RENAN CALDEIRA MENECELLI

**Caracterização de sinais secundários em imagens
mamográficas por redes neurais artificiais para
auxílio ao diagnóstico do câncer de mama**

Dissertação apresentada à Escola de Engenharia de São Carlos da Universidade de São Paulo como parte dos requisitos para obtenção do título de Mestre em Ciências, Programa de Engenharia Elétrica.

Área de concentração: Processamento de Sinais e Instrumentação

Orientador: Prof. Dr. Homero Schiabel

**São Carlos
2013**

Trata-se da versão corrigida da dissertação. A versão original se encontra disponível na EESC/USP que aloja o Programa de Pós-Graduação de Engenharia Elétrica

AUTORIZO A REPRODUÇÃO TOTAL OU PARCIAL DESTE TRABALHO,
POR QUALQUER MEIO CONVENCIONAL OU ELETRÔNICO, PARA FINS
DE ESTUDO E PESQUISA, DESDE QUE CITADA A FONTE.

Mc Menechelli, Renan Caldeira
Caracterização de sinais secundários em imagens
mamográficas por redes neurais artificiais para auxílio
ao diagnóstico do câncer de mama / Renan Caldeira
Menechelli; orientador Homero Schiabel. São Carlos,
2013.

Dissertação (Mestrado) - Programa de Pós-Graduação
em Engenharia Elétrica e Área de Concentração em
Processamento de Sinais e Instrumentação -- Escola de
Engenharia de São Carlos da Universidade de São Paulo,
2013.

1. câncer de mama. 2. achados secundários. 3.
assimetria tecidual. 4. processamento de imagens. 5.
reconhecimento de padrões. I. Título.

FOLHA DE JULGAMENTO

Candidato: Bacharel **RENAN CALDEIRA MENECELLI**.

Título da dissertação: "Caracterização de sinais secundários em imagens mamográficas por redes neurais artificiais para auxílio ao diagnóstico do câncer de mama".

Data da defesa: 25/02/2013

Comissão Julgadora:

Resultado:

Prof. Associado **Homero Schiabel (Orientador)**
(Escola de Engenharia de São Carlos/EESC)

APROVADO

Dr^a. **Simone Elias Martinelli**
(Universidade Federal de São Paulo/UNIFESP)

aprovado

Prof^a. Titular **Agma Juci Machado Traina**
(Instituto de Ciências Matemáticas e de Computação/USP)

APROVADO

Coordenador do Programa de Pós-Graduação em Engenharia Elétrica e Presidente da
Comissão de Pós-Graduação:

Prof. Titular **Denis Vinicius Coury**

Este trabalho teve o apoio financeiro da FAPESP –
Fundação de Amparo à Pesquisa do Estado de São Paulo
Processo nº 2010/13909-9

***D*edico ...**

Aos meus pais, Fram e Jú, e avó Lú que
sempre me dão apoio e incentivo às
escolhas que faço.

AGRADECIMENTOS

À Deus que diariamente guia meu caminho, protegendo-me constantemente e à N. Sr^a de Fátima a qual deposito toda minha confiança.

Ao Prof. Dr. Homero Schiabel por ter investido no meu potencial acadêmico, firmado apoio constante e valorizado meu trabalho ao máximo.

À Prof^a Dr^a Patricia Bellin Ribeiro por ter acreditado na minha capacidade, pelo apoio, empenho, dedicação e toda a ajuda fornecida na realização desta pesquisa e de muitos outros trabalhos que fizemos juntos, além de todo o incentivo para que nunca afastasse-me do caminho acadêmico. Alguém que sempre estará comigo em pensamentos e recordações por toda a vida.

Aos meus pais Jú e Fram pelo incentivo, carinho e amor que expressam por mim diariamente além da incrível confiança em mim depositada.

À avó Lú que tanto se preocupa com a minha felicidade e saúde tornando-me uma pessoa melhor todos os dias.

À toda a minha família pela preocupação com o meu bem-estar, com minha profissão, com meu presente e futuro, expressando continuamente o grande amor por mim: Vó Victória, Kamy, Lalá, Tia Fá, Chico, Di, Tia Con, Tia Tê.

À amiga Karem Daiane Marcomini pela dedicação em estar comigo na caminhada desde a graduação (4 anos) até nos anos de mestrado em São Carlos, além de me aturar, compreender, ajudar, criticar e corrigir, compartilhando a maioria dos momentos juntos.

À Maria Angélica pela edificação de uma grande amizade que só foi possível com o respeito das diferentes opiniões. Pelas pizzas, pipoca doce, Uno, beatniks, Vila Brasil, ... além do socorro com as mais complexas fórmulas, sugestões, críticas e desenvolvimento pessoal a mim proporcionado.

Ao Caio César Benatti pelas baladas, reflexões, distrações, apoio, conselhos. Sem dúvida, sem tudo isso o meu crescimento pessoal não estaria concretizado. De vizinhos de bloco à uma grande amizade verdadeira.

Às colegas do Lapimo: Paula, Renata, Débora e Luciana. Minha gratidão por poder compartilhar momentos e ensinamentos durante esse período.

Aos amigos Bruno Barufaldi e Roger pela convivência durante a pesquisa no Lapimo, cursando disciplinas em comum e ajudando-me no compartilhamento do conhecimento.

Aos grandes e inesquecíveis professores da graduação na USC: Kelton, Elvio, Henrique, André, Cláudio, Luciene, Rosane M. Lima Araújo e Fátima R. Lima Ribeiro. Todos, fontes de inspiração para o meu crescimento profissional.

Aos professores do Departamento de Engenharia Elétrica da USP, Adilson Gonzaga e Marcelo A. C. Vieira (Mogi) por todo o conhecimento e auxílio fornecidos para o desenvolvimento da pesquisa.

Ao Dr. Ladislau Martin Neto por todo incentivo, credibilidade, entusiasmo e preocupação para a produção de resultados e conclusão da pesquisa.

À AIESEC São Carlos, da qual tive orgulho de fazer parte como um agente de mudança em constante desenvolvimento. Muitas coisas tornaram-se mais claras na minha vida: desde o alinhamento do propósito da organização com essa pesquisa até o apoio de cada membro que por ali passou e esteve junto comigo na caminhada final desse trabalho. À Ana Carolina Ussier, Homer, Camila Billerbeck, Fará, Pri, Dani, Tarcila, Isa, Stéfano, Cintia, Léo, Galã, Piauí, Chester, Homer, Aliene e, principalmente, à Naná que acreditou e investiu demais em mim. Sem vocês eu não teria “destruído”!

Ao grande amigo Diogo Andrade que acreditou no meu conhecimento e capacidade profissional, sempre disposto em ajudar, apoiar e criticar minhas decisões.

Aos colegas da pós: Lú, Tati, Camila, Thaís, Marcelo, Fernanda, Raissa, Henrique, Rafael (Prosinha), Amilcar, Samuel, Fabão, Breno, Rodolfo, Wellington, Ricardo (Vences), Lucas, André.

À Santa Casa de São Carlos, em especial ao Dr. Luís que tanto atrapelei e ao Hospital das Clínicas de Ribeirão Preto pela colaboração e auxílio no fornecimento das imagens mamográficas juntos dos respectivos laudos.

À FAPESP pelo apoio financeiro desta pesquisa e pela confiança e credibilidade depositadas em mim.

RESUMO

MENECHELLI, Renan Caldeira. *Caracterização de sinais secundários em imagens mamográficas por redes neurais artificiais para auxílio ao diagnóstico do câncer de mama*. 2013. Dissertação (mestrado) – Escola de Engenharia de São Carlos, Universidade de São Paulo, São Carlos, 2013.

O constante aumento do número de novos casos do câncer de mama vem despertando interesse na elaboração de módulos de esquemas CAD a fim de proporcionar um diagnóstico de maior precisão. Entretanto, a maioria das pesquisas está empenhada em detectar ou classificar fatores primários presentes em imagens mamográficas, como nódulos e microcalcificações. Áreas assimétricas, retração de mamilo, linfonodos axilares, entre outros, são considerados como fatores secundários no diagnóstico do câncer de mama, apesar de poderem alertar para o surgimento não só dessa, mas de outras doenças no futuro. Por isso, essa pesquisa contempla a implementação de um sistema computacional capaz de auxiliar na detecção e classificação, conforme padrão BI-RADS[®], de regiões que contenham sinais secundários capazes de levantar suspeitas da presença ou surgimento do câncer de mama, em imagens mamográficas digitais, utilizando técnicas inteligentes e automáticas de processamento de imagens e redes neurais artificiais. A acurácia alcançada em cada etapa foi: detecção de assimetria de 82,8%, retração de mamilo de 95% e $Az = 0,93$, detecção de linfonodos axilares = 74,9%. Objetiva-se que o resultado do trabalho seja inserido como um dos módulos de um protótipo de esquema CADx em mamografia, a fim de ampliar o conjunto de informações a serem usadas na classificação de cada caso sob análise, visando o aumento da precisão diagnóstica.

Palavras-chave: câncer de mama, achados secundários, assimetria tecidual, processamento de imagens, reconhecimento de padrões.

ABSTRACT

MENECELLI, Renan Caldeira. *Characterization of secondary signals in mammographic images by artificial neural networks to aid diagnosis of breast cancer*. 2013. Dissertação (mestrado) – Escola de Engenharia de São Carlos, Universidade de São Paulo, São Carlos, 2013.

The increase in the number of cases of breast cancer have attracted interest in developing modules of CAD schemes to provider higher diagnostic accuracy. However, most researches are engaged in detect and classify primary factors present in mammographic images such as nodules and microcalcifications. Asymmetric areas, nipple retraction, axillary lymph nodes, among other, are considered as secondary factors to diagnostic the breast cancer, although they may alert for the emergence not only of this but of other diseases in the future. Thus, this research includes the implementation of a computer system able to assist in the detection and classification, according to BI-RADS[®], of regions that containing secondary signals able to arousing suspicion of the presence or appearance of breast cancer in digital mammographic images using intelligent and automatic techniques in the image processing and artificial neural networks. The accuracy obtained in each step was: detection of asymmetry of 82.8%, nipple retraction of 95% and Az = 0.93, detection of axillary lymph nodes = 74.9%. The purpose is that the result of the work is entered as one of the modules of a prototype of CADx schem in mammography in order to extend the range of information to be used in the classification of each case under analysis, aiming to increase diagnostic accuracy.

Key Words: secondary findings, nipple retraction, mammography asymmetry, image processing, pattern recognition.

LISTA DE FIGURAS

Figura 1.1 – Fluxograma genérico de um sistema CADe e CADx (NISHIKAWA, 2010)	29
Figura 1.2 – Tipos de cânceres mais incidentes para ano de 2012, exceto pele não melanoma, baseado nos dados estimativos do Instituto Nacional de Câncer / Ministério da Saúde (INCA/MS)	30
Figura 2.1 – Componentes principais para formação de imagens mamográficas	36
Figura 2.2 – Posíveis direções de captura da mama para produzir imagens mamográficas com maiores detalhes.....	38
Figura 2.3 – Categorias de classificação de densidade mamográfica, conforme BI-RADS.	40
Figura 2.4 - Exemplificação dos tipos de assimetrias padronizados pelo BI-RADS.	42
Figura 2.5 – Exames mamográficos com detecção de assimetrias, conforme BI-RADS®	43
Figura 2.6 – Retração/inversão do mamilo evidenciada em imagens mamográficas de diferentes projeções.....	44
Figura 2.7 - Imagens de projeção MLO com marcação de linfonodos presentes na área correspondente ao músculo peitoral	45
Figura 2.8 - Descrição geral das etapas do algoritmo que compõe o sistema implementado por Tzikopoulos et al. (2011)	49
Figura 2.9 - Descrição geral das etapas desenvolvidas no trabalho de Wu, Besnehard e Marchessoux (2011)	51
Figura 3.1 – Representação gráfica da função sigmóide.....	59
Figura 3.2 – Diagrama de blocos das fases de treinamento e operação da rede neural MLP	60
Figura 4.1 – Principais etapas do sistema inteligente desenvolvido	64
Figura 4.2 – Tela resultante da busca feita na base <i>online</i> de imagens mamográficas do LAPIMO	65

Figura 4.3 – Parcela do conjunto de imagens simuladas do mamilo normal e retraído	66
Figura 4.4 - Etapas para segmentação da região da mama	67
Figura 4.5 - Resultado da segmentação da mama (eliminação de etiquetas e artefatos).....	68
Figura 4.6 – Procedimentos executados para delimitação do músculo peitoral em imagens de projeção MLO.....	69
Figura 4.7 – Obtendo a ROI inicial para delimitação do músculo peitoral.....	70
Figura 4.8 - Finalização da segmentação do músculo peitoral.	70
Figura 4.9 – Exemplificação da área de busca da região do mamilo.	71
Figura 4.10 – Linha com maior número de <i>pixels</i> brancos, dentro da região de busca	72
Figura 4.11 – Obtenção da ROI contendo a região do mamilo. a) ROI traçada a partir da linha com maior número de pixels; b) ampliação da ROI obtida em a).	72
Figura 4.12 – Seleção da ROI da região do mamilo em uma imagem mamográfica com o mamilo retraído. a) imagem original; b) imagem segmentada com a marcação da ROI; c) ROI ampliada.....	73
Figura 4.13 – Fluxo operacional para classificação de regiões da mama.....	75
Figura 4.14 – Correspondência entre uma ROI na imagem binária e imagem mamográfica	76
Figura 4.15 – Interface desenvolvida para classificação manual das ROIs pelo radiologista.....	77
Figura 4.16 – Marcação da ROI a ser classificada na imagem mamográfica completa	77
Figura 4.17 – Conjunto de imagens selecionadas pelo radiologista para cada uma das classes de densidade, conforme BI-RADS®	78
Figura 4.18 – Esquematização das etapas para o desenvolvimento do classificador final	81
Figura 4.19 - Diagrama do fluxo operacional do software classificador de densidade automático	82
Figura 4.20 - Interface do software final de classificação automática de densidade mamográfica.	83
Figura 4.21 – Divisão da mama em (a) quadrantes e (b) ROIs para busca de assimetrias.....	84

Figura 4.22 – Correspondência entre ROIs bilaterais para a detecção de assimetrias	85
Figura 4.23 – Curva ROC para o limiar de 35% de diferença para considerar uma região assimétrica	88
Figura 4.24 – Índice de performance (VP, FP, FN) obtido pela classificação.....	93
Figura 4.25 – Fluxo operacional para o módulo do sistema CAD proposto.....	95
Figura 5.1 – Imagens resultantes das etapas para segmentação da região da mama.	100
Figura 5.2 - Exemplificação das imagens utilizadas para avaliação pelas métricas:	101
Figura 5.3 – Sobreposição dos contornos obtidos na imagem original.....	102
Figura 5.4 – Diferentes técnicas implementadas para delimitação do músculo peitoral em imagens mamográficas MLO.	104
Figura 5.5 – Processo final na delimitação do músculo peitoral:	106
Figura 5.6 - Imagens utilizadas para obtenção das métricas de avaliação.....	106
Figura 5.7 – Imagem original com sobreposição dos contornos de delimitação do músculo peitoral:	107
Figura 5.8 – Gráficos da distribuição gaussiana para os atributos (a) compacidade e (b) relação entre área e perímetro	108
Figura 5.9 – Curva ROC obtida a partir da melhor classificação da MLP para verificar a capacidade de generalização do classificador.....	109
Figura 5.10 - Curva ROC obtida com a classificação das imagens mamográficas para diferenciar mamilo normal e mamilo retraído.....	109
Figura 5.11 - Curva ROC resultante da classificação das imagens mamográficas para testar a generalização do classificador	110
Figura 5.12 – Demonstração dos resultados para detecção de linfonodos no músculo peitoral de imagens MLO	112
Figura 5.13 - Gráficos da distribuição gaussiana para os atributos (a) autocorrelação e (b) variância.....	113
Figura 5.14 – Resultados obtidos com a classificação de regiões da mama pelo trabalho de Wu, Besnehard e Marchessoux (2011).....	114
Figura 5.15 - Curva ROC para a classificação de densidade em 2 classes: baixa densidade e alta densidade.	117
Figura 5.16 – Detecção de assimetria em um par de imagens mamográficas.	119

Figura 5.17 – Tela inicial do módulo CAD desenvolvido	120
Figura 5.18 – Imagens de um exame mamográfico em processamento no módulo CAD desenvolvido	121
Figura 5.19 – Resultado da detecção de assimetrias efetuada pelo módulo.	122

LISTA DE TABELAS

Tabela 4.1 – Comparação entre os atributos descritores de textura para detecção de assimetria em imagens bilaterais de mesma projeção	86
Tabela 4.2 – Verificação do limiar percentual de detecção de diferença assimétrica	87
Tabela 4.3 – Matriz de confusão	91
Tabela 5.1 – Resultados das métricas de avaliação da segmentação da região da mama	100
Tabela 5.2 – Resultado das métricas de avaliação da delimitação do músculo peitoral.....	107
Tabela 5.3 – Resultado obtido com a detecção de linfonodos.....	111
Tabela 5.4 – Resultados obtidos com a classificação das ROIs da mama.....	113
Tabela 5.5 - Resultados obtidos com a classificação das densidades em 4 classes, conforme BI-RADS®	116
Tabela 5.6 – Revisão bibliográfica dos principais métodos de classificação de densidade mamográfica	117
Tabela 5.7 – Tempo médio de processamento do módulo CADx em diferentes computadores.....	119

LISTA DE SIGLAS E ABREVIATURAS

Az	Área sob a curva ROC
API	Interface de programação de aplicações, do inglês: <i>Application Programming Interface</i>
BI-RADS	<i>Breast Image Reporting and Data System</i>
CC	Craniocaudal
CAD	Diagnóstico auxiliado por computador, do inglês: <i>Computer-Aided Diagnosis</i>
CADe	Deteção auxiliada por computador, do inglês: <i>Computer-Aided Detector</i>
CADx	Diagnóstico auxiliado por computador, do inglês: <i>Computer-Aided Diagnosis</i>
CSV	<i>Comma-separated values</i>
EESC	Escola de Engenharia de São Carlos
FP	Falso-positivo
FN	Falso-negativo
IDE	Ambiente de desenvolvimento integrado, do inglês: <i>Integrated Development Environment</i>
INCA	Instituto Nacional de Câncer
LAPIMO	Laboratório de Análise e Processamento de Imagens Médicas e Odontológicas
MLO	Mediolateral Oblíqua
MLP	Perceptron multi-camadas, do inglês: <i>Multi-Layer Perceptron</i>
OMS	Organização Mundial de Saúde, do inglês: <i>World Health Organization</i>
RNA	Redes Neurais Artificiais
ROC	<i>Receiver Operating Characteristic</i>
ROI	Região de interesse, do inglês: <i>Region of Interest</i>
TIFF	<i>Tag Image File Format</i>

USP	Universidade de São Paulo
VP	Verdadeiro-positivo
VN	Verdadeiro-negativo

SUMÁRIO

1 - INTRODUÇÃO	27
1.1 ESQUEMAS DE AUXÍLIO AO DIAGNÓSTICO POR COMPUTADOR	27
1.2 MOTIVAÇÃO.....	30
1.3 OBJETIVOS.....	32
1.3.1 Objetivos específicos.....	32
1.4 DISPOSIÇÃO DO TRABALHO	33
 2 - A IMAGEM MAMOGRÁFICA.....	35
2.1 POSICIONAMENTO PARA CAPTURA MAMOGRÁFICA	37
2.2 O BI-RADS®	38
2.2.1 Densidade mamográfica.....	39
2.2.2 Variação de densidade.....	40
2.3 ASPECTOS SECUNDÁRIOS COMO POSSÍVEIS FATORES DE RISCOS	41
2.3.1 O tecido mamário assimétrico	41
2.3.2 Retração do Mamilo	44
2.3.3 Linfonodos Axilares	45
2.4 PRINCIPAIS TRABALHOS DESENVOLVIDOS NA INVESTIGAÇÃO COMPUTACIONAL DE ACHADOS SECUNDÁRIOS	46
 3 - VISÃO COMPUTACIONAL.....	53
3.1 FUNDAMENTOS DE SISTEMAS VISUAIS ARTIFICIAIS.....	53
3.2 ALARGAMENTO NÃO-LINEAR DE CONTRASTE	54
3.3 SEGMENTAÇÃO.....	54
3.3.1 Limiar de Otsu	55
3.3.2 Entropia de Pun	55

3.4	EXTRAÇÃO DE COMPONENTE CONEXOS.....	56
3.5	DESCRITORES DE PADRÕES	57
3.6	RECONHECIMENTO DE PADRÕES	58
3.6.1	Rede neural Multilayer Perceptron (MLP).....	59
3.6.2	Validação	61
4	- MATERIAIS E MÉTODOS	63
4.1	BASE DE IMAGENS	64
4.1.1	Imagens mamográficas reais	65
4.1.2	Imagens que simulam o mamilo	66
4.2	PROCESSAMENTO DE IMAGENS	67
4.2.1	Segmentação da região da mama e extração do fundo	67
4.2.2	Delineamento do músculo peitoral.....	69
4.2.3	Localização automática da região do mamilo em imagens mamográficas digitalizadas	71
4.2.4	Detecção de Linfonodos	73
4.2.5	Subdivisão automática da região da mama	74
4.2.6	Classificação manual das ROIs por um especialista	76
4.3	SELEÇÃO DE ATRIBUTOS PARA CLASSIFICAÇÃO	79
4.4	CLASSIFICAÇÃO POR REDES NEURAS ARTIFICIAIS	79
4.4.1	Classificação do mamilo	80
4.4.2	Classificação da densidade mamográfica, conforme BI-RADS	82
4.5	DETECÇÃO E CLASSIFICAÇÃO DE ASSIMETRIA TECIDUAL	83
4.5.1	O processo de detecção	83
4.5.2	Classificação de assimetrias.....	88
4.6	MÉTRICAS DE AVALIAÇÃO	89
4.7	SINTESE: MÓDULO DESENVOLVIDO DO ESQUEMA CAD	93
4.7.1	Fluxo operacional do módulo	94
5	- RESULTADOS E DISCUSSÕES.....	99
5.1	SEGMENTAÇÃO DA REGIÃO DA MAMA	99
5.2	DELIMITAÇÃO DO MÚSCULO PEITORAL	103

5.3	CLASSIFICAÇÃO DO MAMILO	108
5.4	DETECÇÃO DE LINFONODOS.....	111
5.5	CLASSIFICAÇÃO DE DENSIDADES EM REGIÕES DA MAMA	112
5.6	CLASSIFICAÇÃO DA DENSIDADE MAMOGRÁFICA, CONFORME BI-RADS®	115
5.6.1	Classificação conforme BI-RADS® em 4 categorias	115
5.6.2	Classificação em 2 classes: baixa densidade e alta densidade	116
5.7	DETECÇÃO DE ASSIMETRIAS	118
5.8	MÓDULO DESENVOLVIDO PARA O ESQUEMA CADx	120
6	- CONCLUSÕES	123
6.1	FUNDAMENTAÇÃO TEÓRICA	123
6.2	METODOLOGIA APLICADA.....	124
6.3	RESULTADOS ALCANÇADOS	125
6.4	CONCLUSÃO.....	127
6.5	PROPOSTA PARA TRABALHOS FUTUROS.....	128
	REFERÊNCIAS BIBLIOGRÁFICAS.....	129

1 - INTRODUÇÃO

É crescente o número de trabalhos desenvolvidos no intuito de colaborar com a avaliação e análise de variados tipos de estruturas capazes de serem visualizadas em mamografias digitalizadas. Ao final desses trabalhos são esperadas alternativas capazes de fornecerem auxílio a um diagnóstico de maior precisão para a ocorrência do câncer de mama (ANGELO, 2007; KOPANS, 2000; ZHENG, 2008). Objetivando o alcance de resultados satisfatórios, diversas propostas são apresentadas para a melhoria de subsídios como: disponibilidade de cursos para especialistas, avanços em tecnologias mamográficas com sistemas de arquivos digitais, desenvolvimento de bancos de dados de imagens mamográficas cuja digitalização ocorre através de sofisticados equipamentos para preservar a qualidade das mesmas, correções efetuadas sobre as imagens digitais para melhorar a visualização de estruturas internas de interesse, módulos de esquemas computacionais CAD (*Computer Aided Diagnosis*), entre outras.

1.1 Esquemas de auxílio ao diagnóstico por computador

O desenvolvimento de esquemas CAD que atuam sobre a imagem mamográfica têm por objetivo apoiar radiologistas na detecção da suspeita de

microcalcificações e massas nodulares presentes nas imagens (ALOLFE et al., 2008; CHENG et al., 2006; DOI, 2007; ZHENG, 2008). De fato, diferentes funções que automatizem computacionalmente o auxílio ao diagnóstico têm sido elaboradas e testadas para ajudar nas decisões em diferentes ramos da medicina desde a década de 50 (DOI, 2007). Atualmente, de acordo com suas funcionalidades específicas, os sistemas CAD são divididos em dois tipos: detecção auxiliada por computador (CAdE – *computer-aided detection*) e diagnóstico auxiliado por computador (CAdx – *computer-aided diagnosis*). Os CAdE objetivam identificar e marcar áreas suspeitas em uma imagem, sendo sua maior potencialidade evidenciar casos clinicamente ocultos de câncer. Já os CAdx têm por finalidade fornecer um relatório com maior detalhamento sobre uma lesão suspeita para ajudar o radiologista a decidir a necessidade do encaminhamento de uma paciente a biópsia, por exemplo (NISHIKAWA, 2010). A Figura 1.1 exibe um fluxograma genérico de esquemas CAdE e CAdx.

A principal diferença notada na Figura 1.1 está na saída de cada um dos sistemas. No CAdE a saída delimita uma possível localização de um achado suspeito, enquanto no CAdx fornece um parecer estatístico sobre a indicação de um achado poder ser benigno ou maligno. A complexidade desses sistemas CAD em implementarem todas as funcionalidades para um diagnóstico preciso de qualquer tipo de imagem torna a sua comercialização inviável, já que fornecem soluções apenas para casos muito específicos e com muitas limitações (DOI, 2007; RANGAYYAN; AYRES; DESAULTELS, 2007).

Em geral, os esquemas CAD fornecem opiniões extraídas de imagens médicas através de técnicas de processamento de imagens, visão computacional, inteligência artificial, reconhecimento de padrões, entre outras, a fim de prover melhorias e extrair informações úteis no suporte ao diagnóstico (BRANCATO et al., 2008; NUNES, 2006; PATROCINIO; SCHIABEL; ROMERO, 2004). Acredita-se que esquemas CAD sejam ferramentas promissoras no combate à taxa de mortalidade feminina pelo câncer de mama através de diagnósticos mais precisos com caráter preventivo (DOI, 2007; RANGAYYAN; AYRES; DESAULTELS, 2007).

Figura 1.1 – Fluxograma genérico de um sistema CAdE e CAdx (NISHIKAWA, 2010)

Visto que são muitos os fatores que podem contribuir para a ocorrência de dificuldades ou falhas na interpretação da imagem mamográfica, o grupo do LAPIMO (Laboratório de Análise e Processamento de Imagens Médicas e Odontológicas) vem desenvolvendo ao longo dos anos pesquisas com intuito de avaliar a qualidade dessas imagens e aplicar técnicas de processamento para realce, detecção e classificação de estruturas como nódulos e microcalcificações (SCHIABEL et al., 2012; SCHIABEL et al., 2011).

1.2 Motivação

Periodicamente, o INCA¹ (2011) disponibiliza estimativas na manifestação e surgimento de variados tipos de cânceres incidentes sobre a população brasileira. Para o ano de 2012, no Brasil, a estimativa é de que 52.680 novos casos sejam diagnosticados, como visto pela Figura 1.2, com um risco estimado de 52 casos a cada 100 mil mulheres. Segundo a Organização Mundial de Saúde (OMS) o câncer de mama é o segundo tipo de maior frequência no mundo dentre os tipos da mesma doença. Anualmente, cerca de 22% dos novos casos de câncer em mulheres são de mama (WHO, 2006).

Figura 1.2 – Tipos de cânceres mais incidentes para ano de 2012, exceto pele não melanoma, baseado nos dados estimativos do Instituto Nacional de Câncer / Ministério da Saúde (INCA/MS)

Diferentemente do que acontece em países desenvolvidos, a realização da mamografia no Brasil, considerando toda a sua extensão territorial, é deficiente e não atinge, principalmente, a população mais pobre. Isso contribui para que doenças relacionadas à mama sejam diagnosticadas em estágios muito avançados (EISENBERG; KOIFMAN; CORDEIRO DE REZENDE, 2001).

Normalmente, a procura tardia por exames e métodos preventivos é o fator mais agravante da doença. Quando diagnosticados em fase inicial, cerca de 50%

¹ Fonte: Instituto Nacional de Câncer (INCA). Disponível em: <<http://www.inca.gov.br>>. Acessado em: 25/11/2011

dos casos possuem chances de serem curados com uma redução de aproximadamente 30% do risco de morte causado por essa doença (ALOLFE et al., 2009; WHO, 2006). Entretanto, estudos revelam que, em um primeiro diagnóstico, o câncer de mama já atinge estágios muito avançados dificultando o tratamento e a cura (WHO, 2006). Dessa forma, é de extrema importância que mulheres sejam submetidas a exames mamográficos rotineiramente, a fim de erradicar a doença em fase inicial e até mesmo determinar fatores que despertam atenção para a possibilidade do surgimento da doença no futuro.

Grande número dos exames mamográficos geram dúvidas para um diagnóstico preciso, levando cerca de 80% dos casos a serem encaminhados a novos exames ou biópsias (BOFF; RUARO; SCHUH, 2008). Conforme Cheng et al. (2006) cerca de 65 a 90% dessas suspeitas revelam ser benignas. Biópsias são física e emocionalmente traumáticas para a paciente, acarretando em gastos financeiros tanto para o sistema de saúde que executa o exame quanto para a paciente, além de tempo e aumento da carga de trabalho de radiologistas, patologistas e cirurgiões. Logo, transtornos como esses poderiam ser evitados se o diagnóstico inicial comprovasse a natureza do tumor (NISHIKAWA, 2010; BOFF; RUARO; SCHUH, 2008; ANGELO, 2007; NAZÁRIO, et al., 2007; KOPANS, 2000).

Os fatores acima descritos ressaltam a importância do desenvolvimento de módulos de esquemas CADx com intuito de aprimorar a precisão dos diagnósticos. A maioria dos módulos desses esquemas descritos na literatura implementa técnicas de processamento de imagens e inteligência artificial, como redes neurais artificiais, para a investigação de possíveis achados suspeitos e com o intuito de serem ferramentas que atuem como um segundo observador da imagem mamográfica. A maioria das pesquisas em esquemas CAD tem seu avanço voltado para a detecção de massas e microcalcificações, além de um número menor de trabalhos que promovem a investigação de achados secundários que possam prevenir ou sugerir o aparecimento do câncer de mama (BOSCH et al., 2006; RANGAYYAN; AYRES; DESAULTELS, 2007). A detecção e classificação de assimetrias devem ser implementadas para aprimorar a acurácia de um esquema CAD (TAHMOUSH; SAMET, 2006).

Assim, torna-se evidente a necessidade da investigação de sinais secundários capazes de alertar o mais precocemente possível o surgimento de neoplasias da mama. Isso possibilitaria aos esquemas CADx um desenvolvimento

aprimorado, principalmente na produção de relatórios descritivos, a fim de cumprirem seu objetivo de funcionar como um segundo observador da imagem mamográfica, em auxílio ao radiologista.

1.3 Objetivos

Em função do exposto, este trabalho tem como objetivo geral o desenvolvimento de um módulo de um esquema CADx capaz de auxiliar a detecção e classificação, conforme BI-RADS®, de sinais secundários, mais importantes do ponto de vista do diagnóstico global da mamografia, entre eles: assimetria tecidual, linfonodos axilares e retração de mamilo. A fonte é um conjunto selecionado de imagens mamográficas digitais, utilizando para tanto técnicas automáticas de processamento de imagens e envolvendo também redes neurais artificiais (RNA).

1.3.1 Objetivos específicos

- Associar diferentes técnicas de processamento de imagens para manipulação de imagens mamográficas de forma automática;
- Aprimorar técnicas já desenvolvidas para a detecção automática de estruturas de interesse em imagens mamográficas;
- Auxiliar radiologistas, através do computador, no manifesto de algum sinal secundário em imagens de mamografias clínicas;
- Desenvolver uma aplicação final com interface intuitiva e de fácil manipulação pelo usuário.

1.4 Disposição do Trabalho

Além deste capítulo introdutório, o conteúdo deste trabalho é sumarizadamente descrito abaixo:

2 – A Imagem Mamográfica: descreve informações sobre o câncer de mama, de exames e diagnósticos além das características mais relevantes presentes nas imagens mamográficas que devem ser consideradas para análise digital. Também é descrito os trabalhos de maior relevância para caracterização de achados mamográficos secundários.

3 – Visão Computacional: resumo das principais etapas de sistemas de visão computacional que inclui técnicas de processamento de imagens, extratores de características e modelo de classificação por rede neural artificial.

4 – Materiais e Métodos: descreve os materiais utilizados na elaboração e desenvolvimento bem como a metodologia empregada.

5 – Resultados e Discussões: relata os resultados obtidos com o desenvolvimento deste estudo e a discussão das técnicas empregadas.

6 – Conclusão – parecer final sobre o trabalho desenvolvido e perspectivas para trabalhos futuros.

Referências Bibliográficas: bibliografia utilizada na elaboração e desenvolvimento deste estudo inicial.

2 - A IMAGEM MAMOGRÁFICA

O câncer de mama pode ter seu surgimento baseado em dois fatores: herança genética ou pela alteração genética de uma única célula. No segundo caso, o controle de mutação celular não pode ser determinado, sendo definido como uma ocorrência de aleatoriedade. A consequência gerada por ambos os casos é dada pela proliferação dessas células anormais quando a primária efetua divisões celulares (KOPANS, 2000).

Dessa forma, pode-se definir que a ocorrência de mutações celulares espontâneas não é previamente definida, podendo ocasionar um diagnóstico tardio da doença.

A detecção precoce possibilitada pelo rastreamento mamográfico permanece como único método que comprovadamente reduz a mortalidade por câncer de mama sendo o método de maior eficácia para detectar a doença em estágios iniciais, clinicamente ocultos (KOPANS, 2008; NEWMAN, 1998).

O rastreamento mamográfico é o método diagnóstico eficaz na redução na mortalidade por câncer de mama. A detecção de um processo maligno antes de sua manifestação clínica é tarefa de grande responsabilidade executada pelo radiologista. Nessa área, os problemas mais observados são a falha ou atraso no diagnóstico de câncer, não comunicação de achados relevantes e falso-positivos (biópsias cirúrgicas desnecessárias) (FENELON, 2003, p. 5).

A qualidade da imagem mamográfica é determinada pelo conjunto de efeitos radiográficos, que inclui: o tubo de raios X, o catodo, a janela do anodo, o ponto focal, a filtração por grades, a distância da imagem à fonte, o filme, o *écran*, a reveladora, o sistema de interpretação (negatoscópio ou monitor de vídeo, por exemplo) e até mesmo o posicionamento da paciente e compressão da mama para aquisição da imagem (CALDAS et al., 2005; NEWMAN, 1998). A Figura 2.1 exibe os principais componentes para a formação da imagem mamográfica.

O departamento de saúde norte-americano *Food and Drug Administration* (FDA)² aprovou em 2000 a utilização do primeiro mamógrafo digital de campo total. Nesse equipamento, semicondutores recebem a radiação e a transformam em sinal elétrico que, por sua vez, é transmitido para um computador. Um dos benefícios fundamentais das imagens digitais resulta em imagens dinâmicas de larga escala com contraste melhorado entre tecidos (FREITAS et al., 2006; KOPANS, 2000).

Fig. 3-1 Técnica de execução de mamografias: visão global dos componentes.

Figura 2.1 – Componentes principais para formação de imagens mamográficas
Fonte: HEYWANG-KÖBRUNNER et al., 1999.

² Fonte: U.S. FDA (*Food and Drug Administration* – EUA). *Center for Devices and Radiological Health*. Disponível em: <<http://www.fda.gov/cdrh/mammography>>. Acessado em: 09/08/2010.

Para que a imagem mamográfica forneça a realidade presente no interior dos tecidos da mama e adjacentes é necessário que se capture corretamente toda a área de interesse a ser investigada, de modo que nenhuma região seja omitida ou desprezada. Como premissa para correta caracterização de aspectos diagnósticos clínicos, interpretação da imagem mamográfica e ideal execução de esquemas computacionais para auxílio ao diagnóstico é fundamental a exigência de qualidade na realização do exame mamográfico. Para isso, alguns fatores são ressaltados: o correto posicionamento da paciente no aparelho mamográfico, compressão da mama feita de forma adequada, revelação dos filmes com uso de elementos químicos específicos para a mamografia, alinhamento do ponto focal do aparelho mamográfico, monitores de vídeo com padrões computacionais de calibração pré-determinados, entre outros (IAEA, 2011).

2.1 Posicionamento para captura mamográfica

Para que lesões sejam encontradas através da mamografia, é necessário possuir uma imagem de alta qualidade e com o posicionamento adequado sem que nenhum espaço tecidual deixe de ser investigado (KOPANS, 2000).

As projeções médiolateral oblíqua (MLO) e crâniocaudal (CC) por se completarem, são as principais formas de posicionamento para visualização do interior da mama (ALVARES; ALMEIDA, 2009; JESKE et al., 2000, p. 41; KOPANS, 2000; PATHMANATHAN, 2008; ZHENG, 2008). A projeção MLO tem o objetivo de comprimir a mama o mais próximo do detector para evitar borramento geométrico, fornecendo uma visualização de toda a mama em uma imagem única. Já a CC forma uma imagem que contém os tecidos mediais, não visualizados pela MLO (KOPANS, 2000; BONTRAGER, 2003). Além dessas, existem outras posições possíveis de direcionamento do mamógrafo para melhor investigação, visualizadas na Figura 2.2.

Figura 2.2 – Posíveis direções de captura da mama para produzir imagens mamográficas com maiores detalhes.

Fonte: Kopans, 2000.

Dessa forma, o posicionamento de captura da imagem mamográfica também deve ser levado em consideração para que a representação digital seja o mais condizente possível com a realidade e possa seguir princípios semelhantes aos utilizados profissionalmente no diagnóstico do câncer de mama.

2.2 O BI-RADS®

Com o objetivo de padronizar a descrição geral de informações presentes nas imagens mamográficas, o Colégio Americano de Radiologia (ACR – *American College of Radiology*) propôs em 1980 a elaboração de um sistema de relatório conhecido por BI-RADS® (*Breast Imaging Reporting and Data System*). O BI-RADS® é caracterizado como sistema, pois utiliza variados componentes de padronização da imagem, sendo os principais (TIMMERS et al., 2012; BURNSIDE et al., 2009; KOPANS, 2000):

- a) Um léxico de descritores;
- b) Um relatório estrutural que inclui avaliação final de categorias de doenças junto de recomendações de gestão de acompanhamento;
- c) Uma ferramenta de consulta, auditoria e reunião de dados (atlas de imagens).

Esse sistema visa proporcionar uma melhor comunicação entre os diversos profissionais envolvidos no diagnóstico da imagem mamográfica, uma vez que um segundo observador é requerido para confirmar a presença (ou não) de estruturas na imagem (BURNSIDE et al., 2009).

Atualmente, o BI-RADS® está na quarta edição, publicada em 2003 e conta com um Atlas de imagens para auxiliar e treinar os profissionais no correto diagnóstico do exame mamográfico (TRIMMERS et al., 2012; BURNSIDE et al., 2009; YOUK, et al., 2008).

Educacionalmente, a padronização de relatórios diagnósticos da imagem mamográfica pode auxiliar em treinamentos e aprendizagem na interpretação de exames mamográficos. O BI-RADS® constitui um extenso documento contendo variadas padronizações de descrição dos achados mamográficos, assim como a recomendação de nomenclaturas e demais fatores para serem escritos nos relatórios diagnósticos. Dessa forma, serão resumidamente descritos a seguir os de maior relevância para essa pesquisa.

2.2.1 Densidade mamográfica

Após vários estudos, concluiu-se que o aumento da densidade predominante da mama está intimamente relacionada com casos de câncer de mama (bem evidentes ou omitidos pela alta densidade). Assim, o BI-RADS® padroniza a densidade mamográfica conforme o percentual de tecido adiposo nela presente (WU, BESNEHARD, MARCHESSOUX, 2011). A Figura 2.3 demonstra as diferentes categorias de densidade, conforme classificação pelo BI-RADS®.

Figura 2.3 – Categorias de classificação de densidade mamográfica, conforme BI-RADS.
a) lipossubstituída (Categoria 1: densidade < 25%); b) fibroglandular (Categoria 2: densidade entre 25% e 50%); c) heterogeneamente densa (Categoria 3: densidade entre 50% e 75%); d) extremamente densa (Categoria 4: densidade > 75%)

Cabe ressaltar que, digitalmente, a densidade mamográfica é caracterizada pela alta intensidade dos níveis de cinza dos *pixels* que compõem a imagem. Logo, estruturas consideradas de alta densidade (com intensidade elevada de níveis de cinza), como nódulos e agrupamento de microcalcificações, possuem sua visualização prejudicada em mamas muito densas (estruturas camufladas pelo tecido mamário).

2.2.2 Variação de densidade

Uma das atualizações do BI-RADS refere-se à recorrente confusão de termos entre densidade mamográfica e descritor de densidade. Dessa forma, o termo foi substituído por assimetria.

Por fim, é comprovado que o uso do BI-RADS® tem contribuído como uma ferramenta fundamental na padronização e identificação de casos malignos do câncer de mama, além de fornecer recomendações eficientes para exames adicionais (TIMMERS et al., 2012; WU, BESNEHARD, MARCHESSOUX, 2011; BURNSIDE et al., 2009).

2.3 Aspectos secundários como possíveis fatores de riscos

“A diferença radiográfica entre o tecido normal e o tecido doente é extremamente tênue; portanto, a alta qualidade do exame é indispensável para alcançar uma resolução de alto contraste que permita essa diferenciação” (CALDAS et al., 2005, p. 295).

A seguir são descritos os principais conceitos sobre assimetria tecidual, retração do mamilo e linfonodos.

2.3.1 O tecido mamário assimétrico

Através da análise da imagem mamográfica pode-se constatar a manifestação não apenas de nódulos, microcalcificações ou doenças da mama, mas também a existência de tecido mamário assimétrico (BOZEK; MUSTRA; GRGIC, 2009; KOPANS, 2000). “A assimetria pode indicar um processo significativo; toda mamografia deve ser avaliada quanto à presença de assimetrias” (KOPANS, 2008). Assim, a assimetria tecidual deve ser incluída em relatórios descritivos de interpretação por possuir características singulares, em que comumente é considerada uma variação normal, encontrada com grande frequência. (KOPANS, 2008).

Segundo Kopans (2000) alguns casos de câncer de mama, apesar de palpáveis, não possuem margens mamograficamente demonstráveis, nem produzem depósitos de cálcio, sendo o único achado visível um volume de tecido mamário radiograficamente de maior densidade e assimetricamente diferenciado quando comparado a imagem de mesma projeção. “As assimetrias sempre devem ser cuidadosamente avaliadas, uma vez que podem ser o primeiro e o único indício de um tumor maligno” (HEYWANG-KÖBRUNNER et al., 1999, p. 131).

A interpretação mamográfica inclui a busca de assimetrias. As mamas devem ser comparadas como órgãos simétricos. Uma pode ser maior que a outra, mas as estruturas internas são geralmente bem simétricas em amplas áreas de análise (KOPANS, 2000, p. 256).

Os diferentes achados assimétricos nas imagens mamográficas, conforme estabelecido pela quarta edição do BI-RADS® em 2003, podem ser descritos como sendo assimetria, assimetria focal ou assimetria global (YOUK et al., 2008). A Figura 2.4 exemplifica e descreve resumidamente cada um dos tipos de assimetrias.

Figura 2.4 - Exemplificação dos tipos de assimetrias padronizados pelo BI-RADS.

a) assimetria; b) assimetria global; c) assimetria focal.

Fonte: YOUK et al., 2008.

Conforme a Figura 2.4, a classificação de regiões assimétricas é embasada nos seguintes critérios:

- a) assimetria: dada pela diferença no volume tecidual vista em apenas uma projeção;
- b) assimetria global: dada pela grande variação tecidual (75% da mama) em pelo menos 1 quadrante de apenas uma das mamas, em ambas as projeções;
- c) assimetria focal: dada pelo isolamento de uma porção tecidual, vista em ambas as projeções, de ocupação menor que um quadrante.

A seguir, a Figura 2.5 demonstra imagens mamográficas diagnosticadas com algum tipo de assimetria.

Figura 2.5 – Exames mamográficos com detecção de assimetrias, conforme BI-RADS®

Apesar do tecido mamário assimétrico comumente representar uma variação normal, sua investigação pode ser significativa na determinação do surgimento de possíveis massas ou doenças da mama no futuro (CHANDRA; KARNAN; SIVAKUMAR, 2010; KOPANS, 2000; SCUT; LANCASTER; MANNING, 2006; YOUK et al. 2008).

2.3.2 Retração do Mamilo

O complexo mamilo-areolar pode ser afetado por uma ampla gama de patologias com aparências semelhantes entre si. Assim, faz-se necessária uma aplicação com eficientes habilidades clínicas e de diagnóstico preciso para diferenciação dessas doenças (NICHOLSON; HARVEY; COHEN, 2009).

Quando investigada como sinal secundário da suspeita de câncer, a retração mamilar geralmente está associada a nódulos grandes que são evidentes no exame mamográfico, podendo corresponder a alterações tardias e evidentes clinicamente (KOPANS, 2008). A Figura 2.6 apresenta uma imagem mamográfica diagnosticada com retração/inversão do mamilo.

Figura 2.6 – Retração/inversão do mamilo evidenciada em imagens mamográficas de diferentes projeções.

A mamografia digital, junto das técnicas de processamento de imagens, possibilita uma investigação mais precisa e sistêmica dos tecidos mamários, pele e mamilo (KOPANS, 2008; NICHOLSON; HARVEY; COHEN, 2009). Contudo, apesar

de visível clinicamente, a inclusão da retração do mamilo em relatórios descritivos de esquemas CADx pode alertar para a necessidade de uma investigação mais detalhada e atenciosa da imagem na busca de achados primários, principalmente de nódulos e microcalcificações, além de assimetrias.

2.3.3 Linfonodos Axilares

Linfonodos costumam ser identificados em imagens mamográficas de projeção médio-lateral oblíqua (MLO), conforme apresentado na Figura 2.7 (KOPANS, 2008; LIM et al., 2004). São normalmente pequenos e com forma oval. Quando possuem alta densidade e de forma redonda ou irregular, os linfonodos são considerados anormais e raramente são visualizados sem a presença concomitante de uma massa (LIM et al., 2004).

Figura 2.7 - Imagens de projeção MLO com marcação de linfonodos presentes na área correspondente ao músculo peitoral

Conforme Haywang-Köbrunner et al. (1999) a maioria dos linfonodos visíveis na imagem mamográfica podem ser diagnosticados com base nas suas principais características:

- costumam ser processos ovais, raramente arredondados, de contornos completamente lisos;
- possuem área de maior transparência localizada centralmente e com forma ovalada.

O envolvimento de linfonodos no diagnóstico é frequentemente reconhecido como um dos mais importantes fatores prognósticos da evolução do câncer de mama. Não obstante, cerca de 20 a 30% das pacientes diagnosticadas com linfonodos negativos podem apresentar recorrência da doença e mortalidade elevada, o que, dessa forma, as tornam dignas de uma abordagem terapêutica intensiva, a qual, normalmente, não é utilizada em tais condições (CLARKE et al., 1998; EISENBERG; KOIFMAN; CORDEIRO DE REZENDE, 2001; KOPANS, 2008; PINDER; ELLIS; EISTON, 1995).

Em alguns casos, nodos axilares palpáveis são a primeira indicação de câncer de mama. Neste caso, a análise mamográfica da mama pode resultar na detecção de uma lesão maligna intramamária que está oculta no exame físico (KOPANS, 2008, p. 530).

Conforme Barbu et al. (2010) a análise de linfonodos é uma tarefa difícil que retém parte do tempo diário do trabalho de um radiologista. A detecção automática dos nodos é um importante estágio de tratamento e monitoramento da evolução ou estadiamento do câncer.

2.4 Principais trabalhos desenvolvidos na investigação computacional de achados secundários

A seguir é apresentada uma breve descrição de trabalhos empenhados em auxiliar, por meio computacional, a detecção de sinais secundários presentes em imagens mamográficas.

O trabalho de Lau e Bischof (1991) fez uso da detecção estrutural de assimetrias entre imagens mamográficas para revelar a existência de tumores. Após alinhadas, foram utilizadas medidas de brilho, rugosidade, e direcionalidade, sobre 10 pares de imagens mamográficas, obtendo resultados de 80% na sensibilidade

com um falso-positivo de 4.9% por par de imagens. Apesar do alto valor percentual, o número de imagens submetidas à técnica é relativamente baixo, colocando em dúvida a capacidade de generalização do método. Em seu trabalho, Miller e Astley (1993) utilizaram análise de textura para segmentação do disco fibro-glandular cujos pares de imagens foram comparados através de descritores de formas e distribuição de níveis de cinza para detecção de assimetrias mamográficas. A acurácia obtida foi de 86,7% com taxa de falso-positivo (FP) de 16,7%. Já Méndez et al. (1998) utilizaram a subtração de imagens e binarização para obter informações de áreas suspeitas que, por sua vez, foram delimitadas pelo algoritmo de crescimento de regiões. O método demonstrou uma taxa de verdadeiro-positivo (VP) de 71% com 0,67 de FP por imagem. Ferrari et al. (2001) desenvolveram um método para analisar assimetrias em imagens mamográficas através da detecção de componentes lineares direcionais utilizando uma representação de multirresolução baseada nas *wavelets* de Gabor. A acurácia de classificação obtida foi de 74,4%. Rangayyan, Ayres e Desautels (2007) estenderam os métodos de Ferrari et al. (2001) utilizando descritores morfológicos e momentos geométricos extraídos das imagens segmentadas e alinhadas entre si. Obteve-se na classificação uma acurácia de 84,4%.

A área de estudo proposta nos trabalhos de Miller e Astley (1993), Ferrari et al. (2001) e Rangayyan, Ayres e Desautels (2007), delimitada apenas pelo disco fibro-glandular, pode contribuir para o aumento de casos de falso-negativo (FN). Isso é devido ao fato de que uma área assimétrica tecidual não necessariamente tende a estar presente apenas na região do disco fibro-glandular, mas sim em toda a extensão da mama. Por outro lado, no trabalho de Méndez et al. (1998) o princípio das técnicas utilizadas parece contribuir para um resultado satisfatório, uma vez que são aplicadas na imagem inteira. Uma desvantagem desse último é a utilização de parametrização fixa que algumas técnicas empregadas exigem, tornando necessária a automatização ou substituição das mesmas.

O trabalho de Silva, Schiabel e Nunes (2003) objetivou a medição percentual da diferença assimétrica bilateral e tecidual em imagens mamográficas digitalizadas. Para isso, as imagens foram segmentadas, alinhadas, subtraídas e classificadas estatisticamente como normais ou assimétricas. A técnica utilizada não é automática, assim como no trabalho de Méndez et al. (1998). Sua aplicação é dada apenas para imagens mamográficas de projeção Craniocaudal (CC) o que interfere

na distinção entre assimetria e assimetria focal. Tahmoush e Samet (2006) propuseram o uso da investigação de similaridade de imagens para determinar assimetrias em mamografias através de filtros, descritores de textura e segmentações. O classificador de Bayes apresentou uma taxa de acerto de 84%. No intuito de investigar a relação entre o volume assimétrico e a presença de câncer de mama, Eltonsy, Elmaghraby e Tourassi (2007) submeteram imagens de 350 casos a técnicas de segmentação automática, extração de medidas da altura, largura e volume da área segmentada e aplicação de análise estatística e redes neurais artificiais. Com os resultados obtidos concluiu-se que tecidos assimétricos podem contribuir com a pré-disposição para o câncer de mama. O trabalho de Karnan e Thangavel (2007) utilizou algoritmos genéticos (AG) para detecção automática da suspeita de assimetrias entre pares de imagens mamográficas. Após detectadas as bordas (extração de etiquetas e artefatos) e a posição do mamilo, as imagens foram alinhadas e subtraídas. As taxas de detecção obtidas foram de 89,6%. Apesar dessa elevada taxa de detecção, AG caracteriza-se por ser uma técnica de extrema aleatoriedade em função de probabilidades. Logo, é comum a obtenção de muitos resultados discrepantes para um mesmo par de imagens, fazendo com que o método seja executado repetidas vezes até atingir uma resposta satisfatória.

Em seu trabalho, Bandyopadhyay e Maitra (2010) concentraram-se nas etapas de segmentação da região anatômica da mama, isolamento da sua borda e investigação da presença de massas e assimetrias em 25 pares de imagens do banco MIAS (*Mammographic Image Analysis Society Digital Mammogram Database*). As técnicas de alargamento de contraste, binarização e crescimento de região foram utilizadas, resultando em aproximadamente 90% de acurácia. Xu et al. (2010) também objetivaram evidenciar áreas assimétricas em pares de mamografias clínicas digitais. Foram selecionadas 60 imagens mamográficas do banco DDSM (*Digital Database for Screening Mammography*). Imagens pertencentes a um mesmo exame e de mesma projeção foram alinhadas e aplicada a subtração digital para detecção de assimetrias. Buscando uma maior caracterização das áreas assimétricas detectadas, aplicou-se um mapa de cores sobre as imagens resultantes da subtração digital. Como resultado final descreve-se a eficiência do método apresentado para auxiliar radiologistas na interpretação de exames mamográficos.

Técnicas de processamento de imagens se fazem presentes na maioria das pesquisas com intuito de detectar ou classificar assimetrias em imagens

mamográficas. Muitas dessas técnicas ainda são dependentes de parametrizações fixas ou exigem a manipulação em tempo real pelo usuário, o que as tornam suscetíveis a erros capazes de prejudicar a qualidade do resultado final.

Recentemente, Tzikopoulos et al. (2011) desenvolveram um esquema completamente automático para segmentação e classificação de imagens mamográficas baseado na densidade e assimetria da mama. Neste trabalho obteve-se uma taxa de acerto de 84,47% na detecção de assimetria utilizando o classificador SVM (*Support Vector Machines*). Foram selecionadas 322 imagens de projeção MLO do banco de dados MIAS, sendo 15 pares identificados como assimétricos. A Figura 2.8 descreve todos os passos do algoritmo desenvolvido.

Figura 2.8 - Descrição geral das etapas do algoritmo que compõe o sistema implementado por Tzikopoulos et al. (2011)

Apesar do algoritmo desenvolvido por Tzikopoulos et al. (2011) caracterizar um sistema robusto, o número de casos descritos com algum tipo de assimetria é baixo, o que prejudica a generalização e eficácia do método. Entretanto, a relação de etapas descritas pela Figura 2.8 parece mostrar-se eficiente, uma vez que integra e otimiza variadas técnicas de trabalhos descritos anteriormente.

No intuito de alertar para o risco do desenvolvimento de câncer de mama Wang et al. (2011) propõem 2 modelos de comparação bilateral e classificação dos pares de imagens mamográficas de projeção CC: o primeiro é baseado na extração de 20 atributos globais das imagens; o segundo consiste na subdivisão de cada imagem em 8 regiões de interesse (ROIs) espelhadas, pareadas e igualmente espaçadas entre si, seguida da extração de 16 atributos locais descritores de textura e subtração dos resultados entre os pares de ROIs. Os melhores atributos foram selecionados por AG e 2 RNAs foram implementadas para prever o risco de casos de câncer em desenvolvimento. Usando o método *leave-one-out* na validação, os modelos resultaram na área sob a curva ROC (Az) de 0,75 (primeiro modelo - área completa) e Az de 0,72 (segundo modelo – comparação entre 8 ROIs).

Apesar de utilizar uma metodologia consistente, fundamentada em trabalhos anteriormente descritos, os resultados de Az obtidos pelo trabalho de Wang et al.(2011) podem comprometer a capacidade de generalização do classificador, tornando-o ineficaz quando imagens são submetidas como conjunto de testes. Além disso, o alinhamento entre as imagens não assegura a comparação de mesmas regiões, uma vez que a formação da imagem mamográfica é influenciada, dentre variados fatores, pelo posicionamento da paciente no aparelho mamográfico.

Por fim, o trabalho de Wu, Besnehard e Marchessoux (2011) objetivou a classificação automática do fundo de mamogramas com base nos padrões de densidade estabelecido pelo BI-RADS® e utilizando a *framework* MCA (*Multi Content Analysis*). Foram utilizadas 70 pares de imagens mamográficas digitais as quais foram subdivididas em ROIs. De cada ROI foram extraídos 40 atributos descritores de textura. A seleção dos atributos que melhor caracterizam cada classe foi feita pelo algoritmo iterativo AdaBoost com implementação de árvores de decisão e a classificação foi realizada através da combinação de 4 árvores de decisão pelo *framework* MCA. Na classificação de 4 categorias, são apresentados os seguintes resultados: VP de 90,3%, verdadeiro-negativo (VN) de 67,8%, FP de 32,1% e FN de 9,62%. A Figura 2.9 apresenta os passos implementados pelo trabalho.

Apesar de não detectar variações assimétricas, a metodologia empregada no trabalho de Wu, Besnehard e Marchessoux (2011) desperta interesse para uma investigação comparativa entre as diferenças de densidade em pares de imagens. Um dos diferenciais do trabalho consiste na ausência do alinhamento entre imagens

de mesma projeção, uma vez que as ROIs são analisadas individualmente em cada imagem, possibilitando a comparação baseada na subdivisão de quadrantes.

Figura 2.9 - Descrição geral das etapas desenvolvidas no trabalho de Wu, Besnehard e Marchessoux (2011)

Assim, acredita-se que as metodologias utilizadas nos trabalhos anteriormente descritos, principalmente aquelas utilizadas por Wu, Besnehard e Marchessoux (2011), Wang et al. (2011) e Tzikopoulos et al. (2011), podem ser integradas para o desenvolvimento de um módulo de esquema CAD para detecção de assimetria tecidual em imagens mamográficas.

Desta forma, a integração, automatização de métodos e uso de técnicas inteligentes, como redes neurais artificiais, devem ser investigadas e aprimoradas a fim de proporcionarem respostas satisfatórias, confiáveis e precisas em um curto período.

3 - VISÃO COMPUTACIONAL

Vários estudos vêm sendo desenvolvidos com a finalidade de auxiliar a melhoria visual e, conseqüentemente, a interpretação de imagens mamográficas, analisando critérios de similaridades entre elas. Isto se dá através da identificação e caracterização de anormalidades encontradas em imagens junto dos atributos que as compõem e que possam servir de parâmetros computacionais similares ao sistema visual humano (NATTKEMPER et al., 2008).

Assim, técnicas voltadas ao processamento e interpretação de imagens contribuem para a caracterização de singularidades das mesmas, assim como a obtenção de padrões a serem submetidos à classificação.

3.1 Fundamentos de sistemas visuais artificiais

Sistemas de visão computacional almejam dotar a máquina da capacidade de enxergar através de sistemas de visão artificial com o reconhecimento e interpretação de imagens da mesma forma, ou com maior minuciosidade, que a visão humana (MARQUES FILHO e VIEIRA NETO, 1999).

São poucos os sistemas que implementam todas as fases devido a grande dificuldade e complexidade apresentada (GONZALEZ E WOODS, 2008). Nesse

trabalho, o sistema desenvolvido apresenta as seguintes etapas aplicadas sob as imagens mamográficas: pré-processamento, segmentação, extração e seleção de características, reconhecimento e interpretação.

3.2 Alargamento não-linear de contraste

Este método foi proposto por Silva Jr, Schiabel e Goes (2006) para melhoria de contraste em mamas densas, considerando o nível de intensidades dos *pixels* da imagem digitalizada. Essa técnica age de forma seletiva, de acordo com a densidade mamográfica.

Uma classificação contínua dos graus de densidade é proposta, o que torna o processo dinâmico e independente da análise tecidual, uma vez que é executada a partir da imagem digitalizada. A classificação é totalmente automática e indiferente à quantificação e resolução da imagem digitalizada (SILVA JR.; SCHIABEL; GOES, 2006, p. 85).

Essa melhoria contribui para a separação mais intensa dos níveis de tonalidade das imagens, o que gera uma melhor visualização das estruturas internas, principalmente de massas presentes, mantendo as características originais de textura e geometria dessas estruturas.

3.3 Segmentação

Essa etapa consiste em binarizar a imagem mamográfica de modo que o fundo (0 – *pixel* preto) seja diferenciado da área do tecido mamário (1 – *pixel* branco) através de um limiar. Para isso, nesse trabalho foram utilizados dois limiares globais, descritos a seguir.

3.3.1 Limiar de Otsu

Esse limiar produz um valor global, a partir da classificação dos *pixels* da imagem em duas classes: fundo e objeto. Para isso, leva em consideração a variância entre as classes e a variância total (GONZALEZ & WOODS, 2010; OLIVETE; RODRIGUES; NASCIMENTO, 2004). Esse método tem uma importante peculiaridade de se basear inteiramente em cálculos realizados sobre o histograma da imagem.

A uniformidade dos objetos presentes em imagens digitais tem importante papel para separação destes com o fundo. O método proposto por Otsu (1979) é caracterizado pela ausência da necessidade de parâmetros (ou seja, automático) e não supervisionado para a seleção de limiar, com as seguintes vantagens:

- processo simples e de baixo custo computacional;
- um limiar adequado é selecionado de forma automática e estável;
- baseia-se no histograma (propriedade global da imagem) e não em propriedades locais características de cada imagem (por exemplo, a busca por vales na técnica *watershed*);
- a generalização do método é favorecida, pois não depende de aprendizado supervisionado (adaptação para qualquer imagem em escala de cinza).

3.3.2 Entropia de Pun

Tem por base a premissa de que a informação pode ser modelada como um processo probabilístico. Então, é definida a entropia do objeto e do fundo da imagem, obtendo um limiar global (PUN, 1981).

Com isso, o limiar global obtido por Pun (1981) é baseado na distribuição da probabilidade de *pixels*, obtido através do histograma da imagem (distribuição de intensidade de *pixels*). A obtenção do limiar ideal é baseado na seguinte equação:

$$f(t) = \frac{H_t}{H_T} \frac{\log P_t}{\log(\max \{\rho_0, \rho_1, \dots, \rho_t\})} + \left[1 - \frac{H_t}{H_T} \right] \frac{\log(1 - P_t)}{\log(\max \{\rho_{t+1}, \rho_{t+2}, \dots, \rho_k\})} \quad (3.1)$$

em que, ρ_i é a probabilidade de i níveis de cinza e H_t representa a entropia dos pixels pretos, definida por:

$$H_t = - \sum_{i=0}^k \rho_i \log \rho_i \quad (3.2)$$

H_T é dada pela entropia total, calculada conforme a equação abaixo:

$$H_T = - \sum_{i=0}^k \rho_i \log \rho_i \quad (3.3)$$

e P_t é a probabilidade, definida por:

$$P_t = \sum_{i=0}^t \rho_i \quad (3.4)$$

3.4 Extração de componente conexos

“A extração de componentes conexos de uma imagem binária é essencial para muitas aplicações automáticas de análise de imagem” (GONZALEZ & WOODS, 2010). O objetivo dessa técnica é detectar e isolar objetos presentes em imagens binárias. Dessa forma, os objetos brancos da imagem binária, neste caso a área correspondente a mama, são selecionados a partir de um ponto inicial, denominado X_0 . Isso é possibilitado através das análise de vizinhança de pixel, que requer um elemento estruturante adequado (GONZALEZ & WOODS, 2010), sendo utilizada a conectividade-8.

3.5 Descritores de padrões

Para a produção de informações utilizáveis ao computador, principalmente em sistemas de visão artificial, os descritores de padrões são encarregados de fornecer propriedades representativas que caracterizam cada imagem ou objeto presentes nas mesmas.

Nesta pesquisa foram extraídos muitos descritores das imagens, entre os quais: 9 momentos regulares (JÄHNE, 2002), 7 momentos invariantes de Hu (1962), 29 descritores de forma e 64 descritores de textura. A seguir são descritos apenas os atributos efetivamente selecionados e utilizados nesse trabalho.

Descritores geométricos: são usuais para a classificação em sistemas de reconhecimento de padrões e para descrever simbolicamente os objetos a um sistema de visão artificial, sendo que se especializam sobre imagens binárias (PRATT, 2001). Os atributos descritores de forma utilizados foram:

- *Compacidade* (Comp): medida derivada da área e do perímetro de um objeto, definido pela Equação 3.5, em que P representa o perímetro e A , a área de uma região.

$$Comp = \frac{P}{4\pi A} \quad (3.5)$$

- *Relação entre área e perímetro* (Rel): busca manter uma proporcionalidade do número de pixels do contorno do objeto em relação a área total, definido pela Equação 3.6, em que P representa o perímetro e A , a área de uma região.

$$Rel = \frac{P^2}{A} \quad (3.6)$$

Descritores de texturas: nas imagens mamográficas digitalizadas as diferenças de densidade são percebidas através da variação no nível de intensidade dos pixels. A descrição de cada atributo utilizado é definida por Castella et al. (2007), Gonzalez et al. (2009), Theodoridis & Koutroumbas (1999) e Haralick et al. (1973). Os descritores de textura extraídos a partir da matriz de co-ocorrência conforme Haralick et al. (1973) e utilizados neste trabalho foram:

$$autocorrelacao = \frac{\sum_i \sum_j (i,j) p(i,j) - \mu_x \mu_y}{\sigma_x \sigma_y} \quad (3.7)$$

$$variancia = \sum_i \sum_j (i - \mu)^2 p(i,j) \quad (3.8)$$

em que p é a imagem de entrada, μ_x , μ_y , σ_x , σ_y correspondem à média e o desvio-padrão, respectivamente, da intensidade de pixels da imagem nos eixos x e y , separadamente, e μ à média de todos os pixels da imagem.

3.6 Reconhecimento de padrões

Algoritmos de aprendizagem são selecionados como uma alternativa para classificação de determinados padrões. Normalmente, trabalham com representação algébrica do mundo exterior, considerando os principais pontos de diferenciação de materiais ou objetos de estudo dos quais serão fundamentais na classificação (NATTKEPPER et al., 2008).

As redes neurais artificiais (RNA) podem ser comparadas a técnicas correspondentes de estatística, do reconhecimento de padrões e do aprendizado de máquina, podendo ser aprimoradas e implementadas em aplicações complexas (RUSSELL & NORVIG, 2004).

Conforme Fausett (1994), a medicina utiliza as técnicas e resultados obtidos com redes neurais para auxílio no tratamento e diagnóstico de doenças. Isso é dado segundo a obtenção de características padrões para cada tipo de caso patológico a ser tratado computacionalmente. Assim, com o processamento dos padrões pela rede neural, é possível obter classificações ou respostas para auxílio de um diagnóstico com maior precisão, evitando falhas.

A rede neural *Multilayer Perceptron* (MLP – Perceptron multi camadas) é um modelo muito adequado para a separação de padrões em diversas classes, generalizando o conhecimento para que novos padrões (diferentes daqueles usados no treinamento) possam também ser classificados em uma das classes pré-definidas (HAYKIN, 2008). Dessa forma, a descrição da MLP é destacada devido à sua utilização no desenvolvimento dessa pesquisa.

3.6.1 Rede neural Multilayer Perceptron (MLP)

A rede neural MLP é caracterizada pela presença de uma ou mais camadas intermediárias (escondidas) de neurônios, situadas entre a camada de entrada e a respectiva camada neural de saída. A MLP é utilizada na resolução de diversos problemas com o uso de um algoritmo de treinamento supervisionado (HAYKIN, 2008).

As principais características da MLP, utilizados por esse trabalho, são resumidamente descritas a seguir, conforme Haykin (2008):

- Várias camadas: além da camada de entrada e da camada de saída, possui uma ou mais camadas intermediárias, também conhecidas como camadas escondidas, em que os neurônios são interligados entre si até aos da camada de saída;
- Arquitetura *feedforward* ou acíclica: possui uma única direção – da camada de entrada para a saída.
- Função Sigmóide: utiliza a função logística para a obtenção de um resultado final no neurônio. Pode assumir um intervalo de valores entre 0 e 1. A Figura 3.1 demonstra a representação gráfica da função logística.

Figura 3.1 – Representação gráfica da função sigmóide

- maior número de neurônios em sua composição geral;
- treinamento com algoritmo de retro-propagação (*backpropagation*) do erro para estimular o aprendizado;
- separação de elementos em mais que 2 classes (resolução de problemas que não são linearmente separáveis).

Segundo Russell & Norvig (2004), as camadas ocultas ou intermediárias em uma rede neural aumentam o espaço de hipóteses que a rede pode representar. Compara-se que, com apenas uma única camada, é possível a representação de qualquer função contínua de entradas com exatidão arbitrária e com duas camadas até funções descontínuas.

Após encontrada uma configuração da rede neural que melhor generaliza o problema investigado, os valores dos pesos sinápticos são fixados e a rede neural passa a funcionar apenas no modo operacional (dispensa a correção do erro na etapa de treinamento). A Figura 3.2 apresenta um resumo do fluxo operacional de cada etapa (treinamento e operação) na utilização da rede neural MLP.

Figura 3.2 – Diagrama de blocos das fases de treinamento e operação da rede neural MLP
Fonte: SILVA, SPATTI, FLAUZINO, 2010.

Contudo, é possível notar que o maior custo operacional está relacionado com a fase de treinamento em busca da configuração de maior generalização do classificador. Nessa etapa, inúmeras combinações do número de neurônios sob diferentes parametrizações (taxa de aprendizagem, EQM, *momentum*, etc) são testadas exaustivamente. Já para a aplicação, o custo computacional é bastante reduzido uma vez que uma nova amostra é inserida e apenas é informada a resposta do classificador, independentemente do acerto.

3.6.2 Validação

A validação representa a forma pela qual se analisa a capacidade da rede neural em generalizar casos novos baseando-se nos exemplos utilizados no treinamento. Deve ser utilizada quando se deseja determinar o melhor número de neurônios da camada intermediária, definir melhor ponto de parada de treinamento e para obter maior generalização da rede neural (HAYKIN, 2008).

O conjunto de treinamento é responsável por adaptar os pesos da rede recalculando-os, e o de validação em medir a capacidade de generalização da rede na apresentação de dados desconhecidos à mesma.

Logo, é necessária uma validação da integridade da resposta definida pela rede neural, e proporcionar uma classificação eficaz generalizada nos possíveis conjuntos de padrões estudados e no auxílio na definição de melhor modelo topológico para treinamento dos padrões. Foram utilizados nessa pesquisa 2 métodos de validação:

- *Cross-Validation*: validação cruzada, é utilizado quando o conjunto de amostras selecionadas para o treinamento da rede neural possui grande quantidade de itens. Possibilita a separação em dois conjuntos: teste – aproximadamente de 70% à 90% dos itens – e validação – cerca de 10% a 30% dos itens (HAYKIN, 2008).
- *Leave-one-out*: o conjunto de validação nunca se repete e possui apenas um item que ainda não foi utilizado no treinamento. Normalmente empregado para conjuntos com pouca variedade de amostras (HAYKIN, 2008).

4 - MATERIAIS E MÉTODOS

Como é possível observar até o momento, todas as técnicas anteriormente descritas (aquisição da imagem mamográfica, processamento e segmentação de imagens e descritores de padrões) tornam-se necessárias para a execução eficaz da etapa de classificação. Essas etapas estão intimamente relacionadas entre si, conforme demonstrado pela Figura 4.1.

Nessa pesquisa foram implementadas técnicas tanto para simplificar, auxiliar e acelerar a obtenção de resultados quanto para suprimir lacunas e adicionar funcionalidades a outros métodos. Também foram realizadas análises, testes e associações de técnicas já desenvolvidas e utilizadas em diferentes trabalhos cujo objetivo trata-se da detecção e classificação de sinais secundários e demais estruturas presentes em imagens mamográficas digitais ou digitalizadas. Dessa forma, imagens que contenham informações de interesse para esta pesquisa foram selecionadas e submetidas a técnicas de processamento de imagens. Sobre os resultados obtidos com o processamento e classificação das imagens, são feitas análises estatísticas capazes de demonstrar a eficácia dos métodos empregados.

Figura 4.1 – Principais etapas do sistema inteligente desenvolvido

A implementação da maioria das técnicas foi feita através da linguagem de programação Java por ser gratuita, compilada e multiplataforma, auxiliada pela biblioteca *ij.jar* fornecida juntamente com o *software* gratuito ImageJ (ABRAMOFF, MAGALHAES, RAM; 2004). A seguir, serão descritos os métodos e recursos utilizados no desenvolvimento dessa pesquisa para a implementação do módulo proposto.

4.1 Base de imagens

Para o desenvolvimento desse trabalho verificou-se a necessidade da utilização de 2 bases de imagens: imagens mamográficas reais e imagens que simulam o mamilo. A descrição das características de cada base é feita através dos itens a seguir.

4.1.1 Imagens mamográficas reais

A seleção de imagens mamográficas foi feita através da base do LAPIMO – SEL/EESC/USP em que os laudos médicos descrevem a presença de algum tipo de assimetria. Essa base possui cerca de 6000 imagens, com 12 bits de resolução de contraste (4096 níveis de cinza), resolução espacial de 0,075 mm a 0,15 mm por *pixel* e armazenadas em formato TIFF (*tag image file format*). Recentemente, uma ferramenta foi aprimorada para o gerenciamento dessa base através da internet, denominada *BancoWeb*, cujo acesso pode ser feito através do endereço <<http://lapimo.sel.eesc.usp.br/bancoweb>> (MATHEUS & SCHIABEL, 2011). A Figura 4.2 exemplifica uma tela resultante de uma busca feita na base *online*.

Figura 4.2 – Tela resultante da busca feita na base *online* de imagens mamográficas do LAPIMO

As imagens do *BancoWeb* possuem laudos radiológicos (e anátomo-patológicos e citológicos, quando existentes), com um mecanismo de busca dinâmico e ágil, interface intuitiva e moderna, além de diversas funcionalidades como: ampliação, recorte de áreas de interesse do usuário, *download* das imagens e detalhamento dos dados do exame proveniente de cada imagem (MATHEUS & SCHIABEL, 2011).

4.1.2 Imagens que simulam o mamilo

No intuito de verificar e estabelecer a possibilidade de diferenciação entre o mamilo normal e retraído, foram elaboradas 200 imagens simuladas. Uma parcela do conjunto das imagens pode ser visualizada na Figura 4.3.

Figura 4.3 – Parcela do conjunto de imagens simuladas do mamilo normal e retraído

As imagens são inspiradas em imagens mamográficas segmentadas, com tamanho arbitrário de 50x160, criadas no *software* CorelDraw 12. Anteriormente, semelhante metodologia também já fora utilizada (MENECELLI; RIBEIRO; SCHIABEL, 2010a) e (MENECELLI; RIBEIRO; SCHIABEL, 2010b) para a classificação de nódulos mamográficos.

4.2 Processamento de imagens

Esta etapa constitui a aplicação das técnicas anteriormente descritas a fim de melhorar a qualidade da imagem mamográfica, selecionar apenas objetos de interesse e viabilizar a execução das etapas seguintes. São descritos nesse item: segmentação da imagem da mama para extração do fundo, etiquetas e demais artefatos; delimitação do músculo peitoral em imagens MLO; obtenção automática da região de localização do mamilo em imagens mamográficas; detecção de linfonodos; e a subdivisão da região da mama para investigações focalizadas para a variação de textura dos tecidos do órgão.

4.2.1 Segmentação da região da mama e extração do fundo

O intuito desse processo é automatizar a segmentação da região da mama e remover o fundo e artefatos, constituindo uma etapa de pré-processamento de imagens mamográficas para esquemas CAD em geral. Os passos executados para a execução dessa tarefa são demonstrados pela Figura 4.4 a seguir.

Figura 4.4 - Etapas para segmentação da região da mama

A aplicação de técnicas de processamento de imagens, extração de atributos e consequente classificação é favorecida quando ocorre aprimoramento do contraste das imagens. Dessa forma, a técnica de alargamento não-linear de contraste, proposta por Silva Jr. et al. (2006), é aplicada na imagem resultante final ao processo de extração de fundo e demais artefatos. Detalhadamente, as etapas desse processo constituem:

- a) alargamento de contraste logarítmico (GONZALEZ & WOODS, 2008; FERRARI et al., 2004);
- b) segmentação com limiar global proposto por Pun (1981);

- c) remoção de linhas que contenham 90% dos pixels da sua extensão total valorados como brancos (intensidade máxima) na imagem binária;
- d) definição da proximidade de localização (direita ou esquerda) da mama na imagem (ELTOUKHY; FAYE; SAMIR, 2010), através das coordenadas do ponto de centro de massa (PRATT, 2007);
- e) baseada na direção obtida na etapa anterior, seleção da área correspondente à mama através de um algoritmo de extração de componentes conectados com uso de um elemento estruturante de conectividade 8 (8-conectados) (GONZALEZ; WOODS; EDDINS, 2009). O ponto inicial (X_0) de execução do algoritmo corresponde à coordenada do ponto de centro de massa;
- f) substituição da área branca na imagem final segmentada pela área correspondente da imagem de entrada a fim de se obter, na imagem final, apenas o tecido mamário;
- g) aplicação sobre a imagem do passo anterior a melhoria de contraste proposta no trabalho de Silva Jr. et al. (2006).

Após estas etapas, a imagem resultante possui apenas a região da mama, como visualizado na Figura 4.5.

Figura 4.5 - Resultado da segmentação da mama (eliminação de etiquetas e artefatos).
a) imagem original; b) imagem segmentada; c) imagem final com melhoria de contraste, conforme Silva Jr.; Schiabel; Goes, 2006.

4.2.2 Delineamento do músculo peitoral

A etapa de delimitação do músculo peitoral em imagens mamográficas de projeção MLO é de fundamental importância nessa pesquisa para que a densidade dessa região, normalmente alta e constante, não influencie, de modo geral, os processos de detecção e classificação de assimetrias.

É bastante extenso o número de técnicas empregadas para a remoção ou delimitação do músculo peitoral em imagens mamográficas de projeção MLO. No entanto, muitas delas dependem de parametrizações e/ou possuem alto custo computacional. Após testes e avaliações de variadas técnicas de segmentação, o método proposto por Barbosa Filho et al. (2011) apresentou os melhores resultados e foi utilizado nessa pesquisa. Resumidamente, os passos aplicados para execução dessa etapa podem ser visualizados pela Figura 4.6.

Figura 4.6 – Procedimentos executados para delimitação do músculo peitoral em imagens de projeção MLO

Inicialmente, é demarcada uma ROI que contenha ao menos uma parcela do músculo peitoral. Essa marcação pode ser visualizada pela Figura 4.7. Os pontos (A,C,D,E) descrevem a ROI desejada e são obtidos a partir de coordenadas da imagem segmentada. O ponto C corresponde ao ponto central da extensão da mama e D ao ponto limite da mama para a primeira coluna de pixels (BARBOSA FILHO et al., 2011).

Figura 4.7 – Obtendo a ROI inicial para delimitação do músculo peitoral

a) imagem segmentada com a marcação dos pontos; b) marcação da ROI traçada a partir dos pontos obtidos pela imagem segmentada; c) ROI da parcela do músculo peitoral.

O procedimento seguinte é obter o limiar de Otsu (1979) para a ROI encontrada. Esse limiar é utilizado na segmentação da imagem mamográfica completa (BARBOSA FILHO et al., 2011). A Figura 4.8 exibe os passos desse processo.

Figura 4.8 - Finalização da segmentação do músculo peitoral.

a) ROI com parte do músculo peitoral segmentado pelo limiar de Otsu (1979); b) imagem mamográfica MLO segmentada com o limiar obtido em (a); c) sobreposição de (b) na imagem MLO de entrada com a área do músculo peitoral demarcada

A imagem do músculo peitoral segmentada (Figura 4.8b) é submetida ao processamento morfológico de abertura (GONZALEZ & WOODS, 2008), com elemento estruturante de disco com raio de 25 pixels. Esse processo visa promover a suavização da borda do músculo peitoral.

Contudo, o músculo peitoral das imagens mamográficas de projeção MLO é delimitado podendo ser utilizado em outras etapas, como detecção de linfonodos axilares, ou mesmo removido, como para a detecção de assimetrias.

4.2.3 Localização automática da região do mamilo em imagens mamográficas digitalizadas

Em muitos trabalhos as imagens mamográficas de mesma projeção são alinhadas através do ponto de localização do mamilo. Contudo, a ROI com a presença do mamilo é obtida através de uma técnica simples de processamento de imagens, baseada no método de Mustra et al. (2009). O método implementado nesta pesquisa para localização da região do mamilo é descrito a seguir.

Primeiramente, delimitamos uma região de busca do mamilo a qual compreende a análise de 30% do total de linhas horizontais da imagem que estejam igualmente distribuídas acima e abaixo da linha horizontal central, conforme exemplificado pela Figura 4.9.

Figura 4.9 – Exemplificação da área de busca da região do mamilo.

Na região delimitada busca-se o *pixel* branco mais distante do ponto central, demonstrado pela linha em vermelho na Figura 4.10, obtido através da distância Euclidiana.

Figura 4.10 – Linha com maior número de *pixels* brancos, dentro da região de busca

O ponto de alinhamento ou localização do mamilo é obtido a partir do último *pixel* branco da linha obtida anteriormente, representando o ponto de maior distância Euclidiana em relação ao ponto central da altura da imagem mamográfica (ou da borda de orientação – direita ou esquerda – da mama), conforme Figura 4.11.

Figura 4.11 – Obtenção da ROI contendo a região do mamilo. a) ROI traçada a partir da linha com maior número de pixels; b) ampliação da ROI obtida em a).

O tamanho da ROI foi escolhido a partir da observação de variadas imagens em relação à área ocupada pelo mamilo. Para imagens cujo mamilo é diagnosticado

como retraído procede-se da mesma forma para marcação da ROI. A região de interesse é obtida em imagens com retração mamilar devido à linha com maior número de pixels brancos estar próxima da região onde ocorre a retração/inversão do mamilo, conforme demonstrado na Figura 4.12.

Figura 4.12 – Seleção da ROI da região do mamilo em uma imagem mamográfica com o mamilo retraído. a) imagem original; b) imagem segmentada com a marcação da ROI; c) ROI ampliada

4.2.4 Detecção de Linfonodos

Como um subprocesso do fluxo principal dessa pesquisa, a detecção de linfonodos é iniciada após a delimitação do músculo peitoral, sendo apenas essa região considerada para esse estudo.

Foi utilizado um total de 100 imagens de projeção MLO contendo linfonodos visíveis e descritos no laudo anátomo-patológico de cada caso. Cada músculo peitoral foi submetido ao processo de detecção de pontos de máxima, integrada ao *software* ImageJ (item *Find Maxima*). Essa técnica cria uma máscara em que os pontos com o máximo local da imagem em estudo são marcados. Internamente, o método utiliza a segmentação de Watershed (GONZALEZ & WOODS, 2008) para definição de pontos ou regiões a serem buscadas (ABRAMOFF; MAGALHAES; RAM, 2004).

Nessa pesquisa utilizamos esse *plugin*, pois ele é capaz de realizar marcação de pontos cuja intensidade de *pixels* é alta, característica essa marcante nos linfonodos uma vez que o fundo do músculo peitoral é normalmente composto por uma densidade constante (baixa variação de intensidade de níveis de cinza).

Apesar da técnica exigir uma parametrização inicial, nesse trabalho o método foi adaptado para atuar de forma automática. O parâmetro exigido refere-se ao nível de tolerância de ruído a ser ignorado na marcação dos pontos. A automatização implementada aqui consiste na adaptação desse parâmetro em função do número de pontos detectados. Enquanto o número de pontos detectados for superior a 10, o parâmetro sofre reajustes até a convergência do método. Essa iniciativa foi utilizada após a observação de que, em uma única imagem, o número de linfonodos visualizados é inferior a 7, em média. Pontos detectados nas bordas do músculo peitoral também são removidos conforme propriedade característica da técnica de *Find Maxima* utilizada (ABRAMOFF; MAGALHAES; RAM, 2004), diminuindo o número de ocorrências de falsos-positivos (FP) na detecção.

Além dessa, outras técnicas com objetivo de segmentar os linfonodos foram testadas, entre elas: limiar global proposto por Otsu (1979), entropia de Pun (1981), entropia de Wu (WU; SONDGE; LU, 1998), *k-means* (TAKAHASHI; BEDREGAL; LYRA, 2005), Fuzzy *c-means* (XU, 2009) e rede neural artificial SOM (MARCOMINI & SCHIABEL, 2012). No entanto, nenhuma dessas técnicas forneceu um resultado mais satisfatório do que a técnica utilizada.

4.2.5 Subdivisão automática da região da mama

Após todo o processo de segmentação da imagem, a região da mama está delimitada pelos *pixels* brancos e será útil para que ROIs externas a ela sejam automaticamente removidas do conjunto. O procedimento envolvido nessa etapa tem seu fluxo operacional demonstrado pela Figura 4.13.

Figura 4.13 – Fluxo operacional para classificação de regiões da mama

Assim, a imagem mamográfica de entrada é dividida em regiões de interesse (ROIs) de tamanho iguais a 256×256 pixels. Uma ROI é efetivamente selecionada para estudo conforme a sua localização correspondente na imagem segmentada, considerando a sua orientação, demonstrada pela Figura 4.14.

Figura 4.14 – Correspondência entre uma ROI na imagem binária e imagem mamográfica

Se uma ROI, na imagem segmentada, possui mais de 95% de pixels brancos em relação a sua área total, então essa ROI é selecionada para a próxima etapa. Do contrário, a ROI é automaticamente excluída.

Com isso, ROIs que contenham informações do fundo da imagem não são analisadas pelos radiologistas que, por sua vez, iriam excluí-las do conjunto. Ao final dessa etapa, 1.945 ROIs foram obtidas a partir de 320 imagens mamográficas iniciais de entrada.

4.2.6 Classificação manual das ROIs por um especialista

Esta etapa tem por objetivo mapear as variações de densidades presentes no tecido mamário. Isso é feito a partir das regiões anteriormente obtidas. Cada região é analisada conforme a variação do nível de intensidade de cinza da imagem.

Para comprovar o acerto na etapa de classificação de cada ROI, é necessário o estabelecimento de um *ground truth*, que consiste em alocar cada ROI em uma das 4 classes de densidade estabelecidas pelo BI-RADS®. Neste trabalho optou-se pela classificação direta de cada ROI pelo radiologista. Para isso, foi desenvolvida uma aplicação com uma interface intuitiva, contendo mecanismos essenciais para a promoção de agilidade e auxílio na tarefa do radiologista, demonstrada pela Figura 4.15.

Figura 4.15 – Interface desenvolvida para classificação manual das ROIs pelo radiologista

Nesta etapa, todas as 1.945 ROIs, automaticamente obtidas anteriormente, são submetidas ao *software* de classificação manual. De acordo com a nomenclatura das ROIs, a imagem mamográfica original é carregada lateralmente e a ROI a ser classificada é demarcada em amarelo, como demonstrado na Figura 4.16, tudo de forma automática.

Figura 4.16 – Marcação da ROI a ser classificada na imagem mamográfica completa

Ao analisar a ROI, o radiologista digita o número da operação a ser realizada, conforme legenda descrita na interface do *software*. Ao digitar a tecla *Enter*, a ROI é classificada de acordo com a opção do radiologista; uma nova ROI é automaticamente carregada e demarcada na imagem mamográfica original, exibida lateralmente. Isso faz com que, em pouco tempo, todas as ROIs sejam classificadas. A Figura 4.17 demonstra parcela do conjunto de ROIs selecionadas pelo radiologista em cada uma das classes.

Figura 4.17 – Conjunto de imagens selecionadas pelo radiologista para cada uma das classes de densidade, conforme BI-RADS®

Cabe ressaltar que foram classificadas manualmente apenas as ROIs em que o radiologista demonstrou certeza da densidade predominante, sendo que as demais foram descartadas. Ao final dessa etapa, um total de 940 ROIs foram selecionadas, sendo 235 para cada uma das 4 categorias de densidade, estabelecidas pelo BI-RADS®.

4.3 Seleção de Atributos para classificação

Previamente à classificação, os atributos extraídos das imagens são analisados estatisticamente para selecionar apenas aqueles que realmente descrevem padrões singulares de cada classe. O método utilizado nesta pesquisa para a parametrização dos dados é através da distribuição gaussiana (PATROCINIO, ANGELO, SCHIABEL, 2010; PATROCINIO et al., 2002), com sua implementação feita pelo *software* Matlab versão 7.7.0 (R2008b). Cada atributo é normalizado e, para todas as classes submetidas à análise, é gerada a curva gaussiana referente à distribuição dos valores. Quanto menor a intersecção das curvas, maior a probabilidade de um atributo representar cada classe singularmente. Além da intersecção das curvas, também analisa-se, no eixo das abscissas, a faixa de distribuição de valores de cada classe para um mesmo atributo e, entre aqueles selecionados, os valores entre eles. Isso evita que sejam selecionados atributos que apresentem uma mesma faixa de valores, representando uma ambiguidade de informações para a entrada da rede neural.

4.4 Classificação por redes neurais artificiais

Com base nos principais periódicos e trabalhos que envolvem classificação por redes neurais artificiais, foi constatado que a MLP (*Multi-layer Perceptron* – Perceptron multi-camadas) é uma técnica inteligente capaz de fornecer resultados satisfatórios. Esta rede neural tem sido utilizada em variadas áreas para solucionar diversos problemas envolvendo a classificação e reconhecimento de padrões.

Dessa forma, as classificações realizadas nessa pesquisa foram feitas através da rede neural artificial MLP com algoritmo de treinamento de *backpropagation* e utilização da função de ativação logística (HAYKIN, 2008). Esta mesma rede neural já foi utilizada anteriormente para a classificação de padrões existentes em imagens com neoplasias mamárias (MENECELLI, RIBEIRO, SCHIABEL, 2010a; MENECELLI, RIBEIRO, SCHIABEL, 2010b). Devido ao elevado número de amostras, a validação cruzada (*cross-validation*) foi utilizada.

Essa rede neural foi utilizada para os seguintes procedimentos de classificação: mamilo normal ou retraído; subregiões da mama, de acordo com a densidade mamográfica predominante; estimacão da densidade mamográfica, conforme categorias estabelecidas pelo BI-RADS®. Também foi utilizada para diminuir casos de Falsos-positivos (FP) na detecção de regiões assimétricas teciduais.

4.4.1 Classificação do mamilo

Devido ao baixo número de imagens mamográficas com laudos indicativos da retração mamilar, esse processo do trabalho foi dividido em duas etapas: (1) verificacão da viabilidade para diferenciar mamilo normal de retraído pela rede neural artificial em 200 imagens simuladas; (2) treinamentos, validacões e testes para o estabelecimento de uma configuracão ideal de generalizacão do classificador neural através de 100 imagens mamográficas. Na Figura 4.18 é possível observar a descriçao resumida de cada uma das etapas.

Foram extraídos 29 atributos descritores de formas de cada imagem (MENECELLI, RIBEIRO, SCHIABEL; 2011; MENECELLI, RIBEIRO, SCHIABEL; 2010a) a fim de verificar um conjunto de características capazes de discriminar o mamilo normal do retraído. Em seguida, a análise da distribuicão gaussiana possibilitou a seleçao dos atributos capazes de diferenciar as classes. Por fim, a rede neural MLP foi utilizada para classificar as imagens (simuladas e reais) e definir uma configuracão final.

Para testes e verificacão da capacidade de generalizacão da rede neural MLP, foram utilizadas 52 imagens mamográficas digitalizadas (26 diagnosticadas

com mamilo normal e 26 diagnosticadas com mamilo retraído – diferentes daquelas utilizadas no treinamento e validação).

Figura 4.18 – Esquematização das etapas para o desenvolvimento do classificador final

4.4.2 Classificação da densidade mamográfica, conforme BI-RADS

A partir da classificação das sub-regiões da mama, conforme a densidade mamográfica predominante em cada imagem, foi facilitada a determinação da densidade mamográfica, conforme as categorias estabelecidas pelo BI-RADS®.

Com as 940 ROIs devidamente classificadas, aplicamos o método de Liu et al. (2011) para estabelecimento da porcentagem estimativa da densidade mamográfica total e classificação, conforme o BIRADS®. Assim, a densidade mamográfica é definida por:

$$d = N / M \quad (4.1)$$

em que N é o número de ROIs classificadas em uma determinada densidade mamográfica e M é o total de ROIs obtidas para a imagem em análise.

A reunião de todas as etapas anteriormente descritas resultou na implementação de um *software* final completamente automático, responsável por fornecer uma taxa percentual em relação à variação de densidade mamográfica. Apenas a imagem mamográfica é passada como parâmetro para o *software*, que realiza uma série de procedimentos, conforme a Figura 4.19.

Figura 4.19 - Diagrama do fluxo operacional do software classificador de densidade automático

Com uma interface simples e intuitiva, a Figura 4.20 demonstra a tela do *software* desenvolvido para determinação da categoria da densidade mamográfica.

Figura 4.20 - Interface do software final de classificação automática de densidade mamográfica.

No intuito de proporcionar maior funcionalidade ao radiologista, o *software* possibilita o carregamento de múltiplas imagens e os resultados das respectivas classificações são detalhadamente salvos em um arquivo *Comma-separated values* (CSV).

4.5 Detecção e classificação de assimetria tecidual

4.5.1 O processo de detecção

Assim como na classificação da densidade mamográfica, conforme BI-RADS[®], esta etapa também é baseada na subdivisão da mama em regiões. Cada quadrante da mama é dividido em quatro regiões, conforme ilustrado na Figura 4.21, sendo estas o alvo de estudo dessa etapa.

Figura 4.21 – Divisão da mama em (a) quadrantes e (b) ROIs para busca de assimetrias

Cabe ressaltar que, primeiramente, o músculo peitoral é demarcado e removido para, depois, as imagens de projeção MLO serem subdivididas em quadrantes. De cada imagem obtém-se em média cerca de 16 ROIs.

Para dar continuidade a este processo é de extrema necessidade que haja pelo menos um par de imagens bilaterais de uma mesma projeção. Então, a imagem da mama direita é espelhada horizontalmente para que ambas (imagem da mama direita e esquerda) possam ser comparadas. De cada uma das 16 ROIs obtidas em cada imagem são extraídos os atributos de autocorrelação e variância, descritos anteriormente. Assim, na Figura 4.22, a seta indica uma das regiões obtidas a qual será comparada na imagem bilateral.

Figura 4.22 – Correspondência entre ROIs bilaterais para a detecção de assimetrias
a) imagem CCD; b) imagem CCE; c) Imagem CCD espelhada; d) imagem CCE

Como os descritores de textura consideram a direção da intensidade de pixel, o espelhamento das imagens para que ambas estejam em uma mesma direção é muito importante. Assim, para alguns dos atributos de textura foi utilizada a matriz de co-ocorrência para sua extração, sob as direções angulares de 0°, 45°, 90° e 135°. Como a direção de uma assimetria não é relevante para o diagnóstico, foi utilizado o valor médio entre os 4 ângulos direcionais de cada atributo de textura.

Em seguida, é feita a comparação dos valores obtidos pelos descritores de textura entre as ROIs correlacionadas bilateralmente entre si. Em cada ROI é fornecido um único valor numérico para o atributo de autocorrelação e outro valor numérico para o atributo variância. A seguir, é demonstrada a formulação matemática de comparação dos atributos em uma mesma ROI, desenvolvida nessa pesquisa.

Inicialmente, o valor de cada atributo é normalizado (entre 0 e 1) perante o total de ROIs obtidas. Em seguida, é obtido o percentual de variação de um mesmo atributo entre ROIs correlacionadas:

$$pV(i, j) = \left[1 - \left(\frac{\min(ROIe; ROI d)}{\max(ROIe; ROI d)} \right) \right] \times 100 \quad (4.2)$$

em que, *ROIe* representa o valor do atributo *j* na região da mama esquerda, *ROI d* o valor do mesmo atributo *j* na região da mama direita, *i* a ROI estudada em ambas as imagens, sendo $1 \leq i \leq 16$ e $1 \leq j \leq 2$.

A Equação 4.2 é aplicada para todas as ROIs e para cada um dos atributos extraídos das ROIs. Em seguida, é obtida a média aritmética da variação percentual final (vpf) entre os atributos, definida por:

$$vpf(i) = \frac{[pVa(i) + pVv(i)]}{2} \quad (4.3)$$

sendo que pVa representa o percentual de variação da autocorrelação em uma determinada ROI, pVv o percentual de variação do atributo variância para a mesma ROI e i a ROI estudada em ambas as imagens, sendo $1 \leq i \leq 16$.

A Tabela 4.1 esboça a obtenção dos atributos para cada ROI de cada imagem assim como os resultados de pV e vpf . Note que as ROIs de número 7 e 16 foram selecionadas como sendo assimétricas.

Tabela 4.1 – Comparação entre os atributos descritores de textura para detecção de assimetria em imagens bilaterais de mesma projeção

	ROIs mama Direita		ROIs mama Esquerda		pV		vpf
	Auto Correlação	Variância	Auto Correlação	Variância	Auto Correlação	Variância	
1	82181,79	82393,13	76639,77	76887,08	6,74361	6,682656	6,7131
2	71083,14	71303,89	69040,24	69263,58	2,873951	2,86143	2,867
3	68133,44	68287,55	59101,86	59213,8	13,25573	13,28757	13,271
4	79523,3	79489,61	75951,82	75992,27	4,491106	4,399735	4,445
5	28789,5	28994,36	27985,97	28167,02	2,791031	2,853426	2,8222
6	76436,38	76488,93	66034,16	66068,3	13,60899	13,62371	13,616
7	736,3217	762,0669	1069,288	1105,119	31,13903	31,04208	31,090
8	37478,84	37700,07	38645,66	38869,53	3,019267	3,008684	3,0139
9	61468,44	61707,31	75618,79	75859,65	18,71274	18,65594	18,684
10	62413,87	62652,48	76053,52	76281,14	17,93428	17,86635	17,900
11	73858,86	73804,38	78602,02	78587,14	6,034405	6,085929	6,0601
12	59938,57	59904,99	61286,72	61287,88	2,199742	2,256398	2,228
13	72842,53	72857,23	71615,82	71613,33	1,684052	1,707312	1,6956
14	28628,47	28646,58	25485,28	25513,43	10,97923	10,93725	10,958
15	33736,36	33828,33	32821,54	32941,08	2,711673	2,622773	2,6672
16	809,0984	818,2156	503,59	510,2151	37,75911	37,64295	37,701

Uma região assimétrica é inicialmente detectada quando a vpf é superior a 35%. Esse percentual fixo foi determinado após a aplicação de alguns testes e verificação da eficiência na detecção, principalmente, de assimetrias focais. Foram

testados alguns limiares e comprovada a capacidade de detecção, conforme a Tabela 4.2 e Figura 4.23.

No entanto, sob uma avaliação inicial dessa técnica por uma médica especialista em laudar imagens mamográficas, foi comprovada a existência de muitas regiões definidas como FP (regiões bilateralmente normais foram marcadas como assimétricas).

Para diminuir o número de FP, cada região em que $35\% \leq vpf \leq 60\%$ e que, inicialmente, foi definida como assimétrica teve seus atributos submetidos como entrada para classificação pela rede neural MLP. A configuração estabelecida pela MLP é a mesma para classificação de subregiões da mama, definida em 4.2.5. A diferença aqui se deve à aleatoriedade do tamanho das ROIs que, no primeiro caso, possuíam tamanhos fixos. Dessa forma, ao submeter uma ROI ao classificador, a densidade predominante será informada conforme as categorias definidas pelo BI-RADS[®]. Se um par de ROIs correlacionadas bilateralmente resultar em classificação de densidades diferentes, então a região que possuir maior percentual de tecido denso (intensidade de *pixel* mais elevada) é marcada como sendo assimétrica. Do contrário, a região é marcada como normal e tem-se a redução de casos de FP, como pode ser constatado pelos dados da Tabela 4.2.

Tabela 4.2 – Verificação do limiar percentual de detecção de diferença assimétrica

Limiar (%)	Com MLP		Sem MLP	
	VP (%)	FP*	VP (%)	FP*
20	39	3,2	30	7,3
30	79	1,1	77	5,5
35	82,8	0,7	84	5,1
40	71	0,3	65	3,2

*FP refere-se à média de regiões detectadas erroneamente por imagem

Figura 4.23 – Curva ROC para o limiar de 35% de diferença para considerar uma região assimétrica

Regiões cujo *vpf* é superior a 60% são automaticamente demarcadas como assimétricas, sem a necessidade de validação pelo classificador neural.

4.5.2 Classificação de assimetrias

O objetivo dessa etapa não consiste em diferenciar os tipos de assimetrias, mas informar o grau de pertinência percentual de detecção para cada um dos tipos. Em uma mesma imagem pode-se encontrar tanto assimetrias focais quanto assimetrias. Dessa forma, essa etapa é responsável por exibir ao usuário o percentual de ocorrência de cada uma delas.

Diferentemente dos processos anteriores, a rede neural MLP não foi utilizada nessa etapa. Para simplificar e aprimorar o tempo computacional de execução da aplicação final, a classificação de assimetrias é realizada através de comparações lógicas:

a) assimetria: quando uma região é marcada na imagem CC e nenhuma região correspondente é marcada na imagem MLO e vice-e-versa. O percentual de assimetria (*pAa*) é obtido pelo índice *pvf*, anteriormente definido;

b) assimetria global: quando 5 ou mais ROIs adjacentes (superior a 1 quadrante) de uma mesma imagem (mama direita ou esquerda) são marcadas como assimétricas em ambas as projeções CC e MLO. O percentual de assimetria global (pAg) é dado pela média aritmética de vpf :

$$pAg = \frac{\left[\sum_{i \leq R}^{i=1} vpf(i) \right]}{R} \quad (4.4)$$

em que R é o número total de ROIs assimétricas em um par de imagens de mesma projeção.

c) assimetria focal: é dada quando uma ROI é classificada como assimétrica em uma imagem de determinada projeção e a região correspondente na outra projeção da mesma direção de imagem é também marcada com assimetria. O percentual de assimetria focal (pAf) é obtido pelo índice pvf , anteriormente definido, para a região marcada.

Exceção ocorre quando uma assimetria global é detectada. Não é necessária a investigação de assimetrias ou assimetrias focais uma vez que mais de 75% da mama já foi considerada assimétrica. Contudo, a classificação da assimetria detectada é capaz de alertar os especialistas para uma investigação mais apurada da imagem mamográfica, além de determinar o grau de pertinência computacional para o achado mamográfico.

4.6 Métricas de Avaliação

No intuito de verificar quantitativamente a eficiência de segmentação tanto da região da mama quanto do músculo peitoral, foram pesquisadas na literatura métricas capazes de fornecer resultados estatísticos do desempenho dos segmentadores. Todas as métricas utilizadas realizam uma comparação entre a imagem resultante de um determinado método computacional de segmentação com uma imagem segmentada por um especialista considerada ideal e denominada padrão ouro ou *ground-truth*. A seguir são descritas as métricas pesquisadas e implementadas nessa pesquisa.

A métrica conhecida como *Area Overlap Measure* (AOM) ou medida de similaridade de Jaccard, é utilizada para avaliar a performance do algoritmo de segmentação (ELTER & HELD, 2010; PEI et al., 2010), descrita como:

$$AOM = \frac{A_{seg} \cap A_{GT}}{A_{seg} \cup A_{GT}} \quad (4.5)$$

em que, A_{seg} são todos os *pixels* da imagem resultante da segmentação e A_{GT} os *pixels* da imagem considerada como *ground-truth*. Para a AOM o valor a ser considerado ideal é igual a 1.

Entretanto, a medida AOM não fornece informações referentes à falta ou excesso de segmentação. Para isso, as medidas *Area Under-Segmentation* (AUM) e *Area Over-Segmentation* (AVM) foram também calculadas, conforme as equações a seguir (PEI et al., 2010):

$$AUM = \frac{A_{GT} - A_{seg}}{A_{GT}} \quad (4.6)$$

$$AVM = \frac{A_{seg} - A_{GT}}{A_{seg}} \quad (4.7)$$

No intuito de verificar a generalização do segmentador, a medida de combinação (CM – *Combination Measure*) é calculada utilizando as informações obtidas pelas métricas AOM, AUM e AVM, conforme a equação abaixo (PEI et al., 2010):

$$CM = [AOM + (1 - AUM) + (1 - AVM)]/3 \quad (4.8)$$

Também na faixa entre 0 e 1, a medida CM deve possuir resultados altos, próximos a 1, para uma segmentação de maior precisão (PEI et al., 2010).

Com a finalidade de verificar e esboçar quantitativamente os resultados da aplicação de diferentes detectores de bordas em imagens, Boaventura (2010) utilizou as medidas Acurácia e Taxa de Erro, definidas por Fawcet (2006). Utilizando

semelhante metodologia, neste trabalho as medidas foram implementadas para verificar a exatidão de uma região segmentada e não apenas a borda. A seguir, são descritos os passos e conceitos utilizados para obtenção das medidas.

A matriz de *pixel* que representa a imagem é o conjunto de amostras em que cada *pixel* pode assumir valores positivo e negativo. Nesse caso, *pixels* positivos referem-se àqueles pertencentes à região buscada (área branca de valor 1) e *pixels* negativos referem-se àqueles do fundo e adversos à área desejada (área preta de valor 0). Dessa forma, ao comparar os *pixels* de uma imagem resultante de um processo de segmentação computacional com a mesma imagem segmentada manualmente e definida como padrão ideal (*ground-truth*), têm-se as seguintes situações distintas:

- *Pixel* pertencente à região desejada na imagem de *ground-truth* (*pixel* positivo): quando é corretamente segmentado (positivo) na imagem resultante, é considerado como Verdadeiro Positivo (VP); se não foi segmentado como pertencente à região buscada (negativo), é considerado Falso Negativo (FN).

- *Pixel* não pertencente à região desejada na imagem de *ground-truth* (*pixel* negativo): quando o *pixel* na imagem resultante é classificado corretamente (negativo), ou seja, não pertencente à região buscada, é considerado Verdadeiro Negativo (VN); se é classificado como pertencente à região buscada (positivo), é considerado Falso Negativo (FN).

Dessa forma, é possível a elaboração de uma tabela de contingência (ou matriz de confusão) que relaciona os dados acima descritos e exemplificada pela Tabela 4.3 (FAWCET, 2006).

Tabela 4.3 – Matriz de confusão

ground-truth segmentada	Pixel Positivo	Pixel Negativo
Pixel Positivo	VP	FP
Pixel Negativo	FN	VN

A partir da matriz de confusão é possível calcular variadas métricas capazes de analisar as características de um detector de borda e de segmentadores de regiões. Neste trabalho são utilizadas as medidas de acurácia (exatidão ou taxa de

acerto) e taxa de erro (TxErr) da imagem segmentada em relação a imagem definida como *ground-truth* pelo especialista.

A acurácia (A) compreende a proporção de *pixels* corretamente segmentados para o total de *pixels* da imagem, definida pela equação abaixo:

$$A = \frac{VP + VN}{FN + FP + VP + VN} \quad (4.9)$$

A faixa de resultados fornecidos pela acurácia está entre 0 e 1, sendo que o valor 1 representa a segmentação ideal.

A taxa de erro é dada pela proporção do número de todos os *pixels* classificados erroneamente para o total de *pixels* classificados (BOAVENTURA, 2010; WITTEN & FRANK, 2005), conforme demonstrado pela equação abaixo:

$$TxErr = \frac{FP + FN}{VP + VN + FP + FN} \quad (4.10)$$

Assim como as demais, esta métrica resulta valores entre 0 e 1. Como o intuito é que se tenham poucos *pixels* classificados erroneamente, resultados próximos ou iguais a 0 são considerados satisfatórios ou ideais (WITTEN & FRANK, 2005).

Outra metodologia proposta por Wirth et al. (2005) possui conceitos semelhantes à aplicação da matriz de confusão, em que se compara a classificação de cada *pixel* das imagens segmentadas com os mesmos *pixels* na imagem de *ground-truth*, porém com uma abordagem específica para imagens mamográficas digitais/digitalizadas. Para isso, as seguintes classes são consideradas: VP (verdadeiro-positivo): *pixels* corretamente classificados, conforme imagem de *ground-truth*; FP (falso-positivo): *pixels* classificados como brancos, porém deveriam ser pretos (*over-segmentation*); FN (falso-negativo): *pixels* classificados como pretos, porém deveriam ser brancos (*under-segmentation*). A Figura 4.24 demonstra como a classificação dos *pixels* é feita, segundo Wirth et al. (2005).

Figura 4.24 – Índice de performance (VP, FP, FN) obtido pela classificação

A medida *quality* foi utilizada por Wirth et al. (2005) para verificar a taxa de acerto final da segmentação que considera valores de excesso e falta de segmentação. O valor considerado ideal deve ser igual a 1 para a medida *quality* cuja definição é dada por:

$$Quality = \frac{VP}{FN + FP + VP} \quad (4.11)$$

Ainda conforme Wirth et al. (2005) o pré-processamento das imagens afetaria os resultados obtidos com as métricas, uma vez que, neste caso, o *ground-truth* é obtido a partir da análise visual das imagens.

4.7 Síntese: módulo desenvolvido do esquema CAD

Como produto final dessa pesquisa, foi desenvolvido um *software* que deverá compor mais um módulo do esquema CAD que vem sendo implementado ao longo dos últimos anos pelo LAPIMO (SCHIABEL et al., 2011). Nesta seção, a título de melhor clareza do que foi desenvolvido no presente trabalho, sintetizam-se todas as etapas do processo de sua elaboração, conforme as descrições mais detalhadas apresentadas nas seções anteriores deste capítulo.

O *software* recebe como parâmetros de entrada as imagens digitais ou digitalizadas provenientes de exames mamográficos. Em seguida, todo o processo de segmentação, detecção e classificação das regiões assimétricas suspeitas é feito de forma automática, com uma interface intuitiva e detalhada para as operações realizadas sobre as imagens. Adicionalmente a etapa de detecção e classificação de assimetrias, a retração do mamilo, detecção de linfonodos em imagens MLO e estimação da densidade mamográfica são realizadas.

A linguagem de programação JAVA e ambiente de desenvolvimento integrado (IDE) NetBeans 7.2 foram escolhidos para o desenvolvimento da ferramenta final. Isso se deve ao fato de que a linguagem independe de sistema operacional para ser executada (multiplataforma), ser compilada (e não interpretada, como no Matlab), flexível, considerada de alta performance e gratuita. Juntamente à linguagem, utilizou-se a biblioteca de desenvolvimento *ij.jar* fornecida junto ao *software* devido à facilidade, disponibilidade gratuita e integração da biblioteca com a linguagem de programação, através de muitos métodos e funcionalidades previamente desenvolvidos. Também foi considerada a capacidade, viabilidade e eficácia do *software* ImageJ, fundamentalmente desenvolvido para análise e manipulação de imagens médicas, que fundamenta o desenvolvimento desta pesquisa.

4.7.1 Fluxo operacional do módulo

O tempo de processamento do sistema completo, desde o início dessa pesquisa, foi planejado para ser otimizado. Isto se deve ao fato de que as saídas obtidas pelos processos e seus respectivos sub-processos, executados para a realização de uma única tarefa, serão reutilizadas para a implementação de etapas consequentes. Assim, tomando-se por base a detecção e classificação de assimetrias como fluxo principal, demais investigações, como a retração do mamilo e detecção de linfonodos, são executadas a partir de um desvio no fluxo principal. A Figura 4.25 ilustra o modelo do fluxo de operação proposto para o desenvolvimento do algoritmo do sistema.

Figura 4.25 – Fluxo operacional para o módulo do sistema CAD proposto

A seguir, descreve-se objetivamente a funcionalidade das tarefas principais a serem executadas internamente pelo sistema:

- Extração de artefatos e ruídos: marcações indesejadas são comumente encontradas nas imagens mamográficas. Logo, é necessária a remoção dessas através de técnicas de processamento de imagens, entre elas, melhorias de contraste, segmentação e análise de vizinhança de pixels. Ao final, tem-se como resultado uma imagem contendo apenas a área correspondente à porção ocupada pela mama.
- Detecção automática da região do mamilo: com finalidade de investigar a normalidade do mamilo, é processada a sua localização para posterior classificação.
- Delimitação do músculo peitoral em imagens MLO: a fim de evitar a detecção FP de áreas assimétricas e delimitar a região de busca de linfonodos, o músculo peitoral é identificado para que apenas a porção correspondente à mama seja analisada. Isso também se faz necessário em função da comparação feita com imagens de projeção CC que, segundo Kopans (2008), é essencial para determinar se uma assimetria é real ou consequência de problemas técnicos na realização do exame mamográfico.
- Subdivisão da região da mama: cada quadrante da mama é dividido em 4 regiões. Serão essas as regiões analisadas para a busca de assimetrias na região bilateralmente correspondente na imagem de mesma projeção.
- Extração de atributos: dentre os principais atributos, são analisados descritores geométricos, de textura, momentos regulares e invariantes.
- Detecção de regiões assimétricas: é feita, inicialmente, pela comparação dos atributos de autocorrelação e variância entre regiões de um mesmo quadrante. Para reduzir o FP, a classificação da densidade de sub-regiões da mama pela rede neural MLP é aplicada e o resultado comparado com a detecção inicialmente realizada.
- Classificação das assimetrias: confronta-se a localização das regiões detectadas entre as imagens de diferentes projeções (CC e MLO). Conforme os resultados obtidos, é informado ao usuário o percentual de regiões consideradas como assimetria, assimetria global, assimetria focal ou ausência de regiões assimétricas.

- Retração de mamilo: com a finalidade de reduzir o custo computacional, o algoritmo de retração ou inversão de mamilo tem início após a etapa de detecção do mesmo. Assim, após determinada a área comum de localização do mamilo em cada imagem mamográfica (região de interesse – ROI), a mesma será submetida à extração de atributos e classificação em normal ou retraído pela rede neural MLP.
- Detecção de linfonodos: também no intuito de otimizar tempo de processamento, a detecção de linfonodo é feita a partir do delineamento do músculo peitoral. A área delimitada é investigada para detecção da presença de linfonodos.
- Categorização da densidade: como a mama foi subdivida em regiões nos processos anteriores, aplica-se a classificação de densidade predominante de cada mama, conforme estabelecido pelo BI-RADS®.

Seguindo os princípios do classificador de assimetrias, a retração do mamilo também foi feita através da rede neural MLP com configuração específica para esta investigação. Dessa forma, a utilização da mesma rede neural em ambas as classificações (de assimetria e retração de mamilo) torna o sistema mais rápido, já que a única alteração a ser feita será a configuração da rede neural artificial para solucionar cada problema distinto.

Logo, ao final do desenvolvimento desse sistema computadorizado para detecção e classificação de aspectos secundários no auxílio à caracterização de casos de câncer de mama, relatórios com maior detalhamento – isto é, uma saída a mais para o esquema CAD como um todo – são produzidos podendo auxiliar profissionais da área médica na precisão de seus diagnósticos.

5 - RESULTADOS E DISCUSSÕES

Essa pesquisa conta com uma variada gama de processos cujos resultados refletem nas etapas subsequentes. Dessa forma, é de extrema importância verificar a viabilidade e capacidade de generalização de cada passo implementado. A seguir, são descritos os resultados obtidos com a aplicação das técnicas e passos anteriormente citados e exemplificados pelo fluxo operacional, na Figura 4.25.

5.1 Segmentação da região da mama

Para verificar a generalização e comportamento do método implementado, 400 imagens mamográficas (200 de projeção CC e 200 de projeção MLO) foram submetidas ao processo de segmentação, descrito no item 4.2.1, com as etapas evidenciadas pela Figura 5.1. A separação da região da mama dos demais elementos presentes nas imagens foi avaliada pelas métricas mais recorrentes na literatura, descritas no item 4.6.

Figura 5.1 – Imagens resultantes das etapas para segmentação da região da mama.

a) imagem original; b) imagem com alargamento de contraste; c) imagem binarizada pelo limiar proposto por Pun (1981); d) inversão da intensidade das linhas cujos *pixels* em (c) possuam mesmos valores em, pelo menos, 90% da extensão total; e) imagem resultante após seleção de *pixels* por extração de componentes conectados; f) imagem final com a remoção de artefatos e melhoria de contraste proposta por Silva Jr. (2006).

A Tabela 5.1 demonstra os resultados obtidos com a avaliação pelas métricas de segmentação.

Tabela 5.1 – Resultados das métricas de avaliação da segmentação da região da mama

métrica imagem	AOM	AUM	AVM	CM	Acurácia	TxErr	Quality
CC	0,92	0,05	0,02	0,94	0,97	0,02	0,92
MLO	0,87	0,10	0,04	0,91	0,94	0,05	0,87
Média do conjunto completo	0,89	0,07	0,03	0,92	0,95	0,03	0,89

Cabe ressaltar que cada imagem mamográfica foi também segmentada manualmente, como demonstrado na Figura 5.2, por uma médica especialista em laudar imagens mamográficas, que atua em parceria com o grupo do LAPIMO; em seguida, foi comparada com a imagem resultante do processo de segmentação.

Figura 5.2 - Exemplificação das imagens utilizadas para avaliação pelas métricas:
a) imagem original; b) imagem segmentada computacionalmente; c) imagem segmentada pela especialista

Observando individualmente os resultados provenientes das métricas para cada imagem, é notório que muitas falhas na etapa de segmentação estão relacionadas com a grande diversidade de etiquetas e artefatos dispostos aleatoriamente na imagem. Também deve-se considerar a presença de ruídos que podem estar relacionados com o processo de aquisição, revelação e digitalização da imagem mamográfica (MENCATTINI et al., 2012). Em seu trabalho, Mencattini et al. (2012) aplicaram metodologias de segmentação da mama específicas para cada tipo de base de imagens, ou seja, dependendo das características das imagens, o processo de segmentação é diferente. Assim, é necessário conhecer as especificações das imagens de entrada, o que pode inviabilizar o método quando essa informação é desconhecida.

Contudo, a metodologia aplicada foi capaz de segmentar todas as imagens submetidas ao processo. Cabe lembrar que as imagens foram obtidas através de diferentes sistemas de aquisição: digitalizadores de filme e sistema CR. A Figura 5.3 apresenta uma sobreposição entre os contornos (*pixels* de borda da imagem binária)

– definido pela especialista e obtido pela técnica computacional automática desenvolvida.

Figura 5.3 – Sobreposição dos contornos obtidos na imagem original
a) marcação do especialista; b) contorno obtido com o método automático de segmentação

Apesar de algumas falhas na segmentação, o resultado obtido pelas métricas demonstram que, de modo geral, a região da mama foi selecionada corretamente na maioria das imagens. O resultado para o conjunto de imagens de projeção MLO é inferior quando comparado com o das imagens CC. Isto está intimamente relacionado ao fato de a área ocupada pela mama nas imagens MLO normalmente ser maior que nas imagens CC, o que acarreta uma grande proximidade de etiquetas e marcações à borda da mama. Porém, os valores médios e desvios-padrões obtidos de $0,89 \pm 0,12$, $0,92 \pm 0,11$, $0,95 \pm 0,05$ e $0,03 \pm 0,04$ para as medidas AOM, CM, Acurácia e taxa de erro, respectivamente, para a totalidade do conjunto de imagens, demonstram a eficácia do método proposto com baixo índice de erro, viabilizando a aplicação de técnicas subsequentes com menor probabilidade de falhas.

Assim, os resultados obtidos nesta etapa são satisfatórios, mostrando a eficiência do método para remoção automatizada do fundo, de etiquetas e demais artefatos estranhos ao tecido mamário na imagem mamográfica. Além disso, no intuito do desenvolvimento de um módulo completamente automático para o

esquema CADx, essa etapa reflete um importante processo na conclusão do objetivo dessa pesquisa, facilitando as demais etapas desse trabalho.

5.2 Delimitação do músculo peitoral

Através da pesquisa de diversos artigos científicos, um total de 10 técnicas foram implementadas no intuito de segmentar e marcar a área ocupada pelo músculo peitoral em imagens mamográficas de projeção MLO.

Nessa etapa foram utilizadas 100 imagens do BancoWeb, a fim de verificar o comportamento e resultados produzidos por cada técnica investigada sobre um amplo conjunto de amostras. As técnicas implementadas e testadas foram: transformada Hough (FERRARI et al., 2004), segmentações com limiar fixo (GARCIA, 2009), método de Otsu (1979), técnica proposta por Barbosa Filho et al. (2011), entropia proposta por Pun (1981), crescimento de região (NAGI et al., 2010), *k-means*, rede neural artificial *Self-Organizing Map* (SOM), algoritmo *cliff detection* (KWOK et al., 2004) e máxima entropia. A Figura 5.4 apresenta uma comparação dos resultados obtidos para uma mesma imagem de projeção MLO.

Alguns dos problemas observados nos resultados da aplicação de algumas técnicas são brevemente descritos abaixo:

- Transformada de Hough: o resultado final é dado por uma linha que se aproxima do músculo peitoral. No entanto, na maioria das imagens mamográficas, o músculo peitoral não possui forma triangular como descrito em muitos trabalhos, o que prejudica a sua detecção uma vez que áreas do tecido mamário são selecionadas erroneamente.
- Limiar de Otsu e entropia de Pun: os resultados não foram satisfatórios dada, principalmente, a grande variação de densidade do tecido mamário, que influencia diretamente na definição do limiar global automático de cada técnica, as quais são baseadas na distribuição de intensidade de histogramas.

Figura 5.4 – Diferentes técnicas implementadas para delimitação do músculo peitoral em imagens mamográficas MLO.

a) Transformada Hough; b) limiar de Otsu (1979); c) método proposto por Barbosa Filho et al. (2011); d) entropia de Pun (1981); e) crescimento de região; f) algoritmo Cliff Detection; g) k-means; h) SOM; i) máxima entropia.

- Método de Barbosa Filho et al. (2011): apesar de apresentar algumas falhas de aproximação dos limites do músculo peitoral, dentre as técnicas testadas foi a que apresentou resultados mais satisfatórios e rápido tempo computacional de execução.
- Crescimento de região: apesar de ser aplicada frequentemente pela literatura correlata, esta técnica é dependente de parametrizações, o que inviabiliza a sua implementação em sistemas automáticos. Mesmo fixando alguns parâmetros, a técnica apresenta resultados muito discrepantes entre um amplo conjunto de imagens.
- Algoritmo *cliff detection*: é uma técnica utilizada por trabalhos mais recentes, porém com algumas limitações. Para que a técnica seja executada até o fim, depende de uma detecção estimativa inicial. Em muitas imagens, a detecção inicial do músculo peitoral falhou, interrompendo o algoritmo e inviabilizando a técnica.
- *K-means*: apesar de apresentar resultados satisfatórios na maioria dos casos, em muitas imagens cujo músculo peitoral possui grande visibilidade (mamas de baixa densidade) a técnica resultou em uma detecção considerada falha.
- *Self-Organization Map* (SOM): apesar de ser uma técnica considerada artificialmente inteligente e automática, os melhores resultados foram obtidos quando submetidas apenas regiões de interesse cuja delimitação do músculo peitoral é bastante notória visualmente.
- Máxima entropia: os resultados obtidos, em sua totalidade, são pouco aceitáveis, sendo que as maiores falhas são percebidas em imagens cuja densidade do músculo peitoral se assemelha à do tecido mamário, resultando em uma segmentação completa da região da mama.

Por isso, nessa pesquisa, optou-se pela utilização do método proposto por Barbosa Filho et al. (2011). A Figura 5.5 exhibe seus passos, resultantes do processamento de imagens para delimitação do músculo peitoral.

Figura 5.5 – Processo final na delimitação do músculo peitoral:

a) imagem mamográfica de entrada; b) músculo peitoral segmentado a partir do limiar de Otsu (1979) obtido pela ROI; c) imagem segmentada sobreposta na imagem de entrada com a marcação do músculo peitoral.

Cada músculo peitoral das 100 imagens testadas também foi segmentado manualmente por uma especialista para estabelecimento do *ground-truth*, como evidenciado pela Figura 5.6.

Figura 5.6 - Imagens utilizadas para obtenção das métricas de avaliação
a) imagem original; b) imagem segmentada pelo método proposto por Barbosa Filho et al. (2011); c) imagem segmentada manualmente pelo especialista

Dessa forma, as métricas anteriormente descritas foram calculadas para avaliar algumas das segmentações testadas. A Tabela 5.2 demonstra os resultados obtidos com aplicação das métricas para avaliar as seguintes técnicas de segmentação: Cliff-detection, método proposto por Barbosa Filho et al. (2011) e *k-means* (MENECELLI & SCHIABEL, 2012).

Tabela 5.2 – Resultado das métricas de avaliação da delimitação do músculo peitoral

métrica segmentador	AOM	AUM	AVM	CM	Acurácia	TxEr	Quality
Cliff-Detection	0,41	0,42	0,68	0,54	0,86	0,13	0,41
<i>k-means</i>	0,70	0,20	0,11	0,79	0,95	0,04	0,70
Método de Barbosa Filho et al. (2011)	0,79	0,13	0,09	0,85	0,97	0,02	0,79

A Figura 5.7 apresenta a imagem original com a marcação sobreposta da delimitação do músculo peitoral feita manualmente pela especialista e obtida pela técnica de segmentação automática proposta por Barbosa Filho et al. (2011) e implementada nesse trabalho.

Figura 5.7 – Imagem original com sobreposição dos contornos de delimitação do músculo peitoral: a) imagem marcada manualmente pela especialista; b) imagem obtida a partir da binarização pela técnica de Barbosa Filho et al. (2011)

É notório que, apesar de fornecer os melhores resultados, conforme os valores obtidos com as métricas de avaliação, considerando todo o conjunto de imagens, o método de Barbosa Filho et al. (2011) ainda apresenta algumas falhas na área segmentada obtida em relação à região desejada, conforme visualizado pela Figura 5.4 (c). Isso reflete a necessidade da busca contínua por métodos de pré-processamento e técnicas automáticas e de maior precisão para o delineamento do músculo peitoral em imagens mamográficas. Contudo, pelos testes executados nessa pesquisa, tal técnica se revelou a mais adequada para lidar com essa questão, aliada à sua rápida execução computacional.

5.3 Classificação do mamilo

Os resultados provenientes dessa etapa são referentes aos procedimentos descritos no item 4.4.1. Inicialmente, foram utilizadas 200 imagens simuladas (100 representando o mamilo normal e 100, o mamilo retraído). A partir da análise dos gráficos da distribuição Gaussiana (v. item 4.3) para as imagens que simulam o mamilo, os atributos compacidade e relação entre área e perímetro foram submetidos como entrada para o classificador neural MLP.

A Figura 5.8 demonstra os gráficos da distribuição Gaussiana dos atributos selecionados.

Figura 5.8 – Gráficos da distribuição gaussiana para os atributos (a) compacidade e (b) relação entre área e perímetro

Os 10 conjuntos aleatórios de treinamento e validação foram compostos por 70 e 30 imagens, respectivamente. A classificação das imagens simuladas resultou em um acerto igual a 97,3% na validação cruzada, com 4 neurônios na camada intermediária. Também foram obtidos os valores de VP de 97,1%, FP de 2,9% e área sob a curva ROC (A_z) igual a 0,97, esta última demonstrada pela Figura 5.9.

Figura 5.9 – Curva ROC obtida a partir da melhor classificação da MLP para verificar a capacidade de generalização do classificador

Após comprovação da eficiência na classificação do mamilo através de imagens simuladas, foram submetidas 100 imagens mamográficas para um novo treinamento e validação pela rede neural MLP, sob as mesmas condições estabelecidas para as imagens que simulam o mamilo. As imagens mamográficas foram automaticamente segmentadas e a ROI do mamilo determinada. Em seguida, os atributos compacidade e relação entre área e perímetro foram extraídos de cada ROI e submetidos como entrada ao classificador neural MLP. Assim como esperado, os resultados obtidos com essa classificação foram também satisfatórios. Obteve-se um acerto de 95% com a validação cruzada, com 7 neurônios na única camada intermediária. Também foram obtidos um VP = 93%, FP = 7% e $Az = 0,93$, como demonstrado pela Figura 5.10.

Figura 5.10 - Curva ROC obtida com a classificação das imagens mamográficas para diferenciar mamilo normal e mamilo retraído

Ainda para constatar a capacidade de generalização do mesmo classificador, foi estabelecida uma configuração topológica para a rede neural com o melhor treinamento/validação obtido com as imagens mamográficas reais. Então, 52 imagens mamográficas reais (26 diagnosticadas com mamilo normal e 26 diagnosticadas com mamilo retraído) que não foram utilizadas nem nos treinamentos e nem nas validações, foram submetidas para a classificação. Logo, essas imagens mamográficas representam o conjunto de teste do classificador, sem a necessidade de um novo treinamento. A classificação dessas 52 ROIs das imagens mamográficas resultou em FP = 19,2%, FN = 0%, VP = 100% e VN = 80,7%, Az = 0,90, demonstrado pela curva ROC na Figura 5.11.

Figura 5.11 - Curva ROC resultante da classificação das imagens mamográficas para testar a generalização do classificador

Em virtude do baixo número de imagens mamográficas diagnosticadas com retração de mamilo e ausência de referências na literatura para investigação computacional dessa especificidade, houve a necessidade da utilização de imagens simuladas para aumentar e verificar a capacidade de generalização do classificador. Essas imagens que simulam o mamilo representam um universo conhecido e controlado (estabelecimento de um *ground-truth* adequado). Se a classificação dessas imagens não fosse viável, dificilmente as imagens mamográficas produziram resultados satisfatórios.

Erros na classificação das imagens mamográficas podem estar relacionados com os processos de segmentação e seleção automática da ROI, uma vez que as

imagens simuladas não são submetidas a tais processos. Além desses, a presença de ruídos (adquiridos em processos de revelação, digitalização ou exposição do filme mamográfico) ou erros dos profissionais técnicos que realizam o exame mamográfico (posicionamento da paciente e compressão da mama inadequados), também influenciam a imagem final obtida.

No entanto, o conjunto de resultados obtidos demonstra a eficiência do método proposto para diferenciar o mamilo normal do retraído em imagens mamográficas e é considerado satisfatório, uma vez que não foi encontrado na literatura o desenvolvimento de trabalhos com semelhantes objetivos. Desse modo, torna-se viável a implementação das técnicas em um sistema CADx a fim de fornecer maior informação em relatórios descritivos sobre achados mamográficos.

5.4 Detecção de linfonodos

A detecção de linfonodos é feita com apenas o músculo peitoral delimitado, como descrito anteriormente. Os resultados obtidos estão apresentados na Tabela 5.3:

Tabela 5.3 – Resultado obtido com a detecção de linfonodos

% média de acerto (VP)	média de FP por imagem	média de FN por imagem
74,9	3,5	0,9

De acordo com a Tabela 5.3, o valor de FP por imagem indica que cerca de 3 a 4 pontos são marcados erroneamente (indicam a presença de linfonodos) em cada imagem; FN por imagem é o número médio de linfonodos que deixam de ser detectados em cada imagem, sendo que nesse caso, cerca de 1 linfonodo não é detectado.

Com o resultado médio de acerto (porcentagem de VP) pode-se notar que ainda é necessária uma melhoria no método de detecção de linfonodos, fato esse também constatado por especialistas em diagnosticar exames mamográficos. A Figura 5.12 demonstra uma parcela do conjunto de imagens com os linfonodos detectados.

Figura 5.12 – Demonstração dos resultados para detecção de linfonodos no músculo peitoral de images MLO

Em testes preliminares notou-se que o aumento do número médio de FP por imagem (maior tolerância a ruídos) contribui para um maior acerto, porém não foi considerado viável uma vez que muitas regiões passam a receber uma análise detalhada que, ao final, comprova-se desnecessária. Também foram aplicadas técnicas de melhoria de imagens como filtragem espacial, porém nenhuma delas mostrou-se capaz de contribuir com o acerto.

5.5 Classificação de densidades em regiões da mama

Descrevem-se aqui os resultados obtidos na classificação automática das regiões da mama, conforme a variação de intensidade. Esses resultados também são comparados nesta seção àqueles descritos no trabalho de Wu, Besnehard e Marchessoux (2011) que tratava diretamente disso. Esta etapa também estará intimamente relacionada com a classificação de densidade mamográfica e detecção de assimetrias teciduais.

Após extração de 80 atributos de cada ROI, analisou-se a distribuição Gaussiana de cada um deles. A partir dessa análise, os atributos de autocorrelação e variância foram selecionados e, portanto, empregados como parâmetros de entrada para o classificador.

A Figura 5.13 demonstra os gráficos da distribuição Gaussiana dos atributos selecionados.

Figura 5.13 - Gráficos da distribuição gaussiana para os atributos (a) autocorrelação e (b) variância

Os 10 conjuntos aleatórios de treinamento e validação foram compostos por 350 e 150 ROIs, respectivamente. A classificação das ROIs resultou em um acerto igual a 94,4% na validação cruzada, com 6 neurônios na camada intermediária. Os valores estatísticos obtidos com a classificação estão na Tabela 5.4.

Tabela 5.4 – Resultados obtidos com a classificação das ROIs da mama

	VP (%)	VN (%)	FP (%)	FN (%)
Classe 1	97,60	74,80	25,20	2,40
Classe 2	96,00	72,20	27,80	4,00
Classe 3	84,80	74,20	25,80	15,20
Classe 4	100,00	73,40	26,60	0,00
Média	94,60	73,65	26,35	5,40

Um número maior de ROIs poderia ser submetido à classificação; porém, diferentemente de Wu, Besnehard e Marchessoux (2011), optou-se por manter proporções iguais entre o número de imagens de cada classe.

A dificuldade no estabelecimento de um limiar visual para classificação manual (estabelecimento do *ground-truth*) entre ROIs de densidades fibroglandular e heterogeneamente densa – categorias 2 e 3 para densidades do BI-RADS® – pode

contribuir com erros na classificação automática pela rede neural MLP. Também a presença de ruídos de qualquer natureza na imagem mamográfica é capaz de afetar diretamente o processo de classificação automática, assim como a aplicação de filtros, uma vez que os descritores de textura aqui selecionados são obtidos através da variação de intensidade do histograma da imagem. Dessa forma, mesmo considerando a experiência do radiologista, o estabelecimento do *ground-truth* está condicionado à subjetividade de interpretação e visualização da imagem mamográfica.

Apesar da média de FP = 26,35 ser considerada alta, esse valor e o de FN são melhores que os resultados obtidos por Wu, Besnehard e Marchessoux (2011). Já os valores de VP e VN são maiores que os daquele trabalho (WU et al., 2011), o que reflete uma maior precisão do classificador das ROIs. A Figura 5.14 exibe os resultados estatísticos alcançados no trabalho de Wu, Besnehard e Marchessoux (2011) para efeito de comparação.

Category	MCA framework	
	Classification rates (%)	Error rates (%)
1. Almost entirely fatty	TP 96.67 - TN 83.61	FP 16.39 - FN 3.33
2. Scattered fibro glandular tissue	TP 86.42 - TN 66.05	FP 33.95 - FN 13.58
3. Heterogeneously dense tissue	TP 82.35 - TN 64.06	FP 35.94 - FN 17.65
4. Extremely dense tissue	TP 96.08 - TN 57.81	FP 42.19 - FN 3.9
Average	TP 90.38 - TN 67.88	FP 32.12 - FN 9.62

Figura 5.14 – Resultados obtidos com a classificação de regiões da mama pelo trabalho de Wu, Besnehard e Marchessoux (2011)

A opção em classificar cada ROI manualmente feita por uma especialista aprimorou o processo de forma geral. A aplicação desenvolvida para classificação manual das ROIs demonstrou-se eficiente, ágil e de fácil usabilidade.

A classificação de subregiões da mama contribui para estabelecimento de um mapeamento com maior detalhamento da sua densidade. Assim, áreas adjacentes a uma determinada estrutura podem auxiliar na identificação do tipo de tecido em que a mesma estrutura está inserida. Também pode contribuir na detecção de assimetrias focais, quando são comparadas imagens bilateralmente.

Adicionalmente, a automatização de todo o processo viabiliza a implementação das técnicas em esquemas CAD, contribuindo para a eficiência do processo e evitando falhas de manipulação humana.

Com os resultados obtidos é possível a viabilização do mesmo método para: (a) detecção de assimetria tecidual; (b) estimativa de densidade mamográfica, conforme categorias estabelecidas pelo BI-RADS®; e (c) análise percentual da densidade de regiões adjacentes a uma estrutura detectada. Além disso, com ROIs de tamanho menor, acredita-se na capacidade de detecção de regiões candidatas à presença de nódulos.

5.6 Classificação da densidade mamográfica, conforme BI-RADS®

A classificação final da densidade média global da mama tem por base os resultados obtidos com a classificação de cada ROI que representa parcela da mama, cujos procedimentos são descritos no item 4.2.5 e os respectivos resultados no item 5.5.

Nessa etapa, foram utilizadas 400 imagens destinadas a testes de classificação, sendo 100 para cada uma das 4 categorias de padrões de densidade estabelecidas pelo BI-RADS®.

Os resultados obtidos com a classificação da densidade mamográfica são divididos através de duas análises: a primeira em relação às 4 classes estabelecidas pelo BI-RADS® e a segunda baseada em alguns trabalhos que consideram apenas 2 classes (categorias 1 e 2 do BI-RADS® como a classe de baixa densidade e as categorias 3 e 4 como a classe de alta densidade). A seguir, são mostrados os resultados obtidos em cada tipo de classificação.

5.6.1 Classificação conforme BI-RADS® em 4 categorias

Com base no método de Liu et al. (2011), foi obtida a média percentual para cada uma das classes, sendo considerada vencedora aquela que apresentou o

maior número de ROIs classificadas. A Tabela 5.5 mostra o percentual de acerto para as imagens submetidas ao *software* final de classificação.

Tabela 5.5 - Resultados obtidos com a classificação das densidades em 4 classes, conforme BI-RADS®

	VP (%)	VN (%)	FP (%)	FN (%)
Classe 1	80,00	74,75	25,25	20,00
Classe 2	99,00	68,25	31,75	1,00
Classe 3	58,00	73,50	26,50	42,00
Classe 4	96,00	66,75	33,25	4,00
Média	83,25	70,81	29,19	16,75

Essa classificação obteve uma elevada porcentagem de acerto. Porém, a porcentagem de casos FP também é considerada elevada. Analisando os resultados dessa classificação, verificamos que o alto índice de FP é devido, principalmente, a erros de classificação da classe 3. Do total de 100 imagens da classe 3, apenas 58 foram corretamente classificadas. A decorrência de tal resultado reflete na dificuldade do especialista em separar corretamente imagens da classe 3 que, em muitos casos, são erroneamente classificadas nas categorias 2 ou 4. Na prática, falhas como essas são características também na determinação da classificação final da densidade mamográfica, conforme relatos de especialistas. Com isso, possíveis erros cometidos nessa etapa inicial influenciam diretamente na capacidade de generalização do classificador final. Por outro lado, as classes 2 e 4 contribuíram para os altos índices de VP. Da classe 2, apenas 1 imagem foi classificada erroneamente e na classe 4, apenas 4 imagens foram categorizadas em outras classes.

5.6.2 Classificação em 2 classes: baixa densidade e alta densidade

Para verificar a taxa de acerto em relação aos principais trabalhos da literatura, também avaliamos a capacidade do classificador em diferenciar imagens com baixa densidade (até 50% de tecido denso) daquelas com alta densidade (maior que 50% de tecido denso).

Para isso, as mesmas imagens foram submetidas ao classificador final, porém com agrupamento de resultados: categoria 1 classificada na categoria 2 consideramos como acerto e vice-e-versa, assim como a categoria 3 classificada na categoria 4 e vice-e-versa.

Os resultados obtidos com essa classificação foram: acerto total de 93,5%, com TP = 92%, VN = 95%, FP = 10%, FN = 16% e Az = 0,95. A curva ROC obtida a partir desse resultado pode ser visualizada pela Figura 5.15.

Figura 5.15 - Curva ROC para a classificação de densidade em 2 classes: baixa densidade e alta densidade.

A classificação mais simples em 2 tipos de classes comprova a capacidade do *software* em diferenciar tecidos de alta e baixa densidades, o que é um resultado bastante satisfatório, principalmente se compararmos com as principais propostas já apresentadas na literatura correlata, resumidamente evidenciada pela Tabela 5.6.

Tabela 5.6 – Revisão bibliográfica dos principais métodos de classificação de densidade mamográfica

Autor	Método	Resultado
Liu et al. (2011)	combinação de análise multiescala com momentos do histograma, a partir de regiões da mama. Utilizou classificador SVM.	86,4%
Petroudi & Brady (2011)	análise de textura para classificação da densidade mamográfica completa.	82%

Chen, Denton e Zwiggelaar (2011)	utilização dos atributos Local Graylevel Appearances (LGA) e Texton II.	91%
Elshinawy et al. (2011)	classificação da densidade mamográfica em 2 categorias: lipossustituída e densa. Fez uso de atributos de textura e classificador SVM.	97,3%
Wu et al. (2011)	classificação automática de ROIs em imagens mamográficas através da extração de atributos de textura.	a média obtida foi: TP=90,38%, TN=67,88%, FP=32,12% e FN=9,62%
Tzikopoulos et al. (2009)	estimação da densidade mamográfica em 3 classes, utilizando diferentes métodos de segmentação.	acerto médio: 67,66%
Subashini et al. (2010)	extração de descritores de textura e uso do classificador SVM para classificar a densidade mamográfica em 3 categorias	95,44%

5.7 Detecção de assimetrias

Foram submetidas ao *software* de detecção de assimetrias um total de 300 casos com 4 imagens cada um de 2 projeções (CC e MLO), totalizando 1200 imagens. As imagens de projeção MLO, primeiramente, foram submetidas à técnica de delimitação do músculo peitoral e, conseqüentemente, o mesmo foi removido da imagem.

A taxa de acerto obtida foi de 84,73% com um FP de 0,39 por imagem. Os resultados da detecção foram comparados com as demarcações obtidas a partir da análise visual de uma radiologista especialista. A Figura 5.16 mostra o par de imagens de entrada e a demarcação das áreas assimétricas, para um mesmo caso.

Figura 5.16 – Detecção de assimetria em um par de imagens mamográficas.

a) e b) imagens originais de entrada; c) e d) imagens com regiões assimétricas demarcadas.

Apesar de todo o processo ser automatizado com uso de técnicas de baixo custo computacional, o tamanho da imagem de entrada influencia diretamente no tempo de execução do *software*. Cabe lembrar também que todo o processo se repete 4 vezes (uma vez para cada imagem de entrada). Os testes foram feitos em computadores com diferentes configurações, sendo especificadas na Tabela 5.7:

Tabela 5.7 – Tempo médio de processamento do módulo CADx em diferentes computadores

Computador	Configuração	Tempo médio de processamento (minutos)*
Notebook Acer	Processador: Intel centrino Core 2 Duo, 2.0GHz. RAM: 2 GB DDR2 HD: Sata 7400 rpm	8
Notebook Sony Vaio	Processador: Intel Core i5 2 ^a geração, 2.4GHz. RAM: 6GB DDR3 HD: Sata 7400 rpm	5
Mackbook Pro	Processador: Intel Core i5 2 ^a geração, 2.5GHz. RAM: 4GB DDR3 HD: Sata 7400 rpm	4
PC Desktop	Processador: Intel Core i3, 3.10 GHz. RAM: 4GB DDR2 HD: Sata 7400 rpm	7

* os valores dos minutos estão aproximados e referem-se a apenas 1 caso a ser investigado (4 imagens).

5.8 Módulo desenvolvido para o esquema CADx

Ao final, como objetivo geral dessa pesquisa, foi implementado um *software* que reúne todas as técnicas acima mencionadas. Essa aplicação é responsável em caracterizar a mama conforme os possíveis aspectos secundários do câncer de mama elencados por este trabalho.

O módulo desenvolvido é classificado em CADe e CADx pois, ao mesmo tempo que realiza detecção de assimetrias e linfonodos (CADe), também classifica o mamilo e a densidade predominante da mama (CADx).

Uma médica mastologista com ampla especialidade em laudar imagens mamográficas testou o *software* desenvolvido e o descreveu como eficaz, de simples utilização e com uma interface intuitiva. O usuário deve apenas efetuar o carregamento de cada imagem do exame mamográfico e pressionar o botão processar, como demonstrado pela Figura 5.17 e Figura 5.18. Todo o restante do processamento é efetuado de forma automática.

Figura 5.17 – Tela inicial do módulo CAD desenvolvido

Figura 5.18 – Imagens de um exame mamográfico em processamento no módulo CAD desenvolvido

Quanto ao desempenho computacional (tempo de resposta) a mesma especialista acredita que poderia ser um pouco mais rápida a execução e exibição de resultados finais. Como sugestão de ferramenta a ser adicionada ao módulo CADx, segundo o parecer dos profissionais da saúde, é a possibilidade de impressão de relatórios descritivos da imagem assim como as próprias imagens com as demarcações efetuadas.

No campo de relatório é possível visualizar o percentual estatístico para a classificação do mamilo e a taxa do tipo de assimetria encontrada na imagem. A detecção de assimetrias pode ser visualizada pelas abas correspondentes a cada projeção de imagem, como visualizado pela Figura 5.19.

Figura 5.19 – Resultado da detecção de assimetrias efetuada pelo módulo.
a) Imagens CC; b) Imagens MLO

6 - CONCLUSÕES

6.1 Fundamentação teórica

Conforme pode ser observado com a pesquisa bibliográfica, é de extrema importância a atenção que deve ser dada aos aspectos secundários de caracterização do câncer de mama. No entanto, por comumente retratar apenas uma variação da normalidade, a esses aspectos não se costuma dar a atenção adequada nas pesquisas que objetivam auxiliar a detecção/classificação computacional desses achados.

Na presente pesquisa, além da importância individual dada a cada aspecto secundário, também houve preocupação quanto à utilização de técnicas de baixo custo computacional, eficazes e recorrentes na literatura para que o *software* desenvolvido seja não só eficiente, mas também de fácil usabilidade. Nesse sentido, também, desde o início do desenvolvimento da pesquisa, houve preocupação em utilizar e automatizar todas as técnicas, evitando falha de manipulação humana na utilização do *software*.

Inúmeras técnicas foram analisadas a fim de prover resultados compatíveis e superiores àqueles presentes na literatura correlata, com a reutilização de códigos

computacionais, como o caso da MLP, para garantir maior aprofundamento, focalizando em cada aspecto estudado.

6.2 Metodologia aplicada

Além das técnicas descritas nesse trabalho, muitas outras foram implementadas e submetidas a testes para verificação e validação dos resultados de cada uma das etapas descritas. Dessa forma, foi grande a preocupação em comparar e selecionar cada método conforme o melhor resultado obtido e sua respectiva aplicação na literatura correlata, além da utilização de métricas e modelos estatísticos de seleção.

Dos trabalhos selecionados na literatura buscou-se implementar as características marcantes e de viabilização comprovada. Assim, foram reunidas aqui as técnicas de maior relevância de cada trabalho, cooperando entre si para a execução do módulo CAD implementado. Um diferencial – pouco utilizado pelos trabalhos aqui referenciados – foi a adição de técnicas inteligentes a fim de buscar uma aproximação mais condizente com a avaliação visual humana.

As imagens que simulam o mamilo segmentado refletem um universo controlável e com respostas plenamente conhecidas, facilitando e contribuindo para a generalização da classificação por redes neurais.

Os processos iniciais que antecedem a detecção e classificação propriamente ditas são fundamentais para a automatização proposta para o módulo CAD desenvolvido. Assim, foi grande a preocupação para que a segmentação da mama, delimitação do músculo peitoral, localização da região do mamilo e subdivisão automática da mama em regiões de interesse fossem executadas para fornecer o melhor resultado possível em fidelidade à avaliação visual humana.

Métodos de classificação, seleção de atributos e mensuração de resultados foram escolhidos conforme a possibilidade de automatização, maior recorrência na literatura e utilização prévia em trabalhos anteriores. Foram capazes de demonstrar eficiência computacional, principalmente pela reutilização da rede neural MLP em diversas etapas.

Também, após o completo desenvolvimento da pesquisa, é perceptível a padronização de processos com a descrição das técnicas implementadas, entre elas:

- Limiarização global: diferentes formas de obtenção de um limiar foram obtidas para o mesmo intuito – segmentar a imagem mamográfica em regiões. As principais etapas de utilização dessas técnicas foram a segmentação da mama e a delimitação do músculo peitoral;
- Métodos de avaliação: utilizados para verificar a eficiência da segmentação da mama e do músculo peitoral;
- Seleção de atributos: a distribuição Gaussiana foi utilizada para verificar a existência de alguma padronização definida pelos atributos extraídos nos processos de: retração/normalidade do mamilo; subregiões conforme a densidade predominante; classificação da densidade mamográfica, conforme categorias definidas pelo BI-RADS®; e capacidade de diferenciar regiões assimétricas na mama;

Classificação pela rede neural MLP: a mesma rede neural com diferentes configurações (alterações nas entradas, pesos sinápticos e números de neurônios) foi utilizada para classificar o mamilo (em normal ou retraído) e subregiões da mama, para estimar a categoria predominante da densidade mamográfica e para eliminar regiões FP na detecção de assimetrias.

6.3 Resultados alcançados

Procurou-se ampliar a eficiência e precisão dos métodos através de constante avaliação comparativa, desde as etapas iniciais de pré-processamento, cujas falhas sempre acabam refletindo diretamente nos resultados das etapas seguintes. Além disso, a presença de médicos especialistas com larga experiência em laudar imagens mamográficas ajudou muito o desenvolvimento e avaliação dos métodos implementados.

Cabe ressaltar também a variedade nas características das imagens utilizadas nesse trabalho. Apesar de obtidas da mesma base, as imagens foram adquiridas através de diferentes aparelhos mamográficos e digitalizadores. Em todas

as imagens, o *software* final produzido apresentou resultados compatíveis e, em alguns casos, superiores a literatura correlata, contribuindo com o alcance do objetivo porposto por esse trabalho. Imagens de outras bases também foram submetidas a testes do módulo CAD desenvolvido, mas não foi possível avaliar os resultados principalmente pela diferença na resolução de contraste: é fundamental que a imagem possua 12 bits de resolução de contraste, em função dos atributos utilizados para o treinamento da rede neural MLP. Também foram perceptíveis algumas diferenças nas características das imagens obtidas pelo mamógrafo com sistema CR. Apesar dos 12 bits de resolução de contraste, a variação da faixa de valores obtidos a partir da extração de atributos foi bastante diferente em relação às imagens digitalizadas. Isso levou à necessidade de novos treinamentos e adaptabilidade das técnicas do módulo CAD desenvolvido a fim de buscar a maior generalização possível.

A utilização de técnicas mais comuns de processamento de imagens, automáticas e de simples processamento, contribuiu para a obtenção de resultados de forma mais rápida. A viabilidade da sua implementação no *software* desenvolvido foi constatada considerando a necessidade de agilidade no tempo de execução do mesmo. Praticamente em todas as etapas foram realizadas avaliações quantitativas para firmar a competência dos processos empregados nas imagens mamográficas. Alguns métodos já desenvolvidos e consolidados na literatura foram também implementados nessa pesquisa, entre eles: detecção automática do músculo peitoral (BARBOSA FILHO et al., 2011), demarcação da região de localização do mamilo (MUSTRA et al., 2009), princípio da classificação da densidade mamográfica (WU et al., 2011), *plugin* do ImageJ para detecção de linfonodos.

Entre as principais contribuições do trabalho está um procedimento computacional desenvolvido em JAVA, totalmente disponível para determinar, como um módulo de um esquema CADx em mamografia, um conjunto de aspectos complementares importantes para o auxílio ao diagnóstico, qual sejam: a segmentação da região da mama com exclusão de etiquetas, marcações e demais artefatos; a classificação da retração/normalidade do mamilo; a comparação bilateral de regiões da mama pela análise valorada de atributos; a confirmação de um limiar percentual para estabelecimento da diferença de regiões por características assimétricas e subsequente classificação das assimetrias detectadas, conforme o BI-RADS®; a detecção de linfonodos axilares em imagens de projeção MLO.

Desta forma, o módulo desenvolvido é capaz de demarcar regiões com diferenças assimétricas teciduais em pares de imagens bilaterais e efetuar a classificação conforme o BI-RADS®, alertar para a presença da retração de mamilo e marcação de linfonodos axilares. Assim, os principais aspectos secundários diagnosticados pelo exame mamográfico são explorados pelo *software* a fim de auxiliar o radiologista para um diagnóstico mais preciso e eficiente.

Desde o início dessa pesquisa, a maior automatização dos métodos foi uma prioridade. Por isso foi desenvolvida uma aplicação de simples manuseio, com pouca obrigatoriedade de interação do usuário e, ao mesmo tempo, com alta funcionalidade para a apresentação detalhada dos achados secundários e notações percentuais referentes aos mesmos.

6.4 Conclusão

É reconhecida a necessidade de meios auxiliares ao diagnóstico do câncer de mama no auxílio, principalmente, da interpretação do caso com vistas a uma maior precisão dos laudos descritivos.

Esse trabalho cumpriu com o objetivo proposto através da implementação de um módulo CAD direcionado à investigação de aspectos secundários ao diagnóstico do câncer de mama. Esse módulo é baseado em técnicas computacionais automáticas, de diferentes áreas, com um baixo custo computacional de execução e confiabilidade de resultados produzidos. Além disso, possui uma interface intuitiva, rápida e a visualização de dados é facilitada para despertar o interesse de utilização por profissionais, estudantes e demais especialistas que manipulam imagens mamográficas em geral.

A associação de diferentes técnicas ocorre desde o processo de segmentação da imagem até o estudo individualizado de ROIs, baseando-se em métodos de diferentes trabalhos da literatura corrente. A manipulação de contraste, proposto por Silva Jr. (2006) e utilizado após a segmentação, contribui para a melhoria no processamento do módulo CADx em geral, principalmente na detecção do músculo peitoral, evidenciando, junto da automatização de métodos – como a

detecção de linfonodos axilares – um aprimoramento para das técnicas implementadas.

Por fim, as avaliações e resultados evidenciam a originalidade e viabilidade da utilização de módulos CADx para caracterização de aspectos secundários em imagens mamográficas. Tais módulos possibilitam o auxílio a todos os profissionais envolvidos nas diferentes etapas de realização do exame mamográfico, favorecendo, ainda, o aumento na precisão de diagnósticos.

6.5 Proposta para trabalhos futuros

Em trabalhos futuros, a partir da metodologia aplicada por essa pesquisa, é de interesse:

- aumentar o conjunto de imagens com assimetrias detectadas e com retração de mamilo para reavaliação do módulo CAD desenvolvido;
- melhorar e aprimorar a detecção do músculo peitoral;
- investigação de outros achados secundários como: retração de pele, detecção de assimetrias em desenvolvimento e distorções de arquitetura;
- adaptação de métodos para detecção automática de nódulos;
- investigação detalhada de tecidos adjacentes à uma estrutura detectada.

Dessa forma, o detalhamento de relatórios descritivos de imagens mamográficas digitais é aprimorado.

REFERÊNCIAS BIBLIOGRÁFICAS

ABRAMOFF, M. D.; MAGALHAES, P. J.; RAM, S. J. Image Processing with ImageJ. **Biophotonics International**, v. 11, n. 7, p. 36-42, 2004.

ALOLFE, M. A.; YOUSSEF, A. B. M.; KADAH, Y. M.; MOHAMED, A. S. Development of a computer-aided classification system for cancer detection from digital mammograms. **National Radio Science Conference – NRSC**, Egito, p. 1-8, 18-20 mar. 2008.

ALVARES, B. R.; ALMEIDA, O. J. Contribuições das incidências mamográficas complementares na investigação do câncer mamário. **Mastologia**, v.19, n. 2, p. 69-75, 30 jun. 2009.

ANGELO, M. F.. **Sistema de processamento de imagens mamográficas e auxílio ao diagnóstico via-internet**. 2007. 126f. Tese de Doutorado, Escola de Engenharia de São Carlos, Universidade de São Paulo - USP, São Carlos.

AZEVEDO-MARQUES, Paulo Mazzoncini de. Diagnóstico auxiliado por computador na radiologia. **Radiol Bras**, São Paulo, v. 34, n. 5, p. 285-293, set./out. 2001. Disponível em: <www.scielo.br/pdf/rb/v34n5/7682.pdf> Acessado em: 05 maio 2009.

BANDYOPADHYAY, S. K.; MAITRA, I. K. Anatomical asymmetry of human breast for indicator of breast cancer. **International Journal of Computer Applications**, v. 9, n. 2, nov. 2010.

BARBOSA FILHO, J. R. B.; BATISTA, L. V.; SCHIABEL, H.; VERÇOSA, L. B. Segmentação do músculo peitoral em imagens mamográficas através do recorte automático de ROI e da binarização de Otsu. VII Workshop de Visão Computacional, 2011, Curitiba, PR. **Anais do VII WVC**, p. 190-195, 2011.

BARBU, A.; MICHAEL, S.; XUN, X.; LIU, D.; ZHOU, S. K.; COMANICIU, D. Automatic detection and segmentation of axillary lymph nodes. **Medical Image Computing and Computer-Assisted intervention**, v. 13, n. 1, p. 28-36, 2010.

BOAVENTURA, I. A. G. **Números Fuzzy em procesamento de imagens digitais e suas aplicações na detecção de bordas**, 2010, 120f. Tese de Doutorado – Escola de Engenharia Elétrica de São Carlos, Universidade de São Paulo, São Carlos, 2010.

BOFF, Ricardo Antônio; RUARO, Simone; SCHUH, Fernando. Abordagem diagnóstica das massas palpáveis da mama. In: BOFF, Ricardo Antônio; WISINTAINER, Francisco. **Mastologia Moderna: abordagem multidisciplinar**. Caxias do Sul: Mesa Redonda, 2006. p. 47-52.

BONTRAGER, K. L. **Tratado de Técnica Radiológica e Base Anatômica**. 5 ed. Rio de Janeiro: Guanabara Koogan, 2003.

BOSCH, A.; MUÑOZ, X.; OLIVER, A.; MARTÍ, J. Modeling and Classifying Breast Tissue Density in Mammograms. **IEEE Computer Society Conference on Computer Vision and Pattern Recognition – CVPR’2006**, v.2, p. 1552-1558, 2006.

BOZEK, J.; MUSTRA, M.; GRGIC, M. A survey of mammographic image processing algorithms for bilateral asymmetry detection. In: **International Symposium ELMAR – ELMAR’2009**, Zadar, Croácia, p. 9-14, set. 2009.

BRAGA, A. P.; CARVALHO, A. C. P. L. F.; LURDERMIR, T. B. **Redes neurais artificiais: teoria e aplicações**. Rio de Janeiro: LTC, 2000.

BRANCATO, B.; HOUSSAMI, N.; FRANCESCA, D.; BIANCHI, S.; RISSO, G.; CATARZI, S.; TASCHINI, R.; DEL TURCO, M. R.; CIATTO, S. Does computer-aided detection (CAD) contribute to the performance of digital mammography in a self-referred population?. **Breast Cancer Res Treat**, v.111, p. 373-376, 2008.

BRUNISIDE, E. S.; SICKLES, E. A.; BASSET, L. W.; RUBIN, D. L.; LEE, C. H.; IKEDA, D. M.; MENDELSON, E. B.; WILCOX, P. A.; BUTLER, P. F.; D'ORSI, C. J. The ACR BI-RADS® experience: learning from history. **Journal of the American College Radiology**, v. 6, n. 12, 2009.

CALDAS, F. A. T.; ISA, H. L. V. R.; TRIPPIA, A. C.; BÍSCARO, A. C. F. P. J.; SOUZA, E. C. C.; TAJARA, L. M. Controle de qualidade e artefatos em mamografia. **Radiol Bras**, v. 38, n. 4, p. 295-300, 2005.

CARLI, A. C. Diagnóstico por imagem em doenças da mama. In: BOFF, R. A.; WISINTAINER, F. **Mastologia Moderna**: abordagem multidisciplinar. Caxias do Sul: Mesa Redonda, 2006. p. 53-80.

CASTELLA, C.; KINKEL, K.; ECKSTEIN, M.; SOTTAS, P.; VERDUN, F.; BOCHUD, F. Semiautomatic mammographic parenchymal patterns classification using multiple statistical features. **Academic Radiology**, v. 14, n.12, p. 1486-1499, 2007.

CHANDRA, J. S.; KARNAN, M.; SIVAKUMAR, R. Detection of masses in digital mammograms. **International Journal of Computer and Network Security (IJCNS)**, v. 2, n. 2, fev. 2010.

CHENG, H. D.; SHI, X. J.; MIN, R.; HU, L. M.; CAI, X. P.; DU, H. N. Approaches for automated detection and classification of masses in mammograms. **Pattern Recognition**, v. 39, n. 4, p. 646-668, 2006.

CLARKE, M.; COLLINS, R.; DAVIES, C.; GODWIN, J.; GRAY, R.; PETO, R. Tamoxifen for early breast cancer: an overview of the randomised trials. **The Lancet**, v. 351, n. 9114, p. 1451-1467, maio 1998.

CURRY III, T. S.; DOWDEY, J. E.; MURRY, R. C. **Christensen's Physics of Diagnostic Radiology**. 4 ed. Philadelphia (USA): Lea & Febiger, 1990.

DOI, K. Computer-aided diagnosis in medical imaging: historical review, current status and future potential. **Computerized Medical Imaging and Graphics**, n. 31, p.198-211, 2007.

EISENBERG, A. L. A.; KOIFMAN, S.; CORDEIRO DE REZENDE, L. M. M. Predictive factors to lymph node involvement on breast cancer. **Revista Brasileira de Cancerologia**, v. 47, n. 4, p. 389-396, out.-dez., 2001.

ELTER, M.; HELD, C.; An improved method for segmentation of mammographic masses. Medical Imaging 2010: Computer-Aided Diagnosis. **Proc. of SPIE 2010**, San Diego (CA), USA, v. 7624, p. 76242I-1—76242I-8, fev. 2010.

ELTONSY, N. H.; ELMAGHRABY, A. S.; TOURASSI, G. D.; Bilateral breast volume asymmetry in screening mammograms as a potential marker of breast cancer: preliminary experience. **IEEE Inter. Conf. on Image Processing – ICIP 2007**, v. 5, p. V-5 - V-8, 2007.

ELTOUKHY, M. M.; FAYE, I.; SAMIR, B. B. Automatic detection of breast masses in digital mammograms using pattern matching. IEEE EMBS Conf. on Biomedical Engineering & Sciences 2010, 2010, Kuala Lumpur, Malaysia. **Proc. of IECBES 2010**, p. 87-92, nov/dec, 2010.

FAUSETT, L.. **Fundamentals of neural networks**: architectures, algorithms and applications. USA: Prentice Hall, 1994.

FAWCET, T. An introduction to roc analysis. **Patter Recognition Letters**, v. 27, p. 861-874, jun. 2006.

FENELON, Sandro. Aspectos éticos-legais em imaginologia. **Radiol Bras**, São Paulo, v. 36, n. 1, p. III-VI, jan./fev. 2003.

FERRARI, R. J.; RANGAYYAN, R. M.; DESAULTELS, J. E. L.; FRÈRE, A. F. Analysis of asymmetry in mammograms via directional filtering with Gabor wavelets. **IEEE Trans. on Medical Imaging**, v. 20, n. 9, p. 953-964, set. 2001.

FERRARI, R. J.; RANGAYYAN, R. M.; DESAULTELS, J. E. L.; BORGES, R. A.; FRÈRE, A. F. Automatic identification of the pectoral muscle in mammograms. **IEEE Trans. on Medical Imaging**, v. 23, n. 2, p. 232-246, fev. 2004.

FERNANDES, A. M. R. **Inteligência artificial**: noções gerais. Florianópolis, Visual Books, 2005.

FREITAS, A. G.; KEMP, C.; LOUVEIRA, M. H.; FUJIWARA, S. M.; CAMPOS, L. F. Mamografia Digital: perspectiva atual e aplicações futuras. **Radiol Bras**, São Paulo, v. 39, n. 4, p. 287-296, jul./ago. 2006.

GARCIA, P. S. M. **Deteccção da assimetria mamária, classificação da densidade e validação dos resultados na web**, 2009, 98 f. Dissertação de Mestrado, Universidade de Mogi das Cruzes, Mogi das Cruzes, 2009.

GOES, C. E.; SCHIABEL, H.; NUNES, F. L. S. Evaluation of Microcalcifications Segmentation Techniques for Dense Breast Digitized Images. **Journal of Digital Imaging**, v. 15, p. 231-233, 2002.

GONZALEZ, R. C.; WOODS, R. E. **Digital image processing**. 3 ed. New Jersey: Prentice Hall, 2008.

GONZALEZ, R. C.; WOODS, R. E. **Processamento digital de imagens**. 3. ed. São Paulo: Pearson, 2010.

GONZALEZ, R. C.; WOODS, R. E.; EDDINS, S. L. **Digital image processing using Matlab**. 2nd ed. [S.l]: Gatesmark, 2009.

GUIMARÃES, L. T. G. **Subtração digital como ferramenta para detecção de tumores em imagens mamográficas de mamas densas: uma abordagem utilizando simulação computacional**, 2009, 119 f. Dissertação de Mestrado - Escola de Engenharia de São Carlos, Universidade de São Paulo, São Carlos, 2009.

HARALICK, R. M.; SHANMUGAM K.; DISTEIN, I. Textural Features for image classification. **IEEE Trans. on Systems, Man, and Cybernetics – SMC**, v. 3, n. 6, p. 610-621, 1973.

HAYKIN, S. **Neural networks and learning machines**. 3 ed. New Jersey: Prentice Hall, 2008.

HAYWANG-KÖBRUNNER, S. H.; SCHREER, I.; DERSHAW, D. D.; FRASSON, A. **Mama: Diagnóstico por imagem**. Rio de Janeiro: Revinter, 1999.

HU, M. K. Visual pattern recognition by moment invariants. **IRE Trans. on Inf. Theory**, v. 8, n. 2, p. 179-187, fev 1962.

INSTITUTO NACIONAL DE CÂNCER (INCA). Ministério da Saúde. **Estimativa 2012: incidência de câncer no Brasil**. Rio de Janeiro: Inca, 2011.

INTERNATIONAL ATOMIC ENERGY AGENCY (IAEA). **Quality assurance programme for digital mammography**. n. 17. Vienna: IAEA, 2011

JÄHNE, B. **Digital image processing**. 5. ed. Berlin: Springer, 2002.

JESKE, J. M.; BERNSTEIN, J. M.; STULL, M. A. Screening and diagnostic imaging. In: WINCHESTER, D. J.; WINCHESTER, D. P. **Atlas of Clinical Oncology: breast cancer**. Canadá: B.C. Decker, 2000.

KARNAN, M.; THANGAVEL, K. Automatic detection of the breast border and nipple position on digital mammograms using genetic algorithm for asymmetry approach to detection of microcalcifications. **Computer Methods and Programs in Medicine**, v. 87, p. 12-20, abr 2007.

KASABOV, N. K. **Foundation for neural networks, fuzzy systems and knowledge engineering**. Cambridge: Bradford Book, 1998.

KOPANS, D. B. **Imagem da Mama**. 2 ed. Rio de Janeiro: Medsi, 2000.

_____. **Diagnóstico por imagem da mama**. 3 ed. Rio de Janeiro: Guanabara Koogan, 2008.

KWOK, S.; CHANDRASEKHAR, R. ; ATTIKIOUZEL, Y.; RICKARD, M. Automatic pectoral muscle segmentation on mediolateral oblique view mammograms, **IEEE Transactions on Medical Imaging**, v. 23, n. 9, p. 1129-1140, set. 2004.

LAU, T. K.; BISCHOF, W. F. Automated detection of breast tumors using the asymmetry approach. **Computer and Biomedical Researches**, v. 24, n. 3, p. 273-295, 1991.

LIM, E. T.; O'DOHERTY, A.; HILL, A. D.; QUINN, C. M. Pathological axillary lymph nodes detected at mammographic screening. **Clinical Radiology**, v. 59, n.1, p. 86-91, jan. 2004.

LIU, Q.; TAN, Y.; WANG, J.; MA, X.; NI, H. Mammogram density estimation using sub-region classification. In: 4th Inter. Conf. On Biomedical Engineering and Informatics – BMEI 2011, Shangai, China. **Proc. of BMEI 2011**, v.4, p. 356-359, out. 2011.

MARCOMINI, K. D.; SCHIABEL, H. Nodules Segmentation in Breast Ultrasound using the Artificial Neural Network Self-Organizing Map. In: World Congress on Engineering (2012 International Conference of Signal and Image Engineering), 2012, Londres, UK. **Proc. of World Congress on Engineering 2012**, 2012. v. II, p. 1152-1155, jul 2012.

MARQUES FILHO, O.; VIEIRA NETO, H. **Processamento digital de imagens**. Rio de Janeiro: Brasport, 1999.

MATHEUS, B. R. N.; SCHIABEL, H. Online mammographic images database for development and comparison of CAD Schemes. **Journal of Digital Imaging**, v. 24, n. 3, p. 500-506, jun. 2011.

MENCATTINI, A.; SALMERI, M.; CASTI, P.; PEPE, M. L. Local active contour models and Gabor wavelets for an optimal breast region segmentation. **Inter. Journal of Computer Assisted Radiology and Surgery**, v. 7, sup. 1, p. 256-258, 2012. In: 14th Workshop on Computer-Aided Diagnosis, in 26th International Congress on Computer Assisted Radiology and Surgery - CARS2012, Pisa, Italy, jul. 2012.

MÉNDEZ, A. J.; TAHOCES, P. G.; LADO, M. J.; SOUTO, M.; VIDAL, J. J. Computer-Aided Diagnosis: automatic detection of malignant masses in digitized mammograms. **Medical Physics**, v. 25, n. 6, p. 957-964, jun. 1998.

MENECHELLI, R. C.; RIBEIRO, P. B.; SCHIABEL, H. Desenvolvimento de um classificador da forma de nódulos mamográficos segmentados utilizando a rede neural artificial Multi-layer Perceptron (MLP). In: VI Workshop de Visão Computacional – WVC, 2010, Presidente Prudente/SP, **Anais do VI WVC'2010**, p. 180-185, jul. 2010a.

_____. Classificação pela forma de nódulos mamográficos de imagens reais digitalizadas e simuladas. In: XXII Congresso Brasileiro de Engenharia Biomédica - CBEB, 2010, Tiradentes/MG. **Anais do XXII CBEB'2010**, p. 717-720, out. 2010b.

_____. Classificação da retração do mamilo em imagens simuladas digitais. XL Congresso Brasileiro de Radiologia, 2011, Recife, PE. **Anais do XL CBR**. out. 2011.

MENECHELLI, R. C.; SCHIABEL, P. B. Avaliação de técnicas computacionais para delimitação automática do músculo peitoral em imagens mamográficas digitalizadas. In: XXIII Congresso Brasileiro de Engenharia Biomédica - CBEB, 2012, Porto de Galinhas/PE. **Anais do XXIII CBEB'2012**, p. , out. 2012.

MILLER, P.; ASTLEY, S. Detection of breast asymmetry using anatomical features. In: ACHARYA, R. S.; GOLDBOF, C. B. Biomedical Image Processing and Biomedical Visualization, **Proc. of SPIE 1993**, v. 1905, p. 433-442, 1993.

MUSTRA, M.; BOZEK, J. GRGIC, M. Nipple detection in craniocaudal digital mammograms. 51st Inter. Symposium ELMAR-2009, **Proc of ELMAR'09**, p. 15-18, Zadar, Croatia, set 2009.

NAGI, J.; KAREEM, S. A.; NAGI, F. AHMED, S. K. Automated breast profile segmentation for ROI detection using digital mammograms. IEEE EMBS Conf. on Biomedical Engineering & Sciences 2010, 2010, Kuala Lumpur, Malaysia. **Proc. of IECBES 2010**, p. 87-92, nov/dec, 2010.

NATTKEMPER, T. W.; DEGENHARD, A.; TWELLMANN T. Breast Tumor Classification and Visualization with Machine-learning Approaches. In: HAYAT M. A. **Cancer Imaging: lung and breast carcinomas**. v. 1-2. New Jersey: Elsevier, 2008.

NAZÁRIO, A. C. P.; REGO, M. F.; OLIVEIRA, V. M. Nódulos benignos da mama: uma revisão dos diagnósticos diferenciais e conduta. **Revista Brasileira de Ginecologia e Obstetrícia**, Rio de Janeiro, v. 29, n. 4, p.211-219, abr. 2007.

NEVES, E. M. de A.; GONZAGA, A.; FRÈRE, A. F.. Atenção visual seletiva para reconhecimento de objetos e análise de cenas. **Multiciência**, São Carlos, v. 4, p. 26-44, 1999. Disponível em: <iris.sel.eesc.usp.br/lavi/pdf/Evelina-P_26_44.pdf> Acesso em: 11 nov. 2009.

NEWMAN, J. Quality control and artifacts in mammography. **Radiologic Technology**, v. 70, n.1, p. 61-76, set. 1998.

NICHOLSON, B. T.; HARVEY, J. A.; COHEN, M. A. Nipple-areolar complex: normal anatomy and benign and malignant processes. **RadioGraphics**, v. 29, n. 2, p. 509-523, abr. 2009.

NISHIKAWA, R. M. Computer-aided Detection and Diagnosis. In: BRICK, U.; DIEKMANN, F. **Digital Mammography**. Medical Radiology: Diagnostic Imaging and Radiation Oncology. Springer: Germany, p. 85-106, 2010.

NUNES, F. L. S.Introdução ao processamento de imagens médicas para o auxílio ao diagnóstico: uma visão prática. In: BREITMAN, K.; ANIDO, R. (org.). **Atualizações em informática**. Rio de Janeiro: PUC-RIO, v. 1, p. 73-126, 2006.

OLIVETE, C.; RODRIGUES, E. L. L.; NASCIMENTO, M. Z. Metodologia de pré-processamento baseada na correção do efeito Heel aplicada em imagens radiográficas da mão. In: III Congresso Latino-Americano de Engenharia Biomédica, 2004, João Pessoa/PB. **Anais do III Congresso Latino-Americano de Engenharia Biomédica**, p. 1375-1378, 2004.

OTSU, N. A threshold selection method from gray-level histograms. **IEEE Trans. on Systems, Man, and Cybernetics**, v. 9, n. 1, p. 62-66, jan. 1979.

PATHMANATHAN, P.; GAVAGHAN, D. J.; WHITELEY, J. P.; CHAPMAN, S. J.; BRADY, J. M. Predicting tumor location by modeling the deformation of the breast. **IEEE Transaction on biomedical engineering**, v. 55, n. 10, p. 2471-2480, out. 2008.

PATROCINIO, A. C. **Classificador automático de achados mamográficos em imagens digitais de mamas densas utilizando técnicas híbridas**. 2004, 94 f. Tese de Doutorado - Escola de Engenharia de São Carlos, Universidade de São Paulo, São Carlos, 2004.

PATROCÍNIO, A. C.; ANGELO, M. F.; SCHIABEL, H. Análise estatística de atributos extraídos de lesões mamárias. In: XXII Congresso Brasileiro de Engenharia Biomédica, 2010, Tiradentes/MG. **Anais do XXII CBEB'2010**, p. 886-890, 2010.

PATROCIONIO, A. C; SCHIABEL, H. Classifying clusters of microcalcification in digitized mammograms by artificial neural network. **IEEE Computer Graphics and Applications**, v. 1, p. 266-272, 2001.

PATROCINIO, A. C.; SCHIABEL, H.; ROMERO, R. A. F. Evaluation of bayesian network to classify clustered microcalcification. **Medical Image – Image Processing. Proceedings of SPIE**, San Diego (CA), USA, v. 5370, p. 1026-1033, 2004.

PATROCINIO, A. C.; SCHIABEL, H.; SANTAELLA, C. H. .M.; ROMERO, R. A. F. Análise de atributos extraídos de nódulos mamográficos para identificação de formatos. In: XVIII Congresso Brasileiro de Engenharia Biomédica - CBEB, 2002, São José dos Campos-SP. **Anais do CBEB'2002**, v.1, p. 317-320, 2002.

PEI, C.; WANG, C.; XU, S. Segmentation of the Breast Region in Mammograms using Marker-controlled Watershed Transform. Inter. Conf. on Information Science and Engineering, **Proc of ICISE'2010**, p. 2371-2374, dez. 2010.

PINDER, S.E.; ELLIS, I. O.; EISTON, C. W. Prognostic factors in primary breast carcinoma. **Journal of Clinical Pathology**, v. 48, n. 11, p. 981-983, nov. 1995.

PRATT, William K. **Digital Image Processing: piks inside**. 3. ed. Nova York: Wiley, 2001.

_____. **Digital image processing: piks scientific inside**. 4th ed. New Jearsey: Wiley, 2007.

PUN, T. Entropic thresholding, the new approach. **Computer Graphics and Image Processing**, v. 16, p. 210-239, jul. 1981.

RANGAYYAN, R. M.; AYRES, F. J.; DESAULTELS J. E. L. A review of computer-aided diagnosis of breast cancer: toward the detection of subtle signs. **Journal of the Franklin Institute**, v. 344, p.312-348, 2007.

RIBEIRO, Patricia Bellin. **Classificação por análise de contornos de nódulos mamários utilizando redes neurais artificiais**. 2006. 76f. Dissertação (Mestrado em Engenharia Elétrica) – Escola de Engenharia de São Carlos, Universidade de São Paulo, São Carlos.

RIBEIRO, P. B.; SCHIABEL, H.; PATROCINIO, A. C. Improvement in ANN Performance by the Selection of the Best Texture Features from Breast Masses in Mammography Images. In: **World Congress on Medical Physics and Biomedical Engineering 2006**, Seoul, Korea: Springer Berlin Heidelberg, vol. 14, p. 2439-2442, 5 jul. 2007.

RUSSELL, S. J.; NORVIG, P. **Artificial Intelligence: a modern approach**. 3. ed. : Prentice Hall, 2009.

SANTOS, V. T.; SCHIABEL, H.; GOES, C. E. A segmentation techniqe to detect masses in dense breast digitalized mammograms. *Journal of Digital Imaging*, maio, v. 15, n. 1, 2002, p. 210-213.

SCHIABEL, H.; MATHEUS, B. R. N.; ANGELO, M. F.; PATROCINIO, A. C.; VENTURA, L. A prototype of mammography CADx scheme integrated to imaging quality evaluation techniques. In: International Symposium on Medical Imaging, 2009, Orlando, FL, USA. **Proc. of SPIE - Medical Imaging 2011**, v. 7963, p. 796323-1 - 796323-12, 2011.

SCHIABEL, H.; MATHEUS, B. R. N.; VERÇOSA, L. B.; ELIAS, S.; RIBEIRO, P. B. Investigations on a CADx scheme in mammography: evaluating the influence of the nodules contour classifier on radiologists analysis. **Inter. Journal of Computer Assisted Radiology and Surgery**, v. 7, sup. 1, p. 258-259, 2012. In: 14th Workshop on Computer-Aided Diagnosis, in 26th International Congress on Computer Assisted Radiology and Surgery - CARS2012, Pisa, Italy, jul. 2012.

SCUTT D.; LANCASTER, G. A.; MANNING, J. T. Breast asymmetry and predisposition to breast cancer. **Breast Cancer Research**, v. 8, n. 2, mar. 2006. Disponível em <<http://breast-cancer-research.com/content/8/2/R14>> Acesso em 12 abr. 2010.

SILVA, K. C.; SCHIABEL, H.; NUNES, F. L. S. Medição de assimetrias em imagens mamográficas. **UNORP**, v. 4, n.2, p. 29-37, jul. 2003.

SILVA JR, E. C.; SHIABEL, H. Um novo modelo de alargamento não-linear de contraste aplicado a imagens de mamas densas. In: XX Congresso Brasileiro de Engenharia Biomédica – XX CBEB, 2006, São Pedro/SP. **Anais do CBEB'2006**, p. 85-88, out. 22-26, 2006.

SILVA JR., E. C.; SCHIABEL, H.; VENTURA, L. . Detection of clusters of microcalcification based on associated differential and morfological filters in full mammogram. In: International Symposium on Medical Imaging, 2009, Orlando, FL, USA. **Proc. of SPIE - Medical Imaging 2009**. Bellingham, Washington, USA: SPIE, 2009. v. 7259. p. 72594G-1-72594G-10, 2009.

SOUZA, Kleber Padovani; PISTORI, Hemerson. Implementação de um Extrator de Características baseado em Momentos da Imagem. In: SIMPÓSIO BRASILEIRO DE COMPUTAÇÃO GRÁFICA E PROCESSAMENTO DE IMAGENS - SIBGRAPI, III, out. 9-12, 2005, Natal. **Anais do SIBGRAPI'05** Campo Grande: UCDB , 2005. Disponível em: <http://www.gpec.ucdb.br/pistori/publicacoes/kleber_sibgrapi2005.pdf> Acessado em 30 maio 2009.

TAHMOUSH, D.; SAMET, H. Using image similarity and Asymmetry to detect breast cancer. Medical Imaging 2006: Image Processing, **Proc. of SPIE 2006**, San Diego, v. 6144, p. S1-S7, fev. 2006.

TAKAHASHI, A.; BEDREGAL, B.R.C.; LYRA, A. Uma versão intervalar do método de segmentação de imagens utilizando o K-means. **Tendências em Matemática Aplicada e Computacional (TEMA)**, v. 6, n. 2, p.315-324, 2005.

THEODORIDIS, S.; KOUTROUMBAS, K. **Pattern recognition**. 4. ed. Londres: Academic Press, 2009.

TIMMERS, J. M. H.; VAN DOORNE-NAGTEGAAL, H. J.; ZONDERLAND, H. M.; VAN TINTEREN, H.; VISSER, O.; VERBEEK, A. L. M.; DEN HEETEN, G. J.; BROEDERS, M. J. M. The breast imaging reporting and Data System (BI-RADS) in the Dutch breast cancer screening programme: its role as an assessment and stratification tool. **European Radiol**, Springer, v. 17, n. 7, 2012.

TZIKOPOULOS, S. D.; MAVROFORAKIS, M. E.; GEORGIU, H. V.; DIMITROPOULOS, N.; THEODORIDIS, S. A fully automated scheme for mammographic segmentation and classification based on breast density and asymmetry. **Computer methods and programs in biomedicine**, v. 102, n.1, Elsevier, p. 47-63, abr. 2011.

WANG, X.; LEDERMAN, D.; TAN, J.; WANG, X. H.; ZHENG, B. Computerized prediction of breast cancer risk: comparison between the global and local bilateral mammographic tissue asymmetry. Medical Imaging 2011: Computer-Aided Diagnosis, **Proc. of SPIE 2011**, Orlando (FL), USA, v. 7963, p. 79631L-1–79631L-7, fev 2011.

WIRTH, M.; NIKITENKO, D.; LYON, J. Segmentation of the breast region in mammograms using a rule-based fuzzy reasoning algorithm, **ICGST-GVIP**, v. 5, n. 2, p. 45-54, Jan. 2005.

WITTEN, I. H.; FRANK, E. **Data Mining**: Practical Machine Learning Tools and Techniques. 2 ed., Morgan Kaufmann, San Francisco, 2005.

WOLBASRST, A. B. **Physics of Radiology**. Appleton & Lange, 1993, 461p.

WORLD HEALTH ORGANIZATION (OMS). **Guidelines for the early detection and screening of breast cancer**. Cairo, 2006.

WU, J.; BESNEHARD, Q.; MARCHESSOUX, C. Automatic classification for mammogram backgrounds based on bi-rads complexity definition and on a multi

content analysis framework. Medical Imaging 2011: Image Processing, **Proc. of SPIE 2011**, Orlando (FL), USA, v. 7962, p. 79623F-1–79623F-12, fev 2011.

WU, L.; SONDGE, M. A.; LU, H. An Effective Entropic thresholding for Ultrasonic Images, **Proc. 14th Inter. Conf. on Pattern Recognition**, v. 2, p. 1552-1554, 1998.

XU, Y. Image decomposition based ultrasound image segmentation by using fuzzy clustering. Symposium on Industrial Electronics and Applications (ISIEA), Kuala Lumpur, 2009. **Proc. of ISIEA'09**, v. 2, p. 6-10, out, 2009.

XU, S.; LIU, H.; XU, X.; SONG, E.; ZENG, J. Bilateral asymmetry detection in mammograms using non-rigid registration and pseudo-color coding. **Electrical and Control Engineering - ICECE, 2010 International Conference on**, p. 544-547, jun. 25-27, 2010.

YOUK, J. H.; KIM, E. K.; KO, K. H.; KIM, M. J. Asymmetric mammographic findings based on the fourth edition of BI-RADS: types, evaluation, and management. **RadioGraphics**, v. 28, n. 7, p. 1-48, nov. 2008.

ZHENG, Bin. Mass detection scheme for digitalized mammography. In: HAYAT M. A. **Cancer Imaging: lung and breast carcinomas**. v. 1-2. New Jersey: Elsevier, 2008.