96,0%	90,9%	7min 12s
Deep Learning	94,0%	6,0%	99,4%	90,7%	98,8%	7min 23s
Decision Tree	95,8%	4,2%	95,1%	98,2%	92,4%	6min 54s
Random Forest	93,2%	6,8%	98,5%	89,8%	98,1%	7min 02s
Gradient Boosted Trees	96,3%	3,7%	97,3%	96,4%	96,3%	8min 00s
Support Vector Machine	84,6%	15,4%	98,9%	74,9%	98,7%	7min 35s

Fonte: A autora (2022)

O teste com melhor performance em relação à acurácia foi o *Gradient Boosted Trees*, com 96,3% de acerto. Apenas um teste apresentou valores inferiores a 90%, o *Support Vector Machine* com 84,6% de acurácia. Conseqüentemente esses testes tiveram a menor e a maior taxa de erro, com 3,7% e 15,4% respectivamente.

A melhor precisão foi alcançada com o *Deep Learning*, com 99%. Todos os testes tiveram valores acima de 90% indicando uma baixa variação nos resultados. O teste com pior precisão foi o *Fast Large Margin* com 94%.

Para a sensibilidade, três testes apresentaram valores inferiores a 90% representando as piores performances. *Support Vector Machine* (74,9%), *Naive Bayes* (85,5%) e *Random Forest* (89,8%). Esses testes comparados aos outros, não apresentaram boa capacidade de identificação das amostras fraudadas a 2%. Por outro lado, *Decision Tree* apresentou a melhor performance com 98,2% de acerto na detecção de fraude quando a amostra realmente continha fraude.

Da mesma forma que na análise de precisão, o *Deep Learning* foi o teste com a melhor capacidade de identificação de amostras não fraudadas, apresentando 98,8% de especificidade e o *Fast Large Margin* o pior resultado, com 90,9%. Em relação ao tempo de análise, o teste com resultado mais rápido foi o *Naive Bayes*.

As curvas ROC e AOC relativas à análise de 2% estão dispostas nas Figuras 4-4 e 4-5.

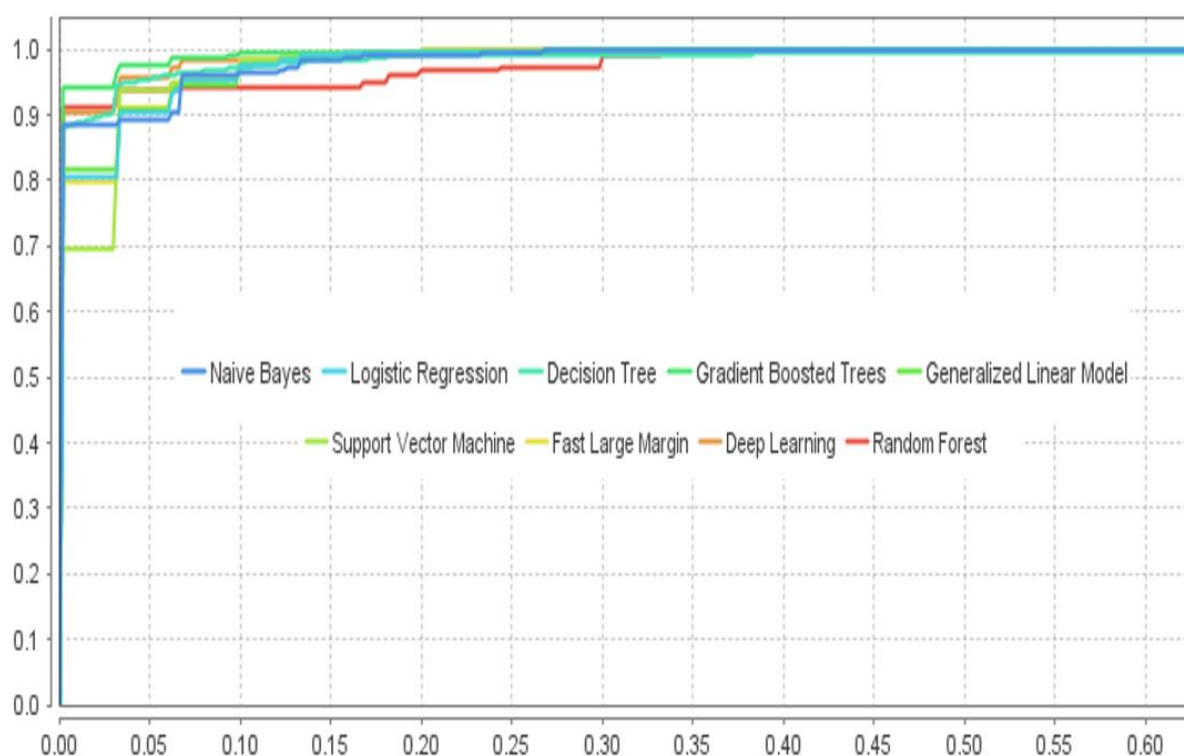


Figura 0-4: Curva ROC relativa aos testes de mineração de dados feita pela plataforma RapidMiner® em amostras com 2% de adição de leite bovino em leite bubalino simulando fraude.

Fonte: A autora (2022)

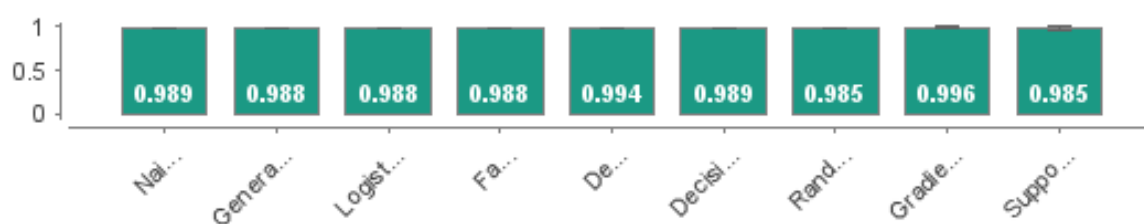


Figura 0-5: Area Under the Curve relativa aos testes de mineração de dados feitos pelo Rapidminer® utilizando dados FTIR relativos à composição do leite em amostras com 2% de adição de leite bovino em leite bubalino simulando fraude. Em ordem: Naive Bayes, Generalized Linear Model, Logistic Regression, Fast Large Margin, Deep Learning, Decision Trees, Random Forest, Gradient Boosted Trees e Support Vector Machine

Fonte: A autora (2022)

Comparando as curvas feitas no gráfico demonstrado na Figura 4-5, o teste com pior performance para a detecção de fraude a 2% é o *Support Vector Machine* que apresenta a curva mais baixa e com menor AUC e maior desvio padrão. O melhor resultado, por outro lado é o do *Gradient Boosted Trees* com maior curva e consequentemente maior área sob a curva.

4.5.3 Fraude 5%:

Os nove testes realizados tiveram os resultados dispostos na Tabela 4-3.

Tabela 0-3: Comparativo dos métodos de mineração de dados analisados pelo RapidMiner® em relação à acurácia, erro, precisão, sensibilidade, especificidade e tempo de análise em amostras de leite com adição de 5% de leite bovino em leite bubalino simulando fraude

Modelo	Acurácia	Erro	Precisão	Sensibilidade	Especificidade	Tempo
Naive Bayes	93,9%	6,1%	98,1%	91,4%	97,5%	6min 54s
Generalized Linear Model	94,5%	5,5%	96,3%	94,2%	95,0%	7min 20s
Logistic Regression	95,3%	4,7%	95,9%	96,0%	94,4%	7min 05s
Fast Large Margin	95,0%	5,0%	95,9%	95,5%	94,4%	7min 17s
Deep Learning	95,3%	4,7%	96,3%	95,5%	95,0%	7min 54s
Decision Tree	95,3%	4,7%	96,8%	95,1%	95,7%	7min 09s
Random Forest	96,3%	3,7%	97,7%	96,0%	96,9%	7min 48s
Gradient Boosted Trees	96,1%	3,9%	98,6%	94,6%	98,1%	7min 57s
Support Vector Machine	95,0%	5,0%	96,8%	94,5%	95,7%	7min 32s

Fonte: A autora (2022)

Todos os testes a nível de 5% de fraude apresentaram bons resultados de acurácia, sendo todos acima de 90%. O teste com melhor performance de acurácia foi o *Random Forest* com 96,3% de acerto do teste e conseqüentemente menor porcentagem de erro. Em contrapartida. A pior performance foi do *Generalized Linear Model* com 94,5%. Em relação ao tempo de análise, o teste com resultado mais rápido foi o *Naive Bayes*.

Todos os resultados apresentaram precisão elevada, acima de 95%, indicando baixa variabilidade dos resultados encontrados. O teste com maior precisão foi o *Gradient Boosted Trees* com 98,6% enquanto *Logistic Regression* e *Fast Large Margin* dividiram o mesmo valor com a pior performance para a métrica, com 95,9%.

Todos os testes apresentaram resultados elevados em relação a sensibilidade, demonstrando a boa capacidade dos algoritmos de detectarem leite com 5% de adulteração. *Logistic Regression* e *Random Forest* apresentaram o melhor nível de sensibilidade com 96% enquanto *Naive Bayes* apresentou o resultado mais baixo dentre os testes comparados com 91,4%.

Em relação à especificidade da mesma forma que a precisão, o melhor resultado foi do teste *Gradient Boosted Trees* com 98,1% e os valores mais baixos foram dos testes *Logistic Regression* e *Fast Large Margin* com 94,4%.

As curvas ROC e AOC relativas à análise de 5% estão dispostas nas Figuras 4-6 e 4-7.

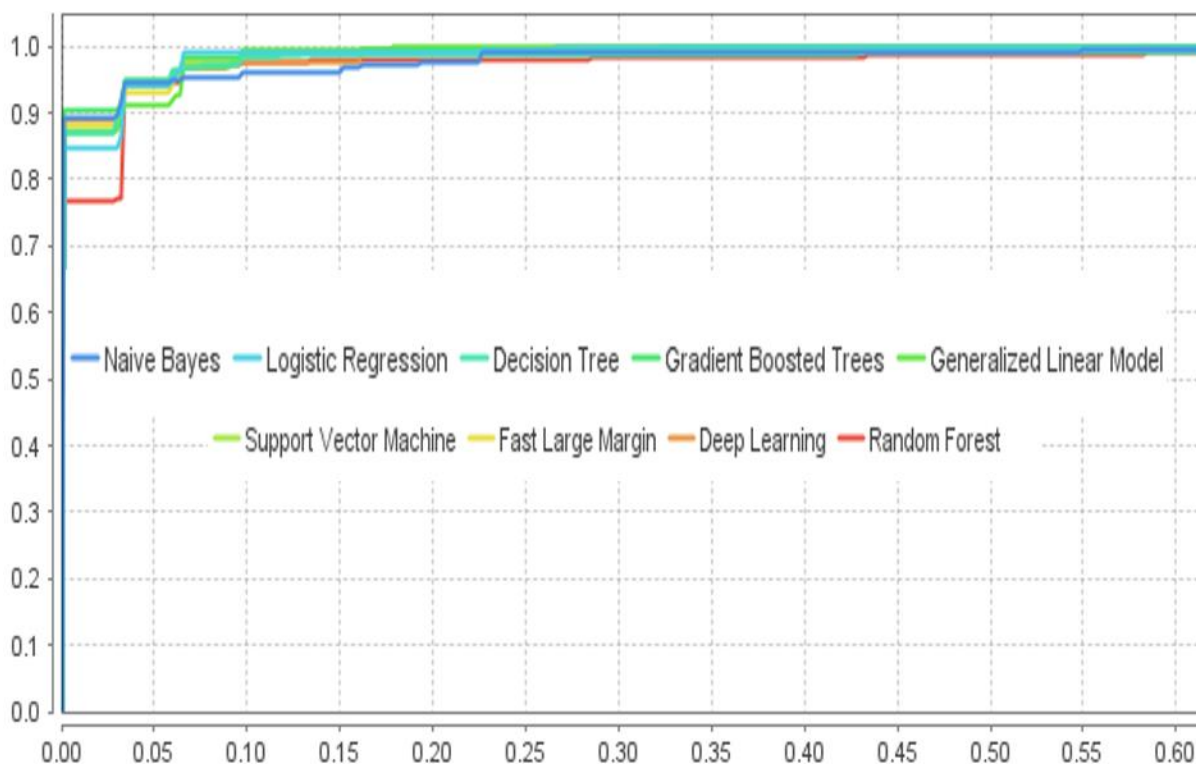


Figura 0-6: Curva ROC relativa aos testes de mineração de dados feita pela plataforma RapidMiner® em amostras com 5% de adição de leite bovino em leite bubalino simulando fraude.

Fonte: A autora (2022)



Figura 0-7: *Area Under the Curve* relativa aos testes de mineração de dados feitos pelo Rapidminer® utilizando dados FTIR relativos à composição do leite em amostras com 5% de adição de leite bovino em leite bubalino simulando fraude. Em ordem: *Naive Bayes*, *Generalized Linear Model*, *Logistic Regression*, *Fast Large Margin*, *Deep Learning*, *Decision Trees*, *Random Forest*, *Gradient Boosted Trees* e *Support Vector Machine*.

Fonte: A autora (2022)

Nota-se observando as curvas ROC que a medida que a porcentagem de fraude vai aumentando, as curvas de forma geral tendem a se aproximar mais do canto superior esquerdo indicando uma melhor performance dos testes. As curvas a nível de 5% de fraude se tornaram menos espaçadas indicando maior similaridade nos resultados dos testes. Apesar das diferenças entre os testes, todos apresentaram bons resultados. Valores acima de 0,9 de AUC são considerados excelentes em relação à precisão do teste (Hosmer e Lemeshow, 2000). O teste com pior performance a nível de 5% foi o *Random Forest*, com curva mais baixa e menor área sob a curva (0,979). Por outro lado, a melhor performance foi a do *Gradient Boosted Trees* com maior curvatura e maior área sob a curva.

4.5.4 Fraude 10%

Os nove testes realizados tiveram os resultados dispostos na Tabela 4-4.

Tabela 0-4: Comparativo dos métodos de mineração de dados analisados pelo RapidMiner® em relação à acurácia, erro, precisão, sensibilidade, especificidade e tempo de análise em amostras de leite com adição de 10% de leite bovino em leite bubalino simulando fraude

Modelo	Acurácia	Erro	Precisão	Sensibilidade	Especificidade	Tempo
Naive Bayes	96,9%	3,1%	96,0%	98,6%	94,6%	7min 03s
Generalized Linear Model	94,0%	6,0%	90,3%	100,0%	86,3%	7min 23s
Logistic Regression	97,1%	2,9%	97,3%	97,6%	96,3%	6min 52s
Fast Large Margin	94,2%	5,8%	91,2%	99,5%	87,2%	7min 04s
Deep Learning	94,0%	6,0%	90,7%	99,5%	86,8%	7min 08s
Decision Tree	94,8%	5,2%	92,7%	98,6%	89,8%	7min 04s
Random Forest	97,4%	2,6%	96,8%	98,6%	95,8%	7min 47s
Gradient Boosted Trees	93,9%	6,1%	90,5%	100,0%	85,8%	7min 54s
Support Vector Machine	97,9%	2,1%	100,0%	96,3%	100,0%	7min 08s

Fonte: A autora (2022)

O teste *Support Vector Machine* apresentou a melhor taxa de acurácia, precisão e especificidade e menor erro, se mostrando, portanto, uma excelente ferramenta de detecção de fraude a 10%. Por outro lado, o teste com resultado mais rápido foi o *Logistic Regression*.

Em relação à análise do gráfico ROC como demonstrado na Figura 4-8, ocorreu uma aproximação das curvas em relação ao valor máximo, indicando uma boa performance do método.

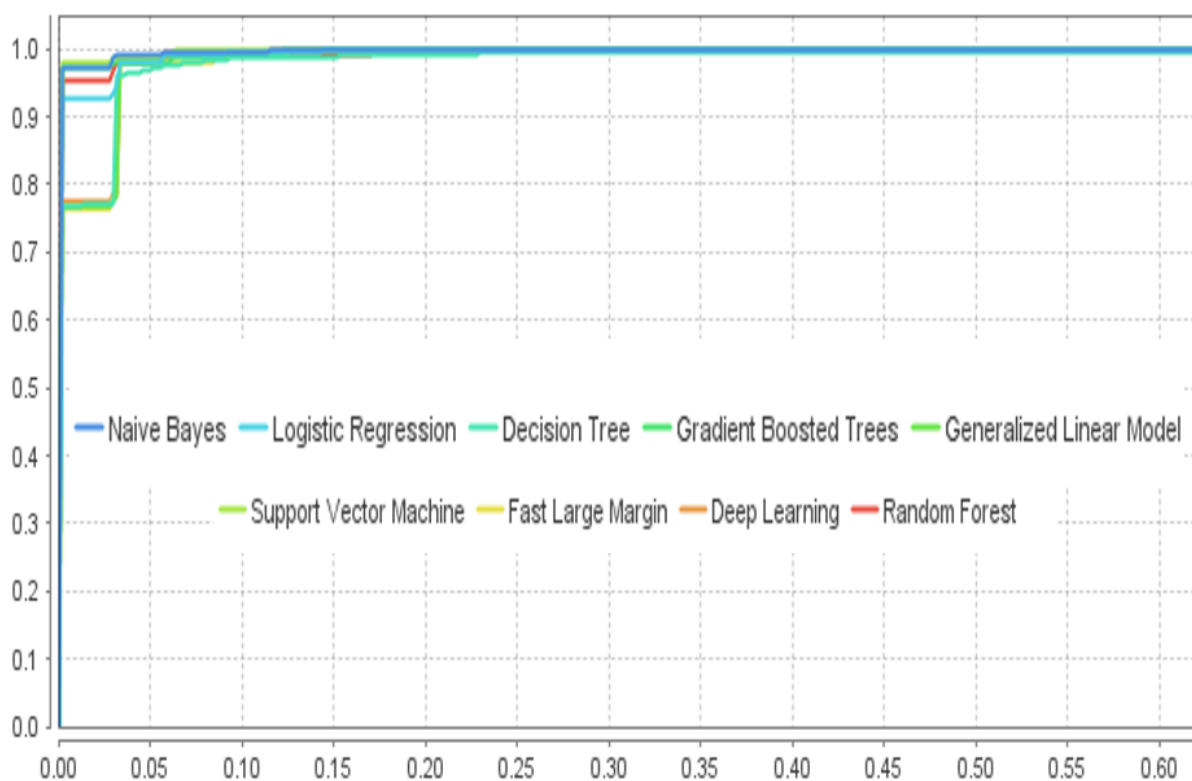


Figura 0-8: Curva ROC relativa aos testes de mineração de dados feita pela plataforma RapidMiner® em amostras com 10% de adição de leite bovino em leite bubalino simulando fraude.

Fonte: A autora (2022)

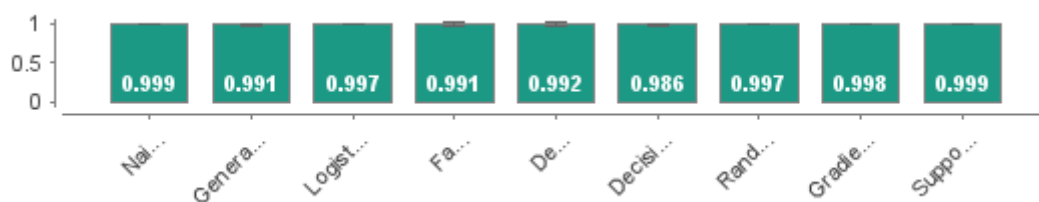


Figura 0-9 *Area Under the Curve* relativa aos testes de mineração de dados feitos pelo Rapidminer® utilizando dados FTIR relativos à composição do leite em amostras com 10% de adição de leite bovino em leite bubalino simulando fraude. Em ordem: *Naive Bayes*, *Generalized Linear Model*, *Logistic Regression*, *Fast Large Margin*, *Deep Learning*, *Decision Trees*, *Random Forest*, *Gradient Boosted Trees* e *Support Vector Machine*.

Fonte: A autora (2022)

Support Vector Machine apresentou a melhor performance na AUC com maior área (0.999), indicando a eficiência do método na detecção das fraudes. Em contrapartida, *Decision trees* apresentou a menor área, destoando no gráfico com a menor curvatura.

4.5.5 Fraude 20%

Os nove testes realizados tiveram os resultados dispostos na Tabela 4-5.

Tabela 0-5: Comparativo dos métodos de mineração de dados analisados pelo RapidMiner® em relação à acurácia, erro, precisão, sensibilidade, especificidade e tempo de análise em amostras de leite com adição de 20% de leite bovino em leite bubalino simulando fraude.

Modelo	Acurácia	Erro	Precisão	Sensibilidade	Especificidade	Tempo
Naive Bayes	96,9%	3,1%	99,0%	95,1%	98,8%	6min 52s
Generalized Linear Model	96,9%	3,1%	99,0%	95,2%	98,8%	7min 17s
Logistic Regression	96,6%	3,4%	97,6%	96,1%	97,1%	6min 47s
Fast Large Margin	96,6%	3,4%	99,0%	94,7%	98,8%	6min 50s
Deep Learning	95,5%	4,5%	99,5%	92,2%	99,4%	7min 47s
Decision Tree	97,4%	2,6%	97,6%	97,6%	97,1%	6min 54s
Random Forest	95,5%	4,5%	99,0%	92,9%	98,8%	7min 52s
Gradient Boosted Trees	96,8%	3,2%	99,0%	95,1%	98,8%	8min 20s
Suport Vector Machine	89,3%	10,7%	100,0%	80,2%	100,0%	7min 06s

Fonte: A autora (2022)

Para as fraudes a 20% de adição, o teste com melhor acurácia foi o *Decision Tree*. O teste apresentou 97,4 de acurácia com menor taxa de erro (2,6%). Apesar de não apresentar as melhores performances em todas os parâmetros, de maneira geral foi o teste que obteve melhor resultado.

Support Vector Machine apresentou os melhores resultados de precisão e especificidade e a melhor sensibilidade foi do método *Decision Tree*. Quanto ao tempo de análise, o teste com resultado mais rápido foi o *Logistic Regression*.

Em relação às curvas ROC e AUC, o método *Decision Tree* também se destacou com a curva mais próxima do nível máximo e a maior área sob a curva, como demonstrado nas Figuras 4-10 e 4-11.

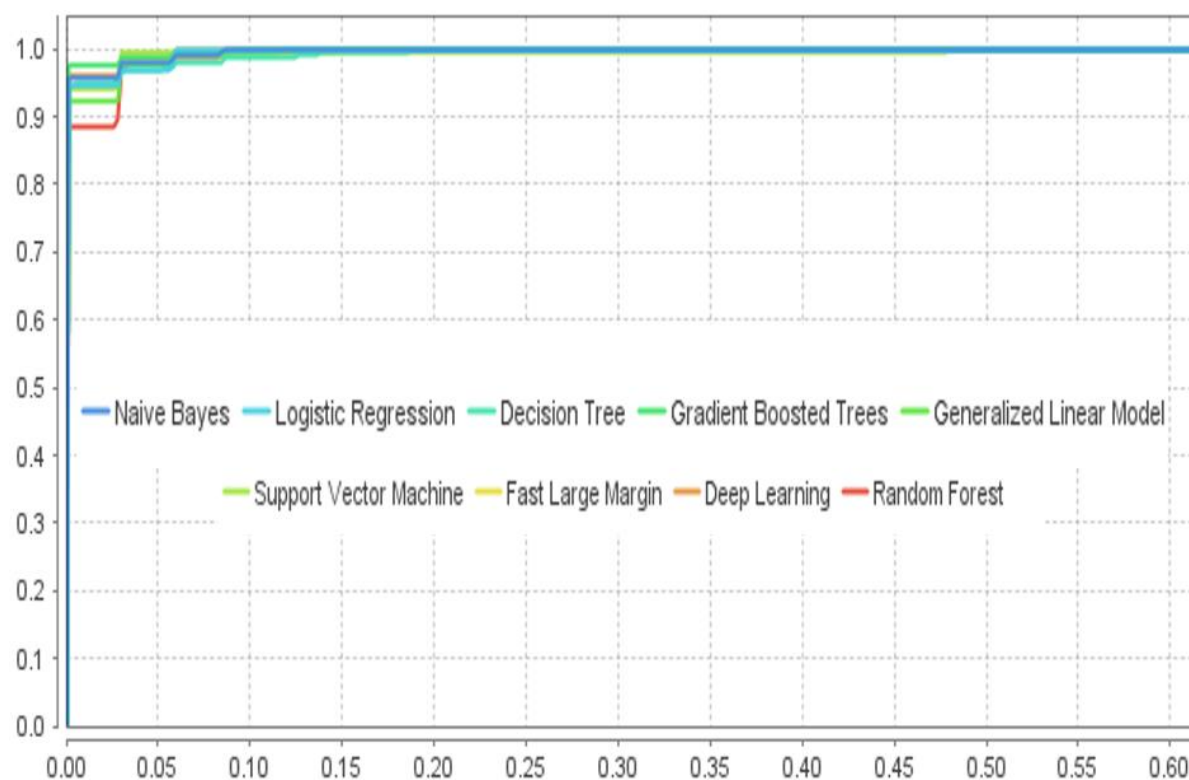


Figura 0-10: Curva ROC relativa aos testes de mineração de dados feita pela plataforma RapidMiner® em amostras com 20% de adição de leite bovino em leite bubalino simulando fraude.

Fonte: A autora (2022)

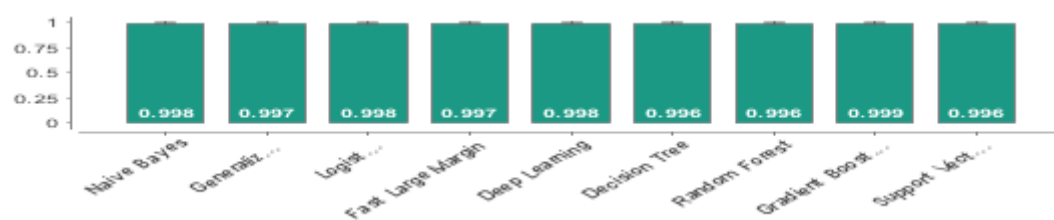


Figura 0-11: *Area Under the Curve* relativa aos testes de mineração de dados feitos pelo Rapidminer® utilizando dados FTIR relativos à composição do leite em amostras com 20% de adição de leite bovino em leite bubalino simulando fraude. Em ordem: *Naive Bayes*, *Generalized Linear Model*, *Logistic Regression*, *Fast Large Margin*, *Deep Learning*, *Decision Trees*, *Random Forest*, *Gradient Boosted Trees* e *Support Vector Machine*.

Fonte: A autora (2022)

4.5.6 Fraude 30%

Os resultados das análises estão dispostos na Tabela 4-6.

Tabela 0-6: Comparativo dos métodos de mineração de dados analisados pelo RapidMiner® em relação à acurácia, erro, precisão, sensibilidade, especificidade e tempo de análise em amostras de leite com adição de 30% de leite bovino em leite bubalino simulando fraude

Modelo	Acurácia	Erro	Precisão	Sensibilidade	Especificidade	Tempo
Naive Bayes	97,4%	2,6%	99,0%	96,0%	98,9%	7min 00s
Generalized Linear Model	97,4%	2,6%	97,5%	97,7%	97,3%	6min 47s
Logistic Regression	97,1%	2,9%	97,0%	97,7%	96,7%	6min 48s
Fast Large Margin	96,3%	3,7%	96,2%	97,2%	95,6%	6min 57s
Deep Learning	98,2%	1,8%	99,5%	97,2%	99,5%	7min 58s
Decision Tree	97,6%	2,4%	97,1%	98,5%	96,7%	6min 55s
Random Forest	96,6%	3,4%	98,5%	95,2%	98,3%	7min 59s
Gradient Boosted Trees	97,6%	2,4%	99,0%	96,5%	98,9%	8min 10s
Support Vector Machine	97,4%	2,6%	99,0%	96,0%	98,9%	7min 15s

Fonte: A autora (2022)

A medida que a porcentagem de fraude aumenta, diminui a porcentagem de erro dos testes, mostrando que quanto maior a fraude, melhor a capacidade de detecção. A acurácia de todos os testes foi elevada, porém o que apresentou uma melhor performance de forma geral foi o *Deep Learning*. Esse teste apresentou maior taxa de acurácia, precisão e especificidade, e uma taxa elevada de sensibilidade, demonstrando uma ótima capacidade de detecção das fraudes. Em relação ao tempo de análise, o teste com resultado mais rápido foi o *Generalized Linear Model*.

No comparativo entre os métodos pela curva ROC (Figura 4-12) e AUC (Figura 4-13), se observa uma maior aproximação entre todos os testes e o valor máximo da curva como evidenciado nas Figuras 4-12 e 4-13.

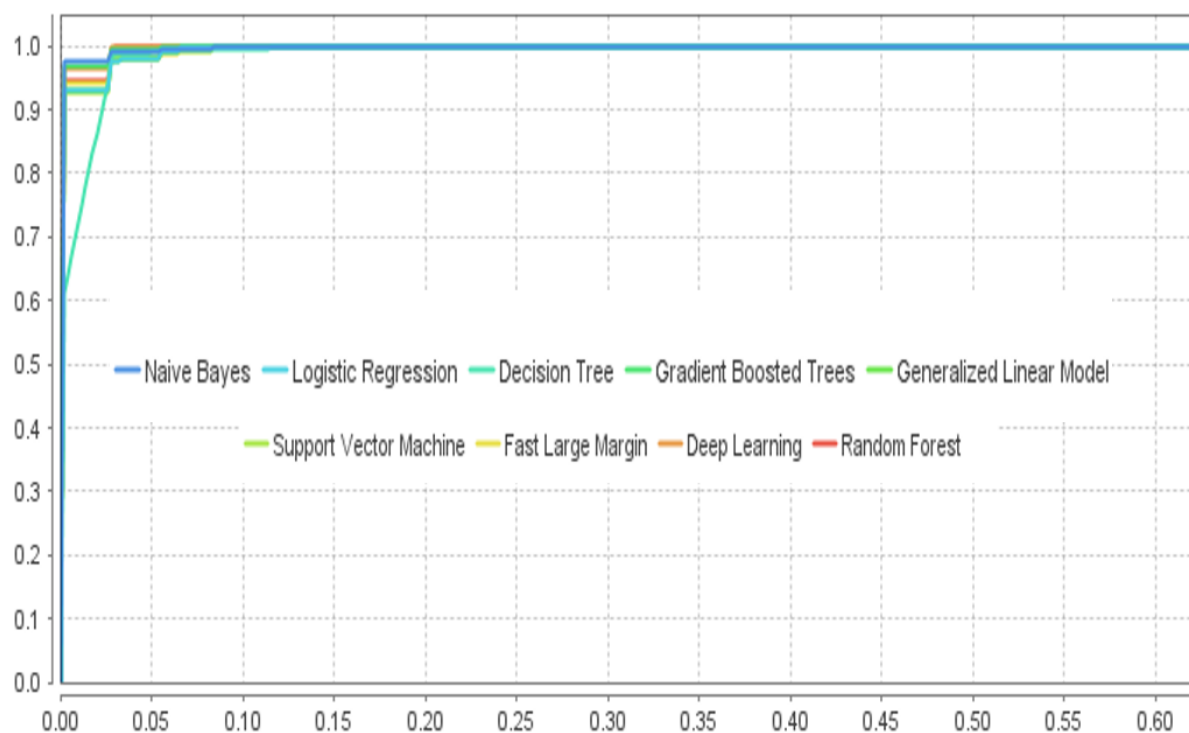


Figura 0-12: Curva ROC relativa aos testes de mineração de dados feita pela plataforma RapidMiner® em amostras com 30% de adição de leite bovino em leite bubalino simulando fraude.

Fonte: A autora (2022)

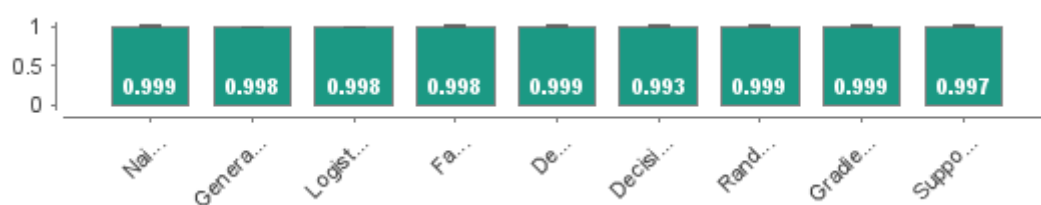


Figura 0-13: Area Under the Curve relativa aos testes de mineração de dados feitos pelo Rapidminer® utilizando dados FTIR relativos à composição do leite em amostras com 30% de adição de leite bovino em leite bubalino simulando fraude. Em ordem: *Naive Bayes*, *Generalized Linear Model*, *Logistic Regression*, *Fast Large Margin*, *Deep Learning*, *Decision Trees*, *Random Forest*, *Gradient Boosted Trees* e *Support Vector Machine*.

Fonte: A autora (2022)

A área sob a curva evidencia a proximidade dos resultados entre os testes. Apenas o teste *Decision Trees* se manteve isolado dos outros testes, porém ainda apresentou uma AUC próxima a 1. Todos os outros apresentaram variação entre 0,997 e 0,999, evidenciando a boa performance dos métodos.

4.5.7 Fraude 40%

Os dados compilados da análise foram dispostos na Tabela 4-7.

Tabela 0-7: Comparativo dos métodos de mineração de dados analisados pelo RapidMiner® em relação à acurácia, erro, precisão, sensibilidade, especificidade e tempo de análise em amostras de leite com adição de 40% de leite bovino em leite bubalino simulando fraude

Modelo	Acurácia	Erro	Precisão	Sensibilidade	Especificidade	Tempo
Naive Bayes	97,4%	2,6%	99,5%	95,4%	99,4%	6min 48s
Generalized Linear Model	98,7%	1,3%	99,0%	98,4%	98,9%	7min 13s
Logistic Regression	98,7%	1,3%	97,5%	100,0%	97,3%	6min 47s
Fast Large Margin	97,4%	2,6%	96,1%	99,0%	95,7%	7min 11s
Deep Learning	98,4%	1,6%	100,0%	96,9%	100,0%	6min 50s
Decision Tree	98,7%	1,3%	99,0%	98,4%	98,9%	6min 56s
Random Forest	99,5%	0,5%	100,0%	99,0%	100,0%	7min 19s
Gradient Boosted Trees	99,2%	0,8%	99,0%	99,5%	98,9%	8min 35s
Suport Vector Machine	99,2%	0,8%	99,0%	99,5%	99,0%	7min 32s

Fonte: A autora (2022)

A 40% de fraude os resultados são mais próximos a 100% em todas as métricas avaliadas. Aumentando a porcentagem de leite bovino nas amostras, a tendência é que ocorra o efeito de diluição dos componentes e torne o leite bubalino mais próximo ao leite bovino. Dessa forma, os testes são capazes de detectar com maior facilidade as diferenças entre os leites sem adição e fraudados. Todos os testes apresentaram acurácia superior a 97% evidenciando a boa capacidade de acerto do teste. O teste com as melhores métricas foi o *Random Forest* com acurácia de 99,5%, 0,5% de erro, 100% de precisão e especificidade e 99% de sensibilidade. Por outro lado, o teste com resultado mais rápido foi o *Logistic Regression*.

Em relação às curvas ROC e AUC, nesse nível de fraude ocorre maior sobreposição das curvas e maior aproximação em relação ao valor máximo, como evidenciado na Figura 4-12. Apenas um teste apresentou um resultado com menor sobreposição, o *Logistic Regression*, que apesar de ter apresentado uma curva mais isolada, apresentou área sob a curva 0,995 como evidenciado nas Figuras 4-14 e 4-15.

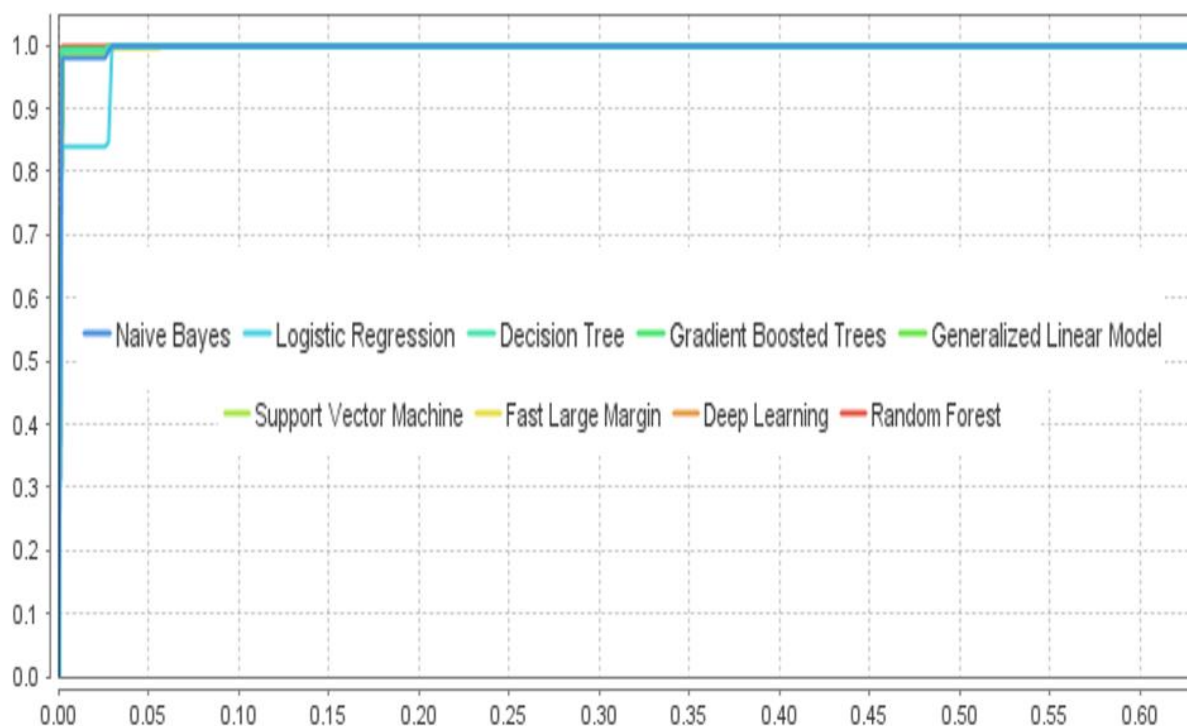


Figura 0-14: Curva ROC relativa aos testes de mineração de dados feita pela plataforma RapidMiner® em amostras com 40% de adição de leite bovino em leite bubalino simulando fraude.

Fonte: A autora (2022)

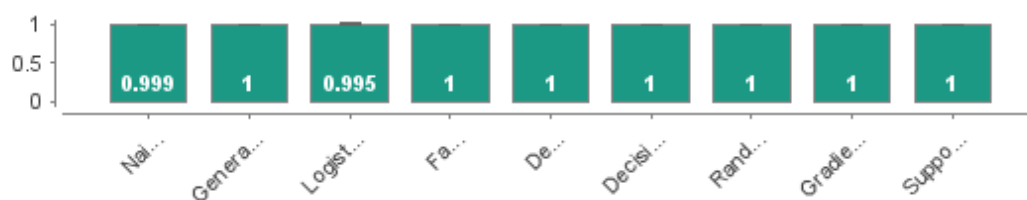


Figura 0-15: *Area Under the Curve* relativa aos testes de mineração de dados feitos pelo Rapidminer® utilizando dados FTIR relativos à composição do leite em amostras com 40% de adição de leite bovino em leite bubalino simulando fraude. Em ordem: *Naive Bayes*, *Generalized Linear Model*, *Logistic Regression*, *Fast Large Margin*, *Deep Learning*, *Decision Trees*, *Random Forest*, *Gradient Boosted Trees* e *Support Vector Machine*.

Fonte: A autora (2022)

4.5.8 Fraude 50%

Os resultados das análises estão dispostos a seguir na Tabela 4-8:

Tabela 0-8: Comparativo dos métodos de mineração de dados analisados pelo RapidMiner® em relação à acurácia, erro, precisão, sensibilidade, especificidade e tempo de análise em amostras de leite com adição de 50% de leite bovino em leite bubalino simulando fraude

Modelo	Acurácia	Erro	Precisão	Sensibilidade	Especificidade	Tempo
Naive Bayes	96,9%	3,1%	96,9%	96,9%	96,8%	7min 06s
Generalized Linear Model	97,6%	2,4%	98,5%	96,9%	98,4%	6min 46s
Logistic Regression	97,6%	2,4%	97,4%	97,9%	97,3%	6min 57s
Fast Large Margin	97,4%	2,6%	96,5%	98,5%	96,3%	6min 53s
Deep Learning	97,6%	2,4%	98,5%	96,9%	98,4%	9min 03s
Decision Tree	98,2%	1,8%	98,5%	97,9%	98,4%	7min 00s
Random Forest	97,9%	2,1%	99,0%	96,9%	98,9%	7min 12s
Gradient Boosted Trees	97,9%	2,1%	99,0%	96,9%	98,9%	8min 23s
Support Vector Machine	97,9%	2,1%	99,5%	96,4%	99,5%	7min 31s

Fonte: A autora (2022)

O modelo *Decision Tree* apresentou melhor acurácia com menor erro e apesar de não apresentar valores máximos nos outros parâmetros, foi um teste com excelente resultado. Os testes apresentaram baixa taxa de erro, mostrando a boa capacidade preditiva dos testes. Além disso, taxas de sensibilidade acima de 96% em todos os testes analisados mostram que todos são perfeitamente capazes de serem utilizados para a detecção das fraudes em leite. Em relação ao tempo de análise, o teste com resultado mais rápido foi o *Generalized Linear Model*.

A curva ROC na fraude 50% (Figura 4-16) apresentou curvas mais espaçadas em relação à fraude 40%, porém os valores da AUC (Figura 4-17) variam entre 0,996 e 1, mostrando a excelente performance dos testes (Hosmer e Lemeshow, 2000).

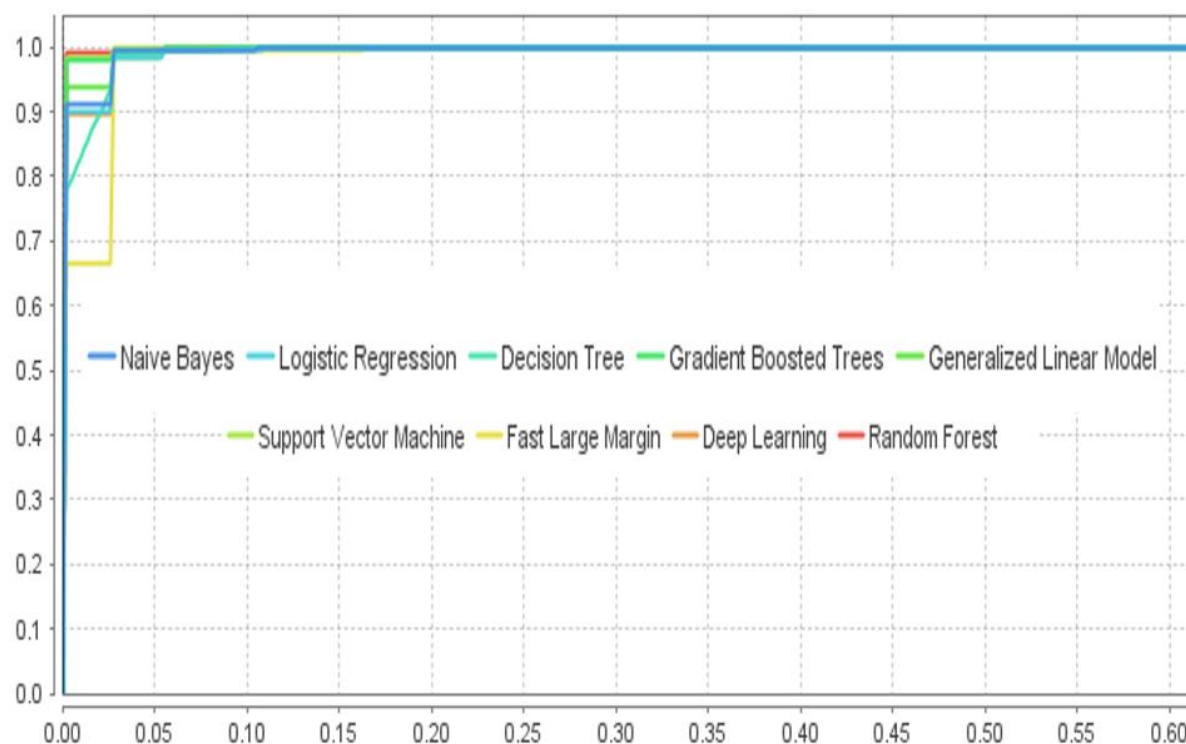


Figura 0-16: Curva ROC relativa aos testes de mineração de dados feita pela plataforma RapidMiner® em amostras com 50% de adição de leite bovino em leite bubalino simulando fraude.

Fonte: A autora (2022)

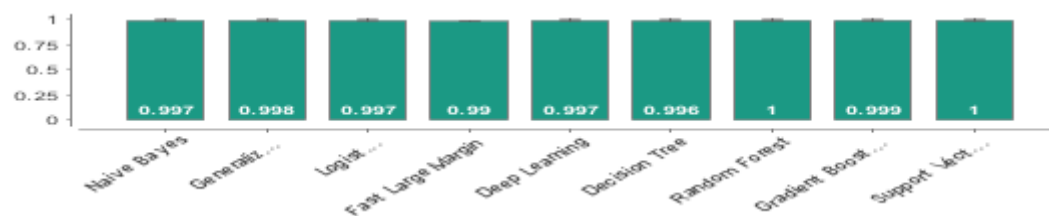


Figura 0-17: *Area Under the Curve* relativa aos testes de mineração de dados feitos pelo Rapidminer® utilizando dados FTIR relativos à composição do leite em amostras com 50% de adição de leite bovino em leite bubalino simulando fraude. Em ordem: *Naive Bayes*, *Generalized Linear Model*, *Logistic Regression*, *Fast Large Margin*, *Deep Learning*, *Decision Trees*, *Random Forest*, *Gradient Boosted Trees* e *Support Vector Machine*.

Fonte: A autora (2022)

4.5.9 Fraude 75%

Os resultados dos testes realizados pelo RapidMiner estão dispostos a seguir na Tabela 4-9.

Tabela 0-9: Comparativo dos métodos de mineração de dados analisados pelo RapidMiner® em relação à acurácia, erro, precisão, sensibilidade, especificidade e tempo de análise em amostras de leite com adição de 75% de leite bovino em leite bubalino simulando fraude

Modelo	Acurácia	Erro	Precisão	Sensibilidade	Especificidade	Tempo
Naive Bayes	99,5%	0,5%	99,5%	99,5%	99,5%	6min 50s
Generalized Linear Model	98,4%	1,6%	100,0%	97,0%	100,0%	7min 13s
Logistic Regression	89,5%	10,5%	83,3%	100,0%	78,4%	7min 07s
Fast Large Margin	99,2%	0,8%	100,0%	98,4%	100,0%	6min 57s
Deep Learning	99,5%	0,5%	100,0%	99,0%	100,0%	8min 25s
Decision Tree	99,2%	0,8%	99,5%	98,9%	99,5%	7min 22s
Random Forest	99,2%	0,8%	99,7%	98,5%	100,0%	7min 19s
Gradient Boosted Trees	99,2%	0,8%	100,0%	98,5%	100,0%	9min 09s
Support Vector Machine	99,7%	0,3%	100,0%	99,5%	100,0%	7min 38s

Fonte: A autora (2022)

Com alta porcentagem de adição de leite bovino ocorre maior diferenciação entre as amostras de leite sem adição e as amostras adulteradas experimentalmente. Como a maior parte da amostra é composta por leite bovino, ocorre maior similaridade com as amostras de leite de vaca sem adição, distanciando dos valores comuns ao leite bubalino. Dessa forma, os testes são capazes de identificar melhor as amostras fraudadas, como demonstrado nas altas taxas de acerto do teste pela acurácia e as baixas porcentagens de erro. Os altos valores de especificidade e sensibilidade demonstram também a excelente capacidade de acerto tanto das amostras não fraudadas, quanto das amostras com adição de leite bovino. Apenas um teste obteve valores discrepantes e menores em relação aos parâmetros avaliados (*Logistic Regression*). Quando ao tempo de análise, o teste com resultado mais rápido foi o *Naive Bayes*.

Essa boa performance é evidenciada nos gráficos ROC (Figura 4-18) e AUC (Figura 4-19) onde ocorreu sobreposição das curvas com muitos resultados de AUC =1. As variações entre os resultados de AUC entre 0,998 e 1 são parâmetros que permitem classificar os testes como altamente precisos (Hosmer e Lemeshow, 2000).

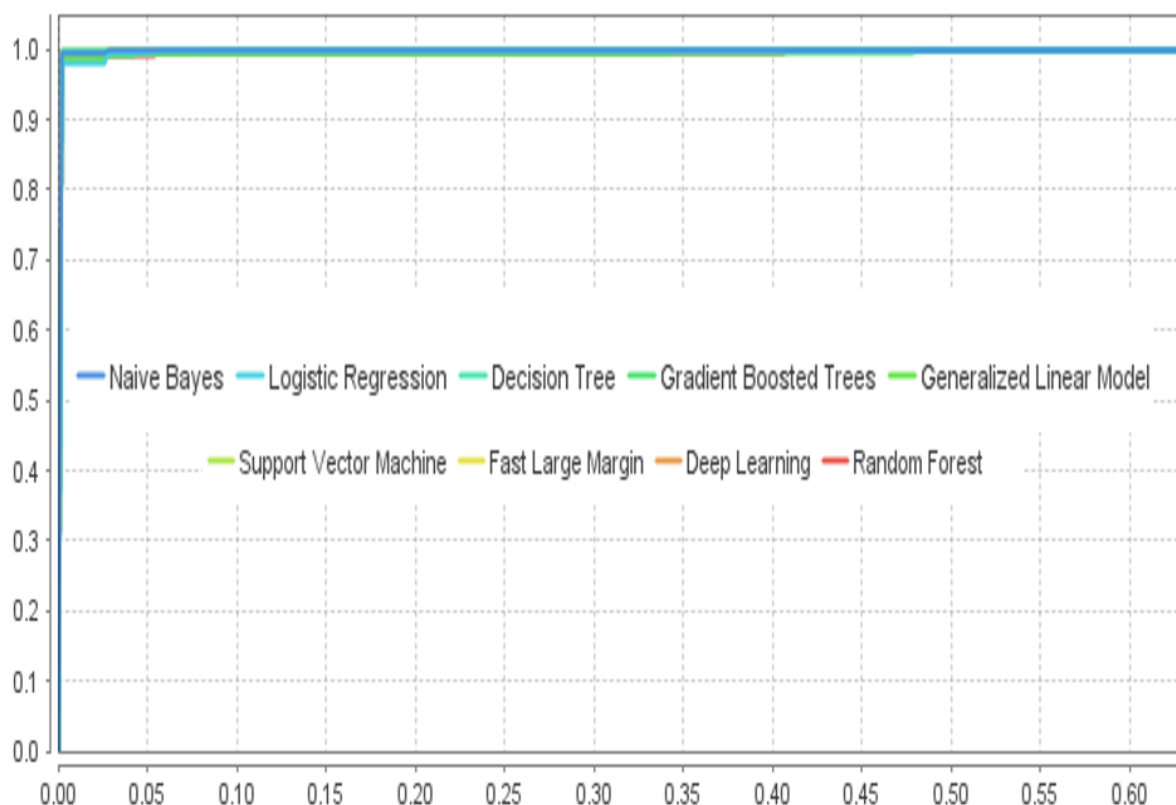


Figura 0-18: Curva ROC relativa aos testes de mineração de dados feita pela plataforma RapidMiner® em amostras com 75% de adição de leite bovino em leite bubalino simulando fraude.

Fonte: A autora (2022)

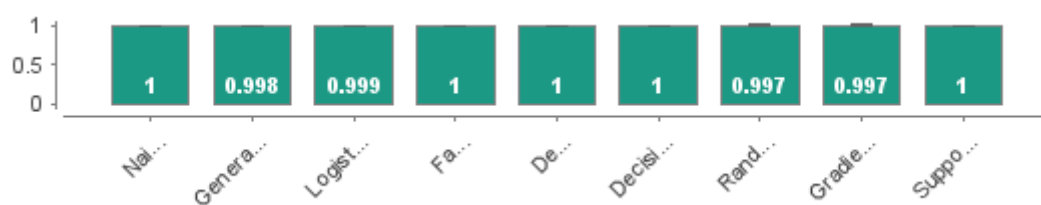


Figura 0-19: Area Under the Curve relativa aos testes de mineração de dados feitos pelo Rapidminer® utilizando dados FTIR relativos à composição do leite em amostras com 75% de adição de leite bovino em leite bubalino simulando fraude. Em ordem: Naive Bayes, Generalized Linear Model, Logistic Regression, Fast Large Margin, Deep Learning, Decision Trees, Random Forest, Gradient Boosted Trees e Support Vector Machine.

Fonte: A autora (2022)

O comparativo entre os diferentes testes realizados com todos os níveis de fraude está evidenciado na Figura 4-20.

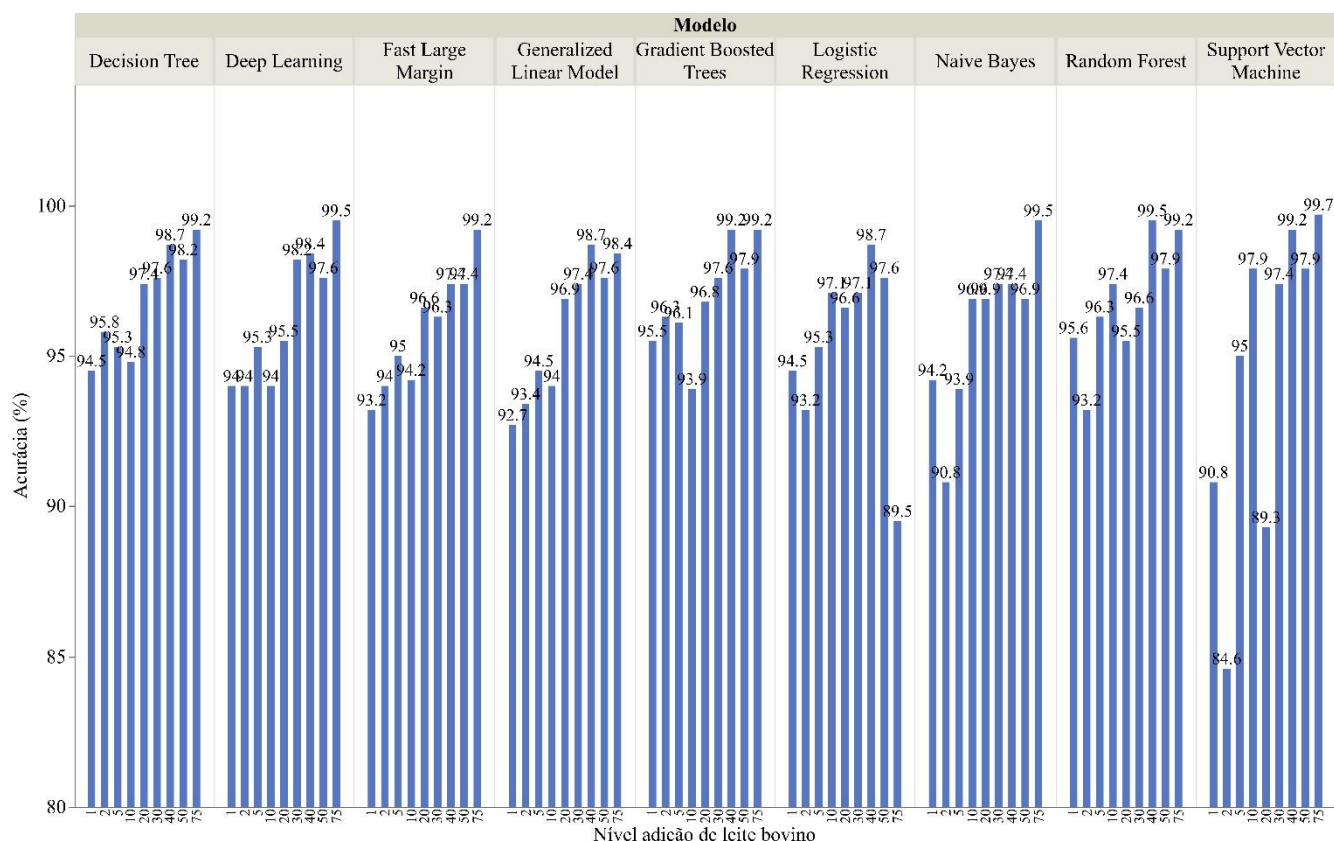


Figura 0-20: Comparativo da acurácia entre os diferentes testes de mineração de dados realizados pelo Rapidminer® em relação aos diferentes níveis de adição de leite bovino ao leite bubalino, utilizando dados obtidos de análise FTIR realizados em amostras obtidas de 2020-2022.

Fonte: A autora (2022)

A partir da observação de todos os testes em um único gráfico como demonstrado na Figura 4-20, é possível observar o aumento gradativo da acurácia na grande maioria dos testes. Todos apresentaram boa acurácia, acima de 80%, com valores muito próximos a 100% à medida que a adição de leite bovino foi aumentando nas amostras. Isso evidencia a boa capacidade dos métodos de detecção de fraude, mesmo em pequenas quantidades de adição.

O RapidMiner tem sido cada vez mais utilizado em estudos de detecção de fraudes. Já foram realizados estudos utilizando essa metodologia para detecção de fraudes tanto nos sistemas financeiros (Lin *et al.*; 2015; Yao *et al.* 2019; Seera *et al.*, 2021), quanto na área da médica com avaliação de licenças médicas indevidas (Brahimi *et al.* 2022) e na medicina veterinária na avaliação de mastite (Ebrahimie *et al.*, 2019)

Lin *et al.* (2015) avaliou fraudes em esquemas de pirâmide usando mineração de dados com o método *Logistic Regression Decision Trees* na detecção de vários fatores de fraude.

Como resultado, a acurácia de predição de fraude foi 90,3%, mostrando a boa capacidade de predição do método.

Outro estudo realizado nessa área de fraude financeira foi realizado por Yao *et al.* (2019) que propôs a identificação utilizando 24 variáveis utilizando os métodos de *Support Vector Machine*, *Logistic Regression* e *Naive Bayes*. Desses modelos, o método *Support Vector Machine* apresentou a maior acurácia para 13 das variáveis avaliadas. No presente estudo, o método que apresentou maior acurácia em mais níveis de fraude foi o *Random Forest*, com melhor resultado em 3 dos 9 níveis. Em seguida com maior acurácia em dois níveis estiveram os métodos de *Support Vector Machine* e *Decision Tree*.

Seera *et al.* (2021) avaliando fraudes em cartões de crédito nos nove métodos avaliados pelo Rapidminer, encontrou bons resultados na acurácia para o método *Naive Bayes* para detecção de fraudes. Em relação à avaliação da área sob a curva, o método que se mostrou mais eficiente foi o *Gradient Boosted Trees*.

Além da área financeira, essa plataforma vem sendo utilizada também na área médica. Brahim *et al.* (2022) avaliando a obtenção de atestados de licença médica falsos através dos mesmos métodos utilizados no presente estudo e encontraram resultados de acurácia para *Naive Bayes*, *Logistic Regression* e *Random Forest* a 99% mostrando a boa capacidade de identificação de fraudes.

Na medicina veterinária, a plataforma já foi utilizada para a análise de predição de mastite sub-clínica com os métodos *Deep Learning*, *Naive Bayes*, *Generalized Liner Model*, *Logistic Regression*, *Decision Tree*, *Gradient-Boosted Tree* e *Random Forest*. O estudo foi realizado utilizando as contagens de células somáticas para a alimentação do programa e os métodos de *Gradient-Boosted Trees* e *Deep Learning* foram os que apresentaram melhores resultados com 84,9% de acurácia (Ebrahimi, *et al.* 2019).

Na área de alimentos, a associação de FTIR com Rapidminer® foi realizada por Freitas *et al.* (2021) na determinação de resíduos de tilosina em leite em pó e obtiveram melhores resultados com o método *Random Forest*, da mesma forma que o presente estudo.

Mu *et al.* (2020) utilizaram um nariz eletrônico e técnicas de aprendizado de máquina na predição do teor de gordura e proteína em leite bovino obtido de fazendas. Os métodos *Support Vector Machine*, *Random Forest* e *Logistic Regression* foram analisados e como

resultado identificaram que o melhor método foi o *Support Vector Machine* com 95% de acurácia.

Lima (2021) avaliou a utilização de técnicas de aprendizado de máquina para a detecção de fraudes por adição de soro de queijo em leite cru. Do mesmo modo que o presente trabalho, os dados foram obtidos através de análises FTIR de amostras de leite adulteradas experimentalmente. Os métodos utilizados na análise de dados foram as *Decision Trees* que obtiveram 96,2% de acurácia. Esse classificador obteve os melhores resultados em duas das nove diluições testadas no presente estudo.

Na área de lácteos, Tavares (2019) utilizou classificadores para a predição de fraudes em leite cru por adição de adulterantes. No estudo foram analisados *Gradient Boosted Trees* e *Random Forest* como modelos, utilizando dados provenientes da análise FTIR de amostras adulteradas experimentalmente. Como resultados, foram encontrados o percentual de classificação de amostras corretas 93,18% e 98,72% para os modelos *Gradient Boosted Trees* e *Random Forest* respectivamente. Apesar dos altos índices encontrados, o modelo que atingiu o maior percentual de acerto foi a rede neural artificial convolucional, com 99,34%.

Outro estudo realizado na área de lácteos utilizando mineração de dados foi realizado por Frizzarin *et al.* (2021) que avaliaram a predição da coagulação do leite e firmeza do coágulo a partir de dados obtidos do FTIR e técnicas de aprendizado de máquina. Os métodos avaliados foram *Random Forests*, *Boosting Decision Trees* e *Support Vector Machines*, sendo este último o método com melhor métrica de acurácia.

A plataforma RapidMiner® possui uma versatilidade enorme de utilização em diversos ramos que podem ser melhor explorados. Sua utilização vem crescendo com o passar dos anos em diversas áreas do conhecimento, mostrando ótimos resultados e a tendência de modernização dos setores. A característica de permitir que vários testes sejam rodados simultaneamente é atrativa em diversos setores. No presente estudo foram avaliados nove modelos de classificação simultaneamente, porém essa quantidade pode ser reduzida com a seleção de modelos mais específicos ao objetivo que se deseja avaliar.

Comparando os modelos de classificação do RapidMiner® com as redes neurais artificiais, os modelos de redes multicamadas *Multilayer Perceptron* e de Função de Base Radial também apresentam resultados com grande acurácia e boa capacidade de predição de adulterações, devendo se levar em consideração a disponibilidade do software e facilidade de análise pela indústria.

4.6 CONCLUSÃO

A plataforma apresenta excelente capacidade de predição de fraudes, com interface intuitiva e fácil observação dos modelos e resultados, sendo uma ótima alternativa de utilização. Porém, é uma ferramenta menos disponível em relação à outras alternativas de análise de dados, como as redes neurais do sistema SPSS, devendo ser considerado a usabilidade em relação ao custo-benefício por parte da indústria.

3. CONCLUSÃO GERAL

De acordo com os estudos realizados, conclui-se que o leite bubalino tem composição diferenciada em relação ao leite bovino, com maiores teores em todos os componentes, o que evidencia a maior qualidade nutricional e de rendimento industrial. A contagem de células somáticas em búfalas possui menor teor em relação às vacas devido a características intrínsecas, resultando em um leite de ótima qualidade.

A inteligência artificial pode ser uma ótima ferramenta a ser utilizada para a detecção de casos de fraude por mistura de leite de diferentes espécies. Pelos estudos realizados foi possível perceber que todas as ferramentas utilizadas para detecção tiveram resultados com boa taxa acerto mesmo em quantidades baixas de adição de leite bovino. As arquiteturas de redes neurais utilizadas foram eficazes na detecção, porém a adição de uma camada oculta nas redes *Multilayer Perceptron* não trouxe ganhos significativos, evidenciando que redes com apenas uma camada são suficientes para um bom resultado. Sendo assim, de acordo com os resultados obtidos, a rede neural de função de base radial apresentou os melhores resultados, seguidos pela *Multilayer Perceptron* de uma camada e *Multilayer Perceptron* de duas camadas respectivamente.

A utilização de classificadores também são formas de se obter um bom resultado na identificação de amostras fraudadas, porém esse método demanda maior tempo de execução em relação à utilização de redes neurais.

Dessa forma, a utilização dos algoritmos pela indústria a fim de se detectar as amostras que forem consideradas “anormais” e indicativas de fraude pode ser uma excelente alternativa aos métodos tradicionais, visto que utiliza dados de análises rotineiras e um tempo menor em relação a análises laboratoriais, para prosseguir com a investigação.

Reforça-se que estes métodos são ferramentas de triagem, que podem identificar possíveis fraudes que devem ser confirmadas por técnicas de análise genética. Porém é uma forma relativamente rápida de se identificar possíveis infratores.

Como todo método, a inteligência artificial apresenta algumas limitações tais como a necessidade de um banco de dados para alimentar o algoritmo

As Redes Neurais *Multilayer Perceptron* e *Radial Basis Function* elaboradas pelo sistema SPSS fornecem um resultado rápido e com boa acurácia quando comparados aos

resultados dos algoritmos analisados pelo *RapidMiner*. Sendo assim, é uma ferramenta mais indicada a ser utilizada por apresentar resultados precisos com um software com maior disponibilidade no mercado e facilidade de utilização, sendo facilmente adaptado a rotina da indústria.

REFERÊNCIAS

- ABRAHAM, A. Artificial neural networks. *Handbook of measuring system design*, John Wiley & Sons, Ltd, 2005.
- AHMAD, S. I.; GAUCHER, I., ROUSSEAU, F. et al. Effects of acidification on physicochemical characteristics of buffalo milk: A comparison with cow's milk. *Food Chemistry Journal*. v. 106 p.11-17. 2008.
- AHMAD, S. F. M.; ANJUM, F. M., HUMA; et al. Composition and physico-chemical characteristics of buffalo milk with particular emphasis on lipids, proteins, minerals, enzymes and vitamins. *The Journal of Animal and Plant Sciences*, 23 (1 Suppl.): p. 62-74, 2013.
- ALHARTHI, H. Healthcare predictive analytics: An overview with a focus on Saudi Arabia, *Journal of Infection and Public Health*, v.11, n. 6, p.749-756 , 2018.
- ALI, M. A.; AHMAD, M.; MUHAMMAD, K. et al. Prevalence of sub clinical mastitis in dairy buffaloes of Punjab, Pakistan. *The Journal of Animal & Plant Sciences*, v. 21, n. 3, p. 477-480, 2011.
- ALISHEROV, F; ISKANDAROV, S. Q., BEKTURDIYEV, S. S *et al.* Development of methods for assessing the performance of teachers using of TUIT-LMS data. *Journal of Physics: Conference Series*. 1889. 2021.
- ALMEIDA, F. Espectroscopia de Infravermelho Próximo com Transformada de Fourier (FT-NIR) na Caracterização de Farinhas para Alimentação Pueril. 2009 *Dissertação* (Mestrado em Engenharia Biológica) - Instituto Superior Técnico, Universidade Técnica de Lisboa, Lisboa, 2009.
- ALVES, R. O. R; TOMÉ, O. C.; PEREIRA, P. C *et al.* Artificial neural networks in the prediction of fraud in integral milk powder by adding whey powder. *Ciência Rural* [online], v. 52, n. 4. 2022
- AGHBASHLO, M., MOBLI, H., RAFIEE, S. The use of artificial neural network to ct exergetic performance of spray drying process: a preliminary study. *Computers and Electronics in Agriculture*, New York, v. 88, p. 32-43, 2012

ANDRADE, K. D., RANGEL, A. H. N.; ARAÚJO, V. et al. Efeito da estação do ano na qualidade do leite de búfalas. *Revista Verde de Agroecologia e Desenvolvimento Sustentável*, v. 6, n. 3, p. 11, 2011.

ANDRIGHETTO, C. Cadeia produtiva do leite de búfala. In: Simpósio da Cadeia Produtiva da Bubalinocultura, 2011, Botucatu. *Anais do Simpósio da Cadeia Produtiva da Bubalinocultura*. Botucatu, 2011.

AMARAL, F. R. Fatores que interferem na contagem de células somáticas e constituintes do leite de búfalas. 2005. 46f. *Dissertação*. (Mestrado em Medicina Veterinária Preventiva) – Escola de Veterinária, Universidade Federal de Minas Gerais, 2005.

AMARAL, F. R.; CARVALHO, L.B.; SILVA, N.; et al. Composição e contagem de células somáticas em leite bubalino na região do Alto São Francisco, Minas Gerais, Brasil. *Revista do Instituto Cândido Tostes*, n. 59, p. 37-41, 200

ARAÚJO, T. P. M., RANGEL, A. H. N.; SOARES, A. D. et al. Influência das estações do ano sobre a composição do leite de búfalas mantido em tanque de resfriamento. *Agropecuária Científica do Semi-Árido*. v.7 n. 1 2011.

ASENSIO, L., I. GONZÁLEZ, T. GARCÍA, *et al.* Determination of food authenticity by enzyme-linked immunosorbent assay (ELISA). *Food Control* v.19 p.1-8. 2008.

AYDIN, S. A short history, principles, and types of ELISA, and our laboratory experience with peptide/protein analyses using ELISA. *Peptides*, v.72 p.4-15. 2015.

AZEVEDO, B. T. Desenvolvimento de Técnicas de Genética Molecular para a Identificação de Fraude de Leite de Vaca em Leite e Produtos lácteos de Búfala. *Dissertação*. Mestrado em Produção Animal Sustentável. Instituto de Zootecnia, APTA/SAA. 2021.

BAILONE, R. L.; BORRA, R.C.; ROÇA, R.O.; et al. Quality of refrigerated raw milk from buffalo cows (*Bubalus bubalis bubalis*) in different farms and seasons in Brazil. *Ciência Animal Brasileira*, v.18, p. 1-12, 2017.

BALESTRIERI, F.; DAMIANI, F.; MARINI, D. *et al.* Artificial neural networks to classify some dairy products. *Journal of Commodity Science*, v. 40, n.1, p. 17-31, 2001.

BARBOSA, L. C. de A. Espectroscopia no infravermelho na caracterização de compostos orgânicos. 1. ed. Viçosa: Editora UFV, 2007.189p.

BARILE, D.; COÏSSON, J. D.; ARLORIO, M.; *et al.* Identification of production area of Ossolano Italian cheese with chemometric complex approach. *Food Control*, v. 17, p. 197–206, 2006.

BARUSELLI, P. S; CARVALHO, N. A. T. Reproductive management and artificial insemination in buffalo. In: Buffalo Symposium of Americas, 1, 2002, Belém. Proceedings ... Belém: Associação Paraense de Criadores de Búfalos, p.119-143. 2002.

BASTIANETTO, E.; ESCRIVÃO, S. C.; OLIVEIRA, D. A. A. Influência das características reprodutivas da búfala na produção, composição e qualidade do leite. *Revista Brasileira Reprodução Animal*. v.29, n.1, p.49-52. 2005

BAUGHMAN, D. R.; LIU, Y. A. Neural Networks in Bioprocessing and Chemical Engineering. 488p. San Diego, CA: Academic Press.1995.

BERNARDES, O. Bubalinocultura no Brasil: situação e importância econômica. *Revista Brasileira Reprodução Animal*. v.31, n.3, p.293-298, 2007.

BEZERRA JÚNIOR, J. D. S., FRAGA, A. B., COUTO, A. D. G. Produção de leite, duração da lactação e intervalo de partos em búfalas mestiças Murrah. *Revista Caatinga*, v. 27, n. 2, p. 184-191, 2014.

BJAOUI, M., SAKLY, H., SAID, M., *et al.* Depth insight for data scientist with RapidMiner an innovative tool for AI and big data towards medical applications. *Proceedings of the 2nd International Conference on Digital Tools & Uses Congress*. 2020.

BOTELHO, B. G.; REIS, N.; OLIVEIRA, L. S.; *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.

BRAGA, A. de P.; CARVALHO, A. C. P. de L. F.; LUDEMIR, T. B. Redes neurais artificiais: teoria e aplicações. Rio de Janeiro: LTC, 2000. 262 p.

BRAHIMI, S.; EL HUSSEIN, M.; AL-REEDY, A. Detection of Undeserved Sick Leaves in Hospitals using Machine Learning Techniques. *Sustainable Computing: Informatics and Systems*, v. 35, p. 100665, 2022.

BRASIL. Código de Defesa do Consumidor. Decreto Presidencial nº 2.181, de 20 de março de 1997, Brasília, DF, 1997

BRASIL. Ministério da Agricultura, Pecuária e Abastecimento. RIISPOA - Regulamento da Inspeção Industrial e Sanitária de Produtos de Origem Animal (RIISPOA). IN: BRASIL. Ministério da Agricultura, Pecuária e Abastecimento. Decreto nº 9.013, de 29 de março de 2017. Diário Oficial da União, 30 de mar. 2017.

BRASIL. Ministério da Agricultura, Pecuária e Abastecimento. Regulamento técnico de Identidade e Qualidade de Leite. IN: BRASIL. Ministério da Agricultura, Pecuária e Abastecimento. Instrução Normativa nº 76, de 29 de dezembro de 2011. Diário Oficial da União, 26 nov. 2018. Seção 1. 2018.

BROOKE-TAYLOR, S; DWYER, K; WOODFORD, K; KOST, N. Systematic Review of the Gastrointestinal Effects of A1 Compared with A2 β -Casein. *Advances in Nutrition*. v. 8 p. 739-748. 2017.

BUZI, K. A.; PINTO, J. P. A. N.; RAMOS, P. R. R.; Análise microbiológica eletroforética do queijo mussarela elaborado a partir de leite de búfala. *Ciência e Tecnologia de Alimentos*, v. 29, n. 1, p.7-11, 2009.

CAMILO, C. O; SILVA, J. C. Mineração de Dados: Conceitos, Tarefas, Métodos e Ferramentas. RT-INF_001-09 - *Relatório Técnico Instituto de Informática Universidade Federal de Goiás*. 29p. 2009

CAMPANILE, G., BERNARDES, O., BASTIANETTO, E., et al. *Manejo de búfalas leiteiras*. 1 ed. São Paulo: ABCB, 80p. 2007.

CARVALHO, B. P. Estudo de misturas de café arábica e robusta usando FTIR e Redes Neurais Artificiais. 2014. 33f. *TCC (Trabalho de Conclusão de Curso) – Tecnologia em Alimentos*, Universidade Tecnológica Federal do Paraná – UTFPR Campus Campo Mourão, 2014.

- CARVALHO, L. O. D. de M.; LOURENÇO JÚNIOR, J. de B. Produção leiteira de bubalinos como opção para a Amazônia. Embrapa Amazônia Oriental-*Artigo Em Anais de Congresso (ALICE)*. 2001.
- CASSOLI, L. D.; SARTORI, B.; ZAMPAR, A.; *et al.* An assessment of Fourier transform infrared spectroscopy to identify adulterated raw milk in Brazil. *International Journal of Dairy Technology*, v. 64, p. 480-485, 2011.
- CAVALLI, J.; PEREIRA, R.G.A. Produção leiteira de búfalos. In: SALMAN, A. K. D.; PFEIFER, L. F. M. (Ed.). *Pecuária leiteira na Amazônia*. Brasília, DF: Embrapa, 2020.
- CERÓN-MUNÓZ, M., TONHATI, H., DUARTE, J., *et al.* Factors affecting somatic cell counts and their relations with milk and milk constituent yield in buffaloes. *Journal of Dairy Science*, v. 85, p. 2885–2889, 2002
- CHAWLA, N.V; BOWYER, K. W.; HALL, L. O. *et al.* “SMOTE: Synthetic Minority Over-Sampling Technique.” *Artificial Intelligence Research*, 2002: 321- 357.
- CHIANESE, L.; P. LAEZZA, L. A.; SMALDONE, C. *et al.* Evaluation of bovine milk in the buffalo mozzarella cheese by two-dimensional electrophoresis. *Scienza e Tecnica Lattiero-Casearia*. v.41 p.315–326. 1990.
- CHOU, C.; LIN, S.; LEE, K. *et al.* Fast Differentiation of Meats from Fifteen Animal Species by Liquid Chromatography with Electrochemical Detection Using Copper Nanoparticle Plated Electrodes. *Journal of chromatography*. B, Analytical technologies in the biomedical and life sciences. v.846. p.230-239. 2007
- CONCEIÇÃO, D. G. Utilização do FTIR aliado à análise quimiométrica como ferramenta de triagem para identificação de adulterantes no leite cru. Itapetinga – BA: UESB, 2018. 73 p. *Dissertação*. (Mestrado em Engenharia e Ciência de Alimentos, Área de Concentração em Ciência de Alimentos).
- COROIAN, A.; TRIF, M; CORIAN, C.O.; *et al.* Qualitative evaluation of buffalo cheese using FTIR spectroscopy. *International Journal of the Bioflux Society*, v. 4, n.2, p.66-70, 2012.
- COSTA FILHO, M. H. B.; de LIMA JÚNIOR, D. M., do NASCIMENTO RANGEL; *et al.* Sazonalidade e variação na qualidade do leite de búfalas no Rio Grande do Norte. *Acta Veterinária Brasilica*, v.8, n.3, p. 201-208, 2014.

COTTENET, G. BLANCPAIN, C. GOLAY, P. A. V. Simultaneous detection of cow and buffalo species in milk from China, India, and Pakistan using multiplex real-time PCR. *Journal of Dairy Science*. v. 94 p.3787–3793. 2011.

CZERWENKA, C.; MÜLLER, L.; LINDNER, W. Detection of the adulteration of water buffalo milk and mozzarella with cow's milk by liquid chromatography-mass spectrometry analysis of β -lactoglobulin variants. *Food Chemistry*, v. 122, p. 901-908, 2010.

DANTAS, et al. Diagnóstico do perfil do consumidor de derivados de leite de búfala na região metropolitana de Belém-PA.V Congresso Internacional das Ciências Agrárias. DOI: 10.31692/2526-7701.VCOINTERPDVAgro.0092. 2020.

DAMÉ, M. C. F. *Búfalo: animal de tração*. Pelotas: Empresa Brasileira de Pesquisa Agropecuária, 2006. 24p

DARWISH, S. F.; ALLAM, H.A.; AMIN, A.S. Evaluation of PCR Assay for Detection of Cow's Milk in Water Buffalo's Milk. *World Applied Science Journal*; v.7 n.4 p.461-467. 2009

DE AMO, S. Técnicas de mineração de dados. *Jornada de Atualização em Informática*, p. 26, 2004.

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

DE MELO, B. A., DA SILVA, S. G. M., DOS SANTOS, et al. Abordagens da mastite em búfalas (*Bubalus bubalis*): uma revisão. In: RODRIGUES N. J. L. *Ciência Animal e Veterinária: Inovações E Tendências*. v.2 202 p. 2022.

DI DOMENICO, M., DI GIUSEPPE, M., RODRÍGUEZ, J.W. *et al.* Validation of a fast real-time PCR method to detect fraud and mislabeling in milk and dairy products. *Journal of Dairy Science*. v. 100, p.106-112, 2017.

DIAS, S. S. Avaliação da autenticidade de derivados de leite de búfala encontrados no varejo do Rio de Janeiro. 2009. 58p. *Dissertação* (Mestrado em Ciência e Tecnologia de Alimentos, Departamento de Tecnologia). Instituto de Tecnologia, Departamento de Tecnologia de Alimentos, Universidade Federal Rural do Rio de Janeiro, Seropédica, RJ, 2009.

- DIAS, S.S, NOGUEIRA, L.C; REIS, R.C.S; *et al.* Avaliação da Disponibilidade e Rotulagem de Derivados de Leite de Búfala nas Diferentes Estações do ano comercializados na Zona Oeste do Rio de Janeiro. *Alimentos e Nutrição*. v.23 p.421-426. 2012
- DRUMMOND, M.G., BRASIL, B.S.A., DALSECCO, L.S. *et al.* A versatile real-time PCR method to quantify bovine contamination in buffalo products. *Food Control*. v.29 n.1 p.131-137. 2013
- DUARTE, J. M. C. TONHATI, H; CERÓN-MUÑOZ, M.F. Efeitos ambientais sobre a produção no dia do controle e características físico-químicas do leite em um rebanho bubalino no estado de São Paulo. *Revista Instituto Laticínio Cândido Tostes*, v. 56 n.5 p.16-19. 2001
- DUBEY, P. C.; SUMAN, C. L.; SANYAL, M. K. *et al.* Factors affecting composition of milk of buffaloes. *Indian Journal of Animal Sciences*, v. 67, n. 9, p. 802-804, 1997
- DÜRR, J.W. Programa nacional de melhoria da qualidade do leite: uma oportunidade única. In: DÜRR, J.W.; CARVALHO, M.P.; SANTOS, M.V. (Eds.) O compromisso com a qualidade do leite no Brasil. Passo Fundo: Editora Universidade de Passo Fundo, 2004. p.38- 55.
- EBRAHIMI, M; MOHAMMADI-DEHCHESHMEH, M; EBRAHIMIE, E; Comprehensive analysis of machine learning models for prediction of sub-clinical mastitis: Deep Learning and Gradient-Boosted Trees outperform other models, *Computers in Biology and Medicine*, v. 114, 2019.
- ETZION, Y.; LINKER, R.; COGAN, U.; *et al.* Determination of protein concentration in raw milk by mid-infrared Fourier transform infrared/attenuated total reflectance spectroscopy. *Journal of Dairy Science*, v. 87, n. 9, p. 2779–2788, 2004.
- FAWCETT, T. An introduction to ROC analysis. *Pattern Recognition Letters*, v. 27, n. 8, p. 861 – 874, 2006.
- FILHO, M. H. B. C; JUNIOR, D. M. L; RANGEL *et al.* Sazonalidade e variação na qualidade do leite de búfalas no rio grande do norte. *Acta Veterinaria Brasilica*, v.8, n.3, p.201-208, 2014.
- FIGUEIREDO, E. L.; JUNIOR, J. D. B. L.; TORO, M. J. U.; Caracterização físico-química e microbiológica do leite de búfala “in natura” produzido no estado do Pará. *Revista Brasileira de Tecnologia Agroindustrial*, v. 4, n. 1, 2010.

- FISTER Jr. I. YANG, X. S., FISTER, I. *et al.* A brief review of nature-inspired algorithms for optimization. *Elektrotehniski Vestnik*, 2013. 1-7. Disponível em: <https://arxiv.org/pdf/1307.4186.pdf>. Acesso em: 19 de dezembro de 2021
- FOX, P. F. et al. *Fundamentals of cheese Science*. New York: Aspen, 2000. 587p.
- FREITAS, A.G.M; MINHO, L.A.C; MAGALHÃES; B. E. A, *et al.*, Infrared spectroscopy combined with random forest to determine tylosin residues in powdered milk, *Food Chemistry*, v. 365, 2021.
- FRIZZARIN, M., GORMLEY, I. C., BERRY, D. P.; *et al.* Predicting cow milk quality traits from routinely available milk spectra using statistical machine learning methods. *Journal of Dairy Science*, v.104 n.7, p. 7438-7447. 2021.
- FUENTE, M. A.; JUAREZ, M. Authenticity Assessment of Dairy Products. *Critical Reviews in Food Science and Nutrition*, v.45, p.563–585, 2005.
- GANGULI N. C. Tecnología de la leche de búfala. *Revista Mundial de Zootecnia*, v.30, p.2-10, 1979.
- GOMES, E. A. S. Aplicabilidade de Algoritmos de Aprendizado de Máquina para Detecção de Intrusão e Análise de Anomalias de Rede. *Dissertação (especialização)* — Universidade Federal de Minas Gerais. 48p. 2019.
- GREGORY, L., ROSSI, R. S., MENDES, J. P. G et al. Ocorrência dos principais agentes bacterianos e parasitários em fezes diarreicas de bezerros búfalos nos estados de São Paulo e Paraná. *Arquivos do Instituto Biológico*, v. 81, p. 180-185, 2014.
- GUTIÉRREZ-TOLENTINO, R., VEGA, S., SCHETTINO-BERMÚDEZ, B., et al. Ácidos grasos y ácido linoleico conjugado en leche orgánica producida en el sureste de México. *Revista MVZ Córdoba*, v. 27, n. 2, p. e2541-e2541, 2022
- HAN, J; KAMBER, M. Data mining: concepts and techniques. Morgan Kaufmann Publishers Inc., San Francisco, CA, USA. 2000
- HAYKIN, S. Redes neurais: princípios e prática, 2001. Bookman Editora, 2001
- HE, H.; GARCIA E.A. “Leraning from Imbalanced Data.” IEEE, 2009.

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

HERMAN, L. Determination of the animal origin of raw food by species-specific PCR. *Journal of Dairy Research*. v.68, p.429–436. 2001

HOSMER, D.W; LEMESHOW, S. Applied Logistic Regression, 2nd Ed. Chapter 5. New York, NY: John Wiley and Sons, Pp. 160 –164, 2000.

HÜNH, S.; FERREIRO, L. Influência da mastite bovina na percentagem da caseína do leite. *Revista do Instituto de Laticínios Cândido Tostes*, Juiz de Fora, v. 35, n. 212, p. 33-35, 1980.

HURLEY, I. P.; COLEMAN, R. C.; IRELAND, H. E. *et al.* Measurement of Bovine IgG by Indirect Competitive ELISA as a Means of Detecting Milk Adulteration. *Journal of Dairy Science*. v.87 p.543–549, 2004

IBGE. Instituto Brasileiro de Geografia e Estatística. Diretoria de Pesquisas, Coordenação de Agropecuária, *Pesquisa da Pecuária Municipal 2020*. 2021

IDF. International IDF Standard 148-2:2006. Milk — Enumeration of somatic cells — Part 2: Guidance on the operation of fluoro-opto-electronic counters. 2006

IDF. International IDF Standard 141:2013. ISO 9622:2013. Milk and liquid milk products — Guidelines for the application of mid-infrared spectrometry. 2013

JAISSWAL, P.; JHA, S. N.; BORAH, A.; *et al.* Detection and quantification of soymilk in cow-buffalo milk using Attenuated Total Reflectance Fourier Transform Infrared spectroscopy (ATR-FTIR). *Food Chemistry*, v. 168, n.1, p. 41-47, 2015.

JAIN, A. MAO, J.; MOHIUDDIN, K. M; *et al.* Artificial neural networks: A tutorial. *Computer*. 1996. 31-44. DOI: <https://doi.org/10.1109/2.485891>

JEYARAJ, A; RAMYA, S; RAJA, M. A study of classification algorithms using Rapidminer. *International Journal of Pure and Applied Mathematics*. p.119. 2018

JORGE, A. M. *et al.* Correlação entre o California Mastitis Test (CMT) e a Contagem de Células Somáticas (CCS) do Leite de Búfalas Murrah. *Revista Brasileira de Zootecnia*, v. 34, n. 6, p. 2039-2045, 2005.

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

KAMILOV, M. M., EGAMBERDIYEV, N. A., CHORSHANBIYEV, A. X. et al. The method of applying the algorithm of calculating grades for finding similar diagnostics in medical information systems. *International Journal of Innovative Technology and Exploring Engineering*, v. 8, n. 6, p. 722-724, 2019.

KHANLARI, G. R. *et al.* Prediction of shear strength parameters of soils using artificial neural networks and multivariate regression methods. *Engineering Geology*, v. 131/132, p. 11-18, 2012.

KHANMOHAMMADI, M., GARMARUDI, A.B., GHASEMI, K, *et al.* Artificial neural network for quantitative determination of total protein in yogurt by infrared spectrometry. *Microchemical Journal*, v.91 n.1, p. 47–52. 2009

KUBISTA, M., ANDRADE, J. M., BENGTSSON, M., *et al* The real-time polymerase chain reaction. *Molecular aspects of medicine*. v. 27, n. 2-3, p. 95-125, 2006.

KUHN, M.; JOHNSON, K. *Applied Predictive Modeling*. 2^aed. New York: Springer. 2018.

LIMA, J. S. Espectrofotometria FTIR (Fourier Transform Infrared) e técnicas de aprendizado de máquina para a detecção de fraude por adição de soro de queijo ao leite cru. Tese. Universidade Federal de Minas Gerais. 2021.

LIN, C.C.; CHIU, A.A.; HUANG, S.Y.; *et al.* Detecting the financial statement fraud: The analysis of the differences between data mining techniques and experts' judgments. *Knowledge-Based Systems*, v. 89, p. 459-470, 2015.

LIN, M. A review of traditional and novel detection techniques for melamine and its analogues in foods and animal feed. *Frontiers of Chemical Engineering in China*, v. 3, p. 427-435, 2009.

LOHUMI, S. H.; LEE, S.; LEE, H.; CHO, B-K. A review of vibrational spectroscopic techniques for the detection of food authenticity and adulteration. *Trends in Food Science & Technology*, v. 46, n.1, p. 85-98, 2015.

- LOPES, F. A. Caracterização da produtividade e da qualidade do leite de búfalas na Zona da Mata Sul de Pernambuco. Recife/PE. Universidade Federal Rural de Pernambuco, 48f. *Tese (Doutorado em Zootecnia)*, 2009.
- LÓPEZ-CALLEJA DÍAZ, I. ALONSO I. G; FAJARDO, V. *et al.* Application of a polymerase chain reaction to detect adulteration of ovine cheeses with caprine milk. *Journal of Dairy Science*. v.225 p.345-349. 2007
- LUDERMIR, T. B. Inteligência Artificial e Aprendizado de Máquina: estado atual e tendências. *Estudos Avançados [online]*. v. 35, n. 101 , p. 85-94. 2021.
- MACEDO, M. P., WECHSLER, F. S., RAMOS, A. D. A. *et al.* Composição Físico-Química e Produção do Leite de Búfalas da Raça Mediterrâneo no Oeste do Estado de São Paulo, São Paulo, SP, *Revista Brasileira de Zootecnia*, v.30, n.3, p.1084-1088, 2001.
- MAFRA, I; ROXO, A; FERREIRA, I. M. P. L. V. O, *et al.* A duplex polymerase chain reaction for the quantitative detection of cows' milk in goats' milk cheese. *International Dairy Journal*. v.17 p.1132-1138. 2007
- MAIMON, O.; ROKACH, L. *Data Mining and Knowledge Discovery Handbook*. Springer: Boston, MA. 2^a ed. 2010.
- MARINI, F. Artificial neural networks in foodstuff analyses: Trends and perspectives A review. *Analytica Chimica Acta*, p. 121–131, 2009.
- MARQUES, J. R. F.; MARTINEZ, A. M.; COSTA, M. R.; *et al.* Genetic diversity of Brazilian buffaloes (*Bubalus bubalis*) using DNA microsatellites. *Archivos de Zootecnia*, v. 60, n. 232, p. 1213-1221, 2011.
- MAYER, H. K. Milk species identification in cheese varieties using electrophoretic, chromatographic and PCR techniques. *International Dairy Journal*. v.15 p.595–604. 2005.
- MAYER, H. K.; D. HEIDLER; C. ROCKENBAUER; *et al.* Determination of the percentages of cows', ewes', and goats' milk in cheese by isoelectric focusing and cation-exchange HPLC of γ - and para- κ caseins. *International Dairy Journal*. v.7 p.619–628. 1997
- MELO, B.A., SILVA. S.G.M., SANTOS, M.T., *et al.* Perfil da mastite subclínica e frequência de micro-organismos isolados de búfalas mestiças (*Bubalus bubalis*). *Research, Society and Development*, Maceió, v. 11, p. e24911427327, 2022.

MENDES, C. G.; SAKAMOTO, S. M. SILVA, J. B. A. *et al.* Análises físico-químicas e pesquisa de fraude no leite informal comercializado no município de Mossoró, RN. *Ciência Animal Brasileira*, v. 11, n. 2, p. 349 - 356, 2010.

MITCHELL, T. *Machine Learning*. S. l.: McGraw Hill, 1997.

MOHANRAJ, M.; JAYARAJ, S.; MURALEEDHARAN, C. Applications of artificial neural networks for refrigeration, air-conditioning and heat pump systems: a review. *Renewable and Sustainable Energy Reviews, Berlin*, v. 16, n. 2, p. 1340-1358. 2012.

MOIO, L., M. L. SASSO, L. CHIANESE, *et al.*. Rapid detection of bovine milk in ovine, caprine and water buffalo milk or cheese by gel isoelectric focusing on Phastsystem. *Italian Journal of Food Science*. v.3 p.185–190. 1990.

MOURA, E. D. O., RANGEL, A.H.D.N., MELO, D.M.C.N., *et al.* Evaluation of microbiological, cellular and risk factors associated with subclinical mastitis in female buffaloes. *Asian-Australasian Journal of Animal Sciences*, v. 30, n. 9, p. 1340-1349, 2017.

MÜHLBACH, P. R. F. Nutrição da vaca em lactação e a qualidade do leite. *Simpósio de Bovinocultura de Leite*, v. 1, p. 25-43, 2003.

MONTEIRO, F. K. De C. Utilização da espectroscopia de infravermelho próximo para quantificação de adição de soro em leite. 86f. *Dissertação* (Mestrado em Ciência e Tecnologia do Leite e Derivados) - Curso de pós-graduação em Ciência e Tecnologia do Leite e Derivados, Universidade Federal de Juíz de Fora. 2015

MU, F., GU, Y., ZHANG, J., *et al.* Milk Source Identification and Milk Quality Estimation Using an Electronic Nose and Machine Learning Techniques. *Sensors*, v.20 n.15, p.4238. 2020

MUGGLI, J. Influence of somatic cell counts on stage of lactations. *Animal Breed Abstract*. 1996.

NADER FILHO, A.; SCHOCKEN-ITURRINO, R. P., ROSSI JUNIOR, O. D. *et al.* Estudo da variação do ponto crioscópico do leite de búfala. *Revista do Instituto de Laticínios Cândido Tostes*, Juiz de Fora, v. 39, n. 234, 1984.

- NAZÁRIO, S. L. S.; BUIOCHI, F.; KITANO, C.; *et al.* Caracterização de leite bovino utilizando ultra-som e redes neurais artificiais. *Revista Controle & Automação*, v. 20, n. 4, p. 627-236, 2009.
- NETO, J. D. B.; BASTIANETTO, E. Diferenças fisiológicas entre bubalinos e bovinos: interferência na produção. *Ciência Animal Brasileira*. 2009.
- NEVES, E. C. A Recent progress concerning buffalo milk technology in Amazon-Brazil. In: *Buffalo Symposium of Américas*, 1. p. 312-316. 2002
- NICOLAOU, 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
- NIEDERMAYER, D. An Introduction to Bayesian Networks and Their Contemporary Applications. *Innovations in Bayesian Networks* Springer, Berlin, Heidelberg, 2008. p. 117-130. 2008.
- 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.
- NÖRNBERG, M. L., PINHEIRO, P. N., DO NASCIMENTO, T. C., *et al.* Compostos bioativos em manteigas: carotenoides e ácidos graxos Bioactive compounds in butters: carotenoids and fatty acids. *Brazilian Journal of Development*, v. 8, n. 2, p. 10270-10288, 2022.
- OLIVEIRA, G.B.; GATTI, M.D.S.; VALADÃO, R.C. *et al.* Detecção da adição fraudulenta de soro de queijo em leite: interferência da atividade de proteases bacterianas. *Revista Instituto Laticínios "Candido Tostes"*, v.64, p.56-65, 2009.
- OLIVEIRA, H.T.V., REIS, R.B., GLÓRIA, J.R., QUIRINO, C.R., & PEREIRA, J.C.C. Curvas de lactação de vacas F1 Holandês-Gir ajustadas pela função gama incompleta. *Arquivo Brasileiro de Medicina Veterinária e Zootecnia*, v.59 n.1, p.233-238. 2007
- OLIVEIRA, L. F. C. Espectroscopia molecular. *Cadernos temáticos de química nova na escola*, v. 4, p. 24-30, 2001.
- PALLANT, J. *Spss Survival Manual: A Step by Step Guide to Data Analysis Using Spss*. Maidenhead: Open University Press/McGraw-Hill, 2010.

PAPPAS, C. S.; TARANTILIS, P. A.; MOSCHOPOULOU, E.; *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

PATEL, R., KUNJADIA, P., KORINGA, P., JOSHI, C., KUNJADIYA, A. Microbiological profiles in clinical and subclinical cases of mastitis in milking Jafarabadi buffalo. *Research in Veterinary Science*, v. 125, p. 94-99, 2019.

PATWARDHAN, N. P.; TORO, V. A.; MAJGAONKAR, S. V. Seasonal variation in chemical composition of milk under heavy rainfall region of Konkan. *Indian Journal of Dairy Science*, 1986.

PEREIRA, D.B.C, SILVA, P.H.F, CARVALHO, A.F, *et al.* Proteínas. In: Cruz, A. G, Zacarchenco PB, PIKE, L. J. The challenge of lipid rafts. *Journal of Lipid Research* 2009, 50, S323.

PEREIRA, R. G. A. Produção de leite e curva de lactação de búfalas mestiças sob dois sistemas de produção em Rondônia. 67f. *Tese (Doutorado em Zootecnia)*, Universidade Federal Rural de Pernambuco, Pernambuco, 2007

PEREIRA, R. G. de A.; TOWNSEND, C. R.; MAGALHÃES, J. A.; *et al.* Estacionalidade de partos de fêmeas búfalas em Rondônia. In: *Embrapa Amapá-Artigo em anais de congresso (ALICE)*. In: SEMINÁRIO DE PESQUISA E EXTENSÃO RURAL, 1., 2007, Porto Velho. Anais... Porto Velho: Fundação Universidade Federal de Rondônia, 2007.

PICCININI, R., MIARELLI, M., FERRI, B., *et al.* Relationship between cellular and whey components in buffalo milk. *Journal of Dairy Research*, v. 73, p. 129-133. 2006.

PIGNATA, M. C; FERNANDES, S. A. D. A., FERRÃO, S. P. B., *et al.* Estudo comparativo da composição química, ácidos graxos e colesterol de leites de búfala e vaca. *Revista Caatinga*. v. 27, n. 4, p. 226-233, 2014.

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

PROVOST, G.M. WEISS F. “The Effect of Class Distribution on Classifier Learning: An Empirical Study.” *Rutgers University*, 2001.

RADOSTITS, O.M., GAY, C.C., HINCHCLIFF, W.K., CNSTABELA, P.D. *Veterinary medicine. A textboock og the dieses of cattle, sheep, goats, pigs and horses*, 10th ed. New York: Elsevier. 673-765 p., 2006.

RAI, P.; MAJUMDAR, G. C.; DASGUPTA, S. Prediction of the viscosity of clarified fruit juice using artificial neural network : a combined effect of concentration and temperature. *Journal of Food Engineering*, v. 68, n. 4, p. 527-533, June 2005.

RAPIDMINER. Identify and Prevent Fraud. Disponivel em: www.rapidminer.com/customer-stories/identify-prevent-fraud/. Acesso em: 15 de maio, 2022.

RECIO, I., M. R; GARCÍA-RISCO, L.; AMIGO, E; Detection of milk mixtures in Halloumi cheese. *Journal of Dairy Science*. v.87 n.1595–1600. 2004

RIBEIRO, D. C. S. Z. Espectroscopia FTIR, HPLC e Redes Neurais Artificiais para determinação analítica da lactose residual e outros açucares no leite. 2021. 125p. *Tese*. (Doutorado em Ciência Animal) – Universidade Federal de Minas Gerais. Belo Horizonte. 2021.

RODRIGUES, N. P. A.; GIVISIEZ, P. E. N.; QUEIROGA, R. C. R. E.; *et al.* Milk adulteration: Detection of bovine milk in bulk goat milk produced by smallholders in northeastern Brazil by a duplex PCR assay. *Journal of Dairy Science*, v. 95, n. 5, p. 2749-52, 2012

ROY, A. *et al.* Deep learning detecting fraud in credit card transactions. In: *Systems and information engineering design symposium (SIEDS)*. p. 129–134. 2018.

SÁ, J. A. S., ALMEIDA, A. C., ROCHA, B. R. P., *et al.* Lightning forecast using data mining techniques on hourly evolution of the convective available potential energy. *Brazilian Congress on Computational Intelligence (CBIC'2011)*, 2011.

SABINO, L. V. Quantificação do percentual de café arábica em misturas com café robusta usando FTIR, NIRS e modelos de dois estágios. 2015. 43 f. *Trabalho de conclusão de curso* – Departamento Acadêmico de Engenharia de Alimentos, Universidade Tecnológica Federal do Paraná. Campo Mourão, 2015.

SAGE, A. P. (Ed.). *Concise Encyclopedia of Information Processing in Systems & Organizations*. New York, NY: Pergamon Press. 1990.

SALA, O. *Fundamentos da espectroscopia Raman e no infravermelho*. 2.ed.- São Paulo: Editora UNESP, 2008.

SALIBA, E. O. S.; GONTIJO NETO, M. M; RODRIGUES, N. M. *et al.* Predição da Composição Química do Sorgo pela Técnica de Espectroscopia de Refletância no Infravermelho Próximo. *Arquivo Brasileiro de Medicina Veterinária e Zootecnia*, v.55, n.3, 2003

SANTINI, G. A.; BERNARDES, O.; SCARPELLI, J. U. Análise das relações comerciais do segmento de processamento de leite e derivados de leite de búfala no estado de São Paulo. *Revista Informações Econômicas*, v. 43, n. 5, 2013.

SANTOS, E. R. Composição físico-química e Contagem de Células Somáticas (CCS) no leite de búfala produzido no município de Parintins, Baixo Amazonas. *Dissertação*. (Mestrado em Ciência Animal e Recursos Pesqueiros) Universidade Federal do Amazonas – UFAM. 52 f 2022.

SANTOS, M.V.; FONSECA L.F.L. Estratégias para controle de mastite e melhoria da qualidade do leite. São Paulo: Manole, 314p. 2007

SANTOS, M. V. Entendendo a média geométrica da contagem de células somáticas. 2002 Disponível em: <https://www.milkpoint.com.br/colunas/marco-veiga-dos-santos/entendendo-a-media-geometrica-da-contagem-de-celulas-somaticas-16220n.aspx#%20>. Acesso em: 25/09/2022

SAPUTRA, R.A.; WIDODO, P.P. Comparação do algoritmo de classificação de mineração de dados para prever tuberculose (Tb): Estudo de caso do centro de saúde pública de Karawang. *Seminário Nacional de Inovação e Tendências (SNIT)*. 2014

SÃO PAULO. Secretaria de Agricultura e Abastecimento. Resolução SAA nº 03, de 10 de janeiro de 2008. Altera dispositivo da Resolução SAA nº 24, de 01 de agosto de 1994. São Paulo, 2008. Disponível em: <https://www.defesa.agricultura.sp.gov.br/legislacoes/resolucao-saa-03-de-10-01-2008,812.html>. Acesso em: 20 dez. 2021.

SCHMIDHUBER J. Deep learning in neural networks: an overview. *Neural networks : the official journal of the International Neural Network Society*, v.61, p.85–117. 2015.

SEERA, M., LIM, C.P., KUMAR, A. *et al.* An intelligent payment card fraud detection system. *Annal of Operations Research*. v. 08. 2021.

SHARIF, T., AHMAD, M.Q., BILAL, YOUSAF, A., *et al.* Effect of severity of sub-clinical mastitis on somatic cell count and lactose contents of buffalo milk. *Pakistan Veterinary Journal*, v. 27, n. 3, p. 142-144, 2007

SHARMA, S. Activation functions in neural networks. *International Journal of Engineering Applied Sciences and Technology*, Vol. 4, Issue 12, ISSN No. 2455-2143, Pages 310-316. 2020.

SILVA, C.L.; SALES, G.A; SANTOS NETO, J.; *et al.* Detecção de fraude em amostras comerciais de queijo bubalino por adição de leite bovino por meio da técnica de Reação em Cadeia da Polimerase (PCR) multiplex. *Revista Instituto Adolfo Lutz. São Paulo*. v.74 n.1 p.21-29. 2015

SILVA, G.C; RIBEIRO, L.F. Os bubalinos no brasil e a produção de leite. *GETEC*, v.10, n.26, p.42-50. 2021

SILVA, I. D.; SILVA, K. F. S. T. Total and differential cell counts in buffalo (*Bubalus bubalis*) milk. *Buffalo Journal*, v. 10, n. 2, p 133-137, 1994.

SILVA, I. N., SPATTI, D. H., & FLAUZINO, R. A. *Redes Neurais Artificiais para engenharia e ciências aplicadas*. São Paulo: Artiliber Editora. 2010

SILVA, L. K. R. Detecção da adição do leite vaca ao leite de búfala por meio da técnica espectroscópica FTIR-ATR e quimiometria. Itapetinga – BA: UESB, 2017. 88 p. *Dissertação – Mestrado em Engenharia e Ciência de Alimentos*. 2017.

SINGH, M. e LUDRI, R.S. Somatic Cell Counts in Marrah buffaloes (*Bubalus bubalis*) during different stages of lactation, parity and season. *Asian-Australasian Journal of Animal Sciences*, v. 14, n. 2, p. 189-192, 2001.

SMOLA, A. J.; BARTLETT, P. E SCHUURMANS, D. “*Advances in large margin classifiers,*” Cambridge, MA, USA: MIT Press. 2000

SOARES, A. D., RANGEL, A. H. N., MEDEIROS, H. R. et al. Nitrogênio uréico e caseína do leite de búfala em diferentes ordens de parto. *Agropecuária Científica No Semiárido*, v. 9, n. 2, p. 94-101, 2013.

SOLLECITO, N. V.; LOPES, L. B.; LEITE, R. C. Sensibilidade antimicrobiana e microorganismos isolados de mastites em búfalos: Breve revisão. *Revista Brasileira de Medicina Veterinária*, v. 33, p. 18-22, 2011.

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

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

TEIXEIRA, L. V.; BASTIANETTO, E.; OLIVEIRA, D. A. A. Leite de búfala na indústria de produtos lácteos. *Revista Brasileira de Reprodução Animal*, v. 29, n. 2, p. 96-100, 2005.

TRIMBOLI, F.; COSTANZO, N., LOPREIATO, V. *et al.* Detection of buffalo milk adulteration with cow milk by capillary electrophoresis analysis. *Journal of Dairy Science*. v.102 p.5962–5970. 2019

VALENTE, G. D. F. S.; GUIMARÃES, D. C.; GASPARDI, A. L. A. Aplicação de redes neurais artificiais como teste de detecção de fraude de leite por adição de soro de queijo. *Revista Instituto Laticínios Cândido Tostes, Juiz de Fora*, v. 69, n. 6, p. 425-432, 2014.

VAPNIK, V. *The nature of statistical learning theory*. Springer science & business media, 1999.

VAZ, A.J. Imunoensaios utilizando conjugados. In: VAZ, A.J; TAKEI, K; BUENO, E.C, editors. *Imunoensaios fundamentos e aplicações*, p.67-83. Rio de Janeiro: Guanabara Koogan. 2012

VERRUMA, M. R.; SALGADO, J. M. Análise química do leite de búfala em comparação ao leite de vaca. *Scientia Agricola*, v. 51, n. 1, p. 131-137, 1994.

VIGUIER, C., ARORA, S., GILMARTIN, N., et al. Mastitis detection: current trends and future perspectives. *Trends in Biotechnology*, v. 27, 8, p.486-493, 2009.

YAO, J.; PAN, Y.; YANG, S.; et al. Detecting Fraudulent Financial Statements for the Sustainable Development of the Socio-Economy in China: A Multi-Analytic Approach. *Sustainability 2019*, v.11, p.1579. 2019.

YOUSSEFI, S.; EMAM-DJOMEH, Z.; MOUSAVI, S. M. Comparison of Artificial Neural Network (ANN) and Response Surface Methodology (RSM) in the prediction of quality parameters of spray-dried pomegranate juice. *Drying Technology*, New York, v. 27, n. 7/8, p. 910-917, Apr. 2009.

YUAN, B. Efficient hardware architecture of softmax layer in deep neural network. In: 2016 29th IEEE International System-on-Chip Conference (SOCC). *IEEE,2016*. p.323-326. 2016

ZAVA, M. *El Bufalo Domestico*. Ed. Orientation Grafica, Buenos Aires, Argentina. 2011.

ZEN, Q. K., REN, F. Z., YANG, B. Z., et al. Study on protein composition of milk from different buffalo breeds and crossbred progenies. In: *Asian buffalo congress on social economic contribution of buffalo to rural areas*, v. 1. p. 287-292, 2006.

ZHANG, H. *The optimality of naive bayes*. In 2004 FLAIRS Conference - AAI, 2004.

ZICARELLI, L. Enhancing reproductive performance in domestic dairy water buffalo (*Bubalus bubalis*). *Society for Reproduction and Fertility Supplement*, v.67, p.443-455, 2010.

