UNIVERSIDADE FEDERAL FLUMINENSE

LINCOLN FARIA DA SILVA

Uma Análise Híbrida para Detecção de Anomalias da Mama usando Séries Temporais de Temperatura

NITERÓI

2015

UNIVERSIDADE FEDERAL FLUMINENSE

LINCOLN FARIA DA SILVA

Uma Análise Híbrida para Detecção de Anomalias da Mama usando Séries Temporais de Temperatura

Tese de Doutorado apresentada ao Programa de Pós-Graduação em Computação da Universidade Federal Fluminense como requisito parcial para a obtenção do Grau de Doutor em Computação. Área de concentração: Computação Visual e Interfaces

Orientadoras:
Profª. D.Sc. Aura Conci e Profª. D.Sc. Débora Christina Muchaluat Saade

NITERÓI

2015

LINCOLN FARIA DA SILVA

UMA ANÁLISE HÍBRIDA PARA DETECÇÃO DE ANOMALIAS DA MAMA USANDO SÉRIES TEMPORAIS DE TEMPERATURA

Tese de Doutorado apresentada ao Programa de Pós-Graduação em Computação da Universidade Federal Fluminense como requisito parcial para a obtenção do Grau de Doutor em Computação. Área de concentração: Computação Visual e Interfaces

Aprovada em 13 de agosto de 2015.

BANCA EXAMINADORA

Profª. D.Sc. Aura Conci (Presidente), UFF

Profª. D.Sc. Débora Christina Muchaluat Saade, UFF

Profª. D.Sc. Bianca Zadrozny, IBM Research Brazil

Prof. D.Sc. Aristófanes Corrêa Silva, UFMA

Prof. D.Sc. Artur Ziviani, LNCC

Prof. Dr.M. Renato de Souza Bravo, UFF

Prof. D.Sc. Célio Vinicius Neves de Albuquerque, UFF

Prof. D.Sc. Leandro Augusto Frata Fernandes, UFF

Niterói

2015

Ao Deus soberano, criador de todo o universo. Esteve sempre ao meu lado, deu-me capacidade e sabedoria necessária para o desenvolvimento deste trabalho. Tudo o que tenho e tudo o que sou devo a Ele. E à minha amada esposa, pois, ainda na graduação, sempre apoiou-me, sempre acreditou em mim, esteve sempre ao meu lado.

Agradecimentos

À Profª. D.Sc. Aura Conci e à Profª. D.Sc. Débora Christina Muchaluat Saade, que confiaram em mim e compartilharam comigo suas experiências e conhecimentos, profissional e acadêmico, na tarefa de orientar-me. Tenho aprendido muito com elas.

Quero agradecer também à minha mãe, Maria das Graças Faria da Silva, e à minha avó, Vitória Ghiotti Faria (em memória), que me educaram, superando situações adversas.

Aos médicos e médicas do Hospital Universitário Antônio Pedro (HUAP), Dr. Alair Augusto Sarmet M. D. dos Santos, Dr. Renato Bravo, Drª Cristina Asvolinsque Pantaleão Fontes, Dr. Amandio Roberto Pereira Henrique, Drª Maria Lúcia de Oliveira Santos e Dr. Alberto Domingues Vianna, que muito contribuíram não só para o desenvolvimento desta tese, mas também para o projeto desenvolvido no Visual Lab de termografia da mama.

Às secretárias da pós-graduação do Instituto de Computação, Teresa Cristina da Silva Cancela e Viviane Aceti, e à funcionária administrativa do Laboratório MídiaCom, Marister Monteiro Luz do Outão, que contribuíram para o projeto, principalmente no início para os testes em voluntárias do protocolo de aquisição das imagens infravermelha.

À secretária do serviço de radiologia, Marilza Lucia Teixeira Albino, ao assistente em administração do Departamento de radiologia, Bryan Marinho Hall, ao funcionário de manutenção, Euzébio Domingues de Souza, pelo apoio prestado no HUAP.

Aos amigos de laboratório, Roger Resmini, Tiago Bonini Borchartt, Giomar Oliver Sequeiros Olivera, Rafael de Souza Marques, João Paulo Scoralick Oliveira, Cassia Isac, Érick Oliveira, Stephenson Galvão, Franciéric Alves, Thiago Alves Elias da Silva, Edgar Moraes Diniz, pela troca de experiências e pelo trabalho em equipe.

Aos alunos de iniciação científica, Gilda de Souza Carvalho, Jéssica Bastos Silva, Pedro Moisés Damasceno, Patrick Barreto dos Santos, Renan Vinagre Câmara, Pollyanna Abreu Mendoza, Jean Nogueira dos Santos Fontes, Larissa Martins dos Santos Blanco, Bruna Salles Reis, Luma Carneiro Soares, Daniela Motta Bello, Juliana Fernandes Ledo, Gustavo Domingos Rodrigues, Amanda de Almeida Souza, Lorena Teixeira da Costa,

Rayane Climaco dos Santos, Roberta Sá e Luciando da Silva Nascimento, pelas atividades executadas no projeto e consequente contribuição para esta tese.

Um agradecimento especial à Fundação de Amparo à Pesquisa e ao Desenvolvimento Científico e Tecnológico do Maranhão (projeto: Visualização, detecção e diagnóstico precoce de câncer de mama baseado em análise de imagens médicas multimodais, coordenado pelo Prof. D.Sc. Aristófanes Corrêa Silva), ao Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) (projeto: Imagens Médicas: Processamento, Análise e Visualização, coordenado pelo Prof. D.Sc. Aristófanes Corrêa Silva), pelo auxílio recebido.

Ao Instituto Nacional de Tecnologia em Medicina Assistida por Computação Científica (INCT-MACC), pelas bolsas cedidas aos alunos de iniciação científica e demais auxílios, e à Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES), pela bolsa de doutorado que recebi durante todo o curso.

À Profª. D.Sc. Rita de Cassia Fernandes de Lima e aos membros de seu grupo de pesquisa, Marcus Costa de Araújo, Ladjane Coelho dos Santos e Mariana Jorge de Andrade Viana, pela cooperação, em especial no início do desenvolvimento desta linha de pesquisa e pelo uso dos dados adquiridos pelo grupo no Hospital das Clínicas da Universidade Federal de Pernambuco (UFPE).

E, por fim, quero agradecer a todos os familiares, amigos, professores e funcionários do Instituto de Computação da Universidade Federal Fluminense (UFF) e do HUAP que, direta ou indiretamente, ajudaram-me e apoiaram-me, em especial ao motorista da Escola de Engenharia da UFF, Jorge Francisco da Cruz, pelo transporte dos equipamentos necessários para a aquisição das imagens no hospital, à minha tia Glacy e ao meu tio Iltom pela ajuda em um momento especial da minha vida.

Resumo

O câncer de mama é o segundo tipo de câncer mais comum no mundo. Mas o diagnóstico e o tratamento em estágios iniciais aumentam as chances de cura da paciente. A temperatura de tecidos cancerosos é geralmente mais alta do que a de tecidos vizinhos saudáveis, tornando a termografia uma opção a ser considerada em estratégias de rastreamento deste tipo de câncer. Esta tese propõe uma metodologia híbrida de análise da Termografia Infravermelha Dinâmica com o objetivo de detectar anomalias da mama, entre elas o câncer, utilizando técnicas de aprendizagem de máquina não supervisionada e supervisionada, o que caracteriza a metodologia como híbrida. Para alcançar esse objetivo, um protocolo de execução Termografia Infravermelha Dinâmica foi estabelecido. A sequência de termogramas capturada de cada paciente executando o protocolo estabelecido é processada e analisada por várias técnicas. Primeiramente, a região das mamas é segmentada e os termogramas da sequência são registrados. Então, séries temporais de temperatura são construídas e o algoritmo *k-means* é aplicado sobre essas séries usando vários valores de k. Índices de validação de agrupamento são aplicados para avaliar os grupos formados para cada valor de k, gerando valores tratados como características na etapa de construção do modelo de classificação. Na fase de avaliação da metodologia, ferramentas de mineração de dados são usadas para resolver o problema de seleção de algoritmos e otimização de parâmetros em tarefas de classificação, mais conhecido como CASH (*Combined Algorithm Selection and Hyperparameter Optimization Problem*). Além dos algoritmos de classificação recomendados pelas ferramentas de mineração de dados, classificadores baseados em *redes Bayesianas*, *redes neurais*, *máquina de vetores de suporte* e *árvore de decisão*, foram testados na avaliação. Os resultados dos testes indicam que a metodologia proposta é capaz de detectar anomalias de mama e contribuir para que o exame termográfico seja adotado em programas de rastreamento organizado de câncer de mama. Dentre os 39 algoritmos de classificação testados, *K-Star* e *Bayes Net* apresentaram acurácia de 100%. Além disso, foi obtida uma acurácia média de 95,71% entre os algoritmos de classificação: *Bayes Net*, *Multi-Layer Perceptron*, *LibSVM* e *J48*.

Palavras-chave: Câncer de mama, Termografia Infravermelha Dinâmica, detecção precoce, aprendizagem de máquina, processamento de imagem.

Abstract

Breast cancer is the second most common type of cancer worldwide. But, diagnosis and treatment in early stages increase the chances of cure of the patient. The temperature of cancerous tissue is generally higher than that of healthy surrounding tissues, making thermography an option to be considered in screening strategies of this cancer type. This thesis proposes a hybrid methodology for analyzing Infrared Thermography Dynamic in order to detect breast abnormalities, including cancer, using unsupervised and supervised machine learning techniques, which characterizes the methodology as hybrid. To achieve this goal, a Infrared Thermography Dynamic execution protocol was established. The captured thermograms sequence from each patient by performing the established protocol is processed and analyzed by several techniques. First, the region of the breasts is segmented and the thermograms of the sequence are registered. Then, temperature time series are built and the *k-means* algorithm is applied on these series using various values of k. Indices of clustering validation are applied to evaluate the formed groups for each value of k, generating values treated as features in the classification model construction step. In the evaluation phase of the proposed methodology, data mining tools were used to solve the *combined algorithm selection and hyperparameter optimization (CASH) problem*. Beyond the classification algorithms recommended by the data mining tools, classifiers based on *Bayesian networks*, *neural networks*, *support vector machine* and *decision tree* were tested on the dataset used for evaluation. Test results support that the proposed analysis methodology is able to detect breast anomalies and help insert the thermography in clinical routines for screening of breast diseases. Among 39 tested classification algorithms, *K-Star* and *Bayes Net* obtained classification accuracy of 100%. Furthermore, among the *Bayes Net*, *Multi-Layer Perceptron*, *LibSVM* e *J48* classification algorithms, an average accuracy of 95.71% was obtained.

Keywords: Breast cancer, Dynamic Infrared Thermography, early detection, machine learning, image processing.

Lista de Abreviaturas e Siglas

ACCT	:	American College of Clinical Thermology;
AD	:	Agrupamento de Dados;
AG	:	Algoritmo Genético;
CAAE	:	Certificado de Apresentação para Apreciação Ética;
CASH	:	Combined Algorithm Selection and Hyperparameter Optimization Problem;
CEP	:	Comitê de Ética em Pesquisa;
DCBD	:	Descoberta de Conhecimento em Banco de Dados;
DMR-IR	:	Database for Mastology Research with Infrared Image;
ECADeG	:	Enabling Collaborative Applications for Desktop Grids;
EMiner	:	ECADeG Miner;
HUAP	:	Hospital Universitário Antônio Pedro;
IM	:	Informação Mútua;
INCA	:	Instituto Nacional de Câncer;
KDD	:	Knowledge Discovery in Databases;
MD	:	Mineração de Dados;
MMTW	:	Módulo Máximo da Transformada de Wavelet
NCI	:	National Cancer Institute
OMS	:	Organização Mundial de Saúde;
ON	:	Óxido Nítrico;
ROC	:	Receiver Operating Characteristic;
ROI	:	Region of Interest;
SVM	:	Support Vector Machine;
TI	:	Termografia Infravermelha;
TID	:	Termografia Infravermelha Dinâmica;
WEKA	:	Waikato Environment for Knowledge Analysis;

Lista de Figuras

Lista de Tabelas

Sumário

Capítulo 1

Introdução

O câncer de mama é o segundo tipo de câncer mais comum, sendo o tipo que mais vitima mulheres no mundo. A estimativa do INCA (Instituto Nacional de Câncer) para os anos de 2014 e 2015 é de 75 mil casos novos da doença [Facina 2014]. Nos Estados Unidos, de acordo com o *National Cancer Institute (NCI)*, o número total de casos novos de câncer de mama, entre mulheres americanas, passará de 283 mil, em 2011, para 441 mil em 2030, um aumento acima de 60%, em 19 anos [Grant 2015]. As taxas de mortalidade continuam elevadas muito provavelmente devido ao diagnóstico da doença em estágios avançados. Entretanto, quando diagnosticado e tratado em estágios iniciais, esse tipo de câncer apresenta um prognóstico relativamente bom [da Silva 2014]. Dessa forma, a detecção precoce da doença torna-se um fator importante, pois o sucesso do tratamento do câncer de mama é inversamente proporcional ao seu tamanho e alastramento na ocasião do diagnóstico.

A detecção precoce do câncer de mama possui dois componentes: o diagnóstico precoce e o rastreamento. No diagnóstico precoce os primeiros sinais e sintomas do câncer são reconhecidos pelo paciente ou pelo profissional de saúde antes do avanço da doença. Isso permite maior eficácia e menor custo do tratamento. Por outro lado, o rastreamento é a realização de exames de triagem em indivíduos assintomáticos com o objetivo de identificar aqueles com alterações sugestivas de câncer de mama, antes de qualquer sintoma e/ou sinal dessa doença. Os indivíduos com tais alterações tem o diagnóstico de câncer confirmado ou descartado, após uma investigação mais detalhada. O rastreamento é dividido em dois tipos: o oportunístico e o organizado. No rastreamento oportunístico exames são solicitados de forma não sistemática em consultas de rotina. De forma diferente acontece no rastreamento organizado, onde exames são solicitados de forma sistemática para uma população de risco, a população-alvo, dentro de um programa es-

truturado [Stein et al. 2009]. No Brasil, programas de rastreamento organizado para o câncer de mama é uma discussão relativamente recente, ao contrário de países como Canadá, Estados Unidos e da União Europeia. Nesse tipo de rastreamento, a busca pela população-alvo é ativa [Silva e Hortale 2012].

Um dos exames utilizado no rastreamento é a mamografia. Considerado como o padrão ouro, a mamografia é utilizada na prevenção secundária, rastreando lesões não palpáveis da mama e, assim, contribuindo com uma redução de aproximadamente 30% da taxa de mortalidade [de Jesus 2005]. A mamografia mostra tumores ainda em estágios iniciais ou suficientemente pequenos para serem percebidos por um médico. Usa doses de radiação ionizante para formar imagens da mama com a finalidade de detectar massas anormais. Os especialistas reconhecem a importância do exame mamográfico no rastreamento do câncer, mas evitam solicita-lo principalmente pela escassez de mamógrafo. Apesar da oferta de mamógrafos pelo SUS (Sistema Único de Saúde) ser satisfatória, a distribuição desses aparelhos pelo país é desigual, a maioria encontra-se nas regiões Sul e Sudeste, e desses, grande parte está nas capitais. Em 2013, eram esperadas pelo INCA (Instituto Nacional do Câncer) 10 milhões de mamografias em mulheres com idade entre 50 e 60 anos, mas apenas 2,5 milhões foram realizadas, índice muito abaixo pelo recomendado pela Organização Mundial de Saúde (OMS), de 7 milhões [de Mastologia 2015]. Assim, o diagnóstico do câncer de mama é retardado pela cobertura insuficiente e a dificuldade no controle e na avaliados dos serviços mamográficos disponíveis pelo SUS [de Castro Mattos et al. 2013]. Além disso, a mamografia obtém as imagens pela radiação das mamas [Thuler 2003] [Gerasimova et al. 2013], cada vez que a mama é exposta a raios-X, o risco de câncer aumenta em 2% e a mama pré-menopausa é ainda mais sensível à radiação [Arabi et al. 2010].

1.1 Motivação

Mediante a dificuldade de acesso ao exame por imagem mais básico de rastreamento do câncer de mama, a mamografia, principalmente pelas mulheres de baixa escolaridade e classe socioeconômica, torna-se necessário definir precisamente a população-alvo para um programa de rastreamento organizado. Nesse sentido, a termografia infravermelha da mama apresenta-se como um exame que pode indicar as mulheres que comporão essa população-alvo. A termografia infravermelha da mama é um exame que detecta e registra a radiação infravermelha emitida pela superfície da mama e produz um termograma. Ela não utiliza radiação ionizante, acesso venoso, ou outro processo invasivo, portanto,

o exame não apresenta dano ou risco algum à paciente. Classificado como um exame funcional, a termografia infravermelha provê informações fisiológicas de funcionamento normal ou anormal dos sistemas vascular, sensorial e nervoso simpático, bem como de processos inflamatórios [Amalu et al. 2008] [Head e Elliott 2002], além de apresentar um custo extremamente baixo quando comparado aos demais exames. A Tabela 1.1 contém a comparação entre a termografia e outros exames, utilizados no rastreamento e na detecção do câncer de mama, em relação à algumas características e entre elas o custo. Dessa forma, a termografia infravermelha da mama pode ser realizada quantas vezes for necessária tornando-se uma aliada no uso racional do exame mamográfico.

Tabela 1.1: Comparação entre a termografia e outros exames [Resmini 2011].

Exame	Termografia	Mamografia	Ultrassonografia	Ressonância magnética
Tipo	Funcional	Estrutural	Estrutural e funcional	Estrutural e funcional
Invasivo	não	muito	muito pouco	pouco
Funcionamento	Detecta a radiação infravermelha de cada ponto da cena	Emite radiação ionizante	Emite e captura ondas de ultrassom	Gera um campo magnético que atua nos elétrons H^+
Desconforto	Mama desnuda	compressão da mama, dor	mama desnuda com gel	aplicação de contraste (se funcional)
Contra indicação	Não existe	Não indicado para mama densa	Muito dependente do operador	Se for alérgico ao contraste ou fóbico
Tipo de imagem	Registra um mapa da temperatura de cada ponto da mama	Destaca as estruturas e calcificações internas da mama	Destaca estruturas e fluxos de líquidos	Destaca estruturas e fluxos de sangue
Custo	x	$100x$	$100x$	$800x$

Visto que a temperatura de tecidos cancerosos é geralmente maior do que a de tecidos saudáveis, a termografia tem sido considerada um método de rastreamento promissor para a detecção do câncer de mama, por gerar imagens que revelam a distribuição de temperatura sobre a superfície de ambas as mamas [Gerasimova et al. 2013] [Borchartt et al. 2012]. A termografia infravermelha de mama é dividida basicamente em duas modalidades: a Termografia Infravermelha Estática (TI) e a Termografia Infravermelha Dinâmica (TID).

Nesse sentido, nosso grupo vem desenvolvendo várias pesquisas. Muitas dessas já foram defendidas na forma de trabalho final de curso e outras ainda estão em desenvolvimento. Em 2010, em sua dissertação de mestrado, Serrano [Serrano 2010] propôs o uso do coeficiente de Hurst e da Lacunaridade para, baseado na assimetria da distribuição das temperaturas das mamas, auxiliar no diagnóstico de doenças da mama, utilizando termografias obtidas por TI. Entretanto, no trabalho de Serrano a segmentação de ambas as mamas foi realizada manualmente. Assim, Motta [Motta 2010], em sua dissertação de mestrado, propôs uma segmentação automática de ambas as mamas nos termogramas obtidos na posição frontal (paciente de frente para a câmera infravermelha) e executando

a TI. Resmini [Resmini 2011], assim como Serrano, desenvolveu, em sua dissertação de mestrado, uma metodologia para auxiliar o diagnóstico de doenças da mama utilizando termogramas obtidos por TI, porém aplicando descritores de textura de imagens e utilizando o segmentador de Motta como parte da segmentação das mamas, nas imagens. Em 2012, Marques [Marques 2012], com o objetivo de aprimorar o segmentador proposto por Motta, desenvolveu uma metodologia de segmentação automática das mamas e axilas nos termogramas frontais obtidos por TI. O principal avanço desse trabalho em ralação ao de Motta foi a detecção automática do contorno das mamas da paciente, descartando completamente o fundo da imagem e o restante do corpo da paciente. Ainda em 2012, Oliveira [Oliveira 2012] desenvolveu uma metodologia de segmentação automática das vistas laterais (90°) das mamas nos termogramas obtidos por TI. Em sua tese de doutorado, Olivera [Olivera 2013] construiu um banco de dados com recuperação de dados baseada no conteúdo, acessível via *web*, para armazenar todos os termogramas obtidos por TI e por TID das pacientes do Hospital Universitário Antônio Pedro (HUAP). Esse banco foi construído não somente para servir aos trabalhos do nosso grupo, mas à qualquer grupo de pesquisa no mundo para que esse possa testar e comparar suas metodologias baseadas em termografia infravermelha de mama. A aquisição e o uso dessas imagens para fins de pesquisa foram aprovados pelo comitê de ética do hospital como parte do projeto intitulado por "Aquisição, Armazenamento e Verificação da Viabilidade do Uso de Imagens Térmicas na Detecção de Doenças da Mama", registrado na Plataforma Brasil, sob o número de Certificado de Apresentação para Apreciação Ética (CAAE) 01042812.0.0000.5243, do Ministério da Saúde, do qual todos os trabalhos do grupo fazem parte. Acrescentando outras classes de características para imagens digitais, além das usadas por Resmini, Borchartt [Borchartt 2013] desenvolveu uma metologia computacional para a classificação de alterações na mama. Além disso, ele desenvolveu uma ferramenta, o *ThermoCAD-UFF*, que reuniu todos as metodologias de segmentação e análise dos termogramas obtidos por TI, propostas nos trabalhos anteriores.

Os trabalhos citados no parágrafo anterior foram todos desenvolvidos utilizando os termogramas obtidos por TI. Entretanto, resultados de trabalhos encontrados na literatura, testes realizados em nosso laboratório e análise gráfica de dados obtidos serviram de argumentos para acreditar que a TID é superior a TI na tarefa de detectar anormalidades da mama utilizando metodologias computacionais. Assim, iniciou-se o desenvolvimento de pesquisas utilizando os termogramas obtidos por TID. A TID é um método para monitorar a resposta dinâmica da temperatura da pele após um estresse térmico. Em vários protocolos de detecção de câncer de mama, o tipo de estresse térmico

mais utilizado é a aplicação de fluxo de ar direcionado às mamas utilizando um ventilador elétrico [Borchartt et al. 2013]. O resfriamento das mamas, teoricamente, melhora o contraste térmico entre tecidos saudáveis e doentes na imagem, pois vasos sanguíneos gerados em função do tumor canceroso não possuem camada muscular e nem regulação neural como vasos embrionários [Amalu 2004]. Esses vasos são somente tubos endoteliais e portanto não contraem em resposta à estimulação do sistema nervoso simpático. Por essa razão, a região da mama com tecidos cancerosos permanece com temperatura quase inalterada, enquanto que a parte saudável da mama é resfriada [Amalu 2004]. A TID é mais rápida e mais robusta quando comparada à TI, que requer condições rígidas de ambiente e tempo significativamente longo para aclimatação da paciente às condições da sala de exame. Por outro lado, a TID é menos dependente das condições e temperatura da sala de exame [Herman 2013].

O primeiro trabalho do grupo utilizando termogramas obtidos por TID foi desenvolvido por Galvão [Galvao 2015]. Em sua tese de doutorado, Galvão propôs uma metologia de registro dos termogramas obtidos em sequência de uma mesma paciente, um passo fundamental na análise da TID, pois se o objetivo é monitorar e quantificar as mudanças da temperatura em cada ponto da mama, é necessários que as imagens da sequência estejam emparelhas, "casadas", e isso é realizado pelo registro das imagens.

Dando continuidade as pesquisas utilizando termogramas obtidos por TID, a presente tese propõe uma metodologia computacional para detectar anormalidades da mama. Mesmo em outros grupos de pesquisa, a TID não tem sido computacionalmente explorada como TI. Nas últimas décadas, muitos trabalhos propondo metologias computacionais para a detecção do câncer de mama utilizando imagens capturadas por TI foram publicados. Ao contrário, trabalhos propondo metologias computacionais para o mesmo propósito utilizando os termogramas obtidos por TID são poucos [Borchartt et al. 2013]. Além disso, esses trabalhos, principalmente os mais antigos, não utilizam técnicas de aprendizagem de máquina para indicar se a paciente possui alterações em uma das mamas ou em ambas, por meio de análises da distribuição e comportamento, no tempo, da temperatura na superfície da pele. Entre os encontrados na literatura, apenas o trabalho de Wishart *et al.* utiliza uma técnica de inteligência artificial para auxiliar na análise dos termogramas [Wishart et al. 2010]. Entretanto, segundo os autores, os resultados obtidos utilizando tal técnica foram: sensibilidade de 48% e especificidade de 74%. Os demais trabalhos propõem metodologias para o desenvolvimento de sistemas de auxílio ao diagnóstico médico, onde a decisão do profissional seria baseada em gráficos ou em imagens pré-processadas computacionalmente [Ohashi e Uchida 1997]

[Ohashi e Uchida 2000] [Anbar et al. 2000] [Anbar et al. 2001] [Parisky et al. 2003] [Button et al. 2004][Saniei et al. 2015][Kaczmarek e Nowakowski 2004][Arora et al. 2008] [Gerasimova et al. 2012] [Gerasimova et al. 2013] [Gerasimova et al. 2014] [Amalu 2004] .

Resumindo, as motivações para o desenvolvimento desta tese foram:

1. a possibilidade de contribuir na definição da população-alvo em programas de rastreamento organizado do câncer de mama para um consequente uso mais racional dos exames atualmente em uso para esse fim, principalmente a mamografia, um exame de difícil acesso para algumas camadas sociais da população brasileira e regiões do país; e

2. a carência de metodologias computacionais utilizando aprendizagem de máquina para detecção de anomalias de mama, entre elas o câncer, utilizando imagens capturadas por TID.

1.2 Objetivo

Mediante as motivações apresentadas, o principal objetivo desta tese é propor uma metologia computacional que detecte anormalidades na mama, utilizando as imagens capturadas por TID, e que essa metodologia possa contribuir na indicação de indivíduos para a população-alvo em programas de rastreamento organizado do câncer de mama.

A Figura 1.1 mostra as séries temporais de temperatura obtidas de uma paciente com câncer de mama e a Figura 1.2 mostra as séries temporais de temperatura obtidas de uma paciente saudável. O eixo x representa o tempo em segundos (5 minutos, tempo de duração do exame), o eixo y contém o índice de cada uma das séries temporais, e o eixo z indica a temperatura em graus Celsius. É possível observar que existem um grupos de séries temporais com temperaturas maiores e com inclinação maior nos momentos iniciais da recuperação da temperatura, após o estresse térmico (resfriamento das mamas por um ventilador elétrico), para a paciente doente (Figure 1.1), ou seja, séries temporais de temperatura que se destacam das demais. O mesmo não é verdadeiro para as séries temporais obtidas da paciente saudável (Figure 1.2). O objetivo da metologia computacional é "perceber" esse grupo que se destaca dos demais nas pacientes com alguma anormalidade de mama, e para isso, aprendizagem de máquina não supervisionada e supervisionada são aplicadas em momentos diferentes da metodologia proposta, o que a caracteriza como

híbrida.

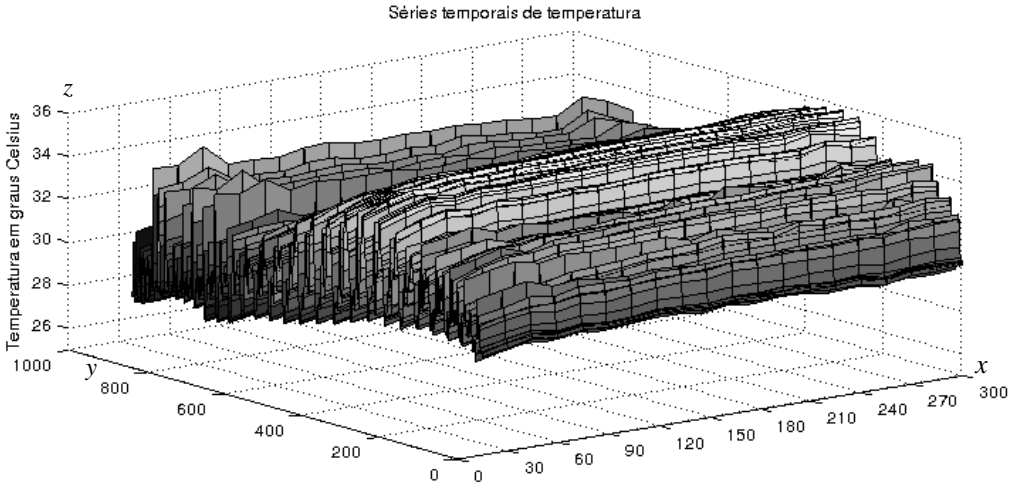

Figura 1.1: Recuperação da temperatura da pele das mamas de uma paciente doente.

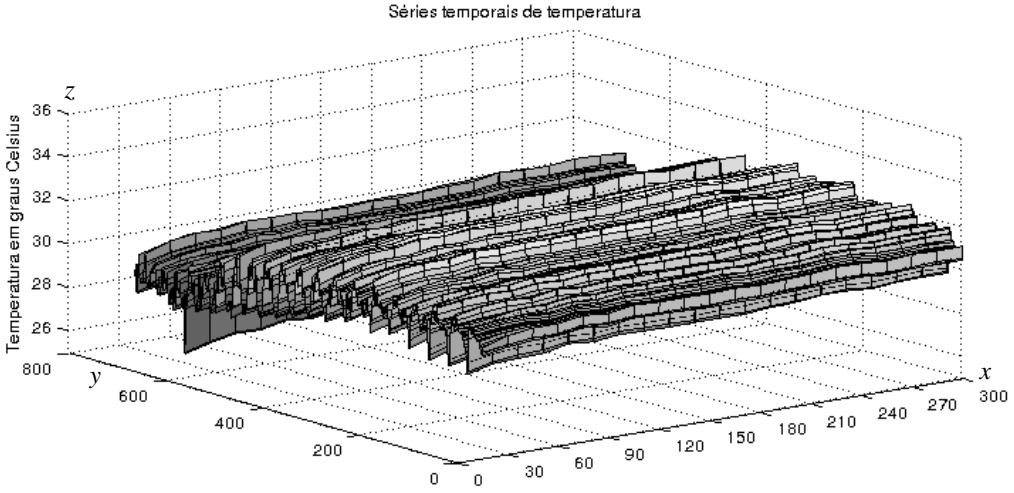

Figura 1.2: Recuperação da temperatura da pele das mamas de uma paciente saudável.

Em busca do objetivo mencionado acima, algumas etapas da metodologia foram construídas ao longo de seu desenvolvimento. A primeira foi o estabelecimento de um protocolo de examinação por TID a ser executado com as pacientes do HUAP [Silva et al. 2013] [Silva et al. 2014b]. Em seguida, a segmentação da região de interesse, no caso ambas as mamas, nos termogramas capturados executando o protocolo estabelecido. Após a segmentação, o registro das imagens e a construção das séries temporais de temperatura [Galvao 2013] [Silva et al. 2014a]. Por último, o agrupamento das séries temporais construídas, a avaliação dos grupos formandos e a classificação da paciente em doente ou saudável [Silva et al. 2015b] [Silva et al. 2015a].

1.3 Contribuições

A principal contribuição desta tese é o desenvolvimento de uma metodologia computacional para a detecção de anormalidades encontradas na mama, examinadas por TID, para que sistemas de exame baseados na metodologia proposta possam ser usados em programas de rastreamento organizado de câncer de mama, contribuindo na definição da população-alvo. Contribuições secundárias são:

1. adaptação do registro computacional de imagens baseado em intensidade de pixel proposto em [Myronenko e Song 2010] para os termogramas obtidos por TID da base de dados *Database for Mastology Research with Infrared Image* (DMR-IR);

2. o estabelecimento de um conjunto de características, utilizado na classificação da paciente;

3. a indicação de um algoritmo de classificação para a construção do modelo de classificação, e suas respectivas configurações de parâmetros;

4. o estabelecimento de um protocolo de aquisição dos termogramas na examinação por TID; e

5. a disponibilidade de todo o material produzido no endereço eletrônico http://visual.ic.uff.br/ (a segmentação das imagens, as imagens registradas e as características computadas para o desenvolvimento de ferramentas de mineração de dados, tais como classificadores e métodos de seleção de características).

1.4 Organização da tese

Esta tese é composta por seis capítulos. O Capítulo 2 aborda os conceitos básicos sobre os quais a metodologia proposta se apoia e por isso necessários para o entendimento da mesma, por parte do leitor. São eles: o câncer de mama e seu processo de crescimento; exames de detecção de câncer de mama, suas características e limitações; a TID e suas propriedades; registro computacional de imagens; a mineração de dados e alguns de seus conceitos; e séries temporais.

O Capítulo 3 contém a revisão bibliográfica realizada no início da pesquisa. Nesse capítulo, trabalhos encontrados na literatura sobre a detecção de câncer e anormalidades de mama, utilizando imagens capturadas por TID, são descritos. Os pontos principais

abordados em cada trabalho, quando possível, são: o protocolo de aquisição das imagens, os métodos e técnicas computacionais de processamento e análise das imagens e/ou dados térmicos; e os resultados obtidos e conclusões. O capítulo é finalizado com uma comparação do trabalho proposto nesta tese com os encontrados na literatura.

A metodologia proposta está detalhada no Capítulo 4. O capítulo é iniciado descrevendo o protocolo de execução da TID utilizado no HUAP. Em seguida, o capítulo descreve: os testes preliminares à metodologia, a segmentação da região de interesse, o registro das imagens capturadas por TID, a construção das séries temporais de temperatura, o agrupamento das séries temporais, a avaliação dos grupos formados pelo algoritmo de agrupamento, e a classificação das pacientes em saudáveis ou doentes.

O Capítulo 5 apresenta as avaliações da metodologia proposta. Nesse capítulo, a base de dados, que contém as imagens utilizadas para formar o conjunto para treinamentos e testes, é descrita. Além disso, esse capítulo apresenta as ferramentas de mineração de dados utilizadas para a resolução do problema de seleção de algoritmos e otimização de parâmetros em problemas de classificação, e a avaliação dos resultados obtidos com os algoritmos e parâmetros recomendados por essas ferramentas. O capítulo finaliza com análises complementares, uma discussão dos resultados obtidos e o resumo da metodologia proposta.

O último capítulo deste texto contém um detalhamento das contribuições principais da metodologia proposta. Além disso, o Capítulo 6 finaliza esta tese com as conclusões finais e apresenta os trabalhos futuros consequências, imediatadas do presente trabalho.

Capítulo 2

Conceitos básicos

Este capítulo aborda os conceitos básicos envolvidos no desenvolvimento da metodologia proposta nesta tese. Em suas primeira seções, o capítulo aborda o câncer de mama, as estratégias de detecção dessa doença, e a termografia infravermelha de mama. Na segunda parte do capítulo, técnicas computacionais tais como aprendizagem de máquina não supervisionada e supervisionada e técnicas de avaliação de resultados são revisadas, finalizando com a definição e principais características de séries temporais.

2.1 Câncer de mama

Referências às doenças da mama são datadas de 1600 a.C., no Egito Antigo. Desde então, o câncer de mama tem sido a forma de câncer mais estudada e descrita na história da medicina. Sua ocorrência e alta taxa de mortalidade instigam os pesquisadores em buscar suas causas e melhorar os resultados de tratamentos [de Jesus 2005]. No Brasil, é o tipo mais frequente e a maior causa de morte por câncer entre as mulheres [Silva e Hortale 2012].

Por mecanismos genéticos ou ambientais, algumas células do corpo começam a se multiplicar de forma descontrolada dando origem a uma neoplasia benigna ou maligna, sendo a maligna também chamada de câncer [de Jesus 2005]. O câncer invasivo possui a capacidade de invadir tecidos próximos, além da membrana que constitui a borda epitelial, ocorrendo um fenômeno conhecido como metástase. Nessa fase, células do tumor, por meio da corrente sanguínea, se espalham por outras partes do corpo como o esqueleto, o fígado, os pulmões, as pleuras, o cérebro e qualquer outra parte do corpo [Richie e Swanson 2003].

No início da tumorigênese, as lesões neoplásicas atravessam uma fase de crescimento não vascular e atingem um tamanho de $3mm$, no máximo (Figura 2.1(a)). O final dessa

fase é marcada por um evento que distingue os tumores de crescimento dos tumores dormentes: "o interruptor angiogênico", o qual é controlado pelo saldo líquido entre reguladores positivos e negativos de crescimento de vascularidade nova. A ativação desse "interruptor" permite ao tumor recrutar vasos sanguíneos circundantes, estimulando-os a formarem vasos novos (angiogênese) que irão supri-lo de oxigênio e de nutrientes (Figura 2.1(b)). A vascularidade do tumor é responsável por seu crescimento (Figura 2.1(c)), pois resolve o problema da limitação de oxigênio e nutrientes, mas por outro lado aumenta o fluxo de sangue na região e, consequentemente, a temperatura [Hoeben et al. 2004].

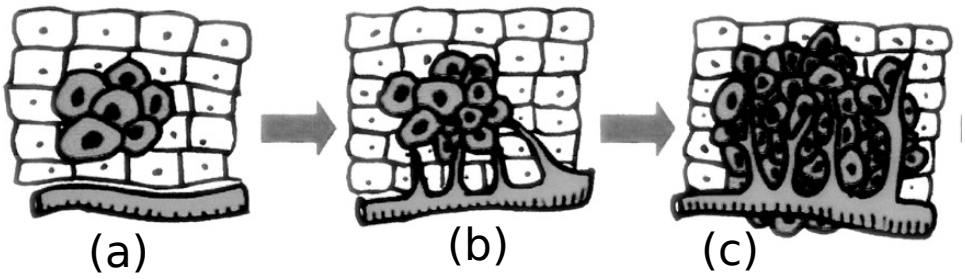

Figura 2.1: Desenvolvimento do câncer de mama: em (a), fase de crescimento não vascular das lesões neoplásicas, em (b), angiogênese e em (c), fase de crescimento vascular [Hoeben et al. 2004].

2.1.1 Exames de detecção de câncer de mama

Quando detectado em estágios iniciais e tratados adequadamente, aproximadamente um terço dos casos de câncer pode ser curado, segundo a *Organização Mundial de Saúde* (OMS). O autoexame das mamas, o exame clínico e os exames por imagens são as técnicas de detecção mais difundidas e estudadas [Silva e Hortale 2012]. O autoexame consiste na inspeção e apalpação das mamas, pela própria mulher, semelhante ao procedimento executado pelo médico, realizado mensalmente, na semana seguinte à menstruação ou em um dia fixo para pacientes que não menstruam mais. O exame clínico é também composto por inspeção visual e palpação, porém executado por um profissional de saúde capacitado. Entre os exames por imagens está a mamografia, um exame radiológico realizado com equipamento dedicado que emite doses de radiação ionizante para formar imagens da mama com a finalidade de detectar massas anormais [Silva e Hortale 2012]. Considerada como o padrão ouro entre os exames, a mamografia é utilizada na prevenção secundária, rastreando lesões não palpáveis da mama e, assim, contribuindo com uma redução de aproximadamente 30% da taxa de mortalidade [de Jesus 2005]. A mamografia mostra tumores ainda em estágios iniciais ou suficientemente pequenos para serem percebidos

por um médico. Durante o exame, as mamas, direita e esquerda, são comprimidas no mamógrafo para gerar uma visualização melhor dos tecidos e aumentar a qualidade da imagem por evitar erros de um possível movimento. A ultrassonografia é outro exame por imagem sendo um bom marcador de características suspeitas e volume do tumor, o que o torna auxiliar importante no rastreio mamográfico. Em mulheres jovens, quando a mama é mais densa, sua utilização é apropriada na diferenciação do conteúdo e superfície interna de nódulos. Ainda na categoria exames por imagem, a ressonância magnética aparece como um complemento à mamografia auxiliando na definição da natureza do achado mamográfico e no planejamento terapêutico. Em alguns casos pode revelar tumores não observáveis à mamografia e/ou ultrassonografia, mas geralmente é aplicada em casos onde existem dificuldades no diagnóstico [de Jesus 2005], porém não detecta a presença de microcalcificações, que são os achados mais frequentes em tumores *in situ* (em até 70% dos casos) [Bravo 1999].

Infelizmente, todos os exames citados no parágrafo anterior apresentam limitações e desvantagens. A ultrassonografia encontra dificuldades para detectar microcalcificações e por isso não é recomendada como um exame de rastreamento, embora evidências preliminares mostrem o contrário. A ressonância magnética gera um número alto de casos de achados falsos positivos devido a sua baixa especificidade, além disso, esse exame não detecta microcalcificações [Minamimoto et al. 2015]. E a mamografia, além de causar muito desconforto e dor à paciente, apresenta algumas limitações na detecção de tumores, por exemplo, em lesões não delineadas pelo método, sem microcalcificações evidentes, e também em pacientes mais jovens, onde as mamas são constituídas de tecido glandular (mais denso ao exame), prejudicando o diagnóstico. A sensibilidade total da mamografia é de somente 75%, e esse valor diminui para 50% para mulheres com mamas densas e heterogêneas [Minamimoto et al. 2015]. Além isso, um estudo, publicado em fevereiro de 2014 [Miller et al. 2014], concluiu que a mamografia anual em mulheres com idade entre 40 e 59 não reduz a mortalidade por câncer de mama. O estudo concluiu que 22% dos casos de câncer de mama invasivo, detectados por triagem e utilizados na pesquisa, não eram casos de câncer de fato.

Uma das alternativas adotadas para contornar as limitações dos exames atualmente em uso na triagem e diagnóstico do câncer de mama é usá-los de forma complementar uns aos outros, pois cada um deles apresenta características específicas. A próxima seção descreve a termografia infravermelha, um exame classificado como biológico ou funcional [Usuki et al. 1990], e que poderia ser incluído no grupo de exames utilizados no rastreamento do câncer de mama.

2.2 Termografia infravermelha

O primeiro registro do uso do diagnóstico termobiológico pode ser encontrado em escritos de Hipócrates próximos ao ano de 480 a.C. Quando uma quantidade de lama era espalhada sobre o paciente, a área que secava primeiro indicava o local da doença. Desde então, pesquisas contínuas e observações clínicas revelaram que certas temperaturas relacionadas ao corpo humano eram de fato indicações de processos fisiológicos normais e anormais [Amalu et al. 2008].

Para qualquer objeto com temperatura acima do zero absoluto ($-273K$) é possível aplicar a Lei de Stefan-Boltzmann [Boltzmann 1884]. Esta lei relaciona a radiação infravermelha emitida da superfície do objeto com sua temperatura pela Equação 2.1:

$$W = A\sigma\varepsilon T^4 \tag{2.1}$$

onde A é a área analisada em m^2, σ é a constante de proporcionalidade ou a constante de Stefan-Boltzmann, ε é a emissividade da superfície do objeto e T é a temperatura absoluta do objeto em *Kelvin*. Em outras palavras, esta lei afirma que a radiação total emitida por um objeto é diretamente proporcional à área do objeto, à sua emissividade e à quarta potência de sua temperatura absoluta.

A emissividade da pele humana é extremamente alta ($\varepsilon \approx 0,98$) e por essa razão a radiação infravermelha emitida por um ponto do corpo do paciente pode ser convertida diretamente para um valor de temperatura. Esse processo é realizado por uma câmera que possui sensores que captam a radiação emitida por todos os pontos de uma cena, os transforma em sinais elétricos e esses sinais em uma imagem chamada de termograma, que representa um mapa da distribuição de temperaturas, ou seja, a matriz de temperatura de todos os pontos da cena [Amalu et al. 2008]. Tal matriz pode ser armazenada em algum tipo de mídia utilizável pelo computador para posterior processamento e análise.

A termografia infravermelha da mama é um exame que detecta e registra a radiação infravermelha emitida pela superfície da mama e produz um termograma. Ela não utiliza radiação ionizante, acesso venoso, ou outro processo invasivo, portanto, o exame não apresenta dano ou risco algum à paciente. Classificado como um exame funcional, a termografia infravermelha provê informações fisiológicas de funcionamento normal ou anormal dos sistemas vascular, sensorial e nervoso simpático, bem como de processos inflamatórios [Amalu et al. 2008] [Head e Elliott 2002], além de apresentar um custo extremamente baixo quando comparado aos demais exames. A Figura 2.2 exibe um termograma onde

a temperatura de cada ponto está representada por uma cor, conforme a escala vertical de cor/temperatura do termograma. Na imagem, os pontos da cena de temperaturas mais baixas estão representados pela cor azul escuro, os pontos da cena de temperaturas intermediárias, estão representados pelas cores azul claro, verde, amarela e vermelha, os pontos da cena de temperaturas mais altas estão representados por cores próximas do branco e o próprio branco.

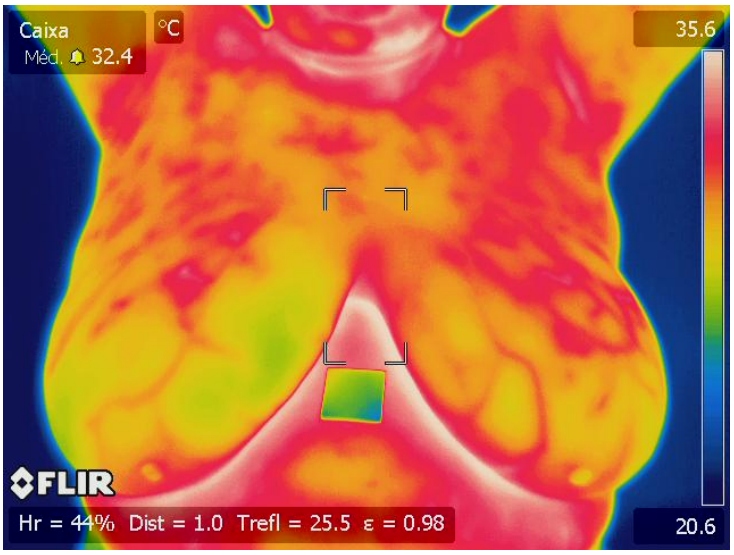

Figura 2.2: Termograma das mamas de uma paciente.

2.2.1 Termografia Infravermelha Estática (TI)

A termografia infravermelha estática (TI) é o mapeamento de temperaturas estáticas, ou seja, é a medição da distribuição de temperaturas da cena no campo visual da câmera em um dado instante de tempo [Amalu et al. 2008]. Na TI de mama, os procedimentos de preparação da sala de exame e da paciente devem ser rigorosamente respeitados para que os termogramas gerados tenham qualidade, isto é, sejam fisiologicamente neutros e livres de artefatos térmicos, prontos para a interpretação e diagnóstico. Tais procedimentos compõem o protocolo de aquisição das imagens que é dividido nas seguintes partes: recomendações à paciente; condições da sala de exame; preparação da paciente; e parâmetros de captura [Amalu et al. 2008] [Silva et al. 2014b].

Geralmente recomenda-se à paciente que não fume e que não consuma cafeína ou bebidas alcoólicas durante um certo tempo antes do exame. Também lhe é recomendado não aplicar loções, desodorantes ou cremes sobre a pele de onde será gerada a imagem (ou as imagens) no dia do exame. Além disso, a paciente deve evitar a exposição ao sol,

estimulação ou tratamento das mamas, exercícios físicos ou banho antes do exame. Todas essas ações são para evitar alterações de temperatura da pele e consequentemente, dados térmicos inconsistentes da área analisada [Silva et al. 2014b] [Amalu et al. 2008].

A TI deve se realizada em um ambiente controlado. A primeira razão para isto é a natureza da fisiologia humana. Influências de um ambiente externo não controlado e fontes de calor próximas à paciente podem alterar a temperatura da pele [Amalu et al. 2008]. Desta forma, as seguintes ações são adotadas: bloquear janelas e qualquer abertura por onde os raios solares e fluxos de ar possam penetrar; utilizar lâmpadas fluorescentes posicionado-as longe do local de aquisição das imagens; manter a temperatura entre 18 °C e 23 °C (isto garante que a paciente se sinta confortável sem transpirar por conta do calor ou tremer devido ao frio), e no momento do exame essa não deve variar mais do que um grau Celsius dentro desse intervalo de temperatura; eliminar possíveis fontes de calor; controlar a quantidade de pessoas no ambiente de exame; direcionar o fluxo de ar do condicionador de ar para longe da paciente; por carpete no piso ou solicitar à paciente que use calçados a fim de evitar o aumento do estresse fisiológico [Silva et al. 2014b] [Amalu et al. 2008].

Estando na sala de exame, a paciente é solicitada a retirar a roupa da parte de cima do corpo, como também cordões, brincos e outros acessórios que possam interferir no exame e, se necessário, a prender os cabelos [Silva et al. 2014b]. Por último, a paciente é submetida a aclimatação por alguns minutos (geralmente de 10 a 15 minutos) até a estabilização térmica da temperatura da pele dentro do ambiente de aquisição. Ao final deste período, as mudanças de temperatura na superfície do corpo ocorrem muito lenta e uniformemente, sem criar diferenças significativas entre regiões anatomicamente homólogas, e a influência térmica da roupa e de ambientes externos já não existe mais. Geralmente a paciente é posicionada a frente da câmera com as mãos sobre a cabeça para obtenção de uma representação anatômica melhorada nas imagens infravermelhas. Em relação às posições das imagens, a mais comum é a frontal (paciente de frente para a câmera), mas, dependendo da anatomia individual da paciente, outras posições são requeridas de tal modo que toda a superfície de ambas as mamas seja mapeada termicamente [Amalu et al. 2008].

A distância entre a câmera e a paciente é um dos parâmetros que fazem parte do protocolo de aquisição. A distância mais comum é de 1 (um) metro, podendo variar dependendo do tamanho da paciente. Outros parâmetros que são considerados no momento da aquisição das imagens são: temperatura do ambiente, umidade relativa do ar e o coeficiente de emissividade da pele humana. Todos esses parâmetros são necessários para que o dispositivo de aquisição (a câmera) possa determinar a temperatura real de cada ponto

da cena [Silva et al. 2014b].

2.2.2 Termografia Infravemelha Dinâmica (TID)

Quando comparada à TI (Seção 2.2.1), a Termografia Infravermelha Dinâmica (TID) é mais rápida e mais robusta. Isto se deve ao fato de que a primeira requer condições rigorosas de controle das condições do ambientes e tempo significamente longo para a aclimatação da paciente às condições da sala de exame. Por outro lado, a TID é muito menos dependente das condições e temperatura do ambiente [Herman 2013]. Enquanto que a TI apresenta-se como uma técnica de observação da distribuição de temperaturas sobre a superfície das mamas, a TID monitora ou quantitativamente mede as mudanças de temperatura sobre tal superfície em um determinado período de tempo [Anbar 2008]. A TID foi conceituada no final da década dos anos 80 por Michael Anbar [Anbar 1987] [Montoro e Anbar 1988]. Anbar notou que mudanças rápidas na temperatura da pele humana produziam informações fisiológicas e patofisiológicas valiosas, informações estas que não podem ser obtidas pelo mapeamento de temperaturas estáticas por TI. Esses conhecimentos foram solidificados no início da década dos anos 90 com base em estudos experimentais [Anbar 2008].

Visto que, sob condições fisiológicas específicas, a temperatura da pele depende do sangue que chega aos tecidos cutâneos e subcutâneos, sendo esse o fluido de troca de calor, tal temperatura reflete uma variedade de processos hemodinâmicos. Esses processos são modulados por mudanças cardiogênicas pulsáteis no fluxo de sangue como também por controle neural do fluxo de sangue na vasculatura e microvasculatura. Considerando que a perfusão sanguínea afeta a temperatura de cada região da pele de maneira peculiar, dependendo da anatomia vascular subjacente, a avaliação desses processos requer o registro de temperatura (aquisição das imagens) da região de interesse em frequências maiores do que a mais alta frequência de modulação mensurável do suprimento de sangue. A TID da região de interesse pode prover informação adicional da anatomia da vasculatura, das mudanças sistêmicas no fluxo sanguíneo devido tanto a função cardíaca quanto ao controle sistêmico neural do fluxo de sangue vascular e perfusão do leito capilar. Além de informação fisiológica e anatômica útil, a TID pode ser clinicamente interessante, pois patologias que afetam quaisquer parâmetros fisiológicos ou anatômicos de suprimento de sangue podem ser diagnosticadas por essa técnica [Anbar 2008].

Lesões neoplásicas são geralmente associadas com angiogênese e subsequente hipervascularização local (Seção 2.1). Isto por si só produz aberrações na anatomia local da

perfusão. Além disso, os vasos sanguíneos neoplásicos recém formados são susceptíveis de serem pouco inervados, quando não totalmente sem terminações nervosas, e portanto respondem ao controle vascular neural de forma anormal, fato este detectável por TID. Porém, uma característica ainda mais específica de lesões neoplásicas, que pode contribuir significativamente para a detecção e tratamento de câncer por TID, é sua produção de Óxido Nítrico (ON) [Anbar 2008].

Em 1978 foi descoberto que humanos e outros animais produzem e transportam quantidades mensuráveis de ON na corrente sanguínea. A função fisiológica desta substância altamente reativa era desconhecida até início da década de 1990, quando foi mostrado que ela é um neurotransmissor que induz a vasodilatação. Doenças neoplásicas estão associadas com hipertermia local da pele sobrejacente e Anbar mostrou em 1994, por análise quantitativa, que este fenômeno está diretamente ligado à vasodilatação local, que por sua vez é consequência da geração de ON por tecidos cancerosos [Anbar 2008].

Visto que foi descoberto por volta de 1994 que o ON é um mensageiro químico na regulação do tono vascular, Anbar especulou que o ON produzido por lesões cancerosas poderia interferir na regulação vasomotora da perfusão sanguínea. Assim, a hipótese de Anbar conjecturava não somente que lesões cancerosas melhoravam a perfusão vascular regional, mas também que a perfusão em tecidos circundantes não seria modulada normalmente. O corolário dessa hipótese é que a TID pode detectar o efeito do ON produzido pelo câncer na modulação neural da perfusão, e portanto detectar lesões cancerosas mais eficientemente do que o monitoramento da hipertermia local. Estudos posteriores comprovaram que a hipertermia induzida por câncer de mama não é devido a hipervascularização local, mas sim a vasodilatação local.

Na prática, após um estímulo térmico, a TID monitora as mudanças dinâmicas da temperatura da pele. Nos métodos desenvolvidos para detecção de câncer de mama por TID, o estímulo térmico mais empregado é o ar direcionado às mamas utilizando um ventilador elétrico. O resfriamento das mamas por ar, teoricamente, melhora o contraste térmico entre os tecidos saudáveis e doentes, nas imagens capturadas. Isto porque, como já comentado, os vasos sanguíneos produzidos por tumores cancerosos praticamente não possuem terminação nervosa como os vasos embriológicos [Amalu 2004]. Esses vasos são apenas tubos endoteliais e por isso não se contraem em resposta a um estímulo simpático. Dessa forma, a região do câncer permanece com a temperatura praticamente inalterada quando a mama é resfriada. Mediante a redução do diâmetro vascular, regiões saudáveis da mama apresentam redução da temperatura [Amalu 2004].

Nas próximas seções, técnicas computacionais utilizadas na análise proposta nesta tese para identificação de pacientes com câncer são descritas.

2.3 Registro de imagens

O registro de imagens é o processo de emparelhamento de duas ou mais imagens de uma mesma cena capturadas em diferentes instantes e/ou de diferentes pontos de captura e/ou por diferentes sensores. As diferenças entre as imagens são provenientes de condições inerentes, no momento da captura. O registro de imagens é um passo crucial em todas as tarefas de análise de imagens onde a informação final é obtida da combinação de várias fontes de dados, como na fusão de imagens, detecção de mudanças e restauração de imagens multicanais [Zitová e Flusser 2003]. O registro de imagens também pode ser considerado como um mapeamento espacial entre duas imagens. O objetivo do registro é encontrar a transformação T ótima, ou a função de mapeamento, que irá alinhar uma imagem à outra de forma que características correspondentes possam ser facilmente relacionadas e as imagens alinhadas possam ser diretamente comparadas, combinadas e analisadas [Myronenko e Song 2010] [Guo et al. 2006] [Hajnal et al. 2001].

Atualmente, imagens médicas capturadas de um paciente em instantes de tempo diferentes são bastante comuns, ou utilizando o mesmo dispositivo (monomodal) ou utilizando dispositivos diferentes (multimodal). Também é comum capturar imagens de um paciente dinamicamente, isto é, ter uma sequência de imagens adquiridas por, frequentemente, muitos quadros por segundos (como acontece na TID, Seção 2.2.2). A quantidade cada vez maior de imagens capturadas, estimula os profissionais de saúde a relacionar tais imagens para a extração de informações clínicas relevantes. Nesse caso, o registro pode ajudar na tarefa de comparar imagens capturadas de forma monomodal, multimodal e dinâmica [Hajnal et al. 2001].

Entre as aplicações de registro de imagens médicas em rotinas clínicas, pode-se incluir: o monitoramento de mudanças sutis devido à progressão de doenças ou tratamento; monitoramento de perfusão ou outras análises funcionais quando não é possível manter o paciente parado em uma posição fixa durante a aquisição dinâmica; e intervenções guiadas por imagens adquiridas previamente e que são registradas para possibilitar o uso dessas pelo cirurgião ou intervencionista em seus trabalhos. O registro de imagens também vem se tornando uma técnica valiosa em pesquisas biomédicas, especialmente em neurociência, onde alguns estudos com imagens estão fornecendo contribuição de como o cérebro

funciona [Hajnal et al. 2001].

Algoritmos de registro de imagens podem ser divididos em três categorias: os que constroem a transformação T usando pontos identificados nas imagens (pontos característicos); os que constroem a transformação T usando superfícies delineadas a partir da imagem; e os que constroem a transformação T usando valores de intensidade de pixels. Os algoritmos de registro usados neste trabalho pertencem à última categoria, que consiste na determinação da transformação T otimizando alguma medida de similaridade calculada diretamente sobre os valores de intensidades dos pixels, ao invés de utilizar estruturas geométricas tais como pontos ou superfícies derivadas das imagens. Nesse caso, a transformação T é determinada iterativamente. Em cada iteração, tais algoritmos transformam a imagem usando a estimativa atual de T e recalculam uma medida de similaridade de pixels. Exceto quando T é uma transformação simples por um número inteiro de pixels, a transformação realizada em cada iteração envolve interpolação entre pontos da amostragem. Dessa forma, a transformação é denotada por Γ e mapeia a posição do pixel e o valor de intensidade associado a essa posição. Definindo a imagem A como a imagem referência, ou a imagem alvo, a imagem B como a imagem iterativamente transformada, ou a imagem móvel, e B^{Γ} a imagem transformada pela transformação Γ, B^{Γ} será definida nas coordenadas da imagem A e os valores dos pixels dependerão do tipo de interpolação usada [Hajnal et al. 2001].

Com maior formalismo matemático, considere as imagens A e B capturadas de um mesmo paciente, com a mesma ou diferentes modalidades de exames. A transformação T (Equação 2.2) realiza o mapeamento espacial que transforma uma posição x_B da imagem B na posição x_A da imagem A [Hajnal et al. 2001].

$$T : x_B \mapsto x_A \Leftrightarrow T(x_B) = x_A \tag{2.2}$$

Considerar a função inversa de mapeamento T^{-1}, que realiza o mapeamento x_A a x_B é algumas vezes importante. Na imagem A, $A(x_A)$ indica o valor de intensidade na posição x_A, similarmente $B(x_B)$ na imagem B. É importante lembrar que as imagens A e B possuem um campo limitado de visão que normalmente não cobre o paciente inteiro. Além isso, este campo de visão é sujeito a ser diferente para as duas imagens, que podem ser consideradas como funções (Equação 2.3 e Equação 2.4) que mapeiam os pontos do campo de visão (ou domínio Ω) para valores de intensidade.

$$A : x_A \in \Omega_A \mapsto A(x_A) \tag{2.3}$$

$$B : x_B \in \Omega_B \mapsto A(x_B) \tag{2.4}$$

Os domínios Ω_A e Ω_B são quase sempre diferentes, pois as imagens possuem campos de visão diferentes na maioria dos casos [Hajnal et al. 2001].

Como as imagens A e B representam uma paciente X, existe uma relação entre as localizações espaciais em A e em B. A imagem A é tal que a posição $x \in X$ é mapeada para x_A, e a imagem B é tal que a posição $x \in X$ é mapeada para x_B. O processo de registro consiste em recuperar a transformação T que mapeia x_A a x_B, sobre o domínio de interesse inteiro, ou seja, T realiza o mapeamento de Ω_A para Ω_B dentro da porção sobreposta dos domínios, denotada por $\Omega^T_{A,B}$. A sobreposição dos domínios, expressada pela Equação 2.5, depende das imagens originais A e B e da transformação espacial T [Hajnal et al. 2001].

$$\Omega^T_{A,B} = \{x_A \in \Omega_A \,|\, T^{-1} \in \Omega_B\} \tag{2.5}$$

2.4 Aprendizagem de máquina

A aprendizagem de máquina é definida como um conjunto de métodos que podem, automaticamente, detectar padrões em um conjunto de dados, e em seguida usar os padrões descobertos para prever dados futuros, ou para realizar outros tipos de tomada de decisão em condições de incerteza. A aprendizagem de máquina é geralmente dividida em dois principais tipos: a supervisionada e a não supervisionada [Murphy 2012].

2.4.1 Aprendizagem de máquina não supervisionada

Na abordagem não supervisionada ou descritiva da aprendizagem de máquina, somente as entradas são fornecidas, ou seja, $T(x_i)\ \ 1 \leq i \leq n$, e o objetivo é buscar por "padrões interessantes" no conjunto de dados. Essa tarefa é chamada de descoberta de conhecimento, um problema muito menos bem-definido, pois não é informando que tipo de padrão será procurado e não há uma métrica de erro óbvia para ser usada (contrariamente acontece na aprendizagem supervisionada, onde é possível comparar a predição y para um dado x ao valor observado). A tarefa de agrupar os dados de um conjunto em grupos é um exemplo clássico de aprendizagem de máquina não supervisionada [Murphy 2012].

2.4.1.1 Agrupamento de Dados (AD)

O Agrupamento de Dados (AD) (termo usado em inglês: *data clustering*) é uma atividade humana importante, pois, desde crianças, somos capazes de agrupar objetos semelhantes e, continuamente e subconscientemente, melhoramos nossos esquemas de agrupamento. Por AD automático, podemos identificar regiões densas e esparsas em um espaço de dados e, portanto, descobrir padrões de distribuição global e correlações interessantes entre os atributos (características) dos dados. Em bases grandes de dados, é muito comum o desconhecimento da classe de cada dado, pois atribuir a classe para todos é uma tarefa muito custosa. O AD, também chamado de análise de grupos, análise de segmentação, análise de taxonomia, classificação não supervisionada ou simplesmente agrupamento, é um método de criação de classes ou grupos de dados (termo usado em inglês: *clusters*) de tal forma que os elementos do conjunto de dados dentro de um mesmo grupo apresentem grande semelhança quando comparados uns com os outros, mas sejam muito diferentes dos elementos em outros grupos. Em outras palavras, é um processo de agrupamento dos elementos de um conjunto de dados em grupos de elementos semelhantes. A semelhança ou a diferença entre os elementos é determinada baseada nos valores de atributos que os descrevem e, frequentemente, medidas de distância são usadas para este fim [Han e Kamber 2006].

O AD tem sido largamente usado em muitas aplicações, incluindo pesquisas de mercado, reconhecimento de padrões, análise de dados, e processamento de imagens. No mundo dos negócios, o AD pode ajudar profissionais de marketing a descobrir grupos distintos em suas bases de clientes e caracterizar grupos baseado em padrões de compras. Na biologia, AD pode ser usado para a obtenção da taxonomia de plantas e animais, categorizando genes com funcionalidades similares, e obter informações sobre as estruturas inerentes a uma população. O AD pode ser usado também para detecção de casos isolados, onde tais casos são mais interessantes do que os casos comuns. Aplicações de detecção de casos isolados incluem detectar fraudes com cartões de crédito e o monitoramento de atividades criminosas no comércio eletrônico. Por exemplo, casos excepcionais em transações em cartões de crédito, tais como compras muito caras e frequentes, pode ser de interesse como possível atividade fraudulenta. Como uma ferramenta de MD, o AD pode ser usado para se obter informações de uma distribuição de dados, para observar as características de cada grupo, e focar em um particular conjunto de grupos para análise futura [Han e Kamber 2006].

O AD é confundido, em algumas situações, com a tarefa de classificação (Seção 2.4.2.2), na qual os dados são associados a classes pré-definidas. Embora a classificação seja um

meio efetivo para distinguir grupos ou classes de dados, esta tarefa requer um conjunto suficientemente grande e com rotulação, na maioria dos casos, custosa dos dados, o qual o classificador usa para modelar cada grupo, na fase de treinamento [Han e Kamber 2006]. No AD, a definição das classes é também um objetivo [Gan et al. 2007]. Ao contrário da classificação, o AD não depende de classes pré-definidas e de exemplos de treinamento já classificados por um especialista. Em aprendizagem de máquina, AD é um exemplo de aprendizagem não supervisionada. Por isso, o agrupamento é considerado uma forma de aprendizagem por observação, e não uma aprendizagem por exemplo. Alternativamente, também pode ser usado como um passo de pré-processamento para outros algoritmos, tais como de extração de características, seleção de características, e classificação, o qual poderia então operar sobre os grupos já detectados e características selecionadas [Han e Kamber 2006].

2.4.1.2 Algoritmos de AD

Existem muitos algoritmos de AD na literatura, e uma categorização rígida dos métodos implementados nesses algoritmos é difícil, pois essas categorias podem se sobrepor, tendo em vista que um método pode ter características de várias categorias. Apesar disso, é importante ter um panorama relativamente organizado desses diferentes métodos. Em geral, os métodos de AD podem ser classificados nas seguintes categorias: métodos de particionamento, métodos hierárquicos, métodos baseados em intensidades, métodos baseados em *grid*, métodos baseados em modelos, métodos de AD de alta dimensão e métodos de AD baseados em restrições. Entretanto, apenas a categoria de métodos de particionamento, a que pertence o algoritmo aplicado neste trabalho, será abordada na próxima seção. Detalhes das outras categorias podem ser encontrados em [Han e Kamber 2006].

Métodos de particionamento

Considerando D um conjunto de dados com n elementos, um método de particionamento constrói k ($k <= n$) partições sobre esse conjunto, onde cada partição representa um grupo. Isto é, ele divide os dados em k grupos, os quais juntos satisfazem os seguintes requisitos: (1) cada grupo deve conter pelo menos um elemento, e (2) cada elemento deve pertencer a um, e somente um, grupo. Note que a segunda condição pode ser relaxada em algumas técnicas de particionamento *fuzzy* [Han e Kamber 2006]. Os grupos são formados otimizando algum critério objetivo de particionamento, tal como uma função de similaridade (proximidade) baseada em distância, de forma que os objetos dentro de um mesmo grupo são próximos ou relacionados entre si (similares) e objetos em grupos diferentes

são distantes ou muito diferentes (dissimilares) uns dos outros, em relação aos atributos (características) do conjunto de dados. Há vários outros tipos de critérios para julgar a qualidade do particionamento. Sendo k o número de particionamentos a ser construído, um método de particionamento cria um particionamento inicial. Então, uma técnica de realocação iterativa é usada na tentativa de melhorar o particionamento, movimentando os elementos de um grupo para o outro [Han e Kamber 2006].

Alcançar um ótimo global em AD baseado em particionamento requer a enumeração exaustiva de todas as possíveis partições. Entretanto, a maioria das aplicações adotam um dos poucos métodos heurísticos populares, tais como (1) o algoritmo *k-means*, onde cada grupo é representado pelo valor médio dos elementos no grupo, e (2) o algoritmo *k-medoids*, onde cada grupo é representado por um dos elementos localizado próximo ao centro do grupo [Han e Kamber 2006]. Nesta tese, entre os algoritmos de particionamento existentes para AD, o *k-means* foi o escolhido.

O algoritmo *k-means*

O algoritmo *k-means*, que utiliza uma técnica baseada em centroide, admite o número de grupos como um parâmetro de entrada e particiona um conjunto de n elementos em k grupos tais que a similaridade resultante intragrupo é alta, mas a similaridade intergrupo é baixa. As medidas de similaridade intragrupo e de dissimilaridade intergrupo, são baseadas nos centroides. O centroide de cada grupo é o valor médio de seus elementos.

No processo de AD pelo *k-means*, primeiramente, k elementos do conjunto de dados são selecionados aleatoriamente, cada um representando o centroide do respectivo grupo que será formado. Para cada um dos elementos remanescentes, a distância para cada um dos centroides é calculada e o elemento é associado ao grupo mais similar a ele (menor distância). Então, um novo centroide é calculado para cada grupo. Esse processo é repetido até que a função critério seja satisfeita. Geralmente, é usado o critério do erro quadrático, definido como:

$$E = \sum_{n=1}^{k} \sum_{p \in C_i} |p - m_i|^2, \tag{2.6}$$

onde E é a soma do erro quadrático para todos os elementos no conjunto de dados; p é o ponto no espaço representando um determinado elemento; e m_i é a média (centroide) do grupo C_i (p e m_i são multidimensionais). Em outras palavras, o erro é a soma dos quadrados das distâncias de cada elemento do conjunto de dados ao centroide do seu respectivo grupo. Esse critério tenta formar k grupos os mais separados e compactos possíveis [Han e Kamber 2006].

Como exemplo, considere um conjunto de objetos localizados no espaço como representado no retângulo mostrado na Figura 2.3(a). Seja $k = 3$, isto é, os objetos serão divididos em 3 grupos. Em conformidade com o Algoritmo 1, 3 objetos são escolhidos aleatoriamente para serem os centroides iniciais dos respectivos grupos, onde os centroides dos grupos estão marcados por "+" (Figura 2.3(a)). Cada objeto é atribuído ao grupo cuja distância de seu respectivo centroide a ele é a menor em relação aos centroides dos outros grupos. As distribuições são representadas por curvas traço-ponto, como mostra a Figura 2.3(a). O próximo passo é a atualização de cada centroide pelo cálculo do valor médio de cada grupo baseado nos objetos correntes no grupo. Então, usando os centroides atualizados, os objetos são novamente redistribuídos aos grupos, como na primeira iteração. A nova distribuição entre os grupos é representada pelas curvas tracejadas na Figura 2.3(b). Este processo de atribuir iterativamente os objetos aos grupos com o objetivo de melhorar o particionamento é referido como *realocação iterativa*, e neste exemplo leva à Figura 2.3(c). O processo tem fim quando não ocorre mais a redistribuição de objetos em algum grupo. Os grupos resultantes são retornados pelo processo de agrupamento [Han e Kamber 2006].

Algoritmo 1 *Algoritmo k-means*

Entrada:
k: número de grupos;
D: um conjunto de dados contendo n elementos.
Saída:
k: um conjunto de k grupos.
Método:
escolher aleatoriamente k elementos de D como os centroides iniciais de grupos;
enquanto existir mudanças **faça**
i) (re)atribuir cada elemento ao grupo para o qual ele é mais similar (mais próximo) baseado no centroide;
ii) atualizar o centroide de cada grupo.
fim enquanto

2.4.1.3 Validação de agrupamento

O objetivo dos métodos de AD é descobrir grupos presentes em um conjunto de dados. No geral, eles buscam por grupos nos quais seus elementos estão próximos uns dos outros (ou seja, são muito similares) e bem separados entre si. Um problema enfrentado nessa tarefa é decidir o número ótimo de grupos que se ajusta ao conjunto de dados. Na maioria das avaliações experimentais de algoritmos de AD, conjuntos de dados 2-D são usados com o objetivo de facilitar a verificação, mesmo que visual, da validade dos resultados pelo

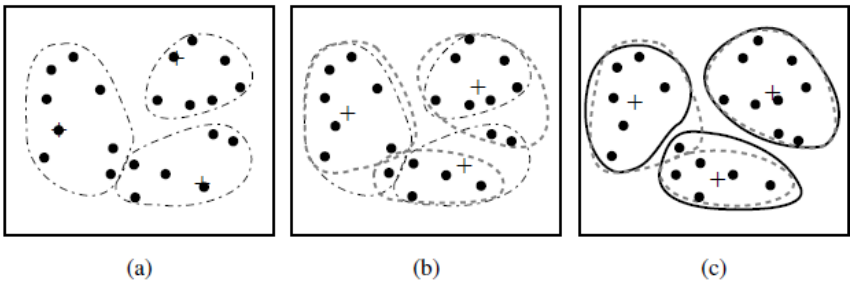

Figura 2.3: Agrupamento por *k-means*: em (a) o resultado da primeira iteração do algoritmo, em (b) o resultado da segunda iteração, e em (c) o resultado final do exemplo [Han e Kamber 2006].

analista humano (isto é, o quão bem o algoritmo de agrupamento descobriu os grupos do conjunto de dados). A visualização do conjunto de dados é uma verificação crucial do resultado do agrupamento. A visualização de um conjunto de dados torna-se difícil quando este possui dimensão maior do que três. Além disso, a percepção de grupos usando ferramentas de visualização disponíveis é uma tarefa difícil para os humanos, que não estão acostumados a espaços de maior dimensão. Cada algoritmo de agrupamento funciona de forma diferente dependendo:

i) das características do conjunto de dados (geometria e densidade de distribuição dos grupos);

ii) dos valores dos parâmetros de entrada [Halkidi et al. 2001].

Na literatura de AD, vários algoritmos têm sido propostos para aplicações diferentes e tamanhos diferentes de conjuntos dados. Entretanto, AD é uma tarefa de aprendizagem não supervisionada, ou seja, não existem classes pré-definidas ou exemplos que possam indicar se os grupos formados pelo algoritmo de agrupamento são válidos. Para avaliar os resultados de diferentes algoritmos de agrupamento ou avaliar os resultados de um mesmo algoritmo de agrupamento, mas com valores de parâmetros diferentes, torna-se necessário encontrar um meio de verificar a qualidade dos grupos formados. Essa tarefa é conhecida como validação de agrupamento ou validação de grupos, um dos assuntos mais importantes em AD e responsável pelo sucesso de suas aplicações. Encontrar o número ótimo de grupos existentes em um conjunto de dado, quando este não é fornecido ao algoritmo de agrupamento, também é uma tarefa, não trivial, da validação de agrupamento [Gan et al. 2007].

A validação de agrupamento é realizada através de medidas de qualidade dos grupos formados pelos algoritmos de AD, e são divididas em duas categorias principais: as base-

adas em critérios externos e as baseadas em critérios internos. A principal diferença entre essas categorias é se informação externa é usada para a validação do resultado do agrupamento ou não. Medidas baseadas em critérios externos avaliam tal resultado observando uma estrutura especificada previamente, que é imposta a um conjunto de dados e reflete sua estrutura intuitiva. É uma medida de concordância, de relacionamento, entre duas partições onde, a primeira partição é a estrutura de agrupamento conhecida previamente, e a segunda é o agrupamento resultante do algoritmo de AD. Por outro lado, medidas baseadas em critérios internos dependem somente de quantidades e características inerentes ao conjunto de dados, e são usadas para medir a qualidade de uma estrutura de agrupamento sem qualquer informação externa. [Gan et al. 2007] [Halkidi et al. 2002] [Liu et al. 2010].

Medidas baseadas em critérios externos são principalmente usadas para selecionar um algoritmo de AD mais adequado para um conjunto de dados específico. Por outro lado, medidas baseadas em critérios internos podem ser usados para escolher o melhor algoritmo de AD, como também o número ótimo de grupos sem qualquer informação adicional. Na prática, informações externas, tais como rótulo de classe, são geralmente indisponíveis em muitos cenários de aplicações. Portanto, nessas situações, somente medidas baseadas em critérios internos podem ser usadas para avaliar o resultado do algoritmo de AD [Liu et al. 2010]. Esse é o caso desta tese, e por isso apenas as medidas baseadas em critérios internos serão abordadas a partir deste ponto e doravante serão chamadas de índices de validação de agrupamento ou simplesmente índices de validação.

Os índices de validação de agrupamento são frequentemente baseadas nos dois critérios seguintes:

I. Compacidade

Pode ser entendido como o quão próximos (compactados) são os elementos em um grupo. Algumas índices avaliam a compacidade de grupo baseando-se na variância, outros índices estimam a compacidade de grupo baseando-se em distâncias, tais como a distância média ou máxima de pares de elementos ou dos elementos ao centro do grupo [Liu et al. 2010].

II. Separação

Pode ser entendido como o quão bem separado um grupo é dos outros. Nesse caso, pode ser usada a distância entre os respectivos centros (centroide) de dois grupos quaisquer ou a distância mínima entre pares de elementos em diferentes grupos. Medidas baseadas em densidade também são usadas em alguns índices [Liu et al. 2010].

Na literatura existem vários índices de validação, porém nesta tese foram aplicados apenas os direcionados a resultados de algoritmos de agrupamento rígido dos dados (cada elemento deve pertencer a um, e somente um, grupo), que é o caso do *k-means*, usado aqui para agrupar as séries temporais de temperatura. São eles: Silhueta; Davies-Bouldin; Calinski-Harabasz; Dunn; Krzanowski-Lai; Hartigan; Homogeneidade; Separação; Hubert-Levin (C-index) e Strehl. Cada um desses índices é aplicado com o propósito de buscar grupos compactos e bem separados, porém eles se diferenciam uns dos outros pela forma como fazem isso e pelas características que analisam nos grupos. As próximas seções descrevem esses índices.

Índice Silhueta

O índice Silhueta [Rousseeuw 1987] [Lewis et al. 2012] [Bolshakova e Azuaje 2003] é uma medida de compacidade e separação dos grupos formados. Além disso, ele pode ser usado com várias medidas de similaridade (proximidade). Para um determinado grupo $G_i \subset X$, onde X é um conjunto de dados, $(i = 1, 2, .., k)$ e k é o número de grupos formados em X, o índice Silhueta fornece para cada $x_{i,j} \in G_i$ a medida de qualidade $s(x_{i,j})$ $(j = 1, 2, ..., g_i$ e g_i é o número de elementos no grupo $G_i)$, conhecida como a *largura de Silhueta*. Essa medida indica a associação de $x_{i,j}$ ao grupo G_i e é definida pela Equação 2.7:

$$s(x_{i,j}) = \frac{b(x_{i,j}) - a(x_{i,j})}{max\{a(x_{i,j}), b(x_{i,j})\}} \tag{2.7}$$

onde $a(x_{i,j})$ é a distância média de $x_{i,j}$ para todos $x_{i,p}$, $p \neq i$, no grupo G_i, $b(x_{i,j})$ é a distância média mínima entre $x_{i,j}$ e todos $x_{l,j} \in G_l$, $(l = 1, ..., k; l \neq i)$, e max é o operador de máximo. Note que $-1 \leq s(x_{i,j}) \leq 1$, para cada elemento $x_{i,j}$.

Se $s(x_{i,j})$ é próximo de (ou igual a) 1, então o elemento $x_{i,j}$ foi "bem-agrupado", *i.e.*, ele foi atribuído a um grupo apropriado. Se $s(x_{i,j})$ é próximo de zero, então $x_{i,j}$ poderia também ser associado ao grupo vizinho mais próximo. Se $s(x_{i,j})$ é próximo de (ou igual a) -1, então $x_{i,j}$ foi "mal-agrupado". Assim, para um dado grupo G_i, é possível calcular o valor $S(G_i)$ pela Equação 2.8, o qual caracteriza as propriedades de heterogeneidade e isolamento de tal grupo.

$$S(G_i) = \frac{1}{g_i} \sum_{j=1}^{g_i} s(x_{i,j}) \tag{2.8}$$

O valor global S do índice Silhueta pode ser usado como um índice de validação de agrupamento do conjunto X

$$S = \frac{1}{k}\sum_{i=1}^{k} S(G_i) \tag{2.9}$$

A Equação 2.9 pode ser aplicada para estimar o número de grupos mais apropriado para X. Nesse caso, o agrupamento com valor máximo de S é considerado o agrupamento ótimo.

Índice Davies-Bouldin

O índice Davies-Bouldin DB [Davies e Bouldin 1979] [Bolshakova e Azuaje 2003] [Gan et al. 2007] é uma relação entre a soma dos valores de dispersão *intragrupo* e a separação *intergrupo*. Então, para se definir esse índice, é necessário a escolha de uma medida de dispersão dos elementos dentro de um grupo, e de uma medida de similaridade entre grupos. A medida de dispersão D de um grupo G deve satisfazer as seguintes propriedades:

1. $D \geq 0$;
2. $D = 0$ se e somente se $x = y$, $\forall x, y \in G$.

A medida D_i de um grupo G_i pode ser definida, por exemplo, pela Equação 2.10:

$$D_i = \left(\frac{1}{g_i}\sum_{x \in G_i} d^p(x, c_i)\right)^{\frac{1}{p}}, \; p > 0, \tag{2.10}$$

onde g_i é o número de elementos em G_i, c_i é o centro (ou o elemento representativo de G_i) e $d^p(x, c_i)$ é a distância entre x e c_i.

A medida de similaridade de grupo $S_{i,j}$ entre os grupos G_i e G_j é definida baseando-se em suas medidas de dispersão D_i e D_j e satisfaz as seguintes condições:

1. $S_{i,j} \geq 0$;
2. $S_{i,j} = S_{j,i}$;
3. $S_{i,j} = 0$ se, e somente se, $D_i = D_j$;
4. se $D_j = D_k$ e $d_{G_i,G_j} < d_{G_i,G_k}$, então $S_{i,j} > S_{i,k}$;
5. se $D_j > D_k$ e $d_{G_i,G_j} = d_{G_i,G_k}$, então $S_{i,j} > S_{i,k}$;

Aqui, D_i, D_j e D_k são as medidas de dispersão dos grupos G_i, G_j e G_k, respectivamente, e d_{G_i,G_j} é a distância (medida de similaridade) entre os dois grupos G_i e G_j, que pode ser considerada como a distância entre seus respectivos centroides e definida pela Equação 2.11:

$$d_{G_i,G_j} = \left(\sum_{l=1}^{d} |c_{il} - c_{jl}|^t \right)^{\frac{1}{t}} \tag{2.11}$$

onde $t > 1$ e d é a dimensão dos pontos de dados em cada grupo. Uma escolha para $S_{i,j}$ seria:

$$S_{i,j} = \frac{D_i + D_j}{d_{G_i,G_j}} \tag{2.12}$$

Logo, o índice DB é definido como:

$$DB = \frac{1}{k} \sum_{i=1}^{k} R_i \tag{2.13}$$

onde k é o número de grupos formados e S_i é definido como:

$$S_i = \max_{j \neq i} S_{ij} \tag{2.14}$$

É importante lembrar que p na Equação 2.10 e t na Equação 2.11 podem ser escolhidos independentemente um do outro. Valores baixos de DB correspondem a grupos que são compactos e cujos respectivos centros são distantes uns dos outros. Portanto, o valor de k que minimiza $DB \in [0, \infty]$ é considerado como o número de grupos ótimo. Visto que as matrizes de dispersão dependem da geometria dos grupos, este índice possui uma base lógica estatística e geométrica.

Índice Caliński-Harabasz

O índice de validação Caliński-Harabasz CH [Caliński e Harabasz 1974] [Gan et al. 2007] [Dudoit e Fridlyand 2002] [Kozak 2012] [Torres et al. 2011] é definido com base nos traços das matrizes de dispersão *intergrupo* e *intragrupo*. Calculado para cada solução de agrupamento possível, o valor de índice máximo alcançado indica o melhor agrupamento dos dados. Seja n o número de elementos no conjunto de dados X e k o número de grupos. Então esse índice pode ser calculado pela Equação 2.15:

$$CH(P_k) = \frac{(n-k)Tr(B)}{(k-1)Tr(W)}, \tag{2.15}$$

onde $Tr(B)$ e $Tr(W)$ são os traços das matrizes B e W, respectivamente. A matriz B contém o quadrado das distâncias entre todos os pares de grupos (*intergrupo*) e a matriz W contém o quadrado das distâncias de todos os elementos dentro de um grupo a seu respectivo centroide (*intragrupo*). Os traços de B e de W estão definidos nas equações:

$$Tr(B) = \sum_{i=1}^{k} |G_i|(z_i - z)^T(z_i - z) \tag{2.16}$$

$$Tr(W) = \sum_{i=1}^{k} \sum_{x \in G_i} (x - z_i)^T(x - z_i), \tag{2.17}$$

onde z e z_i são as médias X e de G_i, respectivamente.

Índice Dunn

O índice Dunn DN [Gan et al. 2007] [Bolshakova e Azuaje 2003] [Lewis et al. 2012] busca identificar conjuntos de grupos compactos e bem separados. Ele é definido pela Equação 2.18:

$$DN = \min_{1 \leq i \leq k} \left\{ \min_{i+1 \leq j \leq k, j \neq i} \left\{ \frac{D(G_i, G_j)}{\max_{1 \leq l \leq k} \{Diam(G_l)\}} \right\} \right\} \tag{2.18}$$

onde k é o número de grupos, $D(G_i, G_j)$ (Equação 2.19) é a distância entre os grupos G_i e G_j, e $Diam(G_l)$ (Equação 2.20) é o diâmetro do grupo G_l.

$$D(G_i, G_j) = \min_{1 \leq i \leq k} d(x, y) \tag{2.19}$$

$$Diam(G_l) = \max_{x,y \in G_l} d(x, y) \tag{2.20}$$

Valores elevados de DN correspondem a agrupamentos de boa qualidade, pois se um conjunto de dados X possui grupos bem separados e compactos, a distância entre grupos será grande e o diâmetro de cada grupo será de valor baixo. Portanto, o valor de k que maximiza $DN \in [0, \infty]$ é tomado como o número ótimo de grupos. As desvantagens desse índice são a sensibilidade alta à presença de ruído e custo computacional alto.

Índice Krzanowski-Lai

Supondo que X seja um conjunto de dados com n elementos e que cada $x \in X$ seja da forma $x = (v_1, v_2, ..., v_p)$. O índice Krzanowski-Lai, KL, [Krzanowski e Lai 1988] é baseado na soma dos quadrados $S = Tr(W_k)$, onde $Tr(W_k)$ é o traço da matriz W_k, que é

a matriz de covariância para um dado agrupamento de X em k grupos. A diferença entre o agrupamento dos dados do conjunto X em k grupos e em $k-1$ grupos é definida pela Equação 2.21:

$$diff_k = (k-1)^{2/p} Tr(W_{k-1}) - k^{2/p} Tr(W_k) \tag{2.21}$$

A introdução da normalização do fator $2/p$ é derivada da observação de que os valores das variáveis de x são assumidas como independentes e uniformemente distribuídas, o agrupamento ótimo dos dados irá reduzir a soma dos quadrados exatamente por esse valor. Os autores afirmam que se existe um solução ótima para o agrupamento em k^* grupos, o valor de $diff_{k^*}$ deverá ser comparativamente grande e positivo. Em contraste, todos os valores de $diff_k$ para $k > k^*$ terão valores bem baixos (até negativos), enquanto que valores para $k < k^*$ serão altos e positivos. Juntando essas informações, o índice KL é definido como:

$$KL = |diff_k| / |diff_{k+1}| \tag{2.22}$$

O número estimado de grupos é o $\max_{2 \leq k}\{KL_k\}$.

Índice Hartigan

O índice Hartigan H [Hartigan 1985] [Dudoit e Fridlyand 2002] possui custo computacional relativamente baixo. Ele pode ser definido pela Equação 2.23:

$$H = (n-k-1)\frac{W(k) - W(k+1)}{W(k+1)} \tag{2.23}$$

onde

$$W(k) = \sum_{i=1}^{k} \sum_{j=1, j \in G_i}^{n} d^2(x_j, c_i) \tag{2.24}$$

Na Equação 2.23, n é número de elementos do conjunto de dados X em análise, W é a soma da distância quadrada de todos os elementos de X ao seus respectivos centroides de grupo e k é o número de agrupamentos. Na Equação 2.24, d é uma medida de distância e c_i é o centroide do grupo G_i. Visto que $W(k)$ é monotonicamente não crescente quando k aumenta, a razão H torna-se uma medida relativa da redução de erro quadrado quando o número de grupos aumenta de k para $k+1$. O termo multiplicador de correção $(n-k-1)$ é um fator de penalidade quando o número de grupos é alto. O valor ótimo de k é o que maximiza H.

Índice Homogeneidade

Seja X um conjunto com n elementos e $G = (G_1, G_2, ..., G_k)$ o resultado do agrupamentos dos elementos de X. A Homogeneidade [Chen et al. 2002] [Sharan et al. 2003] de G é a similaridade média entre pares de elementos de um mesmo grupo G_i, $i = 1, ..., k$ e é definida pela Equação 2.25:

$$H_{avg} = \frac{1}{M} \sum_{x \in G_i, y \in G_j, i \neq j} d(x, y) \tag{2.25}$$

onde M é o número total de pares de elementos de X dentro de um mesmo grupo G_i e d é uma medida de similaridade. O valor máximo de H_{avg} é uma solução ótima para o agrupamento do conjunto X em análise.

Índice Separação

Supondo X um conjunto com n elementos e $G = (G_1, G_2, ..., G_k)$ o resultado do agrupamentos dos elementos de X. O índice de separação [Chen et al. 2002] [Sharan et al. 2003] quantifica a separação dos grupos em G e é definido como a similaridade média entre pares de elementos de diferentes grupos:

$$S_{avg} = \frac{2}{n(n-1) - 2M} \sum_{x \in G_i, y \in G_j, i \neq j} d(x, y) \tag{2.26}$$

Assim como na Equação 2.25, M é o número total de pares de elementos de X dentro de um mesmo grupo G_i e d é uma medida de similaridade. O valor mínimo de S_{avg} é uma solução ótima para o agrupamento do conjunto X em análise.

Índice Hubert-Levin (C-índice)

O índice de validação Hubert-Levin [Bolshakova e Azuaje 2006] é apropriado se os grupos possuem tamanhos similares em relação ao número de elementos. Ele é definido pela equação:

$$C = \frac{S - S_{min}}{S_{max} - S_{min}} \tag{2.27}$$

onde S é a soma das distâncias de todos os pares de elementos de um grupo. Seja l o número desses pares, então S_{min} é a soma das distâncias menores l se todos os pares são considerados. Similarmente, S_{max} é a soma das distâncias maiores l fora de todos os pares. É fácil notar que o numerador na Equação 2.27 será menor se pares de elementos mais próximos estão em um mesmo grupo. Portanto, um pequeno valor de $C(P_k)$ indica

um resultado de agrupamento bom. O denominador possui a função de normalização e assim $C(P_k) \in [0, 1]$.

Índice Strehl

O índice Strehl [Strehl e Ghosh 2000] [Ghosh e Strehl 2002] [Jing et al. 2010] utiliza as similaridades intragrupo e intergrupo para medir a qualidade dos k grupos formados com os n elementos de um conjunto de dados X. A similaridade intragrupo representa a similaridade média entre os elementos de um grupo $G_i \subset G$ (G é o resultado do agrupamento dos elementos de X e $i \in 1, 2, ..., k$) e é definida pela Equação 2.28:

$$intra(G_i) = \frac{1}{(g_i - 1)g_i} \sum_{x,y \in G_i} d(x, y) \tag{2.28}$$

onde g_i é o número de elementos do grupo G_i e $d(x, y)$ é a distância entre dois elementos de G_i. Essa equação não define a similaridade intragrupo para grupos vazios ou com um único elemento. Por definição, a similaridade intragrupo para grupo com um único elemento é igual a 1 (um).

A similaridade intergrupo entre grupos não vazios G_i e G_j, com $i \neq j$ é definida pela Equação 2.29:

$$inter(G_i, G_j) = \frac{1}{g_i g_j} \sum_{x \in G_i, y \in G_j} d(x, y) \tag{2.29}$$

onde g_i e g_j é o número de elementos do grupo G_i e do grupo G_j, respectivamente.

O objetivo é maximizar a similaridade intragrupo e minimizar a similaridade intergrupo. Assim, a qualidade de agrupamento G é definida como a razão entre o total das similaridades intragrupo e o total das similaridades intergrupo. O total das similaridades intragrupo é o somatório de cada uma das similaridades intragrupo ponderada por $g_i - 1$, pois a auto-similaridade (o qual é sempre 1) não deve ser contabilizada. O total das similaridades intergrupo é obtido pelo somatório de cada similaridade $inter(G_i, G_j)$ ponderada por g_i.

$$\frac{\sum_{i=1}^{k} \frac{g_i - 1}{n - k} intra(G_i)}{\sum_{i=1}^{k} \sum_{j=i+1}^{k} \frac{g_i}{n} inter(G_i, G_j)} \tag{2.30}$$

O valor da razão 2.30 está dentro do intervalo [0, ∞).

Sendo G o resultado de um agrupamento dos elementos em X, a qualidade de G é definida pela Equação 2.31:

$$\Gamma(G) = 1 - \frac{(n-k)\sum_{i=1}^{k}\sum_{j=i+1}^{k} g_i\, inter(G_i, G_j)}{n\sum_{i=1}^{k}(g_i - 1)\, intra(G_i)} \tag{2.31}$$

onde $\Gamma(G) \in [0, 1]$

2.4.2 Aprendizagem de máquina supervisionada

Na abordagem supervisionada ou preditiva da aprendizagem de máquina, a tarefa é aprender um mapeamento de entradas x para saídas y, dado um conjunto rotulado de pares de entrada-saída $T(x_i, y_i)\ \ 1 \leq i \leq$, onde T é o conjunto de treinamento e n é o número de exemplos de treinamento. Resumidamente, x_i é um vetor de números D-dimensional, representando as características ou atributos de algo como uma imagem ou uma série temporal. Similarmente, y_i é, quase sempre, uma variável categórica ou nominal de algum conjunto finito. Quando y_i é categórico, o problema é conhecido como classificação ou reconhecimento de padrões, e quando y_i é um valor real, o problema é chamado de regressão [Murphy 2012].

2.4.2.1 Seleção de características

Muitas técnicas de aprendizagem de máquina não são projetadas para lidar com uma grande quantidade de características irrelevantes e por esse motivo a seleção de características torna-se uma etapa importante em tarefas de classificação, para que o modelo de classificação seja construído. Os objetivos da seleção de características são múltiplas, as mais importantes são: evitar *overfitting* e melhorar o desempenho do modelo de classificação construído.

Os métodos de seleção de características podem ser organizados em três categorias, dependendo de como eles combinam a busca das características relevantes com a construção do modelo de classificação, são eles: métodos de filtro, métodos *wrapper* e métodos embutidos. Métodos de filtro avaliam a relevância das características observando apenas as propriedades intrínsecas do conjunto de dados. Na maioria dos casos, a pontuação de relevância de cada características é calculada e as características com as mais baixas pontuações são removidas. Em seguida, o subconjunto com as características que permaneceram é apresentado como entrada ao algoritmo de classificação. As vantagens desses métodos são a facilidade de lidar com conjuntos de dados de alta dimensão e o custo

computacional baixo. Além disse, a seleção de características é executada uma única vez e depois disso, vários algoritmos de classificação podem ser executados com o mesmo subconjunto de características selecionadas. A desvantagem desses métodos é desconsiderar uma iteração com o classificador. Não como os métodos de filtro, os métodos *wrapper* consideram a iteração com o classificador, tornando-os uma abordagem sob medida para um algoritmo de classificação específico. A terceira classe de métodos de seleção de características, os métodos embutidos, são bem semelhantes aos métodos *wrapper*, por considerarem a iteração com o classificador, e por isso o subconjunto de características selecionadas é apropriado a um classificador específico, mas são menos computacionalmente custosos, por estarem embutidos na construção do classificador [Saeys et al. 2007].

2.4.2.2 Classificação

A tarefa de classificação de dados consiste em tentar "aprender" o relacionamento existente entre um conjunto de variáveis de características e uma variável "rótulo". Visto que muitos problemas práticos podem ser expressados como associações entre variáveis de características e variáveis "rótulos", a classificação de dados encontra uma variedade grande de aplicações. O problema de classificação pode ser definido como: *dado um conjunto de dados de treinamento juntamente com os "rótulos" de treinamento associados, determinar o "rótulo" de classe para um dado de teste não rotulado* [Aggarwal 2014].

Considere uma base de dados onde cada elemento X (no contexto da classificação também referenciado como tupla, amostra, exemplo, instância, vetor de dado, ou objeto) seja representado por um vetor de características (ou atributos) *n-dimensional*, ou seja, $X = (x_1, x_2, ...x_n)$, onde x_i é o valor da característica C_i, $1 \leq i \leq n$, calculada sobre os elementos da base de dados. Assume-se que cada elemento X pertença a uma classe pré-definida e isso é indicado pelo atributo "rótulo" de classe. Esse atributo é um valor discreto e não ordenado. Ele é categórico no sentido de que cada valor indica uma categoria ou classe [Han e Kamber 2006].

Classificação de dados é um processo realizado em duas etapas. Na primeira etapa, que é a etapa de aprendizagem (ou fase de treinamento), um classificador (modelo) é construído descrevendo um conjunto pré-determinado de classes ou conceitos. O algoritmo responsável por construir o classificador analisa (ou "aprende" de) um subconjunto da base de dados chamado de conjunto de treinamento. Os elementos desse conjunto, chamados de elementos de treinamento, já possuem um rótulo de classe associado, ou seja, a classe a que cada um pertence já é conhecida. Esta etapa é também conhecida como aprendizagem

supervisionada, pois o rótulo da classe de cada elemento de treinamento é fornecido ao classificador. Diferentemente, na aprendizagem não supervisionada, o rótulo de classe de cada elemento de treinamento não é conhecido, e o número ou o conjunto de classes a ser aprendido pode não ser conhecido antecipadamente também. Por exemplo, se a informação da decisão não está disponível para o treinamento, é possível usar AD para tentar determinar as classes. Esta primeira etapa do processo de classificação também pode ser vista como um mapeamento ou como a determinação de uma função, $y = f(X)$, que pode predizer a classe associada y de um dado elemento X. Por isso é desejável "aprender" um mapeamento (ou determinar uma função) que separe as classes de dados. Tipicamente, este mapeamento é representado na forma de regras de classificação, árvores de decisão, ou funções discriminantes [Han e Kamber 2006].

Na segunda etapa, o modelo de classificação é usado para a tarefa de classificação. Inicialmente, a acurácia preditiva do classificador é estimada. Se o conjunto de treinamento for usado para medir a acurácia do classificador, o resultado será otimista, pois o classificador tende a excessivamente se ajustar aos dados (isto é, durante a aprendizagem ele pode incorporar algumas anomalias do conjunto de treinamento que não estão presentes no conjunto de dados total). Portanto é usado um conjunto de teste formado com elementos de teste e seus rótulos de classe associados. Esses elementos são selecionados aleatoriamente do conjunto total de dados e são independentes dos elementos de treinamento, ou seja, não devem ser usados para construir o classificador [Han e Kamber 2006].

A acurácia de um classificador em um dado conjunto de teste é a porcentagem de elementos do conjunto de teste que são corretamente classificados por ele. Para cada elemento de teste, o atributo classe é confrontado com a predição de classe realizada pelo classificador em análise. Se a acurácia de um classificador é considerada aceitável, o classificador pode ser usado para classificar casos (dados) futuros dos quais não se conhece a classe [Han e Kamber 2006].

A classificação pode necessitar de ser precedida por análise de relevância que, por meio de métodos específicos, busca identificar características que não contribuem para a classificação e excluí-las. Essa fase é conhecida como seleção de características ou redução de dimensionalidade [Han e Kamber 2006]. Existem vários métodos para a construção do modelo de classificação. As seções seguintes abordam os fundamentos dos classificadores testados na metodologia proposta e que são frequentemente usados em outras metodologias.

2.4.2.3 Seleção de algoritmos e otimização de parâmetros

A realização de testes para avaliar o desempenho de um determinado algoritmo de classificação sobre os dados de uma base implica fazer algumas escolhas que não são triviais para pessoas não especialistas em aprendizagem de máquina. O usuário deve escolher um método de seleção de características, um algoritmo de aprendizagem de máquina e os parâmetros para o método e para o algoritmo. Devido à grande quantidade de alternativas para cada uma dessas escolhas (vários métodos de seleção de características, muitos classificadores, assim como um número grande de valores de parâmetros para cada um desses classificadores), combinações entre elas geram um número elevado de possibilidades e quase sempre as escolhas realizadas por um usuário inexperiente não são as melhores possíveis. Para contornar esse problema, a ferramenta Auto-WEKA (*Waikato Environment for Knowledge Analysis*) [Thornton et al. 2013] foi utilizada para auxiliar nas escolhas realizadas.

O Auto-WEKA é uma ferramenta que, automaticamente e simultaneamente, seleciona um algoritmo de aprendizagem de máquina e define seus parâmetros para o usuário. Mais especificamente, dada uma base de dados, o Auto-WEKA explora configurações de parâmetros para vários algoritmos de aprendizagem e recomenda ao usuário o método que provavelmente terá um bom desempenho, usando modelos baseados em técnicas de otimização. Nele são consideradas algumas técnicas de seleção de características (combinando 3 métodos de pesquisa e 8 métodos de avaliação) e 39 classificadores implementados no WEKA [Hall et al. 2009], e configurações de parâmetros para cada uma dessas técnicas e classificadores [Thornton et al. 2013]. Em relação ao *WEKA*, ele é uma coleção abrangente de algoritmos de aprendizagem de máquina e ferramentas de processamento de dados para pesquisadores e profissionais da área. Ele permite experimentar e comparar vários métodos de aprendizagem de máquina em diferentes bases de dados. Entre eles estão algoritmos para regressão, classificação, agrupamento (particionamento), mineração de regras de associação e seleção de características [Hall et al. 2009].

Assim como Auto-Weka, o EMiner [Marques 2014] é uma ferramenta para a resolução do problema de seleção de algoritmos e otimização de parâmetros em tarefas de classificação, mais conhecido como CASH (*Combined Algorithm Selection and Hyperparameter Optimization Problem*). Porém a diferença entre eles está no método usado para a otimização dos parâmetros. Enquanto que o Auto-WEKA utiliza otimização Bayesiana, o EMiner utiliza Algoritmo Genético (AG).

O EMiner está dividido em três fases (Figura 2.4): a definição de valores iniciais, a

otimização de parâmetros e a seleção do algoritmo. Na primeira fase é definido o algoritmo de classificação que será testado e seus valores iniciais de parâmetros, que podem ser obtidos de uma base de dados pública, que contém um número grande de experimentos de mineração de dados, ou aleatoriamente dentro de limites pré-determinados. Esses valores de parâmetros são pré-processados e codificados em diferentes cromossomos para formar a população inicial do AG. A segunda fase é a de otimização dos parâmetros propriamente dito por meio do AG, que aplica os operadores genéticos de seleção, cruzamento e mutação, na população atual, para dar origem a uma população nova. Na última fase, um *rank* dos algoritmos de classificação é criado com base nos resultados da repetição das fases anteriores [Marques 2014].

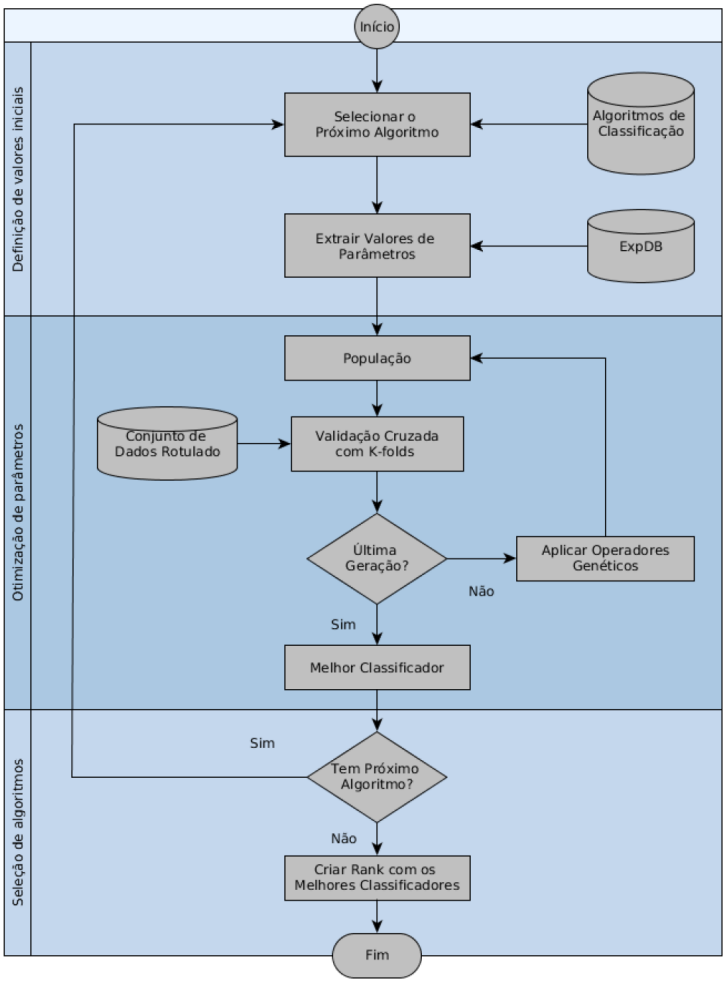

Figura 2.4: Metodologia implementada no EMiner [Marques 2014].

2.4.2.4 Avaliação dos resultados da classificação

Após usar dados do passado, já classificados, para treinar um classificador, é necessário conhecer o comportamento desse classificador na tarefa de classificar dados futuros, dos

quais não se conhece o rótulo de classe, ou seja, conhecer sua acurácia. Os próximos tópicos abordam vários conceitos geralmente envolvidos na avaliação de um classificador.

Treinamento e teste

Para o problema de classificação é natural verificar o desempenho do classificador em termos da taxa de erros. O classificador determina a classe de cada instância: se a classificação está correta, conta-se como um acerto, se não, conta-se como um erro. A taxa de erro é justamente a proporção de erros cometidos sobre o conjunto inteiro de instâncias, e mede o desempenho global do classificador [Witten et al. 2011].

O interesse é conhecer o desempenho do classificador em dados futuros e não nos dados usados na fase de treinamento, pois esses dados já estão classificados, e é exatamente por esse fato é que são usados como dados de treinamento. Além disso, a taxa de erro sobre o conjunto de treinamento é provavelmente um indicador ruim do desempenho do classificador na tarefa de classificar dados não classificados ainda e qualquer estimativa do desempenho baseado no conjunto treinamento será extremamente otimista devido a superespecialização do algoritmo de aprendizagem em relação a esses dados [Witten et al. 2011].

Para se fazer a previsão do desempenho do classificador em dados futuros (dos quais não se conhece o rótulo da classe), é necessário calcular a taxa de erro sobre um conjunto de dados que não participou da etapa de aprendizagem. Este conjunto independente é chamado de conjunto de teste. É importante que os dados de teste não tenham participado na formação do classificador. Tanto os dados de treinamento como os dados de testes devem ser dados representativos do problema em questão [Witten et al. 2011].

Quando existe uma quantidade grande de dados disponível, toma-se uma parte (suficientemente grande) para a fase de treinamento e o restante é usado para a fase de teste. Sendo os dados usados em ambas as fases representativos do problema, a taxa de erro sobre o conjunto de teste fornece uma indicação boa do classificador para dados futuros. Geralmente, quanto mais dados são usados na fase de treinamento, melhor é o classificador construído, apesar de isso não ser verdade quando determinada quantidade de dados para treinamento é ultrapassada. E quanto maior o conjunto de teste usado, mais precisa será a acurácia desse classificador [Witten et al. 2011].

O problema surge quando não existe uma vasta quantidade de dados disponível. Em muitas situações, os dados para o treinamento são classificados manualmente, como também os dados usados na fase de teste. Isso limita a quantidade de dados que pode ser

usada para treinamento, validação e teste. O problema é como aproveitar ao máximo uma base de dados limitada. Dessa base de dados, uma quantidade é mantida para teste. Isto é chamado de procedimento de validação, e o restante é usado para treinamento (e, se necessário, parte disso é reservado para a validação). Assim, o dilema é: para se construir um bom modelo de classificação, o conjunto de treinamento deve ser o maior possível, por outro lado, para se ter uma estimativa precisa do erro, é necessário que o conjunto de teste seja o maior possível [Witten et al. 2011].

Medidas de acurácia do classificador

Como já comentado, a acurácia é melhor estimada sobre um conjunto de teste constituído de instâncias das quais se conhece as respectivas classes, mas que não foram usadas para treinar o modelo. A acurácia de um classificador sobre um determinado conjunto de teste é a porcentagem de instâncias do conjunto de teste que são corretamente classificadas pelo classificador. Na literatura de reconhecimento de padrões, isto é também conhecido como a taxa de reconhecimento global do classificador, isto é, reflete quão bem ele reconhece as instâncias de várias classes [Han e Kamber 2006].

Considerando duas classes, dois conceitos válidos são: instâncias positivas (instâncias da classe principal de interesse, por exemplo, **câncer de mama = sim**) e instâncias negativas (por exemplo, **câncer de mama = não**). *Verdadeiros positivos* referem-se às instâncias positivas que foram classificadas (rotuladas) pelo classificador corretamente (por exemplo, instâncias da classe **câncer de mama = sim** que foram corretamente classificadas como tais), enquanto que *verdadeiros negativos* são instâncias negativas que foram classificadas corretamente pelo classificador (por exemplo, instâncias da classe **câncer de mama = não** que foram corretamente classificadas como tais). *Falsos positivos* são instâncias negativas que foram classificadas incorretamente (por exemplo, instâncias da classe **câncer de mama = não**, mas que o classificador as rotulou como **câncer de mama = sim**). Similarmente, *falsos negativos* são instâncias positivas que foram classificadas incorretamente (por exemplo, instâncias da classe **câncer de mama = sim**, que o classificador as rotulou como **câncer de mama = não**) [Han e Kamber 2006].

A matriz de confusão é uma ferramenta para analisar o quão bem o classificador pode reconhecer instâncias de diferentes classes. Uma matriz de confusão para duas classes é mostrada na Figura 2.5. Considerando m classes, uma matriz de confusão é uma tabela de tamanho m por m, pelo menos. Uma entrada $CM_{i,j}$ indica a quantidade de

instâncias da classe i que foram classificadas pelo classificador como sendo da classe j. Para que um classificador tenha uma acurácia boa, idealmente a maioria das instâncias devem estar representadas ao longo da diagonal principal da matriz, ou seja, nas entradas $CM_{1,1}, CM_{2,2}, ..., CM_{m,m}$, com o restante das entradas próximas de zero. A matriz pode ter linhas e/ou colunas adicionais contendo totais ou taxas de reconhecimento por classe [Han e Kamber 2006].

Matriz de confusão	Classes	
	Com câncer	Saudável
Classificado como com câncer de mama	10	1
Classificado como saudável	1	10
Total de instâncias classificadas corretamente	20	90.91%

Figura 2.5: Matriz de confusão para duas classes.

Existem situações que apresentam necessidade de medidas de acurácia alternativas. Suponha que um classificador foi treinado para classificar instâncias de dados médicos como **câncer de mama = sim** ou **câncer de mama = não**. Se esse classificador apresenta uma taxa de acurácia de 90%, ele pode parecer muito preciso. Entretanto, se apenas 3-4% das instâncias no conjunto de teste são da classe **câncer de mama = sim**, a taxa de 90% pode não ser aceitável, pois o classificador poderia ter rotulado corretamente somente as instâncias da classe **câncer de mama = não**, por exemplo. Assim, torna-se interessante saber o quão bem o classificador pode reconhecer instâncias da classe **câncer de mama = sim** (as instâncias positivas) e o quão bem ele pode reconhecer instâncias da classe **câncer de mama = não** (as instâncias negativas). A sensibilidade (denotada por SSB na Equação 2.32) e a especificidade (denotada por ESP na Equação 2.33) podem ser usadas, respectivamente, para esse propósito. A sensibilidade é também conhecida como a taxa de reconhecimento de *verdadeiros positivos* (isto é, a proporção de instâncias positivas que são corretamente identificadas), enquanto que a especificidade como a taxa de reconhecimento de *falsos negativos* (isto é, a proporção de instâncias negativas que são corretamente identificadas). Além disso, é possível usar a precisão (denotada por PCS na Equação 2.34) para obter a porcentagem de instâncias rotuladas como câncer de mama que de fato são instâncias da classe **câncer de mama = sim**. Essas medidas são definidas como:

$$SSB = \frac{VP}{P} \tag{2.32}$$

$$ESP = \frac{VN}{N} \tag{2.33}$$

$$PCS = \frac{VP}{VP + FP} \tag{2.34}$$

onde VP é o número de *verdadeiros positivos*, P é o número de instâncias positivas, VN é o número de *verdadeiros negativos*, N é o número de instâncias negativas, e FP é o número de *falsos positivos* [Han e Kamber 2006]. É possível definir a acurácia (denotada por ACR na Equação 2.35) em função da sensibilidade e da especificidade [Han e Kamber 2006]:

$$ACR = SSB\frac{P}{P+N} + ESP\frac{N}{P+N} \tag{2.35}$$

k-fold Cross Validation

O que fazer quando a quantidade de dados para treinamento e teste do classificador é limitada? O método *holdout* reserva uma certa quantidade para a fase de teste e usa o restante para a fase de treinamento (e define parte dessa reserva para validação, se requerido). Em termos práticos, é comum deixar fora um terço do conjunto de dados para teste e usar o restante (dois terços) para treinamento [Witten et al. 2011].

Pode acontecer que a amostra dos dados usada para treinamento (ou teste) não seja representativa. Geralmente, não é possível saber se uma amostra é representativa ou não. Mas é interessante verificar se cada classe na base de dados inteira está proporcionalmente representada nos conjuntos de treinamento e de teste. Se por acaso, todas as instâncias de uma determinada classe ficarem fora do conjunto de treinamento, dificilmente o classificador terá um bom desempenho para instâncias dessa classe. E a situação se agravaria pelo fato de que essa classe ficaria mais representada do que as outras no conjunto de teste, pois nenhuma de suas instâncias entrou na fase de treinamento. O procedimento que garante que cada classe seja corretamente representada no conjunto de treinamento e de teste, por amostras aleatoriamente formadas, é conhecido como estratificação ou *holdout* estratificado. Ele fornece uma proteção contra representações desiguais em ambos os conjuntos, de treinamento e de teste [Witten et al. 2011].

O caminho mais geral para eliminar qualquer influência nos resultados causados por uma determinada amostra escolhida aleatoriamente para *holdout* é repetir o processo inteiro, treinamento e teste, várias vezes com amostras aleatórias diferentes. Em cada interação, uma certa quantia, por exemplo, dois terços, do conjunto de dados é selecionada aleatoriamente para o treinamento, possivelmente com estratificação, e o restante é usado para teste. A taxa de erro global é a média das taxas de erros em diferentes iterações. Este é o método de *holdout* repetido para estimação da taxa de erro [Witten et al. 2011].

Em um procedimento *holdout* simples, é possível trocar as funções dos conjuntos de treinamento e de teste, isto é, treinar o classificador com o conjunto de teste e testá-lo com o conjunto de treinamento e tirar a média dos dois resultados, reduzindo assim o efeito de uma representação desigual das classes nos conjuntos de treinamento e teste. Entretanto, isto só é possível com a divisão do conjunto de dados em duas partes iguais entre esses conjuntos, o que geralmente não é o ideal (é preferível usar mais da metade dos dados para treinamento, mesmo em detrimento do conjunto de teste). Porém, uma variante simples do *holdout* forma a base de uma técnica estatística importante chamada *cross-validation*. Nesta técnica, um número de pastas ou partições do conjunto de dados é fixado. Suponha que esse número seja três. Então o conjunto de dados é dividido em três partições aproximadamente iguais; uma de cada vez é usada para teste e as restantes para treinamento. Isto é, dois terços dos dados são usados para treinamento e um terço para teste, e repete-se o procedimento três vezes até que todas as instâncias tenham sido usadas exatamente uma vez na fase de teste. Esse exemplo é um *3-fold cross-validation*, e se a estratificação das classes é adotada (e geralmente é), recebe o nome de *3-fold cross-validation* estratificado [Witten et al. 2011].

Na determinação da taxa de erro de uma técnica de aprendizagem sobre uma base de dados, é comum aplicar *10-fold cross-validation* estratificado. O conjunto de dados é dividido aleatoriamente em 10 partes nas quais cada classe é representada na mesma proporção do conjunto inteiro, aproximadamente. Cada parte é deixada de fora a cada vez e o esquema de aprendizagem treina sobre as nove partes restantes. Então, a taxa de erro é calculada sobre o conjunto que ficou de fora. Assim, o procedimento de aprendizagem é executado 10 vezes, cada hora em um conjunto de treinamento diferente (que não é tão diferente assim dos demais). Finalmente, é calculada a média das 10 estimativas de erros para gerar uma estimativa global [Witten et al. 2011].

O valor de 10 partições é baseado nos resultados de testes exaustivos sobre inúmeras base de dados, com diferentes técnicas de aprendizagem. Os resultados mostraram que com essa quantidade obtém-se a melhor estimativa de erro. Algumas evidências teóricas também sustentam esta conclusão. Embora esses argumentos não sejam conclusivos e o debate continue entre pesquisadores de mineração de dados e de aprendizagem de máquina sobre o melhor esquema de avaliação, *10-fold cross-validation* tem se tornado o método mais usado em termos práticos. Resultados de outros testes mostraram que o uso da estratificação melhora ligeiramente os resultados do classificador avaliado. Assim, a técnica de avaliação padrão em situações onde os dados disponíveis são limitados é o *10-fold cross-validation* estratificado. É importante lembrar que nem a estratificação, nem a divisão em

10 partições são exatas. É suficiente dividir os dados em 10 conjuntos aproximadamente iguais e em cada uma dessas partições representar cada classe em uma mesma proporção, aproximadamente. Entretanto, *5-fold cross-validation* ou *20-fold cross-validation* podem apresentar-se tão bons quanto *10-fold cross-validation* [Witten et al. 2011].

Executar o *10-fold cross-validation* uma única vez pode não ser suficiente para obter-se uma estimativa confiável do erro. Mas executar o *10-fold cross-validation* algumas vezes com o mesmo esquema de aprendizagem e com a mesma base de dados produz, frequentemente, resultados diferentes por conta do efeito da variação aleatória na escolha dos elementos que compõem as pastas. A estratificação reduz a variação, mas certamente não a elimina totalmente. Na busca da estimativa de erro da acurácia, é prática frequente repetir o processo de *cross-validation* dez vezes, isto é, dez vezes *10-fold cross-validation* e determinar a média dos resultados. Porém isso implica executar o algoritmo de aprendizagem 100 vezes nos subconjuntos de dados, que são todos nove décimos do tamanho da base inteira. Assim, obter uma medida boa do desempenho é uma tarefa computacionalmente intensiva [Witten et al. 2011].

Leave-One-Out Cross-Validation

O *10-fold cross-validation* é o caminho mais comum para medir a taxa de erro de um esquema de aprendizagem em uma dada base de dados, mas existem outros métodos com o mesmo propósito [Witten et al. 2011]. *Leave-one-out cross-validation* é simplesmente um *n-fold cross-validation*, onde n é o número de instâncias na base de dados. Cada instância fica de fora por vez, e o esquema de aprendizagem é treinando em todas as instâncias remanescentes ($n - 1$ instâncias). Na fase de teste, o acerto ou o erro do classificador sobre a instância que ficou de fora é registrada e a estimativa final de erro é a média dos n resultados [Witten et al. 2011].

Esse procedimento é atrativo por duas razões: a maior quantidade possível de dados é usada para o treinamento, supostamente aumentando as chances de sucesso do classificador; e o processo é determinístico, ou seja, amostras não aleatórias são usadas. Se o procedimento for repetido k vezes, sempre dará o mesmo resultado. A desvantagem é o custo computacional alto, pois todo o procedimento de aprendizagem é repetido n vezes e isto torna-se inviável para bases de dados grandes. Entretanto, o *leave-one-out cross-validation* oferece a oportunidade de explorar ao máximo bases de dados pequenas, pois a estimativa de erro obtida é a mais precisa possível [Witten et al. 2011].

Área sob a curva ROC

Curvas ROC são ferramentas visuais úteis para comparar dois modelos de classificação. A sigla ROC é uma abreviação de *Receiver Operating Characteristic* (característica de operação do receptor). Curvas ROC vêm da teoria de detecção de sinais que foi desenvolvida durante a II Guerra Mundial para a análise de imagens de radares. Uma curva ROC mostra a relação de perda-e-ganho entre a taxa de *verdadeiros positivos* ou sensibilidade (proporção de instâncias positivas que são corretamente identificadas) e a taxa de *falsos positivos* (proporção de instâncias negativas que são identificadas como positivas incorretamente) para um dado modelo de classificação. Isto é, dado um problema de duas classes, ele permite a visualização da relação de perda-e-ganho entre a taxa de classificação correta das instâncias positivas versus a taxa de classificação das instâncias negativas classificadas como positivas, pelo classificador, para as diferentes porções do conjunto de teste. Qualquer aumento na taxa de *verdadeiros positivos* ocorre à custa do aumento da taxa de *falsos positivos*. A área sob a curva ROC é uma medida da acurácia do modelo [Han e Kamber 2006].

O modelo de classificação deve ter a capacidade de determinar a probabilidade ou a ordenação para a classe predita de cada instância de teste para que a curva ROC seja desenhada. Ou seja, é necessário ordenar as instâncias de teste em ordem decrescente, de tal forma que fique no topo dessa lista as instâncias que têm as maiores chances, segundo o classificador, de pertencerem à classe das instâncias positivas. Classificadores Naive Bayes e de retropropagação são apropriados para este fim, outros classificadores tais como Árvore de Decisão podem ser modificados facilmente para fornecerem a distribuição de probabilidade de classe para cada predição. O eixo vertical da curva ROC representa a taxa de *verdadeiros positivos*. O eixo horizontal representa a taxa de *falsos positivos*. Uma curva ROC para um modelo de classificação é desenhada a partir do canto inferior esquerdo (onde a taxa de *verdadeiros positivos* e a taxa de *falsos positivos* são ambas iguais a zero), verifica-se o rótulo de classe da instância no topo da lista. Se for um *verdadeiro positivo* (isto é, uma instância positiva que foi corretamente classificada), então, na curva ROC, é realizado um movimento para cima e um ponto é marcado. Se, caso contrário, a instância pertence à classe negativa, tem-se um *falso positivo* e na curva ROC é realizado um movimento para a direita e um ponto é marcado. Esse procedimento é repetido para todas as instâncias do conjunto de teste, ou seja, realizando um movimento para cima na curva para um *verdadeiro positivo* ou para a direita para um *falso positivo* [Han e Kamber 2006].

A Figura 2.6 mostra um gráfico com a curva ROC de dois modelos de classificação. O gráfico mostra também uma linha diagonal onde para cada verdadeiro positivo de um dos modelos, implica a mesma probabilidade de se encontrar um falso positivo. Assim, o modelo para o qual a respectiva curva ROC encontra-se mais próxima da linha diagonal é o de menor acurácia. Se o modelo é realmente bom, torna-se mais provável encontrar *verdadeiros positivos* a medida que se percorre a lista ordenada para baixo, logo no início. Nesta fase, a curva se afasta rapidamente do zero. Depois da parte inicial da lista cada vez menos *verdadeiros positivos* são encontrados e cada vez mais *falsos positivos* são encontrados e a curva torna-se mais horizontal [Han e Kamber 2006].

Para se obter a acurácia de um modelo de classificação, a área sob a curva ROC é verificada. Quanto mais esta área se aproxima de 0,5 u.a. (unidade de área), menor é a acurácia do modelo. Para um modelo com acurácia perfeita tal área terá medida igual a 1. Vários pacotes de softwares são capazes de realizar esse cálculo [Han e Kamber 2006].

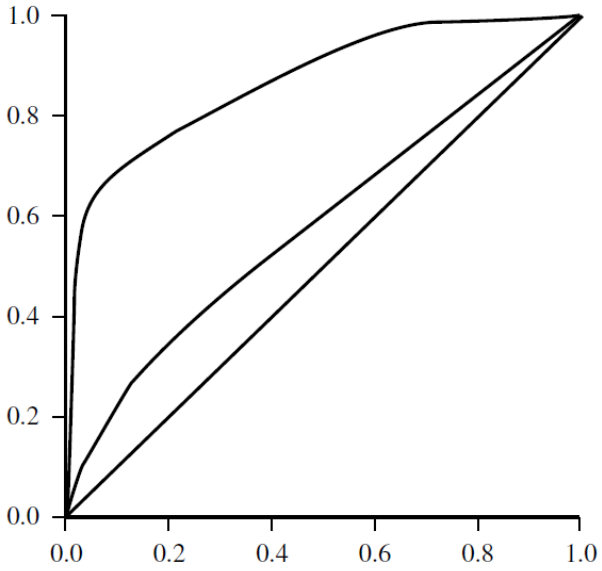

Figura 2.6: Curvas ROC de dois classificadores [Han e Kamber 2006].

2.5 Séries temporais

Séries temporais constituem a forma mais simples de dado temporal. Precisamente, uma série temporal é uma sequência de números reais que representam as medições de uma variável real em intervalos de tempo iguais. Exemplos de séries temporais são: variação dos preços das ações, a temperatura em um determinado lugar, e o volume de vendas. Uma série temporal é discreta se a variável observada é definida em um conjunto finito

de pontos no tempo. A maioria das séries temporais encontradas na análise de grupos (Seção 2.4.1.1) são discretas. Por outro lado, uma série temporal é contínua se a variável observada é definida para todos os pontos no tempo [Gan et al. 2007].

A tarefa de agrupar as séries temporais de um determinado conjunto é encontrar grupos, formados por essas séries, baseados na similaridade entre elas. Desta forma, medidas de similaridade entre séries temporais tornam-se importantes. Como medidas de similaridade (distância) para outros tipos de dados, uma medida de similaridade para séries temporais x e y, denotada por $d(x, y)$, deve obedecer aos seguintes axiomas:

1. $d(x, y) = d(y, x)$ (simetria);
2. $d(x, x) = 0$ (auto-similaridade);
3. $d(x, y) \geq 0$ e $d(x, y) = 0$ se, e somente se, $x = y$ (positividade); e
4. $d(x, y) \leq d(x, z) + d(z, y)$ (desigualdade triangular).

Na literatura existem várias medidas de similaridade de séries temporais e uma delas é a distância de Minkowski. Sendo c o comprimento das séries x e y, e x_j e y_j valores de x e de y no instante de tempo j $(1 \leq j \leq c)$, respectivamente, a distância de Minkowski entre x e y é dada por

$$d_p(x, y) = \left(\sum_{j=1}^{d} (x_j - y_j)^p \right)^{\frac{1}{p}} \tag{2.36}$$

onde p é um número positivo real. Para $p = 2$, a Equação 2.36 passa a representar a medida de similaridade aplicada nesta tese para o agrupamento das séries formadas das imagens capturadas por TID, a distância Euclidiana [Gan et al. 2007].

Câncer de mama, termografia da mama, registro de imagens, técnicas de mineração de dados e séries temporais foram os conceitos abordados neste capítulo e que são a base do desenvolvimento da metodologia proposta nesta tese. O próximo capítulo explora trabalhos na literatura relacionados com o aqui apresentado.

Capítulo 3

Revisão bibliográfica

Trabalhos encontrados na literatura com o objetivo de detectar o câncer de mama utilizando imagens capturadas por TID são descritos neste capítulo. Eles são reportados em ordem cronológica para que seja possível a percepção da evolução desta modalidade de exame no tempo. Para cada trabalho, os pontos principais abordados, quando possível, são: os procedimentos executados para aquisição das imagens (o protocolo de aquisição da TID); as técnicas de processamento e análise dos dados térmicos; e os resultados obtidos e conclusões.

Ohashi e Uchida [Ohashi e Uchida 1997] [Ohashi e Uchida 2000] afirmaram, no final da década de 90, que o desenvolvimento da TID dependia da definição e estabelecimento de métodos padronizados de teste e análise. Nesse sentido, eles estabeleceram um protocolo de aquisição de termogramas. Em uma sala com temperatura de 21 °C a paciente é orientada a colocar as mãos sobre a cabeça e, por 2 minutos, as mamas são resfriadas por um ventilador elétrico. Após esse tempo, uma sequência de termogramas é adquirida com intervalo de 15 segundos entre as imagens. Uma mama é considerada cancerosa se ela apresenta, o que os autores definiram como o critério de decisão, um *padrão de aquecimento positivo* na sequência de termogramas. Porém, esse critério carece de detalhes no texto. Ao que parece, a análise das imagens é realizada visualmente, pois nenhum método computacional é relatado. Também não existe menção às recomendações as quais as pacientes devem seguir, não é informada a distância da câmera à paciente nem quantas imagens são adquiridas e qual o tipo de equipamento usado. Os testes foram realizados com imagens de 728 pacientes coletadas entre 1989 e 1994, onde a taxa de verdadeiros positivos supera os 80%, mas a taxa de falsos positivos ultrapassa 40%.

Em um trabalho publicado em 2000, Anbar *et al.* [Anbar et al. 2000] sugerem que a

TID pode ser usada para detectar a presença de câncer de mama medindo a atenuação da modulação da temperatura da pele. Assim, é possível localizar áreas na superfície da mama que exibem modulação baixa anormal de temperatura. A área da superfície da pele das mamas é dividida em regiões pequenas (2x2 pixels). A temperatura de cada uma dessas regiões é acompanhada nos 2024 termogramas, adquiridos em sequência, dando origem às séries temporais. A distribuição espacial das amplitudes de modulação de temperatura de cada região é representada em um mapa de *bits* multicolorido. Nesse mapa, as regiões de cor preta representam regiões com amplitudes anormalmente atenuadas (amplitudes com valores mais baixos do que um valor de limiarização pré-definido), indicando uma suspeita de câncer em tais regiões. O mapa de *bits* multicolorido é analisado visualmente sem qualquer método de decisão computacional. Apenas quatro casos são relatados no estudo, onde em três, com diagnóstico positivo confirmado, a TID acusou alteração e em um, com suspeita mamográfica, a TID descartou a presença de câncer. Não há descrição dos procedimentos de captura das imagens, mas existem razões para acreditar que são os mesmos descrito no artigo publicado um ano depois, comentado a seguir.

Mais tarde, em 2001, Anbar *et al.* [Anbar et al. 2001], publicaram um artigo com avanços do trabalho descrito no parágrafo anterior. A câmera usada para aquisição das imagens possui um taxa de 100 imagens por segundos. Apenas pacientes com suspeitas mamográficas, que justifiquem a biópsia, foram incluídas nos testes. Após a aquisição dos termogramas, essas pacientes foram submetidas a uma intervenção cirúrgica e classificadas de acordo com a doença: pacientes com câncer invasivo, pacientes com carcinoma ductal *in situ* e pacientes com lesões benignas. Todas assinaram um termo de consentimento. No protocolo de aquisição, as pacientes são orientadas a permanecer sentadas, com as mãos apoiadas sobre a cabeça e sem respirar durante os 11 segundos de aquisição da sequência de imagens de cada posição: frontal, medial e lateral. O motivo da não respiração durante as aquisições é o fato de não existir uma etapa de registro das imagens para atenuar os efeitos dos movimentos de tal ação. A região de interesse, no caso a mama doente, é dividida em pequenas regiões de tamanho 4x4 pixels. A série temporal de uma dada região é constituída pelo valor da temperatura média, dessa região, nas 1024 imagens da sequência de termogramas. A Transformada de Fourier Rápida é aplicada sobre cada série temporal gerando um espectro de energia, para cada uma delas. Características são extraídas desses espectros e análises estatísticas são realizadas sobre eles com o objetivo de discriminar mamas saudáveis e mamas doentes. As conclusões são baseadas somente em gráficos e tabelas, sem uso de técnicas computacionais nas decisões. Especificidade e sensibilidade acima de 95% foram alcançadas em alguns testes, usando 100 pacientes, 34

com câncer e 66 com tumores benignos.

No trabalho de Parisky *et al.* [Parisky et al. 2003], a aquisição da sequência de termogramas é realizada enquanto a mama é resfriada por ar frio e algoritmos buscam por padrões infravermelho que estão associados a tumores malignos e a tecidos de mama saudáveis. Ao final do processo, regiões de interesse são localizadas e recebem uma pontuação numérica (de 0 a 100), em relação à suspeita da existência de lesões malignas. Os termogramas de cada paciente foram adquiridos durante uma única sessão de exame. O paciente é posicionado deitado na cama de exame com ambas as mamas suspensas através de aberturas no topo da cama. Enquanto uma mama é examinada, a outra fica protegida do ar frio por um vestido (camisola) de proteção. O exame é iniciado de forma estática por um breve período e após esse período uma corrente de ar frio é distribuída dentro da câmara de refrigeração, onde a mama a ser resfriada encontra-se suspensa. A sequência de imagens é adquirida antes e durante o resfriamento da mama. O mesmo processo é executado para a outra mama. Todo o processo dura, aproximadamente, 15 minutos com 3 minutos de aquisição das imagens da sequência para cada mama. Para validar a metodologia, sete radiologistas experientes em mamografias e que não participaram da fase de aquisição das imagens, foram convidados para avaliarem os termogramas e determinar um valor numérico, chamado pelos autores de *índice de suspeita*. As imagens de cada paciente foram avaliadas por 3 radiologistas escolhidos aleatoriamente dentre os 7, que foram instruídos a usar as mamografias para localizar as lesões nos termogramas. O número total de lesões avaliadas foi de 875, 187 malignas e 688 benignas. Segundo os autores, sensibilidade 97% e especificidade de 14% foram alcançadas. Para a obtenção desses valores, o valor do *índice de suspeita* de cada caso foi transformado em um resultado positivo ou negativo, de acordo com um valor de limiarização pré-determinado.

Kaczmarek e Nowakowski descrevem uma metodologia para detectar anormalidades na mama, em um artigo publicado em 2004 [Kaczmarek e Nowakowski 2004]. Para testar a metodologia, 3 mulheres com diagnóstico de câncer de mama (por mamografia e biópsia) foram examinadas por TID. Em um ambiente com temperatura de 21 °C, um conjunto de lâmpadas de halogêneo foi usado como uma fonte de aquecimento externo (energia elétrica de 1000W) a uma distância de 50 centímetros, por 30 segundos. As imagens são capturadas durante o resfriamento das mamas. Baseando-se na sequência de termogramas adquirida, imagens sintéticas foram calculadas de acordo com as propriedades térmicas do tecido por meio de modelagem numérica. Segundo os autores, foram detectadas alterações suspeitas. A análise foi realizada por meio de imagens e gráficos somente.

Também em 2004, Button *et al.* [Button et al. 2004] usaram o sistema de TID *BioScanIR* para a detecção de malignidade em tecidos mamários. Esse sistema contém uma câmera com resolução de 256x256 com capacidade de capturar 400 imagens por segundo. Termogramas de vinte e nove pacientes foram adquiridos momentos antes da realização de biópsia das mesmas. Em uma sala com temperatura de 21 °C, as pacientes foram posicionadas deitadas de bruços sobre uma mesa de forma que as mamas ficassem penduradas. Cada mama foi examinada separadamente em uma vista lateral a uma taxa de 10 imagens por segundos, durante 51.2 segundos. Após, as imagens foram revisadas por dois radiologistas. Cada caso recebeu uma das classificações: sem suspeitas; moderadamente suspeita; e altamente suspeita. Esta classificação foi baseada em achados térmicos previamente definidos. A sensibilidade e a especificidade na detecção de cânceres invasivos foi de 92% e de 53%, respectivamente. Nenhum método computacional foi aplicado para analisar as imagens.

Ainda em 2004, Amalu [Amalu 2004] realizou um estudo com termogramas de 500 pacientes sem câncer de mama, retiradas do banco de imagens clínicas de seu grupo de pesquisa. As pacientes desse estudo foram examinados por uma câmera infravermelha de alta resolução de quarta geração da marca FLIR, com largura de banda espectral de 7-13 μm. Em uma sala revestida por carpete, com temperatura entre 19 °C e 20 °C, a paciente é solicitada a submergir suas mãos em 6 polegadas de água com cubos de gelo flutuando (temperatura de aproximadamente 5 °C) por 1 minuto para provocar a vasoconstrição de tecidos saudáveis. Após isso, várias termogramas são capturados. As imagens são analisadas visualmente pelo método padronizado de análise termovascular proposto por Gautherie [Gautherie 1985] e Hobbins [Hobbins 1985], na década de 80. A interpretação é composta de vinte atributos discretos de vascularização e de temperatura da mama. Nesse método, os termogramas são graduados em uma das cinco classificações **Th** (*Thermobiological*). Combinando o padrão vascular e as temperaturas das duas mamas, as imagens são classificadas como **Th1** (normal e sem vascularidade), **Th2** (normal, mas com vascularidade), **Th3** (questionável), **Th4** (anormal) **Th5** (severamente anormal). Todas as pacientes foram graduadas em alguma classificação **Th** baseando-se na análise do padrão vascular e da emissão térmica. As imagens capturadas por TID foram processadas digitalmente e comparadas às imagens capturadas por TIE, onde a paciente permanece por 15 minutos sem a parte de cima da roupa do corpo para que a temperatura da superfície da mama se estabilize e, após esse tempo, as imagens são capturadas. As imagens foram cuidadosamente analisadas em busca de mudanças no tempo da vascularidade e da temperatura, denotando uma resposta ao estresse térmico. Como resultado, 23 pacientes

foram detectadas com câncer de mama comprovados histopatologicamente. Atualmente, o Pacific Chiropractic and Research Center [Amalu 2012], no qual o autor desse artigo é o diretor clínico, realiza uma análise dos gráficos formados pelas séries temporais de áreas suspeitas da mama de termogramas capturados no exame de TID. Nele, um estresse térmico (no caso, um estímulo frio) é administrado e ao final desse, 50 imagens são capturadas por 4 minutos. A Figura 3.1 (a) exibe o gráfico das séries temporais de quatro regiões selecionadas de um termograma classificado como **Th4**. As curvas 1, 2 e 3 são de áreas suspeitas, que começam como áreas de temperaturas altas, mas que as curvas 1 e 2 são as que produzem maior desconfiança por rapidamente se estabilizarem. A única área que apresenta uma resposta normal é a representada pela curva de cor magenta. Quatro áreas semelhantes foram selecionadas na mama esquerda e as séries temporais estão representadas no gráfico da Figura 3.1 (b). Nesse caso, o reaquecimento acontece de forma progressiva e quase idêntica para as quatro regiões, configurando uma resposta normal, ou seja, sem áreas suspeitas. A análise das séries foi realizada visualmente.

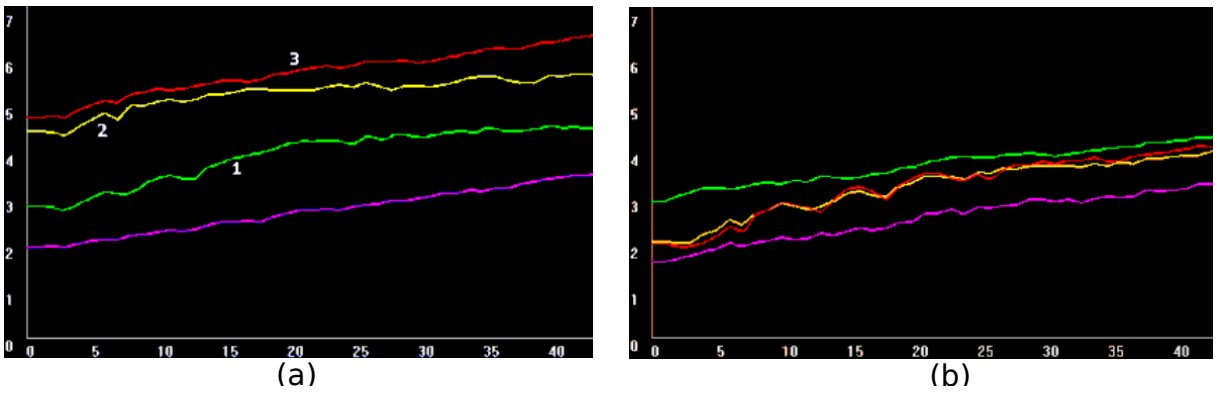

Figura 3.1: Séries temporais de temperatura da superfície da mama: (a) séries de quatro regiões suspeitas na mama direita, (b) séries de quatro regiões não suspeitas na mama esquerda (Fonte: Pacific Chiropractic and Research Center [Amalu 2012])

Arora *et al.* [Arora et al. 2008] propuseram uma metodologia de detecção de câncer de mama no qual, para a aquisição dos termogramas, a paciente é despida da cintura para cima e posicionada em um conjunto de equipamentos dedicados contendo uma cadeira com espelhos para as vistas laterias, um resfriador usando ar e uma câmera infravermelha digital. O exame possui duração de 4 minutos por paciente e uma série de termogramas, com mais de 100 imagens, é capturada durante a administração de um estresse causado pelo frio (ar frio direcionado às mamas). A câmera infravermelha utilizada é uma do tipo matriz de plano focal sem resfriamento, com tamanho de imagem de 320x240 pixels, sensibilidade de $0,08\,^{\circ}\mathrm{C}$, e um espectro de operação variando de 7-12 μm. Por meios computacionais, parâmetros térmicos específicos foram extraídos para realizar a análise

de assimetria térmica entre as mama e buscar por áreas com as maiores diferenças na temperatura, quando comparadas às áreas vizinhas. Uma imagem codificada por cores com regiões suspeitas identificadas foi gerada. Utilizando uma rede neural artificial para avaliar a imagem codificada por cores, a metodologia alcançou sensibilidade de 96,7%, especificidade de 26,5% e valor preditivo negativo de 81,8%. A metodologia foi testada em 94 pacientes, com idade média de 51 anos, sendo 60 com malignidade em pelo menos uma das mamas.

Em um trabalho publicado em 2010, Wishart *et al.* [Wishart et al. 2010] utilizaram termogramas capturados por TID para identificar pacientes com câncer de mama. No exame, a paciente é despida da cintura para cima e apropriadamente posicionada em uma cadeira ergométrica com os braços apoiados no nível dos olhos. Um fluxo de ar de temperatura controlada é direcionado às mamas por 5 minutos enquanto a câmera infravermelha registra uma série de termogramas da superfície da pele em um total de 250 imagens. Por métodos computacionais, parâmetros térmicos específicos, incluindo diferenças de temperaturas e medidas de simetria térmica, foram extraídos das imagens capturadas. Além disso, áreas da mama que exibiram padrões de resfriamento anormal receberam uma codificação por cores. No texto não está detalhado exatamente como, mas uma técnica de inteligência artificial foi aplicada para a identificação de casos com câncer alcançando: 70% de sensibilidade, 48% de especificidade. Os resultados dos testes são baseados em 106 biópsias realizadas em 100 pacientes, onde 65 diagnósticos são positivos (com câncer) e 41 diagnósticos são negativos (lesões benignas).

No trabalho de Gerasimova *et al.* [Gerasimova et al. 2012], TID foi realizado em 46 casos, histopatologicamente comprovados, de tumores de mama entre malignos e benignos, antes da cirurgia. Das regiões com tumor e sem tumor, foram geradas séries temporais da temperatura da pele e sobre essas séries foram aplicados para estudar o comportamento da variação de temperatura: análise de Fourier, transformada de Wavelet e diagrama de fases. No trabalho foi concluído que diagramas de fase caótica correspondem a tecidos saudáveis, enquanto que para tecidos cancerosos é típico a forma irregular no espaço de fase. Além disso, resultados indicam que os sinais de temperatura de tecidos saudáveis são anticorrelacionados, enquanto que no tumor foi observada correlação de ruído térmico. Segundo os autores, isso indica a inabilidade do tecido anormal de se adaptar às influências externas. Além disso, os resultados estão de acordo com a regra de ouro nas ciências biomédicas. Tal regra afirma que sistemas normais e saudáveis são frequentemente muito complexos e que a complexidade diminui quando uma anormalidade ou doença ocorre [Najarian e Splinter 2005].

Em outro trabalho de Gerasimova *et al.* [Gerasimova et al. 2013], análise multifractal das séries temporais, geradas a partir da TID, foi executada para verificar a diferença de comportamento entre tecido com tumor maligno e tecido saudável. O método de máximo módulo da transformada de wavelet foi aplicado para caracterizar as propriedades multifractais de séries temporais oriundas de mamas cancerosas e saudáveis. Os autores concluíram que as propriedades escalares multifractais complexas, observada em séries temporais sobre regulação automática (mamas saudáveis) são drasticamente alteradas na existência de doenças (mamas cancerosas). Nesse estudo, as duas mamas de 9 mulheres foram examinadas, 6 com câncer e 3 saudáveis. Uma câmera detectora fotovoltaica *InSb* foi utilizada. No momento das aquisições das imagens, cada paciente permaneceu sentada com os braços para baixo para evitar o desconforto. Imagens frontais foram capturadas na distância de 1 metro em um ambiente de temperatura controlada entre 20 °C a 22 °C. Cada conjunto de imagens contém 30000 quadros de imagens adquiridos por 10 minutos. Foram usados marcadores de superfície de pele como pontos de referência para o registro das imagens, a fim de eliminar artefatos de movimento na análise. Em pacientes doentes, a região com tumor e a região simetricamente posicionada na outra mama são delimitadas por regiões quadradas de tamanho 8x8 pixels. A análise é realizada somente dentro dessas regiões. A metodologia conseguiu distinguir região com tumor e região saudável. Para mamas saudáveis, foi encontrada uma dimensão multifractal como característica de uma contínua mudança na forma da função de densidade de probabilidade de variação de temperatura através do tempo. Entretanto, sinais térmicos da mama com tumor maligno mostraram estatísticas de variação de temperatura monofractal homogênea como evidência da perda de complexidade. As análises foram realizadas visualmente por meio de gráficos e tabelas.

Recentemente, Gerasimova *et al.* publicou outro trabalho [Gerasimova et al. 2014] usando uma base de dados maior, 33 pacientes com câncer de mama histopatologicamente confirmado e 14 voluntárias saudáveis para controle. Os achados reafirmaram os resultados de [Gerasimova et al. 2013].

Por fim, em um trabalho publicado este ano, Saniei *et al.* [Saniei et al. 2015] desenvolveram uma metodologia para análise de termografias obtidas por TID onde apenas duas imagens são usadas. Em um ambiente revestido por carpete, umidade controlada e temperatura entre 20 °C a 21 °C, cada paciente deve aguardar sentada por 15 minutos sem a parte de cima da roupa, e com os braços afastados do corpo, para a aclimatação. Após esse tempo, a paciente posiciona as mãos sobre a cabeça e um termograma é capturado. Em seguida, ela é instruída a mergulhar as mãos em um recipiente com água e gelo

por 1 minuto (temperatura de aproximadamente 5 °C) para que ocorra a vasoconstrição na superfície da mama. Imediatamente após a retirada das mãos de dentro desse recipiente, outro termograma é capturado. Por técnicas computacionais, ambas as mamas são segmentadas e as imagens são registradas. O próximo passo é a segmentação e a esqueletização do padrão vascular presente nas mamas, nos dois termogramas. Características são extraídas desses padrões vasculares e técnicas semelhantes as usadas para reconhecimento de impressões digitais são aplicadas para comparar tais padrões vasculares, dos termogramas antes e após o estresse térmico. Para quantificar o grau de similaridade entre os padrões vasculares nas duas imagens um valor de pontuação de relacionamento é gerado por meio de uma expressão matemática e esses padrões serão considerados correspondentes se o valor de pontuação estiver abaixo de um determinado valor de limiarização, que é definido empiricamente. Segundo os autores, a metodologia alcançou sensibilidade de 86% e especificidade de 61%. Foram usadas imagens de 50 pacientes, 25 com e 25 sem câncer de mama.

Como visto neste capítulo, a maioria dos trabalhos reportados, principalmente os mais antigos, não utiliza técnicas de aprendizagem de máquina para a identificação automática de pacientes com câncer de mama. Apenas de Wishart *et al.* [Wishart et al. 2010] utiliza uma técnica de inteligência artificial, mas não detalha exatamente como. Os demais trabalhos propõem metodologias para o desenvolvimento de sistemas de auxílio ao diagnóstico médico, onde a decisão do profissional seria baseada em gráficos ou imagem pré-processadas computacionalmente. A Tabela 3.1 contém um resumo dos trabalhos apresentados aqui e sua principais características.

Diferentemente dos trabalhos anteriores, a metodologia proposta aqui aplica técnicas de aprendizagem de máquina sobre as séries temporais, produzidos a partir da TID, para automaticamente identificar pacientes com câncer de mama. Porém, de forma semelhante a esses trabalhos, a metodologia proposta baseia-se na hipótese de que a TID fornece informações térmicas no tempo não presentes na TI, tornando-a mais eficiente. Em alguns trabalhos, as mamas são resfriadas por um ventilador elétrico e apenas uma imagem é capturada ao final no resfriamento [Kapoor e Prasad 2010] [Koay et al. 2004]. Os autores justificam tal procedimento afirmando que o estresse térmico aumenta o contraste de temperaturas entre as regiões quentes e frias das mamas, mesmo em exames por TI.

Tabela 3.1: Trabalhos relacionados e suas caractecterísticas.

Trabalho	Tamanho do conjunto de teste	Tipo de câmera	Estresse Térmico	Uso de aprendizagem de máquina	Taxa de acerto sens./espec.
Ohashi e Uchida, (1997 e 2000)	728	Não informado	Ventilador elétrico	Não	Não informado
Anbar *et al.*, (2000)	4	100 imagens/s	Não informado	Não	75%/–
Anbar *et al.*, (2001)	100	100 imagens/s	Não informado	Não	95%/95%
Parisky *et al.*, (2003)	875	Não informado	Ar frio	Não	97%/14%
Kaczmarek e Nowakowski (2004)	3	Não informado	Lâmpadas de halogêneo	Não	Não informado
Button *et al.*, (2004)	29	400 imagens/s	Não informado	Não	92%/53%
Amalu, (2004)	500	Não informado	Água com gelo	Não	Não se aplica
Arora *et al.*, (2008)	94	2,4 imagens/s	Ar frio	Sim	96%/26%
Wishart *et al.*, (2010)	106	1,2 imagens/s	Ar frio	Sim	70%/48%
Gerasimova *et al.*, (2012)	46	Não informado	Não informado	Não	Não informado
Gerasimova *et al.*, (2013)	18	50 imagens/s	Não informando	Não	Não informado
Gerasimova *et al.*, (2014)	47	50 imagens/s	Não informando	Não	76%/86%
Saniei *et al.*, (2015)	50	Não informado	Água com gelo	Não	86%/61%

Capítulo 4

Metodologia proposta

Este capítulo descreve a metodologia proposta nesta tese para a detecção de anomalias de mama, incluindo o câncer, a partir da examinação de uma paciente por TID [Silva et al. 2014a] [Silva et al. 2015b] [Silva et al. 2015a]. Em resumo, inicialmente um protocolo de execução da TID foi estabelecido para a aquisição das imagens no HUAP [Silva et al. 2013] [Silva et al. 2014b]. Em seguida, a matriz de temperatura de cada imagem capturada é extraída do arquivo gerado pela câmera infravermelha. O próximo passo é a segmentação da região de interesse (termo técnico usado em inglês: *Region of Interest*-(ROI)), que neste trabalho são ambas mamas, com posterior registro das imagens. Em sequência, a ROI é dividida em pequenas regiões quadradas, e a temperatura máxima de cada uma dessas regiões é observada em todos os termogramas da paciente gerando, assim, as séries temporais. Após, o algoritmo *k-means* é aplicado sobre o conjunto de séries temporais de temperatura, formando k grupos ($2 \leq k \leq 10$). Índices de validação de agrupamento são aplicados para avaliar o resultados do agrupamento construído pelo algoritmo *k-means*, para cada valor de k. Os valores obtidos são tratados como características e submetidos à etapa de classificação. A Figura 4.1 mostra o fluxograma das etapas da metodologia proposta e as próximas seções detalham cada uma delas.

4.1 Testes e análises preliminares à metodologia

Em passos iniciais da pesquisa, alguns experimentos foram realizados antes que a metodologia chegasse a forma que será descrito nas próximas seções. Em relação ao protocolo de aquisição da TID, uma revisão bibliográfica foi realizada em busca dos protocolos executados por diversos grupos no mundo. Esses protocolos foram reproduzidos, na medida do possível, em nosso laboratório e testados em voluntárias com algumas adaptações ao

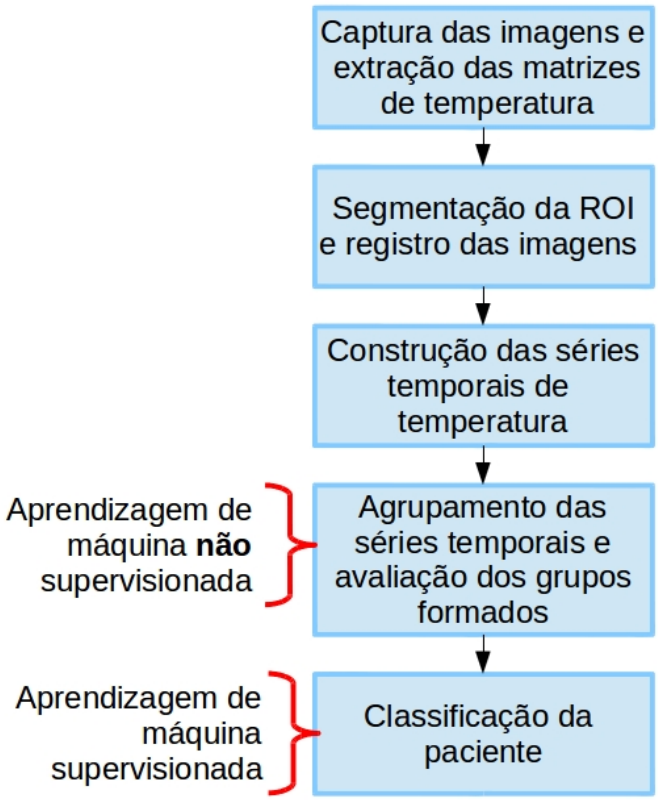

Figura 4.1: Fluxograma das etapas da metodologia proposta nesta tese.

nosso clima e realidade. Após a avaliação visual dos médicos do HUAP dos termogramas produzidos em cada mudança de parâmetros um protocolo foi estabelecido para ser executado no HUAP nas pacientes voluntariadas. O principais parâmetros testados são: o tipo de resfriamento (utilizando um ventilador elétrico ou álcool espalhado na superfície das mamas); tempo de resfriamento (1 minuto, 2 minutos ou 5 minutos); intervalo de aquisição entre as imagens (10 segundos ou 15 segundos); tempo de aquisição das imagens (3 minutos, 5 minutos ou 10 minutos); posicionamento das pacientes (sentada ou não, com as mãos sobre a cabeça ou não); e influência das condições da sala de exame (temperatura da sala, bloqueio de fluxo de ar não controlado direcionado à paciente, uso de material isolante térmico para o fundo da imagem ou não).

A segmentação automática da ROI por métodos computacionais em termogramas obtidos por TI já foi proposta em trabalhos anteriores de membros do grupo de pesquisa do Visual Lab [Motta 2010] [Marques 2012] [Márquez et al. 2015]. Porém os métodos desenvolvidos nesses trabalhos não foram bem sucedidos na tarefa de segmentar a ROI da primeira imagem da sequência de termogramas capturada na execução da TID. Isso se deve ao fato de que eles tomam como base para definir o limite inferior da ROI o contraste de temperatura existente na região das pregas inframamárias nos termogramas, e

esse contraste é menor nos termogramas obtidos por execução da TID, quando comparado ao contraste nos termogramas obtidos por TI. No protocolo proposto neste trabalho, as mamas são resfriadas por um ventilador elétrico por um determinado período de tempo. Essa ação diminui ainda mais o contraste de temperatura (tonalidade), que já é pequeno nesse tipo de imagem. A segmentação da ROI somente da primeira imagem da sequência de termogramas é suficiente, pois todas as imagens de uma sequência são registradas em relação a primeira imagem, ou seja, para cada sequência o registro é realizado 19 vezes onde a primeira imagem sempre é a fixa. Conjecturou-se realizar o registro tendo como imagem fixa o último termograma da sequência que, após os cinco minutos de captura, já é bem semelhante aos termogramas capturados por TI em relação ao contraste de temperatura, pois a superfície da mama encontra-se praticamente em estabilidade térmica de temperatura 2.2.1, e isso possibilitaria o uso do método automático de segmentação da ROI de Marques [Marques 2012]. Mas devido ao tempo de aquisição das imagem, utilizar o último termograma como a imagem fixa não foi a melhor opção. No protocolo de aquisição estabelecido, a paciente permanece em pé por 5 minutos com as mãos sobre a cabeça enquanto as imagens são capturadas na recuperação da temperatura da superfície da pele da mama, pela câmera infravermelha. Algumas pacientes possuem dificuldades de equilíbrio e nos últimos instantes da captura da sequência realizam movimentos mais significativos de balanço, o que torna as últimas imagens bem diferentes da maioria das imagens da sequência. Nessa situação, o registro de imagens considerando como imagem fixa a última da sequência gerou mais ruído do que o registro considerando a primeira imagem da sequência. Dessa forma decidiu-se realizar a segmentação manual da primeira imagem da sequência para obter-se um melhor resultado na etapa de registro dos termogramas. Outra decisão tomada em relação a segmentação da ROI foi deixar de fora as axilas, que na segmentação automática de Marques são consideradas. Em uma das etapas da metologia as series temporais de temperaturas, construídas a partir da observação da temperatura de regiões pequenas da superfície de mama nos 20 termogramas, são agrupadas de acordo com suas semelhanças. As axilas são naturalmente regiões mais quentes e as series temporais de temperatura construídas a partir da superfície dessas regiões se assemelhavam as series temperais de regiões mais quentes de mama por conta de um tumor maligno. Em alguns testes realizados esse fato provocou a classificação de pacientes saudáveis em pacientes doentes por parte da metodologia proposta.

No início da pesquisa, as imagens da sequência de termogramas foram registradas pela metodologia proposto por Galvão [Galvao 2015] em sua tese de doutorado. Mas os resultados ainda não foram satisfatórios, gerando muito ruído nas séries temporais de

temperatura construídas, pois a própria metodologia de registro encontrava-se ainda nas fases iniciais. Então decidiu-se utilizar um método já pronto de registro de imagens. O método utilizado é descrito em [Myronenko e Song 2010]. Porém foi necessário realizar uma adaptação desse registro para as imagens do nosso banco de termografias. Os resultados foram tão satisfatórios que mesmo após a finalização da metologia proposta por Galvão decidiu-se manter os registros já executados.

Após o registro das imagens e construção das séries temporais de temperatura a etapa seguinte da metodologia é o agrupamento de tais séries. Para isso é utilizado o algoritmo de agrupamento de dados *k-means* (Seção 2.4.1.2). Esse algoritmo pode ser executado aplicando-se diversas medidas de similaridade dos dados. Na fase de testes foram aplicadas as seguintes medidas: a distância euclidiana, a distância de Manhattan e a distância correlação de Pearson [Gan et al. 2007]. Mas apesar da simplicidade, a distância euclidiana apresentou os melhores resultados nos testes e por esse motivo decidiu-se utilizar tal distância no algoritmo *k-means* para o agrupamento das séries.

A Figura 4.2 mostra as séries temporais obtidas de uma paciente com câncer de mama e a Figura 4.3 mostra as séries temporais obtidas de uma paciente saudável. O eixo x representa o tempo em segundos (5 minutos, tempo de duração do exame), o eixo y contém o S_k (o índice) de cada uma das séries temporais, e o eixo z indica a temperatura em graus Celsius. É possível observar que existem grupos de séries temporais com temperaturas maiores e com inclinação maior nos momentos iniciais da recuperação da temperatura, após o estresse térmico, para a paciente doente (Figure 4.2), ou seja, séries temporais que se destacam das demais. O mesmo não é verdadeiro para as séries temporais obtidas da paciente saudável (Figure 4.3). A observação de gráficos como esses e resultados de trabalhos relacionados (Capítulo 3) mostravam o potencial da TID na tarefa de detectar anormalidades de mama. Faltava apenas definir como aproveitar a informação temporal da temperatura de cada ponto da mama, fornecida pela TID. Faltava definir como as diferenças percebidas nas séries temporais de temperatura construídas poderiam ser transformadas em características para que pacientes doentes fossem identificadas, e assim alcançar o objetivo, ou seja, propor uma metologia computacional para detectar anormalidades na mama, utilizando as imagens capturadas por TID.

Com o objetivo de detectar o grupo de séries temporais constituído de temperaturas da região da mama com anormalidade, decidiu-se realizar o agrupamento dessas séries e posteriormente avaliar o agrupamento formado. Dada uma paciente qualquer, as suposições foram as seguintes:

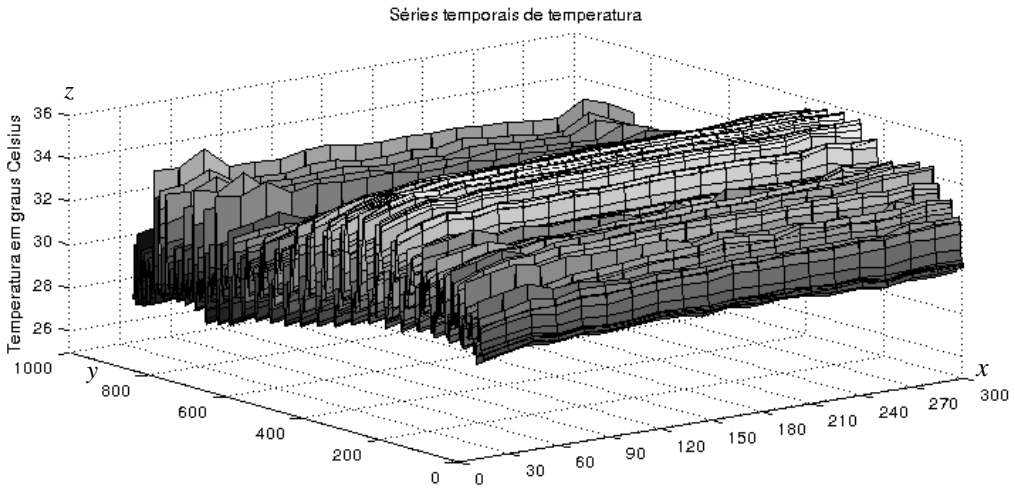

Figura 4.2: Séries temporais de temperatura de uma paciente com câncer de mama.

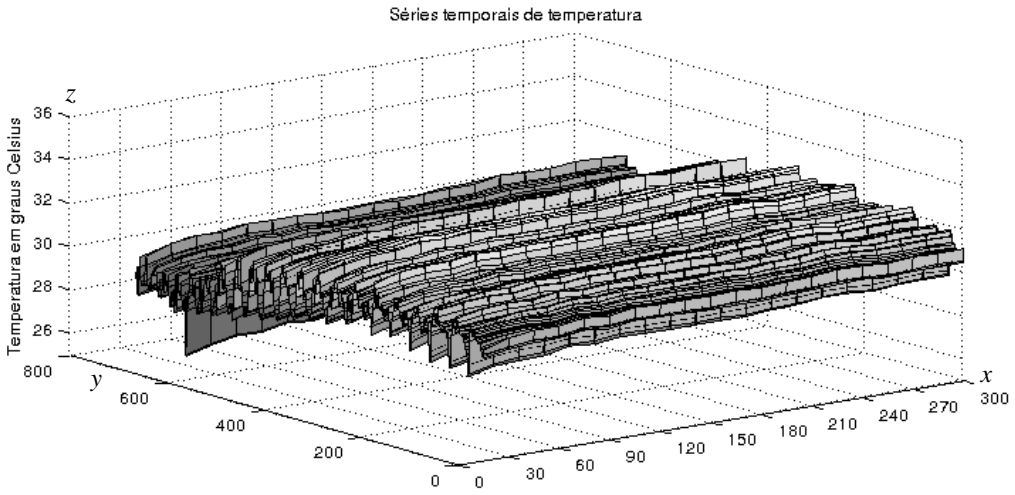

Figura 4.3: Séries temporais de temperatura de uma paciente saudável.

1. Se essa paciente for saudável, as séries temporais construídas a partir de cada ponto da mama são semelhantes (como na Figura 4.3), ou seja, possuem um grau alto de similaridade e então um agrupamento formado por tais séries temporais possui as seguintes características:

   i) os grupos formados não são compactos;

   ii) os grupos formados são próximos (são semelhantes) uns dos outros, em relação a uma medida de similaridade.

2. Por outro lado, se essa paciente for doente, para as séries temporais construídas a partir de cada ponto da mama, existe um grupo que se destaca dos demais por possuir séries temporais com comportamento diferente (como na Figura 4.2), e então um

agrupamento formado por tais séries temporais possui as seguintes características:

i) os grupos formados são mais compactos, se comparados aos da suposição 1.;

ii) os grupos formados são menos próximos (são menos semelhantes) uns dos outros, em relação a uma medida de similaridade, se comparados aos da suposição 1.

Para medir a compacidade de cada grupo formado e a distância uns dos outros, dentro de um agrupamento formado pelo algoritmo *k-means*, medidas (índices) de validação de agrupamento (Seção 2.4.1.3) foram aplicadas na esperança de que essas medidas retornassem valores diferentes entre pacientes doentes e saudáveis. Apenas com 11 pacientes doentes e com 11 pacientes saudáveis, alguns índices de validação de agrupamento, para alguns valores de k (k-grupos) na execução do algoritmo *k-means*, foram calculados. As figuras 4.4 4.5 e 4.6 contém os gráficos de experimentos iniciais. Neles, o eixo horizontal representa a quantidade de pacientes, o eixo vertical o valor do índice, os asteriscos representam os valores do respectivo índice para as pacientes doentes e as circunferências representam os valores do respectivo índice para as pacientes saudáveis. A Figura 4.4 contém os resultados do cálculo do índice Silhueta sobre o agrupamento formado para cada uma das 22 pacientes. Com um valor de limiar igual a $0,4$ (indicado pela seta) é possível separar a maioria das pacientes doentes, da maioria das pacientes saudáveis. A Figura 4.5 é semelhante a Figura 4.4, porém contém os resultados do cálculo do índice Krzanowski-Lai sobre o agrupamento formado para cada uma das mesmas 22 pacientes. Com um valor de limiar um pouco abaixo de $0,4$ (indicado pela seta) Também é possível separar a maioria das pacientes doentes, da maioria das pacientes saudáveis. A mesma análise pode ser feita para o gráfico da Figura 4.6.

Depois de concluir que era possível separar quase que totalmente as pacientes doentes das pacientes saudáveis, escolhendo-se um valor de limiar apropriado para cada índice, surgiu a ideia de usar os valores calculados de cada um desses índices, para cada valor de k, para compor um vetor que representaria uma determinada paciente em uma etapa de classificação. Assim, esse vetor passou a ser o vetor de características da paciente na aprendizagem de máquina supervisionada, caracterizando a metologia como híbrida.

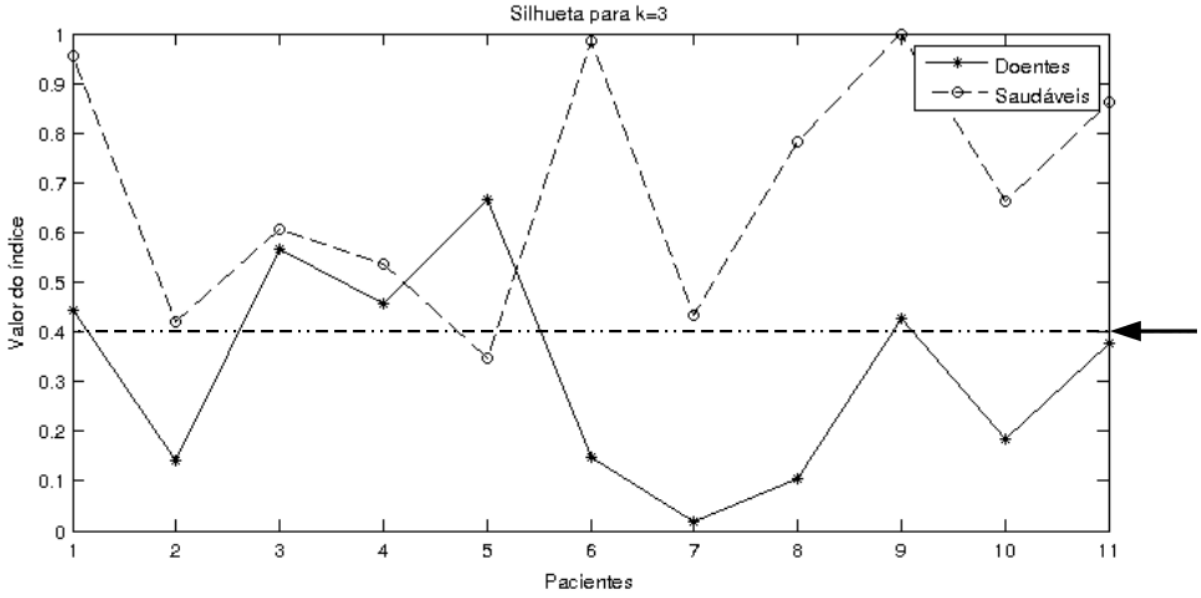

Figura 4.4: Cálculo do índice Silhueta para $k = 3$.

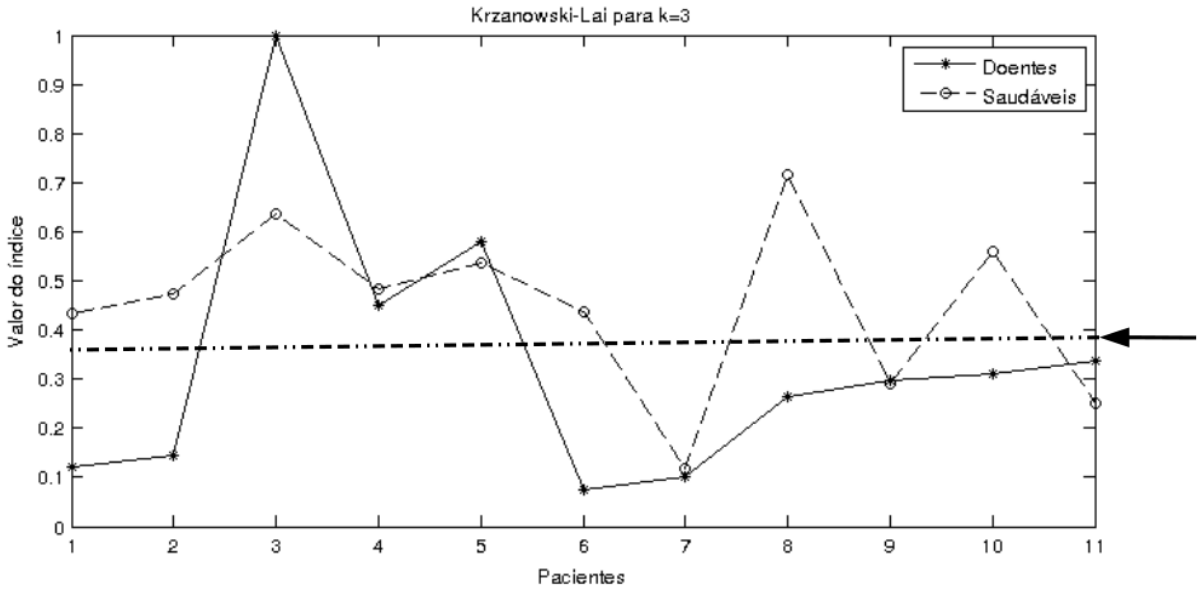

Figura 4.5: Cálculo do índice Krzanowski-Lai para $k = 3$.

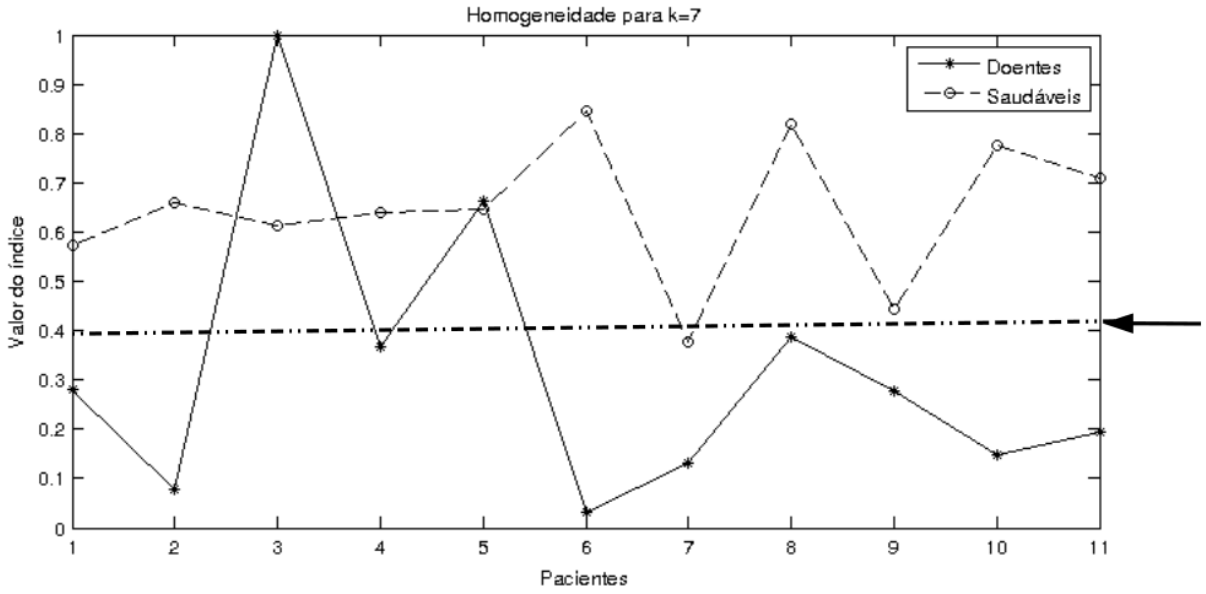

Figura 4.6: Cálculo do índice Homogeneidade para $k = 7$

4.2 Captura das imagens e extração das matrizes de temperatura

A seguir, as orientações do protocolo [Silva et al. 2013] [Silva et al. 2014b] e as respectivas referências para cada uma delas. Alguns trabalhos referenciados estão detalhados no Capítulo 3.

i) *Recomendações à paciente*

Pelo menos duas horas antes do exame, a paciente deve evitar: fumar, ingerir álcool ou cafeína, praticar exercícios físicos ou aplicar algum tipo de creme, óleo ou desodorante na região das mamas e axilas [Kapoor e Prasad 2010] [Koay et al. 2004] [Ng e Kee 2007].

ii) *Condições do ambiente de exame*

A temperatura ambiente é mantida entre 20 °C e 22 °C [Ohashi e Uchida 1997] [Ohashi e Uchida 2000] [Button et al. 2004] [Gerasimova et al. 2013], com ausência de janelas ou aberturas, lâmpadas fluorescentes e fluxo não controlado de ar direcionado à paciente [Kontos et al. 2011].

iii) *Preparação da paciente*

Na sala de exame, a paciente é requisitada a retirar brincos, cordões e/ou outros acessórios que possam interferir na termografia. A temperatura central da paciente é verificada por termômetro clínico e os cabelos presos com uma touca [Kontos et al. 2011].

iv) *Captura das imagens*

A paciente é orientada a apoiar as mãos sobre a cabeça e suas mamas são resfriadas utilizando um ventilador elétrico [Koay et al. 2004] [Kapoor e Prasad 2010] (Figura 4.7(a)) até que a região central do tórax alcance temperatura média de $30,5\,°C$ ou até que 5 minutos sejam transcorridos [Wishart et al. 2010]. Quando uma dessas condições é satisfeita, o ventilador é desligado e os termogramas são capturados. A temperatura média da região central do tórax é monitorada pela ferramenta *Caixa* da câmera (imagem (a) da Figura 4.10). O intervalo de tempo entre as imagens é de 15 segundos [Ohashi e Uchida 1997] [Ohashi e Uchida 2000].

v) *Parâmetros*

As imagens são capturadas na posição frontal, onde a paciente permanece de frente para a câmera [Anbar et al. 2001] [Gerasimova et al. 2013], como mostra a Figura 4.7 (c), na distância de $1\,m$ [Gerasimova et al. 2013] [Kontos et al. 2011], podendo variar de $0,8\,m$ a $1,2\,m$, dependendo do tamanho da paciente. Para pacientes maiores do que a maioria, a câmera é afastada para a distância de $1,2\,m$, para pacientes menores do que a maioria, a câmera é aproximada para a distância de $0,8\,m$. A umidade relativa do ar e a temperatura da sala são registradas por um termo-higrômetro (Figura 4.7(b)) e inseridas como parâmetros nas configurações da câmera, assim como a distância entre a câmera e a paciente, e a emissividade da pele humana ($\varepsilon \approx 0,98$) [Ng et al. 2001] [Acharya et al. 2012].

Figura 4.7: Captura das imagens: em (a) e em (b), respectivamente, o ventilador elétrico e o termo-higrômetro usados, e em (c) o posicionamento da paciente.

Utilizando uma câmera infravermelha e executando o protocolo descrito acima, os termogramas são capturados e armazenados em um banco de dados [Silva et al. 2014b]. A câmera é da marca Flir, modelo SC620 [Flir 2013] (Figura 4.8). Sua sensibilidade é menor do que $0,04\,^{\circ}\mathrm{C}$ e possui faixa de captura padrão de $-40\,^{\circ}\mathrm{C}$ a $500\,^{\circ}\mathrm{C}$, gerando imagens com dimensão de 640x480 pixels, na faixa espectral $7,5 - 13\mu m$.

Figura 4.8: Câmera infravermelha usada para aquisição dos termogramas.

No momento da captura das imagens, essa câmera gera um arquivo que contém a representação espacial da cena e dados radiométricos que são usados para medir, na forma de temperatura, a energia radiante captada pelo sensor da câmera. No caso das câmeras da marca Flir, esses arquivos utilizam a extensão *jpg*, porém programas populares de visualização de imagens não são capazes de modificar e realizar análises sofisticadas de dados radiométricos em tais arquivos, somente programas proprietários da fabricante da câmera. Para contornar essa dificuldade, Borchartt, em sua tese de doutorado [Borchartt 2013], utilizou o SDK da câmera e desenvolveu o *TermoCad* no Visual Lab (http://visual.ic.uff.br/), que, entre várias funcionalidades, gera e armazena, no formato *txt*, a matriz de temperatura da cena capturada pela câmera. A Figura 4.9 é parte de uma das matrizes de temperatura analisadas na metodologia, cada valor representa a temperatura de um ponto na cena, na respectiva posição do plano focal. Visto que a metodologia proposta é desenvolvida baseando-se apenas nas matrizes de temperatura dos termogramas, o *TermoCad* foi executado para gerar tais matrizes. A decisão de usar apenas matrizes de temperatura e não imagens foi tomada quando o uso dos termogramas diretamente no formato *jpg* apresentou várias desvantagens, pois essas imagens possuem apenas 120 tons de cores, e vários artefatos, tais com logomarca do fabricante, legenda de cores/temperatura e diversos parâmetros que atrapalham o processamento e análise das mesmas por métodos computacionais. A Figura 4.10 exibe em (a) um termograma visualizado por um programa comum de visualização de imagens, e em (b) o termograma em tons de cinza gerado a partir da matriz de temperatura extraída pelo *TermoCad*.

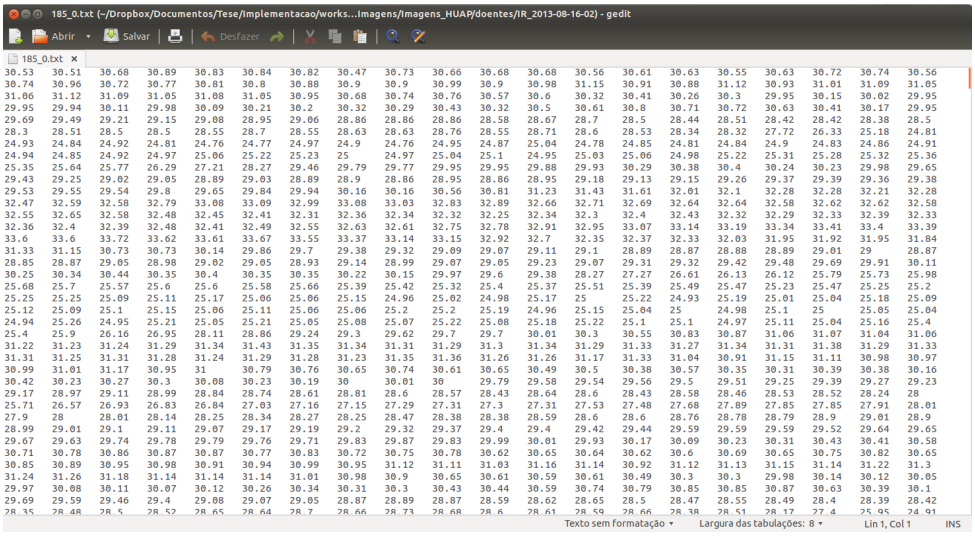

Figura 4.9: Visualização de parte de uma matriz de temperatura no formato *txt*.

4.3 Segmentação da ROI e registro das imagens

Em razão do objetivo desta tese, a ROI inclui somente ambas as mamas da paciente, deixando as axilas de fora, e sua segmentação é realizada manualmente. Os motivos para essas decisões estão detalhados na Seção 4.1. Os trabalhos reportados no Capítulo 3, que propõem análises de termogramas obtidos por TID, não descrevem uma etapa de segmentação automática da ROI, deixando implícito que a segmentação de tal região é realizada de forma manual nesses trabalhos. Na Seção 6.2 uma segmentação automática da ROI é proposta como um trabalho futuro.

Como já foi citado, para cada paciente, uma sequência de 20 termogramas é capturada. Porém, no processo de segmentação da ROI apenas a primeira imagem da sequência é considerada. O motivo para esta escolha está na próxima etapa da metodologia, a etapa de registro das imagens. Todas as imagens da sequência são registradas (emparelhadas) em relação a primeira imagem. Assim, a segmentação da primeira imagem é equivalente a segmentação de todas as outras imagens da sequência. O processo de segmentação da ROI segue os seguintes passos:

i) o primeiro termograma da sequência de termogramas (Figura 4.10(a)) capturada pela câmera é transformado em uma imagem em tons de cinza (Figura 4.10(b)), a partir de sua respectiva matriz de temperatura, pela Equação 4.1, onde p_{ij}, t_{ij}, t_{max} e t_{min} representam, respectivamente, o nível de cinza do pixel na posição (i, j) na imagem transformada, o valor da temperatura na posição (i, j) $(1 \leq i \leq 480,\ 1 \leq j \leq 640)$ na matriz de temperatura, a temperatura máxima presente na matriz e

a temperatura mínima também presente na matriz (esta equação foi adaptada da conversão de escalas termométricas em [Hewitt 2009]);

ii) a imagem em tons de cinza é carregada no programa *ImageJ* [Abramoff et al. 2004] [Schneider et al. 2012] e, uma vez nesse *software*, as ferramentas *Polygon Seletions* e *Fit Spline* são aplicadas para selecionar a região compreendendo as mamas (Figura 4.10(c));

iii) ainda no *ImageJ*, a máscara da região selecionada é criada (Figura 4.10(d)).

$$p_{ij} = 255 \cdot \frac{t_{ij} - t_{min}}{t_{max} - t_{min}} \tag{4.1}$$

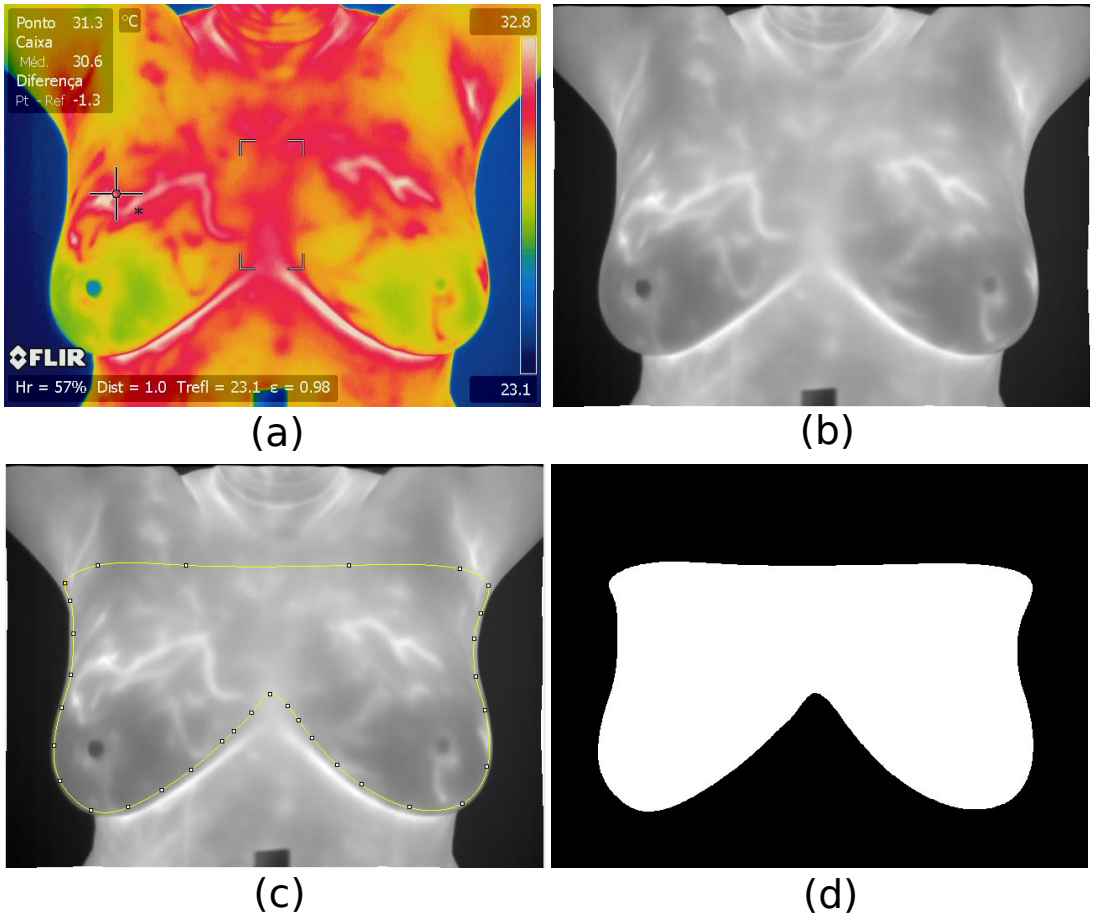

Figura 4.10: Estágios da segmentação da ROI.

A máscara indica ao sistema computacional onde está a ROI na matriz de temperatura de um termograma, pois as coordenadas das temperaturas da superfície das mamas, correspondem às coordenadas dos pixels de cor branca na máscara.

Durante o exame por TID, a paciente executa pequenos movimento involuntários de respiração e balanço. Esses movimentos causam diferenças de uma imagem para outra e consequentemente ruídos nas séries temporais formadas. A Figura 4.11 mostra em (a) o primeiro dos 20 termogramas de uma paciente, em (b) o décimo sétimo termograma da mesma paciente, e em (c) o resultado da subtração da imagem em (b) pela imagem em (a) antes do registro. É possível notar, observando a imagem em (c), a diferença entre esses termogramas causada pelos movimento da paciente durante o exame.

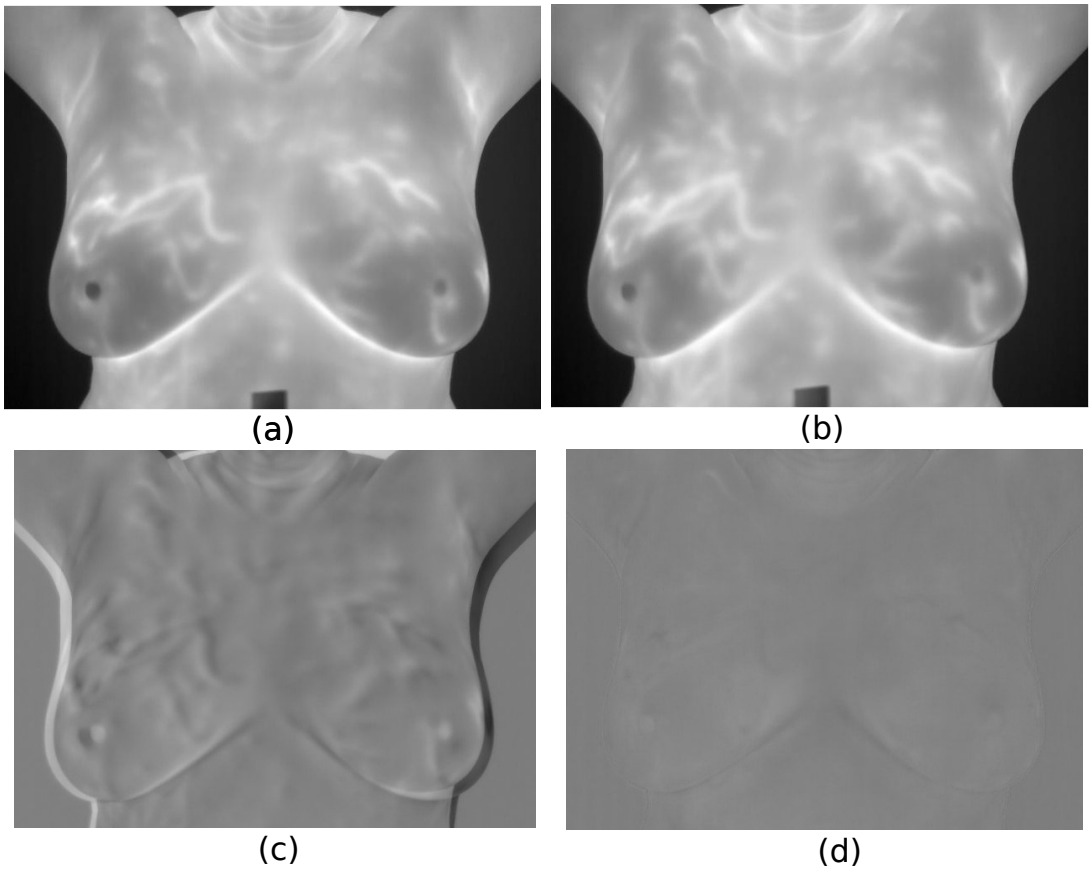

Figura 4.11: Resultado do registro de imagens: (a) a primeira imagem de uma paciente; (b) a décima sétima imagem da mesma paciente; (c) o resultado da subtração dessas duas imagens antes do registro; e (d) o resultado da subtração dessas imagens após o registro.

Com o objetivo de diminuir os efeitos desses movimentos nas séries temporais, o registro de todas as imagens geradas durante o exame é realizado. Como detalhado na Seção 2.3, o registro é um processo no qual as imagens são emparelhadas. Existindo duas imagens (a imagem fixa e a imagem móvel) de uma mesma cena, o registro busca criar um relacionamento entre elas, isto é, o registro pretende alcançar a melhor sobreposição possível para transformar essas imagens independentes em uma imagem comum. Nesta metodologia, o primeiro termograma da sequência é considerado a imagem fixa e os demais são considerados, um por vez, a imagem móvel (que será transformada). Assim, para os

termogramas de uma determinada paciente, o processo de registro é executado 19 vezes, onde todos os termogramas da sequência (exceto o primeiro) são registrados em relação ao primeiro.

As técnicas de registros aplicadas aqui usam a intensidade de pixel para gerar a função de transformação (Seção 2.3), ou seja, usam medidas de similaridade baseadas na intensidade dos pixels. Assim, aplicando a Equação 4.1, as matrizes de temperatura são convertidas para imagens em tons de cinza (imagens (a) e (b) na Figura 4.11), o registro é realizado, e as transformações são transferidas de volta às respectivas matrizes de temperatura.

O registro é executado em dois estágios. No primeiro estágio, Informação Mútua (IM) é usada como medida global de similaridade de intensidade de pixels entre as imagens (fixa e móvel), e a função gerada executa as transformações de translação, rotação e escala, na imagem móvel. O valores de intensidade de pixel de cada imagem são considerados variáveis aleatórias. A IM mede a quantidade de informação que uma variável aleatória contém sobre uma outra e é calculada baseada nas entropias de cada uma das variáveis aleatórias. A entropia de uma variável aleatória discreta X é dada pela Equação 4.2 e a entropia conjunta de duas variáveis aleatórias discretas X e Y é dada pela Equação 4.3. Se X e Y são variáveis aleatórias definindo as intensidades de pixel de duas imagens, respectivamente, então a IM entre essas duas imagens é dada pela Equação 4.4. Nas equações 4.2 e 4.3, $p(x)$ e $p(x, y)$ são funções de densidade de probabilidade e o logaritmo está na base 2 [Mattes et al. 2001].

$$H(X) = -\sum_{x \in X} p(x) log\, p(x) \tag{4.2}$$

$$H(X, Y) = \sum_{x \in X} \sum_{y \in Y} p(x, y) log\, p(x, y) \tag{4.3}$$

$$IM(X, Y) = H(X) + H(Y) - H(X, Y) \tag{4.4}$$

O segundo estágio usa uma medida local de similaridade de intensidade de pixels proposta por Myronenko e Song [Myronenko e Song 2010]. Essa medida considera as distorções complexas de intensidade variando espacialmente na imagem. A Figura 4.12 ilustra os efeitos de cada estágio sobre as imagens: à esquerda, parte da imagem móvel; ao centro, os efeitos do primeiro estágio do registro, onde a região vermelha representa

o quanto a imagem móvel foi transformada (movida) para se aproximar da imagem fixa; à direita, os efeitos do segundo estágio do registro, onde a região vermelha representa o quanto a imagem móvel (resultante do primeiro estágio) foi deformada para se aproximar da imagem fixa. O resultado final do registro pode ser observado na Figura 4.11(d). Tal imagem é o resultado da subtração da imagem em (b) pela imagem em (a) da Figura 4.11, após o registro. A diferença é muito menor do que a da mostrada na imagem em (c), antes do registro.

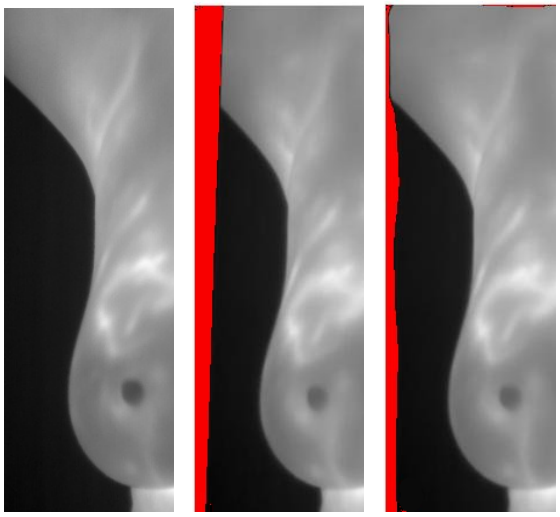

Figura 4.12: Efeitos dos estágios do registro das imagens.

Como citado, o segundo estágio do registro de imagens é um método proposto em [Myronenko e Song 2010]. Porém, a execução somente desse método em algumas imagens não gera bons resultados. Um exemplo de resultado insatisfatório está na Figura 4.13, onde a imagem em (a) é a imagem fixa, a imagem em (b) é a imagem móvel, a imagem em (c) é o resultado do registro executando apenas o segundo estágio e a imagem em (d) é o resultado da subtração da imagem em (c) pela imagem em (a). Isso acontece quando a diferença entre a imagem fixa e a imagem móvel é grande, ou seja, quando a paciente realiza movimentos significativos de uma imagem para a outra, durante a captura da sequência de termogramas.

Para resolver esse problema, foi implementado neste trabalho o registro executado no primeiro estágio. A Figura 4.14 contém o resultado do registro das mesmas imagens da Figura 4.13 quando os registros do primeiro e do segundo estágios são executados em conjunto. As imagens em (a) e em (b) são as mesmas imagens fixa e móvel da Figura 4.13, a imagem em (c) é o resultado do registro implementado no primeiro estágio (a região vermelha indica o quanto a imagem móvel foi deslocada no registro), a imagem em (d) é o resultado do registro do segundo estágio tendo como imagem móvel a imagem em (c),

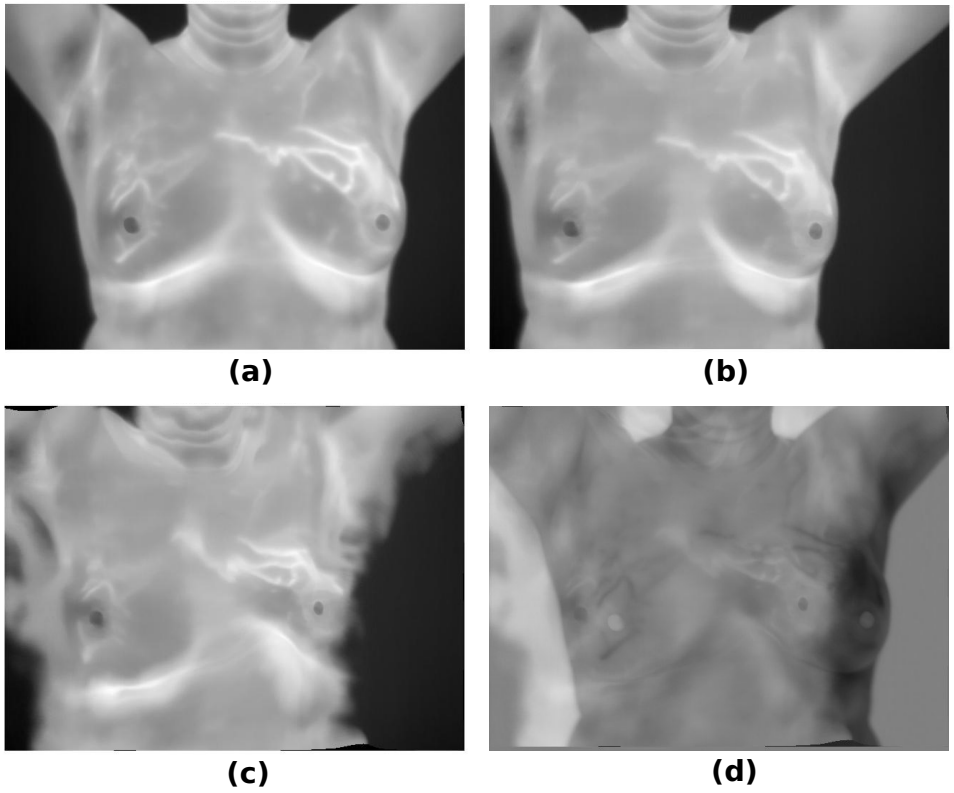

Figura 4.13: Resultado do registro de imagens executando apenas o segundo estágio.

e a imagem em (e) e o resultado da subtração da imagem em (d) pela imagem em (a). É possível perceber o quanto o registro de imagens proposto em [Myronenko e Song 2010] melhora quando o registro implementado no primeiro estágio é executado, comparando a imagem em (c) da Figura 4.13 com a imagem em (d) da Figura 4.14 e o resultado da subtração em (d) da Figura 4.13 com o resultado da subtração em (e) da Figura 4.14.

4.4 Construção das séries temporais de temperatura

Logo após a etapa de registro, a construção das séries temporais (Seção 2.5) de temperatura de uma determinada paciente, examinada por TID (Seção 2.2.2), segue os seguintes passos:

1. a ROI no termograma, correspondendo aos pixels de cor branca na máscara gerada como descrito na Seção 4.3 (Figura 4.15 (a)), é divida em uma "malha" de quadrados R_k de tamanho 11x11 pixels (Figura 4.15 (b)), com $k = 1, 2, ..., p$, onde p é a quantidade de quadrados formados;

2. a temperatura mais alta de cada quadrado R_k é observada em todos os vinte termogramas da paciente (Figura 4.16), produzindo a série temporal $S_k = (t_{k,1}; t_{k,2}; ...; t_{k,20})$.

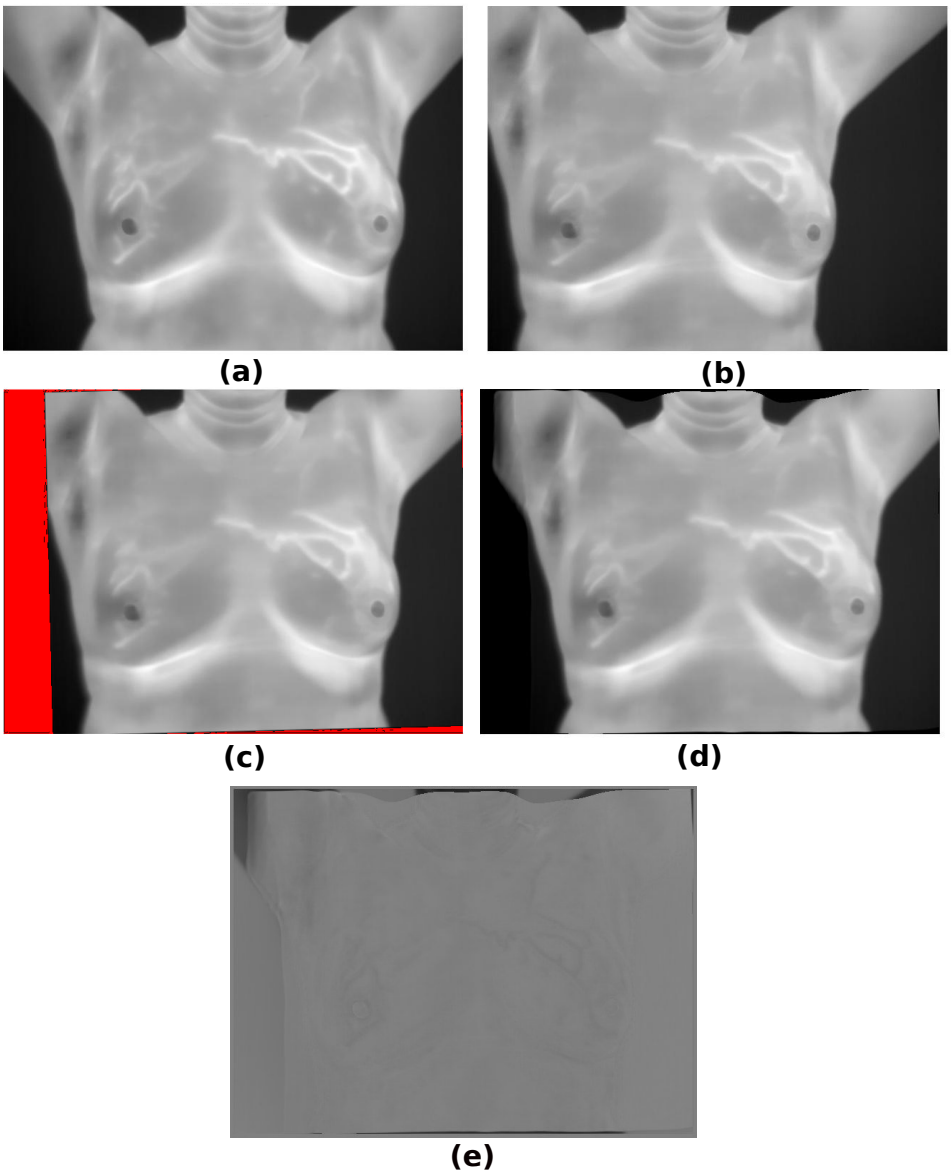

Figura 4.14: Resultado do registro de imagens executando os dois estágios.

É importante destacar que os valores da série S_k estão ordenados cronologicamente, ou seja, $t_{k,1}$ é a temperatura mais alta da região quadrada R_k no primeiro termograma da sequência, $t_{k,2}$ é a temperatura mais alta da região quadrada R_k na segunda imagem da sequência, e assim sucessivamente. Como visto no Capítulo 3, o trabalho descrito em [Anbar et al. 2001] utiliza a temperatura média dos quadrados formados para construir as séries temporais. Testes com séries temporais formadas pela temperatura média foram realizados nesta metodologia, mas os resultados foram inferiores. A temperatura média de uma região quadrada atenua uma das características do tecido canceroso, a presença de temperaturas mais altas quando comparadas a de tecidos circundantes. Além disso, outros tamanhos de R_k foram testados, mas com o tamanho de 11x11 pixels os melhores resultados foram alcançados.

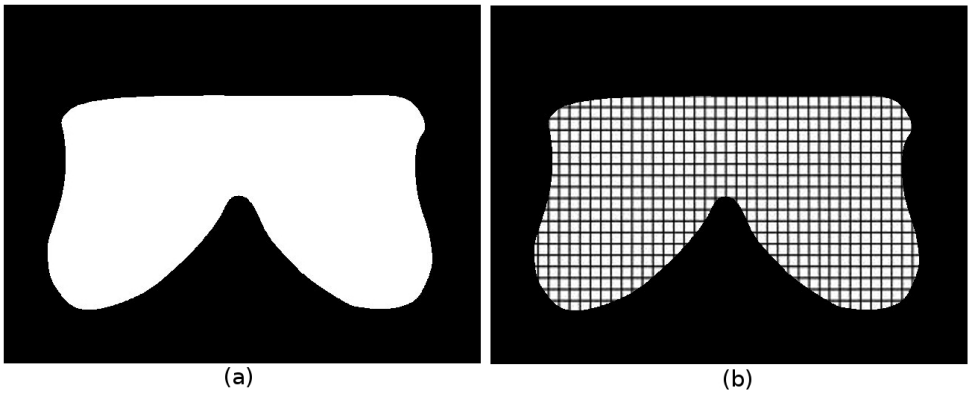

Figura 4.15: Máscara divida em uma "malha" de quadrados de tamanho 11x11 pixels.

4.5 Agrupamento das séries temporais e avaliação dos grupos formados

Para uma determinada paciente i, as séries temporais construídas na etapa anterior formam o conjunto, aqui denominado, X_i. Para separar tais grupos (de regiões saudáveis e de regiões com anomalias), uma técnica de aprendizagem de máquina não-supervisionada (Seção 2.4.1.1) é aplicada por meio de um algoritmo de agrupamento. O algoritmo escolhido para essa tarefa é o *k-means*, maiores detalhes desse algoritmo estão na Seção 2.4.1.2. O *k-means* é executado, tendo como entrada os elementos do conjunto X_i, nove vezes, ou seja, para cada valor de $k \in \{2, 3, ..., 10\}$, gerando o conjunto de resultados de agrupamento $P_i = \{P_{i,2}, P_{i,3}, ..., P_{i,10}\}$, onde $P_{i,2}$ contém 2 grupos de séries temporais, $P_{i,3}$ contém 3 grupos de séries temporais e assim sucessivamente até $P_{i,10}$, que contém 10 grupos. A medida de proximidade utilizada no *k-means* é a *distância Euclidiana* (Seção 2.5).

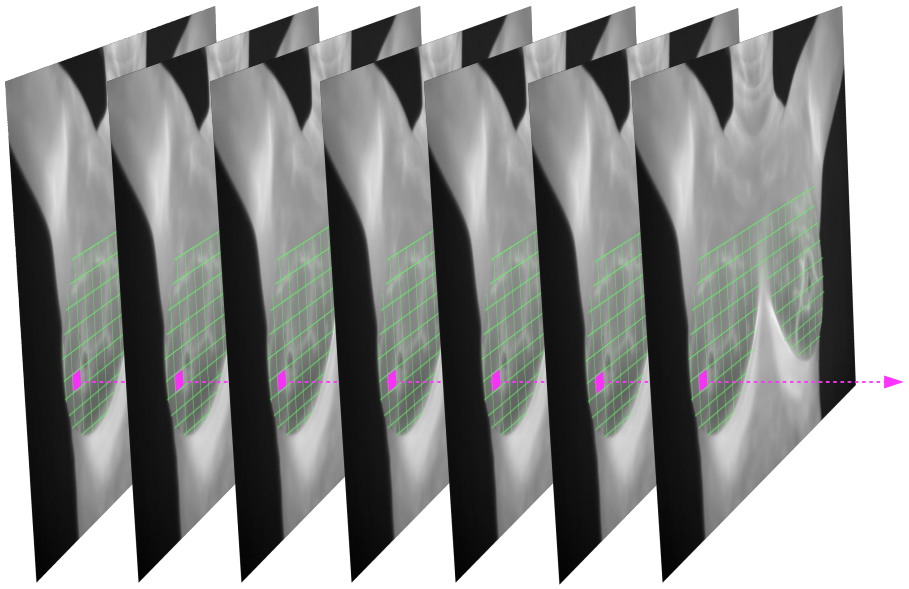

Figura 4.16: Observação da temperatura mais alta de uma determinada região quadrada na mama em todas os termogramas da sequência.

O processamento ocorre na seguinte sequência:

1. o algoritmo *k-means* é executado para $k = 2$ sobre o conjunto X_i (Figura 4.17 (a)), gerando o agrupamento $P_{i,2}$ (Figura 4.17 (b));

2. 10 índices de validação de agrupamento são aplicados para medir a qualidade dos grupos formados no agrupamento $P_{i,2}$, ou seja, para medir o quão compactos e bem separados os grupos são, e os resultados são armazenados no vetor V_i (Figura 4.17 (c)) ;

3. o valor de k é atualizado para $k + 1$ e os passos 1. e 2. são executados novamente.

O processo descrito ocorre até o valor de k igual a 10 ($k = 10$), e ao final o vetor V_i contém 90 valores, ou seja, $V_i = (v_{i,1}, v_{i,2}, ..., v_{i,90})$. Todos os valores do vetor V_i estão no intervalo $[0, 1] \subset \mathbb{R}$, ou seja, $v_{i,j} \in [0, 1] \subset \mathbb{R},\ \forall j \in [2, 10] \subset \mathbb{N}$.

Esses índices são medidas baseadas em critérios internos, uma vez que não se conhece qualquer tipo de estrutura do conjunto de dados X_i (Seção 2.4.1.3). Os índices aplicados são: Silhueta; Davies-Bouldin; Calinski-Harabasz; Dunn; Krzanowski-Lai; Hartigan; Homogeneidade; Separação; Hubert-Levin (Índice C) e Strehl. Todos eles estão descritos na Seção 2.4.1.3. Como descrito anteriormente, para o agrupamento $P_{i,2} \in P_i$ os 10

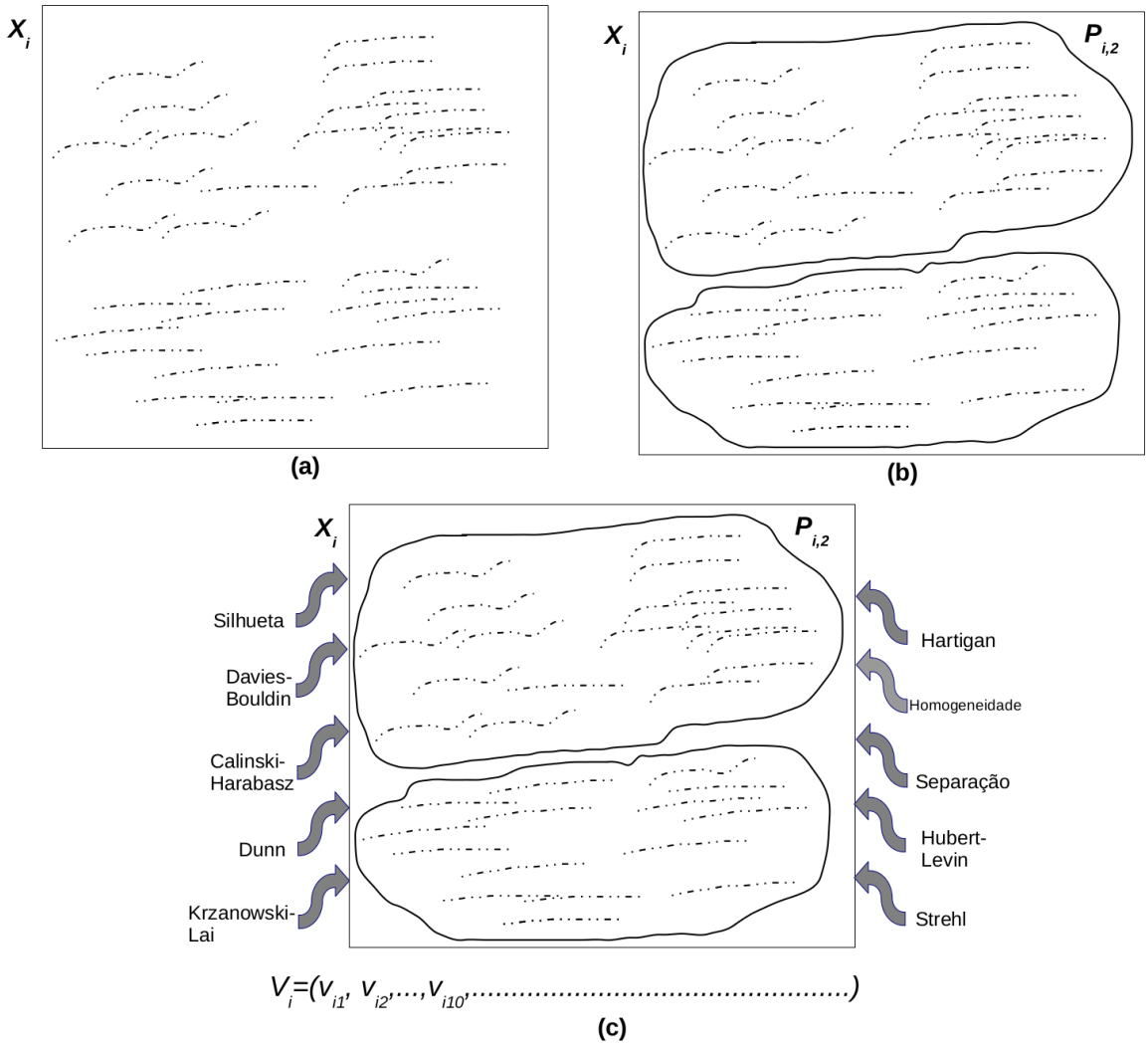

Figura 4.17: Agrupamento e avaliação do agrupamento para $k = 2$.

índices são aplicados, para o agrupamento $P_{i,3} \in P_i$ os mesmos 10 índices são aplicados, e dessa forma, para a paciente i, 90 (noventa) valores de índices (10 índices vezes 9 resultados de agrupamento) são calculados. O vetor $V_i = (v_{i,1}, v_{i,2}, ..., v_{i,90})$ construído é tratado na etapa de classificação como o vetor de características representando a paciente i. Entretanto, nem todos esses valores permanecem no vetor V_i na etapa de classificação. Na avaliação da metodologia proposta nesta tese (Capítulo 5), métodos de seleção de características foram aplicados para determinar as características significativas para a classificação das pacientes entre as classes doente ou saudável. A Tabela 4.1 contém as características selecionadas, são elas: o índice Strehl calculado para $k \in \{3, 4, 6, 7, 8, 9 e 10\}$ Essas 7 características são as propostas nesta tese para serem usadas na etapa de classificação dos dados.

Tabela 4.1: Conjunto das 7 características proposta nesta tese

Índice	Valores de k
Strehl	3, 4, 6, 7, 8, 9 e 10

4.6 Classificação da paciente

Após a aplicação de técnicas de aprendizagem de máquina não-supervisionada para separar as séries temporais e avaliar os grupos formados, a classificação da paciente é realizada considerando-se o vetor de características, construído na etapa anterior, apenas com as características selecionadas (Tabela 4.1). Para uma determinada paciente i, o vetor $V_i = (v_{i,1}, v_{i,2}, ..., v_{i,7})$ é submetido ao modelo de classificação que, de acordo com os valores contidos em V_i, irá classificar a paciente como saudável ou com alguma anomalia de mama.

Na avaliação da metodologia (Capítulo 5), 39 algoritmos de classificação (Seção 2.4.2.2) foram testados, assim como técnicas de seleção de características (Seção 2.4.2.1), combinando 3 métodos de pesquisa e 8 métodos de avaliação. Essa tarefa foi auxiliada por ferramentas de mineração de dados que, além de testar e avaliar os resultados dos algoritmos de classificação e dos métodos de seleção de características, recomendaram o melhor entre eles, respectivamente (ou seja, o de melhor desempenho entre os algoritmos de classificação e o de melhor desempenho entre métodos de seleção de características).

De acordo com os resultados das avaliações, os algoritmos de classificação *k-Star* [Cleary e Trigg 1995] foi o que apresentou os melhores resultados sobre o conjunto de dados utilizado para testes. Por esse motivo, esse algoritmo é o recomendado por esta tese na tarefa de construção do modelo de classificação para classificar pacientes com alguma anomalia e pacientes saudáveis. Esse classificador determinada a classe de cada novo exemplo de acordo com a classe dos seus k vizinhos mais próximos, utilizando uma medida de distância baseada em entropia [Cleary e Trigg 1995].

Este capítulo apresentou a metodologia proposta que se estende do estabelecimento de um protocolo de execução da TID à construção do modelo de classificação das pacientes em doentes ou saudáveis. O próximo capítulo relata as avaliações da metodologia iniciando com a descrição da base de dados utilizada e em seguida detalha os testes realizados. Ele ainda contém as avaliações dos resultados dos testes e uma análise dos resultados.

Capítulo 5

Avaliações da metodologia

Para as avaliações da metodologia proposta nesta tese, foram usados os termogramas de 70 pacientes, 35 sem e 35 com anomalias de mama. Nosso banco de imagens (Seção 5.1) possui atualmente 237 pacientes cadastras, todas com o exame de termografia inserido, mas poucas com dados de outros exames e resultado de biópsia. Porém, o número de pacientes com alguma anomalia de mama ou saudáveis, e que possuíam dados de outros exames e/ou biópsia, estava próximo de 35 para cada caso na ocasião dos testes finais. Então, foi decido usar um conjunto balanceado, ou seja, 35 de cada classe. Neste trabalho, foi considerada como paciente sem anomalias de mama aquela que realizou a mamografia de rastreamento no HUAP e os achados mamográficos foram classificados como BI-RADS de categoria 1 ou 2, e então a paciente recebeu uma recomendação para realizar o exame mamográfico de dois em dois anos. Uma paciente com alguma anomalia de mama é aquela que possui alguma doença em uma ou em ambas as mamas. No conjunto formado para as avaliações da metodologia algumas das anomalias presentes são: fibroadenomas, nódulos, descarga papilar de líquido seroso, tumor filoide de mama, lesão líquida, hiperplasia ductal florida, carcinoma residual, carcinoma papilífero, carcinoma ductal infiltrante e carcinoma lobular *in situ*. Além disso, foi considerada como paciente com alguma anomalia de mama aquela que realizou uma mamografia de rastreamento no HUAP e foi recomendada para complementação com um exame de ultrassonografia, pois os achados mamográficos foram classificados como BI-RADS de categoria 0. O Apêndice A contém a descrição do tipo de anomalia de cada paciente considerada com anomalia de mama e a categoria BI-RADS da mamografia das pacientes consideradas saudáveis, cujos termogramas compuseram o conjunto para as avaliações da metodologia. Os termogramas foram capturados na examinação por TID com a execução do protocolo descrito na Seção 4.2. Para cada uma das 70 pacientes, os 20 termogramas da sequência foram submetidos às

etapas descritas no Capítulo 4. Em seguida, o vetor de características de cada paciente recebeu a informação da classe a qual a paciente pertence. Para uma determinada paciente i, o vetor $V_i = (v_{i,1}, v_{i,2}, ..., v_{i,90})$ recebeu a coordenada $v_{i,91}$, onde $v_{i,91} = 0$ se a paciente possui ambas as mamas saudáveis e $v_{i,91} = 1$ se a paciente possui alguma anomalia em uma das mamas (ou em ambas). O conjunto composto por esses vetores já classificados será denotado por T. O conjunto T foi utilizado por todos os métodos de seleção de características e algoritmos de classificação testados na metodologia. A atribuição do valor da coordenada $v_{i,91}$ para cada vetor do conjunto T foi baseada nas informações clínicas registradas no prontuário arquivado no HUAP, da respectiva paciente. Para facilitar a escolha dos algoritmos de melhor desempenho, as ferramentas de mineração de dados Auto-Weka [Thornton et al. 2013] e EMiner [Marques 2014] [Marques et al. 2015] foram utilizadas. Assim, este capítulo é composto de seções que descrevem os testes e as avaliações dos resultados utilizando os algoritmos e parâmetros recomendados por essas ferramentas e utilizando algoritmos não recomendados, mas frequentemente aplicados em trabalhos de classificação de dados. Entretanto, o capítulo é iniciado apresentando o banco de dados mastológico contendo os termogramas que compuseram o conjunto de dados usado para as avaliações. Discussões das avaliações realizadas e o resumo da metodologia proposta nesta tese estão na Seção 5.6.

5.1 O banco de dados DMR-IR

Os termogramas utilizados na avaliação da metodologia proposta são provenientes do *Database for Mastology Research with Infrared Image* - DMR-IR. O DMR-IR é acessível por uma interface on-line (`http://visual.ic.uff.br/dmi`), destinada ao gerenciamento e recuperação de informações de exames de mama e de dados clínicos das pacientes voluntárias do HUAP. Sua finalidade é oferecer suporte a comunidade científica no desenvolvimento e comparação de metodologias computacionais que auxiliem na detecção e diagnóstico de doenças da mama, principalmente o câncer. A Figura 5.1 exibe uma das páginas de navegação do banco. A aquisição e a utilização desses dados foram aprovados pelo Comitê de Ética em Pesquisa (CEP) do HUAP em 04/06/2012, e o projeto de pesquisa "Aquisição, Armazenamento e Verificação da Viabilidade do Uso de Imagens Térmicas na Detecção de Doenças da Mama", do qual esta tese faz parte, está registrado na Plataforma Brasil, sob o número de Certificado de Apresentação para Apreciação Ética (CAAE) 01042812.0.0000.5243, do Ministério da Saúde. Além dos termogramas, o DMR-IR possui também mamografias (as quatro vistas padrões: médio-lateral oblíqua e

crânio-caudal de cada mama), os laudos das mamografias, resultados de biópsias e ainda informações tais como: idade, queixas, sintomas, hábitos alimentares, histórico pessoal e familiar de doenças e histórico médico. Atualmente são 237 pacientes do HUAP cadastradas no DMR-IR. Maiores detalhes do banco e do protocolo de captura das imagens estão disponíveis em [Silva et al. 2013] e [Silva et al. 2014b].

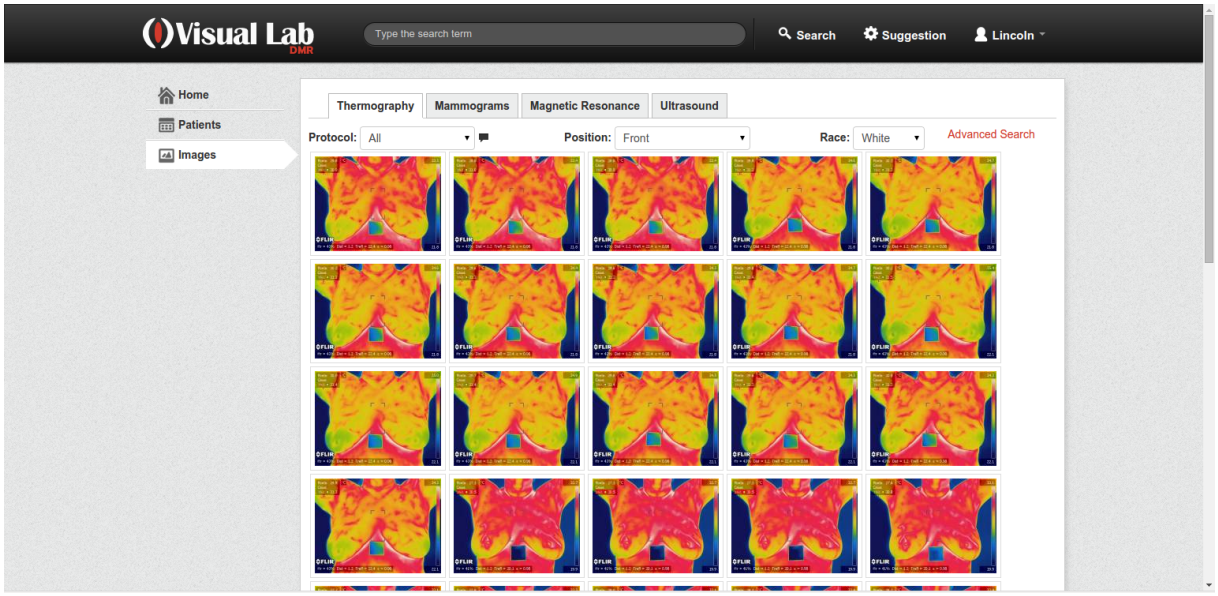

Figura 5.1: Uma das páginas de navegação da DMR-IR.

5.2 Execução do Auto-WEKA, suas recomendações e os resultados obtidos

O Auto-WEKA (Seção 2.4.2.3) foi executado sobre o conjunto T. Os classificadores e métodos de seleção de características testados estão na Tabela 5.1. Como parâmetros de entrada para a execução do Auto-WEKA foram fixados um tempo de execução de 24 horas e a aplicação do método de avaliação *10-fold-Cross-Validation* (Seção 2.4.2.4).

Na Tabela 5.2, encontram-se o método de seleção de características e os valores de seus parâmetros recomendados pelo Auto-WEKA. O algoritmo CFS (*Correlation based Feature Selection*) [Hall 1999] possui como objetivo identificar um subconjunto de características que são altamente correlacionadas com a classe, mas que são pouco correlacionadas umas com as outras. Na parte inferior da mesma Tabela 5.2 estão indicadas as características selecionadas quando as recomendações do Auto-WEKA foram aplicadas no WEKA sobre o conjunto T. Cada "x" marcado nesta tabela indica que o índice em tal linha, calculado para o número de grupos (valor de k) em tal coluna, foi uma característica selecionada.

O algoritmo de classificação recomendado e seus parâmetros estão na Tabela 5.3. Ainda na Tabela 5.3, encontra-se o resultado das avaliação da metodologia proposta quando essas informações fornecidas pelo Auto-WEKA são usadas no WEKA, aplicando as técnicas de avaliação *10-Fold Cross Validation* e *Leave-One-Out Cross-Validation* (Seção 2.4.2.4). As medidas de avaliação aplicadas são: a *sensibilidade*, a *especificidade*, a *precisão*, a *acurácia* (Seção 2.4.2.4) e a *área sob a curva ROC* (Seção 2.4.2.4). A *matriz de confusão* (Seção 2.4.2.4) de cada avaliação com o classificador recomendado também está na Tabela 5.3.

Tabela 5.1: Classificadores e métodos de seleção de características testados no Auto-WEKA

Classificadores
Bayes Net, Naive Bayes, Naive Bayes Multinomial, Gaussian Process, Linear Regression, Logistic Regression, Multi-Layer Perceptron, Stochastic Gradient Descent, Support Vector Machine, Simple Linear Regression, Simple Logistic Regression, Voted Perceptron, K-Nearest Neighbors, K-Star, Decision Table, Repeated Incremental Pruning to Produce Error Reduction (RIPPER), M5 Rules, 1-R, Projective Adaptive Resonance Theory (PART), 0-R, Decision Stump, C4.5 Decision Tree, Logistic Model Tree, M5 Tree, Random Forest, Random Tree, Reduced Error Pruning Tree (REP Tree), Locally Weighted Learning, AdaBoost M1, Additive Regression, Attribute Selected, Bagging, Classification via Regression, LogitBoost, MultiClass Classifier, Random Committee, Random Subspace, Voting, Stacking
Métodos de seleção de características
Best First, Greedy Stepwise, Ranker, Correlation based Feature Selection (CFS) Subset Eval, Pearson Correlation Eval, Gain Ratio Eval, Info Gain Eval, 1-R Eval, Principal Components Eval, RELIEF Eval, Symmetrical Uncertainty Eval.

5.3 Execução do EMiner, suas recomendações e os resultados obtidos

O EMiner (Seção 2.4.2.3), assim como o Auto-WEKA, foi executado sobre o conjunto T. Os classificadores testados estão na Tabela 5.4. Não existe no EMiner a fase de seleção de características como no Auto-WEKA, os classificadores são executados com todas as

Tabela 5.2: Recomendações do Auto-WEKA e as características selecionadas.

Recomendações do Auto-WEKA									
Método de pesquisa de características					Best First				
Parâmetros					-D 2 -N 3				
Método de avaliação de características					CFS Subset Eval				
Parâmetros					-M -P 0 -E 0				
Características selecionadas									
Índices \ valor de k	2	3	4	5	6	7	8	9	10
Silhouette		x							
Davies-Bouldin									
Calinski-Harabasz									
Dunn									
Krzanowski-Lai									
Hartigan									
Homogeneidade	x					x			
Separação									
Hubert-Levin		x					x		x
Strehl		x	x		x	x	x	x	x

características dos vetores em T. Dessa forma, o classificador e seus valores de parâmetros recomendado pelo EMiner foram executados no WEKA em dois testes: no primeiro, foram aplicadas apenas as características selecionadas, indicadas na Tabela 5.2; e no segundo, foram aplicadas todas as 90 características. A Tabela 5.5 contém os resultados das avaliações dos dois testes realizados. Na parte superior dessa tabela estão os parâmetros aplicados no AG, o algoritmo de classificação recomendado e os valores de parâmetros recomendados para ele. Na parte inferior estão os resultados das avaliações do algoritmo e seus valores de parâmetros recomendados (aplicando *10-Fold Cross Validation* e *Leave-One-Out Cross-Validation*) usando somente as características selecionadas e usando as 90 características. As medidas aplicadas na avaliação são: a *sensibilidade*, a *especificidade*, a *precisão*, a *acurácia* e a *área sob a curva ROC*. A matriz de confusão de cada avaliação é também apresentada.

5.4 Avaliação com outros classificadores

No primeiro momento da fase de avaliação da metodologia, o WEKA foi utilizado para executar os algoritmos de classificação e seus respectivos parâmetros recomendados pelo Auto-WEKA e pelo EMiner. O objetivo agora é utilizar o WEKA para avaliar a metodologia com algoritmos de classificação frequentemente utilizados em outros trabalhos

Tabela 5.3: Classificador e seus parâmetros recomendados Auto-WEKA

Recomendações do Auto-WEKA				
Classificação				
Classificador selecionado				K-Star
Parâmetros				-B 74 -E -M n
Avaliação da classificação aplicando *10-Fold Cross Validation*				
Sensibilidade	Especificidade	Precisão	Acurácia	Área ROC
100,00%	100,00%	100,00%	100,00%	1,00
Matriz de confusão			Com anomalia	Saudável
Classificado como com anomalias			35	0
Classificado como saudável			0	35
Avaliação do classificador aplicando *Leave-One-Out Cross-Validation*				
Sensibilidade	Especificidade	Precisão	Acurácia	Área ROC
100,00%	100,00%	100,00%	100,00%	1,00
Matriz de confusão			Com anomalia	Saudável
Classificado como com anomalia			35	0
Classificado como saudável			0	35

Tabela 5.4: Classificadores testados no EMiner

Classificadores no EMiner
J48, Random Tree, Reduced Error Pruning Tree (REP Tree), JRip, Projective Adaptive Resonance Theory (PART) e Ripple Down Rule learner (RIDOR)

baseados nas seguintes técnicas: redes Bayesianas, redes neurais, árvore de decisão e máquina de vetores de suporte (*support vector machine* (SVM)). Os algoritmos executados representando, respectivamente, cada uma dessas técnicas são: o *Bayes Net*, o *Multi-Layer Perceptron*, o *J48* e o *LibSVM*. Eles estão relacionados na Tabela 5.1. Assim como em testes anteriores, os classificadores foram treinados e testados tendo como entrada os vetores de características do conjunto T apenas com as características selecionadas, indicadas na Tabela 5.2. Os parâmetros utilizados em cada classificador são os previamente definidos pelo WEKA, ou seja, nenhuma configuração foi realizada em relação aos parâmetros.

A Tabela 5.6 e a Tabela 5.7 mostram os resultados obtidos. Nessas tabelas, cada linha da primeira coluna indica o algoritmo de classificação usado para construir o modelo de classificação. A avaliação dos modelos construídos estão nas colunas seguintes através das medidas: *sensibilidade*, *especificidade*,*precisão*, *acurácia* e *área sob a curva ROC*. A *matriz de confusão* do modelo com melhor desempenho em cada teste é mostrada na parte inferior da respectiva tabela. Os resultados obtidos na Tabela 5.6 foram avaliados aplicando a técnica de avaliação *10-fold cross validation*, e na Tabela 5.7, aplicando a

Tabela 5.5: As recomendações do EMiner e as avaliações dos testes.

Parâmetros para o AG			
Tamanho da população	Número de gerações	Taxa de crossover	Tipo de população inicial
20	100	0,7	Aleatória

Recomendações do EMiner	
Classificador	Random Tree
Parâmetros	-K, 30, -M, 21.76570735529634, -S, 1093334877

Avaliações dos testes com o classificador recomendado				
Usando as 90 características e aplicando *10-Fold Cross Validation*				
Sensibilidade	Especificidade	Precisão	Acurácia	Área ROC
100,00%	97,14%	98,60%	98,57%	0.99
		Matriz de confusão	Com anomalia	Saudável
		Classificado como com anomalia	35	0
		Classificado como saudável	1	34
Usando as 90 características e aplicando *Leave-One-Out Cross-Validation*				
Sensibilidade	Especificidade	Precisão	Acurácia	Área ROC
88,57%	88,57%	88,60%	88,57%	0,82
		Matriz de confusão	Com anomalias	Saudável
		Classificado como com anomalias	31	4
		Classificado como saudável	4	31
Usando somente as características selecionadas e aplicando *10-Fold Cross Validation*				
Sensibilidade	Especificidade	Precisão	Acurácia	Área ROC
94,28%	88,57%	91,60%	91,43%	0,86
		Matriz de confusão	Com anomalias	Saudável
		Classificado como com anomalias	33	2
		Classificado como saudável	4	31
Usando somente as características selecionadas e aplicando *Leave-One-Out Cross-Validation*				
Sensibilidade	Especificidade	Precisão	Acurácia	Área ROC
94,28%	94,28%	94,30%	94,28%	0,89
		Matriz de confusão	Com anomalias	Saudável
		Classificado como com anomalias	33	2
		Classificado como saudável	2	33

técnica de avaliação *Leave-One-Out Cross-Validation*.

5.5 Análises complementares

Em relação aos índices de validação de agrupamento, o mais discriminativo, ou seja, o que mais esteve presente entre as características selecionadas foi o *Strehl*. Isso aconteceu para $k \in \{3, 4, 6, 7, 8, 9, 10\}$, onde k é o número de grupos formados pelo *k-means*. Além dele, o índice *Hubert-Levin* (para $k \in \{3, 8, 10\}$), o índice *Homogeneidade* (para $k \in \{2, 7\}$)

Tabela 5.6: Avaliação com outros classificadores usando *10-Fold Cross Validation*

Classificadores	Sensibilidade	Especificidade	Precisão	Acurácia	Área ROC
Bayes Net	100,00%	100,00%	100,00%	100,00%	1,00
Multi-Layer Perceptron	94,28%	94,28%	94,30%	94,28%	0,97
LibSVM	94,28%	100,00%	97,30%	97,14%	0,97
J48	94,28%	88,57%	91,60%	91,43%	0,86
Média	95,71%	95,71%	95,8%	95,71%	0,95
Matriz de confusão da classificação com Bayes Net				Com anomalias	Saudável
			Classificado como com anomalias	35	0
			Classificado como saudável	0	35

Tabela 5.7: Avaliação com outros classificadores usando *Leave-One-Out Cross-Validation*

Classificadores	Sensibilidade	Especificidade	Precisão	Acurácia	Área ROC
Bayes Net	100,00%	100,00%	100,00%	100,00%	1,000
Multi-Layer Perceptron	91,43%	94,28%	92,90%	92,86%	0,97
LibSVM	94,28%	100,00%	97,30%	97,14%	0,97
J48	94,28%	91,43%	92,90%	92,86%	0,86
Média	94,99%	96,42%	95,77%	95,71%	0,95
Matriz de confusão da classificação com Bayes Net				Com anomalias	Saudável
			Classificado como com anomalias	35	0
			Classificado como saudável	0	35

e o índice *Silhueta* (para $k = 3$) também estiveram presentes entre as características selecionadas (Tabela 5.2).

Uma investigação foi realizada para averiguar o motivo da superioridade do índice Strehl sobre os outros índices na tarefa de separar as pacientes doentes das saudáveis, o motivo que fez com que ele aparecesse tantas vezes no vetor das características selecionadas. Cada índice, para cada valor de k, foi analisado através de gráficos de linha construídos a partir de pontos que representavam os valores do índice, calculado para cada paciente. Observando esses gráficos foi possível perceber que o índice Strehl de fato consegue praticamente separar os dois tipos de pacientes (doente e saudável), para quase todos os valores de k. Algo semelhante não aconteceu para os outros índices testados. As figuras 5.2, 5.3, 5.4, 5.5, 5.6, 5.7 e 5.8 contém os gráficos com o cálculo do índice Strehl para $k \in \{3, 4, 6, 7, 8, 9, 10\}$, ou seja, as características que estiveram entre as selecionadas. Nos gráficos, o eixo horizontal representa as pacientes, o eixo vertical é a escala de valores do índice Strehl, a linha vermelha representa as pacientes doentes e a linha verde as pacientes saudáveis. Assim, utilizando os valores desse índice quando ele é usado para avaliar os agrupamentos formados por $3, 4, 6, 7, 8, 9$ e 10 grupos de séries temporais é possível praticamente separar as pacientes doentes das saudáveis. Provavelmente

a eficiência desse índice seja consequência de sua formulação, que pondera as similaridades intragrupo e intergrupos pelo tamanho de cada grupo envolvido no cálculo de tais similaridades (Seção 2.4.1.3).

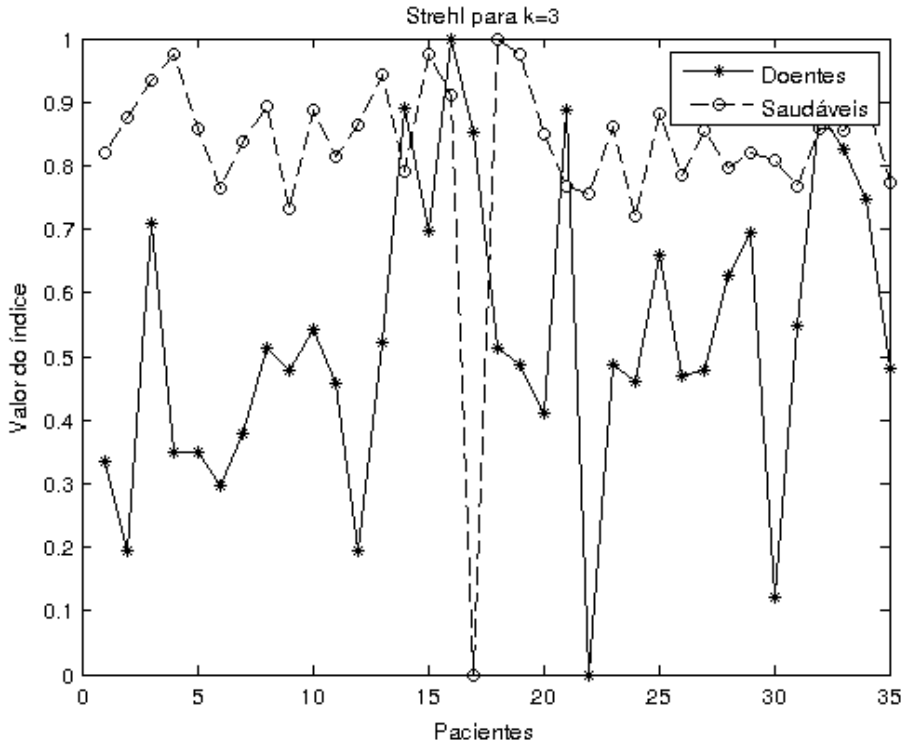

Figura 5.2: Cálculo do índice Strehl para $k = 3$.

Um teste complementar realizado foi considerar como características apenas o valores do índice Strehl para $k \in \{2, 3, 4, 5, 6, 7, 8, 9, 10\}$, formando o vetor $V_i = (v_{i,1}, v_{i,2}, ..., v_{i,9})$, ou seja, 9 características apenas, de uma determinada paciente i. O conjunto T desses vetores para todas as 70 paciente foi submetido novamente ao método de seleção de características com seus respectivos parâmetros indicados na Tabela 5.2. O novo conjunto das características selecionadas está na Tabela 5.8, as características selecionadas são as mesmas da última linha da Tabela 5.2. Esse conjunto de características, com apenas 7, foi usada para construir o modelo de classificação executando apenas o *K-Star* e *Bayes Net*, os de melhores desempenho em testes anteriores, na ferramenta WEKA. Os resultados estão na Tabela 5.9, usando a técnica *10-fold Cross-Validation*, e na Tabela 5.10, usando a técnica *Leave-One-Out Cross-Validation*. Como pode ser observado, somente com as 7 características apontadas na Tabela 5.8 é possível obter o mesmo desempenho na classificação com o modelo construído pelo classificador *Bayes Net*, ou seja, acurácia de 100%. O modelo criado pelo *K-Star* classificou apenas um caso erroneamente, alcançando uma acurácia de 98,57%. É possível concluir que o desempenho da metodologia não diminui

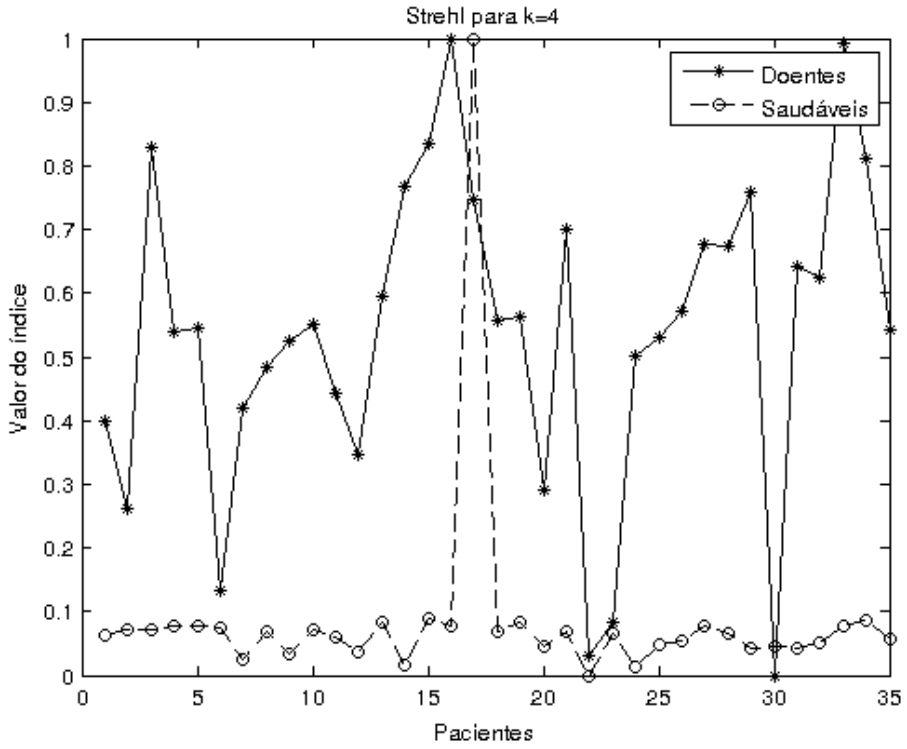

Figura 5.3: Cálculo do índice Strehl para $k = 4$.

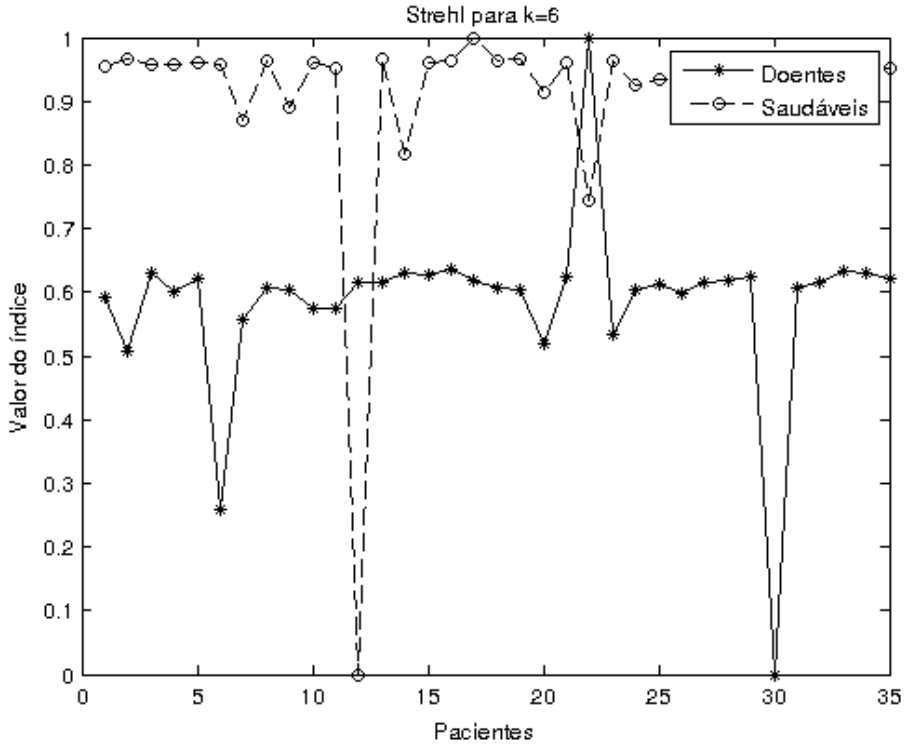

Figura 5.4: Cálculo do índice Strehl para $k = 6$.

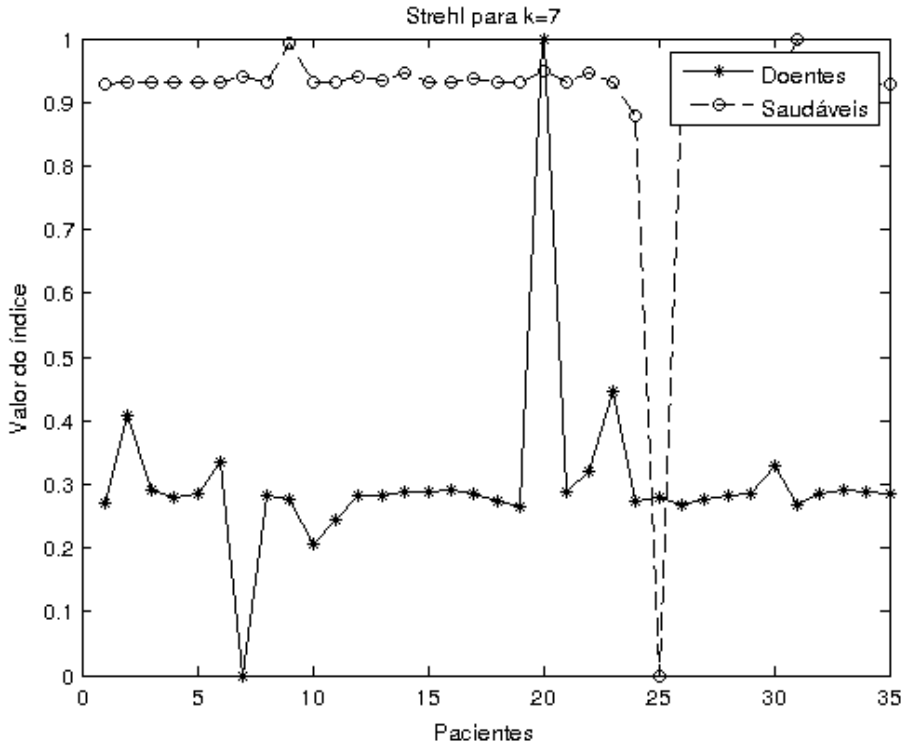

Figura 5.5: Cálculo do índice Strehl para $k = 7$.

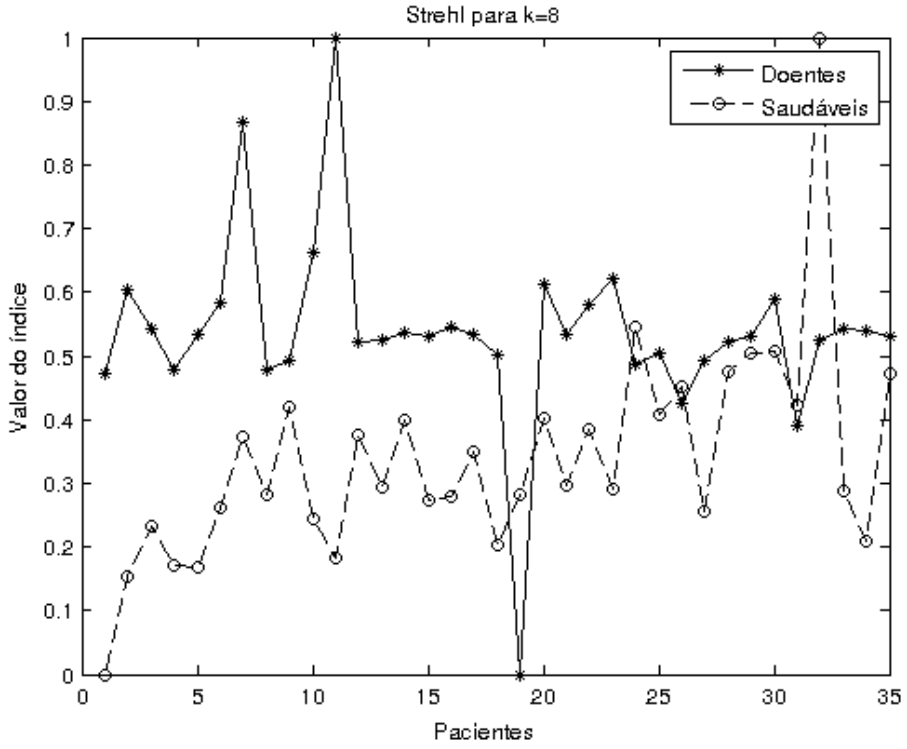

Figura 5.6: Cálculo do índice Strehl para $k = 8$.

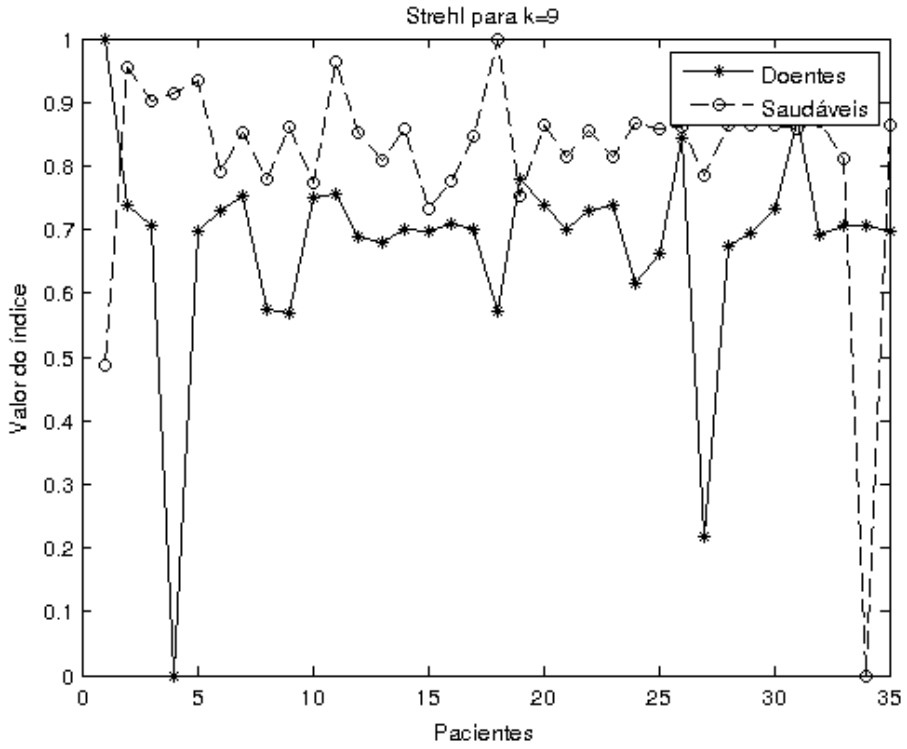

Figura 5.7: Cálculo do índice Strehl para $k = 9$.

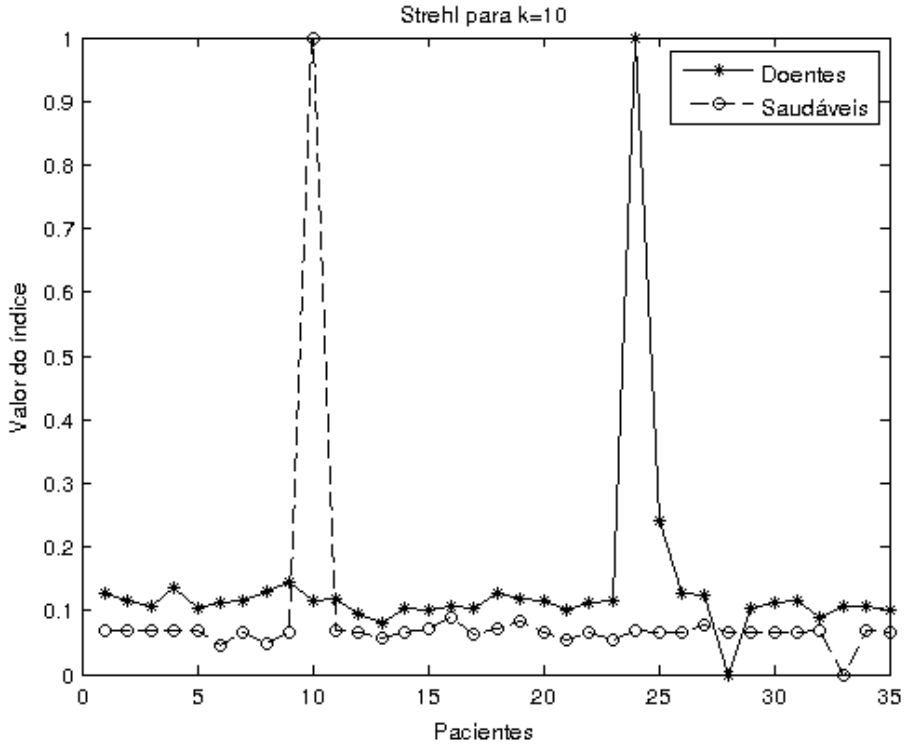

Figura 5.8: Cálculo do índice Strehl para $k = 10$.

quando o número de características diminui de 90 para 9, antes da seleção de características, e de 13 para 7, após a seleção de características. Além disso, a metodologia tornou-se mais simples por diminuir o custo computacional deixando de computar 9 índices para 9 valores de k, ou seja, 81 cálculos a menos, sendo que o cálculos de alguns desses 9 índices é extremamente custoso, como é o caso do índice Dunn.

Tabela 5.8: Conjunto das características selecionadas usando somente o índice Strehl.

Índice	Valores de k
Strehl	3, 4, 6, 7, 8, 9 e 10

Tabela 5.9: Avaliação complementar usando *10-fold Cross-Validation*

Classificadores	Sensibilidade	Especificidade	Precisão	Acurária	Área ROC
Bayes Net	100,00%	100,00%	100,00%	100,00%	1,00
K-Star	97,14%	100,00%	98,60%	98,57%	0,99
Média	98,57%	100,00%	99,30%	99,28%	0,99
Matriz de confusão da classificação com Bayes Net				Com anomalias	Saudável
			Classificado como com anomalias	35	0
			Classificado como saudável	0	35

Tabela 5.10: Avaliação complementar usando *Leave-One-Out Cross-Validation*

Classificadores	Sensibilidade	Especificidade	Precisão	Acurária	Área ROC
Bayes Net	100,00%	100,00%	100,00%	100,00%	1,00
K-Star	97,14%	100,00%	98,60%	98,57%	1,00
Média	98,57%	100,00%	99,30%	99,28%	1,00
Matriz de confusão da classificação com Bayes Net				Com anomalias	Saudável
			Classificado como com anomalias	35	0
			Classificado como saudável	0	35

Para verificar o desempenho dos modelos de classificação construídos com o *K-Star* e *Bayes Net* na ferramenta WEKA para pacientes fora do conjunto das 70 pacientes utilizadas na avaliação da metodologia, um conjunto com os dados de 6 pacientes (3 com e 3 sem anormalidades de mama) foi utilizado. Os dados dessas 6 pacientes são totalmente desconhecidos para os modelos classificação, pois em momento algum estiveram no conjunto de treinamento de tais modelos. O melhor resultado foi obtido com o *K-Star*, onde a classificação resultou em um único falso positivo, o que é menos grave do que um falso negativo em uma tarefa de diagnóstico médico. A acurácia foi de 83,33%.

5.6 Discussões e resumo da metologia proposta

Observando as avaliações de todos os testes realizados, algumas conclusões podem ser construídas. Originalmente, 90 características são calculadas, mas análises complementares 5.5, motivadas pela observação das características selecionadas (Tabela 5.2), revelaram a superioridade do índice Strehl em relação aos outros índices, na tarefa de separar as pacientes doentes das saudáveis, calculado para $k \in \{3, 4, 6, 7, 8, 9, 10\}$.

Em relação a construção do modelo de classificação, a metodologia obteve resultados satisfatórios na execução dos algoritmos de classificação: *Multi-Layer Perceptron*, *LibSVM* e o *J48*. A acurácia média obtida entre os algoritmos *Bayes Net*, *Multi-Layer Perceptron*, *LibSVM* e o *J48* foi de 95,71%, onde a menor obtida foi de 91,43%, com o algoritmo *J48*, aplicando *10-Fold Cross Validation*. O algoritmo de classificação recomendado pelo EMiner, o *Random Tree*, obteve bons resultados, mas inferiores aos obtidos com o algoritmo de classificação recomendado pelo Auto-WEKA. Porém, o número de classificadores testados no EMiner é significativamente inferior quando comparado ao número desses testado no Auto-WEKA. Provavelmente esse seja o motivo do EMiner não recomendar um algoritmo mais eficiente com o conjunto T. Mesmo assim, com o *Random Tree* a metodologia obteve acurácia de 98,57% usando as 90 características e aplicando a técnica *10-Fold Cross Validation* na avaliação. Mas os melhores resultados na classificação foram alcançados com algoritmo de classificação *K-Star*, com os parâmetros indicados na Tabela 5.3 (recomendação do Auto-WEKA), e o algoritmo de classificação *Bayes Net*, com parâmetros pré-definidos. O modelo de classificação construídos por esses algoritmos alcançaram acurácia de 100%, utilizando 13 características selecionadas das 90 originalmente calculadas, e acurácia média de 99,28% utilizando apenas 7 características das 9 posteriormente calculadas utilizando apenas o índice Strehl ($k \in \{3, 4, 6, 7, 8, 9, 10\}$).

Os resultados obtidos não podem ser comparados com os resultados de trabalhos relacionados encontrados na literatura que possuem o mesmo objetivo de detectar anomalias de mama por meio de análises sobre os termogramas obtidos por TID (Capítulo 3), pois esses trabalhos não utilizam aprendizagem de máquina para gerar um modelo de classificação. Eles apenas geram gráficos ou imagens que auxiliam, visualmente, o profissional de saúde. Mas quando os resultados são comparados com os resultados de metodologias que detectam anomalias de mama analisando termografias obtidas por Termografia Infravermelha Estática (TI), é possível perceber a superioridade da metodologia proposta aqui. Os melhores resultados utilizando termogramas obtidos por TI são obtidos por Acharya [Acharya et al. 2012], acurácia de 88,1%, seguido por Borchartt [Borchartt 2013],

acurácia de 88%. Essa superioridade é provavelmente uma consequência do uso da informação temporal utilizada na TID, e que não é utilizada na TI. Como comentado na Seção 2.2.2, os vasos sanguíneos produzidos por tumores cancerosos praticamente não possuem terminação nervosa como os vasos embriológicos. Esses vasos são apenas tubos endoteliais e por isso não se contraem em resposta a um estímulo simpático. Dessa forma, a região do câncer permanece com a temperatura praticamente inalterada quando a mama é resfriada. Mediante a redução do diâmetro vascular, regiões saudáveis da mama apresentam redução da temperatura. Esse fator manifesta-se somente no exame realizado por TID.

Após a conclusão de todos os testes realizados na avaliação e análise dos resultados obtidos, a metodologia proposta encontra-se na Tabela 5.11:

Tabela 5.11: Resumo da metodologia proposta nesta tese.

Etapas	Proposta
Protocolo de aquisição da TID	o protocolo descrito na Seção 4.2
A segmentação das imagens	a segmentação descrita na Seção 4.3
O registro das imagens	o registro descrito na Seção 4.3
A construção das séries temporais de temperatura	a construção descrita na Seção 4.4
O agrupamento dos dados	o agrupamento descrito na Seção 4.5
As características para a construção do modelo de classificação	as 7 características da Tabela 5.8
Os algoritmos de classificação para construção do modelo de classificação	*K-Star*

O motivo de indicar apenas o *K-Star* e não também o *Bayes Net* está na classificação das 6 pacientes desconhecidas aos modelos de classificação, pois com o *K-Star* foi obtido uma acurácia de 83,33% e com o *Bayes Net*, uma acurácia de 33,33%.

Este capítulo apresentou as avaliações dos testes realizados e as análises dos resultados obtidos concluindo com o resumo da metodologia proposta nesta tese. O próximo capítulo contém as conclusões finais, contribuições e trabalhos futuros.

Capítulo 6

Conclusão

O câncer de mama tem sido a causa de morte de muitas mulheres ao redor do mundo. Este trabalho explorou o fato que regiões da mama com anomalias produzem séries temporais de temperatura com alteração de comportamento quando analisadas, para identificar pacientes com tais anomalias após examinação por Termografia Infravermelha Dinâmica (TID). Para alcançar esse objetivo, um protocolo de execução da TID foi estabelecido após um levantamento bibliográfico e vários testes em voluntárias. A região das mamas na sequência de termogramas capturada por uma câmera infravermelha é segmentada e as imagens da sequência são registradas. Então, séries temporais de temperatura são construídas. O algoritmo *k-means* foi aplicado sobre essas séries usando vários valores de k e índices de validação de agrupamento foram aplicados para avaliar os grupos formados, gerando valores tratados como características na etapa de construção do modelo de classificação. Algumas técnicas de seleção de características (combinando 3 métodos de pesquisa e 8 métodos de avaliação) e 39 classificadores, como também suas respectivas configurações de parâmetros, foram testados. A acurácia de 100% foi obtida na classificação com o algoritmo *K-Star* recomendado pelo Auto-WEKA, umas das ferramentas de mineração de dados utilizadas para a resolução do problema de seleção de algoritmos e otimização de parâmetros em problemas de classificação. A metologia também foi avaliada com os algoritmos *Bayes Net*, *Multi-Layer Perceptron*, *LibSVM* e o *J48* e acurácia média entre eles foi de 95,71%, onde a menor obtida foi de 91,43%, com o algoritmo *J48* e a maior foi de 100%, com o algoritmo *Bayes Net*.

Os resultados obtidos não podem ser comparados com os resultados de trabalhos relacionados encontrados na literatura que possuem o mesmo objetivo de detectar anomalias de mama por meio de análises sobre os termogramas obtidos por TID (Capítulo 3), pois esses trabalhos não utilizam aprendizagem de máquina para gerar um modelo de classi-

ficação. Eles apenas geram gráficos ou imagens que auxiliam, visualmente, o profissional de saúde. Mas quando os resultados são comparados com os resultados de metodologias que detectam anomalias de mama analisando termografias obtidas por Termografia Infravermelha Estática (TI), é possível perceber a superioridade da metodologia proposta aqui. Os melhores resultados utilizando termogramas obtidos por TI são obtidos por Acharya [Acharya et al. 2012], acurácia de 88,1%, seguido por Borchartt [Borchartt 2013], acurácia de 88%.

É importante lembrar que uma acurácia de 100% foi alcançada sobre o conjunto de dados utilizado para as avaliações e exaustivos testes, mas a metodologia alcançou uma acurácia de 83,33% na classificação de pacientes fora do conjunto de avaliação, gerando em somente um falso positivo. Dessa forma, os resultados comprovam a habilidade da metodologia proposta nesta tese em detectar anomalias de mama, e sistemas baseados nela podem servir como um exame de rastreamento em clínicas e hospitais. O objetivo é que a termografia infravermelha de mama seja um exame aplicado para determinar a população-alvo em programas de rastreamento organizado de câncer de mama e com isso contribuir na utilização mais racional de exames de rastreamento como a mamografia. Isso torna-se importante mediante a dificuldade de acesso ao exame por imagem mais básico de rastreamento do câncer de mama, a mamografia, principalmente pelas mulheres de baixa escolaridade e classe socioeconômica, devido aos custos do exame e pela distribuição desigual dos mamógrafos em todo território brasileiro. A termografia infravermelha de mama é inofensiva a paciente (é não invasiva, é indolor e não emite radiação) e por esse motivo poderia ser repetida quantas vezes fossem necessárias em uma mesma pessoa, caso seja usada na determinação da população-alvo. Além disso, possui um custo extremamente baixo quando comparado aos demais exames de mama. Portanto, o presente trabalho é uma contribuição para o uso da TID no rastreamento do câncer de mama.

6.1 Contribuições

A principal contribuição desta tese é o desenvolvimento de uma metodologia computacional para detectar anomalias de mama em pacientes examinadas por TID, utilizando técnicas de aprendizagem de máquina, para que sistemas de exame baseados na metodologia proposta possam ser usados em programas de rastreamento organizado de câncer de mama, contribuindo na definição da população-alvo. Existem vários trabalhos que realizam análise de termogramas obtidos por TID (Capítulo 3), mas nenhum utiliza técnicas de aprendizagem de máquina para classificar os dados obtidos. Além disso, os melhor

resultados alcançado entre esses trabalho é uma sensibilidade e uma especificidade de 95%, segundo os autores. Como consequência da contribuição principal, existem algumas contribuições secundárias:

1. adaptação do registro computacional de imagens baseado em intensidade de pixel proposto em [Myronenko e Song 2010] para os termogramas obtidos por TID da base de dados *Database for Mastology Research with Infrared Image* (DMR-IR);
2. o estabelecimento de um conjunto de características, utilizado na classificação da paciente;
3. a indicação de um algoritmo de classificação para a construção do modelo de classificação, e suas respectivas configurações de parâmetros;
4. o estabelecimento de um protocolo de aquisição dos termogramas na examinação por TID; e
5. a disponibilidade de todo o material produzido no endereço eletrônico http://visual.ic.uff.br/ (a segmentação das imagens, as imagens registradas e as características computadas para o desenvolvimento de ferramentas de mineração de dados, tais como classificadores e métodos de seleção de características).

6.2 Trabalhos futuros

A metodologia proposta nesta tese é composta por várias etapas. Assim, trabalhos futuros podem ser focados em melhorar ou modificar cada uma dessas ou realizar uma modificação combinada em todas essas etapas.

Um trabalho futuro imediato seria realizar a classificação por mama e não apenas por paciente. A ROI seria dividida em duas regiões: mama esquerda e mama direita. Cada uma dessas regiões seria submetida as etapas: construção das séries temporais de temperatura; agrupamento das séries temporais de temperatura; avaliação do resultado do agrupamento; e classificação da região (mama) em com alguma anomalia ou saudável. A vantagem em relação a metodologia proposta nesta tese seria: não só indicar se a paciente está com alguma anomalia de mama, mas em qual das mamas ou se está em ambas as mamas.

Um segundo trabalho futuro seria realizar a segmentação automática da ROI por métodos computacionais para termogramas capturados por examinação da TID, visto

que essa tarefa foi realizada manualmente neste trabalho, devido às limitações técnicas de trabalhos anteriores 4.1. Alcançar a segmentação automática é de suma importância para que todo o sistema seja independente de um operador humano, que teria que segmentar manualmente a primeira imagem da sequência de termogramas.

A metodologia proposta apenas indica se a paciente possui um anormalidade de mama, mas não fornece a informação da posição dessa anormalidade, ou seja, em qual mama e em qual quadrante dessa mama a anomalia se encontra. Um outro trabalho futuro seria guardar a posição (x, y), no termograma, de todas as séries temporais de temperatura construídas. Isso possibilitaria, após constatar que a paciente possui uma anomalia de mama, identificar quais são as séries temporais que possuem comportamento diferente das demais e, por meio das posições (x, y) dessas séries, localizar, na termografia, o local, na superfície da mama, da anormalidade. Essa localização facilitaria análises posteriores com outros exames.

Na avaliação da metodologia proposta apenas anormalidade que provocam um aquecimento local estiveram entre os casos usados para os testes. Um trabalho futuro seria verificar o desempenho da metodologia para anormalidades que geram regiões mais frias na superfície da mama. Caso não exista dados suficientes para tal tarefa, dados sintéticos podem ser usando simulando series temporais de temperatura de valores mais quentes e de valores mais frios do que de regiões saudáveis.

O objetivo desse trabalho era determinar se a paciente possui alguma anormalidade de mama ou não, tal objetivo foi alcançado. Porém, como trabalho um trabalho futuro, o objetivo é determinar se a anomalia encontrada é benigna ou maligna. Assim, poderia atribuir-se um grau maior de prioridade para as pacientes com achados malignos para que essas realizassem exames posteriores de rastreamento do câncer de mama. Uma trabalho mais pretensioso seria classificar entre os vários tipos de achados benignos e entre os vários tipos de achados malignos.

Outro trabalho futuro é verificar o quanto o ciclo menstrual (o efeito termogênico da progesterona), as diferentes fases do ciclo reprodutivo feminino, um abscesso local (que provoca o aumento do fluxo de sangue na região afetada), e uso de drogas vasoativas para o tratamento da hipertensão arterial (que produzem vasodilatação) influenciam o comportamento das séries temporais de temperatura construídas. Essas interferências seriam inseridas como uma informação adicional para uma classificação mais precisa da paciente entre doente ou saudável.

Nesta metodologia o volume da mama e a profundidade do achado anormal não é

considerado na hora de classificar a paciente como doente ou saudável. Entretanto, estudos mostram que esses dois fatores influenciam na temperatura proveniente do tumor que chega até a superfície da mama [Bezerra 2007]. Um trabalho futuro seria considerar esses dois fatores na classificação. Para isso, trabalhos de reconstrução das mamas, a partir da termografia capturada, já foram desenvolvidos em nosso grupo de pesquisa por Silva [Silva 2010] em sua tese de doutorado e por Araújo [de Araújo 2014], também em uma tese de doutorado.

Após explorar ao máximo as imagens adquiridas pelo atual protocolo de execução da TID, um novo protocolo poderia ser estabelecido para gerar termogramas mais adequados para as análises computacionais. A principal modificação no novo protocolo seria o desenvolvimento de um equipamento dedicado para o exame por TID. Atualmente, a paciente permanece em pé e com os braços sobre a cabeça durante os cinco minutos de captura de imagens. Isso contribui para que a paciente, inevitavelmente, realize movimentos involuntários. Se esses movimentos fossem reduzidos, ajudaria, e muito, o registro das imagens. Tal equipamento deveria, sem causar qualquer tipo de desconforto e sem interferir na temperatura da superfície de pele a ser examinada, estabilizar o corpo da paciente para que a mesma não realize movimentos de tronco e/ou membros superiores. Os trabalhos em [Parisky et al. 2003], [Arora et al. 2008], e em [Wishart et al. 2010] descrevem em detalhes os equipamentos desenvolvidos para a examinação por TID. Mas observando nossa realidade e necessidade, já existe um planejamento para a construção de tal equipamento, que seria composto de uma cama, apoio para os braços, um suporte para a câmera e um suporte de fixação do ventilador elétrico. Nesse novo equipamento, a paciente permaneceria deitada e com os braços afastados do corpo. a câmera e o ventilador estariam posicionados acima da paciente direcionados à região do tórax. As vantagens em posicionar a paciente deitada na cama (como já acontece em muitos exames de mama) seria: uma maior área da mama visualizada pela câmera e uma menor camada de tecido entre o tumor malignos ou benigno e a superfície da pele imediatamente acima; e eliminação quase que total dos movimentos de tronco e/ou membros superiores causados pelo desequilíbrio da paciente, quando essa está em pé, conforme acontece atualmente.

Referências

[Abramoff et al. 2004] Abramoff, M.; Magalhaes, P. e Ram, S. (2004). Image processing with imagej. *Biophotonics International*, 11(7):36–42.

[Acharya et al. 2012] Acharya, U. R.; Ng, E. Y. K.; Tan, J. H. e Sree, S. V. (2012). Thermography based breast cancer detection using texture features and support vector machine. *Journal of Medical Systems*, 36:1503–1510.

[Aggarwal 2014] Aggarwal, C. C. (2014). *Data Classification: Algorithms and Applications*. CRC Press, Taylor and Francis Group.

[Amalu 2004] Amalu, W. C. (2004). Nondestructive testing of the human breast: The validity of dynamic stress testing in medical infrared breast imaging. *Engineering in Medicine and Biology Society*, 1:1174–1177.

[Amalu 2012] Amalu, W. C. (2012). Pacific Chiropractic and Research Center Infrared Imaging - How is Breast Thermography Performed? http://www.breastthermography.com/breastthermographyproc.htm. acessado em 10 de janeiro de 2015.

[Amalu et al. 2008] Amalu, W. C.; Hobbins, W. B.; Head, J. F. e Elliot, R. L. (2008). Infrared imaging of the breast: A review. *CRC Press, Taylor and Francis Group*, pages 9.1–9.22. Livro: Medical Infrared Imaging, editores: Nicholas A. Diakides e Joseph D. Bronzino.

[Anbar 1987] Anbar, M. (1987). Computerized thermography: The emergence of a new diagnostic imaging modality. *International Journal of Technology Assessment in Health Care*, 3:613–621.

[Anbar 2008] Anbar, M. (2008). Dynamic thermal assessment. *CRC Press, Taylor and Francis Group*, pages 9.1–9.22. Livro: Medical Infrared Imaging, Editores: Nicholas A. Diakides e Joseph D. Bronzino.

[Anbar et al. 2000] Anbar, M.; Brown, C.; Milescu, L.; Babalola, J. e Gentner, L. (2000). The potential of dynamic area telethermometry in assessing breast cancer. 19(3):58–62.

[Anbar et al. 2001] Anbar, M.; Milescu, L.; Naumov, A.; Brown, C.; Button, T.; Carty, C. e AlDulaimi, K. (2001). Detection of cancerous breasts by dynamic area telethermometry. 20(5):80–91.

[Arabi et al. 2010] Arabi, P.; Muttan, S. e Suji, R. (2010). Image enhancement for detection of early breast carcinoma by external irradiation. In *Computing Communication and Networking Technologies (ICCCNT), 2010 International Conference on*, pages 1–9.

[Arora et al. 2008] Arora, N.; Martins, D.; Ruggerio, D.; Tousimis, E.; Swistel, A. J.; Osborne, M. P. e Simmons, R. M. (2008). Effectiveness of a noninvasive digital infrared thermal imaging system in the detection of breast cancer. *The American Journal of Surgery*, 196(4):523–526.

[Bezerra 2007] Bezerra, L. A. (2007). Uso de imagens termográficas em tumores mamários para validação de simulação computacional. Dissertação de Mestrado, Universidade Federal de Pernambuco - UFPE, Recife, Pernambuco - Brasil.

[Bolshakova e Azuaje 2003] Bolshakova, N. e Azuaje, F. (2003). Cluster validation techniques for genome expression data. *Signal Processing*, 83(4):825–833.

[Bolshakova e Azuaje 2006] Bolshakova, N. e Azuaje, F. (2006). Estimating the number of clusters in dna microarray data. *Methods of Information in Medicine*, 45(2):153–157.

[Boltzmann 1884] Boltzmann, L. (1884). Ableitung des stefańschen gesetzes, betreffend die abhängigkeit der wärmestrahlung von der temperatur aus der electromagnetischen lichttheorie. *Annalen der Physik*, 258(6):291–294.

[Borchartt 2013] Borchartt, T. B. (2013). *Análise de Imagens Termográficas para a Classificação de Alteração na Mama*. Tese de Doutorado, Instituto de Computação, Universidade Federal Fluminense, Niterói, RJ, Brasil.

[Borchartt et al. 2013] Borchartt, T. B.; Conci, A.; Lima, R. C. F.; Resmini, R. e Sanchez, A. (2013). Breast thermography from an image processing viewpoint: A survey. *Signal Processing*, 93:2785–2803.

[Borchartt et al. 2012] Borchartt, T. B.; Resmini, R.; Motta, L. S.; Clua, E. W.; Conci, A.; Viana, M. J.; Santos, L. C.; Lima, R. C. e Sanchez, A. (2012). Combining approaches for early diagnosis of breast diseases using thermal imaging. *International Journal of Innovative Computing and Applications*, 4(3-4):163–183.

[Bravo 1999] Bravo, R. S. (1999). *Estudo Radiológico do Carcinoma Ductal in Situ da Mama*. Tese de Doutorado, Universidade Federal do Rio de Janeiro, Rio de Janeiro, RJ, Brasil.

[Button et al. 2004] Button, T. M.; Li., H.; Fisher, P.; Rosenblatt, R.; Dulaimy, K.; Li, S.; O'Hea, B.; Salvitti, M.; Geronimo, V.; Geronimo, C.; Jambawalikar, S.; Carvelli1, P. e Weiss, R. (2004). Dynamic infrared imaging for the detection of malignancy. 49(14):3105–3116.

[Caliński e Harabasz 1974] Caliński, T. e Harabasz, J. (1974). A dendrite method for cluster analysis. *Communications in Statistics*, 3(1):1–27.

[Chen et al. 2002] Chen, G.; Jaradat, S. A.; Banerjee, N.; Tanaka, T. S.; Ko, M. S. H. e Zhang, M. Q. (2002). Evaluation and comparison of clustering algorithms in anglying es cell gene expression data. *Statistica Sinica*, 12:241–262.

[Cleary e Trigg 1995] Cleary, J. G. e Trigg, L. E. (1995). K*: an instance-based learner using an entropic distance measure. In *Proc. 12th International Conference on Machine Learning*, pages 108–114. Morgan Kaufmann.

[da Silva 2014] da Silva, J. A. G. (2014). *Estimativa 2014: Incidência de Câncer no Brasil*. Rio de Janeiro: INCA.

[Davies e Bouldin 1979] Davies, D. L. e Bouldin, D. W. (1979). A cluster separation measure. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, PAMI-1(2):224–227.

[de Araújo 2014] de Araújo, F. A. (2014). *Metodologia para Reconstrução Tridimensional da Geometria da Mama Utilizando Dois Sensores de Profundidade*. Tese de Doutorado, Universidade Federal Fluminense, Instituto de Computação, Niterói, RJ, Brasil.

[de Castro Mattos et al. 2013] de Castro Mattos, J. S.; Caleffi, M. e da Costa Vieira, R. A. (2013). Rastreamento mamográfico no brasil: Resultados preliminares. *Revista Brasileira de Mastologia*, 23(1):22–27.

[de Jesus 2005] de Jesus, J. C. (2005). *Ginecologia Fundamental*. Atheneu.

[de Mastologia 2015] de Mastologia, S. B. (2015). Desigualdade na situação dos mamógrafos e das mamografias no brasil. http://www.sbmastologia.com.br/index/index.php/sala-de-imprensa/-releases-/389-desigualdade-na-situacao-dos-mamografos-e-das-mamografias-no-brasil. acessado em 20 de agosto de 2015.

[Dudoit e Fridlyand 2002] Dudoit, S. e Fridlyand, J. (2002). A prediction-based resampling method for estimating the number of clusters in a dataset. *Genome Biology*, 3(7):36.1–36.21.

[Facina 2014] Facina, T. (2014). Estimativa 2014: Incidência de câncer no brasil. 60(1):63–64.

[Flir 2013] Flir (2013). Sc620 infrared camera system. http://www.flir.com/cs/apac/en/view/?id=41413. acessado em 20 de abril de 2015.

[Galvao 2013] Galvao, S. S. L.; Conci, A. G. S. S. L. S. L. F. (2013). Registro de imagens afim para o protocolo dinâmico de aquisição de imagens térmicas da mama. In *IV Encontro Nacional de Engenharia Biomecânica: ENEBI 2013*, pages 158–159, Vitória, ES, Brazil.

[Galvao 2015] Galvao, S. S. L. (2015). *Registro de Imagens Térmicas da Mama Adquiridas Dinamicamente*. Tese de Doutorado, Universidade Federal Fluminense, Instituto de Computação, Niterói, RJ, Brasil.

[Gan et al. 2007] Gan; Guojun; Ma, C. e Wu, J. (2007). *Data Clustering: Theory, Algorithms, and Applications, ASA-SIAM Series on Statistics and Applied Probability, SIAM*. Philadelphia, ASA, Alexandria, VA.

[Gautherie 1985] Gautherie, M. (1985). New protocol for the evaluation of breast thermograms. *in Thermological Methods, VCH mbH*, pages 227–235.

[Gerasimova et al. 2014] Gerasimova, E.; Audit, B.; Roux, S.; Khalil, A.; Gileva, O.; Argoul, F.; Naimark, O. e Arneodo, A. (2014). Wavelet-based multifractal analysis of dynamic infrared thermograms to assist in early breast cancer diagnosis. *Frontiers in Physiology*, 5(176). (2014).

[Gerasimova et al. 2013] Gerasimova, E.; Audit, B.; Roux, S. G.; Khalil, A.; Argoul, F.; Naimark1, O. e Arneodo, A. (2013). Multifractal analysis of dynamic infrared imaging of breast cancer. *EPL (Europhysics Letters)*, 104(6):68001–p1–68001–p6.

[Gerasimova et al. 2012] Gerasimova, E.; Plekhov, O.; Yu.Bayandin; O.Naimark e G.Freynd (2012). Identification of breast cancer using analysis of thermal signals by nonlinear dynamics methods. *International Conference on Quantitative InfraRed Thermography.* , 11-14 June, Naples-Italy, (2012).

[Ghosh e Strehl 2002] Ghosh, J. e Strehl, A. (2002). Clustering and visualization of retail market baskets. In Pal, N. R. e Jain, L., editors, *Knowledge Discovery in Advanced Information Systems*, AIP. Springer. In press.

[Grant 2015] Grant, B. (2015). NCI: Breast Cancer Cases Could Rise by 50 Percent. http://www.the-scientist.com/?articles.view/articleNo/42757/title/NCI–Breast-Cancer-Cases-Could-Rise-by-50-Percent/. acessado em 28 de abril de 2015.

[Guo et al. 2006] Guo, Y.; Sivaramakrishna, R.; Lu, C.-C.; Suri, J. e Laxminarayan, S. (2006). Breast image registration techniques: a survey. *Medical and Biological Engineering and Computing*, 44(1-2):15–26.

[Hajnal et al. 2001] Hajnal, J.; Hawkes, D. e Hill, D. (2001). *Medical Image Registration.* CRC Press, Taylor and Francis Group.

[Halkidi et al. 2001] Halkidi, M.; Batistakis, Y. e Vazirgiannis, M. (2001). On clustering validation techniques. *Journal of Intelligent Information Systems*, 17(2-3):107–145.

[Halkidi et al. 2002] Halkidi, M.; Batistakis, Y. e Vazirgiannis, M. (2002). Clustering validity checking methods: Part ii. *SIGMOD Rec.*, 31(3):19–27.

[Hall et al. 2009] Hall, M.; Frank, E.; Holmes, G.; Pfahringer, B.; Reutemann, P. e Witten, I. H. (2009). The weka data mining software: An update. *SIGKDD Explorations*, 11(1):10–18.

[Hall 1999] Hall, M. A. (1999). *Correlation-based Feature Selection for Machine Learning.* Tese de Doutorado, Department of Computer Science, The University of Waikato, Hamilton, NewZealand.

[Han e Kamber 2006] Han, J. e Kamber, M. (2006). *Data Mining: Concepts and Techniques.* Morgan Kaufmann Publishers, Burlington, MA 01803, USA, 2rd ed edition.

[Hartigan 1985] Hartigan, J. (1985). Statistical theory in clustering. *Journal of Classification*, 2(1):63–76.

[Head e Elliott 2002] Head, J. e Elliott, R. (2002). Infrared imaging: Making progress in fulfilling its medical promise. *IEEE Engineering in Medicine and Biology Magazine*, 21(6):80–85.

[Herman 2013] Herman, C. (2013). The role of dynamic infrared imaging in melanoma diagnosis. *Expert Review of Dermatology*, 8(2):177–184.

[Hewitt 2009] Hewitt, P. G. (2009). *Fundamentos de Física Conceitual, 1ª. ed.* Bookman, Brasil.

[Hobbins 1985] Hobbins, W. B. (1985). Abnormal thermogram significance in breast cancer. *Interamer. J. Rad.*, 12:33.

[Hoeben et al. 2004] Hoeben, A.; Landuyt, B.; Highley, M. S.; Wildiers, H.; Van Oosterom, A. T. e De Bruijn, E. A. (2004). Vascular endothelial growth factor and angiogenesis. *Pharmacological Reviews*, 56(4):549–580.

[Jing et al. 2010] Jing, L.; Ng, M. K. e Zeng, T. (2010). Novel hybrid method for gene selection and cancer prediction. *World Academy of Science, Engineering and Technology*, 4:02–20.

[Kaczmarek e Nowakowski 2004] Kaczmarek, M. e Nowakowski, A. (2004). Active dynamic thermography in mammography. 8(2):259–267.

[Kapoor e Prasad 2010] Kapoor, P. e Prasad, S. V. A. V. (2010). Image processing for early diagnosis of breast cancer using infrared images. *2nd International Conference on Computer and Automation Engineering*, 13(1):564–566.

[Koay et al. 2004] Koay, J.; Herry, C. e Frize, M. (2004). Analysis of breast thermography with an artificial neural network. *Engineering in Medicineand Biology Society - IEMBS*, 1(1):1159–1162.

[Kontos et al. 2011] Kontos, M.; Wilson, R. e Fentiman, I. (2011). Digital infrared thermal imaging (diti) of breast lesions: sensitivity and specificity of detection of primary breast cancers. *Clinical Radiology*, 66(6):536–539.

[Kozak 2012] Kozak, M. (2012). A dendrite method for cluster analysis by caliński and harabasz: A classical work that is far too often incorrectly cited. *Communications in Statistics - Theory and Methods*, 41(12):2279–2280.

[Krzanowski e Lai 1988] Krzanowski, W. J. e Lai, Y. T. (1988). A criterion for determining the number of groups in a data set using sum-of-squares clustering. *Biometrics*, 44(1):23–34.

[Lewis et al. 2012] Lewis, J.; Ackerman, M. e Sa, V. (2012). Human cluster evaluation and formal quality measures: A comparative study. *Proc. 34th Conf. of the Cognitive Science Society CogSci.*

[Liu et al. 2010] Liu, W.; Meyer, J.; Scully, C.; Elster, E. e Gorbach, A. M. (2010). Observing temperature fluctuations in humans using infrared imaging. *10th International Conference on Quantitative InfraRed Thermography, July 27-30th, Québec*, pages 55–64.

[Marques 2012] Marques, R. S. (2012). Segmentação automática das mamas em imagens térmicas. Dissertação de Mestrado, Universidade Federal Fluminense, Instituto de Computação, Niterói, RJ, Brasil.

[Marques 2014] Marques, R. Z. N. (2014). Uma metodologia para seleção de algoritmos de classificação e otimização de parâmetros. Dissertação de Mestrado, Centro de Ciências Exatas e Tecnologia, UFMA, São Luís, Maranhão Brasil.

[Marques et al. 2015] Marques, R. Z. N.; Coutinho, L. R.; Borchartt, T. B.; Vale, S. B. e da Silva Silva, F. J. (2015). Eminer: a tool for selecting classification algorithms and optimal parameters. *Mexican International Conference on Artificial Intelligence-MICAI 2015*.

[Márquez et al. 2015] Márquez, R. S.; Conci, A.; Pérez, M. G.; Andaluz, V. H. e Mejía, T. M. (2015). Una metodología para la segmentación automática en cads de imágenes térmicas. *IEEE Latim America Transactions*.

[Mattes et al. 2001] Mattes, D.; Haynor, D. R.; Vesselle, H.; Lewellyn, T. K. e Eubank, W. (2001). Nonrigid multimodality image registration. *Medical Imaging 2001: Image Processing, SPIE*, 4322:1609–1620.

[Miller et al. 2014] Miller, A. B.; Wall, C.; Baines, C. J.; Sun, P.; To, T. e Narod, S. A. (2014). Twenty five year follow-up for breast cancer incidence and mortality of the canadian national breast screening study: randomised screening trial. *BMJ*, 348.

[Minamimoto et al. 2015] Minamimoto, R.; Senda, M.; Jinnouchi, S.; Terauchi, T.; Yoshida, T. e Inoue, T. (2015). Detection of breast cancer in an fdg-pet cancer screening program: Results of a nationwide japanese survey. *Clinical Breast Cancer*, 15(2):139–146.

[Montoro e Anbar 1988] Montoro, J. e Anbar, M. (1988). New modes of data handling in computerized thermography. *Proceedings of the 10th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, 10:845–847.

[Motta 2010] Motta, L. S. (2010). Obtenção automática da região de interesse em termogramas frontais da mama para o auxílio á detecção precoce de doenças. Dissertação de Mestrado, Universidade Federal Fluminense, Instituto de Computação, Niterói, RJ, Brasil.

[Murphy 2012] Murphy, K. P. (2012). *Machine Learning: a Probabilistic Perspective*. Massachusetts Institute of Technology.

[Myronenko e Song 2010] Myronenko, A. e Song, X. (2010). Intensity-based image registration by minimizing residual complexity. *Medical Imaging, IEEE Transactions on*, 29(11):1882–1891.

[Najarian e Splinter 2005] Najarian, K. e Splinter, R. (2005). *Biomedical Signal and Image Processing*. CRC Press, 1th edition.

[Ng e Kee 2007] Ng, E. Y. K. e Kee, E. C. (2007). Integrative computer-aided diagnostic with breast thermogram. *Journal of Mechanics in Medicine and Biology*, 7(1):1–10.

[Ng et al. 2001] Ng, E. Y. K.; Ung, L. N.; Ng, F. C. e Sim, L. S. J. (2001). Statistical analysis of healthy and malignant breast thermography. *Journal of Medical Engineering and Technology*, 25(6):253–263.

[Ohashi e Uchida 1997] Ohashi, Y. e Uchida, I. (1997). Some considerations on the diagnosis of breast cancer by thermography in patients with nonpalpable breast cancer. 2:670–672.

[Ohashi e Uchida 2000] Ohashi, Y. e Uchida, I. (2000). Applying dynamic thermography in the diagnosis of breast cancer. *IEEE Engineering in Medicine and Biology Magazine*, 19(3):42–51.

[Oliveira 2012] Oliveira, J. P. S. (2012). Extração automática de região de interesse em imagens térmicas laterais da mama. Dissertação de Mestrado, Universidade Federal Fluminense, Instituto de Computação, Niterói, RJ, Brasil.

[Olivera 2013] Olivera, G. O. S. (2013). *Desenvolvimento de um Banco de Imagens Médicas Acessíiveis via Web com Recuperação de Dados Baseada no Conteúdo.* Tese de Doutorado, Universidade Federal Fluminense, Instituto de Computação, Niterói, RJ, Brasil.

[Parisky et al. 2003] Parisky, Y. R.; Sardi, A.; Hamm, R.; Hughes, K.; Esserman, L.; Rust, S. e Callahan, K. (2003). Efficacy of computerized infrared imaging analysis to evaluate mammographically suspicious lesions. 180(1):263–269.

[Resmini 2011] Resmini, R. (2011). Análise de imagens térmicas da mama usando descritores de textura. Dissertação de Mestrado, Universidade Federal Fluminense, Instituto de Computação, Niterói, RJ, Brasil.

[Richie e Swanson 2003] Richie, R. C. e Swanson, J. O. (2003). Breast cancer: A review of the literature. *Journal of Insurance Medicine*, 35:85–101.

[Rousseeuw 1987] Rousseeuw, P. J. (1987). Silhouettes: A graphical aid to the interpretation and validation of cluster analysis. *Journal of Computational and Applied Mathematics*, 20(0):53–65.

[Saeys et al. 2007] Saeys, Y.; naki Inza, I. e naga, P. L. (2007). A review of feature selection techniques in bioinformatics. *Bioinformatics*, 23(19):2507–2517.

[Saniei et al. 2015] Saniei, E.; Setayeshi, S.; Akbari, M. E. e Navid, M. (2015). A vascular network matching in dynamic thermography for breast cancer detection. *Quantitative InfraRed Thermography Journal*, 12(1):24–36.

[Schneider et al. 2012] Schneider, C.; Rasband, W. e Eliceiri, K. (2012). Nih image to imagej: 25 years of image analysis. *Nature Methods*, 9:671–675.

[Serrano 2010] Serrano, R. C. (2010). Análise da viabilidade do uso do coeficiente de hurst e da lacunaridade no auxílio ao diagnóstico precoce de patologias da mama. Dissertação de Mestrado, Universidade Federal Fluminense, Instituto de Computação, Niterói, RJ, Brasil.

[Sharan et al. 2003] Sharan, R.; Maron-Katz, A. e Shamir, R. (2003). Click and expander: a system for clustering and visualizing gene expression data. *Journal of Classification*, 19(14):1787–1799.

[Silva et al. 2015a] Silva, L.; Sequeiros, G.; Santos, M. L.; Fontes, C.; Muchaluat-Saade, D. C. e Conci, A. (2015a). Thermal signal analysis for breast cancer risk verification. *MEDINFO 2015: eHealth-enabled Health, Ebook Series: Studies in Health Technology and Informatics*, 216:746–750.

[Silva et al. 2013] Silva, L. F.; Marques, R. S.; Carvalho, G. S.; Santos, M. L. O.; Fontes, C. A. P.; Santos, A. A. S. M. D. e Conci, A. (2013). Protocolo de captura de imagens térmicas da mama para construção de um banco público de exames. In *IV Encontro Nacional de Engenharia Biomecânica: ENEBI 2013*, pages 104–105, Vitória, ES, Brazil.

[Silva et al. 2015b] Silva, L. F.; Olivera, G. O. S.; Borchartt, T. B.; Resmini, R.; Santos, A. A. S. M. D.; Fontes, C. A. P.; Muchaluat-Saade, D. C. e Conci, A. (2015b). Uma análise hibrida para identificação de cancer de mama usando sinais térmicos. *WIM-XV Workshop de Informática Médica-Anais CSBC*.

[Silva et al. 2014a] Silva, L. F.; Olivera, G. O. S.; Galvao, S.; Silva, J. B.; Santos, A. A. S. M. D.; Muchaluat-Saade, D. C. e Conci, A. (2014a). Análise de séries temporais de sinais térmicos da mama para detecção de anomalias. *WIM-XIV Workshop de Informática Médica-Anais CSBC*, pages 1818–1827.

[Silva et al. 2014b] Silva, L. F.; Saade, D. C. M.; Sequeiros, G. O.; Silva, A. C.; Paiva, A. C.; Bravo, R. S. e Conci, A. (2014b). A new database for breast research with infrared image. *Journal of Medical Imaging and Health Informatics*, 4(1):92–100.

[Silva e Hortale 2012] Silva, R. C. F. e Hortale, V. A. (2012). Rastreamento do câncer de mama no brasil: quem, como e por quê? *Revista Brasileira de Cancerologia*, 58:67–71.

[Silva 2010] Silva, S. V. (2010). *Reconstrução da Geometria da Mama a partir de Imagens Termográficas*. Tese de Doutorado, Universidade Federal Fluminense, Instituto de Computação, Niterói, RJ, Brasil.

[Stein et al. 2009] Stein, A. T.; de Medeiros Zelmanowicz, A.; Zerwes, F. P.; Biazus, J. V. N.; Lázaro, L. e Franco, L. R. (2009). Rastreamento do câncer de mama: recomendações baseadas em evidências. *Associação Médica do Rio Grande do Sul*, 53(4):438–446.

[Strehl e Ghosh 2000] Strehl, A. e Ghosh, J. (2000). Value-based customer grouping from large retail datasets. *Proc. SPIE Conf. Data Mining Knowl. Discov.*, pages 33–42.

[Thornton et al. 2013] Thornton, C.; Hutter, F.; Hoos, H. H. e Leyton-Brown, K. (2013). Auto-WEKA: Combined selection and hyperparameter optimization of classification algorithms. pages 847–855.

[Thuler 2003] Thuler, L. C. (2003). Considerações sobre a prevenção do câncer de mama feminino. *Revista Brasileira de Cancerologia*, 149(4):227–238.

[Torres et al. 2011] Torres, R.; Oliveira, G.; Xexéo, J.; Souza, W. e Linden, R. (2011). Identification of keys and cryptographic algorithms using genetic algorithm and graph theory. *Latin America Transactions, IEEE (Revista IEEE America Latina)*, 9(2):178–183.

[Usuki et al. 1990] Usuki, H.; Onoda, Y.; Kawasaki, S.; Misumi, T.; Murakami, M.; Komatsubara, S. e Teramoto, S. (1990). Relationship between thermographic observations of breast tumors and the dna indices obtained by flow cytometry. *Biomedical Thermology*, 10(4):282–285.

[Wishart et al. 2010] Wishart, G. C.; Campisid, M.; Boswella, M.; Chapmana, D.; Shackletona, V.; Iddlesa, S.; Halletta, A. e Britton, P. D. (2010). The accuracy of digital infrared imaging for breast cancer detection in women undergoing breast biopsy. *European Journal of Surgical Oncology*, 36(6):535–540.

[Witten et al. 2011] Witten, I. H.; Eibe, F. e Hall, M. A. (2011). *Data mining: practical machine learning tools and techniques*. Morgan Kaufmann Publishers, Burlington, MA 01803, USA, 3rd ed edition.

[Zitová e Flusser 2003] Zitová, B. e Flusser, J. (2003). Image registration methods: a survey. *Image and Vision Computing*, 21(11):977–1000.

APÊNDICE A – DIAGNÓSTICOS DAS PACIENTES

As tabelas seguintes contêm o ID (número de identificação da paciente no banco DMR-IR) e os dados clínicos das pacientes das quais os termogramas foram utilizados para a avaliação da metodologia, descrita no Capítulo 5. Na Tabela A.1 estão as pacientes consideradas com alguma anomalia de mama. A descrição da anomalia foi transcrita dos registros no prontuário arquivado no HUAP, de cada paciente. Na Tabela A.2 estão as paciente que realizaram o exame mamográfico de triagem e receberam a seguinte recomendação: mamografia em 2 anos para ambas as mamas.

Tabela A.1: Pacientes com anomalias

Banco DMI ID	Descrição da anomalia
138	Fibroadenoma em quadrante inferior externo de mama D, com 0,5x 0,6cm, Birads 4
179	Carcinoma ductal infiltrante grau 3 com focos multifocais, representando comprometimento de margem cirúrgica
180	Carcinoma ductal infiltrante com 7cm em união dos quadrantes superiores de mama E
181	Mama direita: Carcinoma Ductal Infiltrante grau 3 com 2cm em União dos quadrantes superiores. Mama esquerda: Carcinoma Ductal Infiltrante grau 3 com 2cm em união dos quadrantes externos e nódulo de 1,5 cm em quadrante superior externo.
192	Carcinoma ductal infiltrante localmente avançado com cerca de 5-10cm em mama E, inicialmente em união de quadrantes inferiores, porém atualmente lesão ulcerada ao redor do complexo aréolo-mamilar(CAM).
198	Carcinoma residual, correspondendo a 60% de celularidade

	em leito tumoral que 2,5x2,0 cm, sendo 2% representado por carcinoma intra-ductal angio-linfática.
202	Carcinoma ductal in situ, localmente avançado, com tamanho inicial de 10 cm em quadrante superior externo de mama direita, submetida a quimioterapia e radioterapia neoadjuvantes até maio/2013. Atualmente com espessamento de 3,5 cm em quadrante superior externo.
204	Descarga papilar de liquido seroso em mama direita Resultado da peça cirúrgica após retirada da mama: Carcinoma ductal in situ de baixo grau
209	Carcinoma ductal infiltrante grau 2, nódulo com cerca de 4cm em união dos quadrantes internos de mama E.
210	Mama E: Carcinoma ductal in situ de alto grau , união dos quadrantes superiores com 1,1- 2 cm
213	Carcinoma papilífero em mama D, quadrante superior externo, nódulo com 1,5cm. Axila esquerda e direita com linfonodos suspeitos
240	Carcinoma ductal infiltrante infiltrando a derme, com graduação histológica Nottingam II, (grau arquitetural 2, nuclear 2 índice mitótico 3)
241	Carcinoma ductal infiltrante, Nottingham grau II, (grau histopatológico=3, grau nuclear=2, índice mitótico=1)
245	Carcinoma ductal infiltrante, grau histológico I de Nottinghan (formação tubular=2; grau nuclear=2; contagem mitótica=1)
249	Tumor filoide de mama, de baixo potencial de malignidade
250	Carcinoma lobular in situ
251	Carcinoma infiltrante, de tipo especial, grau histopatológico final de Nottingham I (formação tubular 1; pleomorfismo nuclear 1; índice mitótico 1)
255	Carcinoma ductal infiltrante , graus histologico de Nottinghan final II; arranjo tubular =3; grau nuclear =2, índice mitótico=1
256	Carcinoma de padrão mucinoso, infiltrando derme.
257	Carcinoma ductal infiltrante , graus histológico I de

	Nottingham (formação tubular =2; grau nuclear=2; contagem mitótica=1), comprometendo toda a derme e hipoderme
259	Carcinoma ductal infiltrante
261	Carcinoma invasivo de tipo não especial. Grau 3 de Nottingham (arranjo tubular=3, pleomorfismo nuclear =3; índice mitótico = 2)
263	Esteatonecrose com área de fibrose hialina e calcificação
264	Tumor filodes maligno com componente estromal maligno heterólogo tipo Lipossarcoma
270	Carcinoma infiltrante do tipo não especial, grau histológico final de Nottingham II, Presença de carcinoma ductal in situ de baixo grau, do tip sólido, correspondendo a cerca de 5% do total da neoplasia.
219	MAMA DIREITA-Categoria 0 (nódulo) (BI-RADS), MAMA ESQUERDA-Categoria 2 (BI-RADS).
222	MAMA DIREITA-Categoria 0 (BI-RADS) (nódulos), MAMA ESQUERDA-Categoria 2 (BI-RADS).
217	MAMA DIREITA-Categoria 0 (nódulo) (BI-RADS), BI-RADS da ultrassonografia - Categoria 1, MAMA ESQUERDA-Categoria 2 (BI-RADS).
291	Fibroadenoma QII - ME.
292	Carcinoma Ductal Infiltrante QQSS - MD.
282	Citologia - lesão líquida (Benigno) QII - ME.
280	Carcinoma Ductal Infiltrante QQSS-ME.
284	Hiperplasia ductal florida/alteração de célula colunar QSL-MD.
281	Tumor filodes QSL-ME
286	Carcinoma Ductal Infiltrante QSL-ME

Tabela A.2: Pacientes sem anomalias segundo o exame mamográfico

Banco DMI ID	Resultado da mamografia (recomendação: fazer mamografia de dois em dois anos)
9	MAMA DIREITA-Categoria 1 (BI-RADS), MAMA ESQUERDA-Categoria 1 (BI-RADS).

30	MAMA DIREITA-Categoria 1 (BI-RADS), MAMA ESQUERDA-Categoria 2 (BI-RADS).
34	MAMA DIREITA-Categoria 2 (BI-RADS), MAMA ESQUERDA-Categoria 1 (BI-RADS).
37	MAMA DIREITA-Categoria 2 (BI-RADS), MAMA ESQUERDA-Categoria 2 (BI-RADS).
40	MAMA DIREITA-Categoria 1 (BI-RADS), MAMA ESQUERDA-Categoria 2 (BI-RADS).
42	MAMA DIREITA-Categoria 1 (BI-RADS), MAMA ESQUERDA-Categoria 1 (BI-RADS).
44	MAMA DIREITA-Categoria 1 (BI-RADS), MAMA ESQUERDA-Categoria 1 (BI-RADS).
49	MAMA DIREITA-Categoria 2 (BI-RADS), MAMA ESQUERDA-Categoria 1 (BI-RADS).
50	MAMA DIREITA-Categoria 1 (BI-RADS), MAMA ESQUERDA-Categoria 1 (BI-RADS).
51	MAMA DIREITA-Categoria 2 (BI-RADS), MAMA ESQUERDA-Categoria 1 (BI-RADS).
55	MAMA DIREITA-Categoria 1 (BI-RADS), MAMA ESQUERDA-Categoria 1 (BI-RADS).
64	MAMA DIREITA-Categoria 1 (BI-RADS), MAMA ESQUERDA-Categoria 1 (BI-RADS).
66	MAMA DIREITA-Categoria 2 (BI-RADS), MAMA ESQUERDA-Categoria 2 (BI-RADS).
68	MAMA DIREITA-Categoria 2 (BI-RADS), MAMA ESQUERDA-Categoria 2 (BI-RADS).
69	MAMA DIREITA-Categoria 2 (BI-RADS), MAMA ESQUERDA-Categoria 2 (BI-RADS).
72	MAMA DIREITA-Categoria 2 (BI-RADS), MAMA ESQUERDA-Categoria 2 (BI-RADS).
80	MAMA DIREITA-Categoria 1 (BI-RADS), MAMA ESQUERDA-Categoria 1 (BI-RADS).
85	MAMA DIREITA-Categoria 2 (BI-RADS), MAMA ESQUERDA-Categoria 2 (BI-RADS).

87	MAMA DIREITA-Categoria 1 (BI-RADS), MAMA ESQUERDA-Categoria 1 (BI-RADS).
108	MAMA DIREITA-Categoria 1 (BI-RADS), MAMA ESQUERDA-Categoria 2 (BI-RADS).
115	MAMA DIREITA-Categoria 2 (BI-RADS), MAMA ESQUERDA-Categoria 1 (BI-RADS).
126	MAMA DIREITA-Categoria 2 (BI-RADS), MAMA ESQUERDA-Categoria 2 (BI-RADS).
129	MAMA DIREITA-Categoria 2 (BI-RADS), MAMA ESQUERDA-Categoria 1 (BI-RADS).
132	MAMA DIREITA-Categoria 1 (BI-RADS), MAMA ESQUERDA-Categoria 1 (BI-RADS).
135	MAMA DIREITA-Categoria 1 (BI-RADS), MAMA ESQUERDA-Categoria 1 (BI-RADS).
137	MAMA DIREITA-Categoria 2 (BI-RADS), MAMA ESQUERDA-Categoria 2 (BI-RADS).
143	MAMA DIREITA-Categoria 1 (BI-RADS), MAMA ESQUERDA-Categoria 2 (BI-RADS).
145	MAMA DIREITA-Categoria 2 (BI-RADS), MAMA ESQUERDA-Categoria 1 (BI-RADS).
166	MAMA DIREITA-Categoria 1 (BI-RADS), MAMA ESQUERDA-Categoria 1 (BI-RADS).
171	MAMA DIREITA-Categoria 2 (BI-RADS), MAMA ESQUERDA-Categoria 2 (BI-RADS).
172	MAMA DIREITA-Categoria 2 (BI-RADS), MAMA ESQUERDA-Categoria 2 (BI-RADS).
220	MAMA DIREITA-Categoria 1 (BI-RADS), MAMA ESQUERDA-Categoria 1 (BI-RADS).
226	MAMA DIREITA-Categoria 2 (BI-RADS), MAMA ESQUERDA-Categoria 2 (BI-RADS).
254	MAMA DIREITA-Categoria 2 (BI-RADS), MAMA ESQUERDA-Categoria 2 (BI-RADS).
272	MAMA DIREITA-Categoria 1 (BI-RADS), MAMA ESQUERDA-Categoria 1 (BI-RADS).