

PONTIFÍCIA UNIVERSIDADE CATÓLICA DE GOIÁS
PROGRAMA DE MESTRADO EM ENGENHARIA DE PRODUÇÃO E SISTEMAS

**CLASSIFICAÇÃO MULTIVARIADA PARA O DIAGNÓSTICO DE MASTITE
BOVINA SUBCLÍNICA EMPREGANDO IMAGEM E RESSONÂNCIA
MAGNÉTICA NUCLEAR**

MARCILON FONSECA DE LIMA

2017



PONTIFÍCIA UNIVERSIDADE CATÓLICA DE GOIÁS
PROGRAMA DE MESTRADO EM ENGENHARIA DE PRODUÇÃO E
SISTEMAS

CLASSIFICAÇÃO MULTIVARIADA PARA O DIAGNÓSTICO DE MASTITE
BOVINA SUBCLÍNICA EMPREGANDO IMAGEM E RESSONÂNCIA
MAGNÉTICA NUCLEAR

MARCILON FONSECA DE LIMA

Dissertação de Mestrado apresentada ao Programa de Pós-Graduação em Engenharia de Produção e Sistemas da Pontifícia Universidade Católica de Goiás, como parte dos requisitos para obtenção do título de Mestre em Engenharia de Produção e Sistemas.

Orientador Dr. Clarimar José Coelho

Bolsista FAPEG

Goiânia

Setembro de 2017

L732c

Lima, Marcilon Fonseca de

Classificação multivariada para diagnóstico de mastite bovina subclínica empregando imagem e ressonância magnética nuclear[manuscrito]/ Marcilon Fonseca de Lima.-- 2017.

61 f.; il. 30 cm

Texto em português com resumo em inglês

Dissertação (mestrado) - Pontifícia Universidade Católica de Goiás, Programa de Pós-Graduação Stricto Sensu em Engenharia de Produção e Sistemas , Goiânia, 2017

Inclui referências 53-58

1. Mastite. 2. Bovino de leite - Doenças. 3. Quimioterapia.
4. Ressonância magnética. I.Coelho, Clarimar José.
II.Pontifícia Universidade Católica de Goiás. III.

Título.

CDU: 636.2:618.19-02(043)

**CLASSIFICAÇÃO MULTIVARIADA PARA O DIAGNÓSTICO DE MASTITE
BOVINA SUBCLÍNICA EMPREGANDO IMAGEM E RESSONÂNCIA
MAGNÉTICA NUCLEAR**

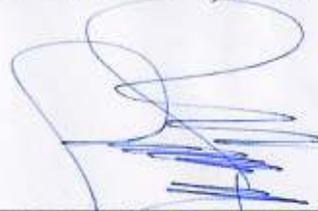
MARCILON FONSECA DE LIMA

Esta Dissertação julgada adequada para obtenção do título de Mestre em Engenharia de Produção e Sistemas, e aprovada em sua forma final pelo Programa de Pós-Graduação em Engenharia de Produção e Sistemas da Pontifícia Universidade Católica de Goiás em setembro de 2017.



Prof. Marcos Lajovic Carneiro, Dr.
Coordenador do Programa de Pós-Graduação em
Engenharia de Produção e Sistemas

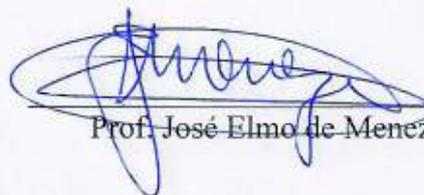
Banca Examinadora:



Prof. Clarimar José Coelho, Dr.
Orientador



Prof. Gustavo Teodoro Laureano, Dr.



Prof. José Elmo de Menezes, Dr.

À memória de minha mãe, Maria Eleuza Fonseca de Lima, que dedicou seus melhores anos em prol de minha educação intelectual, moral e física.

À minha esposa e filhos pela imensa paciência e carinho nesse período de estudos e dedicação.

Com carinho, dedico

AGRADECIMENTOS

- À minha família, por tudo;
- À FAPEG, pela bolsa concedida;
- Ao meu orientador, Dr. Clarimar;
- À minha amiga Priscilla Borges pela amizade, incentivo e auxílio nos artigos e na escrita deste trabalho;
- Ao Prof. Dr. Igor pelo auxílio em vários momentos da pesquisa e aquisição dos dados das amostras;
- Ao Prof. Me Arlindo por ótimas sugestões;
- Ao Prof. Dr. Daniel Lucena;
- Ao Amigo Prof. Dr. Wellington Martins pela amizade e horas dispensadas me ouvindo e contribuindo;
- À colega Elvia, pelo apoio até nas madrugadas, me auxiliando com as bibliografias;
- Ao Agamenon por palavras amigas e imenso auxílio em muitos momentos de pesquisa;
- À Tássia, por toda amizade e as anotações carinhosas em minha qualificação;
- Ao Lucas por me auxiliar nos conceitos de Química;
- Ao Leonardo pelo auxílio nos códigos do *Matlab*®;
- Ao André Luis, colega de estudo e incentivador;
- Aos meus demais colegas de mestrado que sempre me apoiaram e deram uma contribuição sem par a esse trabalho: Bianca, Gyovanna, Viviane Vaz, Zenilda, Daniel e Emerson;
- Aos amigos Prof. Dr. Juliano Geraldo e Prof. Me Felipe Correa por tantas palavras de incentivo e apoio nas horas complicadas;
- Ao Ernani, secretário do programa, sempre pronto para auxiliar em todos momentos;
- Ao amigo Edgard;
- À Pontifícia Universidade Católica de Goiás;

- À Universidade Federal de Goiás;
- Ao Instituto Federal de Goiás
- A todos aqueles que de alguma forma contribuíram para a realização deste trabalho.

“Os problemas significativos que enfrentamos não podem ser resolvidos no mesmo nível de pensamento em que estávamos quando os criamos.”

Albert Einstein

“Não há fatos eternos, como não há verdades absolutas.”

“Es gibt keine ewigen Tatsachen, da es keine absoluten Wahrheiten gibt.”

Friedrich Nietzsche

RESUMO

O objetivo do trabalho é fazer o diagnóstico de mastite subclínica em gado bovino empregando classificação multivariada usando imagens, espectros de ressonância magnética nuclear (*Nuclear Magnetic Resonance*, NMR) e um conjunto de dados de inventário bovino (DIB). A classificação de objetos de interesse nas imagens de leite, NMR e DIB é feita com os algoritmos de projeções sucessivas, análise discriminante linear e análise discriminante por mínimos quadrados parciais. O resultado do diagnóstico obtido por classificação por imagem no melhor caso foi de 75% e com NMR de 78%. Como resultado pode-se observar que há possibilidade de classificação para diagnóstico da mastite por imagem e por NMR, porém é interessante aumentar a quantidade de amostras para um melhor termo.

Palavras-chave: Leite, Quimiometria, Imagem digital, Ressonância Magnética Nuclear, Mastite subclínica.

ABSTRACT

The goal of this study is to diagnose subclinical mastitis in cattle using multivariate classification applying images, Nuclear Magnetic Resonance (NMR) spectra and a bovine data survey set (DIB). The interest objects classification in milk images, NMR and DIB are processed with the algorithm of successive projections, linear discriminant analysis, and partial least squares discriminant analysis. The best diagnostic results obtained by image classification was 75% and 78% with NMR. As a result it can be observed that there is the possibility of classification for the diagnosis of mastitis by image and NMR, however, it is interesting to increase the number of samples for increasing the precision of two evaluated diagnosis methods.

Keywords: Milk, Chemometrics, Digital Image, Nuclear Magnetic Resonance, Subclinical Mastitis

LISTA DE FIGURAS

FIGURA 1 – PLACA PETRI 96 POÇOS COM AS AMOSTRAS DE NATA E LEITE DESNATADO.	19
FIGURA 2 – ESCANEAMENTO DA PLACA PETRI – IMAGEM ADAPTADA DO SITE DA HP (HP, 2017)	19
FIGURA 3 - FLUXOGRAMA GERAL DO TESTE INDIRETO PROPOSTO.....	21
FIGURA 4 – IMAGEM ESCANEADA DAS AMOSTRAS DE LEITE	26
FIGURA 5 – PREPARAÇÃO PARA SEGMENTAÇÃO: (A) IMAGEM MÉDIA DIGITALIZADA; (B) IMAGEM BINARIZADA; (C) MÁSCARA PARA OS CORTES.	27
FIGURA 6 – IMAGENS SEGMENTADAS DAS AMOSTRAS DE NATA (PRIMEIRAS 40 DA ESQUERDA PARA DIREITA E DE CIMA PARA BAIXO) E DE LEITE DESNATADO.	28
FIGURA 7 – HISTOGRAMAS NAS DIMENSÕES R, G, B, CINZA, H, S E V, RESPECTIVAMENTE, DAS AMOSTRAS DE NATA. OS VALORES ZEROS FORAM EXCLUÍDOS DOS DADOS.....	30
FIGURA 8 – HISTOGRAMA DOS ESPAÇOS CINZA E HSV DAS AMOSTRAS DE LEITE DESNATADO. OS VALORES DE FREQUÊNCIA ZERO FORAM EXCLUÍDOS.	31
FIGURA 9 – GRÁFICO DIB DAS AMOSTRAS. AS UNIDADES E ESCALAS OBEDECEM A TABELA DO APÊNDICE A, COLUNAS (1) A (9). CADA COR É UMA AMOSTRA.	32
FIGURA 10 – GRÁFICOS PLS-DA: NÚMERO DE AMOSTRAS X RESPOSTA POR CLASSE (A) CLASSE 1- NEGATIVA, (B) CLASSE 2 - POSITIVA, (C) CLASSE 3 - INDETERMINADA E O (D) COEFICIENTE DE CLASSE X PESO (WEIGHT) DA VARIÁVEL LATENTE 1. OS VALORES DE 1 A 26 SÃO REFERENTES À VALIDAÇÃO E OS VALORES DE T1 A T13 SÃO OS VALORES DE TESTE DA CLASSIFICAÇÃO.	41
FIGURA 11 – GRÁFICO NÚMERO DE VARIÁVEIS X CUSTO DE VALIDAÇÃO NOS ESPAÇOS DE CORES (A) RGB-CINZA-HSV, (B) RGB, (C) R, (D) G, (E) B, (F) CINZA, (G) HSV, (H) H, (I) S, (J) V, (K) RGB-CINZA, (L) RGB-HSV, (M) CINZA-HSV.	45
FIGURA 12 – ALGUMAS IMAGENS DE AMOSTRAS DE LEITE NA FORMA DE NATA E LEITE DESNATADO FILTRADAS COM MAIOR INTENSIDADE PARA AMPLIFICAR DIFERENÇAS.	47
FIGURA 13 – GRÁFICOS AMOSTRA X RESPOSTA CALCULADA NAS CLASSES (A) 1-SUSPEITA, (B) 2- INSUSPEITA E (C) 3-INDETERMINADA. NUMERAÇÃO DE 1 A 30 SÃO AS VARIÁVEIS, SENDO QUE AS T1 A T9 SÃO AS VARIÁVEIS DE TESTE. TÉCNICA PLS-DA PARA NATA.	49

LISTA DE TABELAS

TABELA 1 – DISTRIBUIÇÃO DE QUANTIDADE DE AMOSTRAS PARA CLASSIFICAÇÃO PELO ALGORITMO KS	21
TABELA 2 – CÁLCULO DE DIFERENÇA EM MÉDIA DE IMAGENS.....	25
TABELA 3 – COEFICIENTES w_j DO FILTRO SAVTZKY-GOLAY PARA SUAVIZAÇÃO COM BASE EM POLINÔMIOS DE GRAU 3 E 4. FONTE (CHAU ET AL., 2004)	33
TABELA 4 – CLASSES DE SUSPEITA DE OCORRÊNCIA DE MASTITE DE ACORDO COM A SCC	37
TABELA 5 – DADOS COMPARATIVOS DE TAXA DE CLASSIFICAÇÃO CORRETA (TCC) PARA VALIDAÇÃO E TESTE QUANTO A SUSPEITA DA MASTITE SUBCLÍNICA EM AMOSTRAS DE LEITE CRU RELATIVO AOS DIB.....	40
TABELA 6 – VALIDAÇÃO DA CLASSIFICAÇÃO UTILIZANDO PLS-DA EM AMOSTRAS DE LEITE E DIB	42
TABELA 7 – TABELA DE CONFUSÃO DAS AMOSTRAS TESTE – CLASSIFICAÇÃO PLS-DA PARA DIB	42
TABELA 8 – CLASSIFICAÇÃO DAS AMOSTRAS DE TESTE COM SPA-LDA E DIB.....	43
TABELA 9 – RESUMO DAS TCC NOS MODELOS EM PLS-DA E SPA-LDA PARA AMOSTRAS DE NATA	44
TABELA 10 - RESUMO DAS TCC NOS MODELOS EM PLS-DA E SPA-LDA PARA AMOSTRAS DE LEITE DESNATATO.....	46
TABELA 11 - RESUMO DAS TCC NOS MODELOS EM PLS-DA E SPA-LDA APLICADAS EM RESSONÂNCIA MAGNÉTICA NUCLEAR (NMR) NAS AMOSTRAS DE NATA E LEITE DESNATADO.....	47

LISTA DE ABREVIATURAS E SIGLAS

ACEK	<i>Alternate Current Electrokinetics</i> (Corrente Alternada Eletrocinática)
CMT	<i>California Mastitis Test</i>
EC	<i>Electrical Conductivity</i> (Condutividade Elétrica)
DA	<i>Discriminant Analysis</i> (Análise Discriminante)
DIA	<i>Digital Image Analysis</i> (Análise digital de imagem)
DIB	Dados de Inventário Bovino
DIP	<i>Digital Image Processing</i> (Processamento digital de imagem)
dpi	<i>Dots per inch</i> (pontos por polegada)
DSP	<i>Digital signal processing</i> (Processamento Digital de sinais)
GA	<i>Genetic Algorithm</i> (Algoritmo Genético)
GDT	<i>Gaussian Decision Tree</i> (Árvore de Decisão Gaussiana)
HSV	<i>Hue, Saturation, Value</i> (Matiz, Saturação, Valor)
KNN	<i>K - Nearest Neighbor</i> (Vizinho K Mais Próximo)
KS	Kennard Stone
LDA	<i>Linear Discriminant Analysis</i> (Análise Discriminante Linear)
MIA	<i>Multivariate Image Analysis</i> (Análise Multivariada de Imagens)
MLR	<i>Multivariate Linear Regression</i> (Regressão Linear Múltipla)
NIR	<i>Near Infrared</i> (Infravermelho Próximo)
NMR	<i>Nuclear Magnetic Resonance</i> (Ressonância Magnética Nuclear)
PLS	<i>Partial Least Square</i> (Mínimos Quadrados Parciais)
PLS2	<i>Partial Least Square</i> (Mínimos Quadrados Parciais 2)
PLS-DA	<i>Partial Least Square – Discriminant Analysis</i> (Análise Discriminante pelos Mínimos Quadrados Parciais)
RGB	<i>Red, Green, Blue</i> (Vermelho, Verde, Azul)
RMSEV	<i>Root Mean Square Error of Validation</i> (Raiz Quadrada do Erro Médio Quadrático de Validação)
ROI	<i>Region of Interest</i> (Região de Interesse)

SIMCA	<i>Soft Independent Modeling of Class Analogies</i> (Modelagem Independente e Flexível por Analogia de Alasses)
SC	<i>Somatic Cells</i> (Células Somáticas)
SCC	<i>Somatic Cell Count</i> (Contagem de Células Somática)
SPA	<i>Successive Projections Algorithm</i> (Algoritmo de Projeções Sucessivas)
SPA-LDA	<i>Successive Projections Algorithm and Linear Discriminant Analysis</i> (Análise Discriminante Linear pelo Algoritmo de Projeções Sucessivas)
TCC	Taxa de Classificação Correta
WMT	<i>Wisconsin Mastitis Test</i>

SUMÁRIO

1.	Introdução	14
2.	Materiais e métodos.....	18
2.1.	Dados usados para o diagnóstico da mastite subclínica	18
2.2.	Teste indireto proposto	20
2.3.	Metodologia usada para obter os resultados	23
2.3.1.	Pré-processamento das Imagens Digitais	24
2.3.2.	Segmentação da Imagem	25
2.3.3.	Histogramas de imagens	28
2.3.4.	Pré-Processamento de Dados de Inventário Bovino e sinais NMR	31
2.3.5.	Classificação multivariada dos dados de DIB, RMN e imagens digitais	34
2.3.5.1.	SPA-LDA	35
2.3.5.2.	PLS-DA	37
2.4.	Diagnóstico de Mastite através da Classificação multivariada.....	38
2.5.	Softwares utilizados	39
3.	Resultados e Discussões	40
3.1.	Classificação Multivariada	40
3.1.1.	Classificação Multivariada aplicada em MIA	43
3.1.2.	Classificação Multivariada aplicada em NMR	47
4.	Conclusões	51
	Referências bibliográficas	53
	Apêndices	59
A.	Dados de Inventário Bovino e SCC sobre as amostras	59

1. INTRODUÇÃO

A classificação multivariada para o diagnóstico da mastite subclínica a partir de imagens digitais e ressonância magnética nuclear é uma abordagem para o controle e prevenção da mastite bovina (ČEREŠKIENĖ *et al.*, 2016). A mastite ou mamite é uma das principais doenças do gado leiteiro que se apresenta na forma clínica e subclínica (DUGUMA; TOLOSA; YOHANNES, 2014). A forma clínica apresenta alterações visíveis no leite ou no úbere enquanto a forma subclínica não apresenta alterações visíveis. A forma subclínica provoca mudança na composição e volume do leite produzido e favorece sua disseminação no rebanho porque não apresenta evidência de anormalidade do rebanho (KHAN; KHAN, 2006; TESFAYE; REGASSA; KELAY, 2010). A doença acomete de 20 a 38% do rebanho brasileiro com prejuízos de aproximadamente 2,8 bilhões de litros de leite por ano e correspondente a uma perda produtiva de 12 a 15% (LOPES; LACERDA; RONDA, 2013).

Um sistema de reconhecimento de padrões automático classifica dados em um número de categorias ou classes (BISHOP, 2006). O sistema consiste na análise ou descrição que extrai as características do padrão em estudo e a classificação ou reconhecimento que permite reconhecer um objeto ou um padrão a partir das características derivadas da análise (KOVALEVSKY, 1980). Uma classe é um conjunto de objetos com propriedades comuns importantes e essas propriedades ou características são utilizadas para a classificação, que é o processo de atribuição de um rótulo a um objeto de acordo com alguma representação das propriedades do objeto (ABBURU; BABU GOLLA, 2015). Quando a classificação é baseada em um conjunto de treinamento ou conjunto de padrões já classificados, a estratégia é denominada aprendizagem supervisionada e aprendizagem não supervisionada, caso contrário (KPALMA; RONSIN, 2007).

Na prática, uma classe à qual um objeto pertence, é denotada por algum rótulo de classe. Um classificador é um algoritmo que insere uma representação do objeto e atribui um rótulo da classe (SHAPIRO; STOCKMAN, 2000). Analogamente, o computador adquire a capacidade humana de interpretar os dados presentes na imagem. O processo de classificação (classificador) reconhece informações importantes a partir de detalhes que parecem irrelevantes, aprende a partir de algumas informações, generaliza para aplicações em outras circunstâncias e faz inferências a partir de informações incompletas (APLIN; SMITH, 2008).

Para o reconhecimento de padrões (descrição e classificação) em imagens as características mais empregadas são a cor e a textura. Uma boa quantificação dessas características permite a identificação e a classificação de padrões (KOTSIANTIS, 2007). A

classificação baseada na cor é a forma mais simples. A classificação por textura envolve maior complexidade. Não existe um método genérico ou abordagem formal para uma grande variedade de imagens (MEDJAHED, 2015).

A ressonância magnética nuclear (*Nuclear Magnetic Resonance*, NMR) é uma técnica baseada na propriedade dos núcleos magneticamente ativos que respondem a perturbação induzida numa amostra e aplica um impulso de onda de rádio (HORE, 2015). Quando os núcleos são imersos em campo magnético intenso respondem ao pulso que oscila a uma frequência particular e gera um sinal que é gravado e transformado em um espectro que mostra a intensidade da resposta como uma função da frequência de oscilação (LAMBERT; MAZZOLA, 2004). A frequência de cada núcleo é uma característica da sua posição na molécula e depende do estado físico-químico das substâncias. A resposta depende não apenas da estrutura da molécula à qual o átomo pertence, mas também do ambiente em que a molécula está imersa (KEELER, 2002). A NMR é uma técnica da química analítica com aplicações em medicina (KUPČE; FREEMAN, 2003), agricultura (GAMBHIR; NAGARAJAN, 1999), alimentos (CAPOZZI, 2013) e petroquímica.

A mastite subclínica não apresenta sinais evidentes e seu diagnóstico não é possível sem o uso de testes auxiliares. Um sinal da mastite subclínica no leite é a elevação da contagem das células somáticas (*Somatic Cell Count*, SCC) que indicam a presença de inflamação da glândula mamária da vaca (FORSBÄCK *et al.*, 2010). Em geral, os testes auxiliares usados para o diagnóstico da mastite subclínica são a SCC, o teste de contagem de mastite Califórnia (*California Mastitis Test*, CMT) (JOHN R. MIDDLETON *et al.*, 2004), o teste de mastite Wisconsin ou viscosímetro (*Wisconsin Mastitis Test*, WMT) (SCHULTZE; BRAMLEY, 1982) e há possibilidade de teste de condutividade elétrica (*Electrical Conductivity*, EC) (DUARTE; FREITAS; BEXIGA, 2015).

A contagem de células somáticas individual é um método eletrônico de contagem cada vez mais utilizado no Brasil e que constitui a base do monitoramento da saúde do úbere nos países desenvolvidos (LANGONI, 2000). O teste de CMT também detecta alterações de pH do leite. No caso de mastite, o leite se torna alcalino e, quando em contato com o reagente, um detergente aniônico neutro, apresenta coloração púrpura intensa (MELLENBERGER, 2001). A subjetividade do resultado do CMT representa um aspecto negativo, principalmente no que se refere a dosagens erradas de reagente e de leite e da falta de padronização da leitura, que estará sujeita a variações, de acordo com o profissional que realizar o teste (BENEDETTE *et al.*, 2008). Por sua vez, o teste WMT é um aprimoramento do CMT desenvolvido para diminuir

ou eliminar a subjetividade do teste, pois é executado em tubo graduado com quantidades medidas de leite e de reagentes, o mesmo utilizado no CMT, feitas apropriadamente (COSER; LOPES; COSTA, 2012).

Os processos provenientes da quimiometria utilizando sinais de equipamentos para análise química (FERREIRA *et al.*, 1999), e para diagnóstico da mastite estão em desenvolvimento (SCHABAUER *et al.*, 2014). São usados diferentes tipos de sinais como NMR (SUNDEKILDE, U K *et al.*, 2013), infravermelho próximo (*Near Infrared*, NIR) (TSENKOVA *et al.*, 2006) ou até mesmo imagens digitais (MELO *et al.*, 2015).

Sinais de NMR e análise multivariada estão sendo utilizados para a criação de biomarcadores para o diagnóstico da mastite subclínica. Atualmente estas abordagens não possuem protocolos definidos, mesmo que produzam resultados promissores (SUNDEKILDE, ULRIK K; LARSEN; BERTRAM, 2013). Os biomarcadores são componentes químicos que alteram alguma propriedade do material em análise em resposta a algum fator ambiental ou ao uso de alguma droga podendo ser utilizado para o diagnóstico. No caso da mastite, quando o aumento da SCC implicar em alguma variação em determinado componente, este indicaria a presença da doença (PEDRAZ, 2011). O lactato e acetato são biomarcadores em estudo com a utilização da NMR enquanto butirato, hipurato, isoleucina, fumarato e beta-hidroxibutirato (BHBA) foram observados em alteração de teor com o aumento da SCC, sendo possíveis biomarcadores para diagnóstico da mastite subclínica (SUNDEKILDE, U K *et al.*, 2013).

O uso da espectroscopia em infravermelho próximo (*Near Infrared*, NIR) também possibilita um diagnóstico da mastite em conjunto com a SCC, a condutividade elétrica e o uso da quimiometria. A taxa de classificação correta verificada com a utilização desse método foi entre 93 e 98% no diagnóstico (TSENKOVA *et al.*, 2006).

Pesquisadores da Universidade do *Tennessee* desenvolveram um processo para diagnóstico precoce da mastite utilizando corrente alternada eletrocinética (*Alternate Current Electrokinetics*, ACEK) de baixa voltagem integrada a biosensores, que são dispositivos utilizados para detecção de um objeto observado aliando componente físico-químico e biológico. Foi utilizado o classificador Árvore de Decisão Gaussiano (*Gaussian Decision Tree*, GDT) com o qual foi possível fazer a classificação com taxas de acerto de 100% com tensão de 145 mV e 80% para uma tensão de 40 mV (ZHANG *et al.*, 2015).

Alguns pesquisadores de universidades brasileiras obtiveram ótimos resultados em diagnóstico de mastite utilizando quimiometria por imagem de microscópio fazendo contagem

de células somáticas através de processo de segmentação de imagens (MELO *et al.*, 2015), que é a retirada da imagem de objetos que não sejam de interesse (NUNES; CONCI, 2010). Outro estudo utilizou imagens térmicas em bovinos para detectar a infecção através da temperatura dos úberes. A temperatura dos quartos mamários da vaca quando infectada é mais elevada que naquelas não afetadas. Ainda há a necessidade de estabelecer um protocolo para colocação do processo em uso. A técnica tem potencial para diagnosticar a mastite subclínica com o uso de imagem térmica devido ao baixo custo e não requerer intervenção laboratorial e ainda dispensar coleta de amostras (DIGIOVANI *et al.*, 2016).

Apesar do desenvolvimento de novas tecnologias até o momento não é possível detectar com facilidade o estado subclínico da mastite (SUNDEKILDE, U K *et al.*, 2013). Os métodos auxiliares são abordagens incompletas no percurso da solução do diagnóstico da mastite subclínica.

O objetivo deste trabalho é o desenvolvimento de uma ferramenta de teste auxiliar para o diagnóstico da mastite subclínica a partir da classificação do leite empregando a análise multivariada de imagem (*Multivariate Image Analysis*, MIA), sinais de NMR e dados de inventário bovino (DIB).

O texto está organizado em quatro capítulos. O Capítulo 2 apresenta a definição do problema da pesquisa, os dados e os métodos usados para obter os resultados. O Capítulo 3 traz os resultados da pesquisa e as discussões pertinentes. O capítulo 4 faz os comentários finais e relaciona as conclusões.

2. MATERIAIS E MÉTODOS

Os testes para a contagem de SC são diretos ou indiretos. Os testes diretos fazem a contagem das células por microscópio ou utilizam equipamento eletrônico. Os equipamentos eletrônicos fazem a contagem automaticamente. Feito o depósito da amostra no equipamento, os procedimentos internos realizam a contagem sem interferência do usuário (DOHOO; MEEK, 1982; MILLER; PAAPE; ACTON, 1986). Este trabalho propõe o desenvolvimento de um teste auxiliar empregando método indireto de contagem de SC para o diagnóstico da mastite subclínica. O teste auxiliar proposto faz a contagem indireta da SC a partir da análise de sinais NMR, imagens digitais e inventário bovino provenientes de amostras de leite.

2.1. Dados usados para o diagnóstico da mastite subclínica

As amostras de leite foram adquiridas em 08/10/2016 na fazenda da Universidade Federal de Goiás em quarenta bovinos da raça jersolando, híbrido de *jersey* e holandês. Após a coleta realizada foram anotados dados relativos à quantidade de crias, quantidade de leite ordenhado, idade e data do último parto. Estes dados constam da tabela do apêndice A. O horário da coleta se deu entre as 6h e 7h14min.

As amostras foram centrifugadas por 20 minutos a uma rotação de 4.000 rpm e depois armazenadas na forma de nata e leite desnatado à 5°C (cinco graus Celsius). No laboratório da qualidade do leite da Universidade Federal de Goiás que tem credenciamento no Ministério da Agricultura (BRASIL, 2014) são realizados testes para quantificar os teores de gordura, caseína, proteínas, lactose e ureia. Estas informações compõem o inventário bovino (DIB). Os testes para contagem de células somáticas (SCC) também são obtidos nos laboratórios da UFG. Todos os testes foram realizados de acordo com instrução normativa vigente do Ministério da Agricultura (BRASIL, 2006). A norma específica para obtenção da SCC foi a ISO 13366 (INTERNATIONAL ORGANIZATION FOR STANDARDIZATION, 2006). A norma a ISO 9622 (INTERNATIONAL ORGANIZATION FOR STANDARDIZATION, 2013) foi a norma utilizada para aquisição de teor de gordura, de teor de proteína, de teor de caseína, de teor de lactose e de teor de ureia. Todos estes dados constam da tabela do Apêndice A.

As imagens digitais foram adquiridas em um escâner HP Deskjet F4280 *All-in-one* com resolução de 1.200 x 2.400 pontos por polegada (*dots per inch*, dpi) e profundidade de cor de 48 bits, sendo as digitalizações obtidas com 600 dpi. As amostras de leite em nata e leite desnatado são colocadas em uma placa de *Petri* 96 poços com fundo chato numeradas da esquerda para a direita até a coluna 10. A Figura 1 mostra a organização das amostras na placa para a digitalização com o escâner. As amostras de nata são colocadas nas quatro primeiras filas

(de A até D) e as amostras de leite desnatado nas filas (de E até H) num total de 40 amostras. As duas últimas colunas (11 e 12) não são usadas. As amostras das posições A2, A3, A5, C1, E1, E5, E6, G1 e G10 se perderam antes da digitalização.



Figura 1 – Placa Petri 96 poços com as amostras de nata e leite desnatado.

A placa *Petri* foi colocada no escâner, como pode ser visto na Figura 2, e um anteparo fundo infinito de tecido na cor preta de estúdio fotográfico é colocado sobre as amostras. Esta ação foi realizada para evitar reflexos e facilitar a visualização, pois a almofada de documento do escâner é de cor branca. O *software* utilizado para aquisição das imagens é HP 140.0. As imagens foram gravadas em formato *.png* por ser um formato livre e apesar de ser comprimido mantém todas informações da imagem (SOLOMON; BRECKON, 2013).

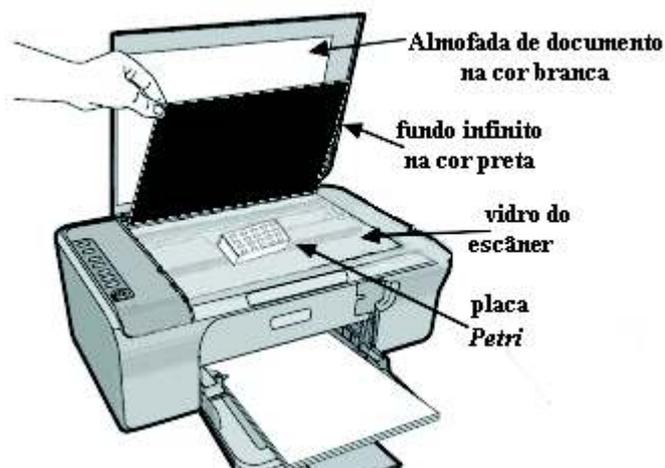


Figura 2 – Escaneamento da placa *Petri* – imagem adaptada do site da HP (HP, 2017)

Os sinais de NMR são obtidos com as amostras de leite em fase gordurosa (nata do leite) e aquosa (leite desnatado) empregando um espectrômetro *Bruker Avance III* 500 MHz para o isótopo de ^1H equipado com uma sonda de 5 mm de detecção inversa com gradiente de campo (*Triple Band Inverse – TBI*) no Instituto de Química (IQ-1) da Universidade Federal de Goiás (UFG), Goiânia-GO. São utilizados os solventes clorofórmio deuterado para a nata e água deuterada para a fase aquosa.

2.2. Teste indireto proposto

O diagnóstico da mastite subclínica empregando o teste proposto considera a quantidade de SC presente na amostra de leite. A quantidade de SC que caracteriza ou não amostras com mastite subclínica é definida a partir da literatura. A contagem de 200.000 SC/mL é utilizada em testes como limite para diagnóstico de mastite (DUARTE; FREITAS; BEXIGA, 2015; HEALD et al., 2000). Dohoo e Meek e Tsenkova e colaboradores (DOHOO; MEEK, 1982; TSENKOVA et al., 2006) admitem 300.000 SC/mL como limite máximo para a amostra estar livre de mastite. Indicam também a possibilidade de explorar outros valores para o diagnóstico da mastite subclínica. A legislação brasileira e União Europeia permitem até 400.000 SC/mL para consumo humano do leite (BRASIL, 2011; SUNDEKILDE, U K *et al.*, 2013). Neste trabalho serão adotados os seguintes valores de SCC. Classe 1: SCC menor ou igual a 200.000 SC/mL é amostra negativa para suspeita da mastite subclínica. Classe 2: SCC entre 200.001 SC/mL e 300.000 SC/mL é amostra indeterminada. Classe 3: SCC maior ou igual a 300.001 SC/mL é uma amostra com suspeita para mastite subclínica.

Os dados para o teste do método proposto provenientes de DIB, NMR e de imagens são divididos em subconjuntos de treinamento, validação e teste. A divisão dos diferentes conjuntos de amostras é feita empregando o algoritmo *Kennard Stone* (KS) baseado na métrica da distância euclidiana (KENNARD; STONE, 1969). O algoritmo KS é aplicado separadamente para cada módulo de dados (DIB, NMR e imagem digital) para leite desnatado e leite gorduroso (nata).

A Tabela 1 mostra o número de amostras, amostras de treinamento, amostras de validação e amostras de teste em cada situação. No caso de amostras destinadas a produção de imagens digitais foram perdidas 5 amostras de leite desnatado e 6 amostras de nata. Para amostras de sinais NMR, uma foi perdida no processo de produção de leite desnatado.

Um fluxograma com a ideia geral do teste indireto proposto para o diagnóstico da mastite subclínica é mostrado na Figura 3. O teste proposto é indireto porque não faz a contagem

de SC diretamente do leite. Como no caso dos testes CMT e WMT, no teste proposto a contagem de SC é feita a partir de imagens no espectro visível (*Red, Green, Blue, RGB*), espectros de NMR e informações de DIB.

Tabela 1 – Distribuição de quantidade de amostras para classificação pelo algoritmo KS

	DIB		NMR		Imagem digital	
	Nata	Leite desnatado	Nata	Leite desnatado	Nata	Leite desnatado
Amostras	40	40	40	39	34	35
Treinamento	20	20	20	20	16	16
Validação	10	10	10	10	10	10
Teste	10	10	10	9	8	9

O módulo *SCC* recebe os valores da SCC e verifica a validade do valor informado através do cálculo da média, amplitude (maior valor menos o menor valor do conjunto de dados) e o desvio padrão. Foram encontradas duas amostras com valores discrepantes em relação à média. As verificações in loco confirmaram que estes valores, apesar de diferentes, são valores normais encontrados em SCC para mastite positiva.

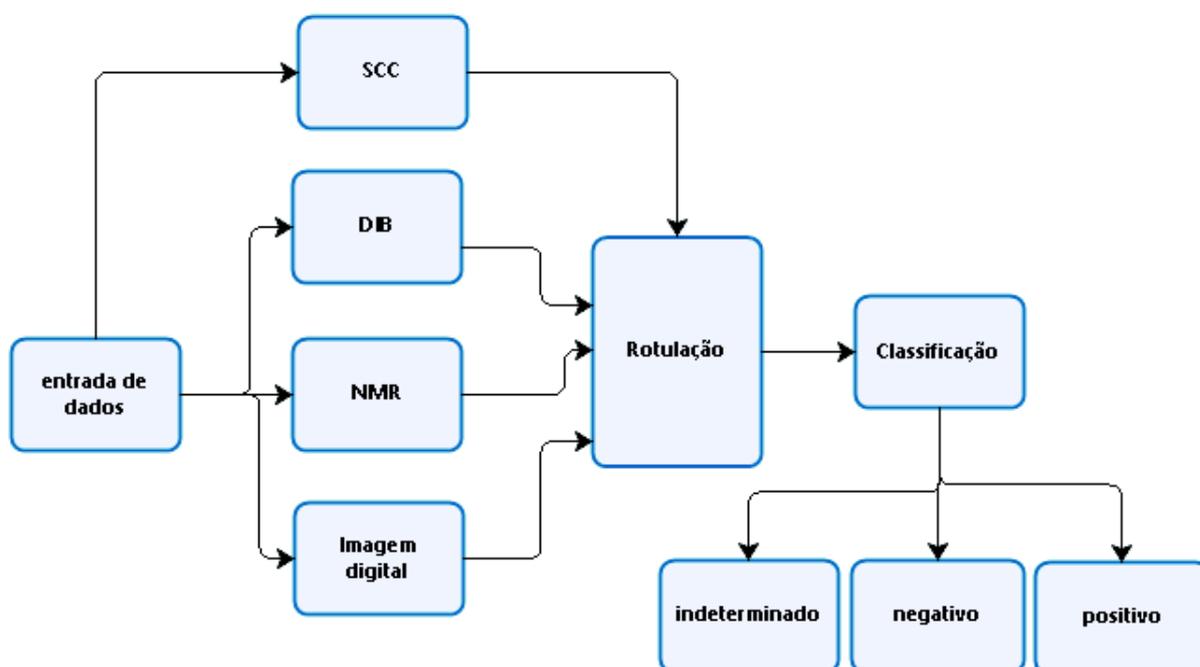


Figura 3 - Fluxograma geral do teste indireto proposto

O módulo *DIB* recebe os dados do inventário bovino e calcula a média, amplitude (maior valor menos o menor valor do conjunto de dados) e desvio padrão e não encontra discrepância para os valores.

O módulo *NMR* recebe os sinais de NMR e aplica filtragem empregando o algoritmo *Savitsky-Golay*. Aplica-se ainda o método do logaritmo e raiz quadrada na tentativa de obter melhores resultados do que com os sinais originais. Porém, nenhuma técnica de filtragem e normalização produziu melhores resultados do que os sinais originais.

O módulo *Imagem digital* recebe as imagens digitais e faz o pré-processamento empregando uma adaptação simplificada da filtragem pela média. A imagem é segmentada usando descritores de contorno e calcula os histogramas da imagem. O histograma é utilizado para fazer a correlação com a SCC no módulo de *Rotulação*.

O módulo de rotulação é um procedimento para associar amostras a uma Classe. As amostras provenientes dos módulos que antecede o módulo de *Rotulação* são armazenadas em uma matriz **A** e dois vetores **b** e **c**, Equação (1). Nas linhas da matriz **A** são armazenadas as amostras provenientes dos módulos que antecedem o módulo de *Rotulação* e nas colunas as variáveis. O vetor **b** armazena a quantidade de SC e o vetor **c** armazena o número da Classe em função da SCC. O módulo de rotulação associa uma linha de **A**, em função do número de SC do vetor **b** e atribui um número de Classe 1, 2 ou 3 obtido do vetor **c**.

$$A = \begin{bmatrix} a_{1,1} & a_{1,2} & a_{1,3} & \cdots & a_{1,n} \\ a_{2,1} & a_{2,2} & a_{2,3} & \cdots & a_{2,n} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ a_{m,1} & a_{m,2} & a_{m,3} & \cdots & a_{m,n} \end{bmatrix}; b = \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_m \end{bmatrix}; c = \begin{bmatrix} 1 \\ 3 \\ \vdots \\ 1 \end{bmatrix} \quad (1)$$

O módulo de *Classificação* utiliza o algoritmo de projeções sucessivas (*Successive Projections Algorithm*, SPA) e análise discriminante linear (*Linear Discriminant Analysis*, LDA), mínimos quadrados parciais (*Partial Least Square*, PLS) e LDA para definir o diagnóstico. Esta etapa recebe as classes rotuladas do módulo *Rotulação* juntamente com o conjunto de dados de treinamento e constrói o Classificador. Este Classificador é construído de acordo com a técnica e de acordo com os dados. A LDA trabalha com a distância de *Mahalanobis* e procura encontrar dados que tenham uma distância menor entre eles e maior entre classes. (XAVIER *et al.*, 2011) e a PLS-DA utiliza a probabilidade da amostra pertencer a uma determinada classe (BALLABIO; CONSONNI, 2013). Em seguida é feita a validação e teste do modelo.

De acordo com a quantidade de acertos na classificação é verificada a acurácia do teste proposto e isso é feito pela taxa de classificação correta, que é a razão entre a quantidade classificada corretamente e a quantidade total dos dados da amostra teste.

A seguir serão apresentados os detalhes do processo de diagnóstico da mastite subclínica empregando o teste proposto.

2.3. Metodologia usada para obter os resultados

Atualmente a análise de imagem tem adquirido a capacidade de realizar análise rápida e não invasiva a baixo custo em produtos, processos e diagnóstico de doenças de pacientes (FACCO *et al.*, 2011). A área de análise de imagem inclui estudos de imagens em escala de cinza ou imagens RGB, análise de imagens coletadas em poucos canais espectrais ou imagens multiespectrais e mais recentemente a análise de imagens hiperespectrais em que a direção espectral é explorada em toda sua extensão (PRATS-MONTALBÁN; JUAN; FERRER, 2011).

A análise de imagem inicialmente foi aplicada a imagens simples pela comunidade da ciência da computação com o objetivo de detectar ruído, segmentar e classificar objetos na imagem (BERNSTEIN, 1976). A partir de 1980, a comunidade de quimiometria começou a usar à análise de imagens e introduziu ferramentas de análise já em uso na espectroscopia clássica que foram adaptadas para a estrutura de imagens (ESBENSEN; GELADI, 1989). A abordagem quimiométrica para análise de imagem usa imagens simples e imagens hiperespectrais que fornecem informação espacial e espectroscópica (GRAHN; GELADI, 2007). A introdução da análise de imagem na quimiometria criou novas aplicações tais como análise exploratória de imagens, controle estatístico multivariado e regressão multivariada de imagem (BHARATI, 2002).

A análise multivariada de imagens (*Multivariate Image Analysis*, MIA) analisa imagens com várias medidas por pixel (menor ponto da imagem digital) como imagens RGB, multiespectral e hiperespectral (UMEA; F. GRAHN, 2007). MIA é empregada para a classificação, segmentação, detecção de defeitos e predição (YU; MACGREGOR, 2003). Tipicamente MIA analisa todos os espectros de pixels individuais tal como a análise de dados espectroscópicos convencionais.

O processamento Digital de imagens (*Digital Image Processing*, DIP) pode ser considerado uma subárea do processamento digital de sinais (*Digital Signal Processing*, DSP) (BERND JAHNE, 2005; DUCHESNE; LIU; MACGREGOR, 2012). O DSP transforma os sinais analógicos em sinais digitais com o objetivo de possibilitar a análise, o processamento e a armazenagem dos sinais (NALON, 2013).

DIP consiste nas etapas de pré-processamento, segmentação e pós-processamento da imagem. Já a análise digital de imagens (*Digital Image Analysis*, DIA) consiste na extração de

atributos e reconhecimento de padrões e classificação. As etapas típicas da DIP são pré-processamento, segmentação e pós-processamento. As etapas de DIA são extração de atributos e reconhecimento de padrões e classificação (IGNÁCIO, 2013).

2.3.1. Pré-processamento das Imagens Digitais

A imagem obtida na etapa de aquisição de imagem (etapa anterior ao pré-processamento) pode apresentar imperfeições como a presença de ruído ou sinal aleatório, brilho inadequado, caracteres interrompidos ou não adequadamente conectados (GONZALEZ; WOODS, 2010). A etapa de pré-processamento tem como objetivo melhorar a qualidade da imagem para que a DIP e a DIA possam ser efetivas. As operações realizadas na etapa de pré-processamento são conhecidas como operações de baixo nível. São feitas diretamente nos valores de intensidade dos pixels sem conhecimento de quais deles pertencem aos dígitos do objeto de interesse, por exemplo dígitos de leite outras informações presentes na imagem. Em geral, a imagem gerada nessa etapa tem uma qualidade melhor que a imagem original produzida na etapa de aquisição de imagem (MARQUES FILHO; VIEIRA NETO, 1999).

A qualidade desejada de imagem para o teste indireto proposto requer que a imagem represente o objeto e tenha pouca variação em relação à aquisição por meio do escâner usado. Isto é, tenha pouca diferença entre as imagens adquiridas e a média dessas imagens. Desta forma, o pré-processamento se dá com a aquisição de 4 imagens, comparação entre elas e a obtenção de uma imagem média.

São adquiridas 4 imagens de cada amostra e uma filtragem pela média é executada utilizando o Matlab®. Essa média, chamada aqui de M_i é a imagem média originada da média de cada pixel das quatro imagens, equação (2). Desta forma M_i é uma matriz com a média de cada pixel em todas dimensões.

$$M_i = \begin{bmatrix} \bar{p}_{1,1} & \bar{p}_{1,2} & \cdots & \bar{p}_{1,n} \\ \bar{p}_{2,1} & \bar{p}_{2,2} & \cdots & \bar{p}_{2,n} \\ \vdots & \vdots & \ddots & \vdots \\ \bar{p}_{m,1} & \bar{p}_{m,2} & \cdots & \bar{p}_{m,n} \end{bmatrix} \quad (2)$$

em que o pixel médio $\bar{p}_{a,b} = \frac{\sum_{j=1}^4 p_{j;a,b}}{4}$; $a = 1 \dots m$; $b = 1 \dots n$ e j é o índice das imagens (I_j).

Para validação da imagem média foram calculadas as diferenças de cada imagem I_j com M_i . O modo de cálculo executado foi somando as linhas das matrizes e dividindo pelo número de colunas obtendo-se assim um vetor com uma média representativa de cada dimensão de cada imagem, Tabela 2. O mesmo processo foi feito com a média M_i e obteve-se o vetor de média

desta imagem. A diferença dessas duas colunas ($\bar{I}_j - Mi$) é a diferença de cada imagem com a média em cada espaço de cor. A última coluna da tabela mostra a diferença percentual, cujo maior valor é -0,34%.

Tabela 2 – Cálculo de diferença em média de imagens.

Imagem	Dimensão	\bar{I}_j	Mi	$\bar{I}_j - Mi$	$\frac{\bar{I}_j - Mi}{\bar{I}_j} \times 100\%$
I_1	R	79,18	78,99	0,18	0,23%
	G	80,08	79,94	0,14	0,18%
	B	86,13	86,01	0,13	0,15%
I_2	R	78,83	78,99	-0,16	-0,20%
	G	79,80	79,94	-0,15	-0,19%
	B	85,85	86,01	-0,16	-0,18%
I_3	R	78,80	78,99	-0,20	-0,25%
	G	79,76	79,94	-0,19	-0,24%
	B	85,83	86,01	-0,18	-0,21%
I_4	R	78,66	78,99	-0,33	-0,42%
	G	79,64	79,94	-0,30	-0,38%
	B	85,72	86,01	-0,29	-0,34%

A diferença média total calculada em valor absoluto é 0,25% (média da última coluna na Tabela 2 em valores absolutos) o que implica em uma imagem média final muito próxima às imagens capturadas pelo scanner e dessa forma adota-se a imagem média Mi para a entrada de dados de imagem digital das amostras.

2.3.2. Segmentação da Imagem

A segmentação da imagem é o processo em que a imagem é separada na região ou regiões de interesse (*Region of Interest*, ROI) trazendo informações importantes para o processo de classificação (ABDULLAH; HAMBALI; JAMIL, 2012). A imagem é dividida basicamente em dois tipos, o segundo plano, também chamado de *background*, e o primeiro plano, também chamado de *foreground*. O primeiro plano é a parte segmentada que é a região de interesse (MERY; PEDRESCHI, 2005).

Há várias técnicas de segmentação que utiliza detecção de bordas ou baseada em frequência, textura, cores, crescimento de regiões, detecção de linhas ou de pontos e limiarização. As técnicas geralmente não são utilizadas isoladamente (GONZALEZ; WOODS, 2010).

Neste trabalho, após a etapa de aquisição e pré-processamento, é feita a segmentação utilizando a técnica de descritores de contorno (GONZALEZ; WOODS, 2010). O fundamento

desta técnica é baseado na detecção de bordas, que por sua vez é fundamentada na mudança abrupta local de intensidade (SOLOMON; BRECKON, 2013).

As imagens adquiridas pelo escâner da placa *Petri* de cada amostra de leite têm um formato circular e, graças ao fundo preto propositalmente colocado no processo de aquisição, há uma mudança abrupta facilmente detectável na imagem, Figura 4.

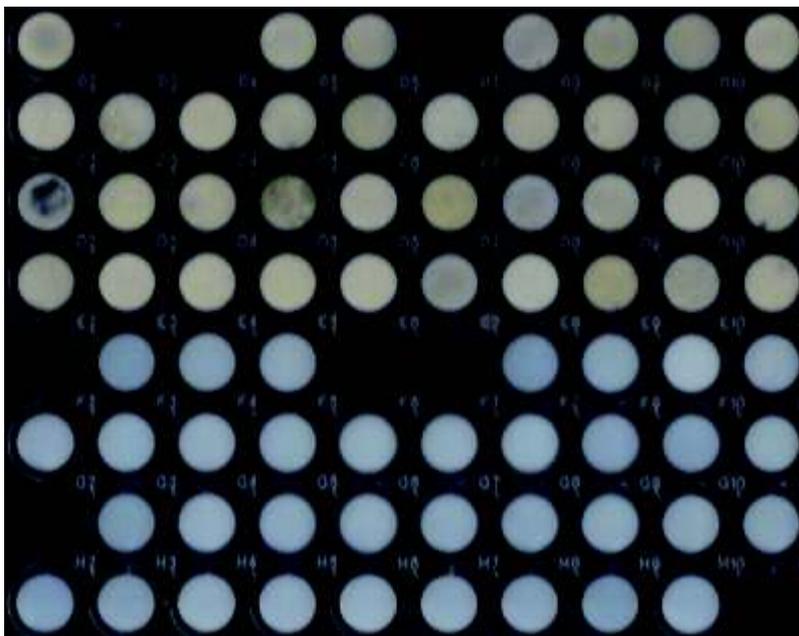


Figura 4 – Imagem escaneada das amostras de leite

A utilização de escala de cinza é uma abordagem comum no processamento de imagens, pois facilita a visualização de mudanças de intensidade na imagem, o que possibilita detectar as bordas com maior precisão. Além disso, diminui a dimensionalidade das matrizes a serem trabalhadas, de cores em RGB para escala de cinza.

A conversão da imagem colorida à escala de cinza, Figura 5(a), poder-se-ia apenas fazer uma média das dimensões R, G e B e desta forma o resultado já seria uma imagem em escala de cinza, pois seria apenas uma dimensão espacial. Porém é sabido que a imagem mais equilibrada para o olho humano é de acordo com a equação (3), em que I_{cinza} é a imagem convertida, I_{cor} é a imagem colorida, os índices n e m representam as posições dos pixels nas matrizes das imagens e r , g e b são os valores no espaço RGB nas respectivas dimensões e os valores dos coeficientes α , β e γ : 0,2989, 0,5870 e 0,1140, respectivamente (SOLOMON; BRECKON, 2013).

$$I_{cinza}(n, m) = \alpha I_{cor}(n, m, r) + \beta I_{cor}(n, m, g) + \gamma I_{cor}(n, m, b) \quad (3)$$

Após a conversão é prático fazer a detecção da borda com uma simples comparação dos valores em escala de cinza e adotar um limite para o início da borda, desta forma a imagem é binarizada Figura 5 (b), ou seja, é transformada de escala e cinza para preto e branco. O objetivo dessa etapa é explicitar os objetos na imagem que não são da ROI. Com a binarização a visualização é facilitada e os cortes podem ser feitos até mesmo manualmente.

Na imagem binarizada é necessário preencher os espaços em branco que não são da ROI. Aqui o formato das imagens a serem segmentadas é conhecido e suas posições são equidistantes, desta forma todas as figuras internas à imagem geral é preenchida ao se verificar que não tem o tamanho, formato e posição adequada, originando a imagem segmentada, Figura 5(c).

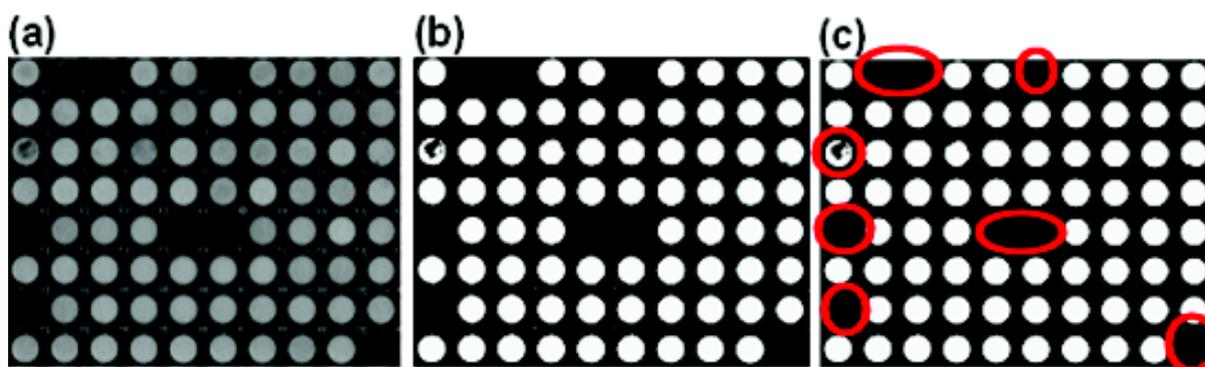


Figura 5 – Preparação para segmentação: (a) imagem média digitalizada; (b) imagem binarizada; (c) máscara para os cortes.

Com a imagem segmentada são feitos os cortes e separação das imagens. Deste processo resultam 80 arquivos com as imagens de cada amostra como pode ser observado na Figura 6. As imagens das amostras perdidas estão em arquivos que foram desconsiderados. Aqui facilmente é visível detectar quais são estas imagens, por estarem pretas parcial ou completamente (destacadas em vermelho na Figura 5 e na Figura 6).

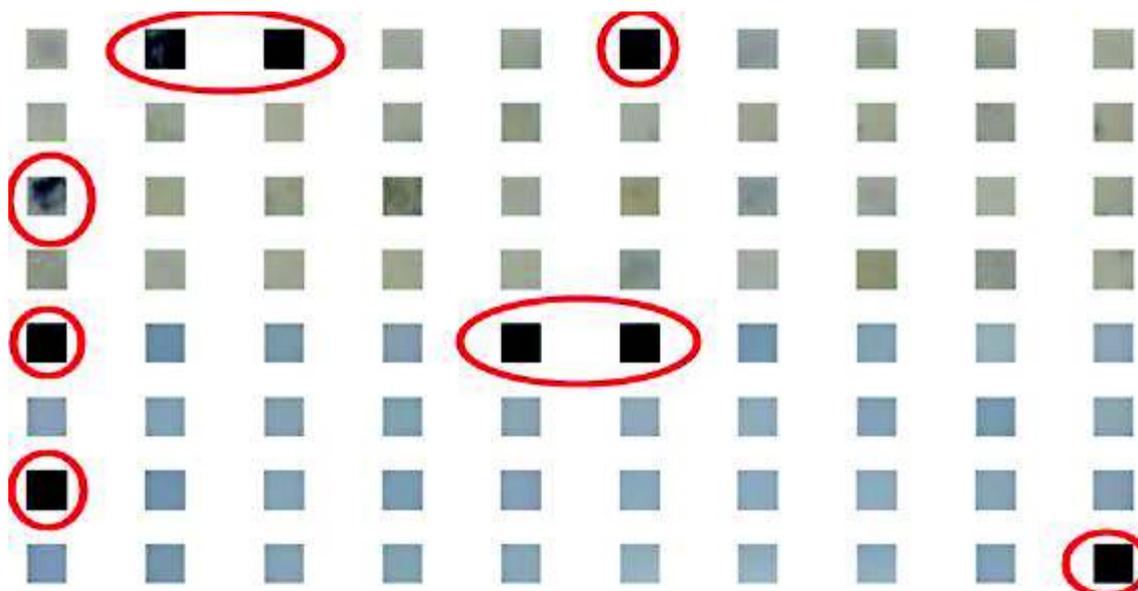


Figura 6 – Imagens segmentadas das amostras de nata (primeiras 40 da esquerda para direita e de cima para baixo) e de leite desnatado.

As imagens segmentadas e divididas em arquivos distintos são utilizadas para a confecção dos histogramas. Não há necessidade de pós-processamento destas imagens, pois os ruídos e saturação foram suavizados durante o pré-processamento quando do cálculo da imagem média.

2.3.3. Histogramas de imagens

O histograma é uma representação gráfica de dados pela frequência relativa destes, ou seja, mostra quantos dados de uma determinada faixa de valores aparece no conjunto total de dados. Um histograma de imagem traz a quantidade de vezes que cada nível de cor repete em uma determinada imagem (GONZALEZ; WOODS; EDDINS, 2009).

Neste trabalho foram obtidos histogramas para os espaços RGB, escala de cinza e HSV, ou seja, dispõe-se de histogramas em 7 dimensões para cada amostra.

O espaço HSV é um espaço de cor perceptual e é utilizado em processos de DPI e DIA e sua utilização se dá por ser considerado o melhor modo com o ser humano percebe as imagens coloridas (WALCZAK, 2001).

Como a aquisição das imagens são no espaço RGB, é necessário converter para o espaço HSV. Os valores das imagens em HSV em suas matrizes correspondentes são baseadas nas coordenadas tridimensionais de um cone. Os valores de H variam de 0 a 360 graus de acordo com a cor em si (matiz), o valor de S varia de 0 a 100 do centro da seção circular do cone para fora indicando a mudança de saturação e o valor de V também varia de 0 a 100 do vértice do

cone até a base indicando a alteração de intensidade (GONZALEZ; WOODS; EDDINS, 2009). Os cálculos da conversão de RGB para HSV são de acordo com as equações (4), (5) e (6) (SOUTO, 2003).

$$H = \begin{cases} 60 x \frac{g - b}{max - min} + 0; se max = r e g \geq b \\ 60 x \frac{g - b}{max - min} + 360; se max = r e g < b \\ 60 x \frac{b - r}{max - min} + 120; se max = g \\ 60 x \frac{r - g}{max - min} + 240; se max = b \end{cases} \quad (4)$$

$$S = \frac{max - min}{max} \quad (5)$$

$$V = max \quad (6)$$

Em que *max* e *min* são os valores máximo e mínimo de R, G e B.

Um conjunto de histograma das imagens de nata pode ser visualizado na Figura 7. Quando em todas as amostras para um determinado canal de cor não havia pelo menos um evento (somatório zero na coluna de variáveis) o canal da cor foi excluído. Isso é feito para minimizar valores zeros redundantes nos dados e melhorar o tempo de execução do programa para o teste proposto. Gráficamente representa apenas um deslocamento no eixo *x* de cada histograma.

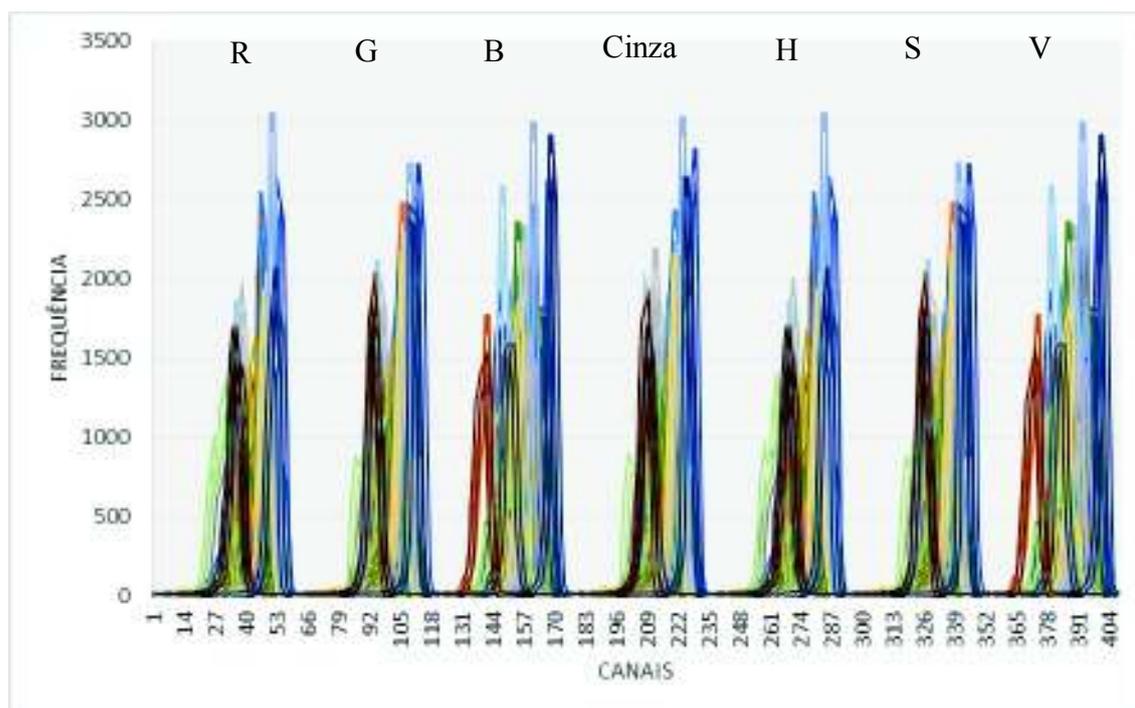


Figura 7 – Histogramas nas dimensões R, G, B, cinza, H, S e V, respectivamente, das amostras de nata. Os valores zeros foram excluídos dos dados.

Os histogramas são formados para nata e leite desnatado formando arquivos com histograma completo com todas dimensões, depois para RGB, Cinza e HSV. São calculados histogramas isolando cada dimensão: R, G, B, H, S e V. Por último, histogramas associando alguns espaços de cores: RGB e cinza, RGB e HSV e, por fim, Cinza e HSV. A Figura 8 mostra o exemplo de como ficou o gráfico do histograma de Cinza e HSV em leite desnatado. O objetivo do cálculo de diferentes histogramas é testar a classificação com cada um destes arranjos e verificar qual deles tem melhor resposta, ou seja, qual deles apresenta um melhor diagnóstico.

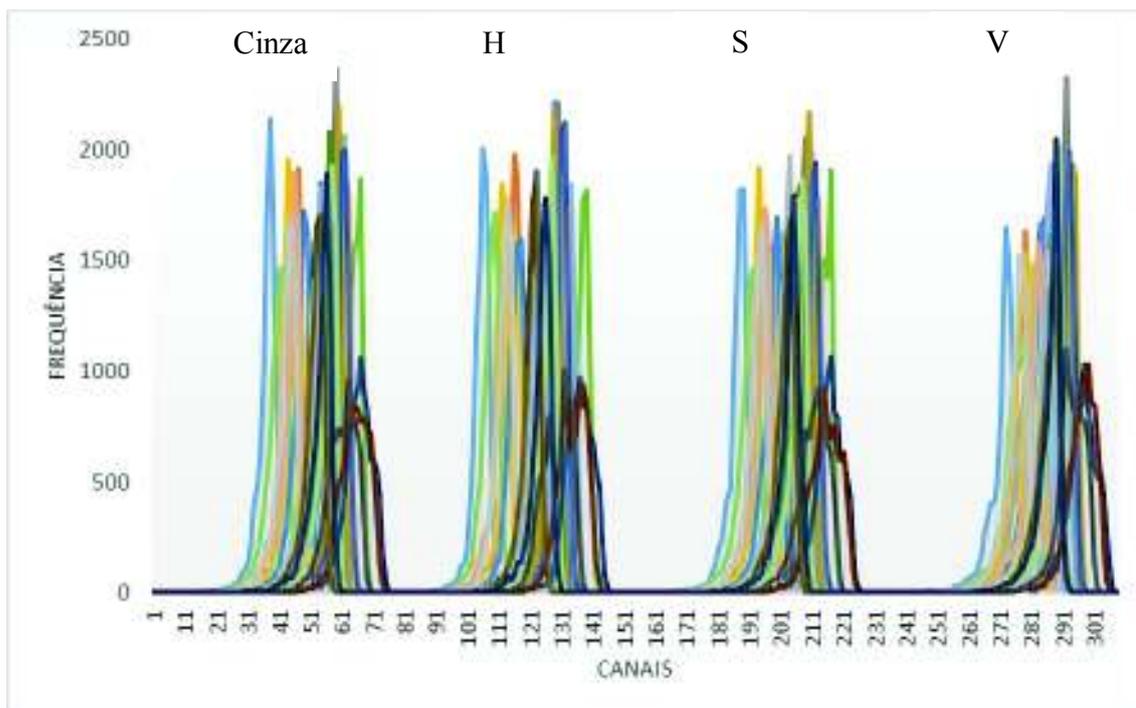


Figura 8 – Histograma dos espaços Cinza e HSV das amostras de leite desnatado. Os valores de frequência zero foram excluídos.

2.3.4. Pré-Processamento de Dados de Inventário Bovino e sinais NMR

Os Dados de Inventário Bovino na etapa *Rotulação* da Figura 3 são pré-processados com a aplicação de técnicas de transformação de dados, raiz quadrada, Equação (7); logaritmo, Equação (8) (ZIMMERMANN, 2004); Z-score, Equação (9) e Min-Max, Equação (10).

$$novo_x = \sqrt{x + k}, k = 0; 0,5 \text{ ou } 1 \quad (7)$$

$$novo_x = \log(x + k), k = 0 \text{ ou } 1 \quad (8)$$

$$novo_x = \frac{x - \bar{x}}{\sigma_x} \quad (9)$$

$$novo_x = \frac{x - \min(x)}{\max(x) - \min(x)} \quad (10)$$

As transformações são utilizadas para melhorar os dados de entrada, já que as variações entre eles não obedecem a critérios visualmente distinguíveis. Pode ser observado na Figura 9 que cada amostra tem uma peculiaridade e não apresenta uma tendência visual para as características (eixo X) do DIB, além das escalas serem muito diferentes. Desta forma as transformações são realizadas na tentativa de minorar os efeitos destas variabilidades e modificar as escalas.

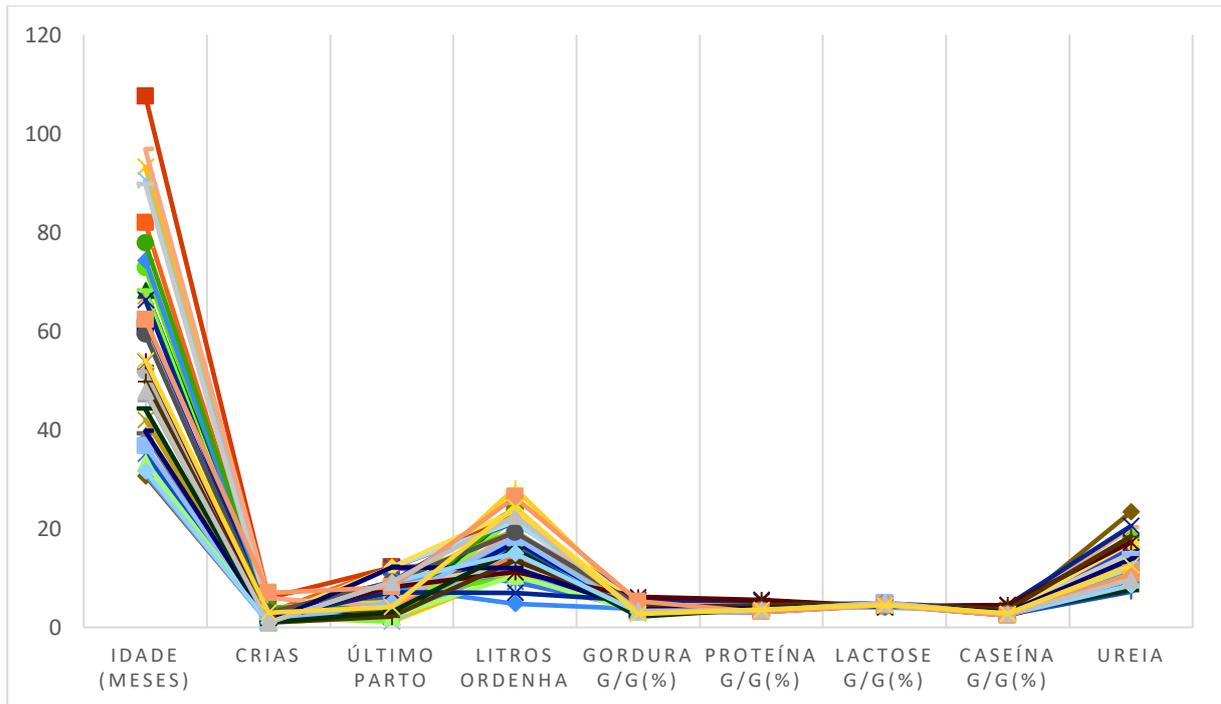


Figura 9 – Gráfico DIB das amostras. As unidades e escalas obedecem a tabela do Apêndice A, colunas (1) a (9). Cada cor é uma amostra.

As Equações (7) e (8) são utilizadas, dentre outras funções, para diminuir a amplitude das escalas das variáveis (ZIMMERMANN, 2004) e aproximar os dados da normalidade (CONAGIN, 1955). A equação (9) centraliza na média e normaliza (WALPOLE *et al.*, 2011), já a Equação (10) normaliza os dados da matriz obtendo novos valores entre 0 e 1.

Os dados utilizados nesse pré-processamento das Equações de (7) a (10) são os do anexo A, das colunas (1) a (9). Apesar dos dados obtidos mostrarem um conjunto de dados em escalas próximas e estarem mais centradas em torno de valores que minimizam a variabilidade, os resultados com a classificação não foram melhores.

O pré-processamento dos sinais NMR utilizaram as mesmas técnicas do item descritas para o DIB e ainda a suavização por Savitsky-Golay (SKOOG; HOLLER; CROUCH, 2007) em que os dados são suavizados por uma regressão polinomial, Equação (11).

$$novo_x = \frac{1}{2m+1} \sum_{j=-m}^m w_j x_{i+j} \quad (11)$$

em que $2m+1$ é a quantidade de pontos em que é feita a média e w_j é o coeficiente obtido na regressão e apresentado na Tabela 3 para regressão polinomial de grau 3 ou 4.

Tabela 3 – Coeficientes w_j do filtro Savitzky-Golay para suavização com base em polinômios de grau 3 e 4. Fonte (CHAU *et al.*, 2004)

Nº de pontos	25	23	21	19	17	15	13	11	9	7	5
$-w_{12}$	-253										
$-w_{11}$	-138	-42									
$-w_{10}$	-33	-21	-171								
$-w_9$	62	-2	-76	-136							
$-w_8$	147	15	9	-51	-21						
$-w_7$	222	30	84	24	-6	-78					
$-w_6$	287	43	149	89	7	-13	-11				
$-w_5$	343	54	204	144	18	42	0	-36			
$-w_4$	387	63	249	189	27	87	9	9	-21		
$-w_3$	422	70	284	224	34	122	16	44	14	-2	
$-w_2$	447	75	309	249	39	147	21	69	39	3	-3
$-w_1$	462	78	324	264	42	162	24	84	54	6	12
w_0	467	79	329	269	43	167	25	89	59	7	17
w_1	462	78	324	264	42	162	24	84	54	6	12
w_2	447	75	309	249	39	147	21	69	39	3	-3
w_3	422	70	284	224	34	122	16	44	14	-2	
w_4	387	63	249	189	27	87	9	9	-21		
w_5	343	54	204	144	18	42	0	-36			
w_6	287	43	149	89	7	-13	-11				
w_7	222	30	84	24	-6	-78					
w_8	147	15	9	-51	-21						
w_9	62	-2	-76	-136							
w_{10}	-33	-21	-171								
w_{11}	-138	-42									
w_{12}	-253										

A aplicação de Savitzky-Golay nos dados de NMR neste trabalho utilizou um polinômio cúbico para até 11 pontos.

Provavelmente os picos existentes nos dados de RMN não permitiram uma suavização que possibilitasse um resultado utilizável na classificação, pois a tendência é distorcer dados reais quando há esses picos mais estreitos (CHAU *et al.*, 2004).

Estas técnicas geralmente resultam em melhores dados de entrada para calibração ou classificação, porém não ocorreu esta melhora quando foram aplicados neste trabalho, por esta razão, os dados utilizados na classificação foram os dados brutos.

2.3.5. Classificação multivariada dos dados de DIB, RMN e imagens digitais

A classificação é uma técnica que investiga as relações dentro de um conjunto de dados para possibilitar que estes sejam definidos por classes em dados com características semelhantes, ou seja atribui regras para identificar objetos a partir de categorias ou classes (GORDON, 1999).

Na construção de classificação com dados multivariados há duas possibilidades: classificação supervisionada ou não supervisionada. Quando se descobre a qual classe pré-existente um determinado subconjunto de dados pertence é a supervisionada e ao agrupar subconjuntos sem uma estrutura pré-existente denomina-se classificação não supervisionada (NAES *et al.*, 2002).

A técnica de classificação multivariada supervisionada é utilizada para construir regras que uma amostra desconhecida possa ser alocada em um subconjunto mais provável, de acordo com as características com as quais essas regras foram construídas. A classificação supervisionada também é conhecida como Análise discriminante (NAES *et al.*, 2002).

Para a análise discriminante é necessário preparar as amostras e dividi-las em subconjuntos, um conjunto de treinamento e um conjunto de validação. O primeiro irá aprender com as características do objeto e definir as regras da classificação e a partir deste modelo definido, os dados de validação testam o modelo e faz o modelo definitivo de classificação. Com o modelo de classificação pronto um novo conjunto pode ser utilizado para a finalidade de classificação, como por exemplo, verificar se uma vaca está com mastite a partir do modelo pronto com os dados treinados e validados de características do leite.

Problemas de colinearidade e sobreposição podem atrapalhar a elaboração dos modelos de classificação. O primeiro traz informações redundantes nos dados e pode gerar um modelo tendencioso e o segundo não discrimina corretamente quais dados pertencem aos subconjuntos que irão preparar o modelo ou regra de classificação.

Há várias técnicas de classificação, como algoritmo genético (*Genetic Algorithm, GA*) (MELANIE, 1999), SIMCA (*Soft Independent Modeling of Class Analogies*) (WOLD, 1976), KNN (*K Nearest Neighbour*) (CHEN *et al.*, 2010).

Para o procedimento de classificação de objetos no teste proposto são usadas as técnicas de SPA-LDA e PLS-DA. A LDA é sensível a colinearidade e não trabalha com um número elevado de variáveis independentes. Como a SPA faz a seleção de variáveis, reduzindo a

dimensionalidade da matriz de dados e ainda trata os dados a partir da ortonormalização, resulta em dados com baixa colinearidade, dessa forma, a técnica em conjunto da SPA com a LDA se apresenta interessante para a classificação dos dados, além disso, a quantidade de amostra não é alta e a LDA é apropriada para esse caso (SOARES *et al.*, 2013). A PLS-DA também pode trabalhar com poucas amostras e trabalha com componentes principais ortonormalizadas, reduzindo a colinearidade (BALLABIO; CONSONNI, 2013), desta forma optou-se por utilizar esta técnica para comparar qual será a mais eficiente para os testes de diagnóstico de mastite subclínica neste trabalho.

2.3.5.1. SPA-LDA

A SPA originalmente foi desenvolvida para seleção de variáveis e sua utilização em conjunto com regressão linear múltipla (*Multiple Linear Regression*, MLR) para calibração multivariada é basicamente a escolha pela SPA do melhor conjunto de variáveis com menor colinearidade através do menor erro médio quadrado de validação (*Root Mean Square Error of Validation*, RMSEV) da MLR (ARAÚJO *et al.*, 2001). A MLR é uma técnica de calibração multivariada em que um modelo é adquirido a partir de uma equação que traz uma relação entre uma variável dependente e variáveis independentes. O RMSEV é raiz quadrada do erro médio quadrático de validação e é calculado conforme a equação (12).

$$RMSEV = \sqrt{\frac{1}{N_{val}} \sum_{i=1}^{N_{val}} (y_{val,i} - \hat{y}_{val,i})^2} \quad (12)$$

em que $y_{val,i}$ e $\hat{y}_{val,i}$ são os valores de referência e de predição da propriedade y de interesse na amostra de validação (SOARES *et al.*, 2013).

A SPA tem como característica a redução da dimensionalidade das variáveis e da colinearidade. A LDA trabalha com número de amostras menor que o número de variáveis e baixa colinearidade, esta junção se apresenta apropriada para uma classificação multivariada (MILANEZ; PONTES, 2015).

Para a classificação supervisionada, a validação é feita verificando os dados de entrada da matriz e os índices de classes. Alterou-se, dessa forma, a função de custo de RMSEV para o risco médio de classificação errada (T), que está descrito na Equação 13 (PONTES *et al.*, 2005).

$$T = \frac{1}{k} \sum_{k=1}^{k_v} t_k \quad (13)$$

em que t_k é o risco de se classificar incorretamente um objeto da k -ésima amostra de validação e seu valor é calculado pela Equação 14.

$$t_k = \frac{r^2(x_k, \mu_{lk})}{\min_{l_j \neq j} r^2(x_k, \mu_{lk})} \quad (14)$$

em que x_k é o objeto da amostra e o numerador é o quadrado da distância de *Mahalanobis* entre esse objeto e a média de sua classe. Já o denominador é o quadrado da distância de *Mahalanobis* entre o objeto e o centro da classe errada mais próxima.

Quanto menor o t_k , melhor, pois dessa forma o objeto estará mais próximo da verdadeira classe a que pertence e distante das demais.

Com a separação das amostras em conjuntos de treinamento, validação e teste e se estas tiverem o mesmo número de variáveis, pode-se produzir o processo de seleção a partir das entradas necessárias (PONTES, 2009): matrizes correspondentes às repostas instrumentais (conjunto de treinamento, conjunto de validação, conjunto externo para teste); índices das classes (conjunto de treinamento, conjunto de validação, conjunto externo para teste) e os números mínimo e máximo de variáveis a serem selecionados (N1 e N2).

O algoritmo de classificação com SPA-LDA está bem descrito na literatura (PONTES *et al.*, 2005; SOARES *et al.*, 2013) e os dados de saída do software utilizado são um vetor contendo a melhor cadeia de variáveis; o risco médio T (definido na equação 14) e os índices das classes previstos pela LDA. Utilizando estes valores é possível verificar a taxa de classificação correta (TCC) definida na equação 15, em que as classes de teste são comparadas com as previstas.

$$TCC(\%) = \frac{\text{Número de acertos de classificação}}{\text{Número de amostras do conjunto teste}} \times 100 \quad (15)$$

Para classificação foram utilizados os dados obtidos pela ressonância magnética nuclear nas amostras de leite na forma gordurosa (nata) e na forma aquosa (leite desnatado) e as imagens digitais com seus dados já convertidos em histogramas e os DIB. Todos conjuntos de dados foram classificados quanto a suspeita de ocorrência de mastite e para tal utilizaram-se os dados da tabela do apêndice A para as classes de suspeita de ocorrência da doença. As classes foram

divididas em positivas, negativas e indeterminadas para suspeita de mastite de acordo com a contagem de células somáticas (SCC). Os valores limites da SCC foram adotados neste trabalho conforme Tabela 4Tabela 1.

Tabela 4 – Classes de suspeita de ocorrência de mastite de acordo com a SCC

SC/mL x 1.000	Status
<200	negativo
>=200 e <300	indeterminado
>=300	positivo

Com os mesmos dados relativos aos histogramas, ressonâncias magnéticas e DIB compôs-se classificações quanto ao diagnóstico de mastite subclínica utilizando-se o algoritmo das projeções sucessivas em conjunto com a análise discriminante linear (SPA-LDA). Com os resultados das classificações atingidos, procedeu-se a elaboração de tabelas comparativas entre as duas técnicas de classificação de acordo com as taxas de classificação corretas (TCC).

2.3.5.2. PLS-DA

Mínimos quadrados parciais (*Partial Least Square*, PLS) é uma técnica originalmente de calibração multivariada, que permite obter uma resposta matemática relativamente às propriedades de interesse do analito estudado de acordo com resultados apresentados instrumentalmente. Essa técnica permite predizer, de acordo com o modelo matemático desenvolvido, alguma propriedade em amostras diferentes das utilizadas na construção do modelo (WESTAD; BEVILACQUA; MARINI, 2013).

A calibração multivariada consiste em construir o modelo, validar e fazer a previsão. A validação é o processo de verificar se o modelo fará a previsão das propriedades de interesse corretamente e pode ser realizada com dados de amostras de validação ou em validação cruzada a partir do próprio conjunto de amostras da calibração (FERREIRA *et al.*, 1999).

Dependendo das características dos dados e dos objetivos da calibração, utiliza-se uma técnica apropriada, como por exemplo a PLS, que é uma técnica de regressão que minimiza efeitos relativamente a ruídos experimentais, colinearidades e não linearidade (LORBER; WANGEN; KOWALSKI, 1987; WOLD, 1985). Os ruídos são problemas com dados obtidos instrumentalmente que afetam as medidas reais. A colinearidade, por sua vez, é a correlação entre variáveis, que pode trazer informações redundantes. Não linearidade, como o próprio nome diz, é a dificuldade de se obter uma informação linear com as respostas instrumentais (variáveis independentes) (HAIR *et al.*, 2009).

Na PLS, as variáveis originais são substituídas por variáveis latentes ou fatores obtidos de uma combinação linear entre as variáveis dependentes e variáveis independentes. Uma maior variância das amostras diminui a possibilidade de colinearidade e a primeira variável latente traz essa informação, já as demais variáveis são obtidas da ortogonalização relativamente à primeira e assim por diante até o limite de variáveis necessárias, que será calculado pela validação cruzada (WESTAD; BEVILACQUA; MARINI, 2013).

Para classificação com análise multivariada a PLS é utilizada em conjunto com análise discriminante (*Discriminant Analysis, DA*), é a PLS-DA. O princípio desta técnica é a utilização da regressão PLS para construir um modelo de classificação relacionando linearmente as variáveis X independentes (resposta instrumental) com as variáveis y dependentes (propriedades de interesse). Porém, para classificação utiliza-se um rótulo para as classes com valores binários 0 ou 1. Para mais de duas classes é utilizado o algoritmo PLS2 (*Partial Least Square 2*) de modo a executar uma modelagem através de componentes principais para obter as variáveis latentes em número calculado pela validação cruzada. Para especificar as classes é escolhido o ponto de menor taxa de erro de classificação (BALLABIO; CONSONNI, 2013; WESTAD; BEVILACQUA; MARINI, 2013).

Exemplos de aplicação da PLS-DA em quimiometria por imagens são alguns relacionados no item 2.3.4, como a classificação de azeite de oliva (MILANEZ; PONTES, 2015); de biodiesel (COSTA *et al.*, 2015); de mel argentino (DOMÍNGUEZ *et al.*, 2014); de café (SOUTO, U. T. DE C. P. *et al.*, 2015) e de bactérias (ALMEIDA *et al.*, 2014).

Neste trabalho foi efetuado o teste de mastite com três fontes de dados, dados de inventário bovino (DIB) contidos no apêndice A, outro com os histogramas e um último com dados de ressonância magnética. Desta forma pode-se avaliar a acurácia e o potencial de cada um no diagnóstico da doença. Os resultados foram comparados utilizando a TCC.

2.4. Diagnóstico de Mastite através da Classificação multivariada

A Figura 3 mostra os módulos do protótipo de software proposto a realizar o diagnóstico da mastite subclínica em que se tem a classificação multivariada como ponto onde há a definição do diagnóstico. O teste diagnóstico identifica através de dados obtidos em laboratório, imagem e inventário levantado in loco na ordenha, de acordo com o nível de células somáticas, o diagnóstico para mastite subclínica em positivo, negativo ou indeterminado para mastite. A classificação multivariada relaciona estes dados através dos algoritmos SPA-LDA e PLS-DA.

Os dados utilizados na classificação multivariada são os dados brutos de NMR e DIB, pois mesmo com o pré-processamento não houve melhoria nos resultados da classificação. No caso dos dados das imagens, foram utilizados os valores obtidos na aquisição dos histogramas.

2.5. Softwares utilizados

Os *softwares* empregados nessa dissertação foram desenvolvidos no Matlab®. Para a classificação foi utilizado o programa desenvolvido por pesquisadores do Instituto Tecnológico da Aeronáutica e Universidade Federal da Paraíba (GALVÃO, 2008). Já a PLS-DA foi executada com um software desenvolvido na Universidade de Milão (BALLABIO e CONSONNI, 2013).

Para aquisição, pré-processamento e tratamento de imagens, como explicado nos itens 2.3.1, 2.3.2 e 2.3.3., foram desenvolvidos programas em Matlab®.

3. Resultados e Discussões

Tendo em vista o escopo da pesquisa, divide-se este capítulo em algumas subseções, nas quais são apresentados e discutidos os resultados obtidos relativamente à classificação em função do diagnóstico de mastite subclínica.

3.1. Classificação Multivariada

Com o intuito de verificar a ferramenta de teste auxiliar para o diagnóstico da mastite subclínica desenvolvida a partir da classificação do leite, fez-se a comparação das técnicas de SPA-LDA e PLS-DA com o uso de imagem (*Multivariate Image Analysis*, MIA), sinais de NMR e dados de inventário bovino (DIB).

Primeiramente foram executados os algoritmos com os dados de inventário bovino e SCC em leite cru. A Tabela 5 traz esses resultados. Os dados de inventário são aquelas constantes do apêndice A: idade da vaca, número de crias, último parto, litros de ordenha, teor de gordura, teor de proteína, teor de lactose, teor de caseína, teor de ureia e SCC.

Tabela 5 – Dados comparativos de taxa de classificação correta (TCC) para validação e teste quanto a suspeita da mastite subclínica em amostras de leite cru relativo aos DIB.

Classe	1 - Negativo		2 - Positivo		3 - Indeterminado		Todas classes	
	PLS-DA	APS-LDA	PLS-DA	APS-LDA	PLS-DA	APS-LDA	PLS-DA	APS-LDA
Validação	42%	75%	33%	0%	0%	0%	37%	60%
Teste	50%	75%	0%	0%	0%	0%	38%	60%

Observando os dados da Tabela 5 é possível constatar que a SPA-LDA tem um melhor desempenho que a PLS-DA na classificação total, tanto na validação com 60% quanto nas amostras teste com 60% de classificação correta. Os negativos puderam ser classificados em 75% na SPA-LDA contra 50% na PLS-DA, já no positivo apenas a SPA-LDA conseguiu fazer 33% de classificação correta enquanto não houve qualquer classificação correta no indeterminado.

Ao executar o programa para a PLS-DA, gráficos de número de amostras pela resposta por classe e o do coeficiente de classe pela variável latente foram adquiridos e estão na Figura 10.

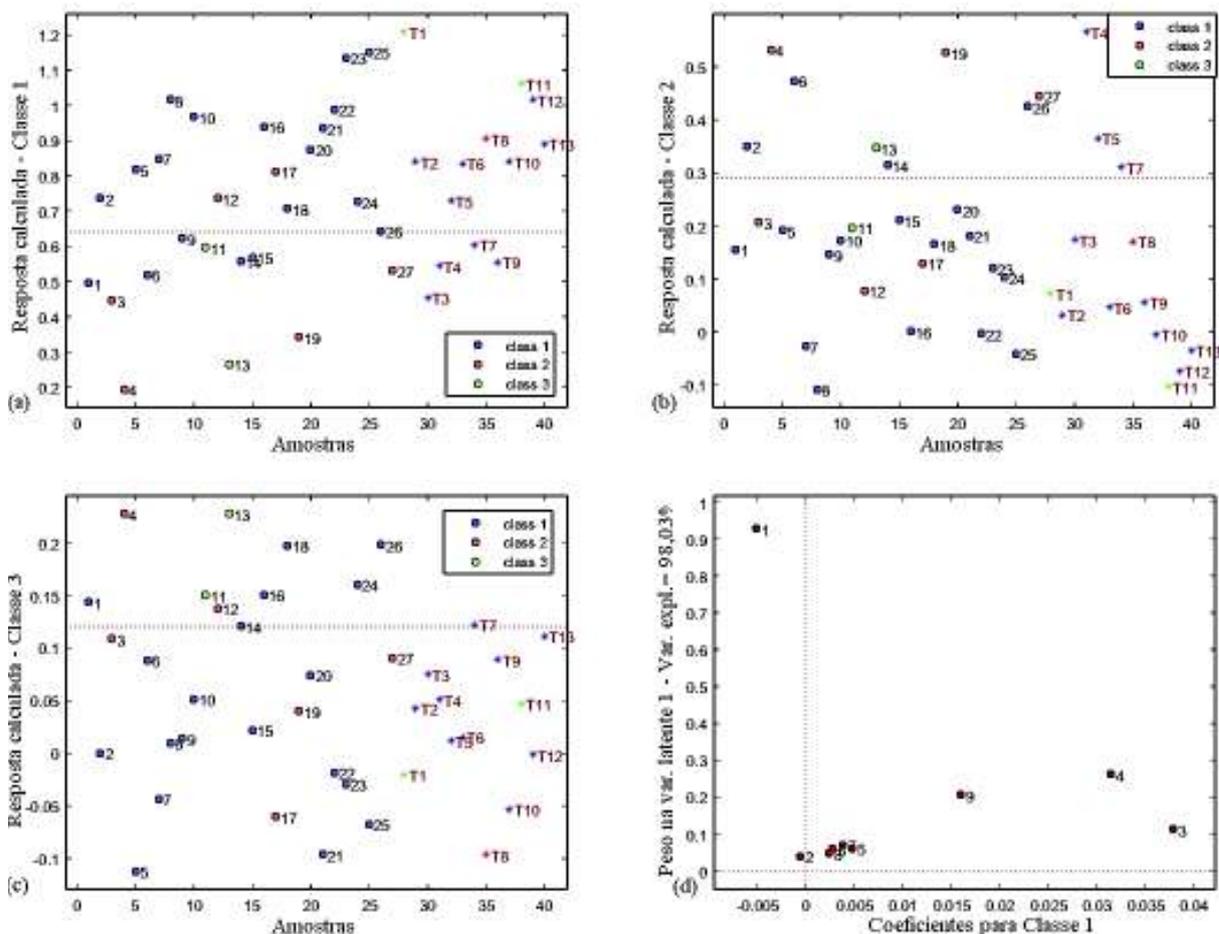


Figura 10 – Gráficos PLS-DA: Número de amostras X Resposta por classe (a) Classe 1- negativa, (b) Classe 2 - positiva, (c) Classe 3 - indeterminada e o (d) Coeficiente de classe X peso (*weight*) da variável latente 1. Os valores de 1 a 26 são referentes à validação e os valores de T1 a T13 são os valores de teste da classificação.

Nos gráficos (a), (b) e (c) da Figura 10 é viável atentar que o modelo realmente tem uma capacidade fraca de classificação, pois há amostras de outra classe acima do limiar calculado (falso positivo classificado ou não atribuído) como por exemplo as amostras 12 e 17 na Figura 10 (a), 4, 13, 19 e 27 na Figura 10 (b) e 4, 11, 12 e 13 na Figura 10 (c). Há também vários valores abaixo da classe a ser classificada (falso negativo ou não atribuído) com as amostras 1, 6, 9, 14 e 15 na Figura 10 (a), as amostras 3, 12 e 17 na Figura 10 (b). A Tabela 6 exibe as amostras classificadas na validação e é observável que as amostras apontadas fora de sua classe nos respectivos gráficos não foram corretamente classificadas.

Também pode ser analisado, pelos gráficos, que as amostras testes, cuja numeração se inicia com T, também não se apresentaram com boa classificação. As cores dos asteriscos indicam as classes a que pertencem e o limiar em cada gráfico continua mostrando quais amostras estão classificadas corretamente.

Tabela 6 – Validação da classificação utilizando PLS-DA em amostras de leite e DIB

Nº da Amostra	Amostra validação	Classificação	Acertos			
			Total	Classe 1	Classe 2	Classe 3
1	1	3	0	0	0	0
2	1	0	0	0	0	0
3	2	0	0	0	0	0
4	2	0	0	0	0	0
5	1	1	1	1	0	0
6	1	2	0	0	0	0
7	1	1	1	1	0	0
8	1	1	1	1	0	0
9	1	0	0	0	0	0
10	1	0	0	0	0	0
11	3	1	0	0	0	0
12	2	0	0	0	0	0
13	3	0	0	0	0	0
14	1	0	0	0	0	0
15	1	0	0	0	0	0
16	1	0	0	0	0	0
17	2	1	0	0	0	0
18	1	0	0	0	0	0
19	2	2	1	0	1	0
20	1	1	1	1	0	0
21	1	1	1	1	0	0
22	1	1	1	1	0	0
23	1	1	1	1	0	0
24	1	0	0	0	0	0
25	1	1	1	1	0	0
26	1	0	0	0	0	0
27	2	2	1	0	1	0
Total	27	14	10	8	2	0
TCC			37%	42%	33%	0%

A tabela de confusão (Tabela 7) mostra a quantidade de acertos por classe, a quantidade de amostras não atribuídas a qualquer classe, mostra as classificações em classes distintas e as não classificações nas classes corretas. Se observar a Figura 10 em conjunto com a tabela é possível verificar as duas classes 3 classificadas como classe 1, assim como uma classe 2 e observar que há quatro amostras classe 1 não atribuídas corretamente, assim como é possível constatar que não há qualquer acerto nas classificações das classes 2 e 3.

Tabela 7 – Tabela de confusão das amostras teste – classificação PLS-DA para DIB

Real/ predito	Classe 1	Classe 2	Classe 3	Não atribuído	Acerto	Total atribuído	Total
Classe 1	5	1	0	4	5	6	10
Classe 2	1	0	0	0	0	1	1
Classe 3	2	0	0	0	0	2	2
Total	8	1	0	4	5	9	13

Agora, analisando a Figura 10 (d), constata-se que as variáveis mais influentes para a PLS-DA são as variáveis 3, 4 e 9, correspondentes à data do último parto, litros de leite ordenhado e teor de ureia no leite. Estes resultados isolados não trazem significado importante à pesquisa, até porque, a classificação foi de 38% de acerto. A classificação utilizando a SPA-LDA poderia ser discutida comparando-se os resultados, porém com 60% de assertividade, ainda há a necessidade de buscar um melhor resultado.

As variáveis selecionadas pela SPA-LDA foram 7: 1, 3, 4, 5, 6, 7 e 9 e verificando no Apêndice A têm as seguintes variáveis: idade, data do parto, litros de ordenha, teor de gordura, teor de proteína, teor de lactose e teor de ureia. A Tabela 8, que traz os resultados da classificação de acordo com as amostras em função das variáveis selecionadas, mostra que não se tem qualquer acerto em suspeita positiva ou indeterminada, além de haver algumas classificações em categorias erradas. Além disso o valor final da acurácia do teste é de 60%, com 75% para negativo e 0% para os demais, o que traz uma dificuldade em se relacionar efetivamente os atributos dos dados de inventário (DIB) com o diagnóstico da mastite.

Tabela 8 – Classificação das amostras de teste com SPA-LDA e DIB

Nº da Amostra	Amostra Teste	Classificação	Acertos			
			Total	Classe 1	Classe 2	Classe 3
T1	2	1	0	0	0	0
T2	1	1	1	1	0	0
T3	2	1	0	0	0	0
T4	1	1	1	1	0	0
T5	1	2	0	0	0	0
T6	1	1	1	1	0	0
T7	1	1	1	1	0	0
T8	2	3	0	0	0	0
T9	1	1	1	1	0	0
T10	1	1	1	1	0	0
Total	10	10	6	6	0	0
TCC			60%	75%	0%	0%

3.1.1. Classificação Multivariada aplicada em MIA

A classificação com o uso de Quimiometria por imagem foi implementada nas amostras de nata e leite desnatado utilizando a PLS-DA e SPA-LDA em função dos histogramas de cores como variáveis e as classes ficaram divididas em negativas (1), positivas (2) e indeterminadas (3). A Tabela 9 traz os resultados dessa construção.

A primeira observação pertinente é que a PLS-DA tem uma performance melhor que a SPA-LDA, relativamente à TCC.

Os gráficos da Figura 11 mostram, pelo custo de validação, que foram escolhidas 16 variáveis para todos os espaços, exceto para o completo (RGB-Cinza-HSV) que teve 12 variáveis escolhidas e o espaço B que teve 15 variáveis.

Por conta da escolha de variáveis implementada pelo KS (*kennard Stone*) e pelo pequeno número de amostras, o algoritmo escolheu subgrupos com apenas duas classes e isso pode ser verificado na Tabela 9, na qual a coluna de indeterminado não pode ser calculada por não haver qualquer amostra naquela classe.

Tabela 9 – Resumo das TCC nos modelos em PLS-DA e SPA-LDA para amostras de NATA

Componentes de cores	Negativo (1)		Positivo (2)		Indeterm. (3)		Todas classes	
	PLS-DA	SPA-LDA	PLS-DA	SPA-LDA	PLS-DA	SPA-LDA	PLS-DA	SPA-LDA
RGB-Cinza-HSV	88%	50%	0%	50%	NC	NC	64%	50%
RGB	75%	67%	0%	0%	NC	NC	55%	50%
R	50%	67%	33%	50%	NC	NC	45%	63%
G	63%	33%	33%	50%	NC	NC	55%	38%
B	63%	17%	0%	50%	NC	NC	45%	25%
Cinza	50%	50%	33%	50%	NC	NC	45%	50%
HSV	75%	67%	0%	0%	NC	NC	55%	50%
H	50%	33%	33%	50%	NC	NC	45%	38%
S	63%	67%	33%	0%	NC	NC	55%	50%
V	63%	50%	0%	50%	NC	NC	45%	50%
RGB-Cinza	88%	50%	0%	50%	NC	NC	64%	50%
RGB-HSV	75%	67%	0%	50%	NC	NC	55%	63%
Cinza-HSV	88%	50%	0%	50%	NC	NC	64%	50%

NC - Não calculado, o denominador (total de amostras da classe atribuída no teste) é zero

Como o objetivo é verificar se há assertividade na escolha das classes, principalmente de positivos e negativos, as duas técnicas não foram muito bem-sucedidas. O que se pode observar é que nos espaços RGB-Cinza-HSV, RGB-Cinza e RGB-HSV houve uma classificação com TCC em negativo de 88% na SPA-LDA, porém nenhum deles conseguiu prever qualquer positivo. Na somatória de todas as classes, esses três grupos classificaram 64% das amostras corretamente.

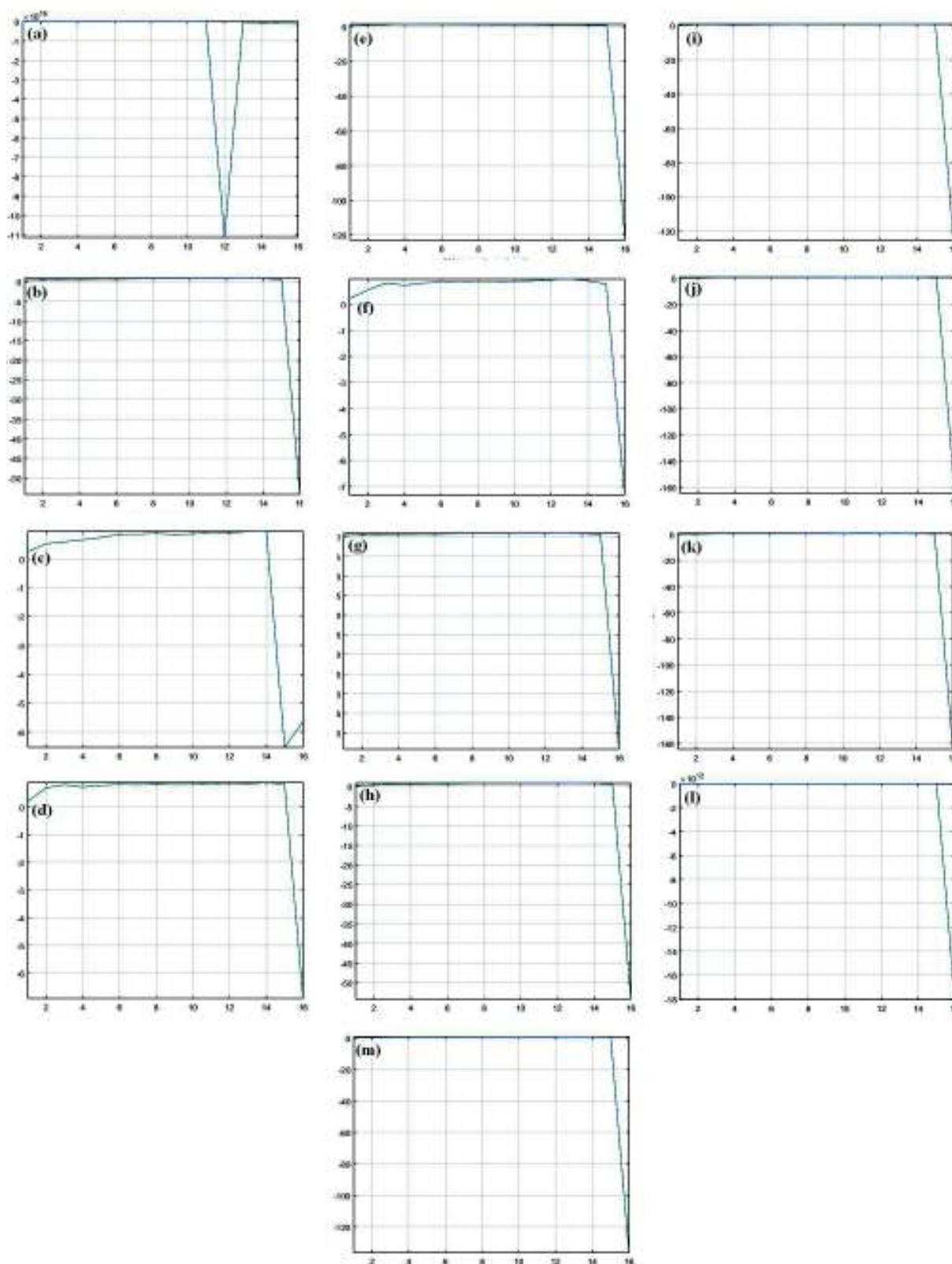


Figura 11 – Gráfico número de variáveis X custo de validação nos espaços de cores (a) RGB-Cinza-HSV, (b) RGB, (c) R, (d) G, (e) B, (f) Cinza, (g) HSV, (h) H, (i) S, (j) V, (k) RGB-Cinza, (l) RGB-HSV, (m) Cinza-HSV.

Pode-se dizer que a classificação em nata é mais precisa em um espectro mais amplo de espaços de cores ao utilizar a SPA-LDA. Os piores resultados foram adquiridos em dimensões isoladas, como R, B, Cinza, H e V, o que reforça o pensamento sobre a necessidade de se utilizarem mais dimensões para uma boa classificação de leite em nata.

Já em amostras de leite desnatado, o comportamento é o oposto ao observado na classificação em nata, como pode ser verificado na Tabela 10, cujos melhores resultados, também auferidos na SPA-LDA foram encontrados nas dimensões únicas de R, Cinza e H.

Tabela 10 - Resumo das TCC nos modelos em PLS-DA e SPA-LDA para amostras de LEITE DESNATATO

Componentes de cores	Insuspeito		Suspeito		Indeterminado		Todas classes	
	PLS-DA	SPA-LDA	PLS-DA	SPA-LDA	PLS-DA	SPA-LDA	PLS-DA	SPA-LDA
RGB-Cinza-HSV	63%	67%	33%	50%	0%	100%	50%	67%
RGB	50%	50%	0%	50%	0%	0%	33%	44%
R	88%	67%	67%	0%	0%	0%	75%	44%
G	50%	33%	33%	0%	0%	0%	42%	22%
B	63%	33%	0%	0%	0%	0%	42%	22%
Cinza	100%	83%	33%	0%	0%	0%	75%	56%
HSV	50%	50%	0%	50%	0%	0%	33%	44%
H	88%	67%	67%	0%	0%	0%	75%	44%
S	50%	33%	33%	0%	0%	0%	42%	22%
V	63%	33%	0%	0%	0%	0%	42%	22%
RGB-Cinza	63%	33%	33%	100%	0%	0%	50%	44%
RGB-HSV	50%	50%	0%	50%	0%	0%	33%	44%
Cinza-HSV	63%	33%	33%	100%	0%	0%	50%	44%

O leite desnatado tem menos componentes químicos e sua coloração é mais homogênea, permitindo inferir de melhor maneira sobre a suspeita em relação à mastite em dimensões únicas, que demonstram a cor (vermelha), matiz e intensidade de cor (cinza) na amostra.

As imagens presentes na Figura 12 dão uma ideia de como SPA-LDA identifica as classes, pois as imagens mais amareladas têm em si uma diferença de intensidade e de coloração, já as imagens das amostras desnatadas visualmente se particularizam pela intensidade (escala de cinza) das cores. As imagens pretas são as amostras perdidas e a numeração é da identificação das amostras.



Figura 12 – Algumas imagens de amostras de leite na forma de nata e leite desnatado filtradas com maior intensidade para amplificar diferenças.

3.1.2. Classificação Multivariada aplicada em NMR

Para a classificação multivariada com a utilização dos dados obtidos pela ressonância magnética (NMR) também foram aplicadas as técnicas de SPA-LDA e PLS-DA. Foram aplicadas em amostras de nata e leite desnatado e os resultados podem ser conferidos na Tabela 11.

Tabela 11 - Resumo das TCC nos modelos em PLS-DA e SPA-LDA aplicadas em ressonância magnética nuclear (NMR) nas amostras de NATA e LEITE DESNATADO

NMR	Negativo (1)		Positivo (2)		Indeterm. (3)		Todas classes	
	PLS-DA	SPA-LDA	PLS-DA	SPA-LDA	PLS-DA	SPA-LDA	PLS-DA	SPA-LDA
Desnatado	100%	71%	0%	0%	0%	0%	78%	56%
Nata	38%	63%	0%	0%	0%	NC	30%	50%

NC - Não calculado, pois o denominador (total de amostras da classe atribuído no teste) é zero

A PLS-DA, para NMR, apresentou desempenho mais satisfatório comparativamente com a SPA-LDA para as amostras de leite desnatado, com um acerto de 100% das negativas e 78% do total. O número de variáveis muito grande pode ser a razão para o melhor desempenho da PLS-DA, pois ela faz a transformação das variáveis em componentes trazendo as características mais importantes consigo. A SPA-LDA precisaria de mais amostras para equilibrar equiparar em precisão a PLS-DA, pois há apenas 40 amostras de nata e 39 de leite desnatado e as amostras para classes indeterminadas e negativas são poucas, o que dificulta ainda mais o trabalho da classificação, basta conferir na Tabela 11 que não houve qualquer

classificação nessas classes. Estando a força de ambas as técnicas nos subconjuntos de maior representatividade de classes, tanto nata quanto leite desnatado.

Entretanto, em relação à nata, a SPA-LDA teve seu desempenho configurado como melhor comparando com o leite desnatado. Porém, o melhor desempenho da SPA-LDA não é assim tão significativo, pois a taxa de acerto foi de apenas 50% e, novamente, não houve qualquer acerto na identificação das amostras positivas e indeterminadas.

Se forem associadas a Tabela 5, a Tabela 9, a Tabela 10 e a Tabela 11, os resultados apontados nas classes suspeitas e indeterminadas corroboram a hipótese de necessidade de um maior número de amostras nas classes positivas e indeterminadas. Obviamente que, nestes casos, se espera que haja menos casos com mastite e gado em estado indeterminado, pois a expectativa e o trabalho realizado nas fazendas é para que o gado esteja sadio. Porém, para melhorar a acurácia das técnicas é necessário que haja uma maior representatividade em todas as classes e isso é mais fácil de se obter com um número maior de amostras como um todo.

Para corroborar a ideia da necessidade de mais amostras é possível visualizar na Figura 13 (a) que a maioria das amostras da classe 1 se colocam acima da linha tracejada, porém algumas amostras classe 2 se colocam neste mesmo setor e outras abaixo. Se fossem mais amostras, apenas duas acima do tracejado não iriam interferir no modelo, porém são apenas seis amostras classe 2 no total.

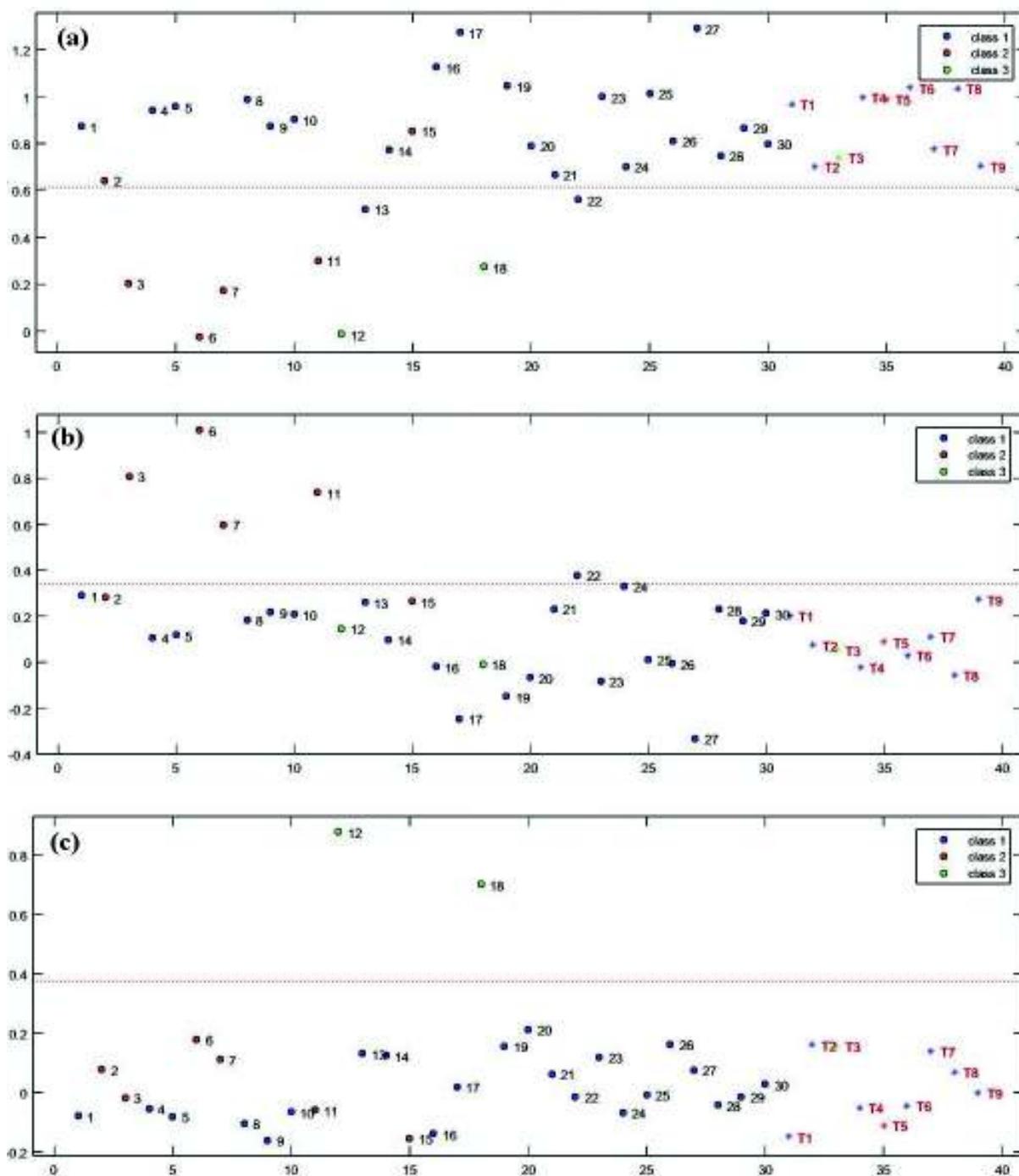


Figura 13 – Gráficos Amostra X resposta calculada nas classes (a) 1-suspeita, (b) 2-insuspeita e (c) 3-indeterminada. Numeração de 1 a 30 são as variáveis, sendo que as T1 a T9 são as variáveis de teste. Técnica PLS-DA para NATA.

Os modelos de classe 1 e classe 3, entretanto conseguem um nível de classificação correta na validação, porém nem há amostra suficiente para ser classificada em amostra de classe 3 e verificar a acurácia do teste, que não classifica classe 2 corretamente e classifica apenas classe 1. Desta forma o teste só diz quando é negativo em suspeita e não consegue determinar se é positivo ou indeterminado, o que traz mais indícios sobre a necessidade de

aumentar o número de amostras, principalmente nas classes 2 e 3, para melhorarem os resultados.

4. CONCLUSÕES

O presente trabalho desenvolveu testes que classificaram amostras de leite cru para diagnóstico de mastite subclínica com taxa de classificação correta de 60% utilizando a SPA-LDA e 38% com a PLS-DA, tendo como variáveis classificadoras os DIB. Os resultados são entendidos como apenas razoáveis para SPA-LDA e não satisfatórios para PLS-DA, o que sugere a necessidade de utilização de outras técnicas ou a repetição das mesmas com um número maior de amostras e que tenham mais variáveis nas classes negativas e indeterminadas.

Na classificação utilizando imagens de nata, a PLS-DA se mostrou mais eficiente, porém com uma taxa de acerto de 64% máxima em três situações, com variáveis nos espaços RGB-Cinza-HSV, RGB-Cinza e Cinza-HSV com uma TCC de 88% na classe negativa. A classe positiva contava com poucas amostras e a classe indeterminada não tinha qualquer amostra escolhida pelo KS, isso traz dificuldades para o treinamento e teste das técnicas. Em todo o caso, a PLS-DA se mostrou mais eficiente quando trabalhou com espaços dimensionais associando RGB, HSV e escala de cinza.

Para a classificação com leite desnatado o resultado foi mais promissor, com 75% de TCC na PLS-DA nas dimensões R, Cinza e H. Neste caso o leite traz menos informações por ter menos componentes, pois há extração da gordura e com ela vários componentes. Desta forma a PLS-DA sugere ser mais propícia para o teste diagnóstico de mastite subclínica.

Com a classificação com NMR só foi possível fazer alguma classificação com o desnatado, com uma TCC de 78% utilizando a PLS-DA. A SPA-LDA só conseguiu classificar corretamente 56% das amostras de teste em leite desnatado e em nata ambas tiveram um desempenho de 50% de acurácia para menos.

Os resultados mostram que houve classificação para diagnosticar a mastite subclínica, porém com taxa de classificação corretas abaixo de 78%. Percebe-se que a maioria dos erros de classificação se deram nas classes positiva e indeterminada, que não continham uma quantidade representativa o suficiente para um bom modelo.

Para futuros trabalhos é interessante, primeiramente, um aumento no número de amostras, tendo um maior número de amostras nas classes positiva e indeterminada.

Outro estudo importante é verificar outras doenças, como a cetose e a possibilidade de diagnóstico com as técnicas apontadas.

Além disso, há a possibilidade de verificar a qualidade do leite com estas técnicas, em relação ao teor dos componentes e classificação quanto a possíveis adulterações. Ainda pode-se pensar em aprofundar com a NMR para estudar biomarcadores como possibilidade para diagnóstico da mastite subclínica e até da cetose.

REFERÊNCIAS BIBLIOGRÁFICAS

ABBURU, S.; BABU GOLLA, S. Satellite image classification methods and techniques: a review. *International Journal of Computer Applications*, v. 119, n. 8, p. 20–25, 2015.

ABDULLAH, S. L. S.; HAMBALI, H.; JAMIL, N. Segmentation of natural images using an improved thresholding-based technique. *Procedia Engineering*, v. 41, n. December, p. 938–944, 2012.

ALMEIDA, V. E. DE *et al.* Using color histograms and SPA-LDA to classify bacteria. *Analytical and Bioanalytical Chemistry*, v. 406, n. 24, p. 5989–5995, 2014.

APLIN, P.; SMITH, G. M. Advances in Object-Based Image Classification. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, v. 37, n. B7, p. 725–728, 2008.

ARAÚJO, M. C. U. *et al.* The successive projections algorithm for variable selection in spectroscopic multicomponent analysis. *Chemometrics and Intelligent Laboratory Systems*, v. 57, n. 2, p. 65–73, 2001.

BALLABIO, D.; CONSONNI, V. Classification tools in chemistry. Part 1: linear models. PLS-DA. *Analytical Methods*, v. 5, n. 16, p. 3790–3798, 2013.

BENEDETTE, M. F. *et al.* Mastite Bovina. *REVISTA CIENTÍFICA ELETRÔNICA DE MEDICINA VETERINÁRIA*, v. 11, n. VI, 2008.

BERND JAHNE. *Digital Image Processing*. 6^a ed. Heidelberg: Springer, 2005.

BERNSTEIN, R. Digital image processing of earth observation sensor data. *IBM Journal of Research and Development*, v. 20, n. 1, p. 40–56, 1976.

BHARATI, M. H. *Multivariate image analysis and regression for industrial process monitoring and product quality control*. 2002. McMaster University, Hamilton, Ontario, 2002.

BISHOP, C. M. *Pattern Recognition and Machine Learning*. 1st. ed. New York: Springer, 2006.

BRASIL. *Instrução Normativa nº 62 de 29 de dezembro de 2011. Diário Oficial da União*. Brasília: Ministério da Agricultura, Pecuária e Abastecimento. , 2011

BRASIL. *Instrução normativa nº 68 de 12 de dezembro de 2006. Diário Oficial da União*. Brasília: Ministério da Agricultura, Pecuária e Abastecimento. , 2006

CAPOZZI, F. The new frontier in foodomics : the perspective of nuclear magnetic resonance spectroscopy increase your profile ? *New food magazine*, v. 16, n. 2, p. 44–48, 2013.

ČEREŠKIENĖ, E. *et al.* Influence of mastitis agents on milk traits of cows. *Veterinarija ir zootechnika (Vet Med Zoot)*, v. 73, n. 95, p. 24–27, 2016.

CHAU, F. *et al.* *Chemometrics from basics to wavelet transform*. New Jersey: John Wiley & Sons, 2004.

CHEN, C. *et al.* Classification in medical images using adaptive metric k-NN. 2010, San Diego: SPIE, 2010.

CONAGIN, A. Transformações Dos Dados Experimentais (*). *Bragantia*, v. 14, n. 15, p. 141–147, 1955.

COSER, S. M.; LOPES, M. A.; COSTA, G. M. Mastite bovina : Controle e Prevenção. *Boletim Técnico UFL*, n. 93, p. 1–30, 2012.

COSTA, G. B. *et al.* Digital image-based classification of biodiesel. *Talanta*, v. 139, p. 50–55, 2015.

DIGIOVANI, D. B. *et al.* Infrared thermography as diagnostic tool for bovine subclinical mastitis detection Artigo. *Revista Brasileira de Higiene e Sanidade Animal*, v. 20, n. 4, p. 685–692, 2016.

DOHOO, I. R.; MEEK, A. H. Somatic cell counts in bovine milk. *The Canadian veterinary journal. La revue vétérinaire canadienne*, v. 23, n. 4, p. 119–25, 1982.

DOMÍNGUEZ, M. A. *et al.* Geographical origin classification of Argentinean honeys using a digital image-based flow-batch system. *Microchemical Journal*, v. 112, p. 104–108, 2014. Disponível em: <<http://dx.doi.org/10.1016/j.microc.2013.09.008>>.

DUARTE, C. M.; FREITAS, P. P.; BEXIGA, R. Technological advances in bovine mastitis diagnosis: an overview. *Journal of veterinary diagnostic investigation*, v. 27, n. 6, p. 665–72, 2015.

DUCHESNE, C.; LIU, J. J.; MACGREGOR, J. F. Multivariate image analysis in the process industries: A review. *Chemometrics and Intelligent Laboratory Systems*, v. 117, p. 116–128, 2012.

DUGUMA, A.; TOLOSA, T.; YOHANNES, A. Prevalence of clinical and sub-clinical mastitis on cross bred dairy cows at Holleta Agricultural Research Center, Central Ethiopia. *Journal of Veterinary Medicine and Animal Health*, v. 6, n. 1, p. 13–17, 2014.

ESBENSEN, K.; GELADI, P. Strategy of multivariate image analysis. *Chemometrics and Intelligent Laboratory Systems*, v. 7, p. 67–86, 1989.

FACCO, P. *et al.* Improved multivariate image analysis for product quality monitoring. *Chemometrics and Intelligent Laboratory Systems*, v. 109, p. 42–50, 2011.

FERREIRA, M. M. C. *et al.* Quimiometria I: Calibração multivariada, um tutorial. *Quimica Nova*, v. 22, n. 5, p. 724–731, 1999.

FORSBÄCK, L. *et al.* Day-to-day variation in milk yield and milk composition at the udder-quarter level. *Journal of dairy science*, v. 93, n. 8, p. 3569–3577, 2010.

GAMBHIR, P.; NAGARAJAN, S. Application of Nuclear Magnetic Resonance in Agriculture. *Proceedings of the Indian National Science Academy : PINSA.*, v. 65, n. 6, p. 731–765, 1999.

GONZALEZ, R. C.; WOODS, R. E. *Processamento Digital de Imagens*. 2. ed. São Paulo: Pearson, 2010.

GONZALEZ, R. C.; WOODS, R. E.; EDDINS, S. L. *Digital Image Processing using Matlab*. Natick: Gatesmark, 2009. v. 2.

GORDON, A. D. *Classification*. . Washington: CHAPMAN & HALL/CRC. , 1999

GRAHN, H. F.; GELADI, P. *Techniques and Applications of Hyperspectral Image*

Analysis. Chichester: John Wiley & Sons, 2007. v. 22.

HAIR, J. F. *et al.* *Multivariate Data Analysis*. Prentice Hall. [S.l: s.n.], 2009

HEALD, C. W. *et al.* A Computerized Mastitis Decision Aid Using Farm-Based Records: An Artificial Neural Network Approach. *J. Dairy Sci.*, v. 83, n. 4, p. 711–720, 2000.

HORE, P. J. *Nuclear Magnetic Resonance*. Oxford: Oxford University Press, 2015.

HP. *Suporte ao cliente HP®*. Disponível em: <<https://support.hp.com/br-pt/document/c01553666>>. Acesso em: 4 maio 2017.

IGNÁCIO, J. DA S. *Processamento e Análise Digital de Imagens em Estudos da Cinética de Recristalização de Ligas Al-Mg-X*. 2013. 106 f. Universidade de São Paulo, 2013.

INTERNATIONAL ORGANIZATION FOR STANDARDIZATION. *ISO 13366-2:2006 - Milk -- Enumeration of somatic cells -- Part 2: Guidance on the operation of fluoro-opto-electronic counters*. ISO 13366-2:2006. [S.l: s.n.], 2006

INTERNATIONAL ORGANIZATION FOR STANDARDIZATION. *ISO 9622:2013 - Milk and liquid milk products -- Guidelines for the application of mid-infrared spectrometry*. [S.l: s.n.], 2013

JOHN R. MIDDLETON *et al.* Use of somatic cell counts and California mastitis test results from individual quarter milk samples to detect subclinical intramammary infection in dairy cattle from a herd with a high bulk tank somatic cell count. *Journal of the American Veterinary Medical Association*, v. 224, n. 3, p. 419–23, 2004.

KEELER, J. *Understanding NMR Spectroscopy*. Cambridge: University of Cambridge, 2002.

KENNARD, R.; STONE, L. A. Computer aided design of experiments. *Technometrics*, v. 11, n. 1, p. 137–148, 1969.

KHAN, M.; KHAN, A. Basic Facts of Mastitis in Dairy Animals: A Review. *Pakistan Veterinary Journal*, v. 26, n. 4, p. 204–208, 2006.

KOTSIANTIS, S. B. Supervised machine learning: A review of classification techniques. *Informatica*, v. 31, p. 249–268, 2007.

KOVALEVSKY, V. A. *Image Pattern Recognition*. 1st. ed. New York: Springer, 1980.

KPALMA, K.; RONSIN, J. *An Overview of Advances of Pattern Recognition Systems in Computer Vision*. *Vision Systems: Segmentation and Pattern Recognition*. [S.l: s.n.], 2007

KUPČE, E.; FREEMAN, R. Reconstruction of the three-dimensional NMR spectrum of a protein from a set of plane projections. *Journal of Biomolecular NMR*, v. 27, n. 4, p. 383–387, 2003.

LAMBERT, J. B.; MAZZOLA, E. P. *Nuclear magnetic resonance spectroscopy: an introduction to principles, applications, and experimental methods*. [S.l.]: Pearson, 2004.

LANGONI, H. Tendências de modernização do setor lácteo: monitoramento da qualidade do leite pela contagem de células somáticas. *Revista de Educação Continuada em Medicina Veterinária e Zootecnia do CRMV-SP*, v. 3, n. 3, p. 57–64, 2000.

LOPES, L. O.; LACERDA, M. S. DE; RONDA, J. B. *Uso de Antibióticos na cura e controle de mastite clínica e subclínica causada por principais microorganismos contagiosos*

em bovinos leiteiros: Revisão de literatura. Disponível em: <http://faef.revista.inf.br/imagens_arquivos/arquivos_destaque/RyKnT9CEMC8Q2de_2013-8-13-18-19-12.pdf>. Acesso em: 2 abr. 2017.

LORBER, A.; WANGEN, L. E.; KOWALSKI, B. R. A theoretical foundation for the PLS algorithm. *Journal of Chemometrics*, v. 1, n. 1, p. 19–31, 1987.

MARQUES FILHO, O.; VIEIRA NETO, H. *Processamento Digital de Imagens*. [S.l.]: Brasport, 1999.

MEDJAHED, S. A. A Comparative Study of Feature Extraction Methods in Images Classification. *International Journal of Image, Graphics and Signal Processing*, v. 7, n. 3, p. 16–23, 2015.

MELANIE, M. *An Introduction to Genetic Algorithms*. Massachusetts: MIT Press, 1999.

MELLENBERGER, R. *California Mastitis Test (CMT)*. Michigan: Michigan State University, 2001.

MELO, G. J. A. DE *et al.* A robust segmentation method for counting bovine milk somatic cells in microscope slide images. *Computers and Electronics in Agriculture*, v. 115, p. 142–149, 2015.

MERY, D.; PEDRESCHI, F. Segmentation of colour food images using a robust algorithm. *Journal of Food Engineering*, v. 66, n. 3, p. 353–360, 2005.

MILANEZ, K. D. T. M.; PONTES, M. J. C. Classification of extra virgin olive oil and verification of adulteration using digital images and discriminant analysis. *Analytical Methods*, v. 7, p. 8839–8846, 2015.

MILLER, R. H.; PAAPE, M. J.; ACTON, J. C. Comparison of Milk Somatic Cell Counts by Coulter and Fossomatic Counters. *J. Dairy Sci.*, v. 69, p. 1942–1946, 1986.

NAES, T. *et al.* *A user-friendly guide to Multivariate Calibration and Classification*. Chichester: NIR Publications, 2002.

NALON, J. A. *Introdução ao processamento digital de sinais*. Rio de Janeiro: LTC, 2013.

NUNES, É. DE O.; CONCI, A. *Algoritmo genético para classificação temática de imagens multiespectrais*. [S.l.: s.n.], 2010.

PEDRAZ, C. G. *Biomarcadores*. Disponível em: <<http://www.dicyt.com/viewNews.php?newsId=20792>>. Acesso em: 29 mar. 2017.

PONTES, M. J. C. *Algoritmo das Projeções Sucessivas para Seleção de Variáveis Espectrais em Problemas de Classificação*. 2009. 123 f. Universidade Federal da Paraíba, 2009.

PONTES, M. J. C. *et al.* The successive projections algorithm for spectral variable selection in classification problems. *Chemometrics and Intelligent Laboratory Systems*, v. 78, n. 1, p. 11–18, 2005.

PRATS-MONTALBÁN, J. M.; JUAN, A. DE; FERRER, A. Multivariate image analysis: A review with applications. *Chemometrics and Intelligent Laboratory Systems*, v. 107, n. 1, p. 1–23, 2011.

SCHABAUER, L. *et al.* Novel physico-chemical diagnostic tools for high throughput

identification of bovine mastitis associated gram-positive, catalase-negative cocci. *BMC veterinary research*, v. 10, n. 156, p. 1–11, 2014.

SCHULTZE, W. D.; BRAMLEY, A. J. Effect of Escherichia coli endotoxin-mediated inflammation of one mammary quarter of the bovine udder on diapedesis into other quarters. *Journal of Dairy Research*, v. 49, p. 381–385, 1982.

SHAPIRO, L.; STOCKMAN, G. *Computer Vision*. Upper Saddle River: Prentice Hall, 2000.

SKOOG, D. A.; HOLLER, F. J.; CROUCH, S. R. *Principles of Instrumental Chemistry*. Belmont: Thomson Brooks/Cole, 2007.

SOARES, S. F. C. *et al.* The successive projections algorithm. *TrAC - Trends in Analytical Chemistry*, v. 42, p. 84–98, 2013.

SOLOMON, C.; BRECKON, T. *Fundamentos de processamento digital de imagens : uma abordagem prática com exemplos em Matlab*. 1. ed. Rio de Janeiro: LTC, 2013.

SOUTO, R. P. *Segmentação de imagem multiespectral utilizando-se o atributo matiz*. 2003. 171 f. Instituto Nacional de Pesquisas Espaciais, 2003.

SOUTO, U. T. DE C. P. *et al.* Identification of adulteration in ground roasted coffees using UV-Vis spectroscopy and SPA-LDA. *LWT - Food Science and Technology*, v. 63, p. 1037–1041, 2015.

SUNDEKILDE, U. K. *et al.* Nuclear magnetic resonance metabonomics reveals strong association between milk metabolites and somatic cell count in bovine milk. *Journal of dairy science*, v. 96, n. 1, p. 290–9, 2013.

SUNDEKILDE, U. K.; LARSEN, L. B.; BERTRAM, H. C. NMR-Based Milk Metabolomics. *Metabolites*, v. 3, n. 2, p. 204–22, 2013.

TESFAYE, G. Y.; REGASSA, F. G.; KELAY, B. Milk yield and associated economic losses in quarters with subclinical mastitis due to Staphylococcus aureus in Ethiopian crossbred dairy cows. *Tropical Animal Health and Production*, v. 42, p. 925–931, 2010.

TSENKOVA, R. *et al.* Near infrared spectra of cows ' milk for milk quality evaluation : disease diagnosis and pathogen identification. v. 370, p. 363–370, 2006.

UMEA, P. G.; GRAHN, H. Multivariate Image Analysis. *Encyclopedia of Analytical Chemistry*. [S.l.]: John Wiley & Sons, 2007. p. 1–23.

WALCZAK, K. Image retrieval using spatial color. In: SKARBEEK, W. (Org.). *Computer analysis of images and patterns : 9th international Conference : proceedings*. [S.l.]: Springer, 2001. p. 743.

WALPOLE, R. E. *et al.* *Probability & Statistics for Engineers & Scientists 9th Ed.* [S.l.: s.n.], 2011.

WESTAD, F.; BEVILACQUA, M.; MARINI, F. Regression. *Chemometrics in Food Chemistry*. Amsterdam: Elsevier, 2013. p. 127–168.

WOLD, H. Partial least squares. *Encyclopedia of Statistical Sciences*, n. 2, p. 581–591, 1985.

XAVIER, A. C. *et al.* Análise Discriminante e Classificação de Imagens 2D de

Ultrassonografia Mamária. 2011, Curitiba: [s.n.], 2011. p. 67–72.

YU, H.; MACGREGOR, J. F. Multivariate image analysis and regression for prediction of coating content and distribution in the production of snack foods. *Chemometrics and Intelligent Laboratory Systems*, 2003.

ZHANG, Z. *et al.* Early mastitis diagnosis through topological analysis of biosignals from low-voltage alternate current electrokinetics. *Proceedings of the Annual International Conference of the IEEE Engineering in Medicine and Biology Society, EMBS*, v. 2015–Novem, p. 542–545, 2015.

ZIMMERMANN, F. J. P. *Estatística aplicada à pesquisa agrícola*. Santo Antônio de Goiás: Embrapa, 2004.

APÊNDICES

A. Dados de Inventário Bovino e SCC sobre as amostras

Amostra	Subclínica (suspeita)	Classe	SCC (x1000 SC/mL)	Idade (meses) (1)	Nº de crias (2)	Último Parto (3)	Litros ordenha (4)	Gordura g/g(%) (5)	Proteína g/g(%) (6)	Lactose g/g(%) (7)	Caseína g/g(%) (8)	Ureia g/g(%) (9)
AA88	Negativo	1	103	59,4	3	2,16	16,4	3,4	3,88	4,79	3	9,56
DP268	Negativo	1	125	81,9	3	12,15	13,6	4,33	4,06	4,57	3,27	13,54
DP352	Positivo	2	598	47,4	1	2,16	12,4	4,2	3,99	4,48	3,13	10,24
DP241	Positivo	2	4163	93,3	4	1,16	11	4,03	4,51	4,18	3,69	14,46
DP369	Negativo	1	147	35,2	1	12,15	9,2	3,88	3,75	4,74	2,92	12,27
DP295	Negativo	1	79	72,8	2	6,15	10,2	4,14	4,56	4,08	3,69	17,91
DP370	Negativo	1	143	35,5	1	9,16	16,4	4,21	3,43	4,67	2,62	7,24
DP342	Negativo	1	70	52,7	3	9,16	24	3,27	3,15	4,53	2,4	8,87
DP366	Negativo	1	159	39,2	1	6,16	12	4,13	3,6	4,53	2,78	10,11
DP378	Negativo	1	114	30,7	1	4,16	17,2	3,94	3,71	4,77	2,97	23,41
DP320	Indeterminado	3	235	60,6	3	2,16	17,4	3,18	3,5	4,69	2,65	14,52
DP301	Positivo	2	319	68,3	4	4,16	22,4	5,07	4	4,7	3,15	10,51
AA69	Indeterminado	3	215	90,6	4	1,16	15,2	2,96	3,53	4,33	2,77	10,43
DP308	Negativo	1	172	67,1	3	4,16	14,2	3,22	3,91	4,41	3,14	15,68
DP343	Negativo	1	167	51,9	2	7,15	11,4	3,31	3,92	4,5	3,08	9,78
DP281	Negativo	1	79	77	4	5,16	28,2	3,75	3,42	4,76	2,61	12,91
DP376	Positivo	2	5690	31,1	1	9,16	11,6	3,19	3,2	4,43	2,42	13,45
DP306	Negativo	1	131	68,2	2	1,16	22,4	3,76	3,96	4,55	3,2	16,18
AA79	Positivo	2	377	74,4	3	8,15	4,8	3,79	4,42	4,05	3,57	13,39
DP213	Negativo	1	114	107,6	6	12,15	23	3,85	3,48	4,5	2,68	12,42
DP364	Negativo	1	149	38,9	1	12,15	12,2	4,08	4,13	4,64	3,27	14,79
DP360	Negativo	1	105	42,1	1	9,15	19,6	3,21	3,61	4,93	2,82	12,1
DP319	Negativo	1	27	60,7	2	12,15	21,2	5,64	4,27	4,67	3,43	16,2
AA77	Negativo	1	119	77,8	4	4,16	23,4	3,86	3,56	4,59	2,79	11,53
DP355	Negativo	1	48	44,1	1	12,15	21,4	3,67	3,94	4,75	3,16	12,02
DP233	Negativo	1	155	96,8	6	4,16	19,2	5,51	4,22	4,37	3,43	20,17

Amostra	Subclínica (suspeita)	Classe	SCC (x1000 SC/mL)	Idade (meses) (1)	Nº de crias (2)	Último Parto (3)	Litros ordenha (4)	Gordura g/g(%) (5)	Proteína g/g(%) (6)	Lactose g/g(%) (7)	Caseína g/g(%) (8)	Ureia g/g(%) (9)
AA71	Positivo	2	591	89,7	3	8,15	12,6	3,65	4,12	4,13	3,34	14,96
DP323	Indeterminado	3	240	59,7	2	12,15	23,6	4,21	4,15	4,52	3,31	17,09
AA97	Negativo	1	33	36,8	1	5,16	18,2	3,02	3,29	4,92	2,52	13,48
DP372	Negativo	1	84	33,4	1	2,16	10,4	3,26	3,81	4,78	2,96	10,45
DP309	Negativo	1	150	66,3	2	7,15	7	5,46	5,39	4,08	4,47	20,59
AA90	Negativo	1	36	53,9	2	8,15	11,2	6,15	5,57	4,26	4,57	17,3
DP324	Negativo	1	77	59,3	3	9,16	19,2	4,56	3,52	4,75	2,72	8,5
DP346	Negativo	1	116	49,8	1	2,16	13,6	3,97	4,47	4,85	3,67	18,44
DP362	Positivo	2	302	39,7	1	12,15	12	3,58	3,88	4,51	3,08	13,86
AA93	Negativo	1	81	44,2	1	3,16	15,8	2,15	3,77	4,73	2,97	7,58
DP375	Negativo	1	192	31,7	1	9,16	15	3,15	3,31	5,05	2,56	8,55
AA67	Indeterminado	3	205	62,3	7	8,16	26,4	5,28	3,16	4,48	2,42	10,88
DP351	Negativo	1	27	47,8	1	9,15	22,6	3,4	3,54	4,9	2,73	9,68
DP333	Negativo	1	33	54,1	3	4,16	24,4	2,71	3,61	4,51	2,78	12,57