



**INSTITUTO FEDERAL DE EDUCAÇÃO CIÊNCIA E TECNOLOGIA
CURSO SUPERIOR DE TECNOLOGIA EM ANÁLISE E DESENVOLVIMENTO DE
SISTEMAS**

BÁRBARA RAYRA DE SOUSA MATOS

**IDENTIFICAÇÃO DE ANOMALIAS NAS MAMAS ATRAVÉS DA EXTRAÇÃO DE
CARACTERÍSTICAS DE TERMOGRAMAS.**

**TERESINA
2014**

BÁRBARA RAYRA DE SOUSA MATOS

**IDENTIFICAÇÃO DE ANOMALIAS NAS MAMAS ATRAVÉS DA EXTRAÇÃO DE
CARACTERÍSTICAS DE TERMOGRAMAS**

Trabalho de Conclusão de Curso de graduação, apresentado à disciplina Trabalho de Conclusão de Curso 2 do Curso Superior de Tecnologia em Análise e Desenvolvimento de Sistemas, do Instituto Federal de Educação Ciência e Tecnologia - IFPI, campus Teresina, como requisito parcial para obtenção do título de Tecnólogo em Análise e Desenvolvimento de Sistemas.

Orientador: Prof. Dr. Thiago Alves Elias

TERESINA

2014

Dados Internacionais de Catalogação na Publicação (CIP) de acordo com ISBD

Matos, Bárbara Rayra de Sousa

M425i Identificação de anomalias nas mamas através da extração de
características de termogramas / Bárbara Rayra de Sousa Matos. - 2014.
48 f.: il. color.

Trabalho de Conclusão de Curso (Tecnologia em Análise e
Desenvolvimento de Sistemas) - Instituto Federal de Educação, Ciência e
Tecnologia do Piauí, Campus Teresina Central, 2014.

Orientador : Prof Me. Thiago Alves Elias.

1. Termografia. 2. Geometria Fractal. 3. SVM(Support Vector Machine). 4.
Histograma. I. Título.

CDD - 004.21

Elaborado por Tanize Maria Sales CRB 3/1024

BÁRBARA RAYRA DE SOUSA MATOS

**IDENTIFICAÇÃO DE ANOMALIAS NAS MAMAS ATRAVÉS DA EXTRAÇÃO DE
CARACTERÍSTICAS DE TERMOGRAMAS**

Trabalho de Conclusão de Curso de graduação,
apresentado à disciplina Trabalho de Conclusão
de Curso 2 do Curso Superior de Tecnologia em
Análise e Desenvolvimento de Sistemas, do
Instituto Federal de Educação Ciência e
Tecnologia - IFPI, Teresina, como requisito parcial
para obtenção do título de Tecnólogo em Análise e
Desenvolvimento de Sistemas.

Aprovada em: ____/____/____

BANCA EXAMINADORA

Prof. Dr. Thiago Alves Elias
Instituto Federal de Educação Ciência e Tecnologia

Prof. Dr. Ricardo Martins Ramos
Instituto Federal de Educação Ciência e Tecnologia

prof. Msc. José Ritomar Carneiro
Instituto Federal de Educação Ciência e Tecnologia

RESUMO

A Termografia tem sido fortemente utilizada em vários campos da medicina há alguns anos. Como ferramenta auxiliar no monitoramento e detecção de câncer de mama, porém, ainda enfrenta diversas limitações. Este trabalho utiliza Termografias de mamas com o objetivo de classificar as pacientes como saudáveis ou doentes. Foram utilizadas 18 pacientes como banco de dados. As imagens foram adquiridas através do banco de dados VISUAL-LAB, que utiliza imagens de pacientes do hospital universitário Antônio Pedro da Universidade Federal Fluminense (UFF). Uma câmera infravermelha da marca FLIR modelo SC-620, com resolução de 640X480, foi utilizada para capturar as imagens das pacientes. Cada imagem térmica possui uma matriz de temperatura correspondente que relaciona cada pixel da imagem a uma temperatura. As regiões de interesse (ROIs) são regiões de uma imagem que apresentam características importantes em uma análise. No caso das termografias as ROIs correspondem às mamas esquerda e direita das pacientes. A segmentação das Regiões de interesse (ROIs) foi realizada manualmente utilizando o GIMP. Partindo do princípio que o corpo é dividido simetricamente pelo plano sagital, espera-se que a distribuição das temperaturas das mamas esquerda e direita também seja simétrica. Nesse contexto, para análise das simetrias das mamas, foram extraídas características divididas em 3 grandes grupos. Os grupos utilizados foram: características baseadas em estatísticas, geometria fractal e histograma. A implementação dos métodos para extração das características foi realizada com a IDE Microsoft Visual C++ e a biblioteca OPENCV, utilizando como classificador o Support Vector Machine (SVM) que é um classificador linear fundamentado em teorias estatísticas. A validação cruzada foi utilizada para avaliar a capacidade de generalização do modelo. Os dados foram analisados usando métricas de avaliação como Sensibilidade, Especificidade, Índice de Youden, Acurácia e Área sob a curva ROC. Os resultados finais alcançaram 77% de acurácia, 69% de sensibilidade 100% de especificidade, 0,6923 de índice de Youden e 0,694 de área sob a curva ROC.

Palavras-chave: Termografia; Regiões de interesse; Estatísticas; geometria fractal; histograma; SVM.

ABSTRACT

For some years, Thermography has been used with more frequency in several medical areas, like auxiliary tool by monitoring and detecting breast cancer, however, it still faces several limitations. This work aims to classify patients as healthy or sick people through breast thermography in 18 patients. The images were acquired by VISUAL-LAB database, which uses the images of patients from University Hospital Antonio Pedro located at Federal Fluminense University (UFF). An infrared camera, FLIR SC-620 model, with resolution of 640X480, was used for capturing the image of the patients. Every thermal image has a correspondent temperature matrix that associates every image pixel in one temperature. The Regions of interests (ROIs) are regions of an image which demonstrate significant characteristics for an analysis. In Thermography, ROIs corresponds to the left and right patients' breasts. The segmentation of the Region of Interest (ROIs) was made through the manual use of GIMP. Considering that the body is divided symmetrically by sagittal plane, it is considered that left and right breast temperature distribution is symmetric too. In this context, analysis of breast symmetry, characteristics were divided in three huge groups. The groups used were: characteristics based in statistics, fractal geometric and histogram. The methods implementation to extraction of characteristics were made through IDE Microsoft Visual C++ and OpenCV library, by using Support Vector Machine (SVM) as classifier, which is a linear classifier based in statistical theories. The cross validation was used to evaluate the model generalization capacity. The data were analyzed through metrics of evaluation like Sensibility, Specificity, Youden index, Accuracy and Area under the curve ROC. The finals results achieved 77% of accuracy, 69% of sensibility, 100% of specificity, 0,6923 of Youden indexes and 0,694 of area under the curve ROC.

Keywords: Thermography; Region of interests; Statistics; Fractal; Geometry; histogram; SVM.

LISTA DE FIGURAS

Figura 1: Plano sagital do corpo humano.....	12
Figura 2:Termografia de uma paciente doente.....	13
Figura 3: Termografia corpo humano.....	16
Figura 4: Bolômetro.....	16
Figura 5: Mama adiposa.....	18
Figura 6: Distância d entre os hiperplanos H1 e H2.....	22
Figura 7: Térmica frontal.....	23
Figura 8: Térmica 45 graus mama direita.....	24
Figura 9: Térmica 45 graus mama esquerda.....	24
Figura 10: Térmica 90 graus mama direita.....	25
Figura 11: Térmica 90 graus mama esquerda.....	25
Figura 12: Termografia original.....	26
Figura 13: Imagem convertida em escala de cinza.....	26
Figura 14: Imagem antes da segmentação	27
Figura 15: Imagem após segmentação.....	28
Figura 16: Imagem binarizada.....	32

LISTA DE TABELAS

Tabela 1: Recomendações da mamografia e do ultrassom.....	12
Tabela 2: Classificação BI-RADS.....	20
Tabela 3: Grupo de características.....	30
Tabela 4: Cálculo Z1 e Z2.....	34
Tabela 5: Método leave one out.....	36
Tabela 6: Matriz de confusão para 2 classes. VP – verdadeiros positivos; VN – verdadeiros negativos; FP- falso positivo; FN- falso negativo.....	36
Tabela 7: Análise dos resultados finais.....	38
Tabela 8: Comparação dos resultados com trabalhos anteriores.....	39

LISTA DE ABREVIATURAS E SIGLAS

ACS	American Cancer Society
CEPEM	Centro de Estudos e Pesquisas da Mulher
FLIR	Foward Looking Infra-Red
INCA	Instituto Nacional de Câncer
RNA	Rede Neurais Artificiais
ROI	Region of Interest
UFF	Universidade Federal Fluminense
WHO	World Health Organization

LISTA DE FÓRMULAS

Formula 1: Equação hiperplano.....	25
Formula 2: Média.....	35
Formula 3: Variância.....	35
Formula 4: Desvio Padrão.....	35
Formula 5: Obliquidade.....	36
Formula 6: Curtose.....	36
Formula 7: Chi-square.....	37
Formula 8: Interseção.....	37
Formula 9: Correlação.....	37
Formula 10: Bhattacharyya.....	37
Formula 11: Lacunaridade caixa de escala r.....	39
Formula 12: Box-counting.....	39
Formula 13: Differential box-counting.....	40
Formula 14: Acurácia.....	42
Formula 15: Sensibilidade.....	42
Formula 16: Especificidade.....	42
Formula 17: Índice de Youden.....	43

SUMÁRIO

1	INTRODUÇÃO.....	10
1.1	MOTIVAÇÃO.....	11
1.2	OBJETIVOS.....	14
1.3	ESTRUTURA DO TRABALHO.....	15
2	REFERENCIAL TEÓRICO.....	16
2.1	TERMOGRAFIA MÉDICA.....	16
2.2	MÉTODOS UTILIZADOS NA DETECÇÃO DE ANOMALIAS.....	18
2.2.1	Mamografia.....	18
2.2.2	Ultrassonografia.....	21
2.2.3	Ressonância Magnética.....	21
2.3	TERMOGRAFIA MAMÁRIA.....	22
2.4	APRENDIZADO DE MÁQUINA.....	22
2.4.1	SVM.....	23
3	3. METODOLOGIA DO TRABALHO.....	26
3.1	AQUISIÇÃO DAS IMAGENS.....	26
3.2	SEGMENTAÇÃO.....	30
3.3	GRUPO DE CARACTERÍSTICAS.....	32
3.4	CARACTERÍSTICAS BASEADAS EM ESTATÍSTICAS.....	33
3.5	CARACTERÍSTICAS BASEADAS EM HISTOGRAMA.....	35
3.6	CARACTERÍSTICAS BASEADAS EM GEOMETRIA FRACTAL.....	36
4	VALIDAÇÃO.....	40
5	RESULTADOS.....	43
6	6. DISCUSSÕES.....	45
7	CONCLUSÃO.....	46
8	TRABALHOS FUTUROS.....	47
9	REFERÊNCIAS.....	48

1 INTRODUÇÃO

O câncer de mama tem sido uma das doenças mais severas a vitimar, principalmente mulheres, no mundo. O crescimento da incidência tem sido objeto de estudo de algumas organizações ligadas à saúde. Segundo(WHO - World Health Organization, 2021) o câncer de mama é a neoplasia mais comum entre mulheres no mundo. Atualmente em casos onde o diagnóstico é feito prematuramente e com a utilização de técnicas e tratamentos recentes, as taxas de curas são bastante elevadas. Entretanto ainda é comum, principalmente em países mais pobres, um diagnóstico tardio.

Ainda não se tem total domínio de todos os fatores que estão ligados ao câncer de mama, no entanto alguns aspectos são conhecidos e aceitos pela comunidade científica como causadores direta ou indiretamente de vários tipos de cânceres. Aspectos ligados à alimentação, sedentarismo, consumo de bebidas, problemas endócrinos, menarca precoce e o cigarro pode aumentar a possibilidade de desenvolver problemas futuros.(ACS – American Cancer Society, 2014). Como não é possível, ainda, prevenir o seu aparecimento, são aconselhadas medidas preventivas tais como: ter um estilo de vida saudável, realizar autoexames periódicos e o aconselhamento de que mulheres acima de 40 anos realizem a Mamografia anualmente. (INCA,2007).

Várias técnicas são utilizadas na detecção de anomalias nas mamas. Dentre elas a Mamografia e o exame clínico da mama, destacam-se como os métodos mais utilizados no diagnóstico precoce. Segundo o INCA (2014) a sensibilidade da Mamografia varia entre 46% a 88% e a sua especificidade varia entre 82% e 99%. Porém, a qualidade do exame está intimamente ligada à estrutura da mama. Sendo assim, a Mamografia quando realizada em mamas densas, característica comum em mulheres mais novas, ou em mamas que possuem próteses de silicone acaba por perder parte da sua eficácia(CEPEM, 2014).

O uso de termogramas mamários como ferramenta de auxílio à detecção de anomalias baseia-se no princípio de que o nosso corpo possui uma distribuição simétrica de temperatura. Tumores de uma forma em geral necessitam de nutrientes para se desenvolverem conseguem isso a partir da criação de novos vasos sanguíneos (neoangiogênese). Esses novos vasos sanguíneos por sua vez aumentam a vascularização o que consequentemente aumenta a temperatura (BORCHARTT, 2013). Esta energia é irradiada para a superfície da pele e pode ser medida utilizando câmeras térmicas.

A termografia de mama caracteriza-se como um exame fisiológico que avalia padrões térmicos no corpo e apresenta algumas vantagens em relação mamografia como o fato de ser um exame rápido, não invasivo e que não utiliza radiações ionizantes. Apresenta, também, um bom potencial por conseguir perceber tumores ainda no momento inicial quando as células cancerígenas estão no processo de neovascularização(VISUAL LAB, 2014).

1.1 MOTIVAÇÃO

A probabilidade de cura de câncer de mama está fortemente influenciada à sua descoberta num estágio inicial. As taxas de sobrevivência, em cinco anos, podem chegar a 100% dependendo do estágio em que a doença for diagnosticada (ACS, 2014). Nesse contexto a eficácia dos métodos utilizados leva em consideração este fator.

O autoexame das mamas é um método simples, de baixo custo, completamente indolor e pode ser realizado pela própria pessoa, no entanto, este método possui uma acurácia muito baixa, uma vez que, tumores perceptíveis ao toque geralmente estão num estágio mais avançado. Este exame não deve ser, portanto, a única medida de rastreamento do câncer de mama. Atualmente existem técnicas bem mais completas como a Mamografia, a Ultrassonografia, a Ressonância Magnética e Tomossíntese mamária. Estas técnicas quando realizadas por profissionais especializados conseguem rastrear anomalias que passariam imperceptíveis pelo autoexame.

A Mamografia atualmente é o exame padrão de rotina, aconselhada pelo Ministério da Saúde e disponibilizada gratuitamente pelo SUS para mulheres com mais de 40 anos (CNS, 2014), indicada tanto como medida de prevenção como medida de rastreamento. No entanto, ainda apresenta desvantagens quando realizada em mulheres mais novas que geralmente possuem mamas densas. Uma vez que esse tipo de mama é formado principalmente por tecido conjuntivo que diminui o contraste da mamografia podendo omitir uma lesão por vezes até palpável (SOCIEDADE BRASILEIRA DE MASTOLOGIA, 2014).

Como medida complementar a Mamografia, a Ultrassom de mamas apresenta-se como outra técnica também bastante utilizada. Em casos onde se encontram alterações, a Ultrassom consegue diferenciar um nódulo de um cisto, por exemplo (INCA, 2007). A seguir a tabela 1 com os dois exames mais populares e suas principais indicações:

Tabela 1: Recomendações da mamografia e do ultrassom

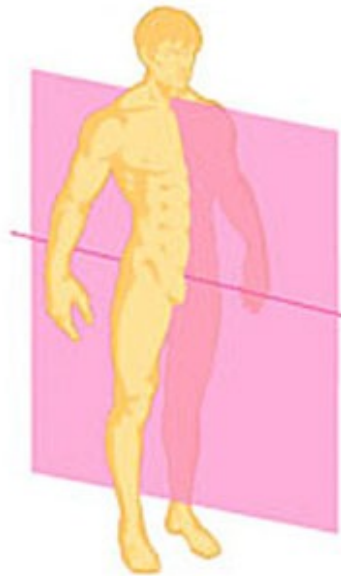
Indicações	Ultrassom	Mamografia
Indicada à pacientes jovens	X	
Exame padrão		X
Diferencia um nódulo de um cisto	X	
Risco com radiação		X
Detecção de neoplasias em fase Inicial		X

Fonte: INCA Instituto Nacional do Câncer. Mamografia da prática ao controle
Mamografia da prática ao controle, Rio de Janeiro, 2007, p. 20-25.

A utilização de termografias das mamas surgiu como mais um método complementar à Mamografia. A grande vantagem da utilização deste método reside no fato de que as chances de se encontrar anomalias nas mamas, antes mesmo que estas possam ser detectadas por uma Mamografia aumentam.

O plano sagital é um plano que divide o corpo humano em lado esquerdo e direito, na análise das duas partes podemos perceber que há uma tendência à simetria entre os dois lados. Seguindo esse princípio consideramos que a distribuição de temperatura entre o lado esquerdo e direito também tenha a tendência de ser a mais simétrica possível. A ilustração a seguir demonstra o plano sagital no corpo humano:

Figura 1: Plano sagital do corpo humano



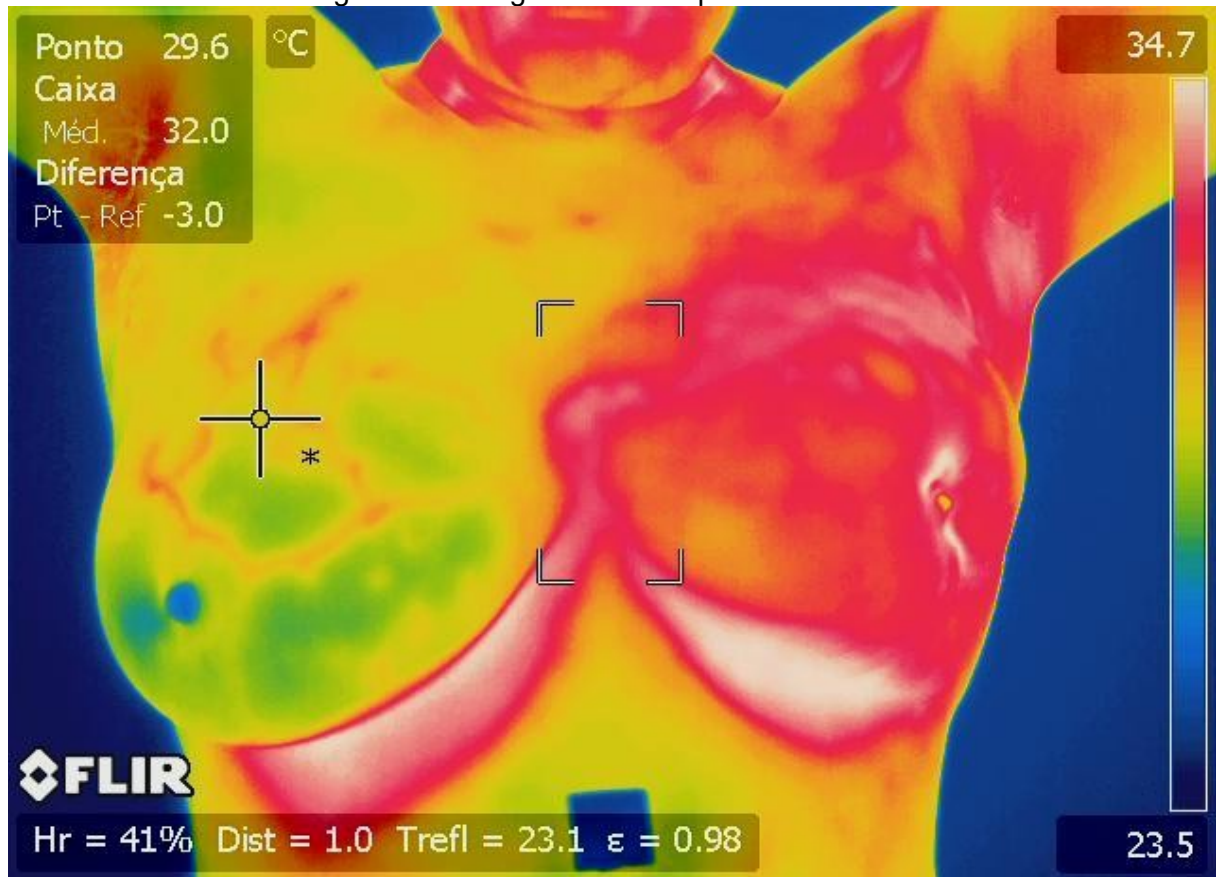
Fonte: Sagittal plane¹.

A neoangiogênese tumoral é um processo que se caracteriza pela formação de novos vasos sanguíneos que vão servir de substrato para o desenvolvimento e crescimento de tumores. Essa maior irrigação sanguínea ocasiona um aumento de temperatura na região. A análise das termografias de mamas também utiliza o princípio de simetria de temperatura das mamas, portanto, uma mama que apresente temperaturas mais elevadas do que a outra pode indicar que o processo de neoangiogênese está ocorrendo.

A variação das temperaturas nas termografias é representada por cores. Cores amarelas tendendo ao vermelho representam temperaturas mais altas do que as que tendem para o azul. A imagem a seguir demonstra a termografia das mamas de uma paciente doente. É possível perceber que a mama esquerda (doente) apresenta temperaturas mais altas que a mama direita:

¹ Disponível em:< <http://weddingdecorateideas.com/sagittal/sagittal-plane.html> > Acesso em 30 de dez. de 2014

Figura 2: Termografia de uma paciente doente.



Fonte: Visual Lab²

1.2 OBJETIVOS

Este Trabalho tem como objetivo geral apresentar um novo método que auxilie na detecção de câncer de mama.

A descrição detalhada dos objetivos deste trabalho é listada a seguir:

- Revisão bibliográfica da utilização da termografia na medicina, métodos de detecção de cânceres utilizados, utilização de termografia mamária no diagnóstico de câncer de mama, breve descrição sobre aprendizado de máquina e SVM;
- Segmentação das imagens;
- Extração das características;

² Disponível em: <visual.ic.uff.br/dmi> Acesso em 30 de dezembro de 2014

- d) Classificação dos resultados das características extraídas utilizando o classificador binário SVM;
- e) Validação do modelo criado pelo SVM;
- f) Apresentação dos resultados obtidos e comparação com resultados de estudos anteriores.

1.3 Estrutura do Trabalho

Este trabalho é composto por 7 capítulos. No início do trabalho foi realizada uma breve contextualização sobre o câncer de mama e sobre a Mamografia. No capítulo 2 é apresentado o referencial teórico utilizado como auxílio no desenvolvimento deste trabalho. O capítulo 3 concentra-se na descrição da metodologia do trabalho. Neste capítulo é demonstrada a metodologia de desenvolvimento de todo o trabalho. No capítulo 4 é apresentada a técnica de validação e as métricas de avaliação dos resultados utilizadas. No capítulo 5 os resultados da extração de todas as características são exibidos e comparados entre si e com outros trabalhos relacionados. No capítulo 6 os resultados obtidos são discutidos. No capítulo 7 é realizada a conclusão do trabalho.

2 REFERENCIAL TEÓRICO

A primeira seção aborda a história do uso da Termografia na área médica, A segunda seção aborda o uso das técnicas mais populares utilizadas no diagnóstico de câncer de mama. A terceira seção trata do uso das termografias na detecção de anomalias nas mamas. A quarta seção aborda o conceito de aprendizado de máquina e Máquina de vetores de Suporte (SVM).

2.1 TERMOGRAFIA MÉDICA

O uso do calor como medida avaliativa de análise do corpo não é uma prática recente. Hipócrates é descrito como primeiro médico a medir as variações de calor do corpo através de uma técnica em que consistia cobrir o corpo com lama e observar onde a mesma secava mais rápido (BERZ,SAUER, 2007).

A emissividade é uma propriedade básica de quaisquer objetos cuja temperatura seja maior que 0 e caracteriza-se pela emissão de energia eletromagnética dos mesmos. Os valores de emissividade variam entre 0 e 1. Um refletor perfeito teria uma emissividade igual a 0 e um corpo negro teria emissividade igual a 1. Esse fenômeno deve-se ao movimento mecânico das moléculas de um corpo cuja intensidade aumenta proporcionalmente com a temperatura do mesmo (BORCHATT, 2014).

O corpo humano emite radiação eletromagnética de comprimento de onda $9,3\mu\text{m}$ (FLIR, 2014). Este comprimento de onda pertence à faixa dedicada à radiação infravermelha. Câmeras com sensores infravermelhos conseguem captar esta energia emitida pelos corpos que por sua vez converte esta energia em um sinal elétrico. Este sinal é processado de forma a mostrar uma imagem no display ao mesmo tempo em que calcula a temperatura de cada pixel (BORCHATT,2014). A figura abaixo mostra a disposição de temperatura do corpo humano, capturada por uma câmera infravermelha, áreas avermelhadas representam regiões de maior temperatura.

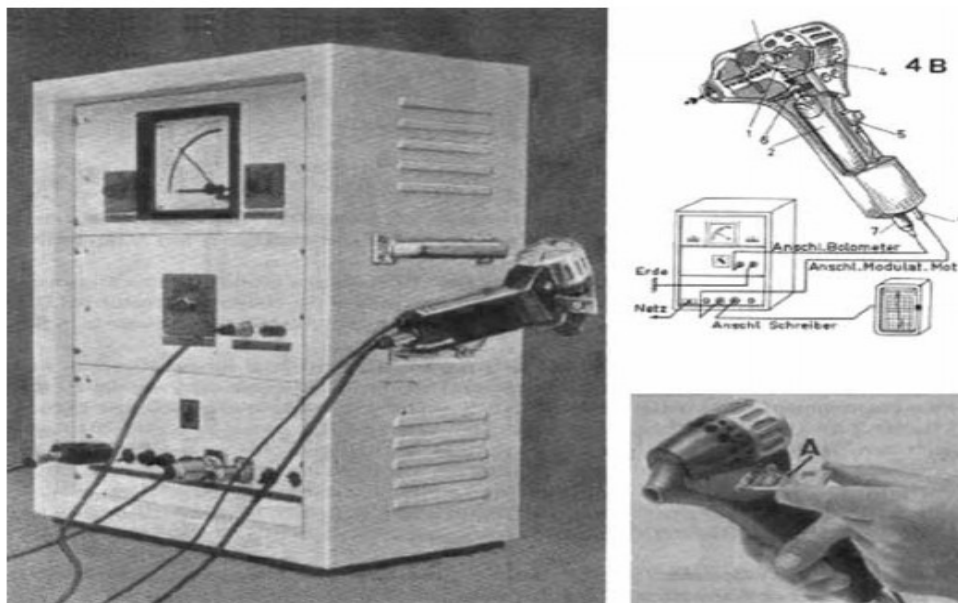
Figura 3: Termografia corpo humano.



Fonte: Thermography ³.

O uso de Termografia infravermelha na medicina começou na Alemanha em 1952, com um trabalho desenvolvido pelo médico Schwamm e pelo físico Reeh, os dois juntos desenvolveram um detector único infravermelho, bolômetro, para medição térmica sequencial de regiões definidas no corpo humano para fins de diagnóstico. (BERZ, R.; SAUER, H, 2007). A figura abaixo apresenta o detector infravermelho desenvolvido pro Schwamm e Reeh.

Figura 4: Bolômetro.



Fonte: Bolometer⁴

3 Disponível em: <<http://thegenesiscenter.com/thermography/>> Acesso em 06 de jan de 2014

Quanto a resolução, estabilidade, reprodutibilidade e medição exata das temperaturas as primeiras câmeras infravermelhas eram muito inferiores em relação com as atuais câmeras (BERZ, SAUER, 2007).

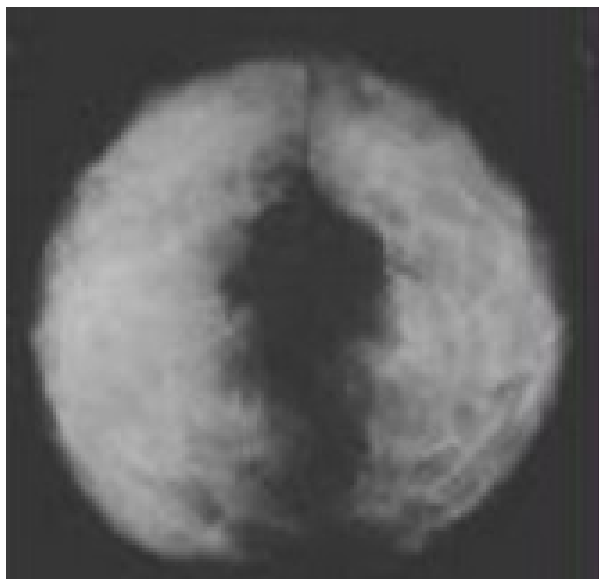
Atualmente a termografia é utilizada em diversos campos da medicina. Em pacientes com câncer, fibromialgia, disfunções vasculares, hérnias, lesões neuro-músculo-esqueléticas, por exemplo, é utilizada como ferramenta para identificar a origem da dor. É utilizada com sucesso, também, na investigação de doenças do trato urinário, doenças cardiovasculares, digestivas, desordens da glândula tireoide e análise de câncer de pele (MEDICAL THERMOGRAPHY, 2014).

2.2 MÉTODOS UTILIZADOS NA DETECÇÃO DE ANOMALIAS

2.2.1 Mamografia

A Mamografia é o exame radiológico das mamas. Durante o exame a paciente deve posicionar a mama no mamógrafo e esperar a emissão dos raios-X. Parte dos raios-X lançados são retidos pela mama o que possibilita a visualização de anomalias. No entanto, mulheres que possuem mamas muito densas absorvem quase que totalmente esta radiação o que prejudica a correta visualização e consequentemente um possível diagnóstico (INCA, 2007). A figura 5 e 6 demonstram exemplos de mamografias de uma mama densa e uma mama adiposa, respectivamente.

Figura 5: Mama densa



Fonte Tipos de mamas⁵

4 BERZ, R.; SAUER, H. "The Medical Use of Infrared-Thermography History and Recent Applications". Thermografie-Kolloquium, 2007

5 Disponível em: < <http://www.radioinmama.com.br/tiposdemama.html> > Acesso em 06 de janeiro de 2015

Figura 5: Mama adiposa.



Fonte: Tipos de mamas

A Mamografia atualmente divide-se em Mamografia convencional e Mamografia Digital. Embora as duas sejam realizadas da mesma forma existem diferenças entre as qualidades de ambas. A Mamografia digital apresenta algumas vantagens em relação à convencional, tais como: menor tempo de exposição à radiação e melhor acurácia na detecção de anomalias em mamas densas (INCA, 2007).

Mamografias de rastreamento é a classificação dada às Mamografias realizadas em pacientes que não apresentam sintomas aparentes, mas que apresentam algumas dessas características:

- a) Mulheres que possuem mais de 35 anos e pertencem ao grupo de risco;
- b) Mulheres assintomáticas com mais de 40 anos;
- c) Antes de iniciar terapia de reposição hormonal;
- d) No seguimento após Mastectomia ou cirurgia conservadora;
- e) A Mamografia diagnóstica é aquela realizada em mulheres que apresentam sinais ou sintomas de câncer. Essas anomalias nas mamas podem ser: nódulos, espessamento ou secreção das mamas (INCA,2007).

Os resultados radiológicos são classificados segundo o padrão BI-RADS (Breast Imaging Reporting and Data System). A seguir a tabela 2 demonstra quais são as possíveis classificações.

Tabela 2: Classificação BI-RADS

BI-RADS	SITUAÇÃO
Nível 1	Sem achados
Nível 2	Achados benignos
Nível 3	Achados provavelmente benignos
Nível 4	Achados suspeitos
Nível 5	Achados altamente suspeitos
Nível 0	Achados que necessitam de avaliação adicional

Fonte: INCA Instituto Nacional do Câncer⁶.

A Mamografia apesar de ser largamente utilizada como método padrão de detecção de anomalias ainda apresenta algumas desvantagens, tais como: exposição à radiação, mesmo que em pequena quantidade, pode ser um exame dolorido para algumas pacientes e sua acurácia está condicionada a aspectos anatômicos da paciente.

2.2.2 Ultrassonografia

A ultrassonografia na maioria das vezes é utilizada como exame complementa a mamografia com exceção em casos onde a paciente tem menos de 30 anos. Nesses casos a ultrassom é a primeira escolha. A ultrassonografia é indicada em casos onde é necessário diferenciar uma lesão sólida de uma lesão líquida, em casos de alterações palpáveis em mulheres com menos de 30 anos sendo a Mamografia a escolha posterior, em casos onde nódulos palpáveis não são detectados numa Mamografia, esses casos acontecem em mamas densas ou porque o nódulo está localizado em uma região onde não é possível ser visualizado na Mamografia. Em alguns casos a Ultrassonografia pode apresentar resultados normais, passando assim, uma falsa sensação de segurança para o médico e a paciente. Em casos de diferenciação entre alterações malignas e benignas, detecções, avaliação de pequenos nódulos detectados na Mamografia de mama adiposa, estudo e acompanhamento de microcalcificações a Ultrassonografia não é indicada (SBM - Sociedade Brasileira de Mastologia, 2009).

⁶ Mamografia da prática ao controle, Rio de Janeiro, 2007, p. 20-25

2.2.3 Ressonância Magnética

A Ressonância Magnética é frequentemente indicada a pacientes que já tenham sido operadas de câncer de mama, que possuam próteses de silicone ou pacientes que possuem mamas densas. O exame exige a administração de uma substância na veia (contraste) e que a paciente fique em repouso por aproximadamente 30 minutos, geralmente de barriga para baixo, dentro do equipamento. Durante o exame imagens da paciente e com a ajuda do contraste é possível identificar possíveis lesões. Posteriormente as imagens são analisadas. A Ressonância Magnética consegue identificar com facilidade lesões nas mamas, no entanto, esta técnica nem sempre é capaz de identificar o grau de malignidade de uma lesão (SOCIEDADE BRASILEIRA DE MASTOLOGIA,2014).

2.3 TERMOGRAFIA MAMÁRIA

A Termografia no campo da Mastologia começou a ser utilizada no início de 1959. Porém com as limitações tecnológicas da época esta técnica acabou sendo desacreditada (WALKER, 2014). No entanto, com o constante aperfeiçoamento de técnicas computacionais de análise de imagens a termografia ressurgiu com um grande potencial, quanto a exame de avaliação de mamas. Sendo usada com relativo sucesso em países de clima frio e temperado (SANTOS, LIMA, 2008)

Estudos atuais utilizando termogramas mamários demonstram que são bastante promissores, de fato, os resultados demonstram uma acurácia próxima a 90% (BORCHARTT,2013).

O uso da Termografia apresenta algumas vantagens, como: praticidade, o fato de ser completamente indolor e seguro. No entanto, a real utilização de Termogramas mamários ainda tem um longo caminho à frente. Diversos são os obstáculos, tais como: A não padronização de um protocolo de aquisição de imagens o que pode representar alterações nas comparações entre estudos e a falta de um banco de dados com imagens mastológicas que sirva com fonte para estudos futuros.

2.4 APRENDIZADO DE MÁQUINA

O aprendizado de máquina é um campo da inteligência artificial que tem como objetivo principal construir sistemas computacionais capazes de adquirir conhecimento de forma automática (LORENA, CARVALHO,2003). Várias técnicas de aprendizado de máquina vêm sendo intensamente estudadas e sua utilização depende da natureza dos casos estudados. De um modo em geral as técnicas de

aprendizagem de máquinas utilizam um modelo indutivo para construir modelos de predição. Do qual a partir de conhecidos exemplos anteriores é possível inferir regras e relações para posteriores casos desconhecidos. Costuma-se dividir o aprendizado de máquina em supervisionado e não supervisionado. Aprendizado de máquina supervisionado possui uma espécie de “guia” onde são apresentados conjuntos de exemplos que já são previamente conhecidos com o objetivo de construir um modelo de predição. No aprendizado de máquina não supervisionado não há nenhum suporte para criação de modelos de predição (LORENA, CARVALHO, 2003). O modelo de predição resultante ou preditor é chamado de classificador. Classificação é o processo de atribuir a um determinado exemplo um rótulo da classe a qual ele pertence (RUSSEL, NORVIG 1995). A escolha de um classificador correto é determinante no desempenho final da aprendizagem de máquina. Durante o processo de classificação de exemplos de treinamento podem ocorrer dois fenômenos que interferem na corretude do resultado de classificação de exemplos posteriores. Overfitting é um fenômeno que ocorre quando o conjunto de exemplos possui um padrão muito específico e o classificador, por sua vez, consegue classificar corretamente todos os exemplos, embora isso pareça ser bom, há uma grande possibilidade do classificador classificar incorretamente exemplos dos testes posteriores que não se pareçam com o conjunto de treinamento. Underfitting ocorre quando o classificador não consegue classificar corretamente o conjunto de treinamento porque este não possui um padrão bem delineado, nesses casos, até para exemplos simples ou com o mínimo de ruído possível o classificador não consegue classificar corretamente o conjunto de treinamento. É normal que durante o processo de escolha de um classificador seja escolhido o que obteve melhor desempenho durante o treinamento, no entanto, essa escolha pode não ser a melhor quando utilizado nos testes posteriores. É necessário, então, que a escolha seja bem equilibrada, que o classificador consiga prever corretamente o maior número de exemplos tanto para treinamento quanto para testes.

2.4.1 SVM

A Máquina de Vetor de Suporte (Support Vector Machine – SVM) utiliza conhecimento provido da Teoria de Aprendizagem Estatística, desenvolvida inicialmente pelo pesquisador russo Vladimir Vapnik (LORENA, CARVALHO, 2003). Estes estudos tinham como objetivos reduzir os casos de erros com exemplos do conjunto de treinamento e de teste, chamados de riscos estruturais. Segundo Vapnik, uma maior redução dos riscos estruturais delimitaria a capacidade de generalização de um modelo de aprendizagem (A. C. Lorena and A. C. P. L. F. Carvalho, 2003). A SVM foi utilizada pela primeira vez por Boser et al (LORENA, CARVALHO, 2003). E ficou conhecida por apresentar altos níveis de acurácia sendo comparada com outras técnicas de classificação como, as Redes Neurais Artificiais (RNAs) (Haykin, 1999) apresentando resultados superiores em alguns casos como, detecção de faces (Hearst et al., 1998) e reconhecimento de texto (Hearst et al., 1998). A classificação realizada pelas SVMs baseia-se na separação por hiperplanos, isto é, seja T uma base de treinamento com n registros. $X \in \mathcal{X}$ e suas respectivas classes $Y \in \mathcal{Y}$, em que X constitui o espaço dos registros e $\mathcal{Y} = \{-1, +1\}$. A

base T será linearmente separável se for possível traçar um hiperplano entre os registros pertencentes às classes -1 e $+1$. (A. C. Lorena and A. C. P. L. F. Carvalho, 2003).

Um conjunto de treinamento é linearmente separável se é possível separar os padrões das classes diferentes por, pelo menos, um hiperplano (NORVIG, 1995):

A equação de um hiperplano é representada pela seguinte fórmula:

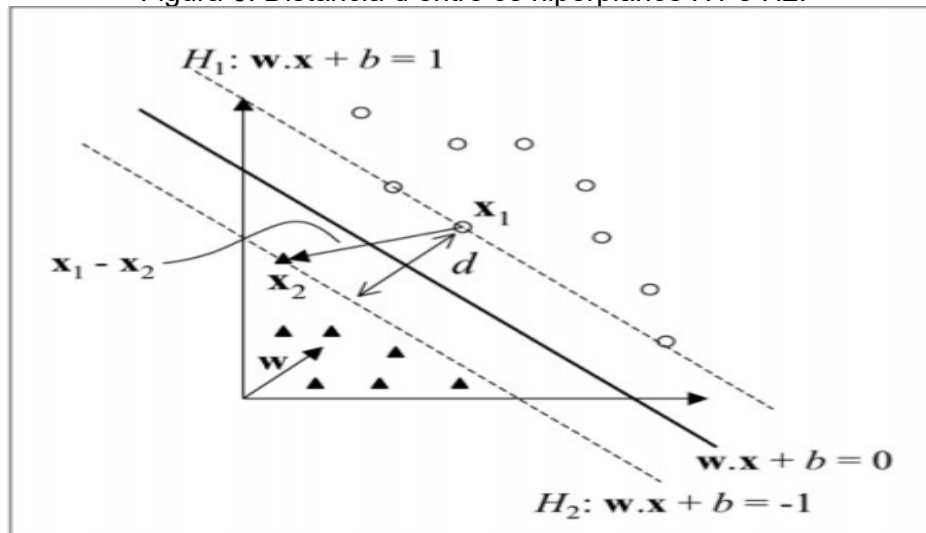
Formula 1: Equação hiperplano

$$\mathbf{w} \cdot \mathbf{x} + b = 0$$

Esta equação divide o espaço em duas entradas $\mathbf{w} \cdot \mathbf{x} + b > 0$ e $\mathbf{w} \cdot \mathbf{x} + b < 0$. Uma função sinal $g(\mathbf{x}) = \text{sgn}(f(\mathbf{x})) = \text{sgn}(\mathbf{w} \cdot \mathbf{x} + b)$, pode ser então aplicada sobre essa função, levando à classificação $+1$ se $f(\mathbf{x}) > 0$ ou -1 se $f(\mathbf{x}) < 0$.

O objetivo da SVM é maximizar a distância d entre os hiperplanos H_1 e H_2 mais detalhes desse processo pode ser visto em Lorena e Carvalho (2003). A imagem abaixo ilustra a separação por hiperplanos feita pela SVM.

Figura 6: Distância d entre os hiperplanos H_1 e H_2 .



Fonte: Lorena e Carvalho⁷

⁷ A. C. Lorena and A. C. P. L. F. Carvalho, Uma introdução às Support Vector Machines. São Carlos: Instituto de Ciências Matemáticas e de Computação. 2003. 66p.

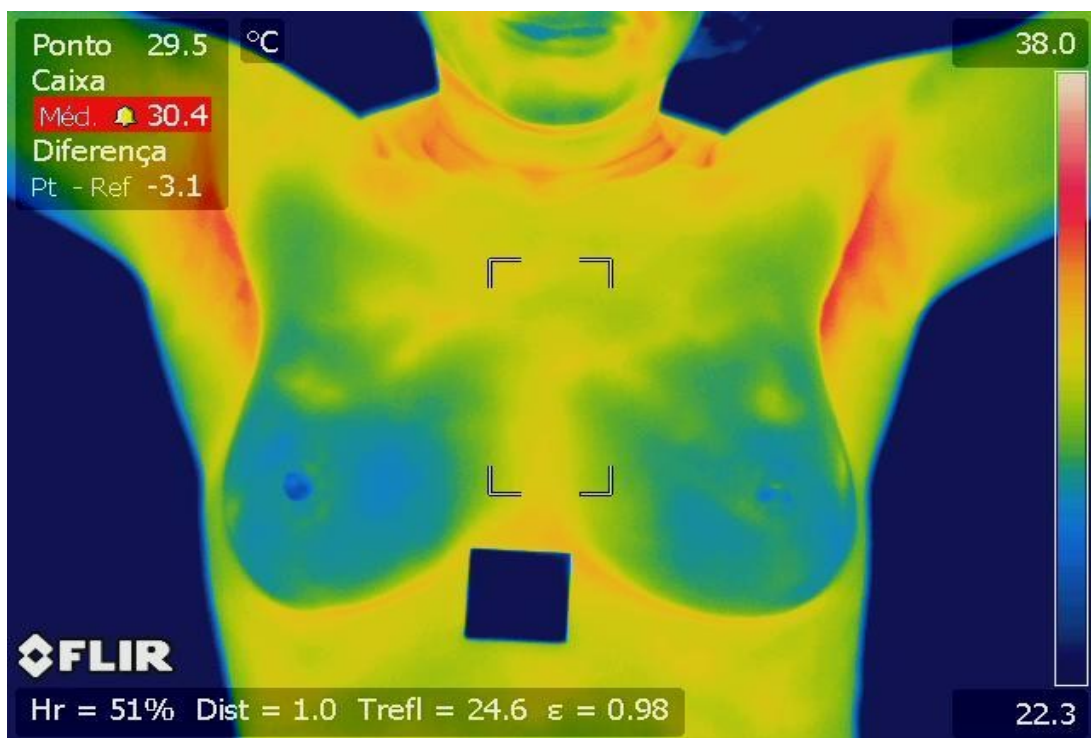
3 3. METODOLOGIA DO TRABALHO

Este capítulo aborda os métodos utilizados no processo de desenvolvimentos do trabalho desde a fase de aquisição das imagens, segmentação, extração das características, classificação e validação.

3.1 AQUISIÇÃO DAS IMAGENS

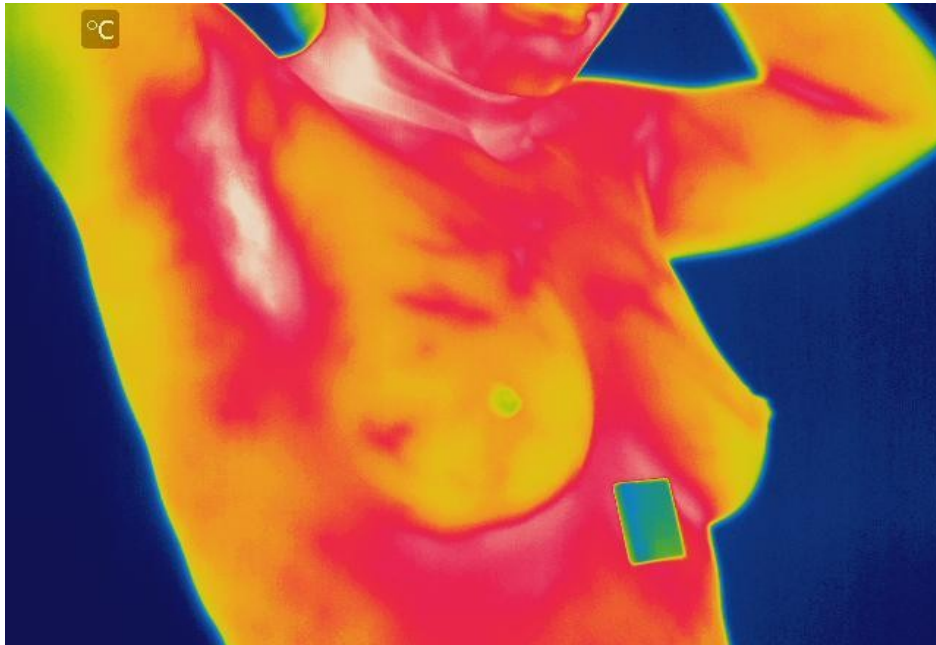
Os termogramas mamários foram disponibilizados pelo banco de dados VISUAL-LAB (2014). O banco de dados foi desenvolvido com a parceria entre pesquisadores da UFF e por médicos do hospital Universitário Antônio Pedro. As imagens foram capturadas usando uma câmera térmica da marca FLIR, modelo SC-620, com resolução de 640X480. Durante o processo de aquisição das imagens as pacientes foram orientadas a permanecer 10 minutos com os braços apoiados no suporte deixando as axilas livres. Após isso a paciente é alocada na posição indicada com os braços apoiados sobre a cabeça. A seguir iniciam-se as sequencias de fotos uma frontal duas com rotação lateral de 45° para esquerda e direita e mais outras duas com rotação lateral de 90° para esquerda e direita. As termografias abaixo ilustram todas as posições assumidas pela paciente durante o processo de aquisição

Figura 7: Térmica frontal.



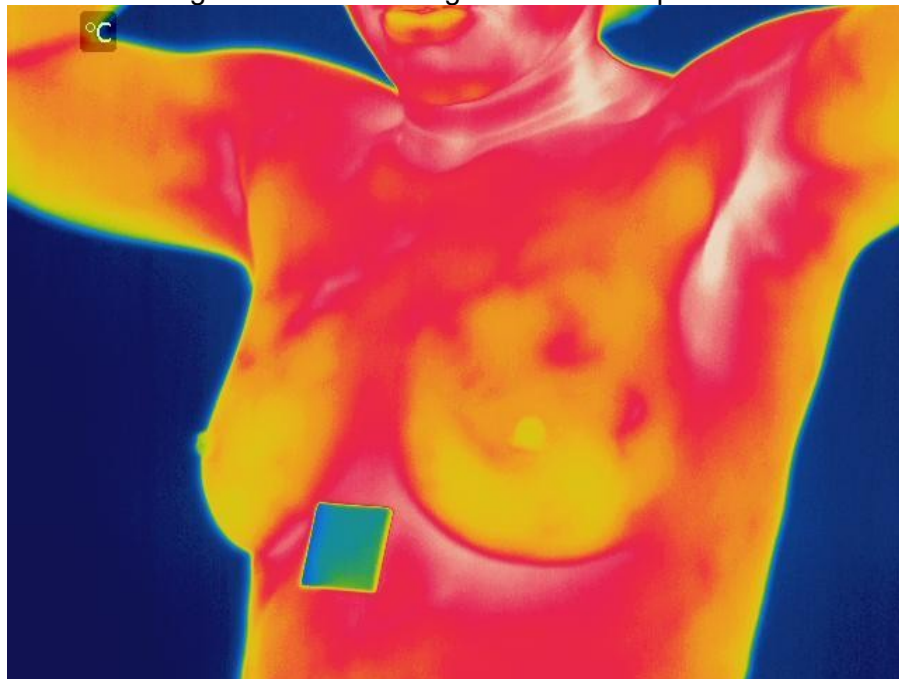
Fonte: Visual Lab.

Figura 8: Térmica 45 graus mama direita



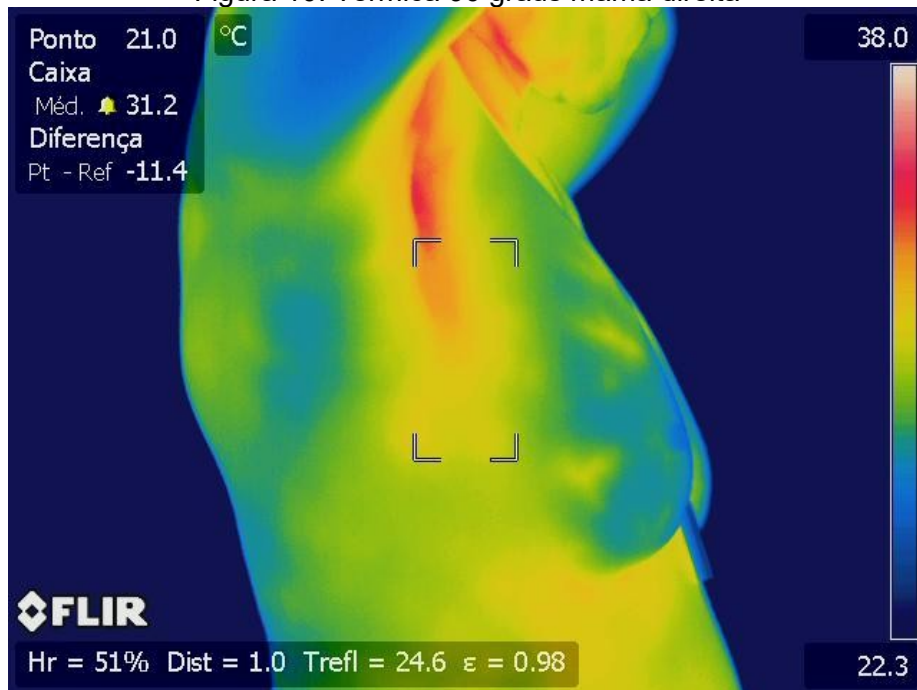
Fonte: Visual Lab.

Figura 9: Térmica 45 graus mama esquerda.



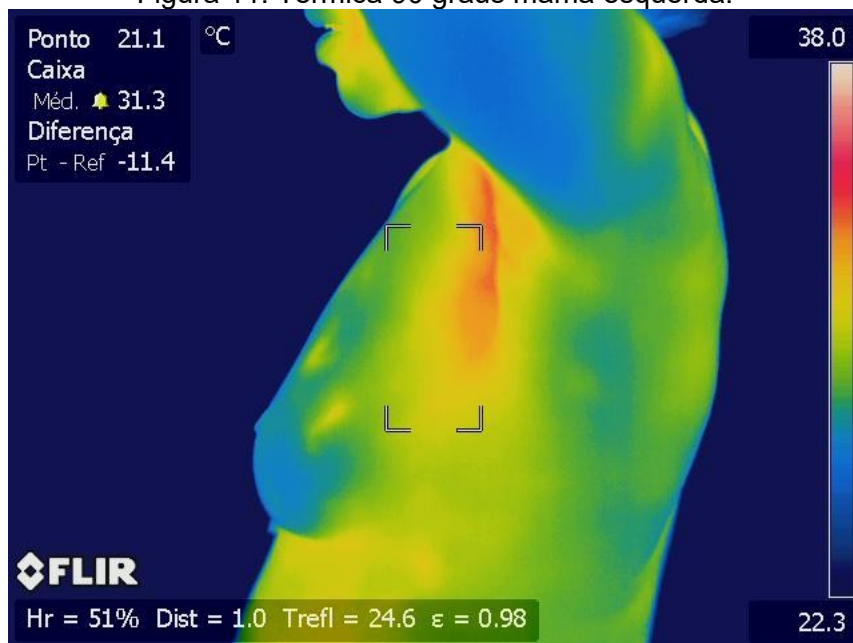
Fonte: Visual Lab.

Figura 10: Térmica 90 graus mama direita



Fonte: Visual Lab.

Figura 11: Térmica 90 graus mama esquerda.



Fonte: Visual Lab.

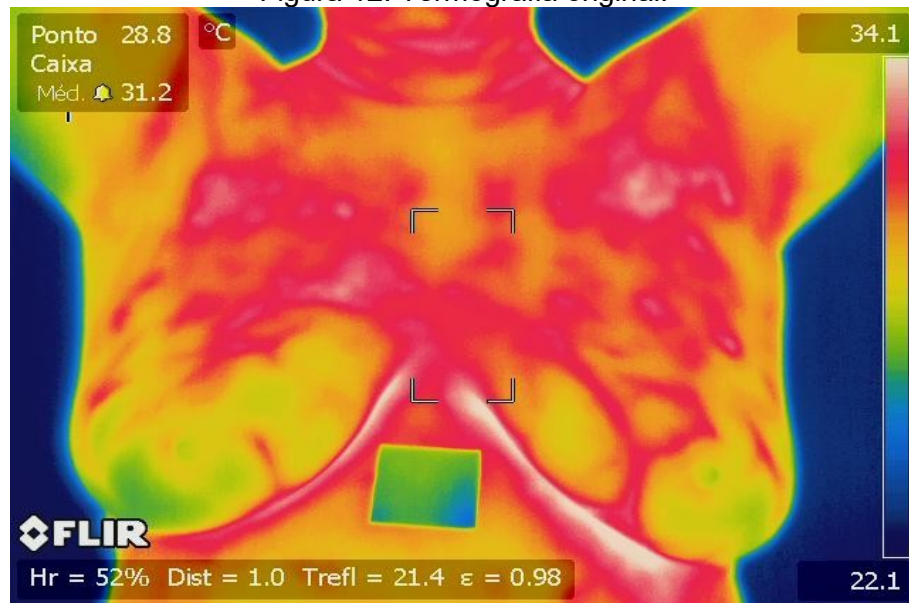
Cada imagem adquirida pela câmera FLIR possui uma matriz de temperatura correspondente. Cada valor da matriz de temperatura corresponde a um valor do pixel da imagem térmica. As imagens térmicas adquiridas, posteriormente passam pelo processo de transformação para níveis de cinza. A transformação da imagem original em níveis de cinza é dada pela seguinte equação (8):

$$\text{Value.val}[i] = ((255 * (\text{value.val}[i] - \text{min})) / (\text{max} - \text{min}))$$

Onde $\text{value.val}[i]$ é o valor do pixel atual e min e max correspondem ao menor e maior valor de temperatura da imagem.

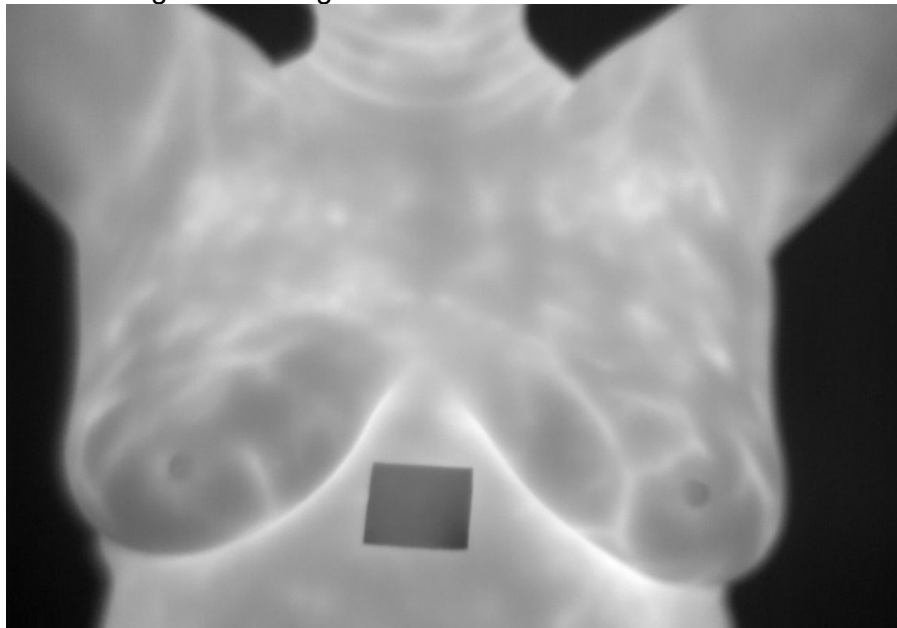
Cada valor de temperatura da imagem original é utilizado na transformação para obtenção do pixel correspondente da imagem em escala de cinza. Exemplo de antes e depois da transformação da imagem térmica em escala de cinza

Figura 12: Termografia original.



Fonte: Visual Lab.

Figura 13: Imagem convertida em escala de cinza.



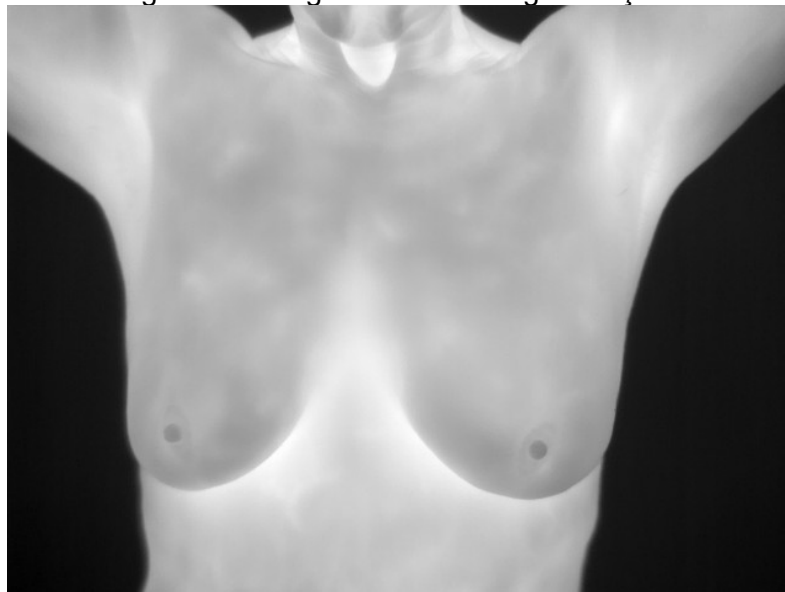
Fonte: Visual Lab.

3.2 SEGMENTAÇÃO

ROI (Region of interest) é uma área da imagem total que apresenta peculiaridades importantes a serem analisadas. A extração da ROI de uma imagem acontece por meio do processo de segmentação que nada mais é do que a delimitação da ROI em relação à imagem total. No contexto deste trabalho as ROIs são as mamas das pacientes. As imagens capturadas das pacientes contêm áreas que não são importantes no desenvolvimento deste projeto, por esse motivo as imagens foram submetidas ao processo de segmentação para extração das ROIs. A segmentação pode ser feita de forma automática ou manual. O software desenvolvido por R (Marques, 2014) foi utilizado no processo de segmentação automática das imagens, mas não apresentou bons resultados, por esse motivo as imagens foram segmentadas manualmente utilizando a ferramenta GIMP. O processo de segmentação das imagens é seguido do processo de extração da matriz de temperatura correspondente a apenas as mamas segmentadas. Este processo é realizado usando cada imagem segmentada como máscara da matriz de temperatura total.

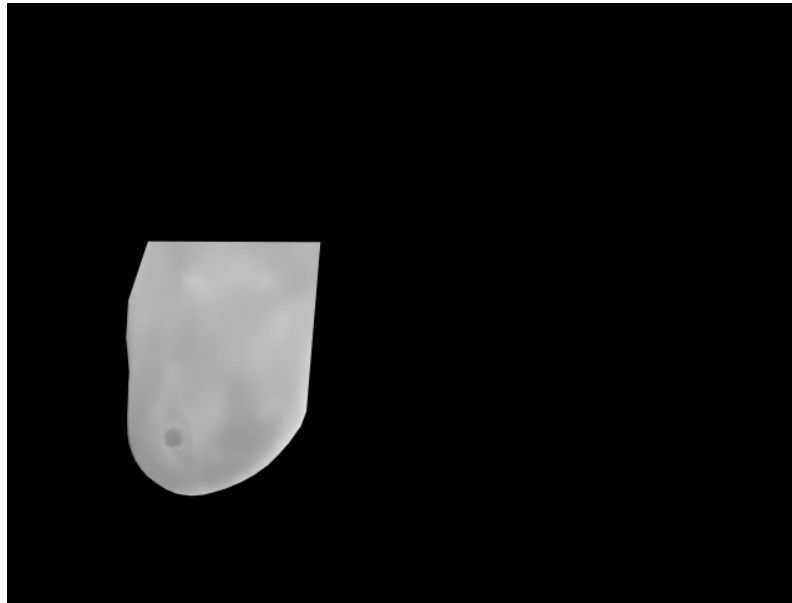
As imagens seguintes apresentam uma imagem total e uma mama direita após o processo de segmentação

Figura 14: Imagem antes da segmentação .



Fonte: Visual Lab.

Figura 15: Imagem após segmentação.



Fonte: Visual Lab.

3.3 GRUPO DE CARACTERÍSTICAS

As características extraídas neste trabalho foram divididas em 3 grandes grupos. As características baseadas em estatísticas utilizadas foram: Média, Variância, Desvio padrão, Curtose e Obliquidade. As características baseadas em comparação de histograma foram: método intersect, Chi-Square, Correlation e Bhattacharyya. As características baseadas em geometria fractal foram: Lacunaridade, Box-Counting e Differential box-counting. Todas as características utilizadas foram extraídas utilizando as imagens segmentadas em escala de cinza e as matrizes de temperaturas correspondente. A implementação dos métodos foi realizada utilizando a IDE Microsoft Visual C++ e a biblioteca OpenCV.

Tabela 3: Grupo de características

Grupos	Características
Características baseadas em Estatísticas	● Média ● Variância ● Desvio Padrão ● Curtose ● Obliquidade
Características baseadas em Histograma	● Intersect ● Chi-Square ● Correlation ● Bhattacharyya
Características baseadas em Geometria fractal	● Lacunaridade ● Box Counting

Fonte: Elaborada pelo autor

Neste trabalho as pacientes são analisadas como doentes ou saudáveis e as características escolhidas são extraídas das mamas esquerda e direita. Após a extração, as características passam pelo processo de redução em que para cada característica extraída da mama esquerda e direita é realizada a diferença absoluta desses valores para encontrar um único resultado. Essa metodologia de comparação das mamas é utilizada para avaliar o quão próximo são os resultados da extração das características para as duas mamas. Quanto mais próximo de 0 mais similar é a distribuição de temperatura de uma mama em relação a outra e mais provável que essas mamas sejam similares entre si o que diminui as suspeitas de um possível tumor.

3.4 Características baseadas em Estatísticas

Características baseadas em estatísticas são comumente utilizadas em trabalhos relacionados pelo fácil entendimento, implementação e bons resultados. Todos os métodos para extração de características baseadas em estatísticas utilizam os valores da matriz de temperatura correspondente à mama utilizada.

A média é um dos métodos mais simples utilizados em estatística, representa a tendência central dos valores de uma distribuição. O cálculo da média, $(\bar{x})$, é feito somando todos os valores da matriz de temperatura da mama, (X_i) , dividindo pela quantidade total de temperaturas da matriz (N_i)

Formula 2: Média

$$\bar{x} = \frac{\sum (x_i)}{N_i}$$

Variância e desvio padrão são medidas utilizadas para analisar o grau de dispersão dos valores de um conjunto. A variância retorna a distancia de cada valor do conjunto em relação ao ponto médio do conjunto. A variância, S^2 , é calculada somando o quadrado das diferenças de cada valor da matriz de temperatura, X , com a média, $\bar{X}$, dividida pela quantidade total de temperatura menos um. O desvio padrão, S , é a raiz quadrada da variância.

Formula 3: Variância

$$s^2 = \frac{\sum_i^N (X - \bar{X})^2}{N - 1}$$

Formula 4: Desvio Padrão

$$s = \sqrt{\frac{\sum_i^N (X - \bar{X})^2}{N - 1}}$$

Numa distribuição estatística, a assimetria é o quanto sua curva de frequência se desvia ou afasta da posição simétrica. A curtose e obliquidade representam medidas de simetria de um conjunto de dados. No caso da obliquidade, coeficientes próximos de zero representam simetria. É calculada dividindo o terceiro momento central, $m_3(\mu)$, pelo desvio padrão, σ . A curtose é calculada dividindo o quarto momento central, $m_4(\mu)$, pelo desvio padrão, σ^4 , menos três e mede a concentração próxima a media de um conjunto.

Formula 5:
Obliquidade

$$Obliquidade = \frac{m_3(\mu)}{\sigma}$$

Formula 6: Curtose

$$Curtose = \frac{m_4(\mu)}{\sigma^4} - 3$$

3.5 Características baseadas em histograma

Para extração de características baseadas em estatísticas, foram utilizadas as imagens das mamas já segmentadas. Cada imagem original foi transformada em uma imagem em nível de cinza, ou seja, os valores dos pixels das imagens podem variar entre 0, preto, e 255, branco.

O histograma de uma imagem é calculado encontrando a distribuição da frequência de cada nível de cinza da imagem. Nos termogramas valores de pixels mais altos representam temperaturas mais altas. Com o histograma é possível analisar a frequência dos valores mais altos nas mamas o que pode representar um sinal de alerta. Neste trabalho os histogramas obtidos das mamas das pacientes são submetidos a técnicas de comparação de histograma. A aplicação das técnicas de comparação de histograma baseia-se na ideia de simetria das mamas. Mamas que apresentam histogramas muito diferentes podem indicar alguma alteração na distribuição de temperaturas das mesmas. Os métodos utilizados para comparação dos histogramas foram: O método de correlação, (1), Chi-Square que é um teste que verifica se a distribuição dos histogramas das mamas esquerda e direita difere significativamente em relação a uma determinada variável, (2), Interseção que é utilizado para medir a distância entre dois histogramas, (3) e Bhattacharyya, que utiliza o coeficiente de Bhattacharyya para comparar duas imagens e é uma medida do tipo divergente que tem uma interpretação geométrica direta (4).

Formula 7: Chi-Square

$$d(H_1, H_2) = \sum_I \frac{(H_1(I) - H_2(I))^2}{H_1(I)}$$

Formula 8: Interseção

$$d(H_1, H_2) = \sum_I \min(H_1(I), H_2(I))$$

Formula 9: Correlação

$$d(H_1, H_2) = \frac{\sum_I (H_1(I) - \bar{H}_1)(H_2(I) - \bar{H}_2)}{\sqrt{\sum_I (H_1(I) - \bar{H}_1)^2 \sum_I (H_2(I) - \bar{H}_2)^2}}$$

Formula 10: Bhattacharyya

$$d(H_1, H_2) = \sqrt{1 - \frac{1}{\sqrt{\bar{H}_1 \bar{H}_2 N^2} \sum_I \sqrt{H_1(I) \cdot H_2(I)}}$$

3.6 CARACTERÍSTICAS BASEADAS EM GEOMETRIA FRACTAL

Mandelbrot definiu fractal como um conjunto para o qual a dimensão de Hausdorff Besicovitch ultrapassa a dimensão topológica. Fractal exemplifica a ideia de auto similaridade o qual o comportamento espacial de um sistema, de um grupo ou de características independe da escala (BURROUGH, 1993; TURCOTTE, 1997)

O conceito de lacunaridade foi introduzido por Mandelbrot para descrever uma propriedade de fractais. Desde então tem sido utilizada também para descrever conjuntos de dados reais que podem ou não ter características fractais. (S. W. Myint, N. Lam, 2005). A lacunaridade pode ser vista como uma medida de complexidade ou heterogeneidade de um conjunto. Quanto maior os valores de lacunaridade de uma imagem maior é o grau de heterogeneidade. O Gliding-box foi o método utilizado para calcular a lacunaridade das mammas proposto por (Allain, Cloitre, 1991).

Considere a representação de uma imagem binarizada a seguir, colocamos uma caixa quadrada de escala 2 no canto superior da imagem.

Figura 16: Imagem binarizada

1	1	0	1	1	1	0	1	0	1	1	0
0	0	0	0	0	1	0	0	0	1	1	1
0	1	0	1	1	1	1	1	0	1	1	0
1	0	1	1	1	0	0	0	0	0	0	0
1	1	0	1	0	1	0	0	1	1	0	0
0	1	0	1	1	0	0	1	0	0	1	0
0	0	0	0	0	1	1	1	1	1	1	1
0	1	1	0	0	0	1	1	1	1	0	0
0	1	1	1	0	1	1	0	1	0	0	1
0	1	0	0	0	0	0	0	0	1	1	1
0	1	0	1	1	1	0	1	1	0	1	0
0	1	0	0	0	1	0	1	1	1	0	1

Fonte: Lacunarity Indices as measures of landscape texture⁸

A massa da caixa, S , é calculada contabilizando os pixels que estão ligados, 1, a cada deslizamento da caixa. No exemplo acima 2 pixels. $n(S, r)$ é a frequência de caixas de escala r que possuem S pixels ligados. $N(r)$ é a quantidade total de caixas de escala r usadas no cálculo e é definida pela seguinte equação (5), onde:

$$N(r) = (M - r + 1) \quad (5)$$

Onde M corresponde ao tamanho da imagem e r é a escala da caixa. $Q(S, r)$ corresponde à probabilidade para cada massa S e caixa de escala r e é calculado dividindo $n(S, r)$ por $N(r)$. O primeiro momento, Z_1 , é calculado multiplicando cada

8 Plotnick, R. E.; Gardner, R. H.; O'Neil, R. V. Lacunarity indices as measures of landscape texture. 1993.

massa, S , por sua respectiva probabilidade, $Q(S,r)$. O Segundo momento, Z_2 , é calculado multiplicando o quadrado da massa S , por sua respectiva probabilidade $Q(S,r)$. A seguir a Tabela 4 demonstra um exemplo hipotético de como é realizado o cálculo do primeiro e segundo momento, Z_1 e Z_2 , para uma caixa de escala $r = 2$.

Tabela 4: Cálculo Z_1 e Z_2

S	n (S,r)	n(S,r)/N(r)=Q(S,r)	$Z_1=S*Q(S,r)$	$Z_2=S*S*Q(S,r)$
0	3	24	0	0
1	35	289	289	289
2	46	380	760	1520
3	29	239	716	2157
4	4	66	264	1057

Fonte: Elaborada pelo autor

A lacunaridade para uma caixa de escala r , Δr , é calculada dividindo o segundo momento, Z_2 , pelo quadrado do primeiro momento Z_1^2

Formula 11: Lacunaridade caixa de
escala r

$$\Delta r = \frac{Z_2}{Z_1^2}$$

A lacunaridade geral da imagem é calculada através do coeficiente angular da função dos logaritmos das caixas de escalas r com o logaritmo das suas respectivas lacunaridades, Δr .

Neste trabalho as imagens não foram binarizadas, em vez disso, o cálculo das massas em cada caixa foi realizado através da diferença entre o máximo e o mínimo valor dos pixels da imagem em cada caixa de escala r , (Karperien, 2007).

O método Box-counting também foi utilizado para calcular a dimensão fractal das termografias mamárias. O procedimento do método consiste em sobrepor uma série de caixas com escalas, ε , decrescentes, sobre uma imagem binarizada e calcular a quantidade de pixels ligados para cada escala, N_ε .

A dimensão fractal é encontrada aplicando o logaritmo para os valores de N_ε e $1/\varepsilon$ e encontrando o coeficiente angular da relação.

Formula 12: Box-counting

$$\dim_{\text{box}}(S) := \lim_{\varepsilon \rightarrow 0} \frac{\log N(\varepsilon)}{\log(1/\varepsilon)}$$

Neste trabalho foram utilizados 6 valores de threshold em cada imagem, resultando em 6, respectivas, dimensões fractais. Esses valores foram calculados seguindo o mesmo princípio do método box-counting. Caixas com escalas que variam de 2^2 a 2^6 foram sobrepostas na imagem em seguida as médias foram

calculadas. O valor da maior média para cada caixa foi utilizado como valor de threshold.

O método Differential box-counting, proposto por Nirupam Sarkar and B. B. Chaudhuri, é uma alternativa no cálculo da dimensão fractal

Consideremos que uma imagem de $M \times M$ pixels pode ter sua escala diminuída para o tamanho $s \times s$ onde $M/2 \leq s < M$ e s é um inteiro. Então nós podemos estimar que $r = s/M$. Considere uma imagem como um espaço de 3 dimensões em que (x,y) denotam a posição do pixel e a terceira dimensão (z) corresponde ao nível de cinza. Seja M_k e M_l o maior e menor valores encontrados respectivamente em uma caixa de escala ε na posição (i,j) , encontramos os valores $k = M_k / \varepsilon$ e $l = M_l / \varepsilon$. A altura relativa da caixa, n_ε , é calculada usando a equação (6):

$$n_\varepsilon(i, j) = l - k + 1 \quad (6)$$

A massa N_ε é calculada da seguinte fórmula, (7):

$$N_\varepsilon = \sum_{i,j} n_\varepsilon(i, j) \quad (7)$$

A dimensão fractal do método Differential Box-counting é calculada com a equação abaixo:

Formula 13: Differential box-counting

$$\dim_{\text{box}}(S) := \lim_{\varepsilon \rightarrow 0} \frac{\log N(\varepsilon)}{\log(1/\varepsilon)}$$

4 VALIDAÇÃO

Com o modelo de predição criado pelo SVM é possível utilizar a validação cruzada como método de avaliação deste modelo. A validação cruzada consiste basicamente em dividir os dados do conjunto em subconjuntos mutualmente exclusivos, denominados conjuntos de treinamento e de validação. O conjunto de treinamento serve como base para criação do modelo de predição e o conjunto de validação é utilizado para verificar o quão assertivo é o modelo criado. Existem algumas técnicas que auxiliam o particionamento do conjunto, como o k-folds, leave on out e holdout. Neste trabalho o leave on out foi a técnica de particionamento escolhida, apesar de apresentar um alto custo computacional sua utilização é indicada quando o banco de dados é pequeno. Seja N o número de dados de um conjunto, o método leave on out, utiliza N-1 elementos do conjunto como treinamento e testa com o elemento restante. Esta técnica se repete N vezes, ou seja, para cada elemento do conjunto. A tabela abaixo ilustra o método leave on out. Cada linha representa um ciclo, células vazias correspondem aos dados usados nos treinos e as pretas os dados usados nos testes.

Tabela 5: Método leave one out

Testes	Treino	Treino	Treino	Treino
Treino	Testes	Treino	Treino	Treino
Treino	Treino	Testes	Treino	Treino
Treino	Treino	Treino	Testes	Treino
Treino	Treino	Treino	Treino	Testes

Elaborada pelo autor

Os resultados podem ser analisados através de várias estimativas. A matriz de confusão, por exemplo, é uma forma de representar o resultado da classificação do modelo. Ela consiste basicamente em uma matriz em que são dispostos os elementos classificados corretamente e os classificados incorretamente. Se a soma dos verdadeiros positivos, VP, com os verdadeiros negativos, VN, for igual à quantidade total de elementos do conjunto então o conjunto foi classificado 100% corretamente.

Tabela 6: Matriz de confusão para 2 classes. VP – verdadeiros positivos; VN – verdadeiros negativos; FP- falso positivo; FN- falso negativo.

Positivo	VP	FP
Negativo	FN	VN

Fonte: Elaborada pelo autor.

A sensibilidade, especificidade e acurácia são outras estimativas resultantes da matriz de confusão. A sensibilidade pode ser avaliada como a capacidade do modelo de identificar os casos verdadeiros positivos e pode ser calculada dividindo a

quantidade dos verdadeiros positivos pela soma dos verdadeiros positivos mais os falsos negativos. A especificidade pode ser representada como a capacidade que um modelo tem de identificar casos verdadeiros negativos e pode ser calculada dividindo a quantidade de verdadeiros negativos pela soma dos verdadeiros negativos com os falsos positivos. A acurácia é a proporção dos acertos do modelo e é calculada somando o total de acertos do modelo dividido pela quantidade total dos dados. Seja N a quantidade total dos dados, a Acurácia, Sensibilidade e Especificidade são calculadas:

Formula 14: Acurácia

$$Acurácia = \frac{VP + VN}{N}$$

Formula 15: Sensibilidade

$$Sensibilidade = \frac{VP}{VP + FN}$$

Formula 16: Especificidade

$$Especificidade = \frac{VN}{VN + FP}$$

A área sob a curva ROC é outra ferramenta utilizada para estimar quantitativamente um teste de diagnóstico. Esta medida é interpretada como a probabilidade do indivíduo portador da doença ter um resultado ao teste de diagnóstico de maior magnitude que aquele não doente (Begg, 1991). Quanto maior a capacidade de classificação de um modelo mais próximo de 1 é o valor da área. O índice de Youden é outra medida utilizada para relacionar a sensibilidade e especificidade é calculado da seguinte forma:

Formula 17: Índice de Youden

$$Y = Sensibilidade + Especificidade - 1$$

5 RESULTADOS

Os dados das pacientes com todas as características extraídas foram submetidos ao SVM. Estes resultados não foram satisfatórios chegando a 0% de acertos. Os resultados, então, foram divididos 3 grupos, cada um correspondendo aos grupos de características.

Os resultados obtidos das características baseadas em geometria fractal também obtiveram um resultado muito baixo com apenas 11% de acertos, pudemos perceber que a lacunaridade e o differential box counting prejudicaram o resultado deste grupo. A característica que melhor apresenta resultado no grupo de geometria fractal foi o box-counting que quando analisado isoladamente alcançou 33% de acertos. Os resultados correspondentes ao grupo de histograma apresentou o pior resultado da análise dos dados representando o pior grupo de características do estudo. E por fim os resultados do grupo de características baseadas em estatísticas que apresentou os resultados mais satisfatórios, alcançando aproximadamente 77% de acertos e obtendo área sob a curva ROC de 0,7778. Neste grupo a característica que melhor apresentou resultado foi a obliquidade. Segue tabela 7 com a apresentação dos resultados finais de cada grupo

Tabela 7: Análise dos resultados finais

Medidas	Geometria Fractal	Estatísticas	Histograma	Todas as características
Sensibilidade	0%	69,23%	0%	0%
Especificidade	22,22%	100%	0%	0%
Índice de Youden	-0,7777	0,6923	0	0
Curva ROC	0,1111	0,778	0	0
Acurácia	12,5%	77,77%	0%	0%

Fonte: Elaborada pelo autor

Podemos perceber que o grupo baseado em estatísticas apresentou os melhores resultados é também o grupo que possui características mais simples de serem entendidas e implementadas. Outros trabalhos também utilizaram essas medidas simples na extração de características, como (Lipari e Head, 1997), (Kuruganti e Qi, 2002), (Koyal et al. 2004) e (Wieceki, 2002). Os resultados obtidos são comparados com alguns trabalhos anteriores utilizando os métodos de avaliação: Sensibilidade, Especificidade, Índice de Youden e Acurácia.

Tabela 8: Comparação dos resultados com trabalhos anteriores

Autores	Sensibilidade	Especificidade	Youden	Acurácia
Acharya et al.	85,7%	90,5%	0,76	88,1%
Koay et al.	60%	100%	0,60	89,5%
Arora et al.	97%	44%	0,41	-
Wishart et al.	48%	70%	0,18	-
Umaderi et al.	66,7%	97,7%	0,64	-
Bonini et al.	91,9%	78,6%	0,71	88,2%
Resultado Final	69,23%	100%	0,69	77,77%

Fonte: Elaborada pelo autor.

É possível perceber que o resultado do trabalho apresentou resultados próximos a outros trabalhos desenvolvidos. Apresentou o terceiro melhor resultado quanto a sensibilidade com 69,23% contra 66,7% de Umaderi et al., 48% de Wishart et al. e 60% Koyal et al. Apresentou uma excelente especificidade, 100%, porém, baixa sensibilidade com apenas 69,23%, quanto ao Índice de Youden apresentou o terceiro melhor resultado, 0,69, ficando atrás somente de Acharya et al., 0,76 e Bonini et al., 0,71.

6 DISCUSSÕES

Quanto ao processo de segmentação das imagens, previamente, seria realizado automaticamente, utilizando como segmentador a ferramenta desenvolvida por Marques (R. S. Marques, 2012). Infelizmente a ferramenta não se mostrou eficaz com o banco de dados utilizado, não conseguindo segmentar perfeitamente a maioria das imagens. Este episódio obrigou a realização do processo de segmentação de forma manual, utilizando apenas uma ferramenta de edição de imagens, GIMP. Todo o processo de segmentação das mamas foi realizado de forma empírica sem ajuda de um profissional de saúde.

As pacientes foram classificadas como doentes ou saudáveis. Não havia a informação de qual das mamas estava como câncer, ou se ambas estavam. Esta condição fez com que os resultados para cada uma das mamas fossem reduzidos para apenas um, para cada característica. Esta abordagem pode ter prejudicado o resultado final uma vez que as mamas não foram vistas isoladamente como doentes ou saudáveis.

Os grupos baseados em histogramas que apresentou os piores resultados não conseguindo classificar corretamente nenhuma das instâncias, na realidade prejudicou os resultados finais. Das características baseadas em Geometria Fractal a única que apresentou resultados diferentes de 0 foi o box-counting, conseguindo classificar corretamente, sozinho, 33% das instâncias, mas, ainda assim, apresentou resultados muito abaixo do esperado.

7 CONCLUSÃO

O banco de dados deste trabalho contou com 18 pacientes adquiridas do banco de dados VISUAL-LAB. As pacientes utilizadas foram classificadas como doentes ou saudáveis. As imagens foram adquiridas utilizando o protocolo de aquisição estático. Os termogramas utilizados foram todos segmentados manualmente utilizando uma ferramenta de edição de imagens. A matriz de temperatura total que cada termograma possuía serviu como máscara para extração da matriz de temperatura apenas das ROIs. As características baseadas em comparação de histogramas não apresentaram resultados satisfatórios na classificação das instâncias chegando a 0% de acertos. Os métodos differential box-counting e Lacunaridade também não apresentaram bons resultados. De fato a única característica do grupo de geometria fractal obteve o mínimo de acertos foi o box-counting com 33% de acertos. As características baseadas em estatísticas apresentaram os melhores resultados deste trabalho, conseguindo alcançar 77% de acertos e área sob a curva ROC de 0,694. O resultado da especificidade alcançou 100% de acertos em detrimento com a sensibilidade, que alcançou 69,23%. Os resultados finais alcançados foram satisfatórios e próximos a resultados de trabalhos anteriores e conseguiu demonstrar a importância da utilização de medidas de estatísticas simples no processo de classificação de termografias de mamas.

8 TRABALHOS FUTUROS

Algumas sugestões para o melhoramento deste trabalho:

- g) Este trabalho foi realizado utilizando um pequeno banco de dados. Uma sugestão para analisar com mais eficácia o modelo criado pelo SVM, seria utilizar um banco de dados com mais pacientes.
- h) Outra linha importante de extensão deste trabalho seria analisar as mamas isoladamente, como doentes ou saudáveis. No entanto, seria necessário o conhecimento prévio da situação de cada uma das mamas que serão analisadas.
- i) As características baseadas em Geometria fractal não tiveram um resultado satisfatório neste trabalho. Outra abordagem de uso das mesmas seria usar caixas de escalas variadas, encontrar as áreas com as maiores médias para cada caixa e aplicar os métodos somente nestas áreas em vez de aplicar sobre toda a região das mamas.
- j) Utilizar outros classificadores, além do SVM, e analisar quais deles conseguiu classificar melhor as características extraídas.

9 REFERÊNCIAS

ACS, American Cancer Society. **What are the risk factors for breast cancer?**

Disponível em:

<<http://www.cancer.org/cancer/breastcancer/detailedguide/breast/cancer-risk-factors>> Acesso em: 10 de setembro de 2014.

INCA Instituto Nacional do Câncer. **Mamografia da prática ao controle**

Mamografia da prática ao controle, Rio de Janeiro, p. 15, 2007.

INCA Instituto Nacional do Câncer. **Deteção precoce do câncer de mama.**

Disponível em:

<http://www.inca.gov.br/conteudo_view.asp?id=1932> Acesso em: 10 de setembro de 2014.

CEPEM, Centro de Estudos e Pesquisas da Mulher. **Entendendo alguns aspectos do câncer de mama.**

Disponível em:

<<http://www.cepem.med.br/perguntas-respostas.asp>> Acesso em: 10 de setembro de 2014.

VISUAL LAB. **Imagens térmicas.**

Disponível em:

<<http://visual.ic.uff.br/dmi/>> Acesso em: 10 de setembro de 2014

ACS, American Cancer Society. **Breast cancer survival by rates.**

Disponível em:

<<http://www.cancer.org/cancer/%20breastcancer/detailedguide/breast-cancer-survival-by-stage>>. Acesso em: 10 de setembro de 2014.

CNS, Conselho Nacional de Saúde. **Dia da Mamografia.**

Disponível em:

< http://conselho.saude.gov.br/ultimas_noticias/2013/02_fev_05_mamografia.html> Acesso em 10 de setembro de 2014.

INCA Instituto Nacional do Câncer. **Mamografia da prática ao controle**

Mamografia da prática ao controle, Rio de Janeiro, 2007, p. 20-25.

Sociedade Brasileira de Mastologia. **Ultrassonografia Mamária.**

Disponível em:

<<http://www.sbmastologia.com.br/cancer-de-mama/rastreamento-diagnostico-cancer-de-mama/ultrassonografia-mamaria-16.htm>> Acesso em 10 de setembro de 2014.

BERZ, R.; SAUER, H. **“The Medical Use of Infrared-Thermography History and Recent Applications”**. Thermografie-Kolloquium, 2007.

Variação do espectro de radiação térmica com a temperatura.

Disponível em: <http://profs.ccems.pt/PauloPortugal/CFQ/Wien/Lei_Wien.html>

T. B. Borchatt, “**Análise de Imagens Termográficas para a classificação de alterações nas mamas**”. Universidade Federal Fluminense, Niterói, RJ. 2013 p.31.

FLIR. **Como funciona um Termovisor.**

Disponível em

<<http://www.flir.com/thermographyamericas/br/view/?id=55706>>. Acesso em 10 de setembro de 2014.

RCC, Radiologia Clínica de Campinas. **Padrões Mamográficos e Densidade Mamária.** Disponível em:<

http://www.radiologiaclinicadecampinas.com.br/info_11.htm > Acesso em 10 de setembro de 2014.

SOCIEDADE BRASILEIRA DE MASTOLOGIA. **Ressonância Magnética de Mama**

Disponível: <http://www.sbmastologia.com.br/cancer-de-mama/rastreamento-diagnostico-cancer-de-mama/ressonancia-magnetica-de-mama-43.htm>> Acesso em 10 de setembro de 2014.

DEBI WALKER, ND, AND TINA KACZOR, ND, FABNO. **Natural Medicine Journal, Breast Thermography: History, Theory, and Use.**

Disponível em:

<<http://naturalmedicinejournal.com/journal/2012-07/breast-thermography-history-theory-and-use>> Acesso em 10 de setembro de 2014.

SANTOS, T. B.; LIMA, RITA DE C. F. **Desenvolvimento de Ferramenta auxílio à investigação das aplicações das termografias em tumores mamários.**

Pernambuco, XVI Congresso Nacional de Iniciação Científica da UFPE, 2008, p.1.

R. RESMINI, T.B. BORCHARTT, 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.

BORCHARTT, T.B. **Análise de Imagens Termográficas para a Classificação de Alterações nas Mamas.** 115p. Dissertação de Mestrado - Universidade Federal Fluminense, Rio de Janeiro, 2013.

A. C. LORENA AND A. C. P. L. F. CARVALHO. **Uma introdução às Support Vector Machines.** São Carlos: Instituto de Ciências Matemáticas e de Computação. 66p, 2003.

R. S. MARQUES. **Segmentação automática das mamas em imagens térmicas,** Dissertação de Mestrado, Universidade Federal Fluminense, Niterói-RJ, 2012.

S. W. Myint, N. Lam. **A study of lacunarity-based texture analysis approaches to improve urban image classification.** 23p. 2005.

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, p.673–676 vol.2, 1997.

P. T. KURUGANTI AND H. Qi. **Asymmetry analysis in breast cancer detection using thermal infrared images in Engineering in Medicine and Biology**, 24th Annual Conference and the Annual Fall Meeting of the Biomedical Engineering Society EMBS/BMES Conference, Proceedings of the Second Joint, pp. 1155–1156 vol.2, 2002.

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.

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**. [21st Annual Conference and the Annual Fall Meeting of the Biomedical Engineering Society] BMES/EMBS Conference. Proceedings of the First Joint, vol. 2, p. 1108, 1999.

PLOTNICK, R. E.; GARDNER, R. H.; O'Neil, R. V. **Lacunarity indices as measures of landscape texture**. 1993.

Sagittal Plane.

Disponível em:< <https://www.teachpe.com/anatomy-physiology/planes-of-movement>> Acesso em 20 de dezembro de 2023.

VISUAL LAB.

Disponível em:<visual.ic.uff.br/dmi> Acesso em 30 de dezembro de 2014

Thermography.

Disponível em: <<http://thegenesiscenter.com/thermography/>> Acesso em 06 de janeiro de 2015

Termografia.

Disponível em:
[<http://www.mundosemdor.com.br/termografia-exame-100-seguro-nao-tem-dor-e-nao-e-invasiva/>].Acesso em 30 de dezembro de 2014.

Medical Thermography.

Disponível em: <<http://www.ebme.co.uk/articles/clinical-engineering/65-medical-thermography>> Acesso em 30 de dezembro de 2014.

Tipos de mama / alterações de pele/ imagens.

Disponível em:< <http://www.radioinmama.com.br/tiposdemama.html>> Acesso em 06 de janeiro de 2015.

WHO – World Health Organization.

Disponível em:

<<https://www.who.int/pt/news/item/03-02-2021-breast-cancer-now-most-common-form-of-cancer-who-taking-action>>.