

TIAGO BONINI BORCHARTT

ANÁLISE DE IMAGENS TERMOGRÁFICAS PARA A CLASSIFICAÇÃO DE
ALTERAÇÕES NA MAMA

Tese apresentada ao Programa de Pós-Graduação em Computação da Universidade Federal Fluminense, como requisito parcial para obtenção do Grau de Doutor. Área de Concentração: Computação Visual.

Orientador: Prof.^a Dr.^a Aura Conci

Niterói

2013

Ficha Catalográfica elaborada pela Biblioteca da Escola de Engenharia e Instituto de Computação da UFF

B726 Borchartt, Tiago Bonini

Análise de imagens termográficas para a classificação de alterações na mama / Tiago Bonini Borchartt. – Niterói, RJ : [s.n.], 2013.

118 f.

Tese (Doutorado em Computação) - Universidade Federal Fluminense, 2013.

Orientador: Aura Conci.

1. Diagnóstico por imagem. 2. Termografia. 3. Mamografia por computação visual. 4. Dimensão fractal. 5. Máquina de vetor de suporte. I. Título.

CDD 006.42

TIAGO BONINI BORCHARTT

ANÁLISE DE IMAGENS TERMOGRÁFICAS PARA A CLASSIFICAÇÃO DE
ALTERAÇÕES NA MAMA

Tese apresentada ao Programa de Pós-Graduação em Computação da Universidade Federal Fluminense, como requisito parcial para obtenção do Grau de Doutor. Área de Concentração: Computação Visual.

Aprovada em junho de 2013.

BANCA EXAMINADORA

Prof.^a Dr.^a Aura Conci – Orientador
Universidade Federal Fluminense – UFF

Prof. Dr. Alexandre Plastino de Carvalho
Universidade Federal Fluminense – UFF

Prof.^a Dr.^a Débora Christina Muchaluat Saade
Universidade Federal Fluminense – UFF

Prof. Dr. Alair Sarmet dos Santos
Universidade Federal Fluminense – UFF

Prof. Dr. Marcello Gattass
Pontifícia Universidade Católica do Rio de Janeiro – PUC-Rio

Prof. Dr. Aristófanés Corrêa Silva
Universidade Federal do Maranhão – UFMA

Prof.^a Dr.^a Rita de Cássia Fernandes Lima
Universidade Federal de Pernambuco – UFPE

Aos meus pais, Evaldo e Clara.

Aos meus irmãos, Zilmar, Berenice e Alessandro.

À minha esposa, Priscilla Karenwen.

AGRADECIMENTOS

Agradeço aos meus pais, meus irmãos, minha irmã e meus sobrinhos, que são a motivação para que eu lute e tente sempre, sem nunca pensar em desistir.

Agradeço a minha esposa pela companhia diária durante o desenvolvimento desta tese, seu apoio e sua motivação foram fundamentais para a finalização deste trabalho e desta fase da minha vida. Te amo.

Agradeço a todos os meus familiares pelo carinho e apoio de sempre, em especial a minha tia Madalena Bonini que sempre acreditou na minha capacidade e não poupou esforços para investir nos meus estudos.

Agradeço a minha orientadora Aura Conci, que desde quando eu estava no mestrado confiou em mim. Durante o doutorado ela esteve sempre me incentivando, potencializando meu crescimento não só como estudante, mas também como pessoa.

Agradeço aos amigos do VisualLab, Roger, Lincoln, Giomar, Rafael, João Gabriel, João Paulo, Thiago Elias, Francieric, Stephenson, pela amizade e pelas grandes colaborações no meu trabalho, pois cada etapa desenvolvida nesta tese, em algum momento, teve o auxílio de algum colega do laboratório.

Agradeço aos demais colegas do Instituto de Computação pelas conversas de corredor, discussões de aula e opiniões no trabalho. Agradeço aos amigos do MediaLab, André Brandão, Giancarlo Taveira, Bruno Moreira, Cristian Ruff, entre outros, pela recepção logo que cheguei em Niterói.

Agradeço aos colegas da UFMA e da UFPE pela colaboração nos projetos e ao pessoal da UFMA pela hospitalidade durante os intercâmbios que fiz em São Luís. Intercâmbios que resultaram em casamento.

Agradeço aos amigos do grupo NinjaGaiden3060 pela amizade, pela “*carioquização*”, pelos sábados de futebol, e tudo mais.

“Quem diz que não pode ser feito nunca deve interromper aquele que está fazendo.” - George Shaw

RESUMO

O câncer de mama lidera o índice de mortalidade em mulheres com idades entre 15 e 54 anos. A maioria dos programas de prevenção é baseada no uso da mamografia para rastreamento. Técnicas de aprendizado de máquina e mineração de dados são usadas para ajudar a reduzir a mortalidade por câncer de mama. Essas técnicas são normalmente aplicadas em imagens mamográficas e bancos de dados clínicos. Com essas técnicas, a informação é obtida com base em textos de dados clínicos (dicionários de dados), classificação BI-RADS e características extraídas das imagens, através da utilização de algoritmos de extração de características e mineração de imagem. Nesta tese é descrito um novo esquema para auxílio à detecção de problemas nas mamas, utilizando imagens termográficas, que pode ser aplicado a vários tipos de protocolos de aquisição. As imagens utilizadas foram adquiridas no Hospital Universitário Antônio Pedro da Universidade Federal Fluminense (UFF) e no Hospital das Clínicas da Universidade Federal de Pernambuco (UFPE). O projeto foi aprovado pelo Comitê de Ética da UFF e no Ministério da Saúde do Brasil. As imagens adquiridas estão disponíveis (por requisição) em um banco de dados público no site do VisualLab. Técnicas automáticas ou manuais foram utilizadas para se obter as regiões das mamas (Região de interesse - ROI) nas imagens. As ROIs foram verificadas com o apoio de um médico. As características utilizadas foram baseadas em medidas estatísticas, histograma, dimensão fractal de Higuchi e em três métodos de geoestatística: Coeficiente de Geary, Índice de Moran e função K de Ripley. A classificação dessas características foi realizada usando a Máquina de Vetores de Suporte (SVM), que é um classificador linear binário não probabilístico baseado em um método de aprendizado de máquina supervisionado. Para validar a classificação realizada pelo SVM foram utilizadas as medidas de especificidade, sensibilidade, acurácia, área sobre a curva ROC e o índice de Youden. Durante os testes foi utilizada uma amostra de 51 termogramas, sendo 37 de pacientes com algum problema na mama e 14 pacientes com mamas saudáveis. Os resultados obtidos pelo SVM foram otimizados utilizando a técnica de algoritmo genético. O método apresentado obteve os seguintes resultados: especificidade de 79%, sensibilidade de 92%, acurácia de 88%, área sob a curva ROC de 0,852 e índice de Youden de 0,705.

Palavras-chave: imagens termográficas; detecção de doenças da mama; dimensão fractal de Higuchi; métodos geoestatísticos; máquina de vetores de suporte.

ABSTRACT

Breast cancer is leading the mortality rate in women with ages between 15 and 54. Most of the prevention programs are based on using mammography for screening. Data mining and machine learning techniques are used to help reducing the mortality by breast cancer, these techniques are usually applied in mammographic images and clinical databases. With those techniques, the information is retrieved based on texts of clinical data (data lexicons), BI-RADS classification and features extracted from images through the use of algorithms for feature extraction and image mining. We describe a new scheme for breast diseases screening, that can be applied to multiple types of acquisition protocols of infrared images. The images from this project are being acquired in the Antônio Pedro University Hospital at the Federal Fluminense University and at the Clinical Hospital of Federal University of Pernambuco. The project was approved by the Ethics Committee of UFF and the Brazilian Ministry of Health. The images acquired are available (by request) on a public database at the VisualLab website. Automatic or manual techniques are used to obtain the region of interest (ROI). The ROIs are verified with a physician's support. The features used were based in statistical measures, histogram, Higuchi's fractal dimension and three geostatistic methods: Geary Coefficient, Moran Index and Ripley's K function. The Support Vector Machine (SVM) technique is used to classify the features. SVM is a supervised machine learning method and non-probabilistic binary linear classifier. We use the specificity, sensibility, accuracy, ROC curve and Youden index to validate the classification realized by the Support Vector Machine (SVM) classifier. During the tests we used a sample of 51 thermograms: 37 non healthy and 14 healthy patients. The results were optimized using the genetic algorithm. The proposed method achieved the following results: specificity of 79%, sensibility of 92%, accuracy of 88%, area under ROC curve of 0.852 and Youden index of 0.705.

Keywords: infrared image; breast screening; Higuchi's fractal dimension; geostatistic methods; support vector machine.

LISTA DE ILUSTRAÇÕES

Figura 1: Número de mortes por câncer de mama em 2008 e 2009 (Fonte: Organização Mundial da Saúde [2]).	15
Figura 2: Mamografias: mama predominantemente composta por tecidos densos (esq.) e composta por tecido adiposo (dir.).	17
Figura 3: Termogramas: paciente com fibroadenoma (esq.) e paciente saudável (dir.).	19
Figura 4: Anatomia da mama feminina (Fonte: Winslow [15]).	22
Figura 5: Etapas do surgimento de uma neoplasia (Fonte: Danum International [18]).	24
Figura 6: Mamografia: (a) Craniocaudal e (b) Médio-lateral Oblíqua (Fonte: Nagy [22]).	26
Figura 7: Ressonância Magnética: (a) realização do exame e (b) imagem adquirida (Fontes: Massada [25] e Alvares e Michell [26], respectivamente).	27
Figura 8: Realização da ultrassonografia da mama: convencional (esq.) e 3D automática (dir.) (Fonte: Gann [30]).	28
Figura 9: Representação do espectro eletromagnético (Fonte: Berkeley Lab [39]).	32
Figura 10: Representação da aquisição IR: (1) ambiente, (2) objeto de interesse, (3) atmosfera e (4) câmera (Fonte: Flir Systems [40]).	33
Figura 11: Aprendizado supervisionado.	35
Figura 12: Três possíveis modelos de classificação de uma mesma base de treino com duas características (Fonte: Lorena e Carvalho [43]).	37
Figura 13: Separação por hiperplanos do classificador SVM (Fonte: Lorena e Carvalho [43]).	38
Figura 14: Exemplo de Curva ROC.	42
Figura 15: Etapas da metodologia para alcançar o objetivo desta tese.	52
Figura 16: Posições capturadas pelo protocolo de aquisição da UFPE.	53
Figura 17: Posições capturadas pelo protocolo de aquisição da UFF.	55
Figura 18: Análise do histograma de 140 termogramas (Fonte: Marques [83]).	58
Figura 19: Termograma original (esq.) e termograma após pré-processamento (dir.).	58
Figura 20: Detecção de pontos simétricos e calculo do ângulo de compensação.	60
Figura 21: Exemplos de termogramas segmentados pelo software de Marques [83].	60
Figura 22: Histograma da frequência de temperatura do termograma IR_3996.	64
Figura 23: Superfície isotérmica: mama inteira (esq.), em n níveis (centro) e apenas no nível mais elevado (dir.).	66
Figura 24: Temperaturas do termograma IR_0089 em formato sequencial.	68

Figura 25: Gráfico $\ln(1/k) \times \ln(L[k])$, com $k_{max} = 100$.	69
Figura 26: Sucolaridade: resultados obtidos pela percolação em 4 direções em uma mesma distribuição, (a) T2B, (b) B2T, (c) L2R e (d) R2L.	70
Figura 27: Parâmetros definidos para os métodos de geoestatística.	73
Figura 28: Exemplo de cromossomo gerado para o algoritmo genético.	84
Figura 29: Exemplo de como são geradas as novas populações.	85
Figura 30: Captura da tela de gerenciamento do ThermoCAD-UFF.	87
Figura 31: Tela de análise de termogramas.	88
Figura 32: Tela de treinamento do classificador.	88
Figura 33: Termograma IR_3996 (HC).	115
Figura 34: Termograma IR_0089 (HC).	115

LISTA DE TABELAS

Tabela 1: Matriz de Confusão.....	39
Tabela 2: Resumo dos principais protocolos de aquisição de termogramas da mama.	46
Tabela 3: Resumo dos principais trabalhos sobre o armazenamento de termogramas.....	47
Tabela 4: Resumo dos principais trabalhos de segmentação de termogramas.	49
Tabela 5: Resumo dos principais trabalhos de análise de características.	51
Tabela 6: Amostra de termogramas utilizados.....	56
Tabela 7: Grupos de medidas analisadas	61
Tabela 8: Distribuição das temperaturas mais elevadas do termograma IR_3996.	65
Tabela 9: Análise dos resultados obtidos com o registro de imagem.	79
Tabela 10: Resultados obtidos com o registro no protocolo dinâmico.	80
Tabela 11: Resultados obtidos pelo <i>Teste 1</i>	80
Tabela 12: Resultados obtidos pelo <i>Teste 2</i> validados por <i>leave-one-out</i>	81
Tabela 13: Resultados obtidos pelo <i>Teste 2</i> validados por <i>17-folds</i>	81
Tabela 14: Resultados obtidos pelo <i>Teste 3</i> validados por <i>leave-one-out</i>	81
Tabela 15: Resultados obtidos pelo <i>Teste 3</i> validados por <i>17-folds</i>	82
Tabela 16: Resultados obtidos pelo <i>Teste 4</i> validados por <i>leave-one-out</i>	82
Tabela 17: Resultados obtidos pelo <i>Teste 4</i> validados por <i>17-folds</i>	82
Tabela 18: Resultados obtidos pelo <i>Teste 5</i> validados por <i>leave-one-out</i>	83
Tabela 19: Resultados obtidos pelo <i>Teste 5</i> validados por <i>17-folds</i>	83
Tabela 20: Resultados obtidos pelo <i>Teste 6</i>	83
Tabela 21: Resultados obtidos pelo <i>Teste 6</i>	83
Tabela 22: Resultados obtidos pelo melhor indivíduo após 1000 iterações.	86
Tabela 23: Resumo de todos os testes realizados.	86
Tabela 24: Comparação de resultados.	92
Tabela 25: Lista de termogramas utilizados do HC – UFPE.	113
Tabela 26: Lista de termogramas utilizados do HUAP – UFF.	113
Tabela 27: Valores gerais dos atributos utilizados no Teste 2 (H: Higuchi, S: Sucularidade, R: K-Ripley, M: Í-Moran, G: C-Geary).....	114

LISTA DE ABREVIATURAS E SIGLAS

3D - Tridimensional
CADE - *Computer Aided Detection* (Detecção auxiliada por computador)
CADx - *Computer Aided Diagnosis* (Diagnóstico auxiliado por computador)
CC - Crânio-caudal
CD - Carcinoma Ductal
CLI - Carcinoma Lobular Invasivo
CLIS - Carcinoma Lobular *in situ*
CT - *Computed Tomography* (Tomografia Computadorizada)
DDSM - *Digital Database for Screening Mammography* (Banco de Dados Digital para Mamografia de Triagem)
DF - Dimensão Fractal
DMI - *Database For Mastological Images* (Banco de Dados para Imagens Mastológicas)
FDA - *U.S. Food and Drug Administration* (Administração de Drogas e Comidas dos Estados Unidos)
FN - Falso-Negativo
FP - Falso-Positivo
HC - Hospital das Clínicas
HUAP - Hospital Universitário Antônio Pedro
INCA - Instituto Nacional do Câncer José de Alencar
IR - Infravermelho
LTI - *Localized Temperature Increases* (Aumento Localizado de Temperatura)
MIAS - *Mammographic Image Analysis Society* (Sociedade de Análise de Imagens Mamográficas)
MLO - Médio-Lateral Oblíqua
NTBS - *NoTouch BreastScan*
OMS - Organização Mundial da Saúde
PEM - *Positron Emission Mammography* (Mamografia por Emissão de Pósitrons)
RM - Ressonância Magnética
RNA - Redes Neurais Artificiais
ROC - *Receiver Operating Characteristic* (Característica de Operação do Receptor)
RNFN - *Rule based Neurofuzzy Networks* (Redes Neurofuzzy baseadas em Regras)
ROI - *Region of Interest* (Região de Interesse)
SBS - *Sentinel BreastScan*
SDK - *Software Development Kit* (Conjunto de Desenvolvimento de Software)
SPSS - *Statistical Package for the Social Sciences* (Pacote Estatístico para as Ciências Sociais)
SUS - Sistema Único de Saúde
SVM - *Support Vector Machine* (Máquinas de Vetores de Suporte)
TAE - Teoria do Aprendizado Estatístico
TIE - Tomografia por Impedância Elétrica
US - Ultrassonografia
VN - Verdadeiro-Negativo
VP - Verdadeiro-Positivo
VPP - Valor Preditivo Positivo

SUMÁRIO

Capítulo 1 – Introdução	15
1.1 Motivação	17
1.2 Objetivos	19
1.3 Estrutura da tese	20
Capítulo 2 – Conceitos básicos	22
2.1 Mama	22
2.1.1 Estrutura da mama	22
2.1.2 O surgimento de neoplasias	23
2.1.3 Principais doenças da mama	24
2.1.4 Imagens médicas aplicadas na mastologia	25
2.2 Termografia	30
2.2.1 Imagens termográficas	31
2.3 Classificadores	34
2.3.1 Seleção do classificador	35
2.3.2 Máquina de vetor de suporte (SVM)	36
2.4 Técnicas de avaliação	38
2.4.1 Validação cruzada	38
2.4.2 Matriz de confusão	39
2.4.3 Sensibilidade e especificidade	40
2.4.4 Acurácia e precisão	40
2.4.5 Área sob a Curva ROC	41
2.4.6 Índice de Youden	42
2.5 Resumo do capítulo	43
Capítulo 3 – Revisão bibliográfica	44
3.1 Aquisição e armazenamento dos termogramas	44
3.2 Segmentação da região de interesse	48

3.3 Auxílio à detecção e diagnóstico de doenças.....	49
Capítulo 4 – Metodologia desenvolvida	52
4.1 Resumo	52
4.2 Aquisição dos termogramas.....	52
4.3 Amostra utilizada	55
4.4 Pré-processamento	56
4.5 Conversão da matriz de temperatura para imagem.....	57
4.6 Correção da posição da paciente na imagem	59
4.7 Segmentação da região de interesse.....	60
4.8 Extração de características.....	60
4.8.1 Características baseadas em estatística (Grupo I).....	62
4.8.2 Características baseadas em histograma (Grupo II)	63
4.8.3 Características baseadas em geometria fractal (Grupo III).....	67
4.8.4 Características baseadas em geoestatística (Grupo IV)	71
4.9 Classificação	74
4.10 Ferramentas utilizadas	75
Capítulo 5 – Testes e resultados.....	76
5.1 Testes	76
5.2 Resultados.....	79
5.2.1 Resultados obtidos no pré-processamento.....	79
5.2.2 Resultados obtidos nos testes.....	80
5.2.3 Otimização dos resultados	84
5.2.4 Resumo dos resultados obtidos nos testes	86
5.2.5 ThermoCAD-UFF.....	86
Capítulo 6 – Trabalhos relacionados.....	89
6.1.1 Comparações.....	91
Capítulo 7 – Discussões finais	94

7.1 Discussão dos resultados	94
7.2 Conclusões	96
7.3 Contribuições	98
7.4 Trabalhos futuros	99
7.5 Publicações	101
7.5.1 Capítulo de livro	101
7.5.2 Artigos em periódicos internacionais	101
7.5.3 Artigos em periódicos nacionais.....	101
7.5.4 Resumos expandidos em periódicos	101
7.5.5 Artigos em anais de congressos nacionais	102
7.5.6 Artigos em anais de congressos internacionais	102
Referências.....	104
Anexo A	112
Anexo B	114
Anexo C	115

CAPÍTULO 1 – INTRODUÇÃO

A última estimativa do Instituto Nacional de Câncer [1] para o biênio 2012/2013, no Brasil, espera 52.680 novos casos de câncer de mama por ano, com um risco estimado de 52 casos a cada 100 mil mulheres. O câncer de mama é o que mais acomete as mulheres em todo o mundo, tanto em países desenvolvidos como em países que estão em desenvolvimento. A Organização Mundial da Saúde [2] calcula que o câncer de mama foi responsável pela morte de aproximadamente 480 mil mulheres no ano de 2008 e 515 mil no ano de 2009. Forouzanfar *et al.* [3] estimam que em 2010 foram aproximadamente 425 mil mortes por câncer de mama. A idade é um dos fatores de risco para o câncer de mama. Até os 50 anos as taxas de incidência aumentam rapidamente de acordo com a idade, como pode ser observado na Figura 1.

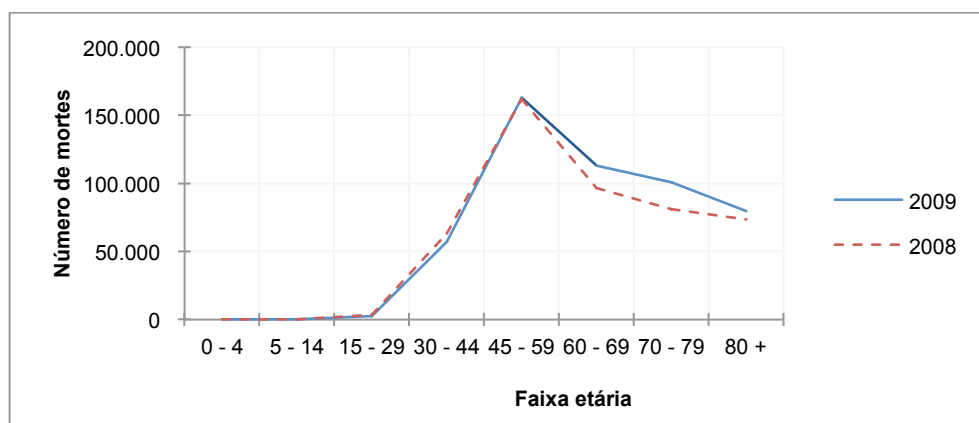


Figura 1: Número de mortes por câncer de mama em 2008 e 2009 (Fonte: Organização Mundial da Saúde [2]).

Além da idade, outros fatores de risco já foram identificados, como por exemplo, características da vida reprodutiva da mulher (menarca precoce, nuliparidade, idade da primeira gestação acima dos 30 anos, anticoncepcionais orais, menopausa tardia e terapia de reposição hormonal), histórico familiar de câncer de mama, alta densidade do tecido mamário (razão entre o tecido glandular e o tecido adiposo da mama) e a exposição à radiação ionizante [1]. O número de ciclos ovarianos está diretamente associado ao risco de desenvolver câncer de mama, reduzindo cerca de 15% a cada ano de retardo do início da menarca, bem como aumentando aproximadamente 3% a cada ano de retardo da menopausa. Já o histórico familiar está relacionado a um aumento de cerca de duas a três vezes no risco de desenvolver esta neoplasia [1].

A prevenção do câncer de mama ainda é difícil devido à grande variedade de fatores de riscos e das características genéticas dos tumores. Novas estratégias de detecção de câncer de mama têm sido estudadas, principalmente abordagens factíveis para países em que grande parte da população sobrevive com baixa renda.

Atualmente, os principais aliados da detecção do câncer de mama são o autoexame, o exame clínico e a mamografia. Entre estes a mamografia é o mais indicado para a detecção precoce de doenças. No Brasil existem inúmeras campanhas em nível nacional para alertar sobre a importância da detecção e tratamento precoce. Estas campanhas estimulam que as mulheres desde cedo façam o autoexame e que tenham, a partir dos 40 anos, a mamografia como um exame de rotina [1], [4]. O Ministério da Saúde, através do Sistema Único de Saúde (SUS), subsidia exames de mamografia para mulheres na faixa etária acima de 40 anos. Entretanto, diversos outros tipos de exames por imagem podem ser realizados, entre eles encontram-se a termografia, ultrassonografia, ultrassonografia tridimensional automática, tomografia computadorizada (CT), ressonância magnética, *Doppler*, tomossíntese, mamografia por emissão de pósitrons (PEM), etc.

Os exames podem ser divididos entre exames de triagem e exames diagnósticos, dependendo da situação em que são utilizados. Exames de triagem buscam encontrar possíveis doenças antes de a pessoa apresentar qualquer sintoma e podem ajudar a detectar doenças em fase inicial. Quanto antes uma doença é detectada, maiores são as chances de tratamento. Se um resultado do exame de triagem é anormal, outros exames são necessários para definir a descoberta, estes são chamados de exames diagnósticos.

Cada tipo de exame tem características específicas, sendo na maioria das vezes complementares, o que torna muitas vezes sua combinação importante para um diagnóstico mais preciso. Por exemplo, os exames de CT, tomossíntese e mamografia são estruturais, isto é, apresentam informações morfológicas e anatômicas das estruturas do corpo. Enquanto que a termografia é um exame biológico ou funcional. A utilização de diversos exames combinados é usual em centros médicos modernos, pois permitem a obtenção de um elevado grau de especificidade e sensibilidade no diagnóstico [5].

As imagens médicas da mama obtidas pelos diferentes tipos de exames apresentam um alto grau de informações relevantes para o especialista que está realizando a sua análise. O uso do computador auxilia no aproveitamento destas informações e facilita a sua interpretação. Este auxílio pode ser na forma de sistemas *CADe* (*Computer Aided Detection*) ou *CADx* (*Computer Aided Diagnosis*), metodologias de detecção e classificação automática de padrões, técnicas computacionais de modelagem, entre outras possibilidades.

1.1 MOTIVAÇÃO

Apesar de o câncer de mama ser considerado um câncer de bom prognóstico se diagnosticado e tratado precocemente, as taxas de mortalidade por este câncer continuam elevadas em todo o mundo, inclusive no Brasil. Esse fato, muito provavelmente, está relacionado à doença ainda ser diagnosticada apenas em estágios avançados [6].

A mamografia é o principal exame utilizado para o diagnóstico em estágios iniciais do câncer de mama, sendo o exame mais usado para diagnosticar este tipo de câncer. No entanto, a mamografia encontra algumas limitações, como a dificuldade de detectar tumores em pacientes jovens [7] ou cânceres sem massas, como o carcinoma de Paget [8]. Esta dificuldade está relacionada à densidade das mamas, pois a mama jovem é principalmente composta de tecido glandular, o que a torna mais densa. Esta alta densidade da mama interfere na identificação de massas e microcalcificações pelos raios-X. É bastante complexa a detecção de alterações em uma imagem resultante da mamografia de uma mama densa, como pode ser visto na Figura 2 (esquerda) [9]. Gradativamente, devido à diminuição dos níveis hormonais, o tecido glandular tende a se atrofiar e ser substituído por tecido adiposo, até o momento em que a mulher atinge a menopausa onde sua mama geralmente está constituída na maior parte por tecido adiposo. Esta nova constituição da mama a torna menos densa facilitando a detecção de regiões suspeitas através da mamografia, como pode ser visto na Figura 2 (direita) [10].

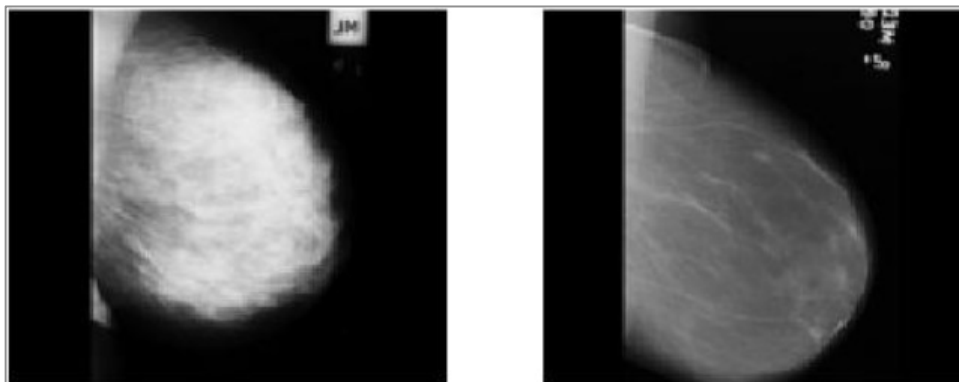


Figura 2: Mamografias: mama predominantemente composta por tecidos densos (esq.) e composta por tecido adiposo (dir.).

Esta limitação da mamografia, juntamente com o crescente número de casos de câncer em mulheres jovens e o alto custo de exames como a CT e o PEM, motivam o desenvolvimento de novas metodologias para a detecção precoce de doenças na mama, dentre as quais, destaca-se a termografia.

A termografia começou a ser utilizada na mastologia em 1959, porém os equipamentos disponíveis na época fizeram com que o método fosse desacreditado e não recomendado para o diagnóstico da mama [11]. Com o avanço tecnológico das câmeras termográficas e das técnicas de processamento e análise de imagens, pesquisas envolvendo a termografia voltaram a ser realizadas. Em 1982, o uso da termografia como exame de triagem para doenças de mama foi aprovado pela *U.S. Food and Drug Administration* (FDA), entretanto, a FDA ressalta que a termografia é um exame complementar à mamografia e não substituto [12].

A análise visual de uma imagem termográfica ou termograma na procura de problemas da mama pode apresentar dificuldades devido a alguns fatores, como:

- i. A baixa resolução de algumas câmeras: as câmeras térmicas com custos mais acessíveis têm resolução de 160x160 pixels;
- ii. O baixo contraste nas imagens: em alguns casos uma variação de temperatura de 1°C pode ser relevante e passar despercebida devido ao contraste mal ajustado na imagem;
- iii. Uma doença em estado inicial de formação pode representar uma variação muito baixa de temperatura, não perceptível de forma visual;
- iv. A falta de medidas, métricas e indicadores bem definidos para a análise dos termogramas: ainda não existem padrões na análise de termogramas, ficando a critério do médico como avaliá-los;
- v. E por fim, fatores humanos, como o cansaço visual e ainda a inexperiência do médico na interpretação dos termogramas.

Para exemplificar essas dificuldades, na Figura 3 são apresentados dois termogramas, um de uma paciente saudável e outro de uma paciente com uma alteração na mama esquerda. É possível perceber a semelhança na distribuição das cores nas mamas de ambas as pacientes nos termogramas.

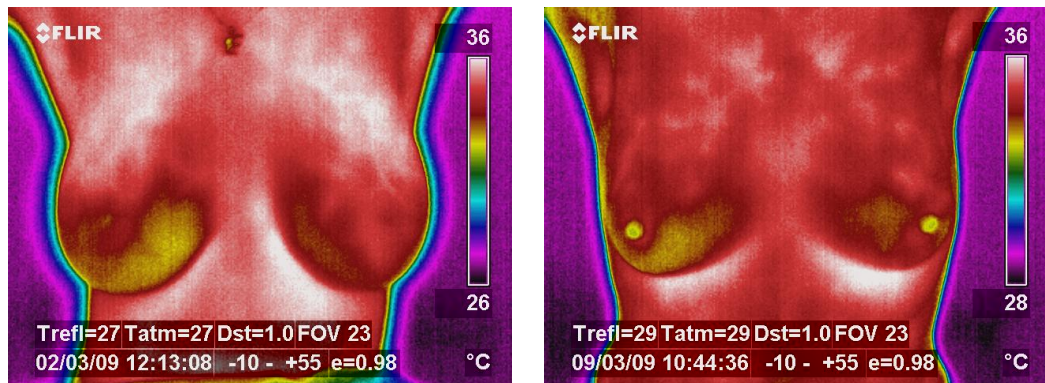


Figura 3: Termogramas: paciente com fibroadenoma (esq.) e paciente saudável (dir.).

Metodologias utilizando técnicas de processamento e análise de imagens podem ser desenvolvidas para facilitar a detecção de alterações da mama por imagens termográficas, com o intuito de verificar as possibilidades do uso da termografia como exame de triagem na mastologia.

1.2 OBJETIVOS

A seguir serão listados os objetivos deste trabalho:

- Realizar uma revisão bibliográfica sobre o uso da termografia da mama;
- Desenvolver uma técnica de registro de imagens para correção da posição corporal das pacientes nas imagens termográficas;
- Desenvolver uma técnica para ajuste automático do contraste nos termogramas;
- Desenvolver uma metodologia de auxílio à classificação de alterações da mama utilizando imagens termográficas;
- Agregar de forma eficiente e utilizável os projetos desenvolvidos anteriormente pelo grupo de pesquisa;
- Selecionar e utilizar um classificador para classificar os termogramas das mamas como saudáveis ou com suspeita de alterações;
- Utilizar e adaptar diversos métodos computados sobre valores de temperaturas, baseados em: medidas estatísticas, histogramas, geoestatísticas e fractal;
- Otimizar os resultados da classificação aplicando uma técnica de seleção de características;
- Desenvolver um sistema CADe (*Computer-Aided Detection*) com as metodologias desenvolvidos.

1.3 ESTRUTURA DA TESE

Esta tese está dividida em seis capítulos principais. Neste capítulo introdutório foi apresentado o tema sobre o qual se foca este trabalho: o uso de técnicas de processamento de imagens e mineração de dados sobre termogramas com o objetivo de auxiliar na detecção precoce de doenças da mama. Também foram apresentadas a motivação e os objetivos desta tese.

No Capítulo 2 são apresentados conceitos básicos para o entendimento desta tese. Serão abordados conceitos relacionados à mama, que é o objeto de estudo, como a sua estrutura interna, como ocorre o surgimento de neoplasias, quais as principais doenças e os principais exames utilizados atualmente para a detecção destas doenças. Em outro tópico será abordada a termografia como exame mastológico, será exemplificado como as imagens termográficas da mama são capturadas e armazenadas, além de como estas imagens são geradas pelas câmeras e processadas por algoritmos computacionais. Ainda no Capítulo 2 serão abordados os classificadores utilizados nesta tese e as técnicas de avaliação de resultados.

O Capítulo 3 apresenta um resumo da ampla revisão bibliográfica realizada no *survey* de Borchardt *et al.* [13], onde mais de 150 trabalhos relacionados ao uso de termografia da mama foram analisados. Estes trabalhos estão separados conforme suas aplicações e esta separação engloba trabalhos relacionados com a (i) aquisição e armazenamento dos termogramas, (ii) o pré-processamento e segmentação das regiões de interesse, (iii) a extração de características, (iv) o uso de classificadores e técnicas de avaliação e (v) a modelagem computacional.

No Capítulo 4 é descrita a metodologia desenvolvida nesta tese. Neste capítulo serão abordadas todas as etapas cumpridas para o desenvolvimento de uma metodologia de auxílio a classificação de alterações nas mamas. Estas etapas abrangem a (i) aquisição dos termogramas, (ii) a seleção da amostra utilizada, (iii) o pré-processamento da amostra, (iv) o registro para a correção da posição corporal das pacientes, (v) a segmentação das regiões de interesse, (vi) extração de características, (vii) a classificação e (viii) a seleção das melhores características. No passo (vi) são descritos todos os métodos computacionais investigados para detectar alterações nas mamas baseados nas suas características térmicas. Os métodos foram agrupados em quatro grupos distintos de acordo com suas especificidades. O Grupo I reúne características baseadas em estatística simples. No Grupo II estão agrupadas as características baseadas no histograma do termograma. No Grupo III agrupam-se as

características baseadas na dimensão fractal. O Grupo IV apresenta as características baseadas em geoestatísticas adaptadas para o uso em termogramas.

No Capítulo 5 são apresentados os testes realizados com a metodologia descrita no Capítulo 4 utilizando as técnicas de avaliação descritas no Capítulo 2. Será apresentada a metodologia utilizada para a realização dos testes e os resultados obtidos em cada um. Também serão apresentados os resultados obtidos com o uso da seleção de características para obter a melhor combinação de características para a classificação de alterações nas mamas. Por fim, os resultados serão analisados.

No último capítulo, são apresentadas as conclusões desta tese, as contribuições, os possíveis trabalhos futuros e as publicações alcançadas durante o seu desenvolvimento.

CAPÍTULO 2 – CONCEITOS BÁSICOS

Neste capítulo são apresentados alguns conceitos necessários para o entendimento desta tese. São abordados primeiramente conceitos relacionados à região da mama, suas doenças e os principais exames utilizados para o diagnóstico, incluindo a termografia.

Em seguida, são abordados conceitos relacionados ao uso de técnicas computacionais no auxílio à detecção e ao diagnóstico de doenças da mama. São vistos conceitos de extração de características, classificadores de padrões e técnicas de avaliação de resultados.

2.1 MAMA

2.1.1 ESTRUTURA DA MAMA

A mama, também chamada de glândula mamária, é composta por tecido conectivo, adiposo e tecido mamário que contém as glândulas que produzem leite. A glândula mamária, como mostrado na Figura 4, é composta por lobos e dutos (tubos por onde o leite passa). Cada mama apresenta de 15 a 20 lobos, que contêm numerosos pequenos lóbulos [14]. Os lóbulos têm conectados às suas extremidades dezenas de pequenos bulbos com a finalidade de produzir leite. Tubos finos, chamados dutos, conectam os lobos, lóbulos e bulbos.

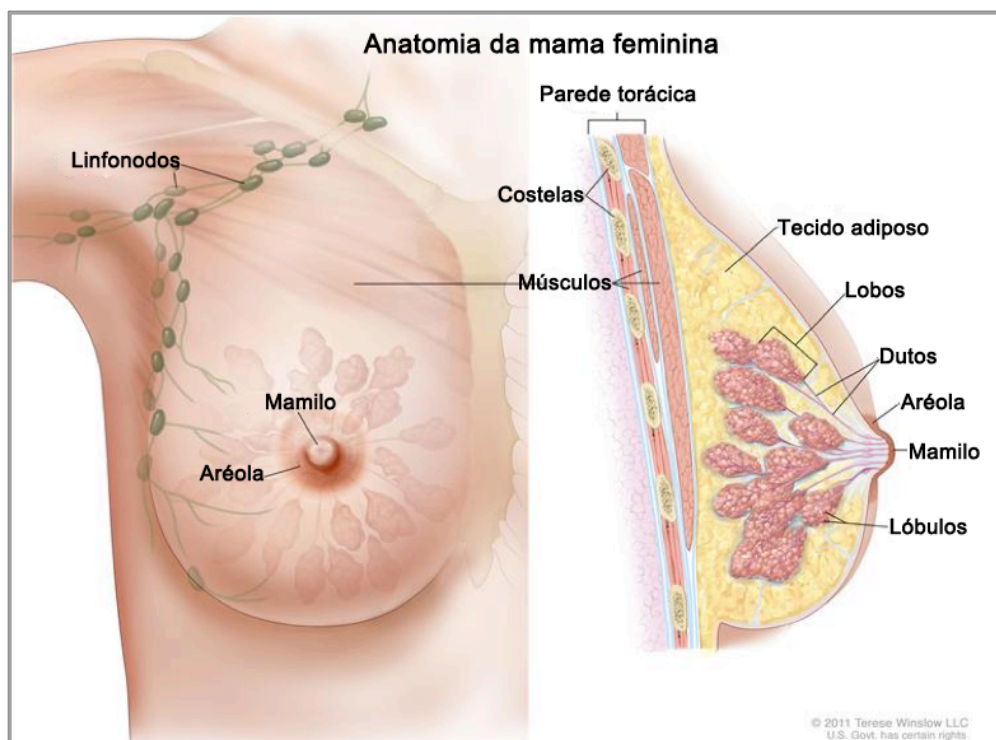


Figura 4: Anatomia da mama feminina (Fonte: Winslow [15]).

Cada mama apresenta ainda vasos sanguíneos e linfáticos. Os vasos linfáticos transportam um líquido transparente responsável por auxiliar as células no combate a infecções e outras doenças. Os vasos linfáticos se unem em uma massa de tecido linfático recoberta por uma camada de tecido conectivo, formando assim os linfonodos, que filtram a linfa (fluido linfático) e armazenam os linfócitos (células brancas do sangue) [16]. Os linfonodos são localizados em toda a extensão dos vasos linfáticos e também são chamados de glândulas linfáticas. Eles são encontrados em grupos em todas as partes do corpo, sendo cada grupo responsável por auxiliar as células contra infecções e doenças em determinada região [16]. Os linfonodos associados às glândulas mamárias estão localizados principalmente na região axilar, subclavicular e no peito.

2.1.2 O SURGIMENTO DE NEOPLASIAS

As neoplasias ou tumores são proliferações descontroladas de células no corpo causadas por mutações genéticas, onde células sofrem alterações no seu código genético e perdem sua função essencial passando a se reproduzir exageradamente. O excesso de reprodução celular no local exige a maior quantidade de nutrientes para a região, assim novos vasos sanguíneos vão sendo criados ao redor das células para alimentá-las. Este processo chamado de angiogênese é natural em situações como cicatrizações de cortes, onde a criação dos novos vasos sanguíneos faz com que a região receba um fluxo maior de sangue [17].

Na formação de neoplasias, a angiogênese auxilia a alimentação das células mutantes, permitindo que elas formem tecidos e posteriormente tumores. A Figura 5 apresenta as fases do surgimento de uma neoplasia, desde o surgimento das primeiras células mutantes até a angiogênese para suprir a demanda de nutrientes na região [18]. A primeira imagem da Figura 5 representa um tumor em fase pré-angiogênese, onde as células tumorais necessitam de suprimento de sangue para crescer. Esse suprimento é solicitado ao organismo e quanto mais nutriente o organismo prover para essas células, mais agressivo será o crescimento do tumor. Na segunda imagem é representada a fase inicial da angiogênese, onde é possível ver vasos capilares terem seu crescimento atraído em direção ao tumor para alimentá-lo. Isto promove um desenvolvimento rápido. A última imagem mostra a angiogênese em fase avançada, onde uma vasta rede de vasos sanguíneos alimenta o tumor, que começa a crescer de forma descontrolada.

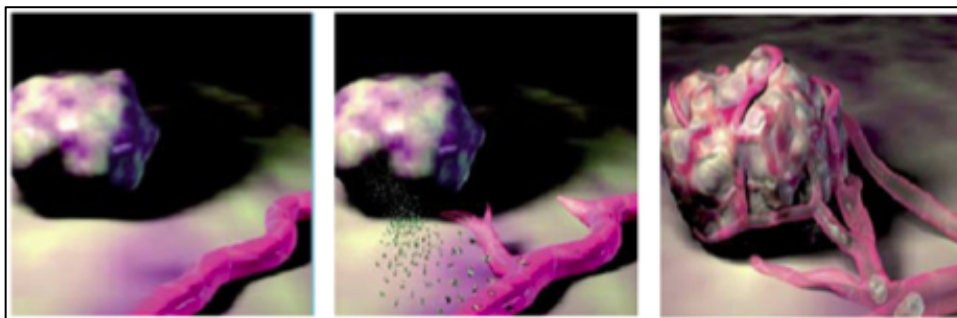


Figura 5: Etapas do surgimento de uma neoplasia (Fonte: Danum International [18]).

As neoplasias dividem-se em benignas e malignas, diferenciando-se pela agressividade e velocidade em que se multiplicam. As neoplasias benignas crescem lentamente e agredem pouco os tecidos vizinhos, enquanto que as neoplasias malignas aumentam de tamanho rapidamente e agredem as células que estão na mesma região buscando por nutrientes [17]. A seguir são descritas as principais doenças da mama.

2.1.3 PRINCIPAIS DOENÇAS DA MAMA

Entre as principais doenças encontradas na mama, principalmente feminina, estão o cisto, o fibroadenoma e o carcinoma, sendo o carcinoma o câncer de mama propriamente dito.

O cisto é uma modificação benigna na estrutura da mama, não sendo cancerígeno ou originário de células mutantes [19]. Cistos são frequentemente diagnosticados em mulheres, mas não representam risco. Um cisto é constituído por uma formação (solitária ou múltipla) de epitélio preenchido por substâncias líquidas ou material semissólido (plasma) [19].

O fibroadenoma é o tumor benigno mais comum em mamas. É formado pela proliferação celular dos tecidos epitelial e estromal (tecido conectivo). É mais comum em mulheres jovens. Manifesta-se através de um nódulo, geralmente indolor, com consistência firme e elástica, e raramente crescem mais do que 3 a 4 cm. Em muitos casos a paciente pode apresentar múltiplos nódulos [20].

O câncer de mama aparece usualmente nos dutos e nos lóbulos podendo ocorrer tanto em homens quanto em mulheres, entretanto casos de câncer de mama em homens são raros. O tipo mais comum de câncer de mama é o carcinoma ductal (CD), que começa nas células dos dutos [21]. O câncer que começa nos lobos ou lóbulos é chamado carcinoma lobular e frequentemente é encontrado em ambas as mamas, ao contrário dos outros tipos de câncer. O carcinoma lobular *in situ* (CLIS) ocorre quando células cancerígenas são encontradas somente nos lóbulos. Quando estas células cancerígenas se espalham a partir dos lóbulos para os

tecidos próximos é chamado de carcinoma lobular invasivo (CLI). Não é muito frequente a evolução de um CLIS para um CLI, entretanto, a existência de um CLIS em uma mama aumenta o risco do desenvolvimento de câncer na outra mama [21].

O câncer inflamatório é um tipo de câncer não tão frequente na mama quanto os outros carcinomas, mas é facilmente identificado através da termografia. No câncer inflamatório a mama apresenta-se vermelha, inchada e bastante quente [21]. A pele da mama pode também apresentar uma textura chamada *peau d'orange* (semelhante à casca de laranja). A vermelhidão e o aumento de temperatura ocorrem devido às células cancerígenas bloquearem os vasos linfáticos da mama.

A doença de Paget é um tipo de câncer de mama que se manifesta como uma lesão do mamilo ou como uma secreção. Esta doença geralmente produz pouco desconforto, o que leva a mulher a ignorá-la por um ano ou mais antes de procurar um médico [8].

2.1.4 IMAGENS MÉDICAS APLICADAS NA MASTOLOGIA

Os primeiros exames indicados para a detecção de alterações da mama são o autoexame e o exame clínico. No autoexame a mulher deve apalpar sua mama verificando a presença de alguma alteração, como secreções, caroços, abaulamentos ou retrações na pele e na regiãoaréolo-mamilar. No caso de alguma suspeita deve procurar um mastologista. É indicado realizar o autoexame ao menos uma vez ao mês [4].

O exame clínico é a palpação da mama, regiões axilares e supra-claviculares por um profissional em mastologia. Nesse exame poderão ser identificadas alterações e em caso de indícios ou suspeitas de doenças será indicado que a paciente realize um exame de auxílio ao diagnóstico por imagem [1].

Os exames baseados em imagem para o auxílio ao diagnóstico de doenças da mama podem ser classificados como estruturais ou funcionais. Os exames estruturais são aqueles que permitem a visualização de estruturas internas da mama, nesta categoria destacam-se a mamografia, o ultrassom e a ressonância magnética. Exames funcionais são os que permitem a visualização do funcionamento dos órgãos e o fluxo de líquidos, como exemplos têm-se o ultrassom (exame tanto estrutural como funcional), a ressonância magnética quando realizada com a utilização de um contraste e a termografia.

Serão detalhados a seguir os principais exames por imagem utilizados para a detecção e diagnóstico de doenças na mama.

2.1.4.1 Mamografia

Na mamografia a paciente tem sua mama comprimida pelo mamógrafo e uma radiação ionizante (feixes de Raios-X) é incidida sobre a mama, geralmente, em duas direções: verticalmente (exposição craniocaudal – CC) e angular (podendo ser médio-lateral – ML – ou médio-lateral oblíqua – MLO), como pode ser visto na Figura 6.

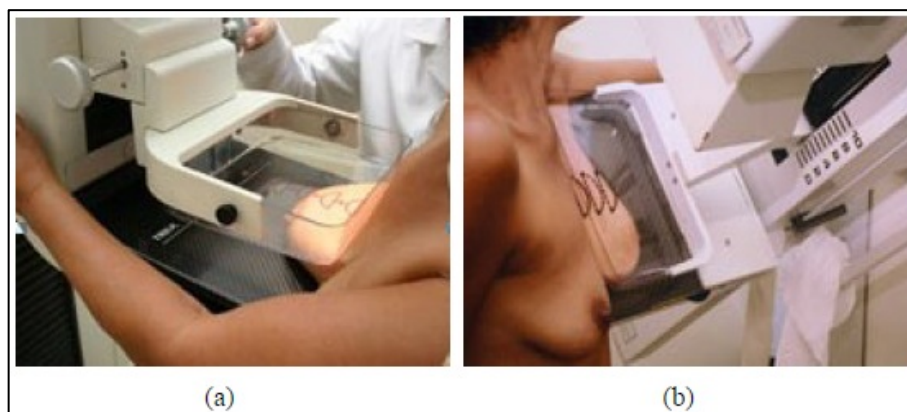


Figura 6: Mamografia: (a) Craniocaudal e (b) Médio-lateral Oblíqua (Fonte: Nagy [22]).

Os feixes de Raios-X emitidos contra a mama são retidos em partes pelos tecidos mamários, sendo esta quantidade proporcional à densidade do tecido. Por este motivo a mamografia não é indicada para pacientes jovens que tenham as mamas densas, pois os tecidos mamários irão reter a maior quantidade da radiação emitida contra a mama, dificultando a visualização de possíveis calcificações [9]. Em mamas pouco densas, como no caso de pacientes com idade superior a 40 anos ou pacientes com bastante tecido adiposo na região das mamas, a mamografia é o método mais indicado, pois identifica com boa precisão as estruturas internas da mama, revelando a presença de calcificações e microcalcificações [9]. A Figura 2 apresenta a diferença da mamografia aplicada em mamas de diferentes densidades.

O uso de próteses de silicone pela paciente também torna difícil à aquisição de imagens com qualidade para o diagnóstico através da mamografia, principalmente pelo risco em comprimir a mama com silicone e pelo fato do silicone “esconder” possíveis estruturas da mama [23].

Atualmente, através da Lei 11.664, de 2008, toda mulher a partir dos 40 anos pode realizar o exame de mamografia gratuitamente pelo Sistema Único de Saúde (SUS) [24].

2.1.4.2 Ressonância magnética

A ressonância magnética (RM) está cada vez mais sendo utilizada como método adjunto aos métodos convencionais de diagnóstico de doenças da mama. A resolução elevada das imagens, em relação aos outros exames, faz com que a RM alcance ótima sensibilidade na detecção de câncer de mama e lesões que poderiam ficar ocultas em exames tradicionais como a mamografia [25]. O maior empecilho para a difusão da RM é o custo ainda bastante elevado do equipamento que reflete em um custo alto para a realização do exame.

Na Figura 7 podem ser vistos o exame sendo aplicado (a) e um exemplo de imagem adquirida pelo método de RM.



Figura 7: Ressonância Magnética: (a) realização do exame e (b) imagem adquirida (Fontes: Massada [25] e Alvares e Michell [26], respectivamente).

Durante o exame, a paciente entra no aparelho. Para analisar o fluxo sanguíneo, é aplicado um *contraste*, substância líquida que é manipulada de forma intravenosa. O contraste desloca-se pelo sistema vascular enquanto o aparelho captura imagens do corpo, mostrando assim o caminho por onde o sangue circula. A RM gera um campo magnético que orienta os núcleos dos átomos de hidrogênio presentes no corpo humano. Em seguida são aplicadas ondas de rádio de frequência variável na região de interesse. Quando a frequência dessas ondas coincide com a frequência natural dos núcleos, ocorre uma absorção de energia por parte do núcleo. Quando o sinal é removido, os núcleos emitem esta energia na forma de sinais fracos, que podem ser detectados pelo aparelho e reproduzidos na imagem [27].

Entre as limitações da RM estão o alto custo da realização do exame, a baixa especificidade para diferenciar tumores malignos e benignos [27], a necessidade da administração de um contraste intravenoso e a ausência de um método padronizado. Geralmente são adquiridas as sequências conhecidas como T1 e T2, com supressão do tecido adiposo, o que causa um aumento no tempo de aquisição do exame [28].

2.1.4.3 Ultrassonografia

A ultrassonografia (US) baseia-se na emissão, reflexão, refração e captação de ondas sonoras de alta frequência para visualizar o interior do corpo. Na mastologia a ultrassonografia é utilizada para diagnosticar a presença de nódulos, sendo capaz de mostrar as estruturas internas da mama e o fluxo sanguíneo. É um exame altamente dependente do profissional que está aplicando, pois é ele quem determina qual o local que será visualizado.

Atualmente é possível realizar o exame de ultrassonografia automática 3D da mama. Este exame fornece a anatomia completa da mama para posterior análise e possibilita ao médico analisar os dados adquiridos por diferentes ângulos e campos de visão, esta possibilidade torna mais fácil a identificação de problemas que na US convencional. Entre outras vantagens do uso da ultrassonografia automática 3D estão a não dependência do profissional que está aplicando e a possibilidade de reproduzir o exame com as mesmas características quando necessário. Entretanto, o grande número de imagens geradas a cada exame aumenta o tempo de análise e interpretação do médico [29]. Na Figura 8 pode ser observada a realização de um exame de US da mama convencional e a ultrassonografia automática 3D.

A US é aconselhada para pacientes com menos de 35 anos, que apresentam as mamas densas e/ou com próteses de silicone, casos onde a mamografia tem sua qualidade prejudicada.



Figura 8: Realização da ultrassonografia da mama: convencional (esq.) e 3D automática (dir.) (Fonte: Gann [30]).

2.1.4.4 Tomossíntese

A tomossíntese da mama é uma técnica de Raio-X recente, onde diversas imagens são adquiridas em diversas direções, não apenas nas direções CC e ML. Durante o exame a mama da paciente permanece imóvel enquanto o canhão de Raio-X gira ao seu redor, disparando

emissões de baixa dosagem em uma gama de ângulos. Geralmente o canhão gira entre 10° e 20° a cada emissão e efetua uma média de 10 a 20 disparos a cada 3° durante o exame. A radiação emitida a cada disparo de Raio-X é equivalente a 5% da radiação utilizada na mamografia convencional [31].

As séries de imagens adquiridas são utilizadas para recriar em 3D o tecido mamário, esta reconstrução facilita a análise do especialista e praticamente elimina os problemas decorrentes de tecidos sobrepostos na mamografia.

O processo de reconstrução das imagens de tomossíntese consiste em analisar as imagens em alta resolução que se encontram paralelas ao plano de suporte da mama. Essas imagens costumam ser reconstruídas em fatias separadas por 1 mm, assim, 5 cm de mama comprimida na tomossíntese pode gerar 50 fatias. A reconstrução pode ser comparada às técnicas utilizadas em tomografia computadorizada (CT) [31].

As vantagens da tomossíntese estão na solução de casos que na mamografia são considerados sobreposição de tecidos. Este problema leva a um falso diagnóstico de microcalcificações ou massas. Como a tomossíntese adquire um número maior de imagens em diferentes ângulos é possível descartar as sobreposições através da análise por outro ponto de vista.

2.1.4.5 Mamógrafo por emissão de pósitrons

O mamógrafo por emissão de pósitrons (PEM – *positron emission mammography*) é uma versão do aparelho de tomografia computadorizada para o corpo inteiro (PET/CT – *positron emission tomography/computer tomography*) dedicado apenas para a região mamária. Este exame ainda encontra-se em fase de testes no Brasil, sendo indicado principalmente para pacientes onde a mamografia não é indicada, casos onde a paciente apresenta as mamas densas, próteses de silicone ou tendência cancerígena devida a mutações genéticas hereditárias [32].

O PEM analisa as mamas através do metabolismo da glicose. Como tumores necessitam de grande quantidade de energia para crescer, consomem um nível maior de açúcar e destacam-se nas imagens geradas pelo exame. Portanto, é indicado que a paciente não ingira carboidratos ao menos 24 horas antes da realização do exame. No momento do exame é injetado uma dose de glicose na corrente sanguínea. O percurso realizado pela glicose é acompanhado pelo equipamento e os locais onde a glicose tende a se concentrar para ser consumida indicam uma possível neoplasia [33].

2.1.4.6 Tomografia por impedância elétrica

As propriedades elétricas variam de material para material, e a proposta da tomografia por impedância elétrica (TIE) é produzir uma imagem do corpo de tal forma que, após o processamento e a reconstrução computacional da imagem, cada "pixel" desta represente a impedância ou a condutividade elétrica dos pontos internos do corpo.

Na TIE, o processo de produção da imagem consiste em primeiramente obter as informações referentes à condutividade elétrica de cada tecido. Para isto, são posicionados eletrodos ao redor da região de interesse e é aplicado um estímulo, que pode ser uma injeção de corrente elétrica ou a aplicação de uma diferença de potencial, para depois medir a resposta obtida do outro lado do corpo, também na forma de corrente ou potencial. Com base nos dados de estímulo e resposta, pode-se estimar a distribuição de condutividade elétrica e reconstruir computacionalmente a imagem do interior do corpo baseada nesta condutividade de cada tecido. As correntes ou tensões aplicadas são de baixa intensidade de forma a não produzir nenhum dano ou sensação indesejável ao paciente.

2.2 TERMOGRAFIA

As imagens por infravermelho não usam radiação ionizante, acesso venoso ou qualquer outro procedimento invasivo. Além desses fatos, é indolor e não tem contato com a superfície da pele. A termografia não provoca qualquer desconforto para a paciente. É um exame muito barato, quando comparado com os métodos tradicionais, tais como mamografia, ultrassom e ressonância magnética.

A termografia também tem vantagens para a detecção de problemas em mamas de mulheres jovens, porque os tecidos densos apresentam dificuldade na visualização por Raios-X. Microcalcificações e massas são geralmente bem visíveis em mamografias apenas de mulheres em idade não reprodutiva onde as mamas apresentam grande parte de tecido adiposo. Nas últimas duas décadas, a termografia atingiu valores de sensibilidade e especificidade média de cerca de 90% para a detecção de tumores de mama [34]. Um termograma anormal é considerado indicador de risco biológico significativo para a existência ou desenvolvimento de tumores mamários [35].

Existem trabalhos que apostam no potencial da termografia, a ponto de pesquisarem sua capacidade de detectar o câncer de mama até dez anos antes do que os métodos tradicionalmente utilizados [34]. Esta afirmativa é embasada no fato de que as primeiras

mutações celulares para a angiogênese de um tumor, teoricamente, devem provocar um aumento na temperatura da região onde está ocorrendo o fenômeno.

A termografia é um exame fisiológico funcional, diferente da mamografia que é um exame estrutural. As imagens são adquiridas através de uma câmera capaz de medir simultaneamente a temperatura em diversos pontos, detectando a radiação infravermelha emitida pelos corpos que estão na cena. A termografia é indicada para pacientes de qualquer idade, não tendo contraindicações e podendo ser repetida quantas vezes forem necessárias, servindo como um exame complementar para a detecção de alterações.

2.2.1 IMAGENS TERMOGRÁFICAS

As imagens termográficas, também conhecidas como termogramas são adquiridas por uma câmera térmica que é sensível à radiação infravermelha. Seus sensores captam a radiação térmica naturalmente gerada por objetos que apresentam temperatura acima do zero absoluto [36]. Todos os objetos que têm temperaturas superiores a este valor (0K ou -273°C) emitem radiação infravermelha a partir da sua superfície [37].

O espectro eletromagnético é dividido em regiões, chamadas bandas, de acordo com o comprimento de onda. Estas bandas são definidas de acordo com os métodos utilizados para produzir e detectar a radiação. Entretanto, todas as bandas sofrem as ações das mesmas leis da física, sendo a única diferença entre elas o comprimento das ondas. A termografia utiliza o espectro de banda infravermelho. Este espectro abrange desde as ondas curtas, próximas ao limite do espectro visível, até ondas longas que se misturam com o comprimento de micro-ondas. A radiação infravermelha abrange aproximadamente os comprimentos de onda entre $0,75\mu\text{m}$ e $100\mu\text{m}$. O corpo humano emite ondas de aproximadamente $9,3\mu\text{m}$ [38]. A Figura 9 mostra uma representação do espectro eletromagnético onde é possível visualizar os limites da radiação infravermelha.

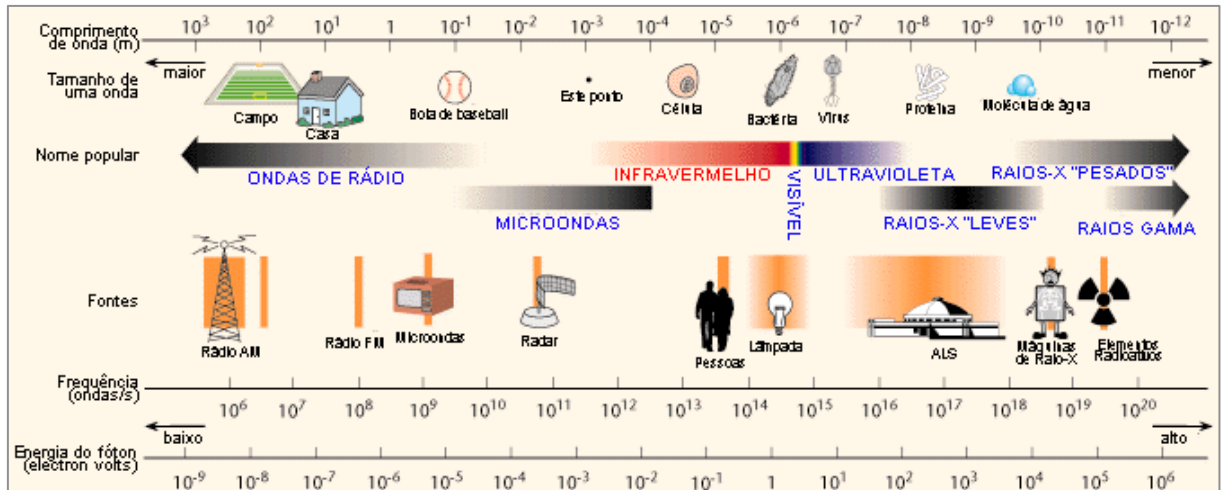


Figura 9: Representação do espectro eletromagnético (Fonte: Berkeley Lab [39]).

A relação entre a radiação total emitida por segundo de um objeto (W) e a sua temperatura absoluta (T) é descrita pela lei de Stefan-Boltzmann [37]. A radiação emitida é diretamente proporcional a quarta potência da temperatura, como pode ser observado na Equação (1).

$$W = A\sigma\epsilon T^4 \quad (1)$$

onde A é a área analisada em m^2 , σ é a constante de proporcionalidade ou constante de Stefan-Boltzmann que equivale a $5,670400 \times 10^{-8} \text{ Js}^{-1}\text{m}^{-2}\text{K}^{-4}$, ϵ é a emissividade do corpo e T é a temperatura absoluta do corpo em Kelvin. A radiação emitida é medida em Joules por segundos por metros quadrados (J/sm^2). A emissividade ϵ é particular a cada material e representa a razão da energia emitida pelo material em relação a energia emitida por um corpo negro (objeto completamente preto capaz de absorver completamente toda radiação recebida) a mesma temperatura. A emissividade pode assumir valores entre 0 e 1. Próximo a zero tem-se, por exemplo, um espelho altamente polido e, próximo a um, materiais opacos. A pele humana tem emissividade de aproximadamente 0,98. A constante de Stefan-Boltzmann é calculada através da Equação (2).

$$\sigma = \frac{2\pi^5 k^4}{15c^2 h^3} \quad (2)$$

onde k é a constante de Boltzmann, c é a velocidade da luz no vácuo e h é a constante de Planck.

A lei de Stefan-Boltzmann é utilizada pelos sistemas de câmeras térmicas para aferir o valor de temperatura através da captura da radiação infravermelho (IR) emitida pela superfície analisada. Entretanto, como as capturas não são realizadas no vácuo, é necessário compensar

a influência da atmosfera e de outros objetos emissores na cena. A Figura 10 representa um cenário real de aquisição de imagem termográfica, onde a câmera recebe três radiações principais: (i) a emitida pelo objeto ($\varepsilon\tau W_{obj}$), onde ε é a emissividade do objeto e τ é o coeficiente de transmitância térmica da atmosfera, (ii) a radiação refletida de outras fontes no ambiente ($(1-\varepsilon)\tau W_{refl}$), onde $(1-\varepsilon)$ é a refletância do objeto, e (iii) a radiação emitida pela atmosfera ($(1-\tau)W_{atm}$), onde $(1-\tau)$ é a emissividade da atmosfera. Assim, a radiação total absorvida pela câmera é $W_{tot} = \varepsilon\tau W_{obj} + (1-\varepsilon)\tau W_{refl} + (1-\tau)W_{atm}$.

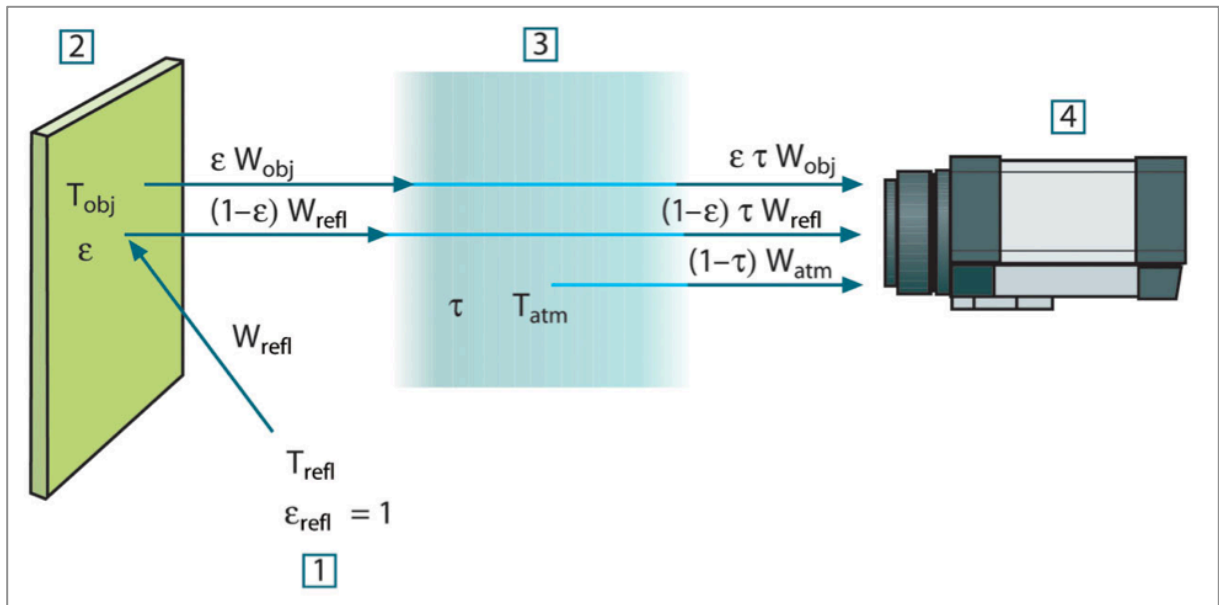


Figura 10: Representação da aquisição IR: (1) ambiente, (2) objeto de interesse, (3) atmosfera e (4) câmera (Fonte: Flir Systems [40]).

Em suma, para calcular a temperatura do objeto de interesse, deve-se subtrair da radiação total capturada pela câmera a radiação refletida do ambiente e a radiação da atmosfera. Devido a isso, durante a captura de uma imagem termográfica é necessário indicar uma série de parâmetros extras para suplantar esse cálculo, tais como: (i) a emissividade do objeto, (ii) a umidade relativa do ar, (iii) a temperatura atmosférica, (iv) a distância do objeto, e (v) a temperatura de objetos considerados fontes de calor presentes no ambiente. O primeiro parâmetro é dependente do objeto analisado. Os parâmetros de ii à v devem ser medidos e configurados na câmera no momento da captura. Caso não exista nenhuma fonte de calor considerável no ambiente, deve-se configurar este parâmetro com o valor da temperatura atmosférica.

2.3 CLASSIFICADORES

O aprendizado de máquina é uma subárea da inteligência artificial relacionada à capacidade do computador aprender e aprimorar seu desempenho em determinada tarefa a partir da repetição da mesma. Há muito tempo o aprendizado de máquina é utilizado em aplicações financeiras, detecção de *spams* em e-mail e detecções de fraudes em cartões de crédito. Nas últimas décadas, o aprendizado de máquina passou a ser utilizado também na área de imagens médicas, sendo amplamente aplicado no auxílio à detecção e diagnóstico de doenças, na recuperação de imagens baseada em conteúdo, no aprimoramento da qualidade das imagens e no monitoramento de sinais cerebrais [41].

Quando o aprendizado de máquina é utilizado no auxílio à detecção ou diagnóstico de doenças são empregados classificadores que modelam técnicas de classificação, também conhecidas como técnicas de tomada de decisão.

As técnicas de classificação podem ser divididas em três abordagens diferentes [42]:

- Supervisionada: onde é fornecido um conjunto de entrada e saída do qual será identificado um padrão comportamental;
- Não supervisionada: onde é fornecido apenas o conjunto de entrada, procurando identificar agrupamentos naturais no conjunto;
- Por reforço: onde há uma interação com o ambiente, recebendo bônus em caso de acerto, ou ônus em caso de erro.

Nesta tese foi utilizada a classificação supervisionados. A classificação supervisionada visa identificar a qual classe um determinado registro pertence. Nesta tarefa, primeiramente o classificador recebe um conjunto de registros fornecidos na forma (x_i, y_i) , onde x_i representa o conjunto de características ou atributos do registro e y_i indica à qual classe pertence, a fim de ‘aprender’ como classificar um novo registro desconhecido (x_i, y_i) , que não sabe-se a classe à qual pertence, com base neste primeiro aprendizado.

Cada registro é representado por um vetor contendo uma série de atributos ou características. Dentro de um conjunto de registros é possível haver falhas principalmente chamadas de ruídos e *outliers*. Os ruídos representam dados incorretos, enquanto que os *outliers* são dados muito distintos dos demais presentes no conjunto [43]. Os ruídos e *outliers* podem influenciar no desempenho do classificador, então é importante eliminá-los do conjunto de registros (corrigindo ou removendo esses registros) ou adotar alguma técnica de classificação robusta o suficiente para minimizar a influência provocada por eles. Nesta tese os ruídos, sempre que detectados, foram corrigidos e os *outliers* analisados caso-a-caso.

A Figura 11 mostra de forma genérica a fase de aprendizado de um classificador supervisionado. É necessária uma base de treino composta por diversos registros, onde as classes as quais pertencem os registros são conhecidas. De cada registro são extraídas características organizadas em vetores identificados pela classe a qual pertence o registro. A partir dos exemplos e suas respectivas classes, o algoritmo de aprendizado de máquina presente no classificador constrói um modelo, que contém uma descrição compacta e funcional dos dados fornecidos. Este modelo ou função será utilizado para classificar futuros registros, nos quais as classes sejam desconhecidas.

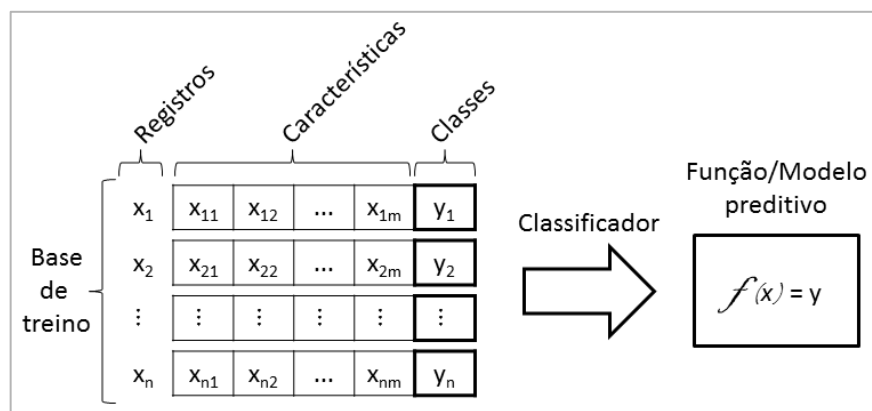


Figura 11: Aprendizado supervisionado.

Para quantificar e qualificar o modelo preditivo gerado pelo classificador são utilizadas técnicas de avaliação que serão discutidas na Seção 2.4. A seguir será abordado o classificador utilizado nesta tese.

2.3.1 SELEÇÃO DO CLASSIFICADOR

Nos trabalhos de Serrano [44] e Borchardt *et al.* [45] mais de 70 classificadores diferentes foram testados para a classificação de alterações em imagens termográficas utilizando características similares as que serão utilizadas nesta tese. Todos os testes foram realizados utilizando o *software Weka* [46].

Dentre todos os classificadores testados, os 10 que obtiveram melhores resultados foram: *support vector machine* (SVM), *naive Bayes simple* (NBS), *rotation forest* (RF), *logitboost alternating decision tree* (LADTree), *classification via clustering* (CVC), *naive Bayes* (NB), *naive Bayes updateable* (NBU), *logistic model trees* (LMT), *simple logistic* (SL), *instance-based classifier KStar* (KS).

Com base nos resultados obtidos por estes testes optou-se por utilizar como classificador nesta tese o SVM. Entretanto, o uso dos outros classificadores e a combinação de múltiplos classificadores podem ser abordados em trabalhos futuros.

2.3.2 MÁQUINA DE VETOR DE SUPORTE (SVM)

A Máquina de Vetor de Suporte (*Support Vector Machine* – SVM) é uma técnica de aprendizado de máquina que busca uma classificação. A SVM foi apresentada primeiramente por Boser *et al.* [47] e atraiu bastante a atenção de pesquisadores da área de mineração de dados por obter resultados com altos índices de acurácia, podendo ser utilizada para modelar situações não-lineares complexas, gerando modelos preditivos interpretáveis.

O aprendizado realizado pela SVM é baseado na Teoria de Aprendizado Estatístico (TAE), proposta por Vapnik e Chervonenkis [48]. Essa teoria quantifica os registros classificados erroneamente e utiliza o princípio da indução para gerar um modelo preditivo que minimize o erro sobre os dados de teste até um limite aceitável.

A Figura 12 ilustra como a TAE pode auxiliar na construção de um modelo de classificação ótimo. O objetivo do classificador é obter um modelo que separe as classes “círculo” e “triângulo”, as hipóteses são representadas nas figuras por meio das bordas ou fronteiras de decisão traçadas entre as classes.

A Figura 12a classifica corretamente todos os registros, inclusive dois que parecem ser *outliers* ou ruídos, porém é uma hipótese complexa para ser representada. Por mais que seja possível estabelecer um modelo como este que acerte todos os casos, essa função tende a apresentar uma alta taxa de erros quando aplicada a registros novos, pois está definida de forma muito específica a base de testes.

A Figura 12c apresenta outra hipótese onde pontos de classes opostas que estão dispostos muito próximos são desconsiderados. Esta hipótese é facilmente construída, porém apresenta um elevado número de registros classificados erroneamente na própria base de treino.

A Figura 12b apresenta uma hipótese intermediária entre as duas anteriores. Este modelo classifica corretamente a maior parte dos registros sem se afixar em pontos individuais. Através do uso da TAE é possível descrever funções similares a esta hipótese, estabelecendo condições matemáticas que auxiliam na criação do modelo considerando a base de treino e a complexidade do modelo.

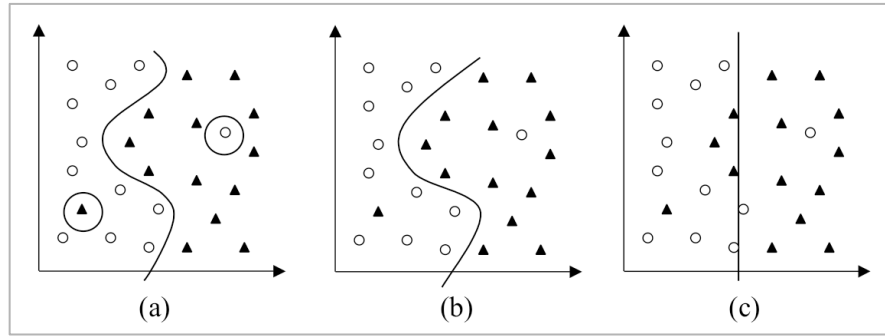


Figura 12: Três possíveis modelos de classificação de uma mesma base de treino com duas características (Fonte: Lorena e Carvalho [43]).

Alguns autores citam como desvantagens da SVM o tempo gasto no aprendizado e seleção de parâmetros para otimizar os resultados e os resultados podem ser afetados pela técnica utilizada para normalizar os dados de entrada [49], [50].

Um classificador SVM originalmente apresenta margens rígidas que definem fronteiras lineares, representadas por hiperplanos para separar as diferentes classes de dados, porém em aplicações reais nem sempre é possível separar de forma linearmente rígida as classes de um problema. Para isto foram propostas margens suaves para os hiperplanos, de modo a permitir que alguns elementos permaneçam entre os hiperplanos ocorrendo alguns erros de classificação.

Em muitos casos de classificação uma separação linear não é eficaz. As SVMs lidam com problemas não lineares mapeando o conjunto de treinamento do espaço original (X), para um novo espaço de maior dimensão (C). Seja $\Phi: X \rightarrow C$ um mapeamento, a escolha apropriada de Φ faz com que o conjunto de treinamento mapeado em C possa ser separado por uma SVM linear [43].

A classificação realizada pelas SVMs baseia-se na separação por hiperplanos, isto é, seja T uma base de treinamento com n registros $x_i \in X$ e suas respectivas classes $y_i \in Y$, em que X constitui o espaço dos registros e $Y = \{-1, +1\}$. A base T será linearmente separável se for possível traçar um hiperplano entre os registros pertencentes às classes -1 e $+1$ [43].

A equação de um hiperplano é representada na Equação (3), onde

$$f(x) = w \cdot x + b \quad (3)$$

Essa equação é capaz de separar o espaço que contém as amostras de X em duas regiões: $w \cdot x + b > 0$ e $w \cdot x + b < 0$. Desta forma, a aplicação de uma função sinal $g(x) = \text{sgn}(f(x))$ consegue distinguir as classes. Essa separação pode ser observada na Figura 13. O

princípio da SVM é maximizar a distância d entre os hiperplanos H_1 e H_2 , a descrição de como é realizada essa maximização pode ser vista com detalhes em Lorena e Carvalho [43].

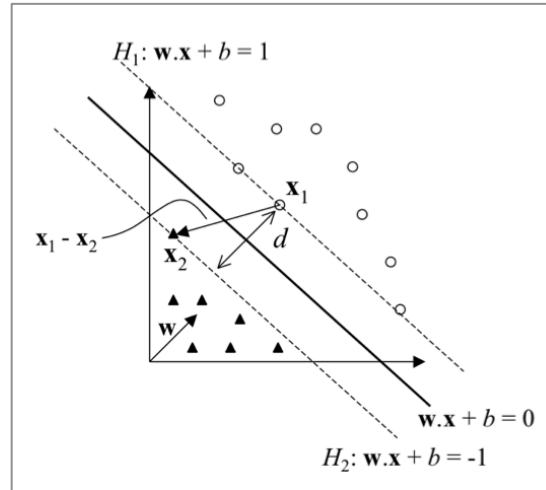


Figura 13: Separação por hiperplanos do classificador SVM (Fonte: Lorena e Carvalho [43]).

2.4 TÉCNICAS DE AVALIAÇÃO

Após a aplicação de classificadores para avaliar as características extraídas é necessário o uso de técnicas de avaliação de resultados para validar o modelo desenvolvido. A seguir serão descritas as principais técnicas utilizadas nesta tese.

2.4.1 VALIDAÇÃO CRUZADA

A validação cruzada é uma técnica de validação utilizada para assegurar que os resultados apresentados por um classificador são independentes do conjunto de dados utilizados. As quatro principais técnicas de validação cruzada são [51, p. 241]:

- (i) *k-fold*: onde a amostra inicial é dividida aleatoriamente em k sub-amostras, destas k sub-amostras $k - 1$ são utilizadas para treinamento e uma é utilizada para validação, repetindo esta rotina k vezes até que todas as sub-amostras sejam utilizadas como validação. Após as k iterações, a acurácia do modelo é calculada pela média dos resultados das k validações;
- (ii) *2-fold* ou *holdout*: é a forma mais simplificada do *k-fold*, onde a amostra inicial é dividida em apenas dois grupos, um utilizado para treinamento e outro para validação, geralmente adota-se 2/3 da amostra para treinamento e 1/3 para validação;

- (iii) Repetição de sub-amostragem aleatória: a amostra inicial é dividida aleatoriamente em dois grupos, treino e validação, e esta rotina é repetida quantas vezes se considerar necessário. No final de todas as iterações a acurácia é calculada pela média dos resultados de todas as iterações realizadas;
- (iv) *Leave-one-out*: é uma abordagem especial da técnica de *k-fold*, onde *k* é igual ao número total de elementos na amostra. Este método apresenta uma investigação completa da variação do modelo, porém tem um alto custo computacional.

2.4.2 MATRIZ DE CONFUSÃO

A matriz de confusão é uma tabela específica para avaliar o desempenho de um algoritmo ou método de classificação. As colunas da matriz de confusão representam classes previstas, enquanto que as linhas representam as classes reais dos elementos analisados [51, p. 301]. Quando se trabalha com uma abordagem de método de classificação, as classes previstas são associadas ao resultado da classificação.

As células da matriz de confusão que correspondem ao cruzamento das informações das classes previstas e das classes reais representam a quantidade de verdadeiros positivos (VP), verdadeiros negativos (VN), falsos positivos (FP) e falsos negativos (FN), como exemplifica a Tabela 1:

		Classes previstas	
		Sim	Não
Classes reais	Sim	VP	FN
	Não	FP	VN

Tabela 1: Matriz de Confusão

Em aplicações médicas geralmente define-se:

- VP: Casos doentes classificados corretamente como casos doentes;
- VN: Casos saudáveis classificados corretamente como casos saudáveis;
- FP: Casos saudáveis classificados erroneamente como casos doentes;
- FN: Casos doentes classificados erroneamente como casos saudáveis.

Diferentes classificadores ou até um mesmo classificador com diferentes configurações de parâmetros podem gerar diferentes matrizes de confusão para um mesmo conjunto de dados. Entretanto, quando estes valores são associados a doenças em que o

tratamento imediato aumenta a possibilidade de cura, como é o caso das doenças de mama, é preferível utilizar a configuração que minimize o número de falsos negativos, desde que possua uma acurácia significativa.

2.4.3 SENSIBILIDADE E ESPECIFICIDADE

Sensibilidade e especificidade são medidas estatísticas de desempenho de técnicas de classificação calculadas utilizando informações presentes na matriz de confusão. A **sensibilidade** mede a proporção de casos positivos classificados corretamente como casos positivos, enquanto que a **especificidade** mede a proporção de casos negativos classificados corretamente como casos negativos [51, p. 314]. Associando esta definição a aplicações médicas, tem-se que a sensibilidade é a capacidade de classificar corretamente casos com doenças, e a especificidade é a capacidade de classificar corretamente casos saudáveis.

A sensibilidade e a especificidade são calculadas conforme as Equações (4) e (5), respectivamente.

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

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

Quando o método de classificação utilizado apresenta um alto valor de sensibilidade, significa que um caso classificado como doente têm uma elevada probabilidade de realmente ser doente. Entretanto, se um método apresenta um alto valor de especificidade, um caso classificado como saudável têm uma elevada probabilidade de realmente ser saudável.

Um método com 100% de sensibilidade é capaz de reconhecer corretamente todos os casos com doenças e um método com 100% de especificidade é capaz de reconhecer todos os casos saudáveis. Porém, deve-se observar se o método é tendencioso. Métodos que classificam todos os elementos de uma amostra como doentes ou como saudáveis de forma errônea apresentarão respectivamente 100% de sensibilidade ou especificidade. Assim, o uso somente da sensibilidade ou da especificidade não é suficiente para determinar se o método é eficaz.

2.4.4 ACURÁCIA E PRECISÃO

Em estatística a **acurácia** consiste no grau de proximidade de um valor medido ou calculado em relação à sua definição real ou a sua referência padrão. Na análise de resultados

de mineração de dados pode-se entender a acurácia como o percentual de amostras classificadas corretamente considerando todas as classes. A **precisão** representa o grau de variação dos resultados obtidos pela classificação. A precisão é calculada de acordo com o desvio padrão de uma série de repetições da mesma análise. Resumindo, a acurácia é a probabilidade de obter resultados corretos, enquanto que a precisão é a probabilidade de repetir os mesmos resultados [52].

Em casos de teste de diagnóstico ou detecção de doenças a precisão também é conhecida como Valor Preditivo Positivo (VPP), que indica a proporção de resultados positivos obtidos que realmente são positivos, isto é, casos onde se obteve um diagnóstico correto. Com esta aplicação a precisão torna-se uma medida crítica para avaliar o desempenho de um método de diagnóstico ou detecção.

A acurácia e o VPP são computadas pelas Equações (6) e (7), respectivamente.

$$Acurácia = \frac{VP + VN}{VP + FP + VN + FN} \quad (6)$$

$$VPP = \frac{VP}{VP + FP} \quad (7)$$

Um método que apresenta 100% de acurácia garante que todos os elementos da amostra foram classificados corretamente, enquanto que um método com 100% de precisão garante que todos os casos classificados como positivos são realmente positivos. Assim, se o método apresentar baixa precisão significa que apresenta um alto número de falsos positivos, não sendo recomendado para o auxílio ao diagnóstico e detecção.

2.4.5 ÁREA SOB A CURVA ROC

As Curvas de Características de Operação do Receptor (Curvas ROC - *Receiver Operating Characteristic*) são uma forma eficiente de demonstrar a relação entre sensibilidade e especificidade. As Curvas ROC foram inicialmente desenvolvidas para demonstrar a relação sinal-ruído na telecomunicação através de ondas de rádio. O sinal era interpretado como sendo verdadeiro positivo e o ruído como falso positivo [53]. A Curva ROC basicamente é um gráfico que associa a sensibilidade com a taxa de falsos positivos (1 - especificidade) para uma série de testes, como pode ser visto no exemplo da Figura 14.

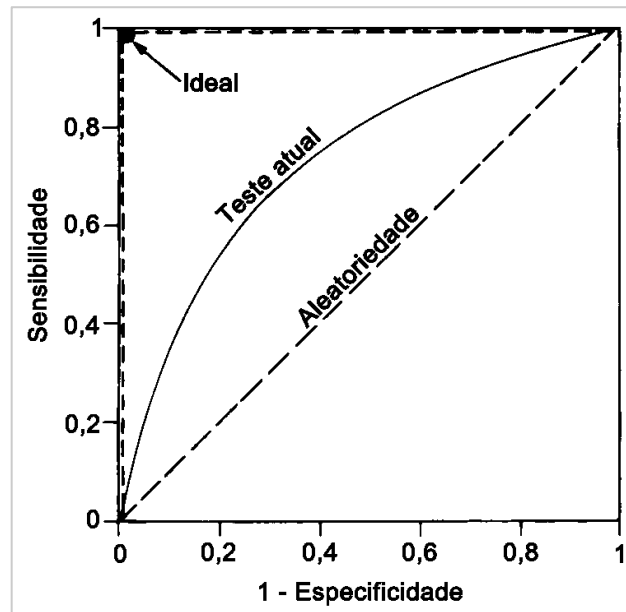


Figura 14: Exemplo de Curva ROC.

A linha diagonal tracejada que divide o gráfico representa uma técnica de classificação aleatória, onde cada caso tem 50% de chance de ser classificado correta ou incorretamente. Um classificador perfeito seria representado no canto superior esquerdo do gráfico, assim este apresentaria uma linha reta no topo indicando que para o ponto ótimo a sensibilidade é igual a um e os falsos positivos são zero. São poucos os problemas de classificação que permitem chegar a uma classificação perfeita, então um bom classificador deve apresentar a sua Curva ROC entre a linha diagonal e a linha reta no topo.

Um modo prático de avaliar a qualidade do classificador que se está usando é calcular a área sob a Curva ROC, esta área indica numericamente o quão próximo da classificação perfeita está o método utilizado. O valor da área varia entre 0 e 1 e é diretamente proporcional a proximidade com a classificação perfeita, caso o valor seja menor que 0,5 significa que uma classificação baseada na aleatoriedade daria melhores resultados.

2.4.6 ÍNDICE DE YODEN

O índice de Youden é uma medida simples para relacionar a sensibilidade com a especificidade, o cálculo deste índice segue a Equação (8).

$$Y = \text{Sensibilidade} + \text{Especificidade} - 1 \quad (8)$$

Este índice foi proposto em 1950 por Youden [54] como uma solução prática para relacionar a sensibilidade e a especificidade em testes diagnósticos. Este índice tem como objetivo medir a performance geral de testes diagnósticos.

2.5 RESUMO DO CAPÍTULO

Inicialmente, neste capítulo, foram abordados os conceitos básicos relacionados a mama, dentre os quais se destacam a estrutura da mama, a ocorrência e o surgimento de neoplasias, as principais doenças encontradas nas mamas e os exames por imagens utilizados na mastologia para a detecção destas doenças.

A seguir, explorou-se os conceitos relacionados a termografia, comparando este exame aos demais exames descritos anteriormente. Também abordou-se a formação das imagens termográficas e os fatores que as influenciam.

O terceiro ponto abordado neste capítulo foram os classificadores, com ênfase no SVM, que foi o classificador selecionado para ser utilizado nesta tese.

Por fim, as técnicas de avaliação foram descritas. As três técnicas de validação cruzada foram descritas na primeira subseção. Em seguida, foi explicada a matriz de confusão e todas as medidas baseadas nela que foram utilizadas nesta tese: sensibilidade, especificidade, acurácia, precisão, curva ROC e índice de Youden.

No próximo capítulo será apresentada a revisão bibliográfica.

CAPÍTULO 3 – REVISÃO BIBLIOGRÁFICA

Técnicas de processamento de imagens e de sinais são aplicadas em dados e imagens médicas como ferramentas para aprimorar a detecção e o diagnóstico de doenças. No artigo “*Breast thermography from an image processing viewpoint: a survey*” escrito durante o desenvolvimento desta tese, Borchardt *et al.* [13] realizaram uma análise crítica de trabalhos sobre o uso de imagens termográficas da mama. A seguir serão apresentados os principais trabalhos encontrados na literatura divididos em 3 áreas: (i) aquisição e armazenamento, (ii) segmentação da região de interesse e (iii) auxílio à detecção e diagnóstico de doenças.

3.1 AQUISIÇÃO E ARMAZENAMENTO DOS TERMOGRAMAS

Os termogramas são sensíveis a variações no ambiente, como por exemplo, a temperatura, a umidade e a circulação de ar na sala onde está sendo realizado o exame. Devido a isso, os exames devem ser realizados sob um protocolo adequado às condições térmicas do local. Um dos principais fatores de inconsistência entre diferentes trabalhos utilizando imagens termográficas da mama é a grande quantidade de protocolos diferentes utilizados para a aquisição das imagens. A seguir serão discutidos alguns protocolos de aquisição presentes na literatura, depois alguns trabalhos relacionados com o armazenamento dos termogramas.

Os protocolos de aquisição podem ser organizados quanto ao comportamento do corpo em relação à transferência de calor: estático ou dinâmico. Outra categorização pode ser empregada de acordo com a repetição da aquisição das imagens: simples, sequencial ou de acompanhamento. Na aquisição estática a paciente está em equilíbrio térmico com o ambiente. Na aquisição dinâmica é aplicado um monitoramento da recuperação da temperatura do corpo após um estresse térmico (aquecimento ou resfriamento) ou um estresse químico (vasodilatação ou vasoconstrição). Na aquisição simples, imagens da paciente são capturadas em um instante de tempo. Este tipo de aquisição é indicado para identificar pontos quentes e frios e medir a assimetria da distribuição de temperatura nas mamas das pacientes [55]. Na aquisição sequencial, uma série de imagens é adquirida sequencialmente durante alguns minutos, que podem ser configurados na própria câmera térmica. Na aquisição de acompanhamento, imagens são capturadas com uma diferença substancial de tempo (entre 3 a 6 meses) com o objetivo de monitorar o progresso de algum tratamento, doença ou para detecção precoce [55]. Alguns protocolos também recomendam as pacientes que antes da

realização do exame evitem o consumo de álcool, cafeína e cigarros, a realização de exercícios físicos e a aplicação de cremes e loções na área a ser examinada, pois estes fatores podem influenciar na medição da temperatura.

A Tabela 2 resume os principais protocolos encontrados na literatura.

Autores - Ano	Aquisição	Temperatura ambiente	Recomendações especiais
Koay <i>et al.</i> [56] - 2004	Simples/Dinâmico: A área das mamas é resfriada por um ventilador por aproximadamente 20 minutos.	22°C.	As pacientes são questionadas se evitaram álcool, cafeína, analgésicos, loções e se não fumaram nas últimas duas horas antes do exame.
Ng e Kee [57] - 2008	Simples/Estático: Paciente aguarda 20 minutos para estabilizar a temperatura do corpo.	20 a 22°C e umidade em 60%.	Redução de fontes de calor no ambiente durante a aquisição das imagens. Pacientes são recomendadas a não usar álcool, cigarros e produtos para a pele.
Arora <i>et al.</i> [58] - 2008	Sequencial/Dinâmico: Paciente permanece 4 minutos com ar frio direcionado nas mamas.	Informação desconhecida.	Informação desconhecida.
Agostini <i>et al.</i> [59] - 2009	Sequencial/Dinâmico: de 50 frames/s a 200 frames/s.	Informação desconhecida.	Informação desconhecida.
Wishart <i>et al.</i> [7] - 2010	Sequencial/Dinâmico: Paciente permanece 5 minutos com ar frio direcionado nas mamas.	Informação desconhecida.	Informação desconhecida.
Luna e Delgado [60] - 2010	Simples/Estático: Paciente aguarda 15 minutos para estabilizar.	18 a 23°C	Informação desconhecida.
Kapoor e Prasad [61] - 2010	Simples/Dinâmico: Paciente permanece de 7 a 10 minutos com ar frio direcionado nas mamas.	22°C.	Imagens capturadas em uma sala escura para minimizar as interferências do ambiente.
Amri <i>et al.</i> [62] - 2011	Simples/Estático: Paciente aguarda de 10 a 20 minutos para estabilizar a temperatura do corpo.	18 a 22°C.	As pacientes são recomendadas a não usar álcool, não praticar exercícios físicos e não passar produtos na pele.
Kontos <i>et al.</i> [63] -	Simples/Estático: Paciente	22°C.	Informação

2011	aguarda de 10 a 15 minutos para estabilizar a temperatura.		desconhecida.
Acharya <i>et al.</i> [64] - 2012	Simples/Estático: Paciente aguarda 15 minutos para estabilizar a temperatura corporal.	20 a 22°C (com variação de $\pm 0.1^\circ\text{C}$), e umidade em $60\% \pm 5\%$.	Informação desconhecida.
ACCT [65] - 2012	Acompanhamento/Estático : Os termogramas são adquiridos a cada 3 meses.	Informação desconhecida.	Informação desconhecida.
Bezerra <i>et al.</i> [66] - 2013	Simples/Estático: Paciente aguarda 10 minutos para estabilizar a temperatura corporal.	25 a 28°C.	Sem recomendações especiais.
Silva <i>et al.</i> [67] - 2013	Simples/Estático: Paciente aguarda 10 minutos até estabilizar a temperatura do corpo. Sequencial/Dinâmico: Paciente permanece com ar frio direcionado às mamas em torno de 5 minutos.	20 a 24°C.	As pacientes são questionadas se evitaram álcool, cafeína, analgésicos, loções e se não fumaram nas últimas duas horas antes do exame.

Tabela 2: Resumo dos principais protocolos de aquisição de termogramas da mama.

Como ainda não existe um protocolo bem definido e aceito unanimemente para a aquisição de imagens termográficas, cada clínica, hospital e grupo de pesquisa adquirem diferentes números de imagens com as pacientes em posições distintas em relação à câmera. Alguns grupos utilizam aparatos, equipamentos e sistemas dedicados a melhor posicionar as pacientes a fim de padronizar as aquisições. Outro ponto a ressaltar é a disponibilidade das imagens adquiridas. Para que as imagens das pacientes possam ser utilizadas em pesquisas científicas, alguns grupos pedem que elas assinem um termo de consentimento autorizando o uso e a publicação de forma anônima de seus termogramas. Também é necessário a submissão de projetos a comitês de éticas.

Na Tabela 3 são resumidos os principais trabalhos que relatam como são gerenciadas as imagens capturadas.

Autores - Ano	Número de imagens por paciente	Número de pacientes/Banco de dados	Equipamento dedicado	Termo de consentimento
Koay <i>et al.</i> [56] - 2004	Informação desconhecida.	19 pacientes - privado.	Não.	Informação desconhecida.
Ng e Kee [57] - 2008	3 imagens: 1 frontal, 2 laterais.	90 pacientes - privado.	Não.	Informação desconhecida.
Arora <i>et al.</i> [58] - 2008	100 imagens frontais por paciente.	92 pacientes - privado.	Sentinel BreastScan.	Sim.
Tang <i>et al.</i> [68]- 2008	Informação desconhecida.	117 pacientes - privado.	TSI thermographic system.	Informação desconhecida.
Agostini <i>et al.</i> [59] - 2009	Vídeo em posição frontal.	Informação desconhecida.	Mesa de exames com o encosto inclinado em 40°.	Sim.
Wishart <i>et al.</i> [7] - 2010	250 imagens por paciente.	100 pacientes - privado.	Sentinel BreastScan.	Sim.
Luna e Delgado [60] - 2010	3 imagens: 1 frontal, 2 laterais.	Informação desconhecida.	Não.	Informação desconhecida.
Antonini <i>et al.</i> [69] - 2011	5 imagens: 1 frontal, 4 laterais.	3 pacientes - privado.	Não.	Sim.
Kontos <i>et al.</i> [63] - 2011	3 imagens: 1 frontal, 2 laterais.	63 pacientes - privado.	Cadeira giratória sem encosto.	Sim.
Bezerra <i>et al.</i> [66] - 2013	8 imagens: 4 frontais, 4 laterais.	220 pacientes - público, através de requerimento.	Não.	Sim.
Silva <i>et al.</i> [67] e Silva [70] - 2013	24 imagens: 1 frontal estática, 19 frontais sequenciais dinâmicas, 4 laterais estáticas.	174 pacientes - público, através de requerimento.	Cadeira com suporte para os braços na altura dos ombros.	Sim.

Tabela 3: Resumo dos principais trabalhos sobre o armazenamento de termogramas

O armazenamento dos termogramas em um banco de dados público é muito importante para o desenvolvimento de pesquisas na área. Quando estas imagens encontram-se acessíveis, de forma organizada, facilitam o acesso de especialistas e pesquisadores

possibilitando o avanço de técnicas de auxílio ao diagnóstico e o aperfeiçoamento de métodos já existentes. Entretanto, em contraste com a maioria dos outros tipos de imagens médicas utilizadas na mastologia (mamografia, ultrassom, ressonância magnética), ainda não se encontra um banco de dados expressivo de imagens termográficas da mama como por exemplo o *Digital Database for Screening Mammography* (DDSM) ou o *Mammographic Image Analysis Society* (MIAS). Os termogramas geralmente são armazenados em bancos privados, disponíveis apenas para acesso interno. Os dois maiores bancos de imagens termográficas de mama, públicos, que se tem conhecimento são da UFPE e da UFF, disponíveis respectivamente em [71] e [72]. Ambos foram utilizados nesta tese.

3.2 SEGMENTAÇÃO DA REGIÃO DE INTERESSE

A segmentação da região de interesse (comumente chamada de ROI - *Region of Interest*) consiste em separar a parte principal do restante da imagem. A segmentação da ROI em termogramas inclui separar toda a região da mama e os gânglios linfáticos de regiões próximas, como as axilas. A segmentação das mamas em imagens termográficas é desafiadora devido a não existência de um padrão no formato de mama entre diversas pacientes, além do baixo contraste natural das imagens termográficas. Estas dificuldades fazem com que muitos autores prefiram técnicas manuais ou semiautomáticas para a segmentação, assim encontram-se poucas técnicas completamente automáticas na literatura.

A Tabela 4 relaciona os principais trabalhos relacionados à segmentação de imagens termográficas da mama encontrados na literatura. Na coluna “*Grau de automação*” tem-se uma escala de 0 a 2, sendo 0 - completamente manual, 1 - semiautomática e 2 - completamente automática. Na coluna “*Região segmentada*” tem-se M para apenas a região das mamas e ML para região das mamas e gânglios linfáticos.

Autores - Ano	Grau de automação	Região segmentada	Validação
Lipari e Head [73] - 1997	1	ML	Informação desconhecida.
Herry e Frize [74] - 2002	0	M	Informação desconhecida.
Zhou <i>et al.</i> [75] - 2004	2	M	Resultados comparados com um método simples de detecção de bordas.
Scales <i>et al.</i> [76] - 2004	1	M	Duas abordagens para a segmentação são apresentadas e comparadas entre si.

Ng e Kee [57] - 2008	0	M	Informação desconhecida.
Qi <i>et al.</i> [77], [78]- 2000 e 2008	2	M	Informação desconhecida.
Jin-Yu <i>et al.</i> [79]- 2009	2	M	Informação desconhecida.
Schaefer <i>et al.</i> [80] - 2009	0	M	Médicos especialistas fizeram a segmentação.
Kapoor e Prasad [61] - 2010	2	M	Informação desconhecida.
Motta <i>et al.</i> [81] - 2010	2	M	As imagens foram segmentadas por 5 especialistas e depois comparadas com a metodologia automática desenvolvida.
Zadeh <i>et al.</i> [82] - 2011	2	M	Informação desconhecida.
Marques [83] - 2012	2	ML	As imagens foram segmentadas por 5 especialistas e depois comparadas com a metodologia automática desenvolvida.

Tabela 4: Resumo dos principais trabalhos de segmentação de termogramas.

3.3 AUXÍLIO À DETECÇÃO E DIAGNÓSTICO DE DOENÇAS

A termografia quando comparada com outros tipos de exames por imagens apresenta um custo muito baixo para a realização do exame. Porém, apresenta algumas desvantagens, como (i) a baixa sensibilidade para detectar tumores pequenos e profundos, (ii) a impossibilidade de diferenciar tumores de pontos quentes correspondentes a inflamações locais e (iii) a subjetividade utilizada pelos médicos para interpretar o diagnóstico através delas.

Técnicas de visão computacional e reconhecimento de padrão podem auxiliar os especialistas a superar essas desvantagens. O primeiro passo para utilizar estas técnicas é a escolha das características que serão extraídas dos termogramas, a fim de compor um vetor de características, e como analisar e comparar esses vetores. Outro passo importante em técnicas de reconhecimento de padrões é a classificação e avaliação das medidas extraídas. No caso de imagens termográficas da mama, este passo é responsável por apontar se a paciente apresenta mamas saudáveis ou não. Assim, para realizar essa classificação, é necessário o uso de técnicas de decisão.

A Tabela 5 apresenta uma revisão de trabalhos relacionados com a extração de característica e classificação de imagens termográficas da mama.

Autores - Ano	Características extraídas	Classificação	Comparação
Lipari e Head [73] - 1997	Média, mediana, desvio padrão, valores mínimos e máximos de temperatura para cada mama e para cada quadrante.	Informação desconhecida.	Informação desconhecida.
Kuruganti e Qi [84] - 2002	Média, variância, obliquidade, curtose, o pixel de maior intensidade, entropia e entropia conjunta.	Cálculo de correlação para análise de assimetria.	Informação desconhecida.
Koay <i>et al.</i> [56] - 2004	Média, desvio padrão, mediana, valores mínimo e máximo de temperatura, obliquidade, curtose, entropia, área e índice de calor.	Rede Neural Artificial (RNA) com <i>Back Propagation</i> .	Informação desconhecida.
Ng e Kee [57] - 2008	Média, moda, pontos de temperatura, mediana e dados biológicos do paciente.	RNA e <i>Rule based Neurofuzzy Networks</i> (RNFN).	Análise de um especialista.
Arora <i>et al.</i> [58] - 2008	<i>Sentinel BreastScan</i> : Sistema proprietário com características desconhecidas.	Três análises: exame de rastreamento, exame clínico e RNA.	Biópsia.
Qi <i>et al.</i> [77], [78] - 2008	Histograma, média, variância, obliquidade, curtose, correlação, entropia e entropia conjunta.	<i>K-means</i> e <i>K-nearest-neighbourhood</i>	Informação desconhecida.
Tang <i>et al.</i> [68] - 2008	Amplitude térmica local através de análise morfológica e processamento de sinais.	Informação desconhecida.	Informação desconhecida.
Wiecek <i>et al.</i> [35] - 1999	Média, desvio padrão, variância, obliquidade, curtose, energia, correlação, diferença de variância e entropia.	RNA.	Informação desconhecida.
Schaefer <i>et al.</i> [80], [85] - 2008, 2009	Características de estatística básica, momentos, histograma, matriz de concorrência cruzada, informação mútua e análise de Fourier.	Sistema de classificação baseado em Fuzzy.	Comparação com resultados de acurácia de outros tipos de exames como a mamografia e a RMI.
Da Silveira Filho <i>et al.</i> [86] - 2009	Medidas de Lacunaridade.	Gráfico com os dados obtidos para cada mama.	Diagnóstico por biópsia.
Serrano <i>et al.</i> [87] - 2010	Medidas de Lacunaridade e coeficiente de Hurst.	Gráfico com os dados obtidos para cada mama.	Diagnóstico por biópsia.
Kapoor e Prasad [61] - 2010	Variação de temperatura, obliquidade e curtose.	Clusterização.	Informação desconhecida.
Wishart <i>et al.</i> [7] -	Informação desconhecida.	Quatro análises: exame de	Biópsia e resultados do <i>Sentinel</i>

2010		rastreamento, exame clínico e duas configurações de RNA.	<i>BreastScan</i> .
Nurhayati <i>et al.</i> [88], [89] - 2010, 2011	Média, variância, obliquidade, curtose e entropia.	Gráfico com as características escolhidas pela análise do PCA.	Informação desconhecida.
Zadeh <i>et al.</i> [82] - 2011	Média, variância, obliquidade e curtose.	<i>K-means</i> e <i>C-means</i> .	Informação desconhecida.
Acharya <i>et al.</i> [64] - 2012	Homogeneidade, energia, entropia, 4 primeiros momentos, 2º momento angular, contraste, média, ênfase em primitivas longas e curtas, percentagem de primitiva, uniformidade do nível de cinza e não uniformidade do nível de cinza.	SVM.	Informação desconhecida.

Tabela 5: Resumo dos principais trabalhos de análise de características.

CAPÍTULO 4 – METODOLOGIA DESENVOLVIDA

Neste capítulo será descrita a metodologia desenvolvida para a detecção de alterações nas mamas através de imagens termográficas. Serão detalhados os procedimentos de aquisição dos termogramas, as etapas de pré-processamento e segmentação dos mesmos, assim como todas as etapas auxiliares desenvolvidas.

4.1 RESUMO

A metodologia desenvolvida nesta tese têm como objetivo detectar possíveis alterações nas mamas, sem classifica-las como benignas ou malignas, apenas indicando a presença ou ausência. A Figura 15 resume os passos usados para alcançar os objetivos desta tese.

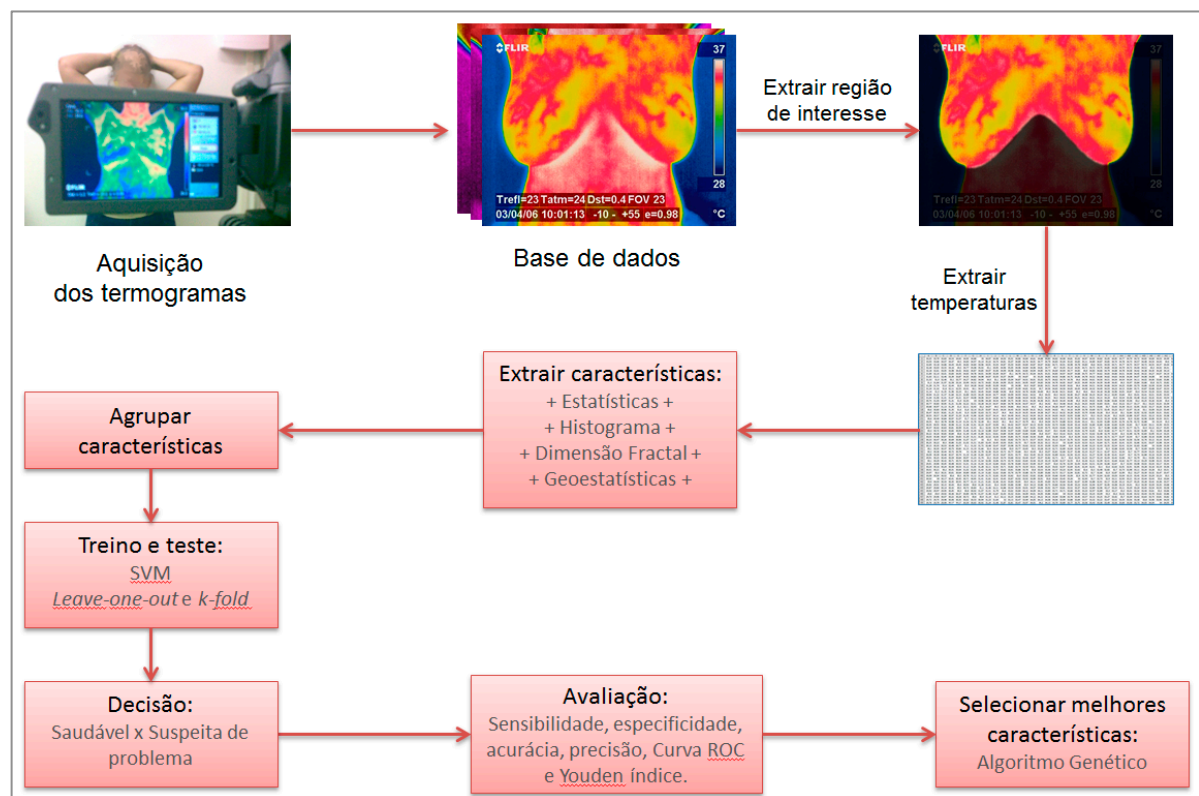


Figura 15: Etapas da metodologia para alcançar o objetivo desta tese.

4.2 AQUISIÇÃO DOS TERMOGRAMAS

Esta tese utiliza os termogramas de mama adquiridos pelo Projeto PROENG PE021-2008 – Processamento e análise de imagens aplicadas à mastologia [90] – em parceria com o

Projeto PROCAD-NF 540/2009 – Processamento e análise de imagens mastológicas [91] – ambos financiados pela Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES).

Atualmente os exames térmicos da mama, referentes aos projetos, são realizados no Hospital das Clínicas (HC) da UFPE no Recife-PE e no Hospital Universitário Antônio Pedro (HUAP) da UFF em Niterói-RJ. As imagens adquiridas no HC seguem o protocolo descrito em Borchardt *et al.* [45] e Bezerra *et al.* [92], os termogramas são capturados baseados em um protocolo estático, onde a paciente aguarda por 10 minutos a aclimação da temperatura do seu corpo antes da aquisição das imagens. Um número mínimo de 8 imagens em diferentes posições são capturadas. A Figura 16 mostra as oito posições mínimas de captura, além destas posições imagens extras podem ser capturadas se necessário. As oito posições são: T1: frontal com as mãos na cintura e os cotovelos afastados do corpo, T2: frontal com as mãos sobre a cabeça, MD: somente a mama direita, ME: somente a mama esquerda, LEMD: lateral externa da mama direita, LEME: lateral externa da mama esquerda, LIMD: lateral interna da mama direita e LIME: lateral interna da mama esquerda. As imagens do HC são capturadas utilizando uma câmera térmica da fabricante FLIR, modelo S45 [93] que apresenta uma sensibilidade de $0,08^{\circ}\text{C}$ e uma faixa de captura padrão de -40°C até 120°C . Os termogramas apresentam resolução de 320×240 pixels, onde cada pixel representa o valor de temperatura de um ponto na cena capturada.

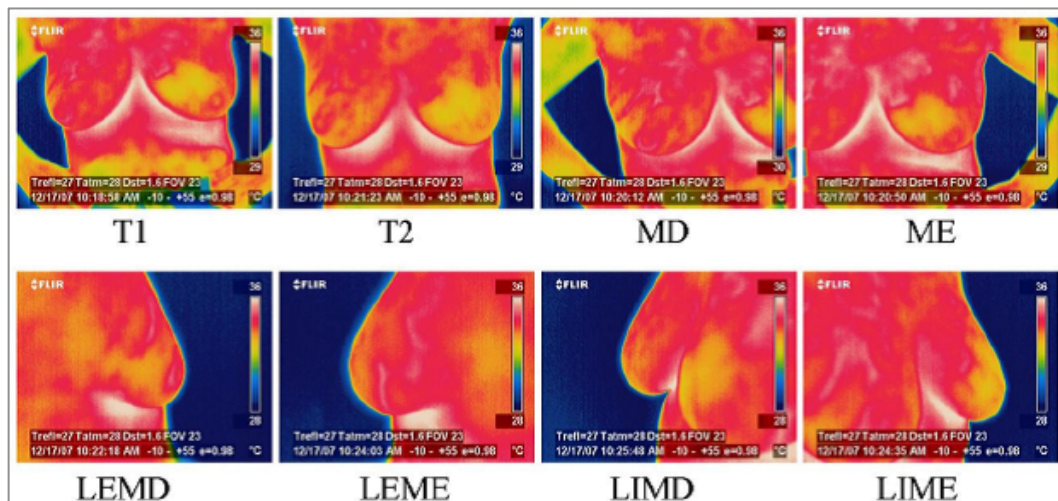


Figura 16: Posições capturadas pelo protocolo de aquisição da UFPE.

Os exames realizados no HUAP seguem o protocolo descrito em Silva [70] e Silva *et al.* [67]. Este protocolo de aquisição foi definido a partir de uma revisão bibliográfica de

diversos tipos descritos na literatura e suas verificações experimentais em pelo menos 20 voluntárias. A forma final considerada levou em conta diversos fatores experimentados e sua eficiência na identificação de achados nas mamas. São obtidas imagens seguindo dois protocolos, um estático e outro dinâmico. No primeiro protocolo, estático, a paciente aguarda por 10 minutos a aclimação da temperatura do seu corpo com os braços apoiados em uma determinada altura que permita que suas axilas fiquem livres. Após este tempo são capturadas 5 imagens em posições distintas. A Figura 17 mostra essas posições, imagens extras podem ser capturadas se necessário. As posições são: TFRON: frontal com as mãos sobre a cabeça; TLL45: lateral esquerda a 45°; TLL90: lateral esquerda a 90°; TLR45: lateral direita a 45°; TLR90: lateral direita a 90°. Após a captura da série de imagens seguindo o protocolo estático, a paciente é submetida a um processo de resfriamento (aplicado por um ventilador direcionado às mamas). Quando a região central do peito da paciente é resfriada a uma temperatura inferior a 30,5°C o ventilador é desligado e imagens são capturadas na posição TFRON durante 5 minutos com intervalos de 15 segundos entre imagens subsequentes para observar como ocorre o reaquecimento da região das mamas. Essa temperatura média é calculada automaticamente pela câmera através de uma função do modelo usado que verifica, em tempo real, a média de temperatura de uma região predefinida pelo operador. Esta região corresponde à área limitada pelos quatro cantos que definem o retângulo centralizado na imagem TFRON mostrada na Figura 17. O valor de 30,5°C e o tamanho adequado da região de verificação foram definidos por experimentos com 10 voluntárias, de modo que todos os termogramas iniciassem a partir de uma mesma temperatura.

Os termogramas no HUAP são capturados por uma câmera térmica da FLIR modelo SC620 [94], que possui uma sensibilidade menor que 0,04°C e faixa de captura padrão de -40°C até 500°C. Os termogramas apresentam dimensão de 640 x 480 pixels. Nestas capturas sempre se considera centralizar as pacientes na cena capturada de modo a ter-se o máximo de informação útil capturada, mantendo a paciente em 2 ou 3 distâncias fixas da câmera, e não uma única distância entre a câmera e a paciente. Uma noção do tamanho real das imagens capturadas e das suas distâncias à câmera pode ser conseguida através do uso da informação do tamanho do quadrado de material isolante que é fixado a paciente no seu eixo médio, pouco abaixo da região da prega inframamária. Este quadrado fica bastante visível na posição TFRON e tem tamanho: 2,5cm x 2,5cm.

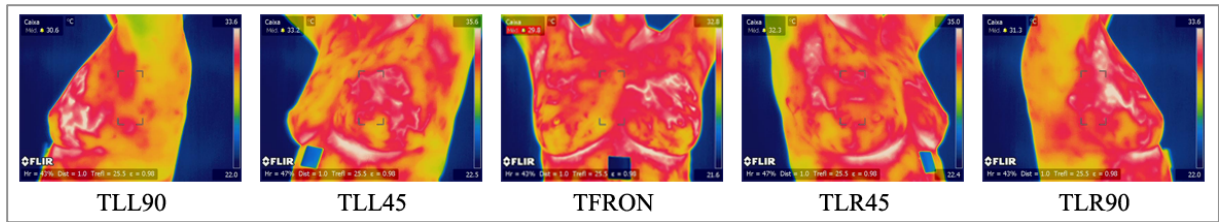


Figura 17: Posições capturadas pelo protocolo de aquisição da UFF.

4.3 AMOSTRA UTILIZADA

A amostra de termogramas da mama utilizadas nesta tese é composta por 69 imagens adquiridas pela UFPE e pela UFF. As imagens provenientes da UFPE foram adquiridas entre 2005 e 2010 no Hospital das Clínicas, com diagnósticos comprovados através de biópsia e laudos apurados por especialistas a partir de mamografias. As imagens da UFF foram adquiridas entre o final de 2012 até a data de fechamento deste texto em abril de 2013 e a classificação de cada imagem foi atribuída baseada no laudo fornecido por especialistas após análise de mamografias, ultrassonografias, exames histológicos ou imuno-histoquímicos.

Por motivos de padronização, as imagens utilizadas nesta tese seguiram o protocolo estático de ambas as fontes, foram utilizados os termogramas da UFPE na posição T2 e os termogramas da UFF na posição TFRON, devido a similaridade entre elas e por existirem metodologias de segmentação automática para estas posições, desenvolvidas por alunos do grupo de pesquisa do projeto.

Do banco de dados da UFPE foram utilizadas imagens de 51 pacientes, 14 saudáveis e 37 com alguma alteração na mama. Das 37 pacientes com alterações da UFPE, 11 apresentavam alterações em ambas as mamas. Entre as alterações nas mamas haviam: 1 linfoma, 1 lipoma, 1 abscesso, 2 esteatonecrose, 8 cistos, 9 fibroadenomas, 10 carcinomas e 19 tumores não classificados. Do banco de dados da UFF foram utilizadas 18 pacientes, 5 saudáveis e 13 apresentando alguma alteração em pelo menos uma das mamas. As alterações nas pacientes da UFF foram todas classificadas como BI-RADS 2 (achados benignos característicos) e as pacientes saudáveis como BI-RADS 1 (mamas simétricas, sem massas, distorções de arquitetura ou calcificações suspeitas). Ambas as classificações BI-RADS (1 e 2) são consideradas mamografias negativas, sem risco de câncer [9]. Devido a diferença na classificação entre as imagens da UFPE e da UFF, optou-se por utilizar as amostras separadamente, explorando principalmente a amostra da UFPE pelo número de casos e pela diversidade de alterações.

A Tabela 6 resume a amostra utilizada. A lista de todas as imagens utilizadas nesta tese está disponível no Anexo A.

	Saudáveis	Patológicas	Total
UFPE	14	37	51
UFF	5	13	18
Total	19	50	69

Tabela 6: Amostra de termogramas utilizados.

4.4 PRÉ-PROCESSAMENTO

Os termogramas adquiridos pelas câmeras térmicas são transferidos para os bancos de dados em seu formato proprietário padrão da fabricante da câmera, denominado JPEG radiométrico. Neste arquivo encontram-se todas as informações de temperatura de cada um dos pontos da cena, além de metadados como a data e horário da captura, a umidade e a temperatura ambiente, a emissividade do corpo, entre outras informações. Também se encontra neste arquivo uma imagem JPEG para visualização, onde a maioria destas informações, acompanhadas de uma legenda de cores/temperaturas e a logomarca da fabricante estão rasterizadas no arquivo, como pode ser observado na Figura 16 em imagens da UFPE e na Figura 17 nas imagens da UFF.

Estas informações presentes na imagem JPEG de visualização atrapalham o processamento e análise dos termogramas por métodos computacionais, pois são informações que não pertencem à ROI e podem levar métodos automáticos de segmentação e extração de características a erros grotescos caso sejam consideradas partes da paciente. Assim, inicialmente é preciso remover todas estas informações da imagem e obter a matriz de temperaturas antes de processá-la.

A solução para obter um arquivo limpo para visualização e a matriz de temperaturas para o processamento foi facilitada pela aquisição, através do Projeto PROCAD-NF (em parceria com a UFMA), do *Software Development Kit* (SDK) da Flir [40]. Este SDK consiste em um componente *ActiveX* que permite o acesso a todas as informações contidas no arquivo JPEG radiométrico, além do controle total das funcionalidades das câmeras Flir, através de códigos escritos em linguagens de programação como *Visual Basic*, *C++* e *C#*.

Com a utilização do SDK foi possível em uma etapa de pré-processamento gerar uma estrutura para armazenar a matriz de temperaturas do arquivo JPEG radiométrico, e a partir

desta matriz foi possível gerar uma imagem do exame completamente limpa, apenas com o corpo da paciente.

4.5 CONVERSÃO DA MATRIZ DE TEMPERATURA PARA IMAGEM

Todas as características foram extraídas das matrizes de temperaturas, porém para se ter uma ideia visual das temperaturas foram geradas imagens em tons de cinza com base na matriz de temperatura. Estas imagens são geradas utilizando a transformação apresentada na Equação (9). Esta equação foi adaptada da conversão de escalas termométricas [95].

$$p_{ij} = 255 \cdot \frac{t_{ij} - t_{min}}{t_{max} - t_{min}} \quad (9)$$

onde p_{ij} e t_{ij} representam respectivamente os pixels da imagem e os valores de temperatura da matriz, t_{max} é a temperatura máxima presente na matriz, este valor em 100% dos casos está contido na área que compreende o corpo da paciente, e t_{min} é uma temperatura limiar calculada automaticamente para eliminar possíveis ruídos do fundo da imagem (região não pertencente ao corpo da paciente) e melhorar o contraste das imagens.

O cálculo de t_{min} é baseado na característica bimodal geralmente apresentada pelo histograma dos termogramas, como pode ser observado na Figura 18, onde os histogramas de uma amostra de 140 termogramas da mama foram analisados. Os valores pertencentes ao entorno da primeira moda (esquerda) representam temperaturas encontradas no fundo do ambiente onde foram realizados os exames, as temperaturas presentes na região da segunda moda (direita) pertencem ao corpo da paciente. Assim, t_{min} é o valor de temperatura com menor quantidade acumulada entre as duas modas do histograma. Desta forma as temperaturas do fundo não são representadas na imagem, com exceção de algum ruído de alta temperatura que possa estar presente no fundo.

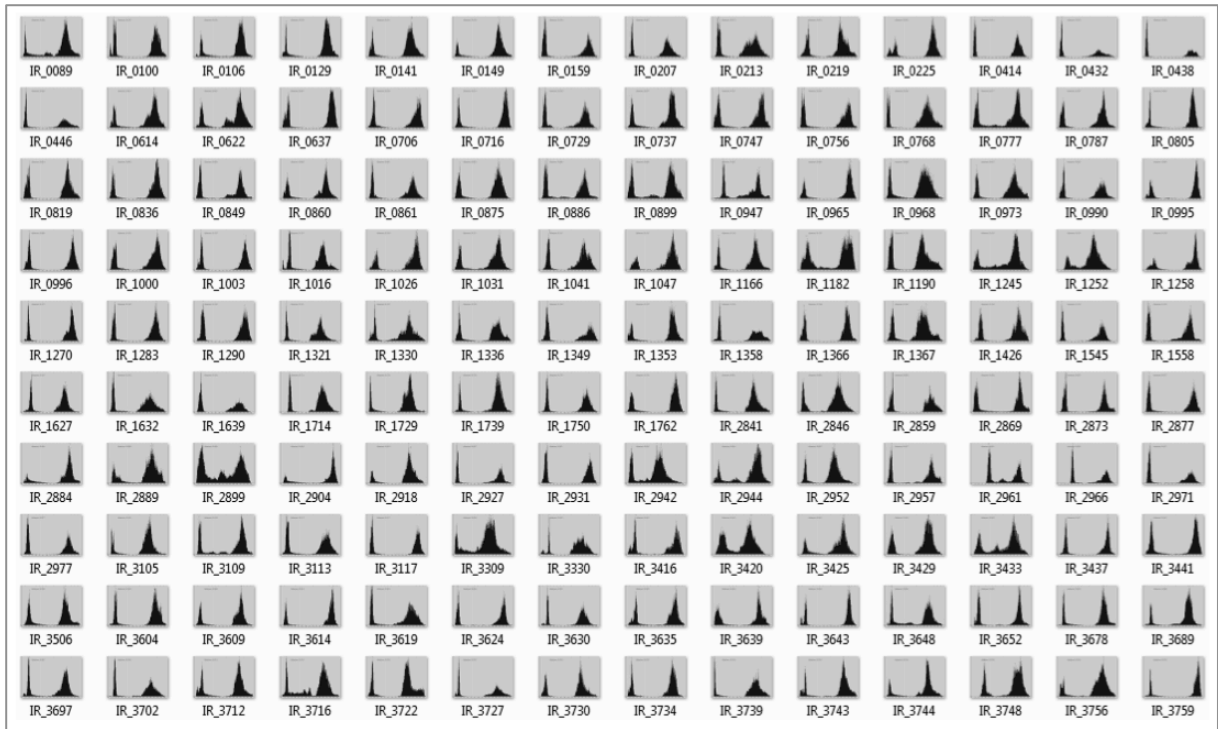


Figura 18: Análise do histograma de 140 termogramas (Fonte: Marques [83]).

A Figura 19 mostra uma comparação entre um termograma original adquirido pela câmera térmica (esquerda) e um termograma após esta conversão com ajuste de contraste. É possível observar que na imagem original tem-se a exibição de valores de temperatura entre 26°C e 35°C. Na imagem pré-processada os valores exibidos estão entre 29,7°C e 35,2°C, o que elimina os pontos de ruído do fundo da sala e melhora o contraste do corpo da paciente na imagem.

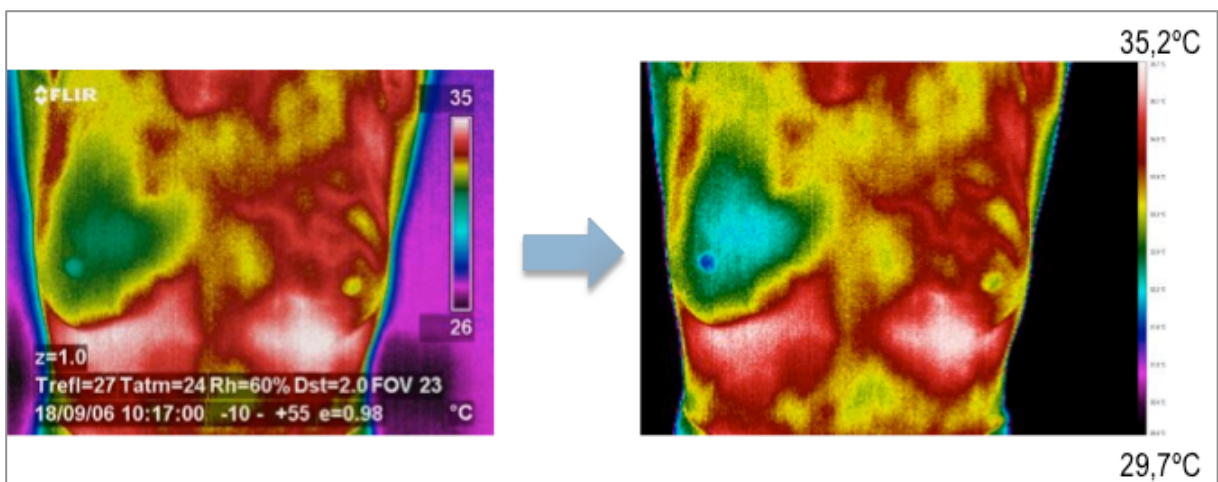


Figura 19: Termograma original (esq.) e termograma após pré-processamento (dir.).

4.6 CORREÇÃO DA POSIÇÃO DA PACIENTE NA IMAGEM

Na termografia, em muitos casos, a aquisição das imagens é prejudicada pelo mau posicionamento das pacientes ou pela dificuldade delas manterem a mesma posição durante o exame. Neste caso, o mau posicionamento da paciente na imagem pode prejudicar no diagnóstico, devido a uma segmentação errônea, e aumentar o número de falsos-negativos. O objetivo desta etapa é a correção após a aquisição das imagens, aplicando técnicas de registro de imagens e transformações geométricas. Esta correção tende a deixar o termograma da paciente mais simétrico lateral e com um posicionamento vertical mais correto, sem perda ou alteração nos valores de temperatura de cada pixel.

O registro desenvolvido baseia-se em 3 passos:

- (i) Detecção de pontos simétricos: Quatro pontos de restrição são detectados em cada imagem. Os dois pontos mínimos das pregas inframamárias e os dois pontos de maior curvatura nas axilas;
- (ii) Compensação angular: É calculado o ângulo existente entre cada um dos pontos de cada par em relação ao eixo horizontal. Então é calculada a média entre os dois ângulos encontrados. No caso de posicionamento perfeito da paciente esta inclinação é zero, o que significa que o paciente está completamente alinhada com o chão ($\alpha = \text{zero}$). Caso seja diferente de zero, é aplicada uma rotação aplicando a matriz de transformação da Equação (10) no termograma, onde c_x e c_y são os pontos centrais da imagem;

$$R_{\alpha} = \begin{bmatrix} \cos \alpha & -\sin \alpha & c_x(1 - \cos \alpha) + c_y \sin \alpha \\ \sin \alpha & \cos \alpha & c_y(1 - \cos \alpha) + c_x \sin \alpha \\ 0 & 0 & 1 \end{bmatrix} \quad (10)$$

- (iii) Compensação linear: Depois da compensação angular é realizado um deslocamento vertical e horizontal na imagem para melhor centralizá-la. É calculado o ponto médio entre os quatro pontos de restrição e este ponto é transladado até o centro da imagem.

A Figura 20 apresenta os pontos simétricos detectados em dois termogramas distintos e o ângulo existente entre eles.

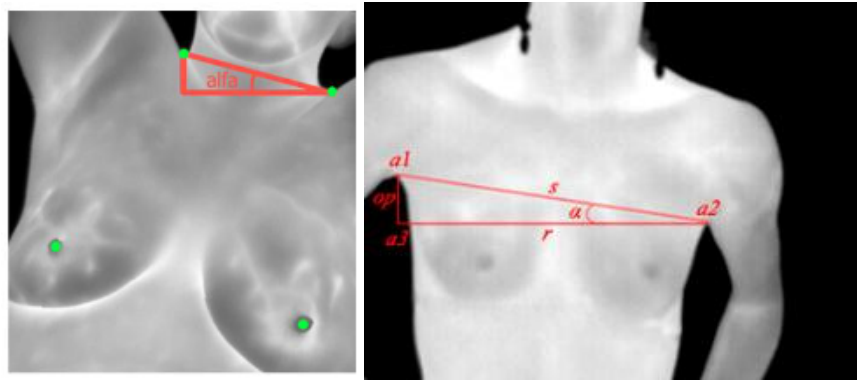


Figura 20: Detecção de pontos simétricos e calculo do ângulo de compensação.

4.7 SEGMENTAÇÃO DA REGIÃO DE INTERESSE

A extração ou segmentação da região de interesse das imagens foi realizada com o auxílio do *software* de segmentação automática desenvolvido por Marques [83], que relata uma acurácia superior a 95% durante os testes realizados na etapa de desenvolvimento. Apesar da alta acurácia, em alguns casos a segmentação automática não foi satisfatória, quando isto ocorreu foram utilizadas segmentações realizadas por especialistas através do aplicativo *Ground Truth Maker*, desenvolvido por Marques [96].

As matrizes de temperaturas obtidas com o auxílio do SDK serviram como entrada para o *software* de segmentação automática, alguns exemplos da segmentação podem ser vistos na Figura 21, onde as linhas delimitam as ROIs.

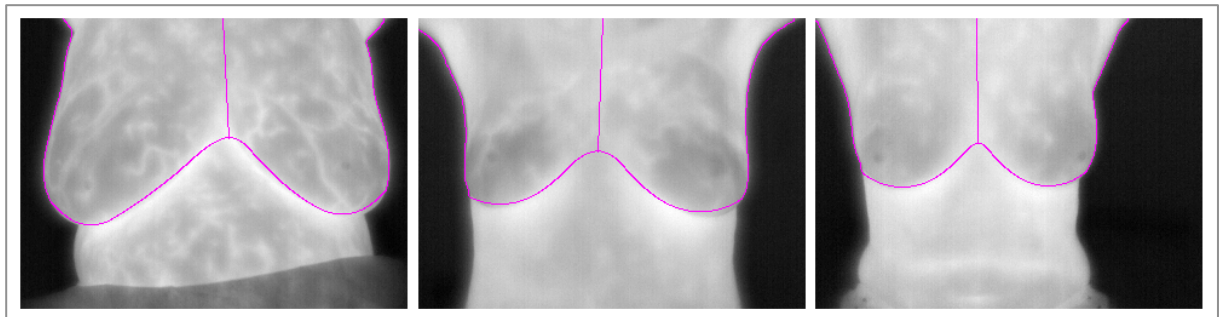


Figura 21: Exemplos de termogramas segmentados pelo software de Marques [83].

4.8 EXTRAÇÃO DE CARACTERÍSTICAS

Nesta seção serão descritos os métodos computacionais utilizados para extrair as características de cada termograma. Os métodos apresentados aqui analisam os termogramas da mama com o objetivo de extrair características baseadas nas informações de temperatura, muitas vezes não observadas pelos especialistas que realizam a análise visual destas imagens.

Desta forma, esses métodos irão fornecer um conteúdo complementar à opinião do especialista.

Os métodos propostos são divididos em grupos conforme a base de seu cálculo. O Grupo I apresenta características baseadas em medidas estatísticas simples, calculadas diretamente sobre os valores das matrizes de temperatura dos termogramas, sendo formado pela média, desvio padrão e amplitude térmica. O Grupo II apresenta características baseadas no histograma do termograma, que foram a quantificação de diferentes níveis em uma posterização da ROI, o temperatura limiar de 1% de área mais quente e a maior área isotérmica de alta temperatura na mama. O Grupo III é baseado na dimensão fractal da região da mama calculada pelo método de Higuchi e pelo cálculo da Sucolaridade. O Grupo IV apresenta características baseadas em geoestatística adaptadas do estudo de análises espaciais de terrenos: coeficiente de Geary, índice de Moran e a função K de Ripley. Além destes quatro grupos básicos de características, são formados grupos complementares para a análise utilizando a combinação destes grupos básicos. A Tabela 7 resume os grupos, os métodos e o número de características por mama utilizadas neste trabalho.

Grupo	Métodos	Nº. de características
I	Média	1
	Desvio Padrão	1
	Amplitude Térmica	1
II	Quantização do nível mais alto	1
	Temperatura limiar de 1% dos pontos mais quentes	1
	Maior área isotérmica	1
III	Dimensão fractal de Higuchi	5
	Sucolaridade	4
IV	Coeficiente de Geary	10
	Índice de Moran	10
	Função K de Ripley	5
V	União das características dos Grupos I e II.	6
VI	União das características dos Grupos I, II e III.	15
VII	Todas as características.	40

Tabela 7: Grupos de medidas analisadas

Nas subseções seguintes serão descritos com detalhes todos os grupos e as características citadas na Tabela 7.

4.8.1 CARACTERÍSTICAS BASEADAS EM ESTATÍSTICA (GRUPO I)

Métodos estatísticos permitem extrair informação de dados, os quais são muitas vezes incompletos, proporcionando informações úteis sobre o problema em estudo. O estudo estatístico permite estimar uma quantidade ou testar uma hipótese acerca de uma grande população, baseando-se em uma pequena amostra.

Sendo assim, o objetivo de utilizar métodos estatísticos neste trabalho é extrair informações das temperaturas das mamas de um grupo de pacientes, e a partir destes dados obter uma melhor compreensão da relação entre os valores de temperatura das mamas e a presença de problemas.

O Grupo I relaciona alguns métodos de extração de características baseadas em estatística simples, frequentemente utilizados na literatura. Em geral, essas características são calculadas sobre os valores dos pixels das imagens, neste trabalho elas foram adaptadas para obter informações sobre a temperatura de cada ponto.

4.8.1.1 Média

A média indica onde mais se concentram os dados de uma distribuição. Nos termogramas a média pode revelar uma temperatura global da região de interesse, sendo um valor que representa de forma simplificada todas as outras temperaturas da região. Esta é a característica mais encontrada em trabalhos com imagens termográficas, como pode ser observado na revisão bibliográfica realizada nesta tese.

A média para um total de n valores de temperatura, com a temperatura de cada elemento denotado por t_i , com $i = 1, \dots, n$, é definida na Equação (11):

$$\bar{t} = \frac{1}{n} \sum_{i=1}^n t_i \quad (11)$$

4.8.1.2 Desvio padrão

O desvio padrão mostra o quanto os dados de uma distribuição estão dispersos em torno da média. Um baixo desvio padrão indica que os valores tendem a estar bem próximos da média, enquanto um alto desvio padrão indica a presença de uma grande dispersão nos valores em relação à média. O desvio padrão é definido na Equação (12).

$$\sigma = \sqrt{\frac{1}{n} \sum_{i=1}^n (t_i - \bar{t})^2} \quad (12)$$

4.8.1.3 Amplitude térmica

A amplitude térmica indica a variação máxima de temperatura na região de interesse. Este valor é calculado a partir da diferença entre a maior e a menor temperatura da região, sendo definido na Equação (13).

$$\Delta t = \max_{1 \leq i \leq n} t_i - \min_{1 \leq i \leq n} t_i \quad (13)$$

Como o cálculo do gradiente baseado em um único pixel pode ser influenciado pelo aparecimento de ruídos no termograma ou por algum ponto do fundo que após a segmentação permaneceu na ROI, o gradiente utilizado foi calculado levando em consideração a média de temperatura de um *kernel* K de área k^2 , onde k^2 é igual a 5% da área total em pixels da região de interesse. Esse valor de k foi definido através de testes e mostrou-se satisfatório para eliminar a influência de ruídos e pontos do fundo. Nos termogramas utilizados com resolução de 320x240 (76800 valores de temperatura), cada mama ocupa em média 22,5% do total (17300 valores), assim o valor de k ficou em média 29, variando entre 27 e 33. Na Equação (14) pode-se observar a matriz que representa o *kernel*, onde os valores de $a_{11}, a_{12}, \dots, a_{kk} = 1$. O *kernel* era movido elemento a elemento pela matriz, e os valores não pertencentes a ROI eram desprezados.

$$K = \frac{1}{k^2} \begin{bmatrix} a_{11} & a_{12} & \dots & a_{1k} \\ a_{21} & a_{22} & \dots & a_{2k} \\ \vdots & \vdots & \ddots & \vdots \\ a_{k1} & a_{k2} & \dots & a_{kk} \end{bmatrix} \quad (14)$$

4.8.2 CARACTERÍSTICAS BASEADAS EM HISTOGRAMA (GRUPO II)

O histograma é uma representação gráfica da distribuição de frequência dos tons em uma imagem. O histograma é muito utilizado na análise e processamento de imagens, auxiliando no ajuste de brilho e contraste, na segmentação de regiões de interesse, na identificação de objetos, na extração de características, entre outras aplicações.

O objetivo da utilização de histogramas neste trabalho é auxiliar a delimitação das regiões mais aquecidas nas mamas, que podem ser indícios de regiões suspeitas de alterações. A partir dos valores de temperatura do termograma é gerado um histograma, que relaciona a frequência em que cada valor de temperatura aparece no termograma. O histograma é calculado sobre as temperaturas de cada mama, e indica a frequência $f(t)$ de cada temperatura. Como as temperaturas são valores reais com duas casas de precisão, para criação dos histogramas os valores foram discretizados a cada 0,1°C. Considerando N o número total de elementos que representam a mama, tem-se que:

$$\sum_{t=t_{min}}^{t_{max}} f(t) = N \quad (15)$$

4.8.2.1 Quantização do nível mais alto

A quantização em múltiplos níveis do histograma permite uma visualização das temperaturas em faixas distintas, onde as frequências das temperaturas pertencentes a cada faixa são acumuladas em um único valor.

Para cada termograma o número de níveis utilizados na quantização é calculado de pela Equação (16), onde o teto da temperatura máxima é subtraído do piso da temperatura mínima, obtendo-se dois limites inteiros. A partir destes limites cada faixa tem comprimento de 1°C.

$$n = \lceil t_{max} \rceil - \lfloor t_{min} \rfloor \quad (16)$$

A Figura 22 mostra como exemplo o histograma calculado e quantizado para o termograma IR_3996 (HC) (imagem no Anexo C), que apresenta a mama direita saudável (linha vermelha tracejada) e um carcinoma na mama esquerda (linha azul contínua). Para este caso foram estabelecidos 7 níveis ($n = 7$), pode-se observar na figura que no último nível a frequência acumulada pela mama esquerda supera a frequência da mama direita.

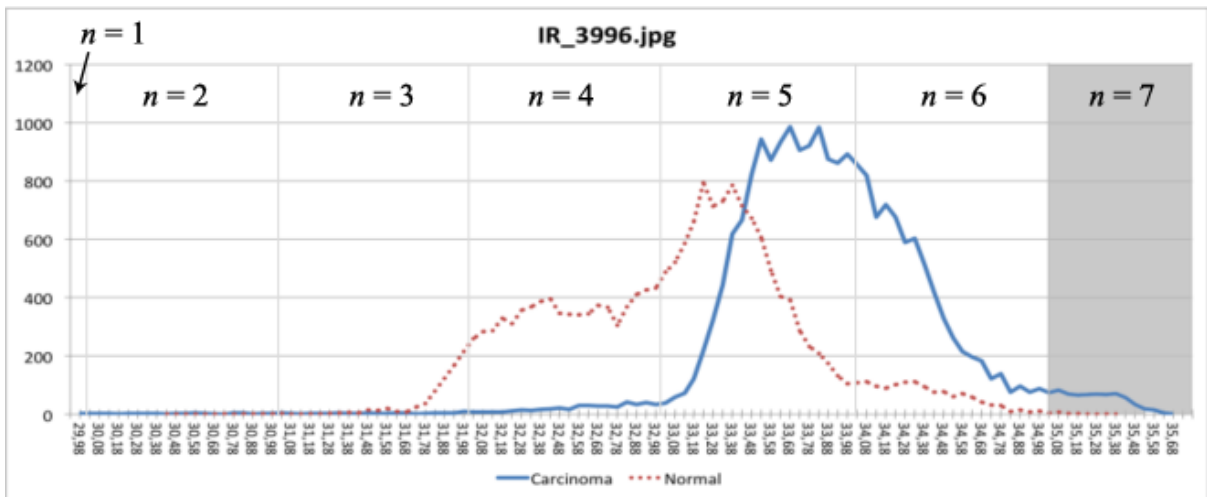


Figura 22: Histograma da frequência de temperatura do termograma IR_3996.

A característica extraída é computada através da Equação (17), onde a frequência acumulada relativa de uma faixa de temperatura ($q(i)$) é resultado do somatório das frequências das temperaturas localizadas entre as temperaturas limites da faixa, dividida pelo número total de pixels da ROI. Nesta tese foram utilizadas apenas a frequência acumulada da

faixa de temperaturas mais elevadas, ou seja, $i = n$, pois são as altas temperaturas que estão relacionadas as doenças da mama.

$$q(i) = \sum_{t=t_{ini}}^{t_{fin}} f(t) \quad (17)$$

4.8.2.2 Limiar de temperaturas elevadas

Outra característica baseada no histograma utilizada foi o limiar que separa 1% das temperaturas mais elevadas, chamado L1. Sendo N o total de pontos da ROI e t_{max} o valor mais alto de temperatura na mama, entre as temperaturas L1 e t_{max} estão 1% dos elementos da matriz. Foram testados diferentes percentuais (1%, 5% e 10%), porém o valor de 1% obteve melhores resultados para diferenciar casos onde as pacientes apresentavam uma mama saudável e outra com alguma alteração. De 21 casos analisados nesta circunstância, utilizando 1% obteve-se em 18 casos um limiar consideravelmente maior para a mama doente (diferença acima de 0,3°C) que para a mama saudável, enquanto que utilizando 5% foram 16 casos e com 10% foram 15 casos.

A Tabela 8 exemplifica como é obtido os valores de limiar utilizados nesta característica. Os valores utilizados na tabela são referentes ao termograma IR_3996. A mama esquerda possui um total de 21470 elementos na matriz de temperatura, enquanto que a mama direita possui 18814.

Temperatura (°C)	Mama esquerda (N = 21470)		Mama direita (N = 18814)	
	Elementos (Acumulado)	% área	Elementos (Acumulado)	% área
35,68	1 (1)	0	0	0
35,63	5 (6)	0	0	0
35,58	15 (21)	0	0	0
35,53	18 (39)	0	0	0
35,48	34 (73)	0	0	0
35,43	57 (130)	0	0	0
35,38	71 (201)	0	1 (1)	0
35,33	68 (269)	1%	1 (2)	0
...
34,63	196 (1668)	7%	31 (222)	1%

Tabela 8: Distribuição das temperaturas mais elevadas do termograma IR_3996.

Na tabela, para cada valor de temperatura do histograma da Figura 22 são apresentados o número de elementos que estão nesta faixa de temperatura e entre parênteses o

valor acumulado desde a temperatura mais alta. Neste exemplo, na temperatura de 35,33°C a quantidade de elementos acumulados para a mama esquerda corresponde a 1% do total de elementos da mama esquerda. Para a mama direita este percentual só é atingido em 34,63°C. Assim, o limiar para a mama esquerda é 35,33°C e para a mama direita é 34,63°C. Neste termograma, assim como em 18 de 21 casos testados, o valor do limiar é maior para a mama que apresenta alguma alteração, neste exemplo a mama esquerda apresenta um carcinoma e a mama direita é saudável e a diferença entre os limiares é de 0,7°C.

4.8.2.3 Superfícies isotérmicas

Uma superfície isotérmica recebe este nome por apresentar em toda sua extensão a mesma temperatura. Devido ao corpo humano não apresentar uma superfície com distribuição simetricamente perfeita de temperatura, nesta característica foram consideradas áreas da superfície da mama que apresentam temperaturas muito próximas.

Esta característica é calculada após o cálculo da característica descrita acima. Após a quantização do histograma em n níveis, como mostrado na Figura 22, são mantidas no termograma apenas as regiões que contêm as temperaturas pertencentes ao nível mais alto de temperatura. Dentre estas regiões, o percentual de área que cada uma ocupa em relação a mama é calculado. É utilizada como característica o percentual da maior área isotérmica de alta temperatura presente no termograma na região da mama. É utilizado o valor em percentual para ficar independente do tamanho da mama e, é utilizada a maior área isotérmica pois grandes áreas quentes nos termogramas representam alterações [8].

A Figura 23 mostra visualmente como é encontrada esta região. A imagem à esquerda representa toda a região da mama (19618 elementos), a imagem central representa a quantização em 5 níveis, a imagem à direita apresenta as regiões isoladas da faixa de mais alta temperatura, a região marcada em vermelho representa a região isotérmica com maior área na mama (196 elementos, 1% do total).

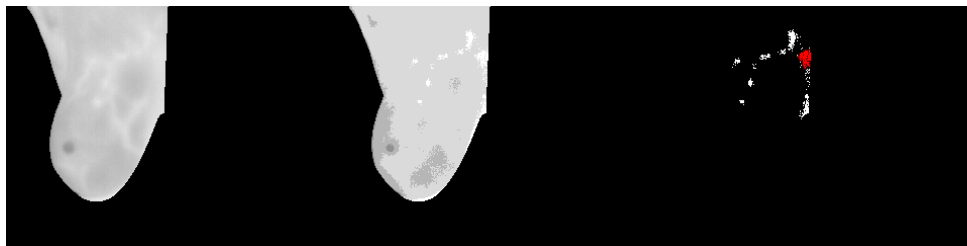


Figura 23: Superfície isotérmica: mama inteira (esq.), em n níveis (centro) e apenas no nível mais elevado (dir.).

4.8.3 CARACTERÍSTICAS BASEADAS EM GEOMETRIA FRACTAL (GRUPO III)

O termo dimensão fractal foi criado por Mandelbrot [97] para estudar fenômenos contínuos em relação ao espaço ou ao tempo que mostram correlação de acordo com uma escala. A dimensão fractal (DF) representa uma medida muito adequada para a análise de geometrias irregulares. A dimensão fractal pode ser avaliada computacionalmente de diversas maneiras. O coeficiente de Hurst, a dimensão *Box counting*, a *differential Box counting* e o coeficiente de Higuchi são algumas destas formas. Nas diversas formas disponíveis, estes avaliadores apresentam valores proporcionais à complexidade e a similaridade própria dos objetos que são avaliados e geralmente são medidas mais adequadas a objetos reais e naturais.

4.8.3.1 Dimensão fractal de Higuchi

A dimensão fractal de Higuchi é utilizada em sinais unidimensionais para medir a sua complexidade segundo o conceito de fractal. Caso o sinal seja fractal, o valor obtido estará entre 1 e 2, quanto mais próximo a 2 mais complexo é o sinal. Este conceito descreve melhor geometrias encontradas na natureza, sendo estas geometrias não triviais de serem representadas sinteticamente [98].

Segundo Gómez *et al.* [99] a dimensão fractal de Higuchi é um método adequado para analisar a DF de sinais biomédicos, pois apresenta algumas vantagens sobre outros métodos, como a não dependência de uma sequência binária de valores necessária em outros métodos [100].

Dada uma série unidimensional de N valores $X = \{x(1), x(2), \dots, x(N)\}$, são geradas k novas séries X_k^m definidas por $X_k^m = \{x(m), x(m+k), x(m+2k), \dots, x(m + \text{int}(\frac{N-m}{k})k)\}$, onde k e m são inteiros, k representa o intervalo discreto entre os pontos e m indica o valor inicial da série analisada sendo $m = 1, 2, \dots, k$. O comprimento de cada nova série gerada é calculada através da Equação (18).

$$L(m, k) = \left\{ \frac{\sum_{i=1}^{\lfloor \frac{N-m}{k} \rfloor} |x(m+ik) - x(m+(i-1)k)| (N-1)}{\lfloor \frac{N-m}{k} \rfloor k} \right\} / k \quad (18)$$

onde $(N-1)/[(N-m)/k]k$ é o fator de normalização da série. O comprimento da série X para o intervalo de observação k é dado pela média das dimensões de todas as k novas séries X_k^m , $L(k) = \{\sum_{m=1}^k L(m, k)\}/k$. $L(k)$ é proporcional a k^{-D} , onde D é a dimensão fractal obtida pelo método de Higuchi [100].

No cálculo da dimensão fractal de Higuchi, os valores de temperatura da mama foram realocadas linha a linha para um vetor. Assim, as temperaturas são apresentadas no formato de um sinal, como pode ser observado na Figura 24 (o termograma IR_0089 pode ser visto no Anexo C).

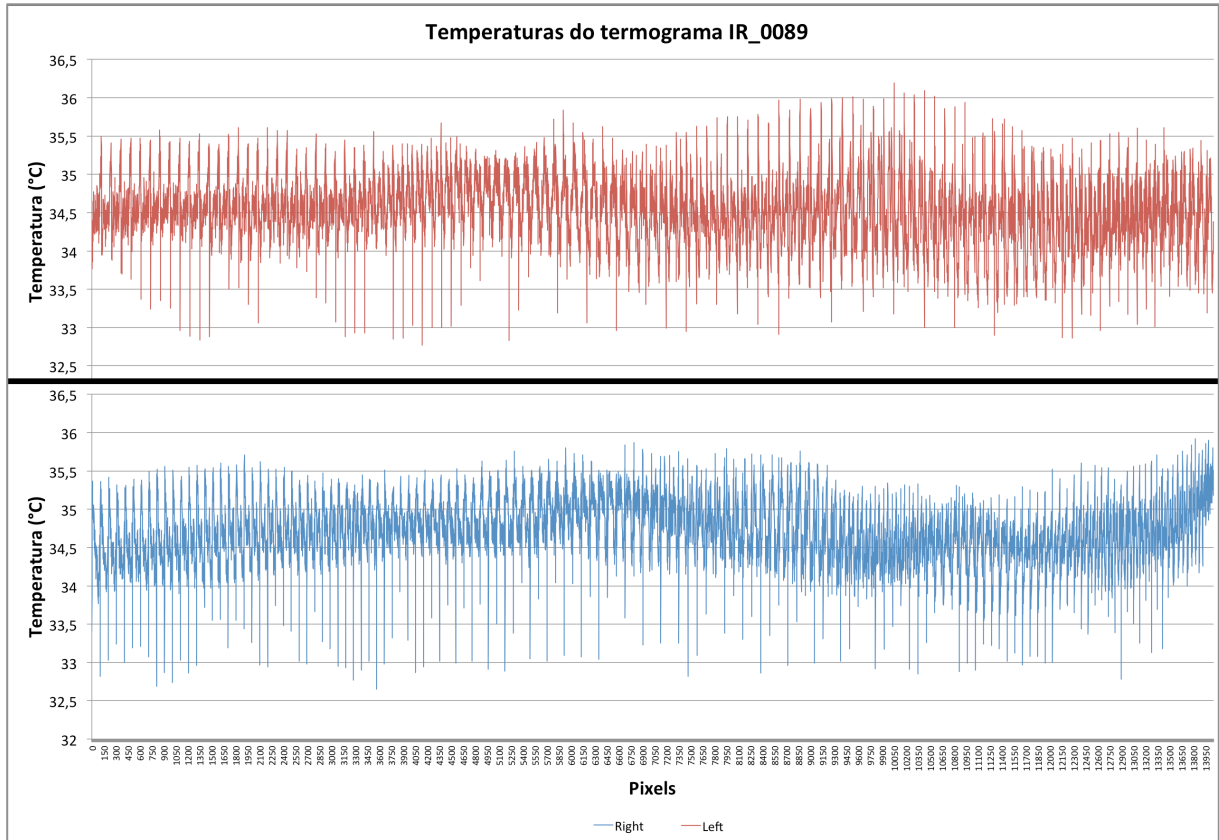


Figura 24: Temperaturas do termograma IR_0089 em formato sequencial.

Para obter o valor da dimensão fractal é necessário analisar o gráfico $\ln(L(k))$ versus $\ln(1/k)$ para $k = 1, 2, 3, \dots, k_{max}$, a angulação da reta que melhor representa a distribuição dos pontos no gráfico, obtida pelo método dos mínimos quadrados, é proporcional a dimensão fractal D . Para avaliar a dimensão fractal de Higuchi pode-se executar o algoritmo para diversos valores de k_{max} , como não foi encontrada nenhuma técnica para selecionar o melhor valor de k_{max} , foram aplicados cinco diferentes valores, $k = 5, 25, 50, 75$ e 100 . Assim, resultaram em 5 características baseadas em Higuchi, denominadas Higuchi(k).

A Figura 25 mostra a distribuição dos pontos no gráfico $\ln(1/k)$ versus $\ln(L(k))$, com $k_{max} = 100$, e a equação da reta obtida por mínimos quadrados para a mama esquerda do termograma IR_0089. A angulação da reta corresponde a dimensão fractal, neste caso 1,724.

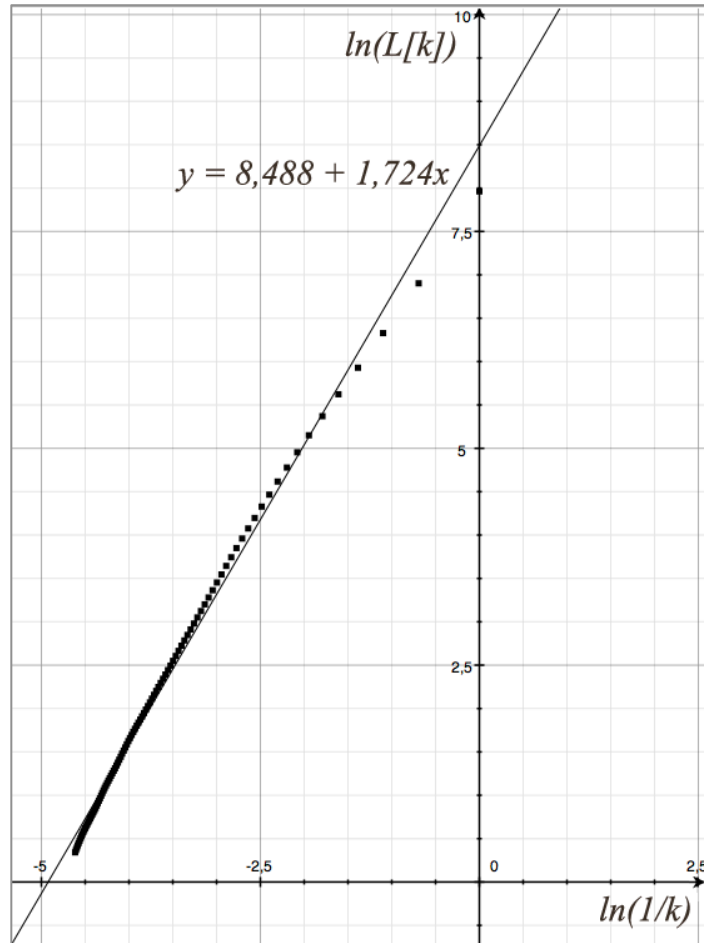


Figura 25: Gráfico $\ln(1/k)$ x $\ln(L[k])$, com $k_{max} = 100$.

4.8.3.2 Sucularidade

A sucularidade mede o grau de percolação de uma distribuição binária (o quanto um fluido pode fluir por essa distribuição) [97]. Para isso, é simulado um fluido percorrendo pela distribuição em diferentes sentidos, onde valores iguais a 0 permitem que o fluido passe e valores iguais a 1 bloqueiam, o fluido se propaga para uma vizinha de 4, recursivamente. Neste trabalho foram utilizados 4 sentidos para a percolação: esquerda para a direita - L2R, direita para esquerda - R2L, de cima para baixo - T2B e de baixo para cima - B2T.

A Figura 26 exemplifica como é calculada a percolação para diferentes direções de uma mesma distribuição binária. Em destaque estão os pixels ocupados pelo fluido após o processo.

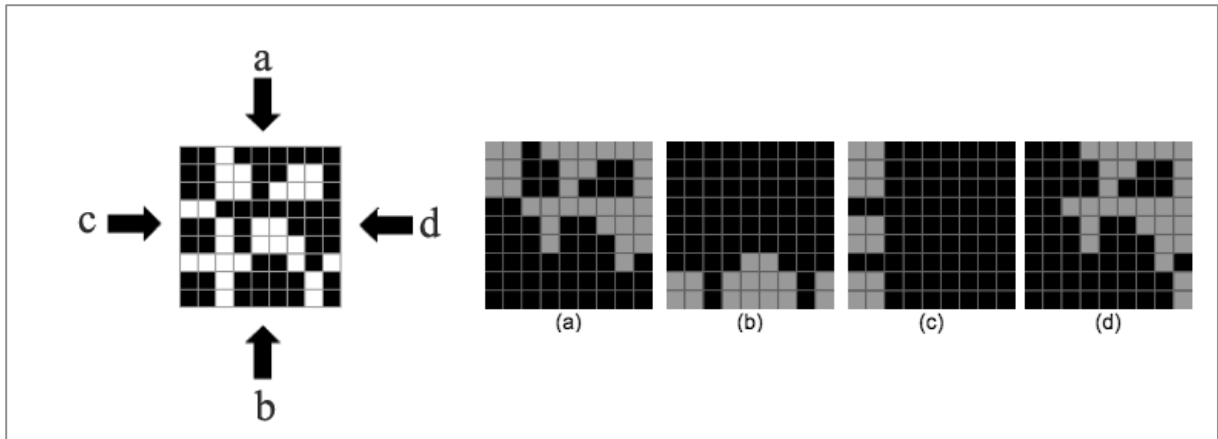


Figura 26: Sucularidade: resultados obtidos pela percolação em 4 direções em uma mesma distribuição, (a) T2B, (b) B2T, (c) L2R e (d) R2L.

O método utilizado para calcular a sucularidade nesta tese foi proposto por de Melo e Conci [101] e pode ser resumido em 4 passos, considerando a distribuição como uma imagem binária:

1. Todos os pixels da imagem são considerados, inclusive as bordas. Se o pixel é 0, significa que o fluido pode passar, então este pixel é marcado. Se o pixel for 1, é considerado um obstáculo. Todos os pixels marcados tem seus 4 vizinhos analisados, recursivamente.
2. A área marcada é dividida em caixas de tamanho igual ($B(n)$, onde n é o número de possíveis divisões da imagem em caixas), como realizado no método de *box-counting*. Após isso, o percentual de ocupação de cada caixa é calculado (OP).
3. Para cada tamanho de caixa, k , é calculada a soma da multiplicação de $OP(B(k))$ pela pressão $PR(B(k), c)$, onde k varia de 1 a n e c é a coordenada x ou y do centroide da caixa. A pressão é acumulada a cada linha ou coluna da imagem, e caso a análise seja no sentido T2B ou B2T o x do centroide é considerado, caso seja R2L ou L2R o y é considerado.
4. Para a sucularidade ser adimensional, como a dimensão fractal, ela é normalizada pelo somatório total de todas as pressões na imagem, como se considerasse toda a imagem como marcada.

A Equação (19) é utilizada para calcular a sucularidade.

$$S(B(k), direction) = \frac{\sum_{k=1}^n OP(B(k)) \cdot PR(B(k), c)}{\sum_{k=1}^n PR(B(k), c)} \quad (19)$$

Foram utilizadas 4 características baseadas em sucularidade para cada mama, uma para cada direção. Essas características foram denominadas Sucularidade(*dir*), onde *dir* são as direções T2B, B2T, L2R e R2L. As temperaturas foram binarizadas utilizando a média de temperatura da mama, isto é, temperaturas acima da média receberam o valor 1 e abaixo da média receberam o valor 0. A média foi escolhida pois valores muito altos de temperatura resultam em baixas percolações para todas as imagens, assim como valores muito baixos resultam em altas percolações para todas as imagens.

4.8.4 CARACTERÍSTICAS BASEADAS EM GEOESTATÍSTICA (GRUPO IV)

Muitos métodos da estatística clássica requerem que as variáveis analisadas apresentem um padrão de distribuição para que a análise realizada apresente bons resultados. Entretanto, em diversas aplicações, as amostras utilizadas não apresentam um padrão de distribuição organizado, como exemplo pode-se citar a termografia, onde em cada imagem a distribuição das temperaturas não segue nenhum padrão de organização.

A geoestatística incorpora não apenas a interpretação do padrão de distribuição como a estatística, mas também a correlação especial dos dados da amostra. Sendo assim, os métodos geoestatísticos são capazes de fornecer ferramentas para entender a aparente aleatoriedade dos dados contidos nas amostras, identificando possíveis estruturas espaciais. A autocorrelação espacial determina que eventos ocorridos (fenômenos) em uma determinada localização tendem a ter comportamento similar em localizações vizinhas. Esta teoria caracteriza o que é chamado de dependência espacial [102].

A seguir serão descritos os métodos de geoestatística utilizados nesta tese: (i) índice de Moran, (ii) coeficiente de Geary e (iii) Função K de Ripley.

4.8.4.1 Índice de Moran

O índice de Moran é a medida geoestatística padrão para quantificar a autocorrelação espacial [102]. A Equação (20) é utilizada para calcular o índice de Moran.

$$I = \frac{N}{\sum_i \sum_j w_{ij}} \frac{\sum_i \sum_j w_{ij} (X_i - \bar{X})(X_j - \bar{X})}{\sum_i (X_i - \bar{X})^2} \quad (20)$$

onde N é o número de ocorrências de um determinado evento; X_i e X_j , são as posições de dois eventos distintos i e j ; w é o peso entre os eventos i e j ; e \bar{X} é a média das posições dos eventos.

O índice de Moran não depende somente do valor de X , mas também da localização espacial dos dados que são quantificados pelo peso w . Este peso oscila assumindo zero ou um.

O peso w pode ser estipulado de diversas formas: pode-se atribuir o valor 1 ao peso w no caso de i ser vizinho de j e 0 no caso contrário; ou pode-se definir w baseado no inverso da distância quadrática $\left(\frac{1}{d^2}\right)_{ij}$ entre a localização de i e j ; além de outras maneiras presentes na literatura [102]. Nesta tese w foi definido como 1 caso o j esteja na direção analisada e dentro da distância máxima d e zero caso contrário.

O valor de I pode ficar entre -1 e 1. Os valores negativos indicam dispersão, -1 indica uma dispersão espacial perfeita. Os valores positivos indicam correlação, 1 indica a ocorrência de uma correlação perfeita. Um valor zero indica um padrão espacial aleatório.

4.8.4.2 Coeficiente de Geary

O coeficiente de Geary é similar ao índice de Moran, porém o índice de Moran é utilizado para calcular a autocorrelação espacial de forma global, enquanto que o coeficiente de Geary é mais adequado para detectar autocorrelações locais [103]. A Equação (21) é utilizada para calcular o coeficiente de Geary.

$$C = \frac{(N - 1) \sum_i \sum_j w_{ij} (X_i - X_j)^2}{2W \sum_i (X_i - \bar{X})^2} \quad (21)$$

onde W é a soma de todos os pesos w_{ij} , as demais variáveis são as mesmas descritas no índice de Moran.

O valor de C situa-se entre 0 e 2, onde 1 significa que não há correlação espacial. Valores inferiores a 1 demonstram o aumento da autocorrelação espacial positiva, enquanto que valores superiores a 1 representam o aumento da autocorrelação espacial negativa (dispersão espacial).

Para o cálculo do índice de Moran e do coeficiente de Geary é necessário estabelecer previamente alguns parâmetros. Como estas medidas levam em consideração a distribuição espacial, é necessário definir até que distância na vizinhança os eventos serão analisados e também em que direções a análise irá proceder. A definição destes parâmetros permite que mesmo utilizando um único método seja possível obter diferentes atributos de acordo com cada configuração empregada. Todos os atributos investigados são complementares na análise.

A Figura 27 demonstra os parâmetros que precisam ser definidos para o cálculo dos métodos de geoestatística, onde X_i é o evento que está servindo como base, d_1, d_2, \dots, d_n , são as distâncias limites para a análise da vizinhança, \vec{v} é a direção analisada e α é a tolerância angular em relação a direção.

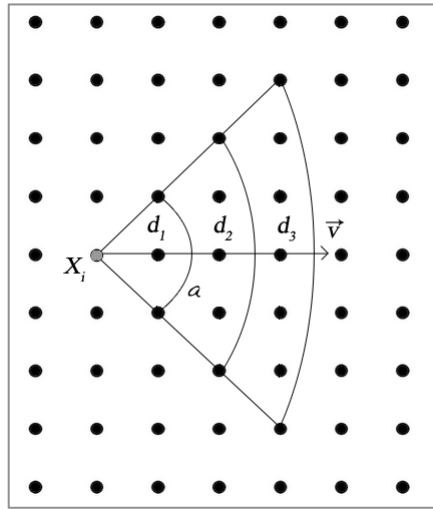


Figura 27: Parâmetros definidos para os métodos de geoestatística.

Nesta tese foram testadas 5 distâncias, definidas como 3%, 6%, 9%, 12% e 15% do total de elementos da matriz pertencentes à ROI, as direções analisadas foram de 0°, 90°, 180° e 270° e a tolerância angular de $\pm 45^\circ$. Entretanto, notou-se que a análise em direções opostas como 0° e 180° ou 90° e 270° retornam os mesmos valores. Assim, foram utilizadas apenas duas direções, 0° e 90°, o que permite uma análise da influência horizontal e vertical, respectivamente, de um evento em relação aos eventos vizinhos. A combinação destes parâmetros gerou 10 características para cada um dos métodos de Moran e Geary (5 distâncias multiplicadas por 2 direções). Estas características foram denominadas $\text{Moran}(d, \alpha)$ e $\text{Geary}(d, \alpha)$, onde d é a distância e α o ângulo que determina a direção.

4.8.4.3 Função K de Ripley

A função K de Ripley [104] é amplamente utilizada na geoestatística, principalmente para análise de imagens de satélite sobre florestas. Ela também é chamada de análise de segunda ordem reduzida, pois se baseia em medidas de segunda ordem. Nesta tese ela foi adaptada para ser utilizada sobre mapas de temperatura ao invés de mapas de vegetação como comumente é utilizada. A principal vantagem do uso da função K de Ripley é a sua capacidade de detectar padrões espaciais em diferentes escalas em uma mesma imagem. A Equação (22) é utilizada para calcular a função K de Ripley para uma determinada distância D .

$$K(D) = \lambda^{-1} \sum_{i=1}^N \sum_{j=1, j \neq i}^N \delta_{ij}(d_{ij}) \quad (22)$$

para $i \neq j$, sendo λ igual a N/A , onde A é a área amostrada, N o total de ocorrências de um determinado evento e $\delta_{ij}(d_{ij})$ uma função que retorna 1 se a distância euclidiana d_{ij} entre os eventos i e j é menor ou igual que D e zero se for maior. Em suma, a função K de Ripley conta o número de ocorrências de um certo evento em um círculo de raio D com centro em i .

Nesta tese foi desenvolvida e utilizada uma função alternativa a função K de Ripley, com o intuito de contar o quão agrupados ocorrem determinados eventos.

Como nos termogramas o evento de interesse são os pontos onde têm-se altas temperaturas, foi investigado o quanto agrupados estão os pontos mais quentes do termograma, pois grandes agrupamentos de pontos quentes podem representar alguma doença. Para estabelecer quais temperaturas representam eventos de interesse foi utilizado o limiar apresentado na Seção 4.8.2.2.

A equação utilizada para calcular esta característica é a mesma apresentada na Equação (22). Porém, o valor de λ foi modificado para $N(N-1)$, assim, caso todos os eventos de interesse estejam a uma distância menor ou igual a D entre eles, o valor de $K(D)$ é igual a 1, caso todos os eventos estejam a uma distância maior que D o valor resultante é zero. O valor $K(D)$ varia entre zero e um de forma inversa à dispersão dos pontos de alta temperatura e direta de acordo com a proximidade das mesmas.

Os valores de D utilizados foram de 3%, 6%, 9%, 12% e 15% da ROI, com o intuito de analisar o agrupamento de regiões quentes com diferentes diâmetros, abaixo de 3% os valores de $K(D)$ ficaram muito baixos para todos os casos e acima de 15% os valores ficaram em 1 ou muito próximos de 1 para a maioria dos casos. Também foram testados valores fixos, como 5, 10, 15, 20 e 25 elementos da matriz de temperatura, porém esses valores não mostraram-se satisfatórios devido a diferença de tamanho entre as mamas de diferentes pacientes e em alguns casos esta diferença pode ocorrer entre as mamas de uma mesma paciente. As características obtidas por este método foram denominadas Ripley(d), onde d é a distância.

4.9 CLASSIFICAÇÃO

O classificador SVM foi utilizado através da biblioteca *LibSVM* [105]. Esta biblioteca permite utilizar três tipos de aprendizado:

- (i) *One-class SVM*: permite a classificação em apenas uma classe;
- (ii) *Support vector classification* (SVC): permite realizar a classificação em duas ou mais classes;

- (iii) *Support vector regression* (SVR): realiza uma regressão ao invés de uma classificação [48].

Nesta tese foi utilizado o aprendizado SVC, para classificar em duas classes. Este aprendizado pode ser configurado como *C-SVC* ou *ν -SVC* (ν ou *nu*). C é um fator de regularização para o espaço de características e ν representa um limite superior sobre a fração de erros de treinamento e um limite inferior da fração de vetores de suporte [105].

Além do tipo de aprendizado, outros parâmetros devem ser configurados de acordo com o *kernel* selecionado para a classificação. O *LibSVM* apresenta quatro *kernels* diferentes, definidos pelas equações:

- (i) linear: $K(x_i, x_j) = x_i^T x_j$
- (ii) poligonal: $K(x_i, x_j) = (\gamma x_i^T x_j + r)^d, \gamma > 0$
- (iii) radial: $K(x_i, x_j) = \exp(-\gamma \|x_i - x_j\|^2), \gamma > 0$
- (iv) sigmoide: $K(x_i, x_j) = \tanh(\gamma x_i^T x_j + r)$

Os parâmetros γ , r e d são configuráveis.

O *kernel* selecionado para uso foi o radial. Esta escolha foi baseada em testes e na sugestão presente no artigo de Hsu *et al.* [106]. O parâmetro a ser configurado neste *kernel* é o valor de γ . Os parâmetros padrões do *LibSVM* são $C = 1$, $\nu = 1$ e $\gamma = 1/n$, onde n é o número de casos na amostra. Estes parâmetros padrões são utilizados caso o usuário não altere-os.

4.10 FERRAMENTAS UTILIZADAS

No processo inicial de carregamento dos termogramas e obtenção das matrizes de temperatura foram utilizadas principalmente funções presentes no SDK [40]. Para a etapa de segmentação dos termogramas foram utilizados o software de segmentação automática de Marques [83] e o *Ground Truth Maker* [96]. O *software* de código aberto *LibSVM* [105] foi utilizado para realizar as classificações das mamas como saudáveis ou patológicas.

Para tornar facilmente utilizável, todas as técnicas e métodos desenvolvidos foram implementados em uma ferramenta desenvolvida em *C#* com o auxílio da biblioteca de código aberto *EmguCV* [107] e o ambiente de desenvolvimento *Visual Studio Professional 2010*.

Os testes foram realizados em um computador com *Windows 7 Professional*, processador *Intel Core i7* (modelo 920 – 2.67GHz), 12GB de memória RAM, placa gráfica *Geforce GTX 260* com 1GB de memória dedicada mais 3GB de memória compartilhada.

CAPÍTULO 5 – TESTES E RESULTADOS

Neste capítulo serão apresentados todos os testes executados durante o desenvolvimento desta tese. Também serão descritas as ferramentas utilizadas. Por fim, serão apresentados os resultados obtidos com a utilização de cada um dos grupos de características propostas e suas combinações, para assim concluir quais métodos detectam melhor as alterações presentes nas mamas.

5.1 TESTES

As características foram extraídas de todos os 69 termogramas da amostra de ambas as mamas de forma independente, isto é, cada mama teve as características extraídas sem levar em consideração as características da outra mama.

Após a extração das características, em alguns testes, foi realizada uma normalização nos valores obtidos, conforme é sugerido por Chang e Lin (2011). Para a normalização todos os valores dos atributos foram escalados entre 0 e 1, através da Equação (23).

$$Y_i = \frac{X_i - X_{min}}{X_{max} - X_{min}} \quad (23)$$

onde Y_i é o valor do atributo X_i escalado entre 0 e 1, X_{min} e X_{max} são os valores mínimos e máximos, respectivamente, obtidos pelo atributo em toda a amostra. Por exemplo, para a média, o valor mínimo e o valor máximo das médias foram utilizados para normalizar a característica média, e assim foi feito para todas as características.

Teste 1:

Inicialmente testou-se a divisão da mama em quadrantes como uma alternativa a extrair as características da mama inteira. Este teste foi chamado de *Teste 1*. Foram extraídas características da mama como um todo e de cada um de quatro quadrantes de área equivalente. Para este teste foram utilizadas as características pertencentes ao Grupo I (estatísticas), Grupo II (histograma) e Grupo III (fractal) e os termogramas adquiridas no HC pelos pesquisadores da UFPE (51 termogramas). Depois de extraídas as medidas de cada mama e de cada quadrante, os valores obtidos na mama esquerda foram subtraídos pelos valores da mama direita e foi utilizado como característica o módulo desta subtração.

O *LibSVM* foi utilizado como classificador, ele foi configurado para o tipo ν -SVC e o *kernel* utilizado foi o radial, com os parâmetros padrões. Foram utilizadas associadas às características dos Grupos I e II (Grupo V) e as características do Grupo III, separadamente.

Apenas estes grupos foram utilizados porquê este teste foi realizado no início do desenvolvimento desta tese, para avaliar a divisão em quadrantes proposta por Resmini [108].

A técnica para validação utilizada foi a *k-fold* com $k = 17$, este valor foi escolhido por ser o maior divisor inteiro do total da amostra, após a realização de todas as interações foram computados os resultados médios de acurácia, sensibilidade, especificidade, índice de Youden e área sob a curva ROC (essas medidas serão utilizadas para todos os testes).

Teste 2:

No segundo teste foram extraídas todas as características de todos os termogramas, separando os valores dos atributos para cada uma das mamas. Com a amostra de 51 termogramas da UFPE, cada mama serviu como um caso para a análise, totalizando um total de 102 casos (54 saudáveis e 48 com alguma alteração).

Os atributos foram utilizados sem nenhuma normalização ou subtração, foram utilizados os valores brutos obtidos pelos métodos desenvolvidos. A tabela apresentada no Anexo 2 mostra os valores mínimos, máximos e a média para cada um dos atributos utilizados.

O *LibSVM* foi utilizado com o tipo *C-SVC* e *kernel* radial com o valor de $\gamma = 0,5$, para os outros parâmetros foram mantidos os padrões do classificador. Este valor de γ apresentou melhores resultados que o valor padrão. Foram testados todos os grupos de características separadamente, utilizando a técnica de *leave-one-out* e também a técnica de *k-fold* com $k = 17$;

Teste 3:

O terceiro teste seguiu como base o segundo, porém com os valores de cada atributo normalizados entre 0 e 1. As configurações do classificador foram o tipo *ν-SVC*, com $\nu = 0,5$, com *kernel* radial e $\gamma = 0,5$. Estes valores apresentaram melhores resultados que os padrões. O tipo do classificador foi modificado pois o tipo *C-SVC* apresentou resultados baixos com os valores dos atributos normalizados. Novamente a amostra total de casos eram de 102, cada caso representando uma das mamas de cada paciente. Foram testados todos os grupos de características separadamente, utilizando a técnica de *leave-one-out* e também a técnica de *k-fold* com $k = 17$;

Teste 4:

No quarto teste utilizou-se a diferença entre os atributos computados para a mama esquerda e para a mama direita de cada paciente, somando um total de 51 casos, cada caso representando uma paciente.

Inicialmente, neste teste, foram utilizadas as diferenças sem realizar a etapa de normalização. O classificador utilizado foi configurado para o tipo ν -SVC, com $\nu = 0,01$ e *kernel* radial com $\gamma = 0,5$. Estes valores apresentaram resultados melhores que os valores padrões. Foram testados todos os grupos de características separadamente, utilizando a técnica de *leave-one-out* e também a técnica de *k-fold* com $k=17$;

Teste 5:

No Teste 5 foram utilizados os atributos do Teste 4, porém normalizados entre 0 e 1. O classificador *LibSVM* foi configurado para o tipo ν -SVC, com $\nu = 0,01$, com *kernel* radial com $\gamma = 0,01$. Estes valores apresentaram resultados melhores que os valores padrões. Novamente foram testados todos os grupos de características separadamente, e validadas com as técnicas de *leave-one-out* e também a técnica de *k-fold* com $k=17$;

Teste 6:

No *Teste 6* foram utilizadas somente as características baseadas em geoestatística. Foram utilizados como atributos a diferença normalizada entre o valor obtido pela mama esquerda e pela mama direita (os mesmos valores utilizados no *Teste 5*). O *LibSVM* foi utilizado como classificador, com os parâmetros padrões. Foram testadas cada uma das medidas de geoestatística separadamente (Grupo IV), isto é, coeficiente de Geary, índice de Moran e função *K* de Ripley. A técnica para validação e a avaliação dos resultados seguiu o que foi estabelecido no *Teste 1*.

Teste 7:

No *Teste 7* foram analisadas as 18 imagens do banco de dados da UFF, adquiridas no HUAP. Como dito anteriormente, resolveu-se testar estes termogramas separadamente pois o banco de dados ainda está em estágio inicial de formação e apresenta poucos casos com alguma informação de diagnóstico. Para estas imagens usaram-se os atributos e as configurações do classificador como no *Teste 5*. Porém, pela amostra ser pequena usou-se apenas a validação com *leave-one-out* e utilizando o Grupo VII com todos os métodos.

5.2 RESULTADOS

5.2.1 RESULTADOS OBTIDOS NO PRÉ-PROCESSAMENTO

Os primeiros resultados obtidos quando considerada a ordem em que as técnicas foram propostas na metodologia, são provenientes da correção da posição da paciente nas imagens. Para avaliar quantitativamente a técnica de registro desenvolvida, usou-se a segmentação automática proposta por Motta *et al.* [81], onde o autor descreve nos resultados que dos 151 termogramas testados, 38 não foram segmentados corretamente. Destas, 18 falhas foram atribuídas ao mau posicionamento das pacientes durante o exame e o restante (20) atribuídas ao baixo contraste das imagens.

Esses 151 termogramas foram registrados pela técnica de registro desenvolvida neste trabalho e novamente passaram pela segmentação automática. Na Tabela 9 são apresentados os resultados obtidos. Pode-se observar nos resultados que houve uma melhora de aproximadamente 16% no número de segmentações realizadas corretamente.

	Amostra	Segmentação correta (%)	Baixo contraste	Mau posicionamento
Sem registro	151	113 (74,8%)	20	18
Com registro	151	137 (90,7%)	8	6

Tabela 9: Análise dos resultados obtidos com o registro de imagem.

Esta mesma técnica de registro foi verificada em termogramas capturados através do protocolo dinâmico utilizado no HUAP, onde uma amostra de quatro pacientes com uma série de seis imagens capturadas com diferença de um minuto cada imagem foram utilizadas. O registro foi aplicado em cada uma das imagens a fim de reduzir o efeito causado pelo movimento da paciente durante o exame, o que pode ser uma dificuldade futura em análises realizadas com este protocolo. Foram calculados os valores do erro médio da subtração pixel-a-pixel das imagens da sequência em relação a imagem inicial, segundo a Equação (24), onde H e W são as dimensões da imagem, p_0 é a imagem inicial e p_s a imagem da sequência que está sendo analisada.

$$E_s = \frac{\sum_{i=0}^{H-1} \sum_{j=0}^{W-1} |p_{0(i,j)} - p_{s(i,j)}|}{H \cdot W} \quad (24)$$

Após adquirir estes valores de erro para cada imagem, foram calculados a média e o desvio padrão entre todas as imagens de cada paciente. A Tabela 10 mostra os resultados

obtidos. Pode-se observar uma redução de mais de 70% (de 31,47 para 11,62) na média do erro no melhor caso (Paciente 2).

	Paciente 1	Paciente 2	Paciente 3	Paciente 4
Média do erro sem o registro	7,91	31,47	17,57	16,39
Média do erro com o registro	5,20	11,62	11,03	8,25
Desvio padrão do erro sem o registro	2,07	12,15	3,71	4,11
Desvio padrão do erro com o registro	2,16	4,61	1,27	1,55

Tabela 10: Resultados obtidos com o registro no protocolo dinâmico.

5.2.2 RESULTADOS OBTIDOS NOS TESTES

Os resultados obtidos pelos testes para a detecção das alterações da mama serão apresentados a seguir.

Na Tabela 11 encontram-se os resultados obtidos pelo *Teste 1*. Observa-se que os Grupos I e II associados (Grupo V) obtiveram melhores resultados que o Grupo III na divisão por quadrantes. E ambos os grupos obtiveram melhores resultados quando calculados sobre a mama inteira que sobre a divisão.

Grupo	Mama	Sensib.	Espec.	Acurácia	ROC	Youden
III	Inteira	39,6%	63%	52%	0,513	0,025
V	Inteira	83,8%	28,6%	68,6%	0,562	0,124
III	Quadrantes	86,5%	14,3%	66,7%	0,504	0,008
V	Quadrantes	75,7%	28,6%	68,6%	0,521	0,042

Tabela 11: Resultados obtidos pelo *Teste 1*.

Na Tabela 12 e na Tabela 13 encontram-se os resultados obtidos pelo *Teste 2*, para a validação com *leave-one-out* (LOO) e *17-folds* (k17), respectivamente. Observa-se que tanto na validação LOO, quanto na k17, que o melhor resultado considerando a área sob a curva ROC e o índice de Youden foi obtido pelo Grupo I. Provavelmente isto deve-se à não normalização dos valores dos atributos, sendo que no Grupo I e no Grupo II havia os maiores valores numéricos. Pode-se observar também que as características baseadas em fractal

(Grupo III), obtiveram um baixíssimo rendimento e obtiveram resultados similar a um modelo de classificação que detecta alterações em 100% dos casos apresentados, sendo eles saudáveis ou não.

Grupo	Sensibilidade	Especificidade	Acurácia	Youden	ROC
I	54,2%	66,7%	60,8%	0,208	0,604
II	56,3%	59,3%	57,8%	0,155	0,578
III	0,0%	100,0%	52,9%	0,000	0,500
IV	39,6%	64,8%	52,9%	0,044	0,522
V	54,2%	63,0%	58,8%	0,171	0,586
VI	39,6%	66,7%	53,9%	0,063	0,531
VII	43,8%	72,2%	58,8%	0,160	0,580

Tabela 12: Resultados obtidos pelo *Teste 2* validados por *leave-one-out*.

Grupo	Sensibilidade	Especificidade	Acurácia	Youden	ROC
I	54,2%	68,5%	61,8%	0,227	0,613
II	54,2%	63,0%	58,8%	0,171	0,586
III	0,0%	100,0%	52,9%	0,000	0,500
IV	39,6%	68,5%	54,9%	0,081	0,541
V	54,2%	64,8%	59,8%	0,190	0,595
VI	41,7%	68,5%	55,9%	0,102	0,551
VII	43,8%	72,2%	58,8%	0,160	0,580

Tabela 13: Resultados obtidos pelo *Teste 2* validados por *17-folds*.

Na Tabela 14 e na Tabela 15 encontram-se os resultados obtidos pelo *Teste 3* para a validação com *leave-one-out* (LOO) e *17-folds* (k17), respectivamente. Observa-se que o melhor resultado obtido foi pelo Grupo V na validação por *17-folds*. Entretanto, obteve-se resultados pouco expressivos em todos os grupos, em alguns casos a área sob a curva ROC ficou abaixo de 0,5, o que indica uma classificação pior que um classificador que indica ou como problema ou como saudável 100% dos casos.

Grupo	Sensibilidade	Especificidade	Acurácia	Youden	ROC
I	58,3%	55,6%	56,9%	0,139	0,569
II	56,3%	57,4%	56,9%	0,137	0,568
III	43,8%	57,4%	51,0%	0,012	0,506
IV	41,7%	66,7%	54,9%	0,083	0,542
V	50,0%	63,0%	56,9%	0,130	0,565
VI	45,8%	57,4%	52,0%	0,032	0,516
VII	41,7%	53,7%	48,0%	-0,046	0,477

Tabela 14: Resultados obtidos pelo *Teste 3* validados por *leave-one-out*.

Grupo	Sensibilidade	Especificidade	Acurácia	Youden	ROC
I	58,3%	55,6%	56,9%	0,139	0,569
II	52,1%	57,4%	54,9%	0,095	0,547
III	47,9%	63,0%	55,9%	0,109	0,554
IV	39,6%	63,0%	52,0%	0,025	0,513
V	58,3%	63,0%	60,8%	0,213	0,606
VI	45,8%	53,7%	50,0%	-0,005	0,498
VII	41,7%	53,7%	48,0%	-0,046	0,477

Tabela 15: Resultados obtidos pelo *Teste 3* validados por *17-folds*.

Na Tabela 16 e na Tabela 17 encontram-se os resultados obtidos pelo *Teste 4* para a validação com *leave-one-out* (LOO) e *17-folds* (k17), respectivamente. Observa-se que o melhor resultado obtido foi pelo Grupo VII em ambas as validações.

Grupo	Sensibilidade	Especificidade	Acurácia	Youden	ROC
I	48,6%	57,1%	51,0%	0,058	0,529
II	43,2%	42,9%	43,1%	-0,139	0,431
III	64,9%	57,1%	62,7%	0,220	0,610
IV	78,4%	50,0%	70,6%	0,284	0,642
V	51,4%	35,7%	47,1%	-0,129	0,435
VI	70,3%	28,6%	58,8%	-0,012	0,494
VII	83,8%	57,1%	76,5%	0,409	0,705

Tabela 16: Resultados obtidos pelo *Teste 4* validados por *leave-one-out*.

Grupo	Sensibilidade	Especificidade	Acurácia	Youden	ROC
I	59,5%	28,6%	51,0%	-0,120	0,440
II	56,8%	64,3%	58,8%	0,210	0,605
III	59,5%	50,0%	56,9%	0,095	0,547
IV	73,0%	42,9%	64,7%	0,158	0,579
V	54,1%	35,7%	49,0%	-0,102	0,449
VI	70,3%	21,4%	56,9%	-0,083	0,458
VII	83,8%	57,1%	76,5%	0,409	0,705

Tabela 17: Resultados obtidos pelo *Teste 4* validados por *17-folds*.

Na Tabela 18 e na Tabela 19 encontram-se os resultados obtidos pelo *Teste 5* para a validação com *leave-one-out* (LOO) e *17-folds*, respectivamente. Observa-se que o Grupo I obteve melhor resultado na validação por *17-folds*, enquanto que na validação por LOO os melhores resultados foram obtidos pelo Grupo V.

Grupo	Sensibilidade	Especificidade	Acurácia	Youden	ROC
I	59,5%	64,3%	60,8%	0,237	0,619
II	59,5%	64,3%	60,8%	0,237	0,619
III	51,4%	35,7%	47,1%	-0,129	0,435
IV	64,9%	42,9%	58,8%	0,077	0,539
V	51,4%	78,6%	58,8%	0,299	0,650
VI	73,0%	28,6%	60,8%	0,015	0,508
VII	67,6%	50,0%	62,7%	0,176	0,588

Tabela 18: Resultados obtidos pelo *Teste 5* validados por *leave-one-out*.

Grupo	Sensibilidade	Especificidade	Acurácia	Youden	ROC
I	67,6%	57,1%	64,7%	0,247	0,624
II	56,8%	50,0%	54,9%	0,068	0,534
III	35,1%	50,0%	39,2%	-0,149	0,426
IV	70,3%	50,0%	64,7%	0,203	0,601
V	32,4%	85,7%	47,1%	0,181	0,591
VI	78,4%	28,6%	64,7%	0,069	0,535
VII	67,6%	50,0%	62,7%	0,176	0,588

Tabela 19: Resultados obtidos pelo *Teste 5* validados por *17-folds*.

Na Tabela 20 encontram-se os resultados obtidos pelo *Teste 6*, observa-se que entre os métodos utilizados o que apresentou melhores resultados, considerando todas as medidas de avaliação, foi o Índice de Moran. Pode-se destacar que o Índice de Moran obteve resultados melhores do que quando utilizado todo o Grupo IV.

	Sensib.	Espec.	Acurácia	ROC	Youden
Índice de Moran	83,8%	78,6%	82,4%	0,812	0,624
Coefficiente de Geary	91,9%	28,6%	74,5%	0,602	0,205
K de Ripley	81,1%	50%	72,5%	0,655	0,311

Tabela 20: Resultados obtidos pelo *Teste 6*.

Na Tabela 21 estão os resultados obtidos pelo *Teste 7*, onde foram utilizadas as imagens da UFF.

	Sensib.	Espec.	Acurácia	ROC	Youden
Grupo VII	61,5%	20,0%	50,0%	0,408	-0,185

Tabela 21: Resultados obtidos pelo *Teste 6*.

selecionado está diretamente relacionada ao valor obtido pela avaliação da função objetivo [111]. Cada novo indivíduo gerado pelo cruzamento recebe carga genética dos pais na proporção de 50% - 50%, 60% - 40% ou 70% - 30%, escolhido aleatoriamente. A nova população gerada apresenta um risco de 1% de sofrer mutação, quando isto ocorre, um dos indivíduos da população é selecionado e até no máximo 5 de seus cromossomos (aproximadamente 11%) são modificados. A Figura 29 mostra um esquema de como são geradas as populações seguintes a partir da população inicial.

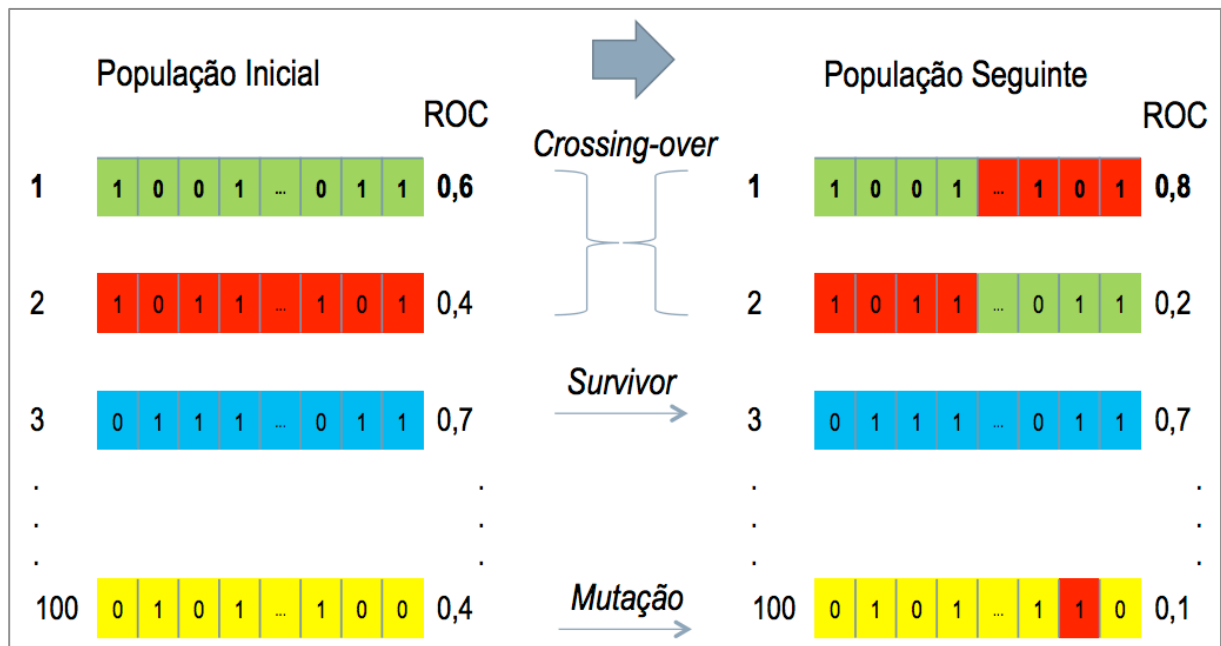


Figura 29: Exemplo de como são geradas as novas populações.

Com a realização de 1000 iterações, isto é, 1000 novas populações foram geradas e avaliadas, obtiveram-se os resultados presentes na Tabela 22. As características de cada grupo presentes neste indivíduo foram:

- Grupo I: Média;
- Grupo II: Nenhuma;
- Grupo III: Higuchi(5), Higuchi(25), Sucolaridade(L2R) e Sucolaridade(T2B);
- Grupo IV: Moran(3%, 0°), Moran(6%, 0°), Moran(9%, 0°), Moran(12%, 0°), Moran(15%, 0°), Geary(3%, 0°), Geary(6%, 0°), Geary(9%, 90°), Geary(12%, 90°) e Geary(15%, 90°).

Este será chamado de *Teste 7*.

	Sensib.	Espec.	Acurácia	ROC	Youden
Algoritmo genético	91,7%	78,6%	88,2%	0,852	0,705

Tabela 22: Resultados obtidos pelo melhor indivíduo após 1000 iterações.

5.2.4 RESUMO DOS RESULTADOS OBTIDOS NOS TESTES

A Tabela 23 apresenta um resumo de todos os testes realizados. Nesta tabela são apresentados a quantidade de casos em cada teste, a especificação do que pretendia-se avaliar em cada um e o grupo ou método que obteve os melhores resultados considerando a área sob a curva ROC.

Teste	Casos	Especificação	Melhor Grupo	Maior Área ROC	Maior Índice Youden
1	51	Divisão em quadrantes vs. Mama inteira.	V	0,562	0,124
2	102	Cada mama um caso, sem normalização.	I	0,613	0,227
3	102	Cada mama um caso, atributos normalizados.	V	0,606	0,213
4	51	Cada paciente um caso, atributos subtraídos da mama esquerda pela mama direita, sem normalização.	VII	0,705	0,409
5	51	Similar ao Teste 4, porém com os resultados da subtração normalizados.	V	0,624	0,247
6	51	Teste para avaliar cada medida geoestatística separadamente.	Índice de Moran	0,812	0,624
7	51	Teste para otimizar os resultados selecionando os melhores atributos.	15 atributos	0,852	0,705

Tabela 23: Resumo de todos os testes realizados.

5.2.5 THERMOCAD-UFF

Durante o desenvolvimento desta tese foi desenvolvido um sistema para auxílio à detecção de alterações na mama. Este sistema foi denominado *ThermoCAD-UFF* e estará disponível publicamente no site do Projeto PROENG [90]. Nesta seção será discutido o protótipo deste sistema.

Os recursos do sistema estão divididos em três principais módulos: gerenciamento, análise e treinamento do classificador. Entre os recursos disponíveis no sistema destacam-se o gerenciamento dos termogramas, com a capacidade de integração com a câmera para aquisição de novas imagens, além da possibilidade de conexão com o banco de dados do projeto hospedado em um servidor da UFF, como mostra a Figura 30.

Na tela de análise apresentada na Figura 31 é possível realizar a segmentação automática dos termogramas, selecionar a paleta de cores para representar as temperaturas, visualizar o valor de temperatura de cada um dos pontos da imagem, utilizar ferramenta de zoom, utilizar a função de auto ajuste descrita na Seção 4.5 e realizar a análise auxiliar do termograma, onde serão extraídas as características utilizadas nesta tese e o termograma será classificado de acordo com os dados previamente treinados pelos testes realizados na tese.

Na tela de treinamento do classificador é possível realizar um novo treinamento no classificador do sistema através de um conjunto de termogramas com diagnóstico conhecido. Neste treinamento é possível selecionar entre os métodos que foram utilizados nesta tese, quais devem ser extraídos e analisados.

Além destas funcionalidades, o sistema apresenta algumas funcionalidades extras, tais como a extração de matrizes de temperatura de termogramas em lote, permitindo salvá-las em diversos formatos, a extração da ROI de diversos termogramas em lote e a extração de características de diversos termogramas que serão salvas em um formato .csv.

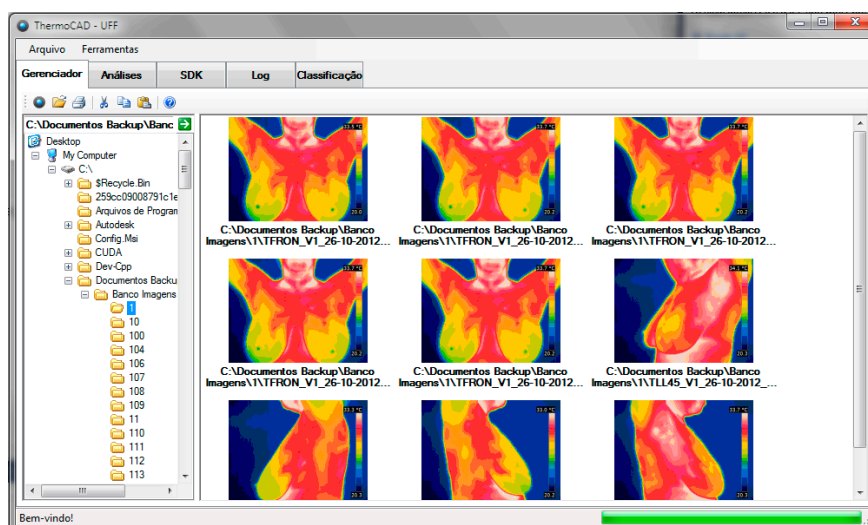


Figura 30: Captura da tela de gerenciamento do ThermoCAD-UFF.

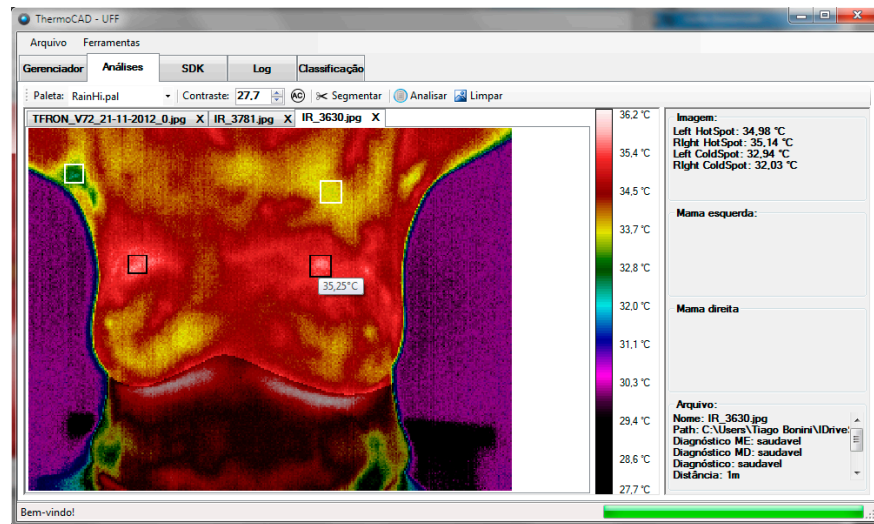


Figura 31: Tela de análise de termogramas.

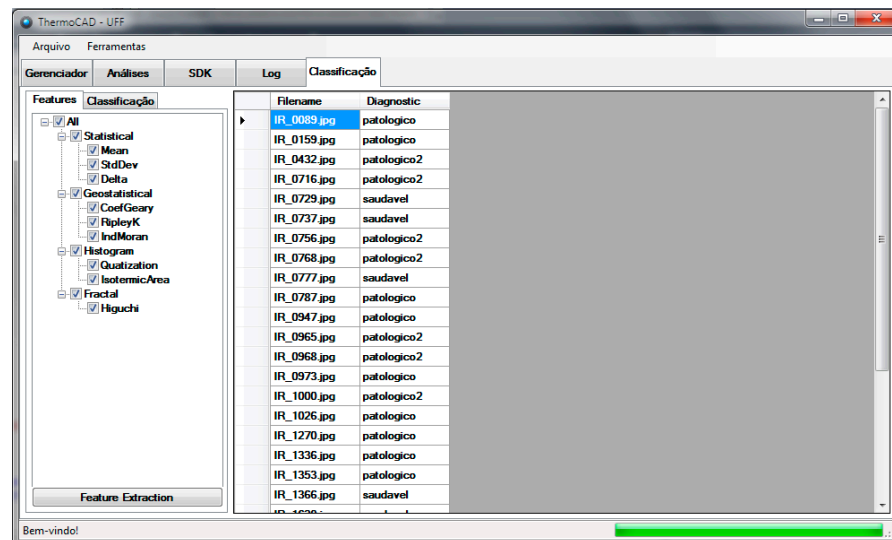


Figura 32: Tela de treinamento do classificador.

CAPÍTULO 6 – TRABALHOS RELACIONADOS

Durante a revisão bibliográfica foram selecionados alguns trabalhos relacionados ao uso da termografia para a detecção e/ou diagnóstico de doenças da mama para servirem de comparação para os resultados obtidos neste trabalho. Os trabalhos relacionados foram analisados considerando as características extraídas, os classificadores utilizados e as ferramentas CAD desenvolvidas. Estes trabalhos serão descritos a seguir, as comparações encontram-se na Seção 6.1.1.

Koay *et al.* [56] desenvolveram um estudo utilizando imagens termográficas de dezenove pacientes, sendo que catorze eram de pacientes saudáveis e cinco de pacientes com alguma doença na mama. Os autores extraíram dez características da imagem termográfica para cada quadrante da mama. Os quadrantes são definidos tendo como ponto comum o mamilo. As características extraídas foram: temperatura média, desvio padrão, mediana, máximo, mínimo, obliquidade (*skewness*), curtose (*kurtosis*), entropia, área e quantidade de calor (entalpia). Após a extração de características, os autores usaram o software estatístico SPSS (*Statistical Package for the Social Sciences*) para determinar a correlação entre as características. Eles reportaram que cinco características são altamente correlacionadas, sendo redundantes e selecionaram para análise as seguintes características: temperatura média, desvio padrão, assimetria, curtose e quantidade de calor.

Os autores descreveram diversas configurações de Redes Neurais Artificiais (RNA) que foram aplicadas sem sucesso para a classificação do subgrupo das cinco características selecionadas. Após reduzir o subgrupo de características para apenas duas (média e desvio padrão), foi utilizada uma RNA com uma única camada oculta reportando apenas dois falsos-negativos. De acordo com os autores, a possível explicação para os fracos resultados obtidos foi o pequeno número de imagens de pacientes com câncer (apenas cinco).

Os resultados obtidos foram: sensibilidade de 60%, especificidade de 100% e acurácia de 89,5%.

Tang *et al.* [68] analisaram termogramas de cento e dezessete pacientes, sendo quarenta e sete com tumores malignos e setenta com tumores benignos. As imagens foram adquiridas utilizando o TSI-21 Thermographic System. Os autores propuseram a análise de uma característica que eles denominaram Aumento de Temperatura Localizada (em inglês, *Localized Temperature Increases* - LTI) nas imagens termográficas da mama. O LTI representa pontos ou padrões vasculares anormais no termograma, com temperaturas muito elevadas.

O processo apresentado foi dividido em dois passos para a detecção e diagnóstico de câncer de mama: (i) busca visual de regiões suspeitas na mama (detecção), e (ii) a medida da amplitude do LTI na região (diagnóstico). Resumindo, o método proposto considera que a possibilidade da presença de um câncer é proporcional a amplitude máxima da temperatura na pele na região suspeita. Os autores utilizaram processamento de sinal morfológico para determinar o LTI e qual o limiar ótimo para separar casos com câncer e normais. Para avaliar os resultados foi utilizada a Curva ROC e o índice de Youden ($Y = \text{Sensibilidade} + \text{Especificidade} - 1$). Utilizando esta metodologia, os autores obtiveram 93,6% de sensibilidade, 44,3% de especificidade e $Y = 0,38$.

Arora *et al.* [58] utilizaram o *Sentinel BreastScan*TM (SBS) para a aquisição e análise das imagens termográficas. Durante a realização do exame a paciente senta-se em uma cadeira, despida da cintura para cima, com os braços em um apoio lateral. Quando o exame começa, as imagens termográficas das pacientes são armazenadas para análise posterior. Durante alguns segundos uma fonte de ar frio é direcionada para as mamas da paciente. O exame tem duração de aproximadamente 3 minutos. O SBS busca identificar no termograma sinais de temperaturas distribuídas de forma anormal. Após a análise um relatório é gerado pelo SBS.

Os autores analisaram noventa e duas pacientes, sendo cinquenta e oito casos com tumores malignos e trinta e quatro com tumores benignos. Cada paciente passa por três tipos de análises, nomeados pelos autores como: (i) *screening*, (ii) clínico e (iii) por rede neural artificial. Na primeira análise as pacientes são ranqueadas de acordo com o risco de apresentar câncer de mama em uma escala de zero (sem risco) até sete (com alto risco). Na análise clínica os autores utilizam a mamografia e o ultrassom para verificar a presença ou ausência de alterações. Na última análise uma RNA é utilizada para identificar a presença de tumores malignos na mama. Os resultados relatados pelos autores no uso de RNA tiveram 97% de sensibilidade, 44% de especificidade e 82% de valor preditivo negativo.

Wishart *et al.* [7] realizaram um estudo sobre a performance das imagens termográficas no auxílio ao diagnóstico de câncer de mama em cem (100) pacientes que seriam submetidas a biopsia. Eles utilizaram o SBS [58] e compararam os resultados obtidos com outros quatro tipos de análises: (i) *screening*, (ii) RNA, (iii) manual, realizada por especialistas e (iv) um método baseado em inteligência artificial desenvolvido pelos autores, denominado *NoTouch BreastScan* (NTBS). As características utilizadas foram idade (<50, 50 a 70, >70), tamanho do tumor ($\leq 2\text{cm}$, $>2\text{cm}$), grau do tumor (1, 2, 3), estado nodal (negativo, positivo), morfologia (ductal, lobular, tubular, misto, DCIS) e invasão vascular (sim, não).

Quando os autores repetiram os experimentos de Arora *et al.* utilizando o SBS, os resultados obtidos mostraram uma baixa sensibilidade nos dois casos testados, 53% no *screening* e 48% no método baseado em RNA, resultados bem inferiores aos publicados. Os resultados obtidos utilizando o NTBS foram 48% de sensibilidade e 70% de especificidade.

Umadevi *et al.* [112] desenvolveram o software chamado *Infrared Thermography Based Image Construction* (ITBIC) para a interpretação de termogramas da mama. O sistema desenvolvido captura três imagens de cada paciente (frontal, lateral esquerda e lateral direita). Após a captura, as áreas de maior temperatura nas imagens são extraídas e uma imagem simplificada apenas com essas áreas é gerada para a análise. A análise é realizada a partir de decisões lógicas baseadas nas áreas quentes que foram extraídas no passo anterior.

Os autores reportam os seguintes resultados obtidos pelo ITBIC, na interpretação das imagens termográficas de cinquenta pacientes: valor preditivo positivo de 80%, valor preditivo negativo de 95,6%, sensibilidade de 66,7% e especificidade de 97,7%.

Acharya *et al.* [64] apresentaram um estudo com cinquenta termogramas de mama, sendo 50% normal e 50% com câncer de mama. Os autores extraíram das imagens um grupo de dezesseis características baseadas em textura: homogeneidade, energia, entropia, os quatro primeiros momentos calculados da matriz de co-ocorrência, segundo momento angular, contraste, média, ênfase em primitivas curtas, ênfase em primitivas longas, uniformidade do nível de cinza e percentagem de primitivas. Porém, os autores selecionaram apenas quatro características como clinicamente significante em comparação as outras: primeiro momento, terceiro momento, percentagem de primitivas e uniformidade do nível de cinza. O classificador utilizado foi o SVM.

Os resultados relatados pelos autores foram 85,7% de sensibilidade, 90,5% de especificidade e 88,1% de acurácia.

6.1.1 COMPARAÇÕES

Os resultados obtidos nesta tese foram comparados com os resultados obtidos pelos trabalhos relacionados descritos acima. Esta comparação pode ser vista na Tabela 24, onde são comparadas as avaliações de sensibilidade (Sens.), especificidade (Espec.), acurácia (Acur.) e o índice de Youden. Além destes avaliadores, a tabela também apresenta a amostra utilizada por cada trabalho, onde é especificado a relação entre o número de casos saudáveis pelo número de casos patológicos ou o número de casos benignos pelo número de casos malignos. As amostras utilizadas por Koay *et al.* [56], Umadevi *et al.* [112], Acharya *et al.* [64] e por esta tese estão separadas entre casos saudáveis e mamas com problemas. As

amostras utilizadas por Tang *et al.* [68], Arora *et al.* [58] e Wishart *et al.* [7] estão separadas entre casos benignos e casos malignos.

	Amostra	Sens.	Espec.	Acur.	Youden
Koay <i>et al.</i> (2004)	19 (14/5)	60%	100%	89,5%	0,60
Tang <i>et al.</i> (2008)	117 (70/47)	93,6%	44,3%	-	0,38
Arora <i>et al.</i> (2008)	94 (34/60)	97%	44%	-	0,41
Wishart <i>et al.</i> (2010)	106 (41/65)	48%	70%	-	0,18
Umadevi <i>et al.</i> (2010)	50 (44/6)	66,7%	97,7%	-	0,64
Acharya <i>et al.</i> (2012)	50 (25/25)	85,7%	90,5%	88,1%	0,76
Grupo VII (Teste 4)	51 (14/37)	83,8%	57,1%	76,5%	0,41
Índice de Moran	51 (14/37)	83,8%	78,6%	82,4%	0,62
Resultado Otimização	51 (14/37)	91,9%	78,6%	88,2%	0,71

Tabela 24: Comparação de resultados.

Analisando os resultados obtidos pelos métodos utilizados nesta tese em relação aos resultados encontrados na literatura, pode-se observar que os resultados obtidos são competitivos. Entretanto, a comparação direta entre os valores obtidos nas medidas de avaliação dos diferentes métodos não é completamente leal, pois cada trabalho utiliza uma base de dados diferente, com diferentes distribuições das amostras, isto é, algumas amostras apresentam um número maior de casos saudáveis, enquanto outras apresentam um número maior de casos com doenças. Este desbalanceamento da amostra pode influenciar nos resultados da classificação.

Os resultados obtidos pelas características selecionadas pelo algoritmo genético nesta tese foram melhores que a maioria dos trabalhos encontrados na literatura. Esses resultados ficaram muito próximos aos resultados obtidos por Acharya *et al.* [64]. A acurácia de ambos os trabalhos ficaram com uma diferença de apenas 0,1%. Na sensibilidade o trabalho desenvolvido nesta tese alcançou valores melhores, 91,9% contra 85,7% de Acharya *et al.* [64]. Porém, obtive uma especificidade menor, 78,6% contra 90,5%. Após discussões, acredita-se que os motivos pelo qual os resultados desta tese não superaram os encontrados por Acharya *et al.* [64], seja devido a amostra utilizada apresentar diversos tipos de alterações, desde lipomas e cistos até tumores não classificados, enquanto que a amostra utilizada por Acharya *et al.* [64] além de ser perfeitamente balanceada (25 casos saudáveis e 25 com alterações), todos os casos com alterações eram de pacientes com tumores malignos

confirmados. Outras observações interessantes são o SVM que foi utilizado como classificador por ambos os trabalhos e a amostra de termogramas com números muito próximos de casos, 51 e 50.

De todas as combinações de métodos utilizados nesta tese os métodos selecionados pela otimização utilizando o algoritmo genético foram os que apresentaram melhores resultados. Esta combinação de métodos apresenta resultados próximos aos melhores presentes na literatura, porém para obter esta combinação de métodos foi preciso dispêndio de um longo tempo computacional processando cada uma das combinações presentes nas populações do algoritmo genético, além de ter que repetir este processo muitas vezes.

CAPÍTULO 7 – DISCUSSÕES FINAIS

Neste capítulo serão apresentadas as conclusões deste trabalho, os possíveis trabalhos futuros a serem desenvolvidos e a lista de publicações obtidas durante o desenvolvimento da tese.

7.1 DISCUSSÃO DOS RESULTADOS

Nesta seção serão discutidos os resultados obtidos nesta tese, desde o pré-processamento até os testes de classificação de alterações.

Sobre a técnica de registro de imagem, para a correção do mau posicionamento das pacientes, pode-se verificar bons resultados obtidos. Isto indica que mesmo sendo uma técnica relativamente simples, que utiliza conceitos de matrizes de transformações geométricas, ela pode trazer melhorias na etapa de segmentação quando utilizada como pré-processamento. Este melhor posicionamento melhora os algoritmos de segmentação automática de termogramas, pois o mau posicionamento das pacientes durante a captura é um dos principais problemas encontrados nas falhas na definição automática da ROI. E por consequência, os métodos computacionais baseados em comparações de simetrias bilaterais alcançam melhores resultados quando a posição da paciente está correta na imagem analisada. No teste realizado com o segmentador automático de Motta *et al.* [81], observou-se um aumento de segmentações corretas, após o registro, de 113 para 137 casos, no total de 151. Esta melhora representa um ganho de 16% no número de segmentações automáticas. No teste realizado com as imagens sequenciais do protocolo dinâmico da UFF, a média do erro gerado na comparação entre os termogramas sequenciais de cada paciente reduziu consideravelmente. Este erro é gerado pela movimentação da paciente durante a execução do exame. No melhor caso obteve-se uma redução de 70% no valor do erro. O erro foi reduzido de 31,47 para 11,62.

Sobre o *Teste I* pode-se ressaltar que a divisão da mama em quadrantes para o cálculo das características não contribuiu para alcançar melhores resultados, inclusive apresentou o efeito contrário. Analisando os resultados obtidos pelos Grupos III e V no índice de Youden, pode-se observar que os valores obtidos aumentam de 0,008 para 0,025 e de 0,042 para 0,124, respectivamente, quando utilizada a mama inteira. Em muitos trabalhos é observada a divisão da mama em regiões menores para a realização de análises mais localizadas, porém no caso dos termogramas a divisão em quadrantes não foi satisfatória.

Analisando os *Testes 2 e 3* observa-se que a não normalização dos atributos antes da avaliação pode comprometer os resultados, o mesmo conjunto de dados foi utilizado para ambos os testes, sendo a normalização dos valores dos atributos a única diferença. Em geral o *Teste 2* obteve resultados abaixo do *Teste 3*. Entretanto, se analisado apenas o melhor resultado de cada teste, o *Teste 2* obteve com o Grupo I o melhor valor de área sob a Curva ROC (ROC = 0,613) em relação ao *Teste 3*. Em ambos os testes apenas um grupo de características alcançou um valor de ROC acima de 0,6. A maioria dos grupos apresentaram um valor de ROC muito próximo a 0,5, sendo estes resultados muito baixos. Estes resultados indicam o que era esperado, isto é, a normalização dos valores dos atributos realmente melhora os resultados em uma classificação [106]. Entretanto, o baixo valor obtido na curva ROC indica que esta abordagem de analisar as mamas separadamente, não alcança bons resultados com as características utilizadas.

Analisando os *Testes 4 e 5* em relação aos *Testes 2 e 3*, pode-se perceber que quando realizada a subtração dos valores dos atributos das mamas de uma mesma paciente, os resultados alcançados na classificação são melhores do que quando cada mama é analisada separadamente. Quando o *Teste 4* é analisado em relação ao *Teste 5* ocorre o mesmo que na análise entre os *Testes 2 e 3*. Em geral o *Teste 5* que utiliza valores de atributos normalizados obtém resultados melhores, porém o *Teste 4* obteve o melhor resultado entre os dois, com o uso do Grupo VII (ROC = 0,705). Destes testes pode-se concluir que utilizando estas características, a subtração entre os valores dos atributos obtidos em cada mama é uma abordagem melhor que utilizar cada mama separadamente.

O *Teste 6* mostrou que é possível obter resultados relevantes na classificação de alterações da mama em termogramas adaptando funções amplamente utilizadas em geoestatística. O Índice de Moran se destacou alcançando os melhores resultados entre os três métodos quando analisados separadamente. Os resultados alcançados pelo Índice de Moran estão entre os melhores obtidos em todos os testes realizados nesta tese. Conclui-se que caso seja necessário escolher apenas um método para realizar a classificação de alterações nas mamas por imagens termográficas, o Índice de Moran é o mais indicado.

O *Teste 7* apresentou resultados fracos, após discussões no grupo de pesquisa, esses resultados foram atribuídos a pouca diferenciação existente entre os casos saudáveis e com alterações, pois esta separação é apenas entre os BI-RADS 1 e 2. Outro fator, talvez menos relevante, é o baixo volume de casos com classificação, a amostra era de apenas 18 imagens. Espera-se que após o banco de dados da UFF apresentar mais casos com classificações de BI-RADS maiores (4 ou 5), os resultados de testes similares a este melhorem.

Os resultados alcançados pelos métodos selecionados através do uso do algoritmo genético não foram ótimos, mas foram os melhores resultados produzidos. Quando o fator mais importante para a aplicação é a capacidade de classificar corretamente os casos saudáveis e com alterações a busca da otimização é fundamental. Mesmo que a tentativa de alcançar resultados ótimos seja um processo que demanda muito tempo de processamento, para analisar inúmeras combinações de métodos, o ganho alcançado em acurácia, sensibilidade e especificidade na classificação compensa. Nas características selecionadas pode-se observar que o grupo que mais colaborou na seleção foi o Grupo IV, das medidas geoestatísticas. Das 20 características selecionadas, 10 pertenciam ao Grupo IV, 9 ao Grupo III e 1 ao Grupo I. A partir desta seleção, pode-se concluir que apesar do índice de Moran obter bons resultados como mostrado no *Teste 6*, existem combinações com outros métodos que permitem alcançar resultados ainda melhores.

O sistema desenvolvido que contempla os métodos e a metodologia utilizada nesta tese pode ser muito útil para a análise de termogramas, entretanto este sistema ainda não foi testado com usuários especialistas em mastologia e radiologistas, visto que estes são os principais usuários a serem alcançados por um sistema CAD para termografia.

7.2 CONCLUSÕES

Neste trabalho foram utilizadas um total de 69 imagens de dois bancos de dados distintos, a lista de imagens utilizadas pode ser encontrada no Anexo A. Do total de 69 termogramas utilizados, 50 eram de pacientes que apresentavam algum problema em pelo menos uma das mamas e apenas 19 eram de pacientes saudáveis. Utilizando esta amostra foi possível chegar a resultados competitivos em comparação a outros trabalhos presentes na literatura.

Os termogramas utilizados seguiram dois protocolos diferentes de aquisição, o primeiro aplicado no Hospital das Clínicas pelo grupo do projeto na UFPE e o segundo aplicado no Hospital Universitário Antônio Pedro pelo grupo da UFF. A combinação na amostra de imagens de diferentes protocolos indica que os métodos propostos nesta tese não estão condicionados a apenas um estilo de captura de termogramas e nem a termogramas produzidos por um único modelo de câmera, visto que câmeras diferentes são utilizadas em cada hospital.

A segmentação das mamas dentro dos termogramas foi realizada após a aplicação de uma técnica de registro de imagem, desenvolvida para corrigir a posição das pacientes nos

exames. A técnica de segmentação utilizada foi desenvolvida por Marques [83] e é automática. Em alguns casos (16 termogramas) a segmentação automática não foi satisfatória, então apenas nestes termogramas foram utilizadas ROIs segmentadas através do aplicativo *Ground Truth Maker*, por radiologistas e especialistas em mastologia do HUAP.

A utilização de medidas estatísticas simples e de medidas baseadas em histograma, como descritores para a detecção de regiões suspeitas em termogramas, alcançou resultados relevantes em alguns casos. Visto que estas são medidas fáceis de serem obtidas computacionalmente e simples de serem entendidas por profissionais da área da saúde, que em geral não estão habituados com o cálculo de medidas complexas, como por exemplo, as medidas geoestatísticas e de dimensão fractal. Os melhores resultados alcançados com o uso destes grupos de atributos foram de 51,4% de sensibilidade, 78,6% de especificidade e 58,8% de acurácia, com índice de Youden de 0,299 e área sob a curva ROC em 0,650. Estes resultados foram obtidos pelo Grupo V que reúne todas as características de estatística e histograma no *Teste 5*, onde é realizada a subtração dos atributos obtidos para cada mama e estes valores são normalizados para cada atributo.

A utilização da dimensão fractal de Higuchi e da sucularidade não apresentaram resultados tão bons quanto eram esperados, pois obtiveram em média baixa sensibilidade, chegando a 0% em alguns casos. Esta má atuação pode estar atrelada à má escolha das escalas para o método de Higuchi, pois não foi encontrada uma base teórica que auxiliasse na escolha das mesmas, sendo estas definidas empiricamente.

As medidas geoestatísticas utilizadas apresentaram bons resultados em relação aos outros grupos de métodos, com destaque para o Índice de Moran que obteve 83,8% de sensibilidade, 78,6% de especificidade, 82,4% de acurácia e área sob a curva ROC de 0,812. Estes resultados indicam que o uso destas medidas representam uma boa abordagem para estudos na área de detecção e diagnóstico de alterações da mama através da análise de temperaturas.

O classificador SVM, já utilizado em diversas aplicações computacionais, obteve bons resultados sendo utilizado para a detecção de doenças de mama a partir de atributos extraídos dos termogramas. A união do SVM com a técnica de algoritmo genético para seleção dos melhores atributos fez com que os resultados obtidos por este trabalho melhorassem consideravelmente. Os resultados obtidos pela classificação, utilizando apenas os atributos selecionados, ficaram entre os melhores encontrados na literatura.

Por fim, foi possível desenvolver uma ferramenta para o auxílio à detecção de regiões suspeitas nas mamas em imagens termicas utilizando a metodologia desenvolvida nesta tese.

A ferramenta *ThermoCAD-UFF* conta com funcionalidades para a segmentação da região de interesse, a extração de medidas sobre as temperaturas e a classificação realizada pelo SVM, utilizando o modelo treinado neste trabalho.

A termografia da mama é um exame mais fácil, barato e simples de ser realizado que outros. O aumento do poder de detecção de doenças nas mamas através de metodologias computacionais poderá vir a permitir que a termografia seja utilizada como um exame de triagem realizado em conjunto com o exame clínico. A partir do resultado da termografia, o médico pode sugerir a realização de um exame mais aprofundado, que apresente as estruturas internas da mama. Com a apresentação das metodologias desenvolvidas nesta tese, espera-se que a termografia possa se firmar como um exame e passe a ser utilizado em um número maior de clínicas e hospitais no Brasil.

7.3 CONTRIBUIÇÕES

Ao final desta tese é possível destacar que os objetivos listados na introdução foram alcançados e geraram contribuições importantes. Estas contribuições serão listadas a seguir na ordem em que aparecem na tese.

- Revisão bibliográfica sobre o uso da termografia da mama: Esta revisão foi resumida no Capítulo 3 e publicada no periódico internacional com *qualis* A1, o *Signal Processing* (ver Seção 7.5.2).
- Técnica de registro de imagem para correção da posição corporal das pacientes nos termogramas: Esta técnica foi apresentada na Seção 4.6 e publicada primeiramente nos anais de um congresso nacional e posteriormente, mais aprimorada, nos anais de um congresso internacional.
- Técnica de ajuste automático do contraste nos termogramas: Esta técnica foi apresentada na Seção 4.5 e publicada juntamente com a técnica de registro.
- **Metodologia de auxílio à classificação de alterações na mama através de termogramas:** Esta é a principal contribuição apresentada nesta tese, descrita no Capítulo 4. Apesar de existirem outras metodologias de classificação na literatura, poucas têm resultados tão significativos quanto a metodologia desenvolvida. Além disso, não foi encontrado nenhum trabalho que descreve-se a metodologia de forma tão detalhada, desde a aquisição dos termogramas até a seleção das melhores características, como foi feito nesta tese.

- União dos trabalhos desenvolvidos anteriormente pelo grupo de pesquisa: Esta é outra contribuição importante, pois nesta tese foi possível agregar diversos projetos que já haviam sido desenvolvidos pelo grupo de pesquisa, desde os protocolos de aquisição desenvolvido por Silva *et al.* [67], os bancos de imagens, as segmentações de Motta *et al.* [81] e Marques [83], [96], os testes de classificação realizados por Resmini [108], até a seleção de classificadores baseada no trabalho de Serrano [44].
- Adaptação de diversos métodos computacionais para extração de características de termogramas: Diversos métodos baseados em estatística, histograma, fractal e geoestatística foram descritos na Seção 4.8. Todos estes métodos foram computados sobre matrizes de temperaturas.
- Técnica de seleção de características associando o algoritmo genético com o classificador *LibSVM*: Os testes mostraram que os resultados melhoraram após a aplicação da seleção de características descrita na Seção 5.2.3. Esta técnica pode ser utilizada em outras aplicações.
- Um sistema *CADe (Computer-Aided Detection)* com as metodologias desenvolvidos: Ao final desta tese foi desenvolvido o *software ThermoCAD-UFF*, que engloba todas as metodologias e técnicas desenvolvidas neste trabalho.

7.4 TRABALHOS FUTUROS

Este trabalho faz parte de um projeto com diversas linhas de estudo sobre o uso de imagens termográficas na mastologia. A seguir serão listadas sugestões de trabalhos futuros que podem melhorar os resultados obtidos por este trabalho, além de possibilitar uma continuidade do mesmo por outros membros do projeto.

- Aumentar o número de pacientes no banco de dados com diagnóstico fechado, principalmente no DMI, hospedado pela UFF, onde o número de casos fechados ainda é muito baixo. Atualmente, este trabalho vêm sendo desenvolvido por alunos de iniciação científica e em breve estará disponível um banco de imagens termográficas ainda mais completo.
- Incluir informações do prontuário da paciente como características, muitas informações do prontuário são relevantes na detecção de doenças, pois existem fatores de risco que aumentam a predisposição das pacientes apresentarem carcinomas ou outras lesões.

- Comparar os resultados obtidos pelo classificador *SVM* com outros classificadores amplamente utilizados na literatura, como redes neurais, árvores de decisão e classificadores *Fuzzy*. Um método prático para obter essas comparações é a utilização de *softwares* especializados como o *Weka* [46], que apresenta uma grande variedade de classificadores já implementados para uso.
- Este trabalho se restringiu a classificar os termogramas apenas como saudáveis ou com suspeita de lesão. Com uma amostra maior de termogramas, onde estejam identificadas a posição e o tipo de lesão pode-se aprofundar a classificação e possivelmente classificar os termogramas quanto à lesão apresentada.
- Conforme um tumor ou câncer cresce, estudos mostram que ele consome mais nutrientes no organismo e necessita de uma vascularização maior para alimentá-lo. Assim, com a evolução do tumor acredita-se que ele destaque-se cada vez mais na termografia. Um trabalho interessante é realizar uma análise que comprove que esta evolução do tumor está relacionada com o aumento da temperatura, para isso é necessário fazer um acompanhamento temporal através de termogramas de pacientes já diagnosticadas com câncer.
- O protocolo de aquisição utilizado no HUAP realiza uma série de imagens sequenciais, para acompanhar o reaquecimento das mamas das pacientes após um processo de resfriamento. Outro estudo importante a ser realizado é tentar identificar a relação entre a velocidade do reaquecimento das mamas e a presença de lesões.
- Sobre os termogramas do protocolo sequencial, já foram realizados alguns trabalhos, um dos principais trabalhos aborda o registro dos termogramas, de forma a posicionar corretamente o corpo da paciente na mesma posição em todos os termogramas da série para a análise. Outro trabalho desenvolvido realiza a remoção de ruídos nas imagens utilizando bases *Wavelets*. Entretanto, neste protocolo ainda existem muitas possibilidades de trabalhos à serem desenvolvidos.
- Os métodos propostos podem ser adaptados para trabalhar com outros tipos de imagens mastológicas. Como atualmente o banco de imagens da UFF também está armazenando mamografias e ultrassons da mama, um estudo a ser desenvolvido é a análise conjunta de características extraídas de diferentes exames.

7.5 PUBLICAÇÕES

Durante o desenvolvimento desta tese foram publicados: um capítulo de livro, 3 artigos completos em periódicos internacionais com corpo de revisores, um em revista nacional, 2 resumos em periódicos internacionais e diversos trabalhos em congressos. A grande maioria deles relacionados à análise de imagens termográficas como ferramenta para a detecção de alterações na mama. A seguir serão listadas essas publicações:

7.5.1 CAPÍTULO DE LIVRO

L.A. Bezerra, M.M. Oliveira, M.C. Araujo, M.J.A. Viana, L.C. Santos, F.G.S. Santos, T.L. Rolim, P.R.M. Lyra, R.C.F. Lima, **T.B. Borchardt**, R. Resmini, A. Conci, “Infrared Imaging for Breast Cancer Detection with Proper Selection of Properties: From Acquisition Protocol to Numerical Simulation.” In: E.Y.K. Ng; U.R. Acharya; R.M. Rangayyan; J.S. Suri. (Org.). *Multimodality Breast Imaging: Diagnosis and Treatment*. 1ed. Bellingham, WA: Society of Photo Optical, v. PM227, p. 285-332, 2013.

7.5.2 ARTIGOS EM PERIÓDICOS INTERNACIONAIS

T.B. Borchardt, A. Conci, R.C.F. Lima, R. Resmini, A. Sanchez, “Breast thermography from an image processing viewpoint: A survey.” *Signal Processing* (In Press), 2012. (<http://www.sciencedirect.com/science/article/pii/S0165168412002794>)

T.B. Borchardt, R. Resmini, L.S. Motta, E.W.G. Clua, A. Conci, M.J.A. Viana, L.C. Santos, R.C.F. Lima, A. Sanchez, “Combining approaches for early diagnosis of breast diseases using thermal imaging.” *International Journal of Innovative Computing and Applications*, v. 4, p. 163-183, 2012.

A.C.B. Romero, P.H.P. Aguiar, **T.B. Borchardt**, A. Conci, “Quantitative Ventricular Neuroendoscopy Performed on the Third Ventriculostomy: Anatomic Study.” *Neurosurgery*, v. 68, p. ons347-ons354, 2011.

7.5.3 ARTIGOS EM PERIÓDICOS NACIONAIS

R. Resmini, A. Conci, **T.B. Borchardt**, R.C.F. Lima, A.A. Montenegro, C.A.P. Fontes, “Diagnóstico precoce de doenças mamárias usando imagens térmicas e aprendizado de máquina.” *Revista Eletrônica do Alto Vale do Itajaí*, v. 1, p. 55-67, 2012.

7.5.4 RESUMOS EXPANDIDOS EM PERIÓDICOS

S. Vasconcelos, **T.B. Borchardt**, R. Resmini, A. Conci, C.A.P. Fontes, R.C.F. Lima, “On the Breast Reconstruction by Thermal Images.” *Thermology International*, v. 20, p. 134-135, 2010.

A. Conci, **T.B. Borchardt**, R. Resmini, R.C.F. Lima, C.A.P. Fontes, “A New Method to Aid to the Breast Diagnosis Using Fractal Geometry.” *Thermology International*, v. 20, p. 135-136, 2010.

7.5.5 ARTIGOS EM ANAIS DE CONGRESSOS NACIONAIS

- 2013
 - (1) J.G. Gazolla, R. Resmini, **T.B. Borchardt** e A. Conci, “On eficiente registration for breast infrared images”, no. 63, In: ENEBI 2013: Vitoria, 2013.
 - (2) T.A.E. da Silva, **T.B. Borchardt**, L.F. da Silva, J.G.F.M. Gazolla, R. Resmini, F.A. de Araújo, A. Conci, C.A.P. Fontes, M.L.O. Santos, A. Sarmet, “Breast disease and thermal imaging”, no. 59. In: ENEBI 2013: Vitoria, 2013.
- 2012
 - (3) S.V.C. Soares da Silva, P.S. Aguiar Jr., P.H.B. Diniz, A. C. Silva, A. C. de Paiva, **T.B. Borchardt**, A. Conci, R.C.F. de Lima, “Detecção de Regiões Suspeitas da Mama em Imagens Termográficas baseada em Assimetria e Índice de Diversidade Beta”, XXIII Congresso Brasileiro de Engenharia Biomedica, XXIII CBEB, 4pp.
 - (4) P.S. Aguiar Jr., S.V.C.S. da Silva, P.H.B. Diniz, A. C. Silva, A. C. de Paiva, **T.B. Borchardt**, A. Conci, R.C.F. de Lima, “Detecção de Regiões Suspeitas da Mama em Imagens Térmicas Usando Assimetria e Spatiogram”, XXIII Congresso Brasileiro de Engenharia Biomédica, XXIII CBEB, pp.1122-1125. Trab 560-1
 - (5) R. Resmini, **T.B. Borchardt**, A. Conci, R.C.F. Lima, “Auxílio ao Diagnóstico Precoce de Patologias da Mama Usando Imagens Térmicas e Técnicas de Mineração de Dados.”, In: Computer on the Beach, 2012, Florianópolis-SC. Computer on the Beach 2012: Anais do Evento., 2012. v. 1. p. 305-314.
 - (6) **T.B. Borchardt**, R. Resmini, R.S. Marques, A. Conci, A.C. Silva, R.C.F. Lima, “Registro de imagem para correção da posição corporal em mamotermogramas.” In: VII Congresso Nacional de Engenharia Mecânica - CONEM2012, 2012, São Luís. Anais CONEM2012, 2012.
 - (7) R. Resmini, **T.B. Borchardt**, A. Conci, e R.C.F. Lima, “Extraction of features based on geostatistical measures using thermal images of breast”, VII Congresso Nacional de Eng. Mecânica, Sao Luis , 31/07 a 3/08 de 2012
 - (8) G.O.S. Olivera, **T.B. Borchardt**, A. Conci, L.A.F. Fernandes, e R.C.F. Lima, “Técnicas de mineração de dados aplicadas a imagens térmicas mastológicas”, VII Congresso Nacional de Eng. Mecânica, S. Luis , 31/07 a 3/08 de 2012.
- 2011
 - (9) M.A. Silva, R.S. Marques, J.P. Scoralick, R. Resmini, **T.B. Borchardt**, A. Conci, “Segmentação dinâmica da pele humana utilizando o espaço de cor LUV.” In: 3º Encontro Nacional de Engenharia Biomecânica, 2011, Foz do Iguaçu. Anais ENEBI 2011, 2011.

7.5.6 ARTIGOS EM ANAIS DE CONGRESSOS INTERNACIONAIS

- 2013
 - (1) M. Sheeny, **T.B. Borchardt**, J. McKay, A. Conci, “Using wavelets for denoising infrared medical images”, XVI International Workshop on Wavelets, Differential Equations, Mechanics and Financial Mathematics.Facultad de

Matemática y Computación, Universidad de La Habana, Habana 10400, Cuba. 2013.

- 2011

- (2) **T.B. Borchardt**, R.H.C. Mello, J.G.F.M. Gazolla, R. Resmini, A. Conci, A. Sanchez, E.A. Vieira, “On the reduction of cloud influence in natural aerial images” In: 18th International Conference on Systems, Signals and Image Processing, 2011, Sarajevo. Proceedings of IWSSIP 2011. Sarajevo: Faculty of Electrical Eng., 2011. p. 279-282.
- (3) A. Conci, **T.B. Borchardt**, R. Resmini, L.S. Motta, M.J.A. Viana, L.C. Santos, R.C.F. Lima, E.M. Diniz, A.M. Santos, A.C. Paiva, A.C. Silva, “Biomedical use of infrared images.”, In: 18th International Conference on Systems, Signals and Image Processing, 2011, Sarajevo. Proceedings of IWSSIP2011, 2011. p. 1-4.
- (4) **T.B. Borchardt**, R. Resmini, A. Conci, A.M. Santos, A.C. Silva, E.M. Diniz, A.C. Paiva, R.C.F. Lima, “Thermal feature analysis to aid on breast disease diagnosis.”, In: International Congress of Mechanical Engineering, 2011, Natal-RN. Proceedings of 21st International Congress of Mechanical Engineering, 2011

- 2010

- (5) A. Conci, R.C.F. Lima, C.A.P. Fontes, S. Vasconcelos, **T.B. Borchardt**, R. Resmini, “On the breast reconstruction by thermal images.” In: International Consensus and Guidelines on Medical Thermology, 2010, Fortaleza. Proceedings of International Consensus and Guidelines on Medical Thermology, 2010. p. 135-135.
- (6) A. Conci, R.C.F. Lima, C.A.P. Fontes, R. Resmini, **T.B. Borchardt**, “A new method to aid to the breast diagnosis using fractal geometry.”, In: International Consensus and Guidelines on Medical Thermology, 2010, Fortaleza. Proceedings of International Consensus and Guidelines on Medical Thermology, 2010. p. 136-136.

REFERÊNCIAS

- [1] INCA, *Estimativa 2012: Incidência de Câncer no Brasil*. Rio de Janeiro: Divisão de Tecnologia da Informação, 2011.
- [2] Organização Mundial da Saúde, “Global Health Observatory Database: Cause-specific mortality 2008,” 2011. [Online]. Available: <http://apps.who.int/ghodata/>. [Accessed: 11-May-2012].
- [3] M. H. Forouzanfar, K. J. Foreman, A. M. Delossantos, R. Lozano, A. D. Lopez, C. J. L. Murray, and M. Naghavi, “Breast and cervical cancer in 187 countries between 1980 and 2010: a systematic analysis,” *The Lancet*, vol. 378, no. 9801, pp. 1461–1484, 2011.
- [4] IBCC, “O câncer de mama no alvo da moda,” 2011. [Online]. Available: <http://www.ocancerdemamanoalvodamoda.com.br>. [Accessed: 11-May-2012].
- [5] A. Tejerina, *Aula de habilidades y simulación em patología de la mama*. Madrid, Spain: ADEMÁS Comunicacion Gráfica, 2009.
- [6] Â. F. Barros, G. Uemura, and J. L. S. de Macedo, “Atraso no diagnóstico e tratamento do câncer de mama e estratégias para a sua redução,” *Femina*, vol. 40, no. 1, Feb. 2012.
- [7] G. C. Wishart, M. Campisi, M. Boswell, D. Chapman, V. Shackleton, S. Iddles, A. Hallett, and P. D. Britton, “The accuracy of digital infrared imaging for breast cancer detection in women undergoing breast biopsy,” *Eur J Surg Oncol*, vol. 36, no. 6, pp. 535–540, Jun. 2010.
- [8] H. Usuki, T. Ikeda, Y. Igarashi, I. Takahashi, A. Fukami, T. Yokoe, H. Sonoo, and K. Asaishi, “What Kinds Of Non-palpable Breast Cancer Can Be Detected By Thermography?,” *Biomed Thermology*, vol. 18, no. 4, pp. 8–12, 1998.
- [9] INCA, *Mamografia: da prática ao controle*. Rio de Janeiro: Ministério da Saúde, 2007.
- [10] S. V. Silva, “Reconstrução da geometria da mama a partir de imagens termográficas,” Tese de Doutorado, Universidade Federal Fluminense, Niterói, RJ, 2010.
- [11] J. R. Keyserlingk, P. D. Ahlgren, E. Yu, N. Belliveau, and M. Yassa, “Functional infrared imaging of the breast,” *IEEE Engineering in Medicine and Biology Magazine*, vol. 19, no. 3, pp. 30–41, Jun. 2000.
- [12] J. Ventura, “FDA: Breast thermography not a substitute for mammography,” *U.S. Food and Drug Administration*, 2011. [Online]. Available: <http://www.fda.gov/NewsEvents/Newsroom/PressAnnouncements/ucm257633.htm>. [Accessed: 20-Dec-2012].
- [13] T. B. Borchardt, A. Conci, R. C. F. Lima, R. Resmini, and A. Sanchez, “Breast thermography from an image processing viewpoint: A survey,” *Signal Processing*, no. 0, 2012.
- [14] Masonic Cancer Center, “Breast Cancer Treatment and Pregnancy,” *University of Minnesota*, 2012. [Online]. Available: <http://www.cancer.umn.edu/cancerinfo/NCI/CDR62970.html>. [Accessed: 07-Jan-2013].
- [15] T. Winslow, “Medical and scientific illustration,” 2013. [Online]. Available: <http://www.teresewinslow.com/>. [Accessed: 14-Jan-2013].

- [16] National Cancer Institute, “Adult Non-Hodgkin Lymphoma Treatment,” 2012. [Online]. Available: <http://www.cancer.gov/cancertopics/pdq/treatment/adult-non-hodgkins>. [Accessed: 07-Jan-2013].
- [17] Medicina Celular, “Manual de Bases Técnicas APAC/ONCO,” 2013. [Online]. Available: http://www.medicinacelular.com.br/aulasdownload/aula13/13c_neoplas.mansia.htm. [Accessed: 08-Jan-2013].
- [18] Danum International, “Optical Scanning - A New Light In Diagnostic Breast Imaging,” 2013. [Online]. Available: <http://www.danum.com/comfortscan.html>. [Accessed: 08-Jan-2013].
- [19] Grupo de Pesquisas em Mastologia, *Doenças da Mama - Guia Prático Baseado em Evidências*, 1st ed. Rio de Janeiro: Atheneu, 2011.
- [20] D. M. Pires, “Patologias Benignas da Mama.” [Online]. Available: <http://mastologiabh.site.med.br/index.asp?PageName=Patologias-20Benignas-20da-20Mama>. [Accessed: 10-Jan-2013].
- [21] A. M. Zelmanowicz, “Câncer de mama,” *ABC da Saúde*, 2001. [Online]. Available: <http://www.abcdasaude.com.br/artigo.php?611>. [Accessed: 10-Jan-2013].
- [22] V. Nagy, “Como Fazer Mamografia: Você Sabe Como É Feito Este Exame? - Saúde e Dicas | Hiper Feminina,” 2012. [Online]. Available: <http://www.hiperfeminina.com/saude/como-fazer-mamografia>. [Accessed: 27-Jun-2013].
- [23] S. Sedicias, “Mamografia com prótese de silicone,” *Tua Saúde*, 2012. [Online]. Available: <http://www.tuasaude.com/mamografia-com-protese-de-silicone/>. [Accessed: 10-Jul-2013].
- [24] L. I. L. da Silva, *Lei nº 11.664, de 29 de abril de 2008*. 2008.
- [25] J. Massada, “Recomendada ressonância magnética em casos de risco de cancro de mama,” *Ciência Hoje*, 30-Mar-2007. [Online]. Available: <http://www.cienciahoje.pt/index.php?oid=21217>. [Accessed: 06-Apr-2013].
- [26] B. R. Alvares and M. Michell, “Magnetic resonance imaging in the evaluation of breast cancer,” *Radiologia Brasileira*, vol. 36, no. 6, pp. 373–378, Dec. 2003.
- [27] E. M. R. Rocha, M. L. D. Dantas, N. R. Medeiros, and A. B. G. Gomes, “Sequência de difusão em ressonância magnéticanuclear mamária: aspectos técnicos e aplicações,” *Femina*, vol. 40, no. 5, Oct. 2012.
- [28] D. de D. Leopoldino, G. D’Ippolito, A. S. de A. Bezerra, and T. S. Gracio, “Technical aspects of contrast-enhanced magnetic resonance imaging of the breast: literature review,” *Radiologia Brasileira*, vol. 38, no. 4, pp. 287–294, Aug. 2005.
- [29] E. Y. K. Ng, U. R. Acharya, R. M. Rangayyan, and J. S. Suri, *Multimodality Breast Imaging: Diagnosis and Treatment*. SPIE Press, 2013.
- [30] C. Gann, “Device May Help Spot Breast Cancer,” *ABC News*, 11-Apr-2012. [Online]. Available: <http://abcnews.go.com/Health/fda-reviews-automated-breast-ultrasound-device/story?id=16119592>. [Accessed: 21-May-2013].
- [31] A. Smith, “Full-field breast tomosynthesis,” *Radiol Manage*, vol. 27, no. 5, pp. 25–31, Oct. 2005.

- [32] K. Schilling, D. Narayanan, J. E. Kalinyak, J. The, M. V. Velasquez, S. Kahn, M. Saady, R. Mahal, and L. Chrystal, "Positron emission mammography in breast cancer presurgical planning: comparisons with magnetic resonance imaging," *Eur J Nucl Med Mol Imaging*, vol. 38, no. 1, pp. 23–36, Jan. 2011.
- [33] D. A. B. Bonifacio, "Modelagem de sistema de detecção para mamografia por emissão de pósitrons utilizando detectores cintiladores monolíticos," 05-Oct-2011. [Online]. Available: <http://www.teses.usp.br/teses/disponiveis/43/43134/tde-26042012-155440/es.php>. [Accessed: 10-Jul-2013].
- [34] E. Y.-K. Ng, "A review of thermography as promising non-invasive detection modality for breast tumor," *International Journal of Thermal Sciences*, vol. 48, no. 5, pp. 849–859, May 2009.
- [35] B. Wiecek, S. Zwolenik, A. Jung, and J. Zuber, "Advanced thermal, visual and radiological image processing for clinical diagnostics," in *[Engineering in Medicine and Biology, 1999. 21st Annual Conference and the 1999 Annual Fall Meeting of the Biomedical Engineering Society] BMES/EMBS Conference, 1999. Proceedings of the First Joint*, 1999, vol. 2, p. 1108 vol.2–.
- [36] B. F. Jones, "A reappraisal of the use of infrared thermal image analysis in medicine," *IEEE Trans Med Imaging*, vol. 17, no. 6, pp. 1019–1027, Dec. 1998.
- [37] L. Boltzmann, "Ableitung des Stefan'schen Gesetzes, betreffend die Abhängigkeit der Wärmestrahlung von der Temperatur aus der electromagnetischen Lichttheorie," *Annalen der Physik*, vol. 258, no. 6, pp. 291–294, 1884.
- [38] P. J. S. C. Portugal, "Variação do espectro de radiação térmica com a temperatura." [Online]. Available: http://profs.ccems.pt/PauloPortugal/CFQ/Wien/Lei_Wien.html. [Accessed: 30-Apr-2013].
- [39] Berkeley Lab, "Electromagnetic Spectrum." [Online]. Available: <http://www.lbl.gov/MicroWorlds/ALSTool/EMSpec/EMSpec2.html>. [Accessed: 30-Apr-2013].
- [40] Flir Systems, "ThermoVision SDK - User's Manual." 05-Aug-2009.
- [41] M. Wernick, Y. Yang, J. Brankov, G. Yourganov, and S. Strother, "Machine Learning in Medical Imaging," *IEEE Signal Processing Magazine*, vol. 27, no. 4, pp. 25–38, Jul. 2010.
- [42] S. S. Haykin, *Neural networks: a comprehensive foundation*. Prentice Hall, 1999.
- [43] A. C. Lorena and A. C. P. L. F. Carvalho, "Uma introdução às Support Vector Machines," *RITA*, vol. XIV, pp. 43–67, 2007.
- [44] R. C. Serrano, "Análise da viabilidade do uso do coeficiente de Hurst e da lacunaridade no diagnóstico precoce de patologias da mama," Dissertação de Mestrado, Universidade Federal Fluminense, Niterói, RJ, 2010.
- [45] T. B. Borchardt, R. Resmini, L. S. Motta, E. W. G. Clua, A. Conci, M. J. A. Viana, L. C. Santos, R. C. F. Lima, and A. Sanchez, "Combining approaches for early diagnosis of breast diseases using thermal imaging," *International Journal of Innovative Computing and Applications*, vol. 4, no. 3, pp. 163–183, Jan. 2012.
- [46] M. Hall, E. Frank, G. Holmes, B. Pfahringer, P. Reutemann, and I. H. Witten, "The WEKA data mining software: an update," *SIGKDD Explor. Newsl.*, vol. 11, no. 1, pp. 10–18, Nov. 2009.

- [47] B. E. Boser, I. M. Guyon, and V. N. Vapnik, "A Training Algorithm for Optimal Margin Classifiers," in *Proceedings of the 5th Annual ACM Workshop on Computational Learning Theory*, 1992, pp. 144–152.
- [48] V. Vapnik and A. Y. Chervonenkis, "On the Uniform Convergence of Relative Frequencies of Events to Their Probabilities," *Theory of Probability and its Applications*, vol. 16, no. 2, pp. 264–280, 1971.
- [49] C. O. Camilo and J. C. da Silva, "Mineração de Dados: Conceitos, Tarefas, Métodos e Ferramentas," Relatório Técnico RT-INF_001-09, Aug. 2009.
- [50] S. de Amo, "Classificador SVM - Support Vector Machine." [Online]. Available: www.deamo.prof.ufu.br/arquivos/ClassificadorSVM.ppt. [Accessed: 03-Jul-2013].
- [51] T. Hastie, R. Tibshirani, and J. H. Friedman, *The elements of statistical learning data mining, inference, and prediction*. New York: Springer, 2009.
- [52] Diffen, "Accuracy vs Precision - Difference and Comparison." [Online]. Available: http://www.diffen.com/difference/Accuracy_vs_Precision. [Accessed: 10-Jul-2013].
- [53] E. Z. Martinez, F. Louzada-Neto, and B. B. Pereira, "A Curva ROC para testes diagnósticos," *Cadernos Saúde Coletiva*, vol. 1, no. 11, pp. 7–31, 2003.
- [54] W. J. Youden, "Index for rating diagnostic tests," *Cancer*, vol. 3, no. 1, pp. 32–35, 1950.
- [55] B. F. Jones and P. Plassmann, "Digital infrared thermal imaging of human skin," *IEEE Engineering in Medicine and Biology Magazine*, vol. 21, no. 6, pp. 41–48, 2002.
- [56] J. Koay, C. Herry, and M. Frize, "Analysis of breast thermography with an artificial neural network," *Conf Proc IEEE Eng Med Biol Soc*, vol. 2, pp. 1159–1162, 2004.
- [57] E. Y. K. Ng and E. C. Kee, "Advanced integrated technique in breast cancer thermography," *J Med Eng Technol*, vol. 32, no. 2, pp. 103–114, Apr. 2008.
- [58] N. Arora, D. Martins, D. Ruggerio, E. Tousimis, A. J. Swistel, M. P. Osborne, and R. M. Simmons, "Effectiveness of a noninvasive digital infrared thermal imaging system in the detection of breast cancer," *Am. J. Surg.*, vol. 196, no. 4, pp. 523–526, Oct. 2008.
- [59] V. Agostini, M. Knaflitz, and F. Molinari, "Motion artifact reduction in breast dynamic infrared imaging," *IEEE Trans Biomed Eng*, vol. 56, no. 3, pp. 903–906, Mar. 2009.
- [60] J. G. V. Luna and F. G. Delgado, "Feasibility of New-generation Infrared Imaging Screening for Breast Cancer in Rural Communities," *US Obstetrics & Gynecology*, vol. 1, pp. 52–55, 2010.
- [61] P. Kapoor and S. V. A. V. Prasad, "Image processing for early diagnosis of breast cancer using infrared images," in *2010 The 2nd International Conference on Computer and Automation Engineering (ICCAE)*, 2010, vol. 3, pp. 564–566.
- [62] A. Amri, A. Saidane, and S. Pulko, "Thermal analysis of a three-dimensional breast model with embedded tumour using the transmission line matrix (TLM) method," *Comput. Biol. Med.*, vol. 41, no. 2, pp. 76–86, Feb. 2011.
- [63] M. Kontos, R. Wilson, and I. Fentiman, "Digital infrared thermal imaging (DITI) of breast lesions: sensitivity and specificity of detection of primary breast cancers," *Clin Radiol*, vol. 66, no. 6, pp. 536–539, Jun. 2011.

- [64] U. R. Acharya, E. Y. K. Ng, J.-H. Tan, and S. V. Sree, "Thermography based breast cancer detection using texture features and Support Vector Machine," *J Med Syst*, vol. 36, no. 3, pp. 1503–1510, Jun. 2012.
- [65] American College of Clinical Thermology, "Breast Thermography - What is Breast Thermography?," 2012. [Online]. Available: http://www.thermologyonline.org/Breast/breast_thermography_what.htm. [Accessed: 22-May-2013].
- [66] L. A. Bezerra, M. M. Oliveira, M. C. Araújo, M. J. A. Viana, L. C. Santos, F. G. S. Santos, T. L. Rolim, P. R. M. Lyra, R. C. F. Lima, T. B. Borchardt, R. Resmini, and A. Conci, "Infrared Imaging for Breast Cancer Detection with Proper Selection of Properties: From Acquisition Protocol to Numerical Simulation," in *Multimodality Breast Imaging: Diagnosis and Treatment*, vol. PM227, Washington, USA: Ng, E.Y.K.; Acharya, U.R.; Rangayyan, R.M.; Suri, J.S., 2013, p. 572.
- [67] L. F. Silva, R. S. Marques, G. Carvalho, M. Santos, C. Fontes, A. Santos, and A. Conci, "Protocolo de captura de imagens térmicas da mama para construção de um banco público de exames.," in *Anais do 4º ENEBI*, Vitória-ES, 2013.
- [68] X. Tang, H. Ding, Y. Yuan, and Q. Wang, "Morphological measurement of localized temperature increase amplitudes in breast infrared thermograms and its clinical application," *Biomedical Signal Processing and Control*, vol. 3, no. 4, pp. 312–318, Oct. 2008.
- [69] S. Antonini, D. Kolaric, I. A. Nola, Z. Herceg, V. Ramljak, T. Kulis, J. K. Holjevac, and Z. Ferencic, "Thermography surveillance after breast conserving surgery #x2014; Three cases," in *ELMAR, 2011 Proceedings*, 2011, pp. 317–319.
- [70] L. F. Silva, "Captura de imagens térmicas no HUAP (Rotina diária e protocolos)," Universidade Federal Fluminense, Relatório Técnico 1, 2013.
- [71] UFPE, "Banco de imagens térmicas e prontuários eletrônicos," 2010. [Online]. Available: <http://150.161.110.168/termo>. [Accessed: 22-May-2013].
- [72] UFF, "Database for Mastological Image," 2012. [Online]. Available: <http://visual.ic.uff.br/dmi/>. [Accessed: 22-May-2013].
- [73] C. A. Lipari and J. F. Head, "Advanced infrared image processing for breast cancer risk assessment," in *Proceedings of the 19th Annual International Conference of the IEEE Engineering in Medicine and Biology Society, 1997*, 1997, vol. 2, pp. 673–676 vol.2.
- [74] C. L. Herry and M. Frize, "Digital processing techniques for the assessment of pain with infrared thermal imaging," in *Engineering in Medicine and Biology, 2002. 24th Annual Conference and the Annual Fall Meeting of the Biomedical Engineering Society EMBS/BMES Conference, 2002. Proceedings of the Second Joint*, 2002, vol. 2, pp. 1157–1158 vol.2.
- [75] Q. Zhou, Z. Li, and J. K. Aggarwal, "Boundary extraction in thermal images by edge map," in *Proceedings of the 2004 ACM symposium on Applied computing*, New York, NY, USA, 2004, pp. 254–258.
- [76] N. Scales, C. Herry, and M. Frize, "Automated image segmentation for breast analysis using infrared images," *Conf Proc IEEE Eng Med Biol Soc*, vol. 3, pp. 1737–1740, 2004.

- [77] H. Qi, P. T. Kuruganti, and W. E. Snyder, "Detecting Breast Cancer from Thermal Infrared Images by Asymmetry Analysis," in *Medical Infrared Imaging*, vol. 1, Diakides, N.A. e Bronzino, J.D., 2008, p. 11.–11.14.
- [78] H. Qi, W. E. Snyder, J. F. Head, and R. L. Elliott, "Detecting breast cancer from infrared images by asymmetry analysis," in *Proceedings of the 22nd Annual International Conference of the IEEE Engineering in Medicine and Biology Society, 2000*, 2000, vol. 2, pp. 1227–1228 vol.2.
- [79] Z. Jin-Yu, C. Yan, and H. Xian-xiang, "IR Thermal Image Segmentation Based on Enhanced Genetic Algorithms and Two-Dimensional Classes Square Error," in *Second International Conference on Information and Computing Science, 2009. ICIC '09*, 2009, vol. 2, pp. 309–312.
- [80] G. Schaefer, M. Závisek, and T. Nakashima, "Thermography based breast cancer analysis using statistical features and fuzzy classification," *Pattern Recognition*, vol. 42, no. 6, pp. 1133–1137, Jun. 2009.
- [81] L. S. Motta, A. Conci, R. C. F. Lima, and E. M. Diniz, "Automatic segmentation on thermograms in order to aid diagnosis and 2D modeling," in *Proceedings of 10th Workshop em Informática Médica*, Belo Horizonte, MG, 2010, vol. 1, pp. 1610–1619.
- [82] H. G. Zadeh, I. A. Kazerouni, and J. Haddadnia, "Distinguish Breast Cancer Based On Thermal Features in Infrared Images," *Canadian Journal on Image Processing and Computer Vision*, vol. 2, no. 6, pp. 54–58, 2011.
- [83] R. S. Marques, "Segmentação automática das mamas em imagens térmicas," Dissertação de Mestrado, Universidade Federal Fluminense, Niterói-RJ, 2012.
- [84] P. T. Kuruganti and H. Qi, "Asymmetry analysis in breast cancer detection using thermal infrared images," in *Engineering in Medicine and Biology, 2002. 24th Annual Conference and the Annual Fall Meeting of the Biomedical Engineering Society EMBS/BMES Conference, 2002. Proceedings of the Second Joint*, 2002, vol. 2, pp. 1155–1156 vol.2.
- [85] G. Schaefer, T. Nakashima, and M. Zavisek, "Analysis of Breast Thermograms Based on Statistical Image Features and Hybrid Fuzzy Classification," in *Advances in Visual Computing*, G. Bebis, R. Boyle, B. Parvin, D. Koracin, P. Remagnino, F. Porikli, J. Peters, J. Klosowski, L. Arns, Y. K. Chun, T.-M. Rhyne, and L. Monroe, Eds. Springer Berlin Heidelberg, 2008, pp. 753–762.
- [86] O. T. da Silveira F, R. C. Serrano, A. Conci, R. H. C. De Melo, and R. C. F. Lima, "On Using Lacunarity for Diagnosis of Breast Diseases Considering Thermal Images," in *16th International Conference on Systems, Signals and Image Processing, 2009. IWSSIP 2009*, 2009, pp. 1–4.
- [87] R. C. Serrano, J. Ulysses, S. Ribeiro, R. C. F. Lima, and A. Conci, "Using Hurst Coefficient and Lacunarity to Diagnosis Early Breast Diseases," in *Proceedings of 17th International Conference on Systems, Signals and Image Processing*, Rio de Janeiro, 2010, pp. 550–553.
- [88] O. Nurhayati, A. Susanto, T. S. Widodo, and M. Tjokronagoro, "Principal component analysis combined with first order statistical method for breast thermal images classification," *International Journal of Computer Science and Technology*, vol. 2, no. 2, pp. 12–18, 2011.

- [89] O. D. Nurhayati, T. S. Widodo, A. Susanto, and M. Tjokronagoro, "First Order Statistical Feature for Breast Cancer Detection Using Thermal Images," *World Academy of Science, Engineering & Technology*, vol. 70, pp. 1040–1042, Nov. 2010.
- [90] Universidade Federal de Pernambuco and Universidade Federal Fluminense, "Projeto PROENG," *Processamento e Análise de Imagens Aplicadas à Mastologia*, 2008. [Online]. Available: <http://visual.ic.uff.br/proeng/>. [Accessed: 28-Apr-2013].
- [91] Universidade Federal Fluminense and Universidade Federal do Maranhão, "Projeto PROCAD-NF," *Processamento e análise de imagens mastológicas*, 2009. [Online]. Available: <http://visual.ic.uff.br/procadnf/>. [Accessed: 28-Apr-2013].
- [92] L. A. Bezerra, M. M. Oliveira, T. L. Rolim, A. Conci, F. G. S. Santos, P. R. M. Lyra, and R. C. F. Lima, "Estimation of breast tumor thermal properties using infrared images," *Signal Processing*, vol. 93, no. 10, pp. 2851–2863, Oct. 2013.
- [93] Flir Systems, "ThermaCAM TM S45 - Infrared Camera," *Technical Specifications*, 2004. [Online]. Available: http://yellotec.com/pdf/S45_Datasheet.pdf. [Accessed: 28-Apr-2013].
- [94] Flir Systems, "Flir SC620 - The High Performance infrared inspection system," *Technical Specifications*, 2009. [Online]. Available: http://www.flir.com/uploadedFiles/Thermography_APAC/Products/Product_Literture/FLIR%20SC620%20Infrared%20Camera%20Datasheet_au.pdf. [Accessed: 28-Apr-2013].
- [95] P. G. Hewitt, *Fundamentos de Física Conceitual*, 1st ed. Bookman, 2009.
- [96] R. S. Marques, *Ground Truth Maker*. 2012.
- [97] B. Mandelbrot, "How Long Is the Coast of Britain? Statistical Self-Similarity and Fractional Dimension," *Science*, vol. 156, no. 3775, pp. 636–638, May 1967.
- [98] T. Higuchi, "Approach to an irregular time series on the basis of the fractal theory," *Physica D: Nonlinear Phenomena*, vol. 31, no. 2, pp. 277–283, Jun. 1988.
- [99] C. Gómez, A. Mediavilla, R. Hornero, D. Abásolo, and A. Fernández, "Use of the Higuchi's fractal dimension for the analysis of MEG recordings from Alzheimer's disease patients," *Med Eng Phys*, vol. 31, no. 3, pp. 306–313, Apr. 2009.
- [100] R. Esteller, G. Vachtsevanos, J. Echauz, and B. Litt, "A comparison of waveform fractal dimension algorithms," *IEEE Transactions on Circuits and Systems I: Fundamental Theory and Applications*, vol. 48, no. 2, pp. 177–183, 2001.
- [101] R. H. C. De Melo and A. Conci, "Succolarity: Defining a method to calculate this fractal measure," in *15th International Conference on Systems, Signals and Image Processing, 2008. IWSSIP 2008*, 2008, pp. 291–294.
- [102] S. Shekhar and H. Xiong, *Encyclopedia of GIS*. Springer, 2008.
- [103] R. C. Geary, "The Contiguity Ratio and Statistical Mapping," *The Incorporated Statistician*, vol. 5, no. 3, pp. 115–146, Nov. 1954.
- [104] B. D. Ripley, "Modelling Spatial Patterns," *Journal of the Royal Statistical Society. Series B (Methodological)*, vol. 39, no. 2, pp. 172–212, gener 1977.
- [105] C.-C. Chang and C.-J. Lin, "LIBSVM: A library for support vector machines," *ACM Trans. Intell. Syst. Technol.*, vol. 2, no. 3, pp. 27:1–27:27, May 2011.
- [106] C. Hsu, C. Chang, and C. Lin, *A practical guide to support vector classification*. 2010.

- [107] Emgu, *EmguCV: OpenCV in .NET*. EmguCV, 2013.
- [108] R. Resmini, “Análise de imagens térmicas da mama usando descritores de textura,” Dissertação de Mestrado, Universidade Federal Fluminense, Niterói, RJ, 2011.
- [109] J. H. Holland, *Adaptation in natural and artificial systems*. Cambridge, MA, USA: MIT Press, 1992.
- [110] J. Zhou, T. Bai, and C. Suo, “The SVM optimized by culture genetic algorithm and its application in forecasting share price,” in *IEEE International Conference on Granular Computing, 2008. GrC 2008*, 2008, pp. 838–843.
- [111] J. J. Grefenstette, “Optimization of Control Parameters for Genetic Algorithms,” *IEEE Transactions on Systems, Man and Cybernetics*, vol. 16, no. 1, pp. 122–128, 1986.
- [112] V. Umadevi, S. V. Raghavan, and S. Jaipurkar, “Interpreter for breast thermogram characterization,” in *Conference on Biomedical Engineering and Sciences (IECBES)*, Kuala Lumpur, 2010, pp. 150–154.

ANEXO A

Lista de termogramas utilizados na tese:

Termograma	Mama Esquerda	Mama Direita	Situação
IR_0089.jpg	Saudável	Patológica	Patológica
IR_0159.jpg	Saudável	Patológica	Patológica
IR_0432.jpg	Patológica	Patológica	Patológica
IR_0716.jpg	Patológica	Patológica	Patológica
IR_0729.jpg	Saudável	Saudável	Saudável
IR_0737.jpg	Saudável	Saudável	Saudável
IR_0756.jpg	Patológica	Patológica	Patológica
IR_0768.jpg	Patológica	Patológica	Patológica
IR_0777.jpg	Saudável	Saudável	Saudável
IR_0787.jpg	Saudável	Patológica	Patológica
IR_0947.jpg	Patológica	Saudável	Patológica
IR_0965.jpg	Patológica	Patológica	Patológica
IR_0968.jpg	Patológica	Patológica	Patológica
IR_0973.jpg	Saudável	Patológica	Patológica
IR_1000.jpg	Patológica	Patológica	Patológica
IR_1026.jpg	Saudável	Patológica	Patológica
IR_1270.jpg	Patológica	Saudável	Patológica
IR_1336.jpg	Saudável	Patológica	Patológica
IR_1353.jpg	Saudável	Patológica	Patológica
IR_1366.jpg	Saudável	Saudável	Saudável
IR_1620.jpg	Saudável	Saudável	Saudável
IR_1632.jpg	Saudável	Patológica	Patológica
IR_2841.jpg	Patológica	Saudável	Patológica
IR_2942.jpg	Patológica	Saudável	Patológica
IR_3433.jpg	Saudável	Saudável	Saudável
IR_3437.jpg	Patológica	Patológica	Patológica
IR_3441.jpg	Patológica	Saudável	Patológica
IR_3604.jpg	Saudável	Patológica	Patológica
IR_3630.jpg	Saudável	Saudável	Saudável
IR_3689.jpg	Saudável	Saudável	Saudável
IR_3702.jpg	Saudável	Saudável	Saudável
IR_3727.jpg	Patológica	Saudável	Patológica
IR_3781.jpg	Saudável	Saudável	Saudável
IR_3809.jpg	Saudável	Patológica	Patológica
IR_3830.jpg	Saudável	Saudável	Saudável
IR_3838.jpg	Patológica	Saudável	Patológica
IR_3849.jpg	Saudável	Saudável	Saudável
IR_3927.jpg	Patológica	Patológica	Patológica

IR_3996.jpg	Patológica	Saudável	Patológica
IR_4075.jpg	Saudável	Saudável	Saudável
IR_4149.jpg	Patológica	Saudável	Patológica
IR_4875.jpg	Patológica	Saudável	Patológica
IR_4910.jpg	Saudável	Patológica	Patológica
IR_4978.jpg	Patológica	Saudável	Patológica
IR_5563.jpg	Patológica	Saudável	Patológica
IR_5929.jpg	Patológica	Saudável	Patológica
IR_7464.jpg	Patológica	Patológica	Patológica
IR_8143.jpg	Patológica	Patológica	Patológica
IR_8227.jpg	Patológica	Saudável	Patológica
IR_8269.jpg	Saudável	Saudável	Saudável
IR_8285.jpg	Saudável	Patológica	Patológica

Tabela 25: Lista de termogramas utilizados do HC – UFPE.

Termograma	Mama Esquerda	Mama Direita	Situação
TFRON_V21_14-11-2012_0.jpg	Patológica	Patológica	Patológica
TFRON_V24_15-11-2012_0.jpg	Saudável	Saudável	Saudável
TFRON_V26_31-10-2012_0.jpg	Patológica	Patológica	Patológica
TFRON_V29_31-10-2012_0.jpg	Patológica	Saudável	Patológica
TFRON_V34_7-11-2012_0.jpg	Saudável	Patológica	Patológica
TFRON_V38_7-11-2012_0.jpg	Patológica	Patológica	Patológica
TFRON_V41_12-11-2012_0.jpg	Patológica	Saudável	Patológica
TFRON_V43_14-11-2012_0.jpg	Saudável	Saudável	Saudável
TFRON_V44_14-11-2012_0.jpg	Saudável	Saudável	Saudável
TFRON_V45_14-11-2012_0.jpg	Saudável	Saudável	Saudável
TFRON_V53_14-11-2012_0.jpg	Saudável	Patológica	Patológica
TFRON_V54_14-11-2012_0.jpg	Saudável	Saudável	Saudável
TFRON_V55_14-11-2012_0.jpg	Saudável	Patológica	Patológica
TFRON_V56_14-11-2012_0.jpg	Saudável	Patológica	Patológica
TFRON_V68_21-11-2012_0.jpg	Patológica	Patológica	Patológica
TFRON_V69_21-11-2012_0.jpg	Patológica	Saudável	Patológica
TFRON_V70_21-11-2012_0.jpg	Patológica	Patológica	Patológica
TFRON_V75_10-12-2012_0.jpg	Patológica	Patológica	Patológica

Tabela 26: Lista de termogramas utilizados do HUAP – UFF.

ANEXO B

Valores mínimos, máximos e médios dos atributos utilizados no *Teste 2*:

Atributo	Média	Máximo	Mínimo		Atributo	Média	Máximo	Mínimo
M(3,0)	0,108	0,378	-0,047		Média	33,73	35,84	30,80
M(3,90)	0,127	0,540	-0,204		Desvio Padrão	0,644	1,360	0,320
M(6,0)	0,040	0,327	-0,122		Delta de temperatura	2,572	6,060	1,280
M(6,90)	0,065	0,452	-0,221		Quant.	0,100	0,165	0,035
M(9,0)	0,022	0,319	-0,173		Limiar	35,20	37,52	33,24
M(9,90)	0,050	0,379	-0,221		Isoterma	0,011	0,065	0,000
M(12,0)	0,017	0,319	-0,191		H(5)	1,545	1,719	1,273
M(12,90)	0,045	0,353	-0,221		H(25)	1,512	1,630	1,310
M(15,0)	0,015	0,299	-0,190		H(50)	1,579	1,701	1,393
M(15,90)	0,042	0,352	-0,221		H(75)	1,634	1,735	1,447
G(3,0)	0,949	1,204	0,499		H(100)	1,693	1,786	1,547
G(3,90)	0,946	1,330	0,558		S(R2L)	0,373	0,682	0,041
G(6,0)	0,982	1,295	0,501		S(T2B)	0,328	0,751	0,000
G(6,90)	0,971	1,330	0,627		S(L2R)	0,285	0,741	0,005
G(9,0)	0,989	1,359	0,503		S(B2T)	0,502	0,682	0,159
G(9,90)	0,976	1,330	0,610		R(3)	0,306	0,671	0,097
G(12,0)	0,989	1,391	0,503		R(6)	0,530	0,952	0,149
G(12,90)	0,977	1,330	0,603		R(9)	0,641	0,998	0,195
G(15,0)	0,985	1,387	0,503		R(12)	0,695	1,000	0,259
G(15,90)	0,976	1,330	0,601		R(15)	0,723	1,000	0,287

Tabela 27: Valores gerais dos atributos utilizados no Teste 2 (H: Higuchi, S: Sucularidade, R: K-Ripley, M: Í-Moran, G: C-Geary)

ANEXO C

Imagens referenciadas na tese:

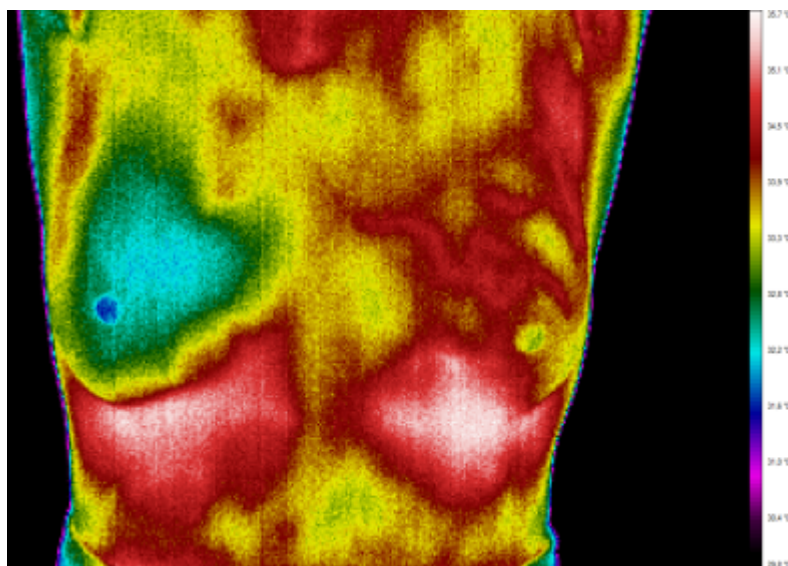


Figura 33: Termograma IR_3996 (HC).

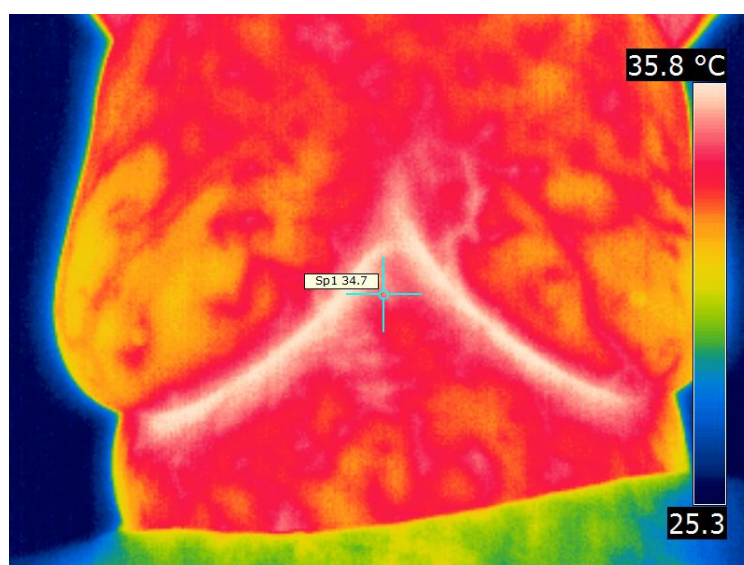


Figura 34: Termograma IR_0089 (HC).