

UNIVERSIDADE TECNOLÓGICA FEDERAL DO PARANÁ
DAINF - DEPARTAMENTO ACADÊMICO DE INFORMÁTICA
CURSO DE BACHARELADO EM SISTEMAS DE INFORMAÇÃO

BRUNO PEDROSO SCHMOCKEL
CELSON LUIZ CHANE JUNIOR

**APLICAÇÃO DE TÉCNICAS DE RECONHECIMENTO DE
PADRÕES EM IMAGENS DE TERMOGRAFIA PARA AUXÍLIO NO
DIAGNÓSTICO DE CÂNCER DE MAMA**

TRABALHO DE CONCLUSÃO DE CURSO

CURITIBA
2018

BRUNO PEDROSO SCHMOCKEL
CELSON LUIZ CHANE JUNIOR

**APLICAÇÃO DE TÉCNICAS DE RECONHECIMENTO DE
PADRÕES EM IMAGENS DE TERMOGRAFIA PARA AUXÍLIO NO
DIAGNÓSTICO DE CÂNCER DE MAMA**

Proposta de Trabalho de Conclusão de Curso apresentado ao Curso de Bacharelado em Sistemas de Informação da Universidade Tecnológica Federal do Paraná, como requisito parcial para a obtenção do título de Bacharel.

Orientadores: Leyza Baldo Dorini e
Ricardo Dutra da Silva
DAINF - Departamento Acadêmico de Informática -UTFPR

CURITIBA
2018



TERMO DE APROVAÇÃO

APLICAÇÃO DE TÉCNICAS DE RECONHECIMENTO DE PADRÕES EM IMAGENS DE TERMOGRAFIA PARA AUXÍLIO NO DIAGNÓSTICO DE CÂNCER DE MAMA”

por

Bruno Pedroso Schmockel, Celso Luiz Chane Junior

Este Trabalho de Conclusão de Curso foi apresentado como requisito parcial à obtenção do grau de Bacharel em Sistemas de Informação na Universidade Tecnológica Federal do Paraná - UTFPR - Câmpus Curitiba. O(a)(s) aluno(a)(s) foi(ram) arguido(a)(s) pelos membros da Banca de Avaliação abaixo assinados. Após deliberação a Banca de Avaliação considerou o trabalho _____.

<hr/> <p><Profa. Leyza Baldo Dorini> (Presidente - UTFPR/Curitiba)</p>	<hr/> <p><Prof. Bogdan Tomoyuki Nassu> (Avaliador(a) 1 - <Instituição>)</p>
<hr/> <p><Prof. Rodrigo Minetto> (Avaliador 2(a) - Instituição)</p>	<hr/> <p><Profa. Leyza Baldo Dorini> (Professora Responsável pelo TCC – UTFPR/Curitiba)</p>
<hr/> <p><Prof. Leonelo Dell Anhol Almeida> (Coordenador do curso de Bacharelado em Sistemas de Informação – UTFPR/Curitiba)</p>	

“A Folha de Aprovação assinada encontra-se na Coordenação do Curso.”

RESUMO

SCHMOCKEL&CHANE, Bruno&Celso. Aplicação de Técnicas de Reconhecimento de Padrões em Imagens de Termografia para Auxílio no Diagnóstico de Câncer de Mama. 2018. 43 f. Trabalho de Conclusão de Curso – Curso de Bacharelado em Sistemas de Informação, Universidade Tecnológica Federal do Paraná. Curitiba, 2018.

O câncer de mama é o tipo de câncer que mais afeta mulheres do mundo. Novas formas para detectar a doença em seus estágios iniciais são fundamentais para tratamentos mais eficientes e com maior chance de cura. Atualmente o exame padrão para a detecção de câncer de mama é a mamografia, um exame que além de doloroso é invasivo e necessita expor a paciente a radiação ionizante. As imagens termográficas são uma alternativa para a detecção de anomalias no funcionamento do corpo, que necessitam somente da captura feita através de uma câmera especial. Tais anomalias, entretanto, são subjetivas até mesmo para o profissional bem treinado. O método aqui apresenta uma acurácia de 86% com uma sensibilidade de 100%. para um conjunto amostral de 67 imagens.

Palavras-chave: Termografia. Câncer de mama. Detecção de doenças de mama. Reconhecimento de padrões.

ABSTRACT

SCHMOCKEL&CHANE, Bruno&Celso. Application of pattern recognition techniques for supporting breast cancer diagnosis. 2018. 43 f. Trabalho de Conclusão de Curso – Curso de Bacharelado em Sistemas de Informação, Universidade Tecnológica Federal do Paraná. Curitiba, 2018.

Breast cancer is the type of cancer that most affects women worldwide. New means of screening methods to detect the disease in early stages are fundamental for more efficient treatments and better recovery rates. Currently, mamography is the standart exam for breast cancer detection, a painfull and invasive exam witch exposes the pacient to ionizing radiation. Thermographic images are an alternative for detection of anomalies in the body behavior, that only needs a special camera. Although identifying such anomalies is subjective even for trained professionals, the work here presented has an accuracy of 86% with a sensibility of 100% for 67 pacient images.

Keywords: Thermography. Breast cancer. Detection of breast diseases. Pattern recognition.

LISTA DE FIGURAS

Figura 1 – Termograma de uma paciente com câncer de mama.	11
Figura 2 – Fluxo de trabalho do projeto.	16
Figura 3 – Exemplo de imagens.	17
Figura 4 – Exemplo de segmentação ideal para os casos de teste. (a) Imagem de entrada (Visual Lab, 2018) e (b) Segmentação.	18
Figura 5 – Histograma logarítmico da imagem esquerda da Figura 6.	19
Figura 6 – Exemplo da remoção de fundo	19
Figura 7 – Exemplo de detecção de bordas de Canny.	20
Figura 8 – Exemplo de limites laterais externos.	20
Figura 9 – Exemplo de limites de uma área de interesse.	21
Figura 10 – Contorno da borda da curva inframamária representada pela curva em destaque (mais grossa).	21
Figura 11 – Exemplo de limites estabelecidos pelas curvas das mamas.	22
Figura 12 – Exemplo de remoção de parte inferior do tórax e divisão das mamas.	23
Figura 13 – Ilustração das características GLCM utilizadas.	23
Figura 14 – Exemplos de segmentações bem sucedidas, para casos de teste com diferentes tamanhos de mama e estilos de marcadores.	26
Figura 15 – Exemplos de segmentações que não foram bem sucedidas.	27
Figura 16 – Exemplos de imagens desafiadoras: (a) termograma assimétrico, (b) baixo contraste, (c) marcador muito proeminente e (d) definição borrada.	27
Figura 17 – Exemplo de segmentação assimétrica.	28
Figura 18 – Exemplo de extração de bordas de imagem com baixo contraste corporal.	28
Figura 19 – Exemplo de segmentação com marcador problemático.	29
Figura 20 – Exemplo de segmentação de imagem borrada.	30

SUMÁRIO

1 – INTRODUÇÃO	8
2 – REVISÃO DE LITERATURA	10
2.1 Termografia	10
2.2 Reconhecimento de padrões	11
2.3 Extração de Características	11
2.3.1 Histograma	12
2.3.2 Matriz de Co-ocorrência de Escala de cinza	12
2.4 Algoritmos de classificação	13
2.5 Trabalhos relacionados	14
3 – METODOLOGIA	16
3.1 Aquisição de Base de Dados	17
3.2 Segmentação	18
3.2.1 Remoção do Fundo	18
3.2.2 Detecção de Bordas	19
3.2.3 Limites da Região de Interesse	20
3.2.4 Detecção da Curva Inframamária	21
3.3 Extração de Características	23
3.4 Classificação	24
3.5 Formas de Avaliação	25
4 – Resultados	26
4.1 Resultados da segmentação	26
4.2 Classificação	30
4.2.1 Avaliação das características	30
4.2.2 SVM	31
4.2.3 KNN	32
5 – CONSIDERAÇÕES FINAIS	33
5.1 TRABALHOS FUTUROS	33
Referências	34

Apêndices	37
APÊNDICE A–Lista de artigos revisados	38

1 INTRODUÇÃO

O câncer de mama é o segundo tipo de câncer mais comum entre mulheres no Brasil, chegando a uma estimativa de 59.700 novas ocorrências entre 2018 e 2019 ([Instituto Nacional de Câncer, 2018](#)), sendo a causa mais frequente de morte por câncer no sexo feminino no país ([PINHEIRO et al., 2013](#)). A sua detecção durante o estágio inicial é essencial para aumentar a taxa de sobrevivência, possibilitando que tratamentos eficientes e apropriados possam ser aplicados.

Atualmente, a mamografia, o ultrassom e a ressonância magnética são as principais técnicas para a detecção antecipada da presença de um câncer ([SATHISH et al., 2017](#)). Embora a mais utilizada seja a mamografia, ela possui diversos fatores negativos, tais como a necessidade de repetidas exposições à radiação, alto número de falsos positivos e falha em detectar um a cada cinco casos de câncer, particularmente em mulheres com mamas mais densas ([American Cancer Foundation, 2018](#)). Além disso, cerca de 90% das pacientes que passam pelo exame de mamografia reportam algum tipo de desconforto ([FREITAS WALDIR FAGANELO FIORI, 2006](#)). Estabelecer um método alternativo para o auxílio ao diagnóstico pode permitir sua utilização em regiões mais remotas, além de oferecer uma possibilidade às mulheres que não realizam o exame pelo desconforto causado. Neste contexto, a termografia mamária apresenta-se como uma ferramenta complementar para auxílio ao diagnóstico de abnormalidades da mama, tornando-se vantajosa por ser um exame não invasivo e que não expõe a paciente a radiação, bem como pela facilidade de transporte do equipamento, além de ser mais barata que o ultrassom e a ressonância ([WALKER; KACZOR ND, 2012](#)).

A detecção de evidências da existência do câncer através da termografia se dá a partir da análise da distribuição de temperatura na região da mama. Como tumores possuem uma vascularização acima do normal, isso ocasiona uma alteração da temperatura em comparação àquela de regiões saudáveis, fazendo com que a assimetria entre os padrões de calor entre as mamas seja uma das características que podem apontar para a existência de um tumor ([SATHISH et al., 2017](#)). Estudos recentes ([PAVITHRA et al., 2018](#)) apontam cada vez mais para o fato de que boas taxas de sensibilidade e especificidade podem ser atingidas quando as técnicas mais apropriadas são empregadas.

Este trabalho tem como objetivo auxiliar no diagnóstico da presença de câncer de mama em imagens termográficas com uso de técnicas de reconhecimento de padrões. Como objetivos específicos tem-se:

- Auxiliar no processo de análise de imagens de termografia, visando identificar automaticamente regiões de interesse que possam tanto ser analisadas manualmente por um especialista quanto utilizadas para extração de características.
- Desenvolver uma abordagem baseada em reconhecimento de padrões em imagens utilizando características extraídas de termogramas para dar evidências da presença de câncer

de mama em pacientes, minimizando a ocorrência de falsos positivos e principalmente de falsos negativos.

Para extrair as características utilizadas para compor os descritores utilizados no treinamento dos algoritmos de classificação SVM e k-NN, este trabalho propôs uma abordagem de segmentação de imagens termográficas da região da mama, tendo como base o trabalho de [Sathish et al. \(2017\)](#). O diferencial está no fato que as imagens são segmentadas somente de maneira automatizada e, por isso, um esforço foi realizado para melhorar o algoritmo e o tornar mais robusto. Este trabalho também diverge de [Sathish et al. \(2017\)](#) nas características ideais propostas para obter os melhores resultados de classificação.

Testes experimentais realizados em parte de uma base pública de imagens, disponibilizada pelo projeto PROENG do VisualLab da UFF, obtiveram 86% de acurácia, 100% de sensibilidade e uma especificidade de 78%. Considerando o mesmo conjunto de imagens e região de interesse, mas as características sugeridas por [Sathish et al. \(2017\)](#), as taxas foram menores: 86% de acurácia, 80% de sensibilidade e 89% de especificidade.

Este trabalho está organizado como se segue. O Capítulo 2 apresenta uma revisão da literatura, contendo uma explanação rápida dos conceitos utilizados, bem como uma listagem de trabalhos semelhantes, cujos métodos foram considerados na elaboração deste trabalho. No Capítulo 3, é apresentada a metodologia dos experimentos realizados neste trabalho. No Capítulo 4 os resultados dos experimentos são listados e discutidos. Por fim, o Capítulo 5 apresenta as conclusões, bem como considerações finais e sugestões de trabalhos futuros.

2 REVISÃO DE LITERATURA

Este capítulo apresenta alguns conceitos relacionados ao tema de estudo deste projeto, visando dar ao leitor uma visão genérica do tema sendo estudado. Para uma fundamentação teórica mais aprofundada, os autores recomendam a leitura da bibliografia indicada durante o texto.

2.1 Termografia

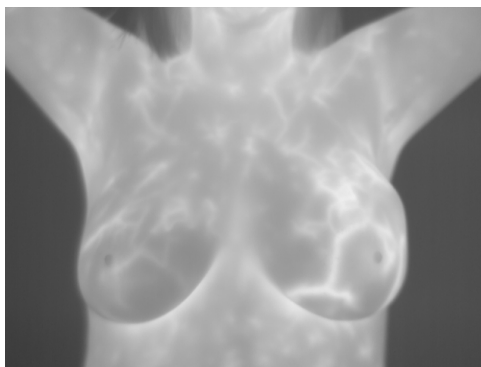
A termografia infravermelha é uma técnica de monitoramento rápida, não invasiva, não destrutiva e que não envolve contato. Seu objetivo é capturar o espectro de calor de um objeto de interesse em forma de imagens, chamadas de termogramas, que podem ser obtidas através do uso de câmeras específicas para detectar radiação infravermelha.

Todo material que possua temperatura acima do zero absoluto irradia uma quantidade de energia infravermelha, que tem relação direta com sua temperatura. Os sensores das câmeras termográficas captam esta energia e convertem as ondas infravermelhas para sinal digital, que é utilizado para criar os termogramas (DEY; ASHOUR; ALTHOUPETY, 2017). Em resumo, termogramas são representações na escala de comprimento de onda visível das ondas térmicas.

Embora o corpo humano possua uma temperatura variável, ela se mostra simétrica entre os lados do corpo em indivíduos saudáveis (UEMATSU, 1985). Portanto, a detecção de padrões assimétricos de temperatura em certas regiões do corpo pode auxiliar no diagnóstico de doenças que, de outra maneira, necessitariam de exames invasivos. No caso do câncer de mama, esta assimetria é causada pela produção de óxido nítrico pelo nódulo cancerígeno, que impede o controle do sistema nervoso sobre o fluxo dos vasos sanguíneos. Isto causa vasodilatação, que provoca um aumento de modo anormal na temperatura da região (BORCHARTT et al., 2013; FOK; NG; TAI, 2002).

Neste trabalho, desenvolve-se uma abordagem para analisar a distribuição de temperatura na região das mamas, relacionando-a com assimetria térmica entre o lado esquerdo e o direito do corpo. Tal assimetria é uma forma de se observar uma discrepância no funcionamento biológico, comparando-se tecidos da própria pessoa. A Figura 1 é exemplo de um termograma cuja paciente possui uma marca assimétrica na mama esquerda e que em exame de mamografia foi confirmada como portadora de câncer de mama.

Cabe ressaltar que termogramas também são de grande utilidade na medicina em áreas como transplantes de rim, dermatologia, diabetes, doenças vasculares periféricas (LAHIRI et al., 2012) e detecção de lesões de exercício (HILDEBRANDT; RASCHNER; AMMER, 2010).



Fonte: (Visual Lab, 2018)

Figura 1 – Termograma de uma paciente com câncer de mama.

2.2 Reconhecimento de padrões

A descoberta automática de (ir)regularidades nos dados através de algoritmos computacionais é o propósito do campo de reconhecimento de padrões. Através do uso destas (ir)regularidades é possível tomar ações como classificar e categorizar os dados (BISHOP, 2006). No campo de análise de imagens, estas técnicas podem ser utilizadas para classificação, visão computacional, reconhecimento de caracteres, comunicação (compressão) e diagnóstico médico (FU; ROSENFELD, 1976).

No contexto deste trabalho, são utilizados os classificadores SVM e *K-Nearest-Neighbours* (Seção 2.4). Eles são treinados com características estatísticas extraídas do histograma e de Matrizes de Co-ocorrência de Escala de Cinza (Seção 2.3.2), todas calculadas considerando apenas a região de interesse de ambas as mamas de cada termograma.

2.3 Extração de Características

A textura é o aspecto de uma superfície. Em visão computacional, é uma imagem, ou parte dela, que segue um padrão específico, com um padrão estatístico ou de repetição (TAMURA; MORI; YAMAWAKI, 1978). Texturas podem ser utilizadas para identificar objetos e regiões de interesse em imagens (HARALICK; SHANMUGAM; DINSTEN, 1973). Neste trabalho, o termo textura se refere às características da temperatura da superfície da pele, tais como uniformidade, aspereza, intensidade, densidade e regularidade.

A forma mais comum de descrição de características de textura é a abordagem estatística, tendo como vantagem o fato de poder ser aplicada à texturas que não apresentam boa regularidade. Tal abordagem pode ser classificada em duas categorias. A estatística de primeira ordem baseia-se em informações de histogramas de primeira ordem, onde não é possível obter uma diferenciação espacial da distribuição dos níveis de cinza. A estatística de segunda ordem, como matrizes de co-ocorrência, possibilita tal diferenciação, sendo possível relacionar espacialmente os valores de uma imagem.

2.3.1 Histograma

O histograma é uma função que conta o número de observações de cada intervalo de uma classe. Em imagens de escala de cinza, o histograma de frequência é comumente utilizado para avaliar a distribuição das intensidades de cor. Estatísticas de primeira ordem aplicadas a histogramas de frequência utilizadas neste trabalho são:

- Média: a média aritmética é dada pela soma de todos os valores dividida pela sua quantidade. A média descreve o centro da distribuição, podendo permitir a comparação com centros de massa. A média é descrita pela Equação 1;

$$\mu = \frac{1}{n} \sum_{i=1}^n f_i x_i. \quad (1)$$

- Variância: é a média dos quadrados das diferenças entre os valores e o valor da média. Ela mede o espalhamento dos dados e é descrita pela Equação 2;

$$\theta = \frac{1}{n} \sum_{i=1}^n (x_i - \mu)^2. \quad (2)$$

- Entropia: é uma medida da diversidade de dados, sendo definida pela Equação 3, em que d_i é a distribuição do valor no histograma;

$$E = - \sum_{i=1}^n x_i \log(x_i). \quad (3)$$

- Curtose: descreve a semelhança na distribuição das amostras do histograma (ou o achatamento da curva de distribuição). É calculada através da divisão do quarto momento central pelo quadrado da variância, como definido na Equação 4;

$$K = \frac{1}{n} \sum_{i=1}^n \left(\frac{x_i - \mu}{\sigma} \right)^4 - 3. \quad (4)$$

- Obliquidade: indica a assimetria da curva de distribuição dos dados, dada pela Equação 5.

$$S = \frac{1}{n} \sum_{i=1}^n \left(\frac{x_i - \mu}{\sigma} \right)^3. \quad (5)$$

2.3.2 Matriz de Co-ocorrência de Escala de cinza

A matriz de co-ocorrência de escala de cinza (*Gray Level Co-occurrence Matrix*) é um tipo de histograma de segunda ordem. Ela consiste de uma matriz de frequências relativas que descreve a distribuição espacial dos níveis de cinza de uma imagem, dadas uma distância e uma direção. Esta estrutura de dados é ideal para a descrição de texturas, dado que estas são caracterizadas pela relação espacial que os níveis de cinza da imagem tem uns com os outros com base na relação angular na distância (HARALICK; SHANMUGAM; DINSTEN, 1973).

Características podem ser extraídas de matrizes de co-ocorrência obtendo valores estatísticos de primeira e segunda ordem. Neste trabalho, estatísticas de primeira ordem empregadas foram média, variância, entropia, curtose e oblíquidade, e de segunda ordem são as que seguem:

- **Contraste:** medida que descreve o quão diferentes são valores vizinhos da imagem. Um vizinho é definido pela direção e distância da matriz de co-ocorrência. Calculado conforme Equação 6:

$$\mu = \sum_{i,j=0}^{255} M_{i,j}(i-j)^2. \quad (6)$$

em que M é a matriz de co-ocorrência.

- **Homogeneidade:** medida que descreve o quão homogêneos são os vizinhos da matriz de co-ocorrência. Calculada conforme Equação 7:

$$\mu = \sum_{i,j=0}^{255} \frac{M_{i,j}}{1+(i-j)^2}. \quad (7)$$

em que M é a matriz de co-ocorrência.

- **Correlação:** medida calculada conforme Equação 8 e que descreve o quanto os pixels de uma imagem estão correlacionados com seu vizinho:

$$\mu = \sum_{i,j=0}^{255} M_{i,j} \frac{(i-\mu_i)(j-\mu_j)}{\sqrt{(\sigma_i^2)(\sigma_j^2)}}. \quad (8)$$

em que M é a matriz de co-ocorrência, e μ e σ denotam a média e o desvio padrão das distribuições marginais, respectivamente.

- **Energia:** medida que descreve o quão uniforme é uma textura. É calculada a partir do somatório dos valores ao quadrado, conforme Equação 9:

$$\mu = \sum_{i,j=0}^{255} P_{ij}^2. \quad (9)$$

em que M é a matriz de co-ocorrência,

2.4 Algoritmos de classificação

Este trabalho utiliza dois algoritmos. O primeiro, chamado Máquina de Vetores de Suporte (do inglês, *Support Vector Machine* - SVM), é um classificador linear binário não probabilístico que busca criar um hiperplano que separa dois grupos de dados em um espaço multi-dimensional (VAPNIK, 1982).

O segundo, chamado em português de k vizinhos mais próximos (do inglês *k-nearest neighbors* k -NN), é um algoritmo de aprendizado supervisionado que se baseia na distância dos pontos no hiperespaço de características. Quando um novo ponto é adicionado ao espaço

de características, os k pontos mais próximos definem o grupo ao qual pertence o novo ponto, através de uma função de pesos baseada nas características semelhantes entre os elementos nos k vizinhos e o elemento a ser classificado, bem como na quantidade de elementos de um mesmo grupo contido nos k elementos mais próximos (YANG; LIU, 1999).

2.5 Trabalhos relacionados

Esta seção se concentra em discutir os trabalhos encontrados mais relevantes que apresentaram uma combinação de protocolos bem documentados e semelhantes às necessidades do projeto realizado na UTFPR; uso de características variadas, explorando sua relevância ao problema, e explanando características mais eficientes; descrição aprofundada dos métodos de classificação utilizados; e apresentação dos resultados estatísticos, principalmente a acurácia, sensibilidade, especificidade. Uma lista extensa dos artigos avaliados está disponível no Apêndice A.

Um dos trabalhos mais próximos realizados é a tese de doutorado de Silva (2015), que utiliza a mesma base deste trabalho (<http://visual.ic.uff.br/en/proeng/>). Neste trabalho, são utilizadas 20 termogramas por paciente, que constituem um protocolo dinâmico. Estes termogramas são segmentados manualmente sob um conjunto de regras previamente estabelecidas. Características estatísticas, de textura, agrupamento e comparação de histogramas são utilizadas, bem como características específicas para análise de imagens dinâmicas, como índices de diversidade, estatística espacial e lacunaridade. Estas características são jogadas em séries temporais, com um valor para cada imagem de uma paciente. Das séries temporais é calculado a amplitude e a raiz quadrada da soma ao quadrado. As informações extraídas das séries temporais são passadas para um classificador SVM. Os testes resultaram numa acurácia de 100% para um conjunto selecionado e balanceado de 64 imagens.

ACHARYA et al. utilizam uma câmera de vídeo para realizar a aquisição das imagens, entretanto imagens estáticas são usadas para a classificação. Seu protocolo estipula exposição ao ambiente com temperatura entre 20° e 22° Celsius por 15 minutos. As imagens são adquiridas com a câmera posicionada a um metro da paciente. Das 16 características estudadas, as escolhidas para o classificação momentos estatísticos e características extraídas de Matriz de Coocorrência em Escala de Cinza (GLCM; do inglês *Gray Level Co-occurrence Matrix*). O *dataset* utilizado consiste de 36 imagens, 18 positivas e 18 negativas. Usando *Support Vector Machine* (SVM) foi documentado 85.71% de sensibilidade e 90.48% de especificidade.

Em Kapoor e Prasad (2010), a paciente é exposta de peito aberto a um ventilador por 10 minutos antes da captura de imagens (uma frontal e duas laterais) em uma sala escura. As áreas de interesse são extraídas usando detecção de bordas e Transformada de Hough. Os pixels são então agrupados com o uso de características como curtose e obliquidade. O objetivo da clusterização é criar grupos baseados na simetria ou assimetria da temperatura corporal das pacientes. Por se tratar de um trabalho focado no desenvolvimento de método, não há resultados apresentados.

Por fim, [Sathish et al. \(2017\)](#), utilizando a base da Universidade Federal Fluminense (<http://visual.ic.uff.br/en/proeng/>). Detector de Canny e Descritor de Projeção Horizontal (HPP) são utilizados para encontrar as áreas de interesse. Assim como em ([ACHARYA et al., 2012](#)), um SVM é utilizado para classificar as imagens por características de entropia, curtose, obliquidade e variância aplicadas sobre o histograma da imagem. Além disso, são utilizadas GLCM's em quatro direções, e delas são extraídas a média, variância, entropia, contraste, correlação, energia e homogeneidade. Foram obtidos resultados com acurácia de 90%, sensibilidade de 87,5% e especificidade de 92.5%.

3 METODOLOGIA

A metodologia deste trabalho foi dividida em cinco partes principais, como ilustrado na Figura 2. A primeira, explicada na Seção 3.1, trata da aquisição dos dados, bem como o seu tratamento e normalização. A base utilizada é de livre acesso e de autoria do PROENG (Visual Lab, 2018).

A segunda parte consiste na segmentação da região de interesse. A Seção 3.2 detalha a abordagem proposta, que tem como base os métodos propostos por Sathish et al. (2017). Cabe ressaltar que todas as segmentações foram feitas de maneira automática e nenhum ajuste manual foi aplicado.

A extração de características para aprendizado, explicada na Seção 3.3, foca em texturas e estatísticas dos histogramas, como realizado por Acharya et al. (2012) e Kapoor e Prasad (2010).

As características são então utilizadas em algoritmos de aprendizagem, mais especificamente *Support Vector Machine* (SVM) e *k-nearest neighbours* (k-NN). Os resultados são avaliados considerando a acurácia, sensibilidade e especificidade (Seção 3.5), considerando *k-folding* para validação cruzada.

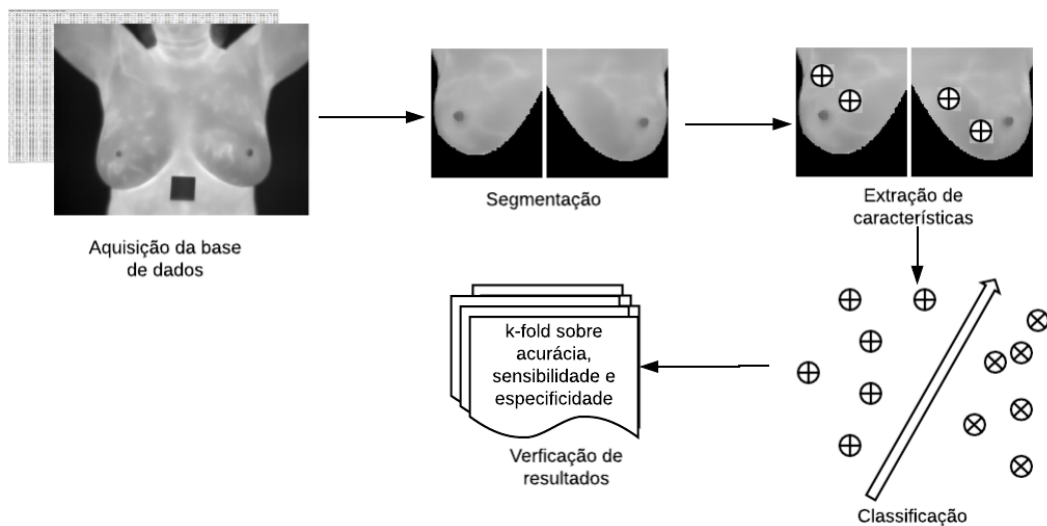


Figura 2 – Fluxo de trabalho do projeto.

Fonte: Autoria própria

3.1 Aquisição de Base de Dados

O *Database For Mastology Research* (DRM) é uma plataforma online do PROENJ, um projeto do VisualLab da Universidade Federal Fluminense. Ele mantém dados coletados sobre imagens mastológicas, visando viabilizar estudos sobre detecção de câncer de mama (Visual Lab, 2018). O banco é composto por imagens térmicas frontais e laterais obtidas por protocolo dinâmico de 287 diferentes pacientes.

Para baixar as imagens, procurou-se inicialmente recuperar as imagens no formato PNG disponíveis em um banco de dados conectado ao site da PROENJ (Visual Lab, 2018). Entretanto, havia inconsistências entre as identificações dos arquivos baixados e os requisitados. Em contato com os responsáveis, foi informado que isso era causado por uma migração de sistema e que a alternativa seria a utilização das matrizes de temperatura em formato .TXT disponíveis na página de cada paciente. Para garantir a integridade dos arquivos, foram baixados aproximadamente 2800 arquivos de texto de 122 pacientes de maneira manual.

Destas pacientes, acabou-se utilizando apenas uma imagem de cada uma (dado que a abordagem proposta neste trabalho considera termografia estática). Destas, apenas 67 passaram pela segmentação automática com resultados satisfatórios, sendo que 30% destas são pacientes com câncer. Problemas encontrados na segmentação serão explanados na Seção 4.1.

Como os dados são disponibilizados no formato de matrizes de temperaturas em arquivo texto, para convertê-los em imagens foi utilizada a transformação proposta por Silva (2015), descrita na Equação 10:

$$value[i] = ((255 * (value[i] - min)) / (max - min)) \quad (10)$$

em que $value[i]$ é o valor de cinza em um pixel i , e max e min são o máximo e mínimo global entre todas as imagens analisadas, respectivamente.

A Figura 3 ilustra alguns exemplos de imagens da base.

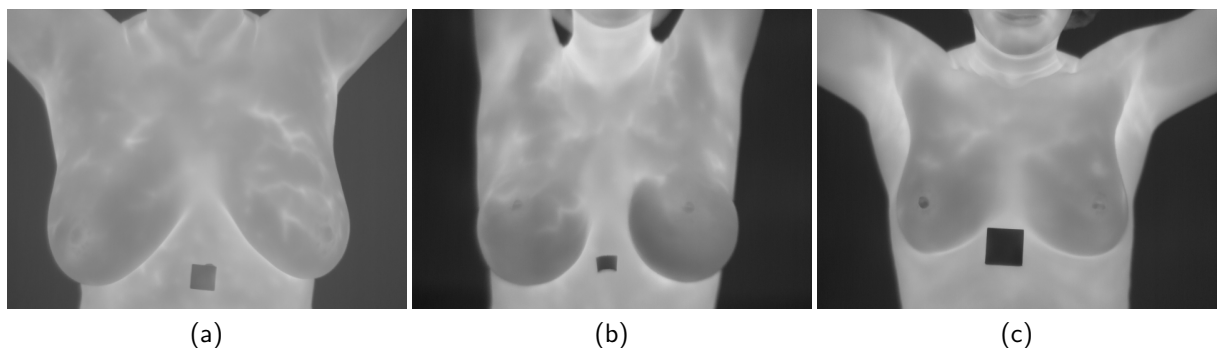


Figura 3 – Exemplo de imagens.

Fonte: (Visual Lab, 2018)

3.2 Segmentação

Nesta seção são explicados os algoritmos utilizados para segmentação das imagens, visando manter apenas a região de interesse que será utilizada na extração de características. Esta região será dividida em duas em um ponto de simetria, de forma a permitir a comparação da distribuição da temperatura nos dois lados do corpo. A Figura 4 ilustra o resultado esperado.

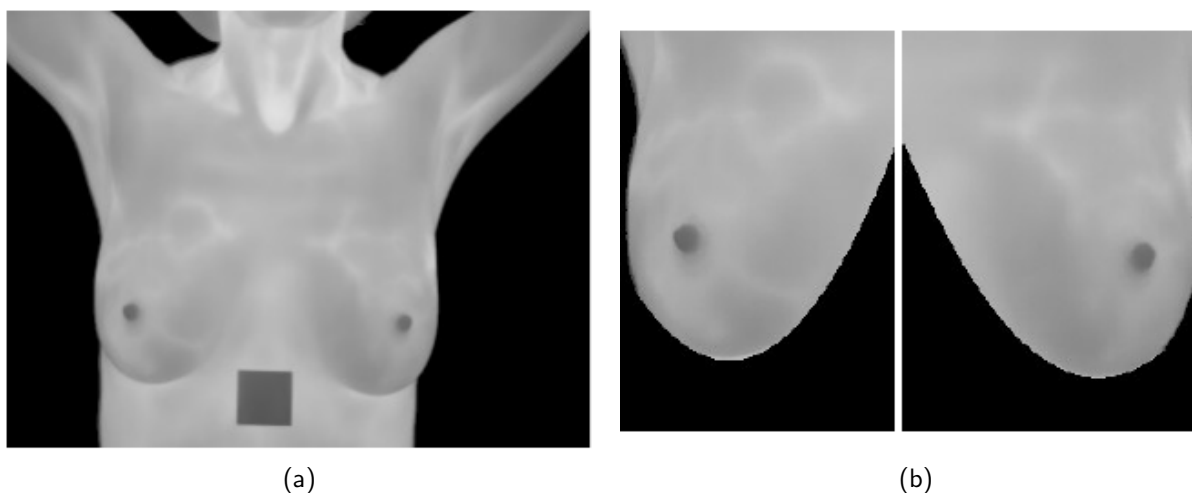


Figura 4 – Exemplo de segmentação ideal para os casos de teste. (a) Imagem de entrada (Visual Lab, 2018) e (b) Segmentação.

Fonte: Autoria própria.

Após diversos testes experimentais realizados, o processo com melhor sucesso para as imagens do banco PROENJ-UFF segue o seguinte passo-a-passo:

1. Remove-se o fundo da imagem;
2. Utiliza-se o filtro de Canny para detecção das bordas;
3. Encontra-se os limites horizontais e verticais da região de interesse;
4. Encontra-se a curva que define a região inframamária e a utiliza para dividir a imagem e apagar a região inferior do tórax.

Na sequência serão discutidos cada um destes passos, visando detalhar a metodologia considerada no trabalho.

3.2.1 Remoção do Fundo

A remoção do fundo das imagens foi realizada através da análise do histograma do termograma. Observou-se que existe um pico na região correspondente aos menores valores de intensidade, o qual é logo seguido de uma queda. Tal região define o fundo da imagem, dado que este está sempre mais frio do que as pacientes e, portanto, é composto pelos valores de menor intensidade. Um exemplo de histograma pode ser visto na Figura 5 .

A remoção do fundo é realizada pela detecção e remoção de todos os pixels relacionados a esta região inicial do histograma. Um exemplo pode ser visto na Figura 6.

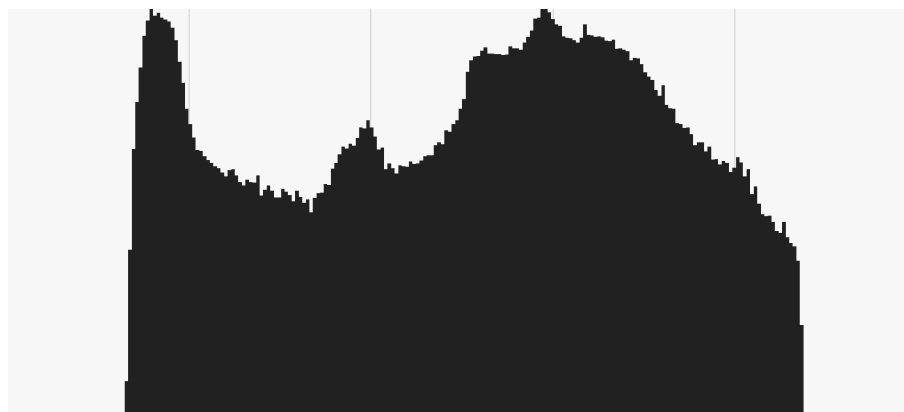


Figura 5 – Histograma logarítmico da imagem esquerda da Figura 6.

Fonte: Autoria própria.

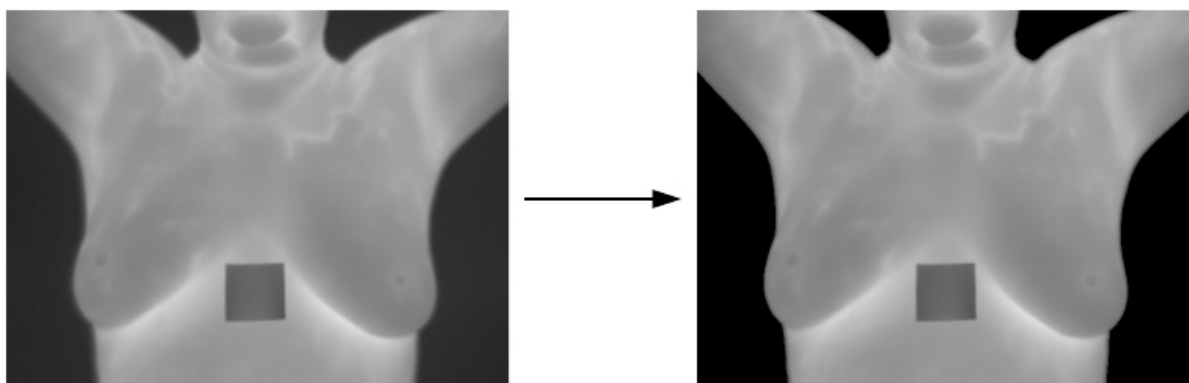


Figura 6 – Exemplo da remoção de fundo

3.2.2 Detecção de Bordas

O detector de bordas de Canny é um algoritmo para detecção de bordas proposto por Canny em 1986, o qual utiliza a magnitude dos gradientes dos pixels para identificar bordas fortes (ou seja, pixels que são definitivamente bordas) e o ângulo destes gradientes em relação a bordas fracas (ou seja, pixels que estão acima de um mínimo aceitável, mas que não se tem certeza se são bordas ou não), visando preencher buracos entre os pixels de bordas fortes (CANNY, 1986).

Como as imagens da base utilizada apresentam pouca definição nas bordas que delimitam o corpo, nenhum filtro Gaussiano foi aplicado (para que detalhes importantes não se perdessem). Por esse mesmo motivo, os parâmetros de Canny são relativamente baixos, com limite superior de 40 e inferior de 20. Para fortalecer mais a conexão entre as bordas encontradas, ainda é feita uma operação morfológica de dilatação com um *kernel* circular de tamanho 3×3 . Um exemplo da aplicação deste algoritmo é apresentado na Figura 7.

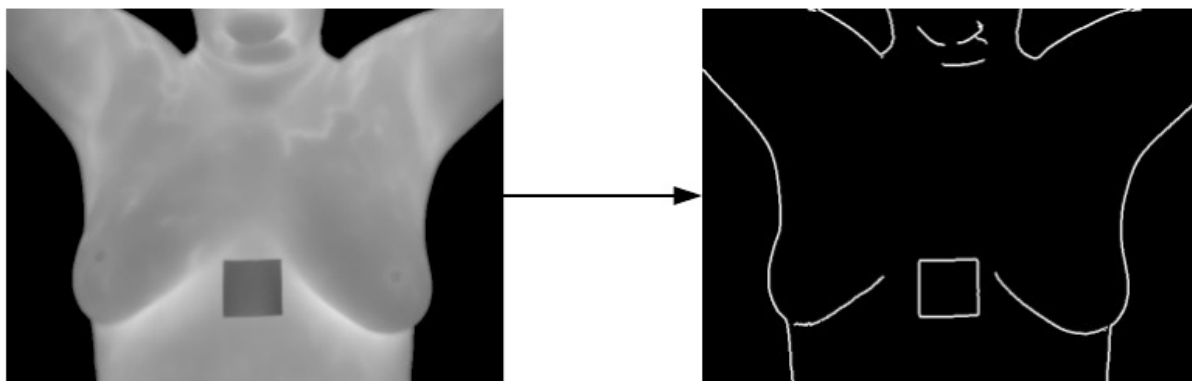


Figura 7 – Exemplo de detecção de bordas de Canny.

Fonte: Autoria própria.

3.2.3 Limites da Região de Interesse

Na abordagem proposta, a região de interesse é delimitada por limites verticais e horizontais, detectados a partir das bordas identificadas segundo descrito na seção anterior.

Para encontrar os limites verticais, é realizada uma linha de varredura no sentido horizontal, partindo dos lados esquerdo e direito. A cada iteração verifica-se se a linha atravessa a borda mais de uma vez e, caso sim, a coluna atual será a escolhida como limite. Caso nenhuma das linhas cumpra o critério, significa que a mama não excede o limite do tórax, e então esta região será o limite utilizado. Para tal, é realizada uma nova linha de varredura, dessa vez contando o número de pixels da cada coluna. A primeira coluna que passar um *threshold* ou a coluna com maior número de pixels será selecionada, como ilustrado na Figura 8.

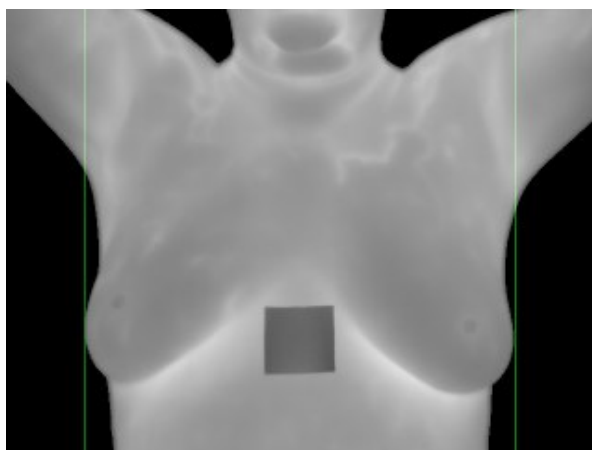


Figura 8 – Exemplo de limites laterais externos.

Fonte: Autoria própria.

Das bordas horizontais, a superior é a mais simples, sendo marcada no ponto correspondente a 60% da altura da imagem (porcentagem determinada experimentalmente). A inferior é delimitada pela linha inframamária, que pode ser definida através da contagem do número de pixels brancos que cada linha da imagem de borda possui. Como a linha que encontra a curva

inferior do seio possui um alto número de pixels, basta localizar a primeira linha onde o número de pixels brancos é mais alto do que a média em uma leitura de baixo para cima (SATHISH et al., 2017). O resultado final pode ser visto na Figura 9.

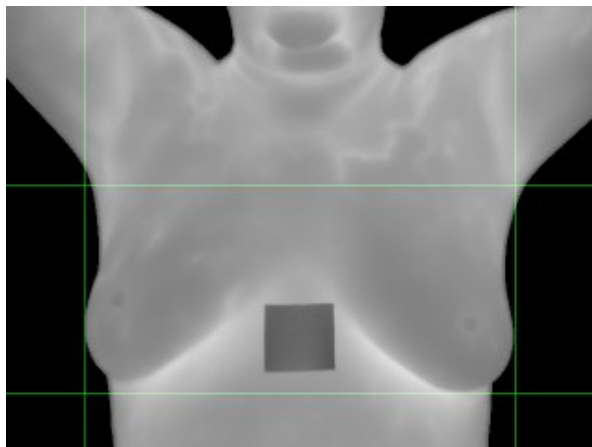


Figura 9 – Exemplo de limites de uma área de interesse.

Fonte: Autoria própria.

Algumas imagens que possuem marcador corporal abaixo da região das mamas acabaram detectando o limite inferior abaixo da região ideal. Isto acabou não representando um problema pelo final da segmentação remover toda região abaixo da curva das mamas, como é explicado na próxima seção.

3.2.4 Detecção da Curva Inframamária

Para encontrar a linha de simetria da imagem foi utilizado o método sugerido por Sathish et al. (2017), o qual ajusta as bordas que representam a curva de cada mama por uma equação de segundo grau e busca o ponto em que elas se interceptam. Para definir estas curvas, são escolhidos os componentes conexos da imagem de bordas que possuam maior tamanho horizontal. Na Figura 10, as curvas da mama (que estão realçadas com linhas mais grossas) foram identificadas utilizando este método.

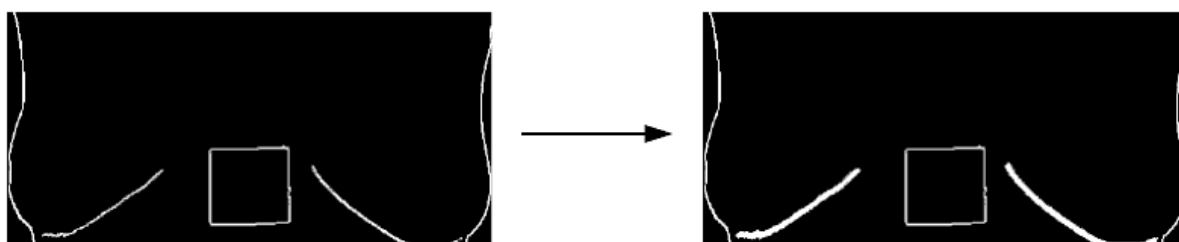


Figura 10 – Contorno da borda da curva inframamária representada pela curva em destaque (mais grossa).

Fonte: Autoria própria.

A linha de simetria será definida no ponto de encontro das projeções (SATHISH et al., 2017). O ajuste da curva busca a função polinomial descrita na Equação 11 que melhor descreva o conjunto de pontos que define as bordas das mamas.

$$y(x, \mathbf{w}) = w_0 + w_1x + w_2x^2 + \dots + w_Mx^M = \sum_{j=0}^M w_jx^j \quad (11)$$

em que M é a ordem do polinômio, \mathbf{w} é o vetor que denota os coeficientes w_0, \dots, w_M e x, \dots, x^M os pontos. A curva deve obrigatoriamente formar uma parábola com concavidade voltada para cima, como ilustrado na Figura 11.

Para determinar a curva, é realizada a minimização do erro segundo a Equação 12:

$$E(\mathbf{w}) = \frac{1}{2} \sum_{n=1}^N \{y(x_n, \mathbf{w}) - t_n\}^2 \quad (12)$$

em que t denota os pontos dados (*target*) e \mathbf{w} é o valor que define uma curva em relação aos pontos originais (BISHOP, 2006). O valor de $E(\mathbf{w})$ não é negativo e se aproxima de zero se e somente se o valor descrito pela função $y(x_n, \mathbf{w})$ se aproxima do valor de treinamento t_n .

O resultado do processo pode ser observado na Figura 11.

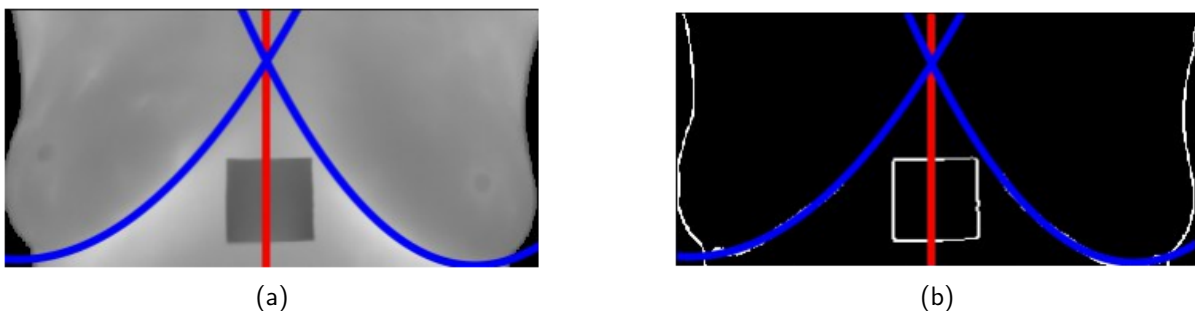


Figura 11 – Exemplo de limites estabelecidos pelas curvas das mamas.

Fonte: Autoria própria.

Como buscamos parábolas bem definidas, podemos nos limitar a buscar equações do segundo grau e assim nosso M .

Para encontrar a função da curva a partir dos componentes conexos, é utilizado a função *polyfit* da biblioteca *numpy* do *Python*. Para encontrar o ponto comum entre as curvas foi calculado o zero da função da subtração das duas equações, utilizando *fsolve* da biblioteca *scipy.optimize* do *Python*.

As equações encontradas também são utilizadas para remover da imagem a parte inferior e central do tórax que não definem as mamas. Para isso, basta remover todos os pixels que não estão contidos nas parábolas definidas pelas equações. A Figura 12 mostra um exemplo da segmentação final.

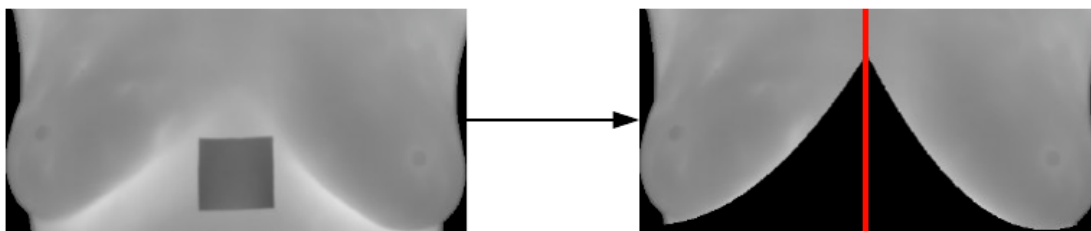


Figura 12 – Exemplo de remoção de parte inferior do tórax e divisão das mamas.

Fonte: Autoria própria.

3.3 Extração de Características

Valores estatísticos escolhidos com base em [Sathish et al. \(2017\)](#) são extraídos do histograma e da GLCM. São construídas 4 matrizes de co-ocorrência para cada lado da segmentação (ou seja, para as regiões segmentadas como mamas esquerda e direita), sendo uma para cada direção (vertical, horizontal e duas diagonais de 45°) - com distância 1 entre os pixels. Destas matrizes, serão recuperados os valores estatísticos citados na Seção 2.3.2: média, variância, entropia, contraste, correlação, energia e homogeneidade. A Figura 13 ilustra a construção e extração das características das matrizes de co-ocorrência.

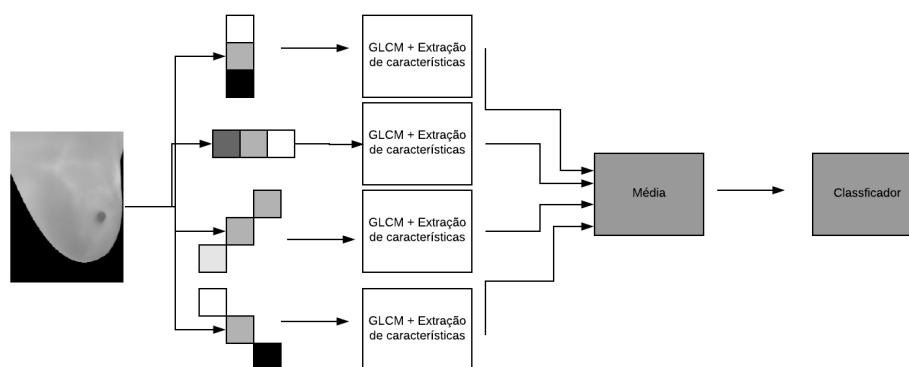


Figura 13 – Ilustração das características GLCM utilizadas.

Fonte: Autoria própria.

Do histograma são extraídas as características citadas na Seção 2.3.1: média, variância, assimetria, curtose e entropia, também considerando-se cada lado da segmentação. Para desconsiderar o fundo, valores que representam 0 foram removidos do histograma e das matrizes de co-ocorrência.

O vetor de características final de uma amostra é formado tanto pelos valores do lado direito quanto do lado esquerdo da segmentação.

3.4 Classificação

Foram considerados dois algoritmos de classificação, mais especificamente SVM e k-NN, utilizando as características detalhadas na seção anterior. Os testes foram aplicados em uma base contendo 22 imagens de pacientes portadores de câncer e 45 imagens de pacientes saudáveis, totalizando 67 amostras. Destas, 80% são reservadas para treinamento e 20% para teste.

Foram treinados cinco classificadores com parâmetros escolhidos utilizando a função *GridSearchCV* do Python, a qual realiza uma busca por força bruta do melhor resultado de classificação, dado um método de avaliação e uma lista de parâmetros possíveis.

Foram considerados um classificador k-NN e quatro variações do SVM com diferentes *kernels* disponibilizados na função *NuSVC* da biblioteca *sklearn*, sendo eles: Linear, RBF, Polinomial e Sigmoide.

Abaixo, seguem os parâmetros informados na função *GridSearchCV* para cada classificador:

- k-NN: quantidade de vizinhos, algoritmo para cálculo da distância, número de nós (informação utilizada nos algoritmos que reduzem o espaço de busca para cálculo da distância) e atribuição de pesos diferentes para distâncias.

```

1 KNN = {'n_neighbors': [1,2,3,4,5,6,7],
2       'algorithm' : ['auto', 'ball_tree', 'kd_tree', 'brute'
3                     ],
4       'leaf_size': [30,40,50,60,70],
5       'weights': ["uniform", "distance"]
6       }

```

- SVM: todos os *kernels* recebem os mesmos parâmetros de *nu* (limite superior da fração dos erros de margem) e *gamma* foi passada aos *kernels* em que ela se aplica. Para o *kernel* polinomial foi passado o grau da função.

```

1 SVM_Linear = {'kernel': ['linear'], 'nu': [0.2,0.3,0.4,0.5]}
2 SVM_RBF = {'kernel': ['rbf'], 'nu': [0.2,0.3,0.4,0.5],
3           'gamma': np.logspace(-6,1,6)}
4 SVM_Sigmoid = {'kernel': ['sigmoid'], 'nu': [0.2,0.3,0.4,0.5],
5              'gamma': np.logspace(-6,1,6)}
6 SVM_Polinomial = {'kernel': ['poly'], 'nu': [0.2,0.3,0.4,0.5],
7                 'gamma': np.logspace(-6,1,6), 'degree': [1,2,3,4,5]}

```

A escolha dos melhores resultados utiliza um balanceamento entre sensibilidade e acurácia, e a validação cruzada é feita utilizando *Stratified KFold* com 20 *splits*.

3.5 Formas de Avaliação

Para avaliar os resultados das classificações são necessários alguns conceitos:

- VP: verdadeiro positivo - resultado positivo na classificação e verdadeiro na realidade (classificação correta);
- VN: verdadeiro negativo - resultado negativo na classificação e negativo na realidade (classificação correta).
- FN: Falso negativo - resultado negativo na classificação e positivo na realidade (classificação incorreta).
- FP: Falso positivo - resultado positivo na classificação e negativo na realidade (classificação incorreta).

A acurácia é a taxa de acertos sobre o total de resultados, definido pela Equação 13. Visa avaliar o quão correta foi a classificação em relação ao universo total. Em outras palavras, consiste na percentagem absoluta de classificações corretas.

$$Acurácia = \frac{(VP + VN)}{(VP + VN + FP + FN)}. \quad (13)$$

A sensibilidade, ou taxa de verdadeiros positivos, é definida pela Equação 14. Se uma pessoa está doente, o quão certo é o exame? Considerando que o exame tem como objetivo detectar uma doença letal, minimizar o número de falsos negativos (pessoas doentes não detectadas pelo exame) é fundamental.

$$Sensibilidade = \frac{VP}{(VP + FN)}. \quad (14)$$

A especificidade ou taxa de verdadeiros negativos, é definida pela Equação 15. Se uma pessoa não está doente, o quão certo é o exame? Após garantir uma boa sensibilidade é necessário garantir que aqueles que não possuem a doença não sejam expostos a mais exames. No caso do câncer de mama, após a confirmação por exame, é necessária a realização de biópsia para confirmação. Logo, aumentar a especificidade é garantir que não sejam feitos exames em pessoas sem a doença.

$$Especificidade = \frac{VN}{(VN + FP)}. \quad (15)$$

O *k-folding* é forma de validação cruzada em que os dados são divididos em k grupos. Um deles é usado como teste e os demais $k-1$ grupos são usados para treinamento. O processo é repetido k vezes alternando os grupos de treinamento e teste e os resultados acumulados (ZIEGLER et al., 2005). Este método permite uma avaliação com base em um conjunto de amostras mais significativo.

4 Resultados

Este capítulo apresenta os resultados obtidos nas fases de segmentação e classificação, bem como os principais problemas encontrados.

4.1 Resultados da segmentação

A Figura 14 ilustra alguns exemplos em que o procedimento para segmentação descrito anteriormente obteve sucesso.

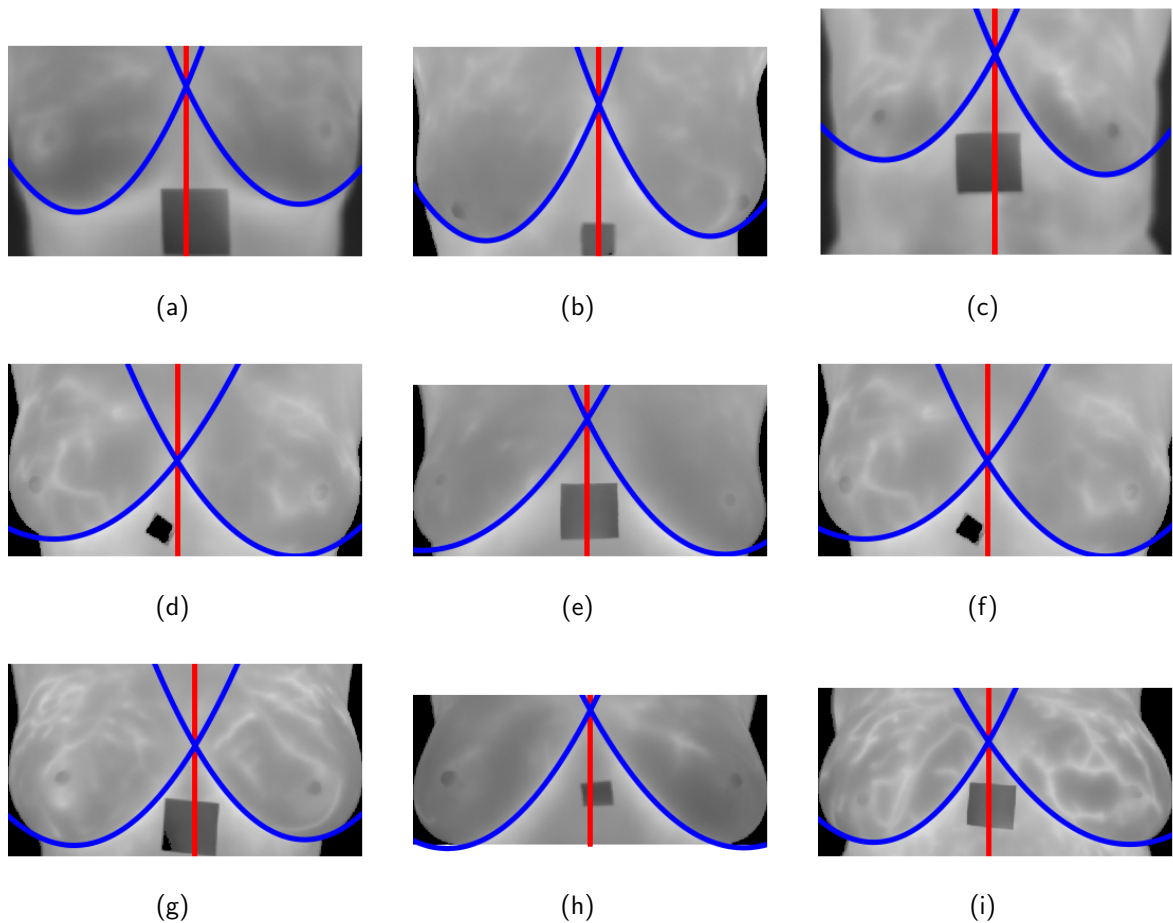


Figura 14 – Exemplos de segmentações bem sucedidas, para casos de teste com diferentes tamanhos de mama e estilos de marcadores.

Fonte: Autoria própria.

Entretanto, como mencionado anteriormente, das 135 imagens obtidas da base da UFF, apenas 67 obtiveram resultados satisfatórios de segmentação automática. Destas, 21 são de pacientes com câncer e 46 de pacientes saudáveis. Exemplos em que o processo não obteve resultados satisfatórios são apresentados na Figura 15.

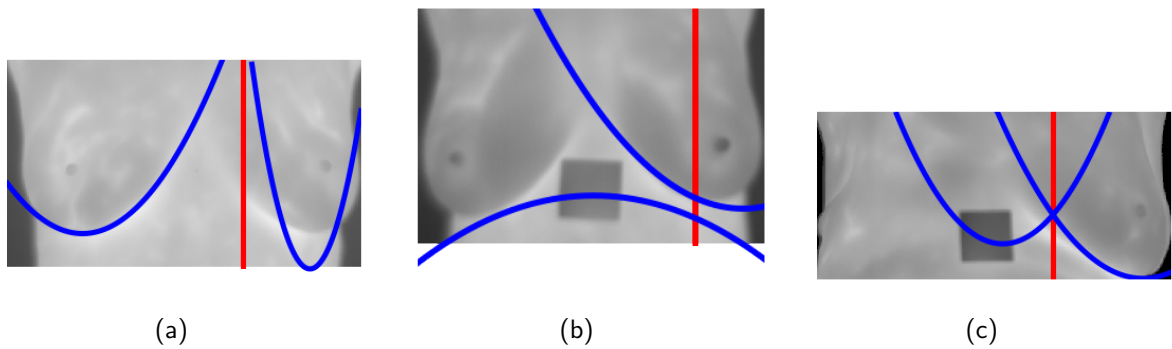


Figura 15 – Exemplos de segmentações que não foram bem sucedidas.

Fonte: Autoria própria.

Os problemas mais comuns encontrados, os quais inviabilizaram a utilização de parte da base, foram quatro: assimetria muito significativa entre mamas como na Figura 16(a); falta de contraste nas bordas do corpo devido a temperatura muito homogênea ou mamas muito pequenas, como na Figura 16(b); presença de marcadores proeminentes, como na Figura 16(c); e imagens com definição borrada, aparentando estar fora de foco, como na Figura 16(d).

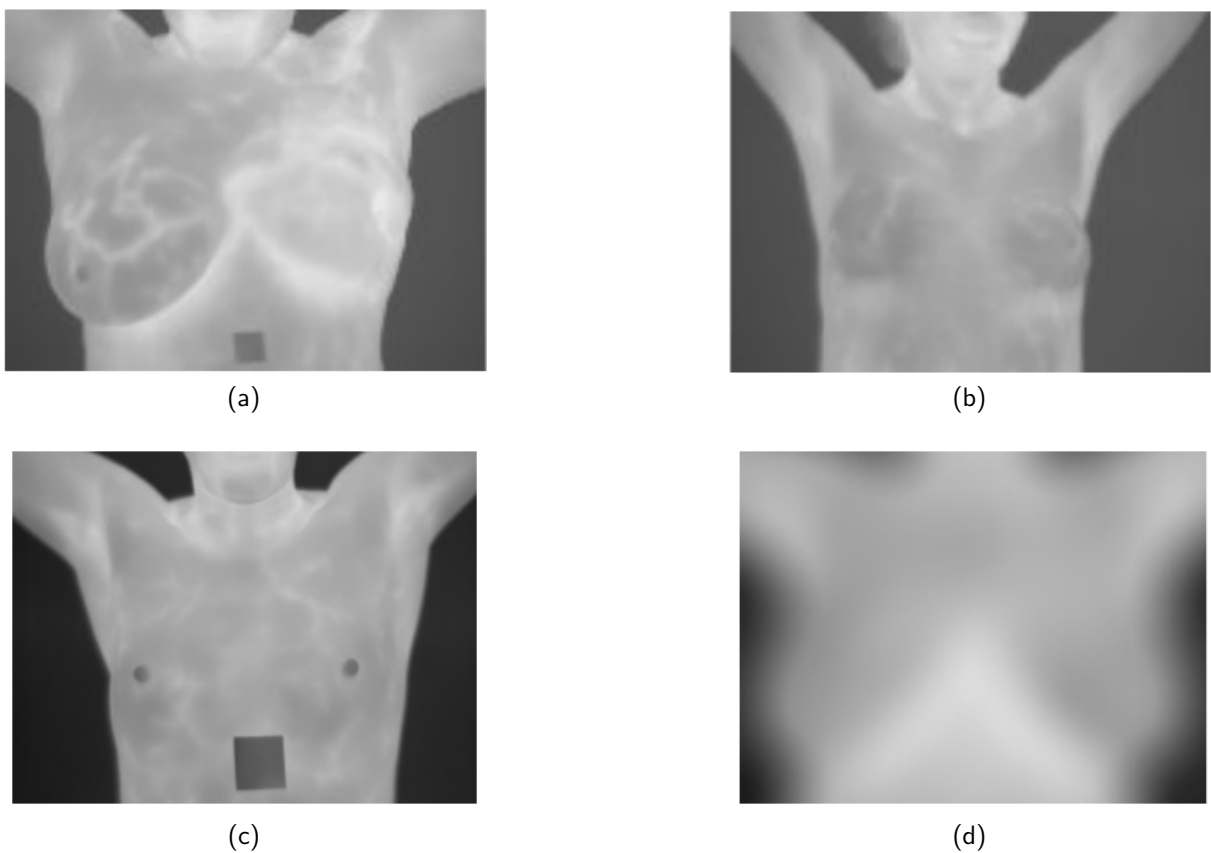


Figura 16 – Exemplos de imagens desafiadoras: (a) termograma assimétrico, (b) baixo contraste, (c) marcador muito proeminente e (d) definição borrada.

Fonte: Autoria própria.

Estes problemas não se encontram de maneira exclusiva nas imagens, ou seja, uma amostra pode ter assimetria e marcador proeminente, por exemplo. As imagens que apresentaram as segmentações mais irregulares apresentam um combinação deles.

Casos em que a discrepância entre a formas da mamas se mostrou grande demais foram desconsiderados. Em primeiro lugar, pelo fato da segmentação nestes casos ser mais complicada, dificultando a criação de um algoritmo automático e genérico. Em segundo lugar, pelo fato de que mamas muito assimétricas podem enviesar nossas características, que são baseadas em assimetria. Além disso, como alguns casos de assimetria presentes são criados por pacientes que passaram pelo processo cirúrgico de mastectomia (remoção do tecido mamário), não se pode checar a simetria de tecidos que já não são simétricos. Um exemplo de segmentação de um caso de assimetria pode ser visto na Figura 17.

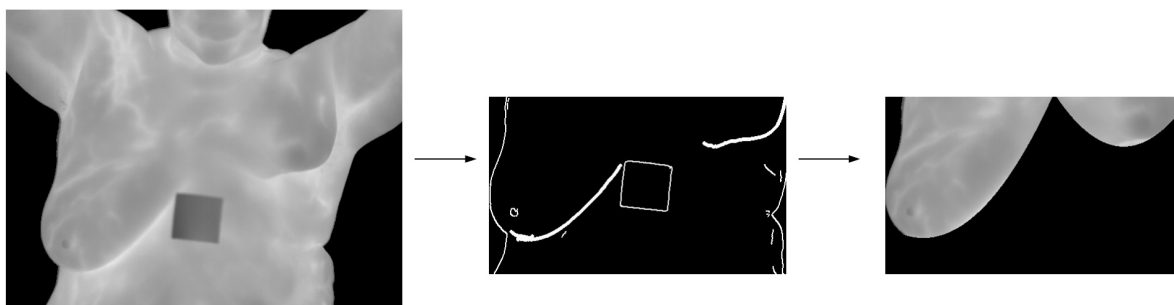


Figura 17 – Exemplo de segmentação assimétrica.

Fonte: Autoria própria.

Termogramas que possuem baixo contraste definindo as bordas do corpo e das mamas representam um problema por dificultarem a criação de um algoritmo genérico de segmentação, dado que necessitam de métodos ou parâmetros diferenciados. Este problema pode ocorrer quando as mamas de uma paciente são muito pequenas ou quando a temperatura do corpo é muito homogênea. A Figura 18 mostra um exemplo.

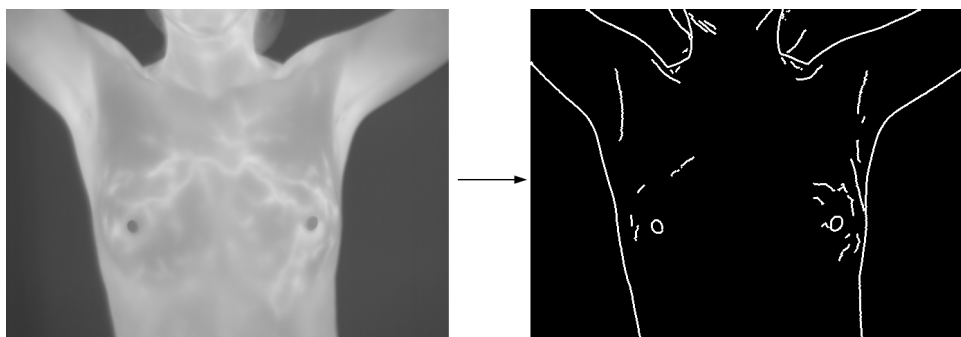


Figura 18 – Exemplo de extração de bordas de imagem com baixo contraste corporal.

Fonte: Autoria própria.

Como pode ser visto, as bordas definindo a região inframamária não foram detectadas, o que inviabiliza a aplicação dos métodos propostos neste trabalho. Para contornar este

problema, novas etapas de segmentação devem ser criadas, utilizando métodos diferentes para este tipo de imagem. Este tipo de imagem também pode causar inconsistências menores, como na Figura 15(a), que apresenta um corte incorreto da mama esquerda.

Outro problema de contraste encontrado foi o de que algumas imagens possuem contraste muito baixo entre a temperatura da paciente e o fundo. Isto causou um problema em que o fundo não pode ser segmentado no mesmo padrão das outras imagens. Como a ideia do trabalho é a de que o segmentador seja completamente automático, estes exemplos não foram utilizados nas etapas de classificação. Para contornar o problema, novas etapas no algoritmo de segmentação precisariam ser criadas.

Algumas das imagens da base possuem marcadores colados no corpo das pacientes, visando possibilitar uma noção de escala e posição na análise de imagens termográficas dinâmicas. Como o trabalho está analisando imagens estáticas, não existe a necessidade de manter a noção de escala e posição de amostras de uma mesma paciente. Portanto, no presente trabalho estes marcadores acabaram sendo um problema para os métodos propostos quando seu tamanho o fazia sobressair em relação às bordas das mamas. Neste caso, o algoritmo de segmentação acaba detectando as bordas do marcador como curvas correspondentes à região inframamária. Este problema é agravado em casos de imagens com baixo contraste corporal ou com definição borrada. Também existem casos que foram desconsiderados em que um marcador foi posicionado em cima da região de interesse. Um exemplo de segmentação incorreta pode ser visto na Figura 19 e também na Figura 15(b).

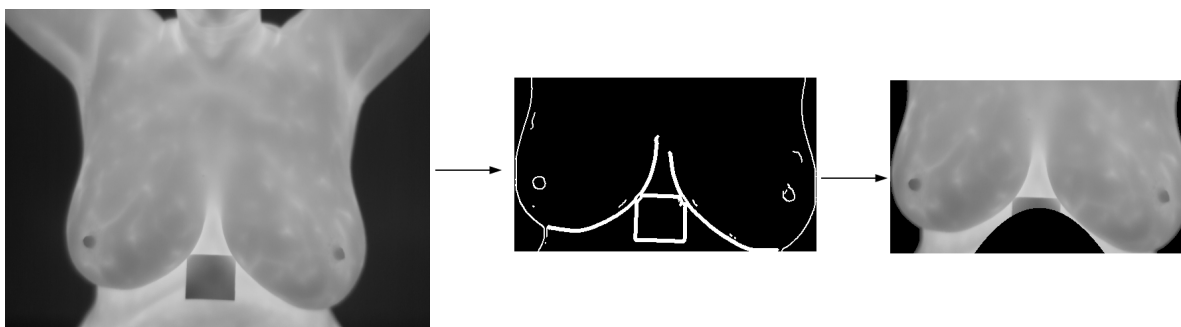


Figura 19 – Exemplo de segmentação com marcador problemático.

Fonte: Autoria própria.

Já os casos em que a imagem aparenta estar borrada também foram desconsiderados por serem mais difíceis de segmentar e por apresentarem valores muito discrepantes nas características extraídas. Alguns casos necessitariam de um tratamento especial antes de passar pelo detector de bordas, ou de parâmetros diferenciados neste. Outros não já não podem ser recuperados, pois estão borrados demais para se discernir qualquer tipo de borda. O exemplo da Figura 20 ilustra como a falta de definição causa instabilidade no detector de bordas, resultando em detecção de curvas e segmentações errôneas.

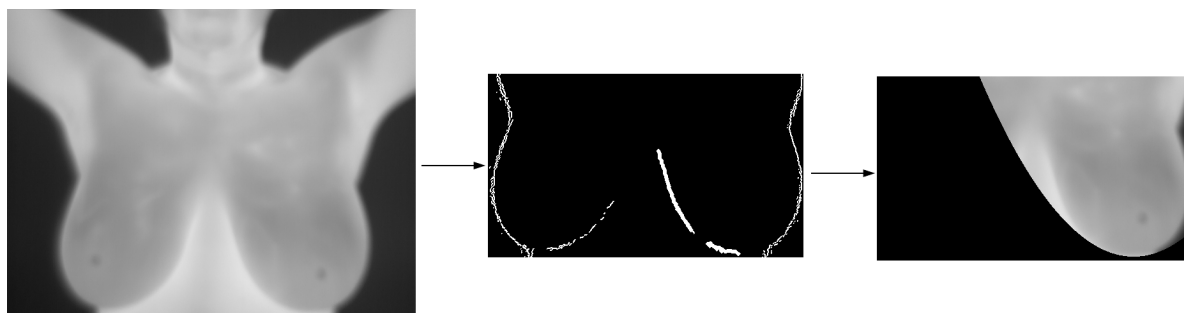


Figura 20 – Exemplo de segmentação de imagem borrada.

A Tabela 1 mostra a relação da porcentagem das imagens removidas em relação a todas as imagens coletadas, divididas entre as categorias citadas nesta seção.

Problema	Porcentagem de amostras afetadas
Borrado	4,92%
Contraste	20,49%
Assimetria	7,38%
Marcador	10,66%

Tabela 1 – Porcentagem de amostras removidas em relação a base completa.

4.2 Classificação

Na sequência são apresentados os resultados do processo de classificação considerando um descritor composto por medidas extraídas do histograma e de matriz de co-ocorrência.

Como descrito anteriormente, foram treinados cinco classificadores com parâmetros escolhidos utilizando a função `GridSearchCV` do Python, sendo um k-NN e quatro variações do SVM com diferentes *kernels*. Os testes foram aplicados em uma base contendo 22 imagens de pacientes portadores de câncer e 45 imagens de pacientes saudáveis, totalizando 67 amostras. Destas, 80% são reservadas para treinamento e 20% para teste.

4.2.1 Avaliação das características

Com base em [Sathish et al. \(2017\)](#), a seleção das características ideais para o treinamento foi realizada comparando-se a média e o desvio padrão dos valores entre as duas classes. A Tabela 2 mostra esta relação para cada uma das características propostas. As selecionadas para compor o vetor de características estão em negrito, e foram escolhidas com base na diferença direta entre os valores da classe de pacientes doentes e saudáveis. Toda característica que possui diferença na média ou no desvio padrão maior do que 0.04 foi selecionada.

Características	Saudável		Doentes	
	média	desvio padrão	média	desvio padrão
Média (histograma)	0,2547	0,2099	0,3501	0,2618
Variância (histograma)	0,3116	0,2401	0,2568	0,2262
Entropia (histograma)	0,4106	0,2544	0,4071	0,3028
Curtose (histograma) :	0,2497	0,1941	0,2734	0,2227
Obliquidade (histograma) :	0,3261	0,2384	0,2791	0,2457
Média (GLCM) :	0,255	0,2101	0,35	0,2617
Variância (GLCM) :	0,3195	0,2615	0,2931	0,2391
Entropia (GLCM) :	0,2827	0,2257	0,3104	0,2436
Energia (GLCM) :	0,3073	0,2261	0,3213	0,2577
Contraste (GLCM) :	0,2223	0,2337	0,2543	0,2443
Homogeneidade (GLCM)	0,2422	0,2141	0,3726	0,2815
Correlação (GLCM)	0,1029	0,1772	0,2276	0,2636

Tabela 2 – Características analisadas no desenvolvimento deste trabalho. Aquelas que estão em negrito foram escolhidas para compor o descritor.

As características selecionadas divergem das propostas em [Sathish et al. \(2017\)](#). Isto possivelmente se deve ao fato da base utilizada ser diferente, apesar de possuir a mesma origem. Vale notar também que a segmentação realizada neste trabalho não considera regiões que se encontram abaixo das curvas das mamas, removendo informações desnecessárias à extração de características, o que acaba resultando em informações mais precisas (mas que alteram as medidas calculadas).

4.2.2 SVM

A Tabela 3 mostra os resultados de classificação para o SVM, considerando os parâmetros descritos na Seção 3.4.

Os melhores resultados foram obtidos utilizando o *kernel* polinomial, assim como em [Sathish et al. \(2017\)](#), tendo 85,71% de acurácia, 100% de sensibilidade e uma especificidade de 78%. Ele foi escolhido por possuir a melhor combinação de sensibilidade e acurácia comparando-se aos demais, dando prioridade à sensibilidade (que é o valor que denota quantas pacientes com câncer foram classificadas corretamente).

Parâmetros do classificador				Avaliação			
kernel	nu	grau	gama	Acurácia	Sensibilidade	Especificidade	F1-Score
linear	0,4	-	-	78,57%	80,00%	78,00%	79,00%
poly	0,3	1	0,01584	85,71%	100,00%	78,00%	86,00%
rbf	0,2	-	1,00E-06	35,71%	100,00%	0,00%	19,00%
sigmoid	0,3	-	0,00063	92,86%	80,00%	100,00%	93,00%

Tabela 3 – Resultados da classificação utilizando características propostas neste trabalho

Outro resultado notável foi o do *kernel* sigmoide, que atingiu 92,86% de acurácia, 80%

de sensibilidade e uma especificidade de 100%, similar ao melhor resultado obtido em [Sathish et al. \(2017\)](#) que obteve 90% de acurácia, 87.5% de sensibilidade e 92.5% de especificidade.

O pior resultado foi obtido pelo *kernel* RBF, em que todos os pacientes do conjunto de teste foram classificados como doentes. Isso pode ser observado pelo fato dele possuir sensibilidade de 100% e especificidade de 0%.

A Tabela 4 contém os resultados de classificação utilizando as características propostas em [Sathish et al. \(2017\)](#) e a mesma base de imagens deste trabalho. Comparando ambas as tabelas, podemos observar que apesar dos resultados se mostrarem mais consistentes, houve uma notável queda nos valores de sensibilidade.

Parâmetros do classificador				Avaliação			
kernel	nu	grau	gama	Acurácia	Sensibilidade	Especificidade	F1-Score
linear	0,2	-	-	78,57%	60%	89,00%	78,00%
poly	0,4	3	2,51188	85,71%	80,00%	89,00%	86,00%
rbf	0,2	-	2,51E+00	85,71%	80,00%	89,00%	86,00%
sigmoid	0,5	-	0,0158489	85,71%	80,00%	89,00%	86,00%

Tabela 4 – Resultados obtidos com a seleção de características propostas por [Sathish et al. \(2017\)](#)

Testes realizados utilizando todas as características disponíveis obtiveram resultados piores e menos consistentes.

4.2.3 KNN

As Tabela 5 mostra o resultado da classificação utilizando k-NN. Ele se mostra inferior ao obtido pelo SVM em todos os quesitos. Testes também demonstraram maior inconsistência na classificação. Isso se dá provavelmente pelo fato da base de dados possuir poucos doentes em comparação à saudáveis.

Parâmetros do classificador				Avaliação			
n_neighbour	leaf size	weights	algoritm	acc	sens	esp	f-1
auto	30	1	uniform	71,43%	60%	78%	71%

Tabela 5 – Resultado do *GridSearch* para o KNN utilizando as características escolhidas neste trabalho

Resultados para todas as diferentes combinações de características citadas na Seção 4.2.2 foram exatamente iguais ao mostrado na Tabela 5. Testes com os tipos de algoritmo isolados também não resultaram em diferenças significativas e, portanto, não serão apresentados aqui.

5 CONSIDERAÇÕES FINAIS

A detecção de evidências da existência do câncer através da termografia tem sido explorada em diversos trabalhos na literatura. Como tumores possuem uma vascularização acima do normal, isso ocasiona uma alteração da temperatura em comparação àquela de regiões saudáveis, fazendo com que a assimetria entre os padrões de calor entre as mamas seja uma característica relevante.

Este trabalho teve como objetivo desenvolver uma abordagem para auxiliar no diagnóstico de câncer de mama, tendo-se como base o uso de técnicas de reconhecimento de padrões e imagens de termografia. Uma contribuição importante foi a identificação automática de regiões de interesse que futuramente possam tanto ser analisadas manualmente por um especialista quanto utilizadas para extração de características.

Tornar o algoritmo de geral geral o suficiente para se manter consistente em um grande conjunto diversificado de imagens se mostrou um grande desafio. Pequenos ajustes mostraram-se essenciais para ajustar a casos mais amplos e, assim, incluir mais imagens para os estudos de classificação.

O desbalanceamento entre o número de pacientes doentes e saudáveis também se mostrou problemático em alguns experimentos, principalmente para o k-NN.

Os resultados são promissores para uma segmentação realizada de maneira totalmente automatizada, ao contrário de (SATHISH et al., 2017), que realiza uma segmentação de forma manual em parte dos casos de teste.

Com uma sensibilidade de 100% e uma acurácia de 86% este é um exemplo isolado de como a termografia mamária, se expandida e melhorada, pode ser utilizada como pré-exame ou alternativa de exame para mulheres que não podem se expor a radiação ionizante ou não têm acesso aos sistemas de saúde que oferecem a mamografia.

5.1 TRABALHOS FUTUROS

Existem vários pontos na segmentação que podem ser trabalhados com mais tempo, como um tratamento mais robusto a interferência de marcadores na detecção das curvas do corpo, bem como a utilização do caráter dinâmico da base para uma segmentação que utilize múltiplas imagens.

A segmentação também pode ser aprimorada para identificar sub-regiões da mama, sendo que estudos apontam que 80% dos casos de câncer ocorrem na região superior externa da mama, região com mais tecido mamário (LEE, 2005). Poder comparar regiões exclusivamente pode levar a um aumento da sensibilidade, mas principalmente de especificidade da classificação.

Outro ponto a ser explorado é realizar a segmentação usando aprendizado baseado em redes neurais, como apresentado em Koay, Herry e Frize (2004).

Referências

ACHARYA, U. R. et al. Thermography based breast cancer detection using texture features and support vector machine. **Journal of Medical Systems**, v. 36, n. 3, p. 1503–1510, Jun 2012. ISSN 1573-689X. Disponível em: <<https://doi.org/10.1007/s10916-010-9611-z>>. Citado 3 vezes nas páginas 14, 15 e 16.

American Cancer Foundation. **Limitations of Mammograms**. 2018. Disponível em: <<https://www.cancer.org/cancer/breast-cancer/screening-tests-and-early-detection/mammograms/limitations-of-mammograms.html>>. Citado na página 8.

BISHOP, C. M. **Pattern Recognition and Machine Learning (Information Science and Statistics)**. Berlin, Heidelberg: Springer-Verlag, 2006. ISBN 0387310738. Citado 2 vezes nas páginas 11 e 22.

BORCHARTT, T. B. et al. Breast thermography from an image processing viewpoint: A survey. **Signal Processing**, v. 93, n. 10, p. 2785 – 2803, 2013. ISSN 0165-1684. Signal and Image Processing Techniques for Detection of Breast Diseases. Disponível em: <<http://www.sciencedirect.com/science/article/pii/S0165168412002794>>. Citado na página 10.

CANNY, J. A computational approach to edge detection. **IEEE Transactions on Pattern Analysis and Machine Intelligence**, PAMI-8, n. 6, p. 679–698, Nov 1986. ISSN 0162-8828. Citado na página 19.

DEY, N.; ASHOUR, A. S.; ALTHOUPETY, A. S. Thermal imaging in medical science. In: **Recent Advances in Applied Thermal Imaging for Industrial Applications**. [S.l.]: IGI Global, 2017. p. 87–117. Citado na página 10.

FOK, S. C.; NG, E. Y. K.; TAI, K. Early detection and visualization of breast tumor with thermogram and neural network. **Journal of Mechanics in Medicine and Biology**, v. 02, n. 02, p. 185–195, 2002. Disponível em: <<https://www.worldscientific.com/doi/abs/10.1142/S0219519402000344>>. Citado na página 10.

FREITAS WALSIR FAGANELO FIORI, F. J. d. F. R. E. G. R. M. S. R. J. G. d. O. R. Discomfort and pain during mammography. 2006. Citado na página 8.

FU, K.-S.; ROSENFELD, A. Pattern recognition and image processing. **IEEE Transactions on Computers**, C-25, n. 12, p. 1336–1346, Dec 1976. ISSN 0018-9340. Citado na página 11.

HARALICK, R. M.; SHANMUGAM, K.; DINSTEN, I. Textural features for image classification. **IEEE Transactions on Systems, Man, and Cybernetics**, SMC-3, n. 6, p. 610–621, Nov 1973. ISSN 0018-9472. Citado 2 vezes nas páginas 11 e 12.

HILDEBRANDT, C.; RASCHNER, C.; AMMER, K. An overview of recent application of medical infrared thermography in sports medicine in austria. **Sensors**, v. 10, n. 5, p. 4700–4715, 2010. ISSN 1424-8220. Disponível em: <<http://www.mdpi.com/1424-8220/10/5/4700>>. Citado na página 10.

Instituto Nacional de Câncer. **ESTIMATIVA 2018 Incidência de Câncer no Brasil Síntese de Resultados e Comentários**. 2018. Disponível em: <<http://www.inca.gov.br/estimativa/2018/sintese-de-resultados-comentarios.asp>>. Citado na página 8.

KAPOOR, P.; PRASAD, S. V. A. V. Image processing for early diagnosis of breast cancer using infrared images. In: **2010 The 2nd International Conference on Computer and Automation Engineering (ICCAE)**. [S.l.: s.n.], 2010. v. 3, p. 564–566. Citado 2 vezes nas páginas 14 e 16.

KOAY, J.; HERRY, C.; FRIZE, M. Analysis of breast thermography with an artificial neural network. In: **The 26th Annual International Conference of the IEEE Engineering in Medicine and Biology Society**. [S.l.: s.n.], 2004. v. 1, p. 1159–1162. Citado na página 33.

LAHIRI, B. et al. Medical applications of infrared thermography: A review. **Infrared Physics & Technology**, v. 55, n. 4, p. 221 – 235, 2012. ISSN 1350-4495. Disponível em: <<http://www.sciencedirect.com/science/article/pii/S1350449512000308>>. Citado na página 10.

LEE, A. Why is carcinoma of the breast more frequent in the upper outer quadrant? a case series based on needle core biopsy diagnoses. 2005. Disponível em: <<https://www.ncbi.nlm.nih.gov/pubmed/15767185>>. Citado na página 33.

PAVITHRA, P. et al. The effect of thermography on breast cancer detection. v. 9, p. 10–16, 01 2018. Citado na página 8.

PINHEIRO, A. B. et al. Câncer de mama em mulheres jovens: análise de 12.689 casos. **Revista brasileira de cancerologia**, v. 59, n. 3, p. 351 – 359, 2013. Disponível em: <http://www.inca.gov.br/rbc/n_59/v03/pdf/05-artigo-cancer-mama-mulheres-jovens-analise-casos.pdf>. Citado na página 8.

SATHISH, D. et al. Asymmetry analysis of breast thermograms using automated segmentation and texture features. **Signal, Image and Video Processing**, v. 11, n. 4, p. 745–752, May 2017. ISSN 1863-1711. Disponível em: <<https://doi.org/10.1007/s11760-016-1018-y>>. Citado 11 vezes nas páginas 8, 9, 15, 16, 21, 22, 23, 30, 31, 32 e 33.

SILVA, T. A. E. da. **Uma metodologia de auxílio ao diagnóstico de doenças de mama a partir de termografias dinâmicas**. Tese (Doutorado) — Universidade Federal Fluminense, 2015. Citado 2 vezes nas páginas 14 e 17.

TAMURA, H.; MORI, S.; YAMAWAKI, T. Textural features corresponding to visual perception. **IEEE Transactions on Systems, Man, and Cybernetics**, v. 8, n. 6, p. 460–473, June 1978. ISSN 0018-9472. Citado na página 11.

UEMATSU, S. Symmetry of skin temperature comparing one side of the body to the other. **Thermology**, v. 1, p. 4–7, 1985. Disponível em: <<https://ci.nii.ac.jp/naid/10026733414/en/>>. Citado na página 10.

VAPNIK, V. **Estimation of Dependences Based on Empirical Data: Springer Series in Statistics (Springer Series in Statistics)**. Berlin, Heidelberg: Springer-Verlag, 1982. ISBN 0387907335. Citado na página 13.

Visual Lab. **Processamento e Análise de Imagens Aplicadas à Mastologia**. 2018. Disponível em: <<http://visual.ic.uff.br/proeng/#>>. Citado 5 vezes nas páginas , 11, 16, 17 e 18.

WALKER, N. D.; KACZOR ND, F. T. Breast thermography: History, theory, and use. **Natural Medicine Journal**, v. 4, n. 7, p. 351 – 359, 2012. Citado na página 8.

YANG, Y.; LIU, X. A re-examination of text categorization methods. In: **Proceedings of the 22Nd Annual International ACM SIGIR Conference on Research and Development in Information Retrieval**. New York, NY, USA: ACM, 1999. (SIGIR '99), p. 42–49. ISBN 1-58113-096-1. Disponível em: <<http://doi.acm.org/10.1145/312624.312647>>. Citado na página 14.

ZIEGLER, C.-N. et al. Improving recommendation lists through topic diversification. In: **Proceedings of the 14th International Conference on World Wide Web**. New York, NY, USA: ACM, 2005. (WWW '05), p. 22–32. ISBN 1-59593-046-9. Disponível em: <<http://doi.acm.org/10.1145/1060745.1060754>>. Citado na página 25.

Apêndices

APÊNDICE A – Lista de artigos revisados

Paper	J. Koay, C. Herry, M. Frize, Analysis of breast thermography with an artificial neural network, Engineering in Medicine and Biology Society— IEMBS 1 (1) (2004) 1159–1162	E.Y.K. Ng, E.C. Kee, Integrative computer-aided diagnostic with breast thermogram, Journal of Mechanics in Medicine and Biology 7 (1) (2007) 1–10.	N. Arora, D. Martins, D. Ruggerio, E. Tousimis, A.J. Swistel, M.P. Osborne, R.M. Simmons, Effectiveness of a noninvasive digital infrared thermal imaging system in the detection of breast cancer, The American Journal of Surgery 196 (4) (2008) 523–526.
Tipo de aquisição	Unitária: a área foi exposta a um ventilador por 20 minutos	Unitária/estática: 20 minutos de exposição	Dinâmico: 4 minutos de exposição ao ar frio
temperatura da sala	22 °C	20° até 22° C	não informado
Recomendações de captura	For data collection, the patient was asked to avoid alcohol, caffeine, pain medication, lotions, and stop smoking two hours before the test	Reduction of heat fonts during acquisition. Patient recommendations of no use of alcohol, cigarettes, skin products	não informado
Posição dos braços	Erguidos	não informado	Erguidos
Número de capturas por paciente	não especificado. Provavelmente uma frontal	3 imagens:	100 frontal images with lateral mirrors
Base publica/n pacientes	não / 19 pacientes	No, 90 patients	No, 92 patient
equipamento de posicionamento	não	Não	Yes: sentinel BreastScan
Features extraídas	Mean, standard deviation, median, maximum, minimum, skewness, kurtosis, entropy, area and heat content	Mean, mode, points of temperature, median and biological data of patient	Proprietary System. Both assymetrie and local contrast
Técnicas de Segmentação	Canny to find edges, Manual adjustment in all images. Breasts were Divided in four quadrants automaticaly	não informado	não informado
Considerações		Linear regression and correlation	não informado
evaluation	Artificial neural network (ANN) with back propagation	ANN, RNFN	Three way: screening, clinical and ANN
Resultado	não informado	não informado	96.7 sensitivity, 26.5 Specificity

Paper	V. Agostini, M. Knaflitz, F. Molinari, Motion artifact reduction in breast dynamic infrared imaging, IEEE Transaction on Biomedical Engineering 56 (3) (2009) 903–906.	U.R. Acharya, E.Y.K. Ng, J.H. Tan, S.V. Sree, Thermography based breast cancer detection using texture features and support vector machine, Journal of Medical Systems (2010) 01–08	F.G. Delgado, J.G.V. Luna, Feasibility of new-generation infrared screening for breast cancer in rural communities, US Obstetrics and Gynecology, Touch Briefings 5 (2010) 52–56
Tipo de aquisição	Dinâmico: 200 frames	Unitario/estatico:15 minutos para estabilizar	Unitario/estatico:15 minutos para estabilizar
temperatura da sala	não informado	20° to 22° C, e umidade 60%, 75%	18° até 23° C
Recomendações de captura	Placement of 12 markers in the patient skin for images stabilization and motion artifacts reduction. Keep arms up	Patients required to rest for 15 minutes and to wear a loose gown, Ensure that patients are in the period between 5th to 12th and 21st day after menstrual cycle.	
Posição dos braços	Erguidos	Erguidos	Erguidos
Número de capturas por paciente		frontal. Provavelmente 1	3 imagens: 1 frontal, 2 laterais oblíquas
Base publica/n pacientes	Not cited in the paper	No, 50 patient	Não mencionado
equipamento de posicionamento	Yes: examination table with the backrest inclination of 40 degrees. 2.2m distance and 1.7 m height		Não
Features extraídas	não informado	Homogeneity, energy, entropy, 4 first moments, entropy, angular second moment, contrast, mean, short and long runs emphasis, run percentage, grey level and run length non-uniformity	não informado
Técnicas de Segmentação	não informado	não informado	não informado
Considerações	Esse paper não tem intenção de classificar as imagens obtidas, mas sim estabelecer um padrão para maior qualidade de captura de vídeo	Three-fold stratified cross validation was used to test the SVM classifier. 36 images for training (18 normal and 18 malignant)	não informado
evaluation	não informado	Support vector machine (SVM)	não informado
Resultado	não informado	85.71 sensitivity, 90.48 Secificity 81.07 Positive Predictive Acc best results	não informado

Paper	P. Kapoor and S.V.A.V. Prasad, (2010) Image processing for early diagnosis of breast cancer using infrared images, 2nd International Conference on Computer and Automation Engineering, vol. 3, 1, pp. 564–566.	A. Amri, A. Saidane, S. Pulko, Thermal analysis of a threedimensional breast model with embedded tumour using the transmission line matrix (TLM) method, Computers in Biology and Medicine 41 (1) (2011) 76–86.	M. Kontos, R. Wilson, I. Fentiman, Digital infrared thermal imaging (DITI) of breast lesions: sensitivity and specificity of detection of primary breast cancers, Clinical Radiology 66 (2011) (2011) 536–539.
Tipo de aquisição	Unitario: 10 minutos de exposição ao ar frio	Unitario/estatico: 10 a 15 minutos para estabilizar	Unitario/estatico: 15 minutos para estabilizar
temperatura da sala	22° C	18° to 22° C	22° C
Recomendações de captura	Capture in a dark room for minimum interference	Patient recommendations of no use of alcohol consumption, physical exercises, skin products	The examination room had no windows and the vents were muffled, pointing away from the patient. The ceiling lights were fluorescent and invariably kept well away from the scanned area.
Posição dos braços	Erguidos	não informado	não informado
Número de capturas por paciente	3 images: 1 frontal, 2 laterais oblíquas	não informado	3 images: 1 frontal, 2 lateral
Base publica/n pacientes	não informado	não informado	No, 63 patients
equipamento de posicionamento	não informado	não informado	Yes: backless chair with rotation. 100 cm distance
Features extraídas	Skewness, temperature variation, kurtosis, Cumulative histogram. The asymmetry can be measured by the maximum ratio between temperature difference and area.	não informado	não informado
Técnicas de Segmentação	Canny edge detector, Hough Transform to find lower boundaries	não informado	não informado
Considerações	This paper proposes an approach to diagnose breast cancer based on infrared images, that's all.	Utilização de modelagem 3d	based on the colour gradient between adjacent areas and differences between the two breasts. Images were reported by a radiology consultant with extensive experience in breast imaging (R. W.) and by a senior surgical research fellow with experience in the use and interpretation of the technique (M.K.). The images were double-read with consensus reached on image classification
evaluation	Clusterization	não informado	Clinical
Resultado	não informado	não informado	25 Sensitivity, 85 Specificity 24 Positive Predictive Value and 86 Negative

Paper	G.C. Wishart, M. Campisid, M. Boswella, D. Chapman, V. Shackleton, S. Iddlesa, A. Halletta, P. D. Britton, The accuracy of digital infrared imaging for breast cancer detection in women undergoing breast biopsy, European Journal of Surgical Oncology 36 (6) (2010) 535–540.	American College of Clinical Thermology (ACCT) Breast Screening Procedure, Available in: /http://www.thermologyonline.org/Breast/breast_thermography_procedure.htmS, accessed in 31 January 2012.	Qi, Hairong & Snyder, Wesley & F. Head, Jonathan & L. Elliott, Robert. (2000). Detecting breast cancer from infrared images by asymmetry analysis. 2. 1227 - 1228 vol.2. 10.1109/EMBS.2000.897952	T.B. Borchardt, R. Resmini, A. Conci, A. Martins, A.C. Silva, E.M. Diniz, A. Paiva and R.C.F. Lima, (2011) Thermal feature analysis to aid on breast disease diagnosis, Proceedings of 21st Brazilian Congress of Mechanical Engineering—COBEM2011, 24–28 October 2011, Natal, Brazil, pp. 1–8
Tipo de aquisição	Dinâmico: 5 min com ar gelado apontado para a área	estatico	estática	visão frontal
temperatura da sala	não informado	não informado	nãp especificado	não informado
Recomendações de captura	Arms supported at eye level in a chair	não informado	não informado	não informado
Posição dos braços	Erguidos	não informado	Erguidos	Afastados do corpo a ~45°
Número de capturas por paciente	250 images per patient	não informado	não informado	Provavelmente 1, porém não é explicitado. Imagens frontais
Base publica/n pacientes	No, 100 patients	não informado	no, não especificado	não informado
equipamento de posicionamento	Yes: Sentinel BreastScan	não informado	não informado	não informado
Features extraídas	None	não informado	Supervisionado: Mean, variance, skewness and kurtosis. The peak pixel intensity of the correlated image. Entropy and Joint Entropy.	Range of temperature, mean, standard deviation and the last bin of a quantization of ten bins
Técnicas de Segmentação	None	não informado	Edge Detection(Canny) -> Four Feature Cuver(Hough) -> Segments	Manual
Considerações	The computer analysed the raw data using 'artificial intelligence' and produced a "Patient Report" he images were also reviewed by an independent thermography expert. these same images were interpreted using a novel artificial intelligence programme called NoTouch BreastScan	não informado	When the images are relatively symmetrical, small asymmetries may indicate a suspicious region	não informado
evaluation	Four way: screening, clinical and two ANN	não informado	k-means and c-means	SVM
Resultado	Screening Sensitivity: 53 Ann Sensitivity: 48 Clinical Sensitivity: 73	não informado	não informado	95.83 Sensitivity, 25 Specificity 85.71 Accuracy

Paper	J.G. Schaefer, M. Zavissek, T. Nakashima, Thermography based breast cancer analysis using statistical features and fuzzy classification, Pattern Recognition 42 (6) (2009) 1133–1137.	M.C. Araújo, R.C.F. Lima, F. Santos, Desenvolvimento de um banco de dados como ferramenta auxiliar na detecção precoce de câncer de mama 301 Iberian-Latin-American Congress on Computational Methods in Engineering, 1, Armacao dos Buzios, RJ, Brazil, 2009	Sathish, D., Kamath, S., Prasad, K. et al. SIVIP (2017) 11: 745. https://doi.org/10.1007/s11760-016-1018-y
Tipo de aquisição	visão frontal	Método Estático	visão frontal
temperatura da sala	não informado	não informado	não informado
Recomendações de captura	não informado	cada paciente é submetida a um período de aclimação à sala de exame, após o qual são tomadas as imagens.	smoothed using a Gaussian filter
Posição dos braços	não informado	não informado	não informado
Número de capturas por paciente		1 imagem frontal por paciente	frontal view static breast thermogram temperature matrices 1/2 1/2
Base publica/n pacientes	não. dataset of 146 thermograms (29 malignant and 117 benign cases)	Sim.	sim (http://visual.ic.uff.br/en/proeng/#), Brazil. 80
equipamento de posicionamento	não informado	não informado	não informado
Features extraídas	Basic statistical features, moments, histogram features, cross co-occurrence matrix, mutual information, and Fourier analysis	não informado	Histogram and grey level cooccurrence matrix-based texture features are extracted from the segmented images
Técnicas de Segmentação	Manual	não informado	Canny edge detection. y morphological dilation operation. . HPP approach is used for the detection of the inframammary line
Considerações	symmetry typically identifies healthy subjects	Artigo se ocupa primariamente com a estrutura e regras do banco de dados em si. Não trata muito sobre a aquisição. (TODO Descobrir como poder adquirir permissão de acesso)	a fully automatic segmentation of right and left breast for asymmetry analysis
evaluation	Fuzzy rule-based classification system	não informado	SVM RBF classifier
Resultado	classification accuracy of about 80%	não informado	n accuracy of 90%, sensitivity of 87.5% and specificity of 92.5%