ORGANIZAÇÃO AUTOMÁTICA DE BANCOS DE MAMOGRAFIAS NO PADRÃO DE DENSIDADE BI-RADS

Silvia Cristina Martini Rodrigues

Tese apresentada à Escola de Engenharia de São Carlos, Universidade de São Paulo, como parte dos requisitos para obtenção do Título de Doutora em Engenharia Elétrica.

DEDALUS - Acervo - EESC
31100050060

Orientadora: Prof[a]. Dr[a]. Annie France Frère Slaets

São Carlos
2004

EESC-USP
Biblioteca
CAMPUS DE S. CARLOS

FOLHA DE JULGAMENTO

Candidata: Licenciada **SILVIA CRISTINA MARTINI RODRIGUES**

Tese defendida e julgada em 30-08-2004 perante a Comissão Julgadora:

aprovado

Profa. Titular **ANNIE FRANCE FRÈRE SLAETS (Orientadora)**
(Escola de Engenharia de São Carlos/USP)

aprovado

Prof. Dr. **CLÓVIS SIMÃO TRAD**
(Faculdade de Medicina de Ribeirão Preto/USP)

Aprovado

Prof. Dr. **PAULO MAZZONCINI DE AZEVEDO MARQUES**
(Faculdade de Medicina de Ribeirão Preto /USP)

Aprovado

Prof. Dr. **NELSON DELFINO D'ÁVILA MASCARENHAS**
(Universidade Federal de São Carlos/UFSCar)

APROVADA

Profa. Associada **AGMA JUCI MACHADO TRAINA**
(Instituto de Ciências Matemáticas e de Computação/USP)

Prof. Associado **MURILO ARAUJO ROMERO**
Coordenador do Programa de Pós-Graduação
Em Engenharia Elétrica

Profa. Titular **MARIA DO CARMO CALIJURI**
Presidente da Comissão de Pós-Graduação

Aos meus Pais, pelo esforço, compreensão, apoio e amor dados em todos os momentos de minha vida.

Ao Evandro, meu marido e amigo, por ter me ajudado durante o meu crescimento interior e por estar sempre presente em todos os momentos de minha vida.

À Camila e Marília, minhas filhas, por todos os momentos que vocês me proporcionaram e proporcionam, por estarem sempre presentes como filhas e como minhas melhores amigas.

Obrigada a todos vocês por fazerem parte da minha vida. Amo vocês!

- À Profa. Dra. Annie France Frère Slaets, minha orientadora e amiga pela orientação, apoio e incentivo nas horas difíceis.
- Aos meus amigos do laboratório @LADIM, por toda a ajuda neste trabalho.
- Ao Prof. Dr. Paulo M. A. Marques por ter fornecido as imagens pertencentes à base de dados do Hospital Universitário de Ribeirão Preto (FMRP–USP).
- Ao Ms. Marcelo Honda por ter disponibilizado a base de dados de Ribeirão Preto.
- Aos professores, Dr. Clóvis Trad e Dr. José Antonio Rodrigues (FMRP–USP) por terem participado das avaliações e análises das imagens utilizadas.
- Ao Prof. Dr. Fujita da Universidade Gifu – Japão, por ter cedido uma das bases de dados utilizadas.
- Aos Professores, Dr. Caetano Traina e Dra. Agma Traina por terem cedido a ferramenta *FastMapDB*.
- Ao Humberto, pertencente ao Laboratório GBDI, por ter me auxiliado no uso do *FastMapDB*.
- As secretárias e funcionários do Departamento de Engenharia Elétrica, pelo auxílio e amizade.
- As meninas da biblioteca, pelo auxílio na procura dos artigos e principalmente pelo bate papo tão gostoso.
- A todos meus familiares, pelo apoio durante todos esses anos.
- À FAPESP por ter me concedido uma bolsa de estudo e reserva técnica para a realização desse projeto.

A DEUS por ter me permitido tudo isto!

RESUMO

RODRIGUES, S. C. M. (2004). Organização Automática de Banco de Mamografias no Padrão de Densidade BI-RADS. Tese (Doutorado). Escola de Engenharia de São Carlos, Universidade de São Paulo, São Carlos (2004).

Este trabalho apresenta um método computacional que classifica as mamografias no padrão de densidade BI-RADS, visando auxiliar a detecção precoce do câncer de mama, seja essa realizada por analise visual ou por auxilio computadorizado. A classificação das mamografias em bancos padronizados objetiva eliminar conflitos entre laudos mamográficos de diferentes profissionais, bem como quanto à conduta medica a ser seguida. Entretanto, o estabelecimento de bancos feito visualmente e principalmente em períodos diferentes dificulta sua uniformização, proporcionando uma classificação muito subjetiva e relativamente grosseira em conseqüência a grande variação entre e inter observadores. O método desenvolvido permitiu classificar as imagens independentemente da subjetividade própria à observação visual de quem organizou o banco ou da técnica de exposição aos raios X utilizada. Os resultados foram superiores a 92% mesmo para bancos de imagens totalmente diferentes. Esses resultados foram obtidos respeitando-se as possíveis diferenças de interpretações de diversas equipes médicas. Alem do estabelecimento de banco de mamografias com limiares entre as composições bem quantificadas, com esta ferramenta, tanto os estagiários poderão ser treinados para classificar as imagens no padrão de densidades do BI-RADS, respeitando as particularidades locais, quanto os resultados dos CAD poderão ser comparados.

Palavras Chave: Bancos de Imagens, Processamento de Imagens, Imagens Mamográficas, densidades mamográficas, BI-RADS.

ABSTRACT

RODRIGUES, S. C. M. (2004). Automatic organization of mammography database of the density patterns described in the BI-RADS. Ph. D Thesis. Escola de Engenharia de São Carlos, Universidade de São Paulo, São Carlos (2004).

This thesis presents a computational method that classifies the mammography into the composition of the breast tissue density patterns described in the BI-RADS protocol, intended to help in the early detection of breast cancer, either if this detection happens to be realized by visual analysis or by computerized support. The classification of the mammography in standardized database intends to eliminate issues between mammography awards of distinct professionals and the correct medical conduct to be followed. However, the determination of database only visually, especially in different periods, difficult it's to standardize, causing an extremely subjective classification and relatively superficial in consequence of the large inter-and intraobserver variability. The method allows classifying the images independently of the subjective quality of the visual analysis from who organized the database or from the technique of the exposition to X-ray employed. The results were superior of 92% even to database totally distinct. These results were obtained respecting eventual differences of interpretation from several medical groups. Beside the establishment of mammography database with thresholding between the well quantified categories, this methodology will consent to probationers to be trained for classify the images according to the composition of the breast tissue density patterns described in the Bi-RADS, respecting its local particularity. Likewise, with this methodology, the results from CAD would be compared.

Key Works: Image Database, Image Process, Mammography Image, Mammographic Density, BI-RADS.

SUMÁRIO

RESUMO i
ABSTRACT ii
LISTA DE FIGURAS v
LISTA DE TABELAS viii
LISTA DE SIGLAS ix
CAPÍTULO 1 1
1. 1 MOTIVAÇÃO 1
1.2 JUSTIFICATIVA 3
1.3 OBJETIVO 4
1.4 CONTRIBUICÕES DESTE TRABALHO 4
1.5 APRESENTAÇÃO DA TESE 5
CAPÍTULO 2 6
ASPECTOS E CONCEITOS RELEVANTES PARA A TESE 6
2.1 A GLÂNDULA MAMÁRIA 6
2.2 OS SISTEMAS DE AUXÍLIO AO DIAGNÓSTICO 11
2.2.1 MAMOGRAFIA 11
2.2.2 O SISTEMA DE DIAGNÓSTICO AUXILIADO POR COMPUTADOR (CAD) 13
2.2.3 RELEVÂNCIA DOS SISTEMAS CAD 15
2.3 PADRÕES MAMOGRÁFICOS 17
2.3.1 PADRÃO DE WOLFE 17
2.3.2 PADRÃO DE DENSIDADE MAMOGRÁFICA 22
2.3.3 PADRÃO SCC – (SEIS CATEGORIAS DE COMPOSIÇÃO) 23
2.3.4 DENSIDADE E PADRÃO DE WOLFE JUNTOS COMO FATOR DE RISCO 25
2.3.5 PADRONIZAÇÃO DOS LAUDOS MAMOGRÁFICOS – BI-RADS 26
CAPÍTULO 3 28
ESTADO DA ARTE - MÉTODOS COMPUTADORIZADOS PARA CLASSIFICAÇÃO DA DENSIDADE DAS IMAGENS MAMÁRIAS 28
3.1 CLASSIFICAÇÃO SEMI-AUTOMÁTICA DA DENSIDADE ATRAVÉS DE TÉCNICA DE LIMIARIZAÇÃO 28
3.2 CLASSIFICAÇÃO AUTOMÁTICA DA DENSIDADE ATRAVÉS DE MEDIDA LOCAL DE ASSIMETRIA 29
3.3 CLASSIFICAÇÃO SEMI-AUTOMÁTICA DA DENSIDADE (WOLFE) ATRAVÉS DA ANÁLISE DE FOURIER 30
3.4 CLASSIFICAÇÃO AUTOMÁTICA DA DENSIDADE ATRAVÉS DE LIMIARIZAÇÃO E TÉCNICAS DE FRACTAIS 33
3.5 CLASSIFICAÇÃO AUTOMÁTICA DA DENSIDADE ATRAVÉS DA TÉCNICA AUTOMÁTICA DE LIMIARIZAÇÃO GLOBAL 35
3.6 CLASSIFICAÇÃO AUTOMÁTICA DA DENSIDADE ATRAVÉS DE UMA METODOLOGIA ESTATÍSTICA BASEADA EM UMA ANÁLISE DE PROBABILIDADE UTILIZANDO O χ^2 35
3.7 CLASSIFICAÇÃO AUTOMÁTICA DA DENSIDADE ATRAVÉS DO LIMIAR ÓTIMO DE KITTLER E ILLINGWORT'S 37

3.8 CLASSIFICAÇÃO AUTOMÁTICA DA DENSIDADE ATRAVÉS DE ESTATÍSTICAS DA IMAGEM, MÉTODO DA VARIAÇÃO, MORFOLOGIA MATEMÁTICA, TEXTURA DE BORDA, E SUBTRAÇÃO GAUSSIANA. 39
CAPÍTULO 4 41
METODOLOGIA 41
4.1 BANCO DE IMAGENS 41
4.1.1 BANCO DE IMAGENS DA FACULDADE DE MEDICINA DE RIBEIRÃO PRETO (FMRP) 41
4.1.2 BANCO DE IMAGENS DA UNIVERSIDADE DE GIFU (JAPÃO) - FUJILAB 42
4.2 MÉTODOS DE PRÉ PROCESSAMENTOS UTILIZADOS 43
4.2.1 DIVISÃO DAS IMAGENS EM REGIÕES 43
4.2.2 DETECÇÃO DA BORDA DA MAMA 43
4.3 MÉTODOS DE CLASSIFICAÇÃO DE DENSIDADES UTILIZADOS 45
4.3.1 EXTRAÇÃO DOS PARÂMETROS ESTATÍSTICOS 45
4.3.2 CLASSIFICAÇÃO COM FASTMAPDB 47
4.3.3 MÉTODO DE CLASSIFICAÇÃO SEMELHANTE AO USADO PELO MÉDICO 47
4.3.3.1 O MÉTODO 47
4.3.3.2 CÁLCULO DO SEGUNDO MOMENTO ANGULAR (SMA) 49
4.3.3.3 FLUXOGRAMA UTILIZADO 51
CAPÍTULO 5 53
RESULTADOS 53
5.1 DIVISÃO DA IMAGEM E DETECCÃO DA BORDA 53
5.2 CLASSIFICAÇÃO DAS IMAGENS DO BANCO DE TREINAMENTO DA FACULDADE DE MEDICINA DE RIBEIRAO PRETO NO PADRÃO DE DENSIDADE MAMÁRIA BI-RADS 57
5.2. 1 PRIMEIRA CLASSIFICAÇÃO 57
5.2.2 CLASSIFICAÇÃO COM O FASTMAP 59
5.2.3 CLASSIFICAÇÃO DAS IMAGENS UTILIZANDO DENSIDADE E HOMOGENEIDADE 65
5.3 VALIDAÇÃO DO MÉTODO DESENVOLVIDO 81
5.3.1 APLICACÃO NO BANCO DO FUJILAB 81
5.3.2 APLICAÇÃO NO BANCO DE TREINAMENTO (FMRP-II) 94
5.4 ESTABELECIMENTO DE UM BANCO DE IMAGENS PADRÃO 98
5.5 DISCUSSÃO 102
CAPÍTULO 6 103
CONCLUSÕES 103
6.1 DETECÇÃO DA BORDA DA MAMA 103
6.2 CLASSIFICAÇÃO NO PADRÃO DE DENSIDADE BI-RADS 103
6.3 CLASSIFICAÇÃO EM OUTROS PADRÕES 104
6.4 CONCLUSÕES FINAIS 106
6.5 PROPOSTA PARA TRABALHOS FUTUROS 106
REFERÊNCIAS 108

LISTA DE FIGURAS

Figura 1 – Relações Anatômicas Gerais. 7
Figura 2 – Fáscias Importantes da Região Mamária. 8
Figura 3 – Músculos da Região Mamária. 8
Figura 4 – Visão frontal da mama (três tipos de tecidos). 9
Figura 5 - a) Posicionamento para imagem CC, b) – Imagem mamográfica da mama na posição CC. 12
Figura 6 – a) Posicionamento para imagem MLO, b) Imagem radiográfica da mama na posição MLO. 13
Figura 7- Classificação N1, ... 19
Figura 8- Classificação P1. 19
Figura 9- Classificação P2. 19
Figura 10-Classificação DY. 19
Figura 11- Classificação QDY. 20
Figura 12 - Seis tipos de classificação da densidade mamária (SCC). A) Nenhuma Densidade; B) Densidade < que 10% da área total da mama; C) Densidade de 10-25% da área total da mama; D) Densidade de 25-50% da área total da mama; E) Densidade de 50-75% da área total da mama; F) Densidade > que 75% da área total da mama. 24
Figura 13 – Composição segundo BI-RADS – Composição I. 27
Figura 14 – Composição segundo BI-RADS - Composição II. 27
Figura 15 – Composição segundo BI-RADS - Composição III, ... 27
Figura 16 – Composição segundo BI-RADS - Composição IV. 27
Figura 17 - (a) A linha pontilhada sinaliza a borda da mama, enquanto a linha sólida identifica o tecido fibroglandular. (b) O histograma mostra o limiar selecionando a i_{borda} e a densidade i_{DY}. A área no histograma i_{borda} é a área A, descrita na eq. (1). 28
Figura 18 – Divisão da mama em três Regiões. 31
Figura 19 – Regiões da mama selecionadas pelo computador para serem analisadas. .. 32
Figura 20 – Imagem de uma superfície que pode ser associada com uma imagem de mama. Na representação digital, o nível de cinza de cada "pixel" determina a altura da superfície. As imagens de cima representam densidades menores, enquanto as de baixo representam maior densidade. 33
Figura 21 – Distribuição da assimetria, a) negativo e b) positivo. 34
Figuras 22 – a) imagem normal de mama. b) imagem com transformada logarítmica aplicada, note que a região apontada com a seta da esquerda representa área de tecido adiposo, enquanto a seta da direita tecido denso. 36
Figuras 23 – a) Resultado da primeira interação com o SL = 0,10, b) Resultado da segunda interação com o SL = 0,001. 36
Figuras 24 – a) Mamograma de uma mulher de 71 anos, b)segmentação da densidade por limiar regular global, c) Variação da imagem, d) Segmentação da mesma imagem pelo método desenvolvido. 38
Figuras 25 – a) Mamograma de uma mulher de 43 anos, b)segmentação da densidade por limiar regular global, c) Variação da imagem, d) Segmentação da mesma imagem pelo método desenvolvido. 38

Figura 26 – Fluxograma para classificar as imagens nas composições de densidades mamárias descritas pelo método BI-RADS 52
Figura 27 – Mamograma digitalizado do FMRP antes do processamento. 53
Figura 28 – Região de interesse do mamograma original selecionada utilizando o IMAGE J. 54
Figura 29 – Imagem da mama dividida em 25 regiões. 55
Figura 30 – Imagem binarizada da mama. 55
Figura 31 – Mamografia com o fundo preto (zero) e a região da mama com as características originais. 56
Figura 32 – Imagem com a detecção da borda da mama. 56
Figura 33 – Distribuição das Composições I, II, III e IV, utilizando o FastMap. 60
Figura 34 – Distribuição da Composição I e a classe II, com visualização no FastMap. 61
Figura 35– Distribuição da Composição I e classe II utilizando somente os valores de xa, xb e razão a/b (FastMap). 62
Figura 36 – Distribuição da Composição IV, com as composições I, II e III agrupadas na classe I (FastMap). 63
Figura 37 – Distribuição das Composições I, II, III e IV, utilizando somente os valores de xa, xb e razão a/b sendo que as I, II e III foram chamadas de I, (FastMap). 64
Figura 38 a) imagem original e b) mamografia com densidade 1 marcada pelo limiar 1. 66
Figura 39 – Distribuição dos valores de porcentagens da mamografia ocupada pela densidade 1 para todas as imagens do banco da FMRP-I. 67
Figura 40 –Distribuição dos valores de porcentagens de densidade 1 para as imagens do banco da FMRP-I com Composição I. 67
Figura 41 - Distribuição dos valores de porcentagens de densidade 1 para as imagens do banco da FMRP-I com Composição II. 68
Figura 42 –Distribuição dos valores de porcentagens de densidade 1 para as imagens do banco da FMRP-I com Composição III. 68
Figura 43 –Distribuição dos valores de porcentagens de densidade 1 para as imagens do banco da FMRP-I com Composição IV. 69
Figura 44 - a) imagem original e b) mamografia marcada com densidade 2 70
Figura 45 – Distribuição dos valores de porcentagens de densidade 2 para as imagens do banco da FMRP-I com Composição II, III e IV. 71
Figura 46 – Distribuição dos valores de porcentagens de densidade 2 para as imagens do banco da FMRP-I com Composição II. 71
Figura 47 – Distribuição dos valores de porcentagens de densidade 2 para as imagens do banco da FMRP-I com Composição III. 72
Figura 48 – Distribuição dos valores de porcentagens de densidade 2 para as imagens do banco da FMRP-I com Composição IV. 72
Figura 49 – Valores do Segundo Momento Angular obtidos processando as imagens não classificadas como composição I e II do banco da FMRP-I. 73
Figura 50 – Valores do Segundo Momento Angular obtidos processando as imagens da Composição III do banco da FMRP-I. 74
Figura 51 – Valores do Segundo Momento Angular obtidos processando as imagens da Composição IV do banco da FMRP-I. 74
Figura 52 - a) imagem original do FujiLab, b) mamografia com densidade 1 marcada pelo limiar 1. 82
Figura 53 – Distribuição de porcentagens de densidades 1 para todas as imagens do banco FujiLab. 83

Figura 54 – Distribuição de porcentagens de densidades 1 para as imagens do banco FujiLab laudadas como Composição I. 83
Figura 55 – Distribuição de porcentagens de densidades 1 para as imagens do banco FujiLab laudadas como Composição II. 84
Figura 56 – Distribuição de porcentagens de densidades 1 para as imagens do banco FujiLab laudadas como Composição III. 84
Figura 57 – Distribuição de porcentagens de densidades 1 para as imagens do banco FujiLab laudadas como Composição IV. 85
Figura 58 - a) imagem original do FujiLab, b) mamografia com densidade 2 marcada pelo limiar 2. 86
Figura 59 –Distribuição de porcentagens de densidade 2 para as todas as imagens do banco Fujilab. 87
Figura 60 –Distribuição de porcentagens de densidade 2 para as imagens do banco Fujilab laudadas como Composição II. 87
Figura 61 –Distribuição de porcentagens de densidades 2 para as imagens do banco Fujilab laudadas como Composição III. 88
Figura 62 –Distribuição de porcentagens de densidade 2 para as imagens do banco Fujilab laudadas como Composição IV. 88
Figura 63 - Histograma dos valores de SMA para as Composições laudadas como III e IV do FujiLab. 89
Figura 64 - Histograma dos valores de SMA para a Composição laudadas como III do FujiLab. 90
Figura 65 - Histograma dos valores de SMA para a Composição laudadas como IV do FujiLab. 90
Figura 66 – Imagem do FujiLab da Composição III, classificada pelo algoritmo como Composição II. 93
Figura 67 – Imagem do FujiLab da Composição IV, classificada pelo algoritmo como Composição III. 93
Figura 68 – **a)** imagem mais representativa da Composição I, **b)** imagem mais representativa da Composição II, **c)** imagem mais representativa da Composição III e **d)** imagem mais representativa da Composição IV considerando os critérios da equipe médica da FMRP. 99
Figura 69 – **a)** imagem mais representativa da Composição I, **b)** imagem mais representativa da Composição II, **c)** imagem mais representativa da Composição III e **d)** imagem mais representativa da Composição IV do banco FujiLab. 101

LISTA DE TABELAS

TABELA 1- Características estatísticas da imagem 45
TABELA 2 – Relação das imagens com as porcentagens de densidade 1 e 2, os SMAs e a concordância entre o laudo e as regras para o banco FMRP-I..... 76
TABELA 3 – Resultados de VP, FN e FP. 80
TABELA 4 – Resultado da classificação dada pelo médico e a concordância com o laudo, banco do FujiLab..... 91
TABELA 5– Resultados de VP, FN e FP para o FujiLab. 92
TABELA 6 – Relação das imagens com as porcentagens de densidade 1 e 2, os SMAs e a concordância entre o laudo e as regras..... 94
TABELA 7– Resultados de VP, FN e FP para o FMRP-II..... 98

LISTA DE SIGLAS

ACR - Colégio Americano de Radiologia

BI-RADS - Breast Imaging Reporting and Data System

CAD – Diagnostico Auxiliado por Computador

CC – Crânio Caudal

ECM - Exame Clínico das Mamas

ESPECIFICIDADE - o número de decisões negativas verdadeiras pelo número de casos verdadeiramente negativos

FDA - Food and Drug Administration

FMRP – Faculdade de Medicina de Ribeirão Preto

FN – Falso Negativo

FP – Falso Positivo

FUJILAB – Laboratório do Prof. Dr. Fujita, Japão.

INCA – Instituto Nacional do Câncer

LO - Médiolateral Oblíqua

MM – Médiolateral

ROC – Receiver Operating Characteristic

SENSIBILIDADE - o número de decisões verdadeiras positivas por número de casos verdadeiramente positivos.

SMA – Segundo Momento Angular

VP - Verdadeiro Positivo

VN – Verdadeiro Negativo

CAPÍTULO 1

1. 1 MOTIVAÇÃO

Em 2000, foram registradas 8.390 mortes decorrentes do câncer de mama. Dos 402.190 novos casos de câncer com previsão de serem diagnosticados em 2003, o câncer de mama será o segundo mais incidente entre a população feminina, sendo responsável por 41.610 novos casos e 9.335 óbitos, (INCA – Instituto Nacional do Câncer, 2004) .

A detecção precoce é uma das maiores armas contra o câncer, sendo realizada, principalmente, através da mamografia. Segundo Homer (1997), estudos feitos em mulheres americanas mostram que, quando o câncer de mama é detectado no estágio inicial, a chance de cura é de 93%, mas quando se encontra num estágio intermediário, essa taxa diminui para 72%, chegando a apenas 18% de sobrevida quando a doença está em estágio adiantado. Outros estudos, realizados pelo INCA, demonstram que sessenta por cento dos casos identificados em estágios avançados, sendo tumores com mais de cinco centímetros de diâmetro, são incuráveis. Quando é possível remover o tumor, ainda são necessários dez anos de não incidência da doença para que a pessoa seja considerada curada. A tendência é de que pelo menos 10% das vítimas apresentem novamente o problema.

Um dos fatores importantes para a detecção precoce é o padrão parenquimal, sendo que a porcentagem da área da mama ocupada por densidade discrimina melhor o risco de câncer, segundo afirmam Byrne (2002), Heine e Malhotra (2002). Além do mais, a densidade mamográfica dificulta a detecção do câncer seja essa feita visualmente ou com ajuda de sistemas computadorizados (CAD). Entretanto a classificação do padrão de densidade feita pelos radiologistas depende da qualidade da imagem e da comparação de exames anteriores. Para diminuir essa subjetividade, o Colégio Americano de Radiologia – American College of Radiology (1995), junto com

outros órgãos, padronizou o laudo mamográfico. Essa padronização obteve o nome de BI-RADS (Breast Imaging Reporting and Data System) e tem como maior objetivo à ajuda na comparação de exames por diferentes radiologistas e condutas a serem tomadas frente a achados mamográficos específicos. A padronização visa eliminar conflitos entre laudos mamográficos de diferentes profissionais, bem como quanto à conduta médica a ser seguida. Uma das características utilizadas no padrão BI-RADS é a descrição do tipo de tecido mamário sendo a densidade, dividida em quatro composições que variam desde a mama composta principalmente por tecido adiposo até a mama radiograficamente densa. Essa informação como melhora significativamente a sensibilidade deve ser incorporada ao modelo do diagnóstico. Quando uma imagem é classificada como pertencente à Composição I, é fácil encontrar na análise visual ou no processamento uma anormalidade. Em compensação, quando uma imagem pertencente à Composição IV, o grau de dificuldade é maior, sendo que esse mamograma deve ser visto com alto grau de suspeita.

Para auxiliar o diagnostico precoce do câncer mamário, vários centros de pesquisas estão desenvolvendo "softwares" especialistas. O primeiro sistema CAD (CAD-Computer-Aided Diagnosis) para diagnóstico do câncer de mama foi desenvolvido por Franklin and Angerman (1983), que usaram um sistema "inteligente" para demonstrar a importância de se ter uma ajuda no diagnóstico. A proposta do sistema CAD é de contribuir para diminuir a taxa de resultados falsos positivos e falsos negativos. Utilizando a resposta do computador como uma segunda opinião, o radiologista melhora os acertos nos diagnósticos diminuindo com isso os exames de biopsias desnecessários. Os pesquisadores vêm apresentando resultados que alcançam cerca de 90% de êxito, tanto na detecção de nódulos e microcalcificações, quanto na classificação dos tumores em malignos e benignos. Entretanto, já é bem estabelecido na comunidade científica, que estes algoritmos somente serão aplicáveis ao tipo de imagens que o pesquisador utilizou, proporcionando resultados inferiores para outras mamografias.

Para solucionar este problema, já foram apresentadas em fóruns internacionais, algumas propostas tais como: utilizar um banco de imagens único para todas as pesquisas ou estabelecer a mesma tecnologia para a aquisição das mamografias. Entretanto, todas foram rejeitadas por se mostrarem inviáveis. A única padronização conseguida até agora foi o protocolo BI-RADS. Desde então, vários hospitais e centros

de pesquisas têm proposto que as imagens sejam classificadas também em função da composição da densidade, descrita no BI-RADS.

Vários trabalhos mostram que a taxa de acerto do diagnostico depende muito da prática de leitura das mamografias realizadas pelos radiologistas. Segundo o *Journal of the National Cancer Institute* (BOCARADIOLOGY – 2002), os radiologistas que lêem mais de 300 mamografias por mês obtêm resultados mais precisos do que aqueles que lêem menos. O artigo sugere que países que requerem volumes altos de leituras, como o Reino Unido, no qual um mínimo de 5.000 mamografias são lidas pelo radiologista por ano, obtêm melhores resultados diagnósticos do que países como os EUA, que requerem um volume anual de leituras de apenas 480 mamografias. A Dra. L. Esserman, da University of California em San Francisco, EUA, avaliou a precisão diagnóstica de 204 radiologistas a quem pediu para interpretar 60 mamografias, 13 das quais diagnosticavam câncer. Estes radiologistas com baixo, médio e alto volumes liam 100 ou menos, 101 a 300 e mais de 300 mamografias por mês, respectivamente. Num nível de especificidade de 90%, a sensibilidade para diagnosticar o câncer de mama variou de 78,5% para os radiologistas com mais de 300 leituras a 64,8% para os radiologistas com baixo volume.

1.2 JUSTIFICATIVA

As considerações anteriores mostram que a classificação das mamografias no padrão de densidade BI-RADS deve auxiliar a detecção precoce do câncer de mama, seja essa realizada por analise visual ou por auxilio computadorizado. Um banco de imagem também é relevante para o ensino e treinamento na área de radiologia de mama. Entretanto o estabelecimento do banco feito visualmente e principalmente em períodos diferentes dificulta sua uniformização sendo a classificação muito subjetiva e relativamente grosseira em conseqüência a grande variação entre e inter observadores, (CHANG et al, 2002).

Vários pesquisadores desenvolveram técnicas computacionais para a detecção de densidade com o intuito de tornar cada vez mais segura a classificação da composição mamária. Taylor et al (1984) desenvolveram uma metodologia semi-automática que selecione todas as imagens de mama com características de tecido adiposo, separando do banco todas as imagens de fácil compreensão, aumentando o tempo disponível do

radiologista para as imagens que apresentam maior dificuldade. Byng et al (1996) e Karssemeijer (1998), utilizaram limiarização e técnicas de fractais para relacionar o brilho da imagem com a variação da textura da mama. Utilizando o músculo peitoral como referência, Heine e Velthuizen (2000) estimaram a variância global por comparações estatísticas, aplicando um procedimento de detecção baseado em análise de probabilidade $\chi 2$. Chang et al (2002) propuseram uma quantificação objetiva e reproduzível da composição do tecido mamário usando estatística da imagem, morfologia matemática, textura e subtração gaussiana. Entretanto, essas técnicas são semi-automáticas, requerendo a intervenção manual para a classificação das imagens.

A real necessidade de uma ferramenta que permite classificar os mamogramas, quantificando os limiares entre as composições em função dos critérios da equipe medica de uma maneira automática, delineou o objetivo deste trabalho.

1.3 OBJETIVO

Desenvolver um método que permite organizar automaticamente as imagens mamográficas nos padrões de densidade descritos pelo Colégio Americano de Radiologia (ACR), BI-RADS, independentemente da subjetividade própria à observação visual de quem organizou o banco ou da técnica radiográfica utilizada, respeitando as características da equipe medica.

1.4 CONTRIBUICÕES DESTE TRABALHO

O método desenvolvido proporciona:

- Um banco organizado no padrão BI-RADS com as características (limiares) estabelecidas pela equipe medica.
- Estabelecimento de banco de mamografia padrão.
- Um banco didático, podendo tanto ser usado em sala de aula como também no treinamento e aprendizagem dos residentes.

- Uniformizar o banco de imagens dando sugestão ao médico para uma classificação mais próxima do padrão utilizado no serviço.
- O desenvolvimento de algoritmos (CAD) específicos para as dificuldades apresentadas em cada composição do tecido mamário.

1.5 APRESENTAÇÃO DA TESE

A tese divide-se em 6 capítulos. A seguir, são apresentadas breves descrições de cada um dos capítulos.

- **Capítulo 1-** Motivação, justificativa, objetivo e apresentação da tese.
- **Capítulo 2-** Aspectos e Conceitos Relevantes para a Tese. Neste capítulo há uma breve descrição da glândula mamária, os sistemas de auxílios ao diagnostico (mamografia e CAD) e suas relevâncias. Também os tipos de padrões mamográficos, padrão de Wolfe, densidade mamográfica, padrão SCC (Seis categorias de composição), densidade e padrão de Wolfe juntos como fator de risco e padronização dos laudos mamográficos (BI-RADS).
- **Capítulo 3-** Estado da arte. São apresentados métodos computadorizados para a classificação da densidade das imagens mamarias.
- **Capítulo 4-** Metodologia. Neste capitulo são apresentados os bancos de imagens utilizados no trabalho, o pré-processamento e a detecção da borda da mama, método de classificação utilizado (extração dos parâmetros estatísticos e classificação com o FastMapDB) e o método semelhante ao usado pelo medico.
- **Capítulo 5-** Resultados. São apresentados os resultados obtidos em cada etapa do trabalho, validação do método desenvolvido, estabelecimento de um banco de imagens padrão e a discussão.
- **Capítulo 6-** Conclusões. Neste capitulo são apresentadas às conclusões das etapas realizadas, conclusões finais, trabalhos futuros e referências.

CAPÍTULO 2

ASPECTOS E CONCEITOS RELEVANTES PARA A TESE

2.1 A GLÂNDULA MAMÁRIA

A anatomia da mama, abordando as características consideradas atualmente relevantes para o diagnóstico, já foi bem descrita por (PINOTTI, 1991). A mama é a maior glândula da pele, derivada do ectoderma, apresentando características comuns com as glândulas apócrinas desenvolvidas nas proximidades dos pelos. O volume da mama depende da quantidade de gordura e do tecido conjuntivo que a preenche, e não da quantidade de tecido glandular que possui, e, por esse motivo, seu volume maior ou menor nunca é expressão de sua capacidade funcional.

Geralmente, o tecido glandular emite dois prolongamentos que se dirigem, respectivamente, para a região ínfero-externa, que pode ser o prolongamento curto, e (outro) para a região súpero-externa da mama, que pode ir até a axila em maior ou menor profundidade, e é chamado prolongamento axilar (Figura 1). O prolongamento glandular axilar pode fazer-se por baixo do bordo caudal do músculo peitoral maior, motivo pelo qual passa desapercebido à exploração médica.

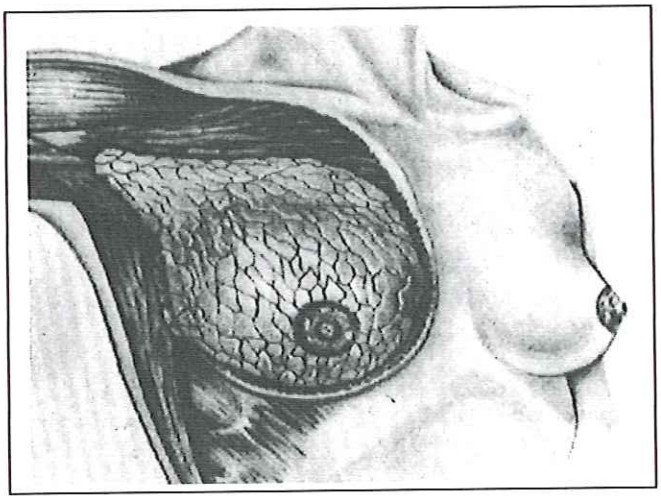


Figura 1 – Relações Anatômicas Gerais.
Fonte: PINOTTI (1991).

No vértice da mama situa-se o mamilo, um pequeno botão de pele rugosa fortemente pigmentada, que se eleva sobre sua superfície cutânea. O mamilo é rodeado por uma pequena zona circular chamada aréola.

Sobre a superfície da aréola surge uma série de "botõezinhos" – de 10 a 15 – determinados pela presença de glândulas cutâneas subjacentes – as glândulas de Montgomery – que são glândulas mamarias rudimentares.

A glândula mamaria é constituída de dois compostos principais: um parênquima epitelial de origem ectodérmica e um suporte conjuntivo de origem mesodérmica. O parênquima está formado pelos condutos galactóforos (que conduzem leite) com seus alvéolos, que se reúnem em feixes, em número variável formando os lóbulos mamários. O número de lóbulos varia de 15 a 20 na mulher adulta.

Entre a camada profunda da fáscia (faixa de tecido fibroso) superficial e a fáscia profunda que cobre o músculo peitoral e outros músculos da parede torácica, se encontra o espaço retromamário. As projeções da camada profunda da fáscia superficial cruzam o espaço retromamário e se fundem com a fáscia profunda ou fáscia peitoral, constituindo o ligamento suspenso posterior da mama (Figura 2).

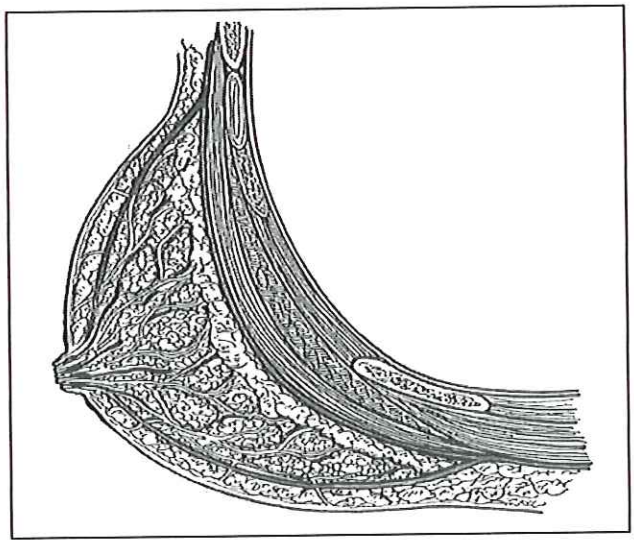


Figura 2 – Fáscias Importantes da Região Mamária.
Fonte: PINOTTI (1991).

O músculo grande peitoral abre-se em leque, inserindo-se no gradeado costal e no esterno (Figura 3). As metástases nesse músculo são extremamente raras.

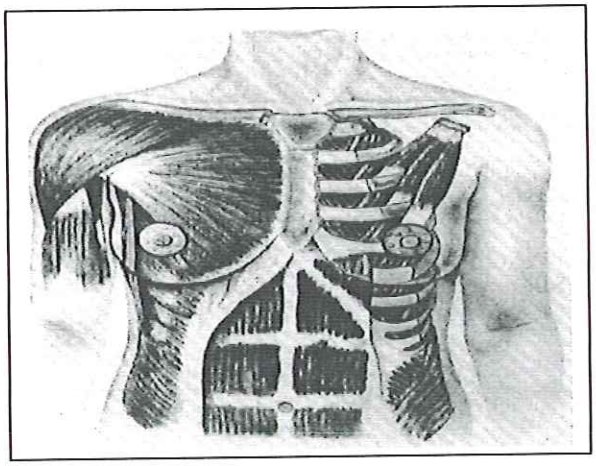


Figura 3 – Músculos da Região Mamária.
Fonte: PINOTTI (1991).

O tecido mamário pode ser dividido em três tipos principais: (1) glandular, (2) fibroso ou conjuntivo e (3) adiposo (Figura 4).

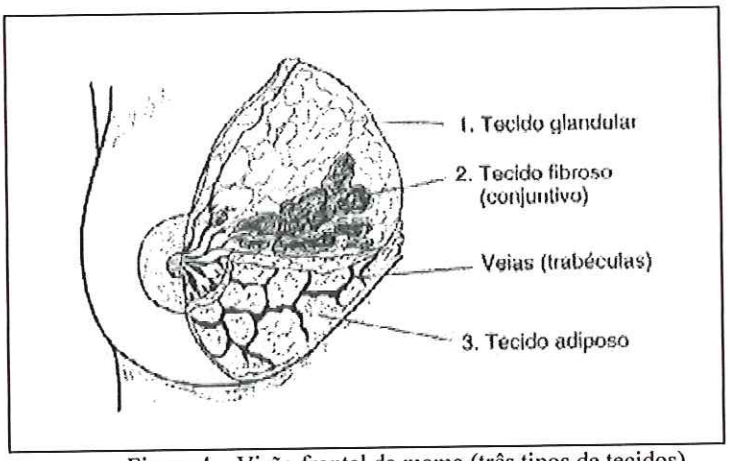


Figura 4 – Visão frontal da mama (três tipos de tecidos).

Fonte: PINOTTI (1991).

O tecido conjuntivo forma o chamado estroma que pode ser classificado como:

- Intralobular – caracterizado por ser rico em células, fibroblastos, células plasmáticas e linfócitos.

- Interlobular – caracterizado por ser pobre em conteúdo celular e rico em fibras e colágeno. Serve de armação a toda glândula mamaria, pois as suas fibras se dispõem em forma de tabiques entre os lóbulos e enviam prolongações e lâminas até a derme. Estas lâminas atuam como ligamentos suspensores chamados de Ligamentos de Cooper, são projeções denteadas periféricas de tecido mamário em processos fibrosos que atingem a camada superficial se fundindo com a mesma. Eles estabelecem uma relação íntima entre a mama e a pele, podendo determinar a retração dessa última no carcinoma e em outras lesões da mama acompanhadas por fibrose.

As mulheres mais jovens apresentam mamas com maior quantidade de tecido glandular, o que torna esses órgãos mais densos e firmes. Ao se aproximar da menopausa, o tecido mamário vai-se atrofiando e sendo substituído progressivamente por tecido adiposo, até se constituir, quase que exclusivamente, de gordura e resquícios de tecido glandular na fase pós-menopausa.

A densidade relativa da mama é afetada basicamente pelas características inerentes da paciente, estado hormonal, idade e gestações. A glândula mamaria sofre alterações cíclicas associadas à elevação e diminuição das secreções hormonais durante o ciclo menstrual, durante a gravidez e a lactação, e alterações graduais que ocorrem em toda a vida da mulher (BONTRAGER, 1999). Entretanto, genericamente falando, as mamas podem ser divididas em três grandes categorias, dependendo das quantidades de tecido fibroglandular em relação ao tecido adiposo.

- **Mama fibroglandular**: a mama jovem geralmente é muito densa, pois contem quantidade relativamente pequena de tecido adiposo. A faixa etária comum para a categoria fibroglandular varia da pós-puberdade ate cerca de trinta anos. Entretanto, as mulheres com mais de trinta anos que nunca amamentaram provavelmente também pertencerão a este grupo geral como também mulheres grávidas ou lactantes de qualquer idade.

- **Mama fibroadiposa**: à medida que a mulher envelhece e ocorrem mais alterações nos tecidos mamários, há uma mudança gradual da pequena quantidade de tecido adiposo para uma distribuição mais igual de tecido gorduroso e fibroglandular. Portanto, em um grupo que varia dos 30 aos 50 anos de idade, a mama não é tão densa quanto num grupo jovem. Radiologicamente, ela tem densidade média e requer menos exposição que o tipo fibroglandular. Varias gestações na vida reprodutiva de uma mulher acelerarão a evolução de sua mama para essa categoria fibroadiposa.

- **Mama adiposa**: ocorre após a menopausa, comumente a partir dos 50 anos de idade. Após a idade reprodutora da mulher, a maior parte do tecido glandular da mama sofre atrofia, sendo convertido em tecido adiposo em um processo denominado involução. Radiologicamente é necessária uma exposição ainda menor nesse tipo que nos dois primeiros.

2.2 OS SISTEMAS DE AUXÍLIO AO DIAGNÓSTICO

2.2.1 MAMOGRAFIA

A mamografia ainda é o método mais eficaz e mais barato para a detecção precoce do câncer de mama. O exame radiológico dos tecidos moles das mamas é considerado um dos mais importantes procedimentos para o rastreamento do câncer ainda impalpável de mama. Este método de detecção permite a observação de aspectos suspeitos em mamas nas quais não consegue-se palpar nódulos ou áreas irregulares. Segundo Dance (1996), a mamografia ainda é o único método com os requisitos necessários para a detecção de lesões pequenas.

A sensibilidade da mamografia é alta (46% a 88%), ainda que sejam registradas perdas entre 10 a 15% dos casos de câncer detectáveis ao exame físico. A especificidade varia entre 82% e 99% e depende da qualidade do exame. A sensibilidade é muito menor em mulheres jovens. A mamografia, devido à sua pouca eficácia em mulheres com menos de 40 anos e mais de 70, em termos epidemiológicos e de saúde pública, não deve ser utilizada em programas maciços, e sim ser indicada no seguimento das mulheres de alto risco ou com suspeitas de doenças mamárias, (INFOCANCER e INCA 2004).

O rastreamento do câncer de mama feito pela mamografia, com periodicidade de um a três anos, reduz significativamente a mortalidade em mulheres de 50 a 70 anos. Nas mulheres com menos de 50 anos, existe pouca evidência deste benefício. O Instituto Nacional de Câncer recomenda que o Exame Clínico das Mamas (ECM) seja realizado a cada três anos pelas mulheres com menos de 35 anos, a cada dois anos pelas mulheres entre 35 e 39 anos e anualmente pelas mulheres entre 40 e 49 anos. As mulheres na faixa etária entre 50 e 70 anos devem submeter-se ao exame semestralmente.

Segundo Souen (1998) o maior problema em relação ao método é o grande número de falsos suspeitos que faz com que realizem-se biópsias desnecessárias. Na literatura, do percentual de achados malignos ou pré-malignos, somente 20% a 30% dos casos suspeitos são comprovados, o que significa que 70% a 80% das mulheres, foram submetidas a atitudes cirúrgicas desnecessárias, com a implicação dos danos que disto decorre.

Segundo Dance (1996) a imagem mamográfica quando comparada às demais radiografias é a que requer o mais alto padrão técnico na sua execução. Isto se deve a própria estrutura dos tecidos que compõem o órgão (tecidos de densidade muito semelhantes) e a geometria bastante particular do que é radiografado (compressão e localização). Uma imagem radiografada é considerada de boa qualidade quando apresenta nitidez de detalhes e visibilidade das estruturas anatômicas de interesse. A nitidez está associada às propriedades geométricas da imagem sendo uma medida do limite de resolução da imagem radiográfica final. Já a visibilidade, associada às propriedades fotográficas da imagem, é controlada pelos fatores de técnicas radiográficas que contribuem para a densidade ótica e o contraste da imagem.

Um exame mamográfico normal consiste de 4 mamogramas, sendo duas projeções de cada mama. O posicionamento pode ser:

- CC – Crânio Caudal, o segredo desse posicionamento reside na elevação do sulco inframamário (Figura 5a). A pele que reveste a parede torácica sob a mama desliza cranialmente por vários centímetros, permitindo a elevação da face inferior da mama. Em seguida se deve ajustar a altura do mamógrafo, podendo assim ter uma boa visão do corpo mamário. A Figura 5b mostra a imagem mamográfica neste posicionamento.

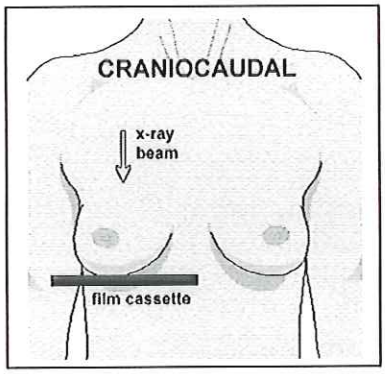


Figura 5a

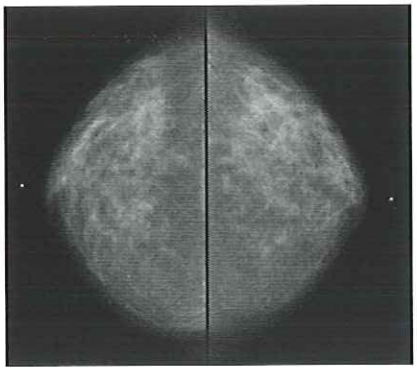


Figura 5b

Figura 5 - a) Posicionamento para imagem CC, b) – Imagem mamográfica da mama na posição CC.
Fonte: AMOROSA (2002).

- MLO e MM - Médiolateral Oblíqua (Figura 6a) e Médiolateral, nessas projeções o prolongamento axilar do corpo mamário envolve a maior parte do músculo grande peitoral. Os feixes de raios X devem ser perpendiculares à margem

lateral desse músculo. Os músculos peitorais, juntamente com a face lateral da mama são amplamente mobilizados em sentido mediais. Esse posicionamento do corpo possibilita a visualização por inteiro do corpo mamário. Os ângulos mais freqüentemente usados são de 45° no MLO e 90° no posicionamento MM. A Figura 6b mostra a imagem mamográfica neste posicionamento.

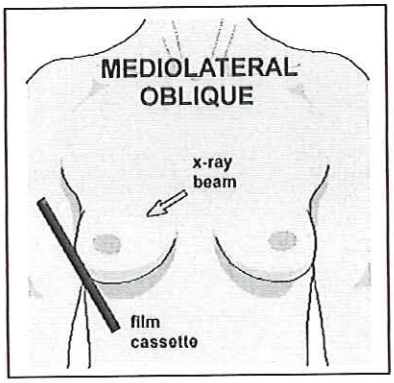


Figura 6a

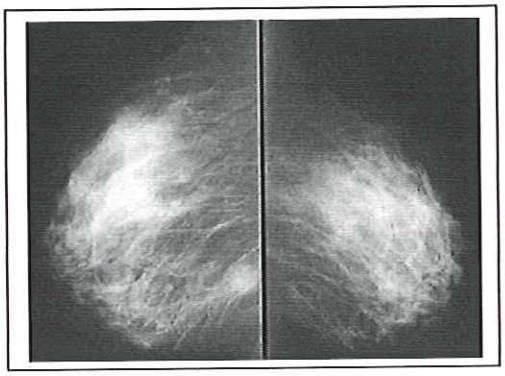


Figura 6b

Figura 6 – a) Posicionamento para imagem MLO, b) Imagem radiográfica da mama na posição MLO.
Fonte: AMOROSA (2002).

2.2.2 O SISTEMA DE DIAGNÓSTICO AUXILIADO POR COMPUTADOR (CAD)

Vários centros de pesquisas estão desenvolvendo "softwares" para auxiliar os médicos no diagnóstico do câncer. O primeiro sistema CAD (CAD-Computer-Aided Diagnosis) para diagnóstico do câncer de mama foi desenvolvido por Franklin and Angerman (1983) apud, Doi (1991), que usaram um sistema "inteligente" para demonstrar a importância de se ter uma ajuda no diagnóstico. A proposta do sistema CAD é de contribuir para diminuir a taxa de resultados falsos positivos e falsos negativos. Utilizando a resposta do computador como uma segunda opinião, o radiologista melhora os acertos nos diagnósticos diminuindo com isso os exames de biopsias desnecessários.

Segundo Doi (1999) no início da década de 90, os pesquisadores tentaram transmitir os conceitos básicos dos CAD e iniciaram a aplicação para detecção e caracterização de lesões. Estes métodos e técnicas baseados na visão do computador e na inteligência artificial, alcançaram nos últimos anos uma nova fase. A 1ª unidade comercial para detecção de lesões de mama em mamogramas desenvolvida pela "Tecnology R2 Inc, Los Altos, CA", foi aprovada em junho de 1998, pelo FDA (Food and Drug Administration) para compra e venda podendo assim ser utilizada em clínicas nos EUA.

Segundo Giger (1998), que pertence a um dos grupos de maior atuação em pesquisas de sistemas CAD, o do professor Kunio Doi de Chicago, o método permite direcionar o radiologista a dar uma maior atenção a uma suposta lesão, para assim prevenir erros ou enganos, ou mesmo o alertar para probabilidades de malignidade. Segundo Doi (1999) o sistema CAD pode ser considerado como uma segunda opinião dada com o auxílio de um computador. É então importante notar que o computador é usado somente como um instrumento para fornecer informações adicionais para o trabalho clínico, podendo auxiliá-lo a tomar a decisão final no diagnóstico de um paciente. Portanto, o conceito básico do CAD é claramente diferente do que o termo "*diagnóstico automático*", preconizado nas décadas de 60 e 70, deixava entender. O objetivo do CAD na radiologia é de melhorar a precisão do diagnóstico. Sendo que o médico interpreta a imagem usando a resposta do computador como um guia. Esse resultado computadorizado pode ser muito útil porque os diagnósticos dos radiologistas são feitos baseados em julgamentos subjetivos além do que, lesões tais como nódulos de pulmão em radiografias de tórax e microcalcificações ou massas em mamogramas são difíceis de identificar.

A visão computacional envolve o uso de um computador para o processamento dos dados e a extração das características importantes das imagens. Dependendo de sua natureza, tais características podem ou não ser visíveis ao olho humano. O desenvolvimento do sistema de visão computacional requer uma informação à priori sobre o tipo de imagem médica em questão, o conhecimento de várias técnicas de processamento computacional e do método de análise de decisão. Hoje, estão disponíveis computadores velozes que possibilitam o processamento de imagens médicas em tempo real.

De um modo geral o sistema CAD geralmente apresenta três etapas básicas baseadas em 3 tecnologias diferentes.

A 1ª etapa é o pré-processamento de imagem para o realce e extração da lesão. Várias técnicas tem sido empregadas dependendo dos diferentes tipos de lesões e incluem filtros baseados em análise de Fourier e transformadas de "Wavelet", assim como filtros morfológicos e redes neurais artificiais (ANNs).

A 2ª etapa é a quantização das características das imagens, tais como, contraste e forma do objeto selecionado no 1° passo. Isto é necessário para encontrar características ímpares que permitem distinguir uma lesão de outra ou de estruturas anatômicas normais. É útil definir na fase inicial do desenvolvimento do CAD, no mínimo as características de imagens que já foram reconhecidas e descritas pelos radiologistas. Isto porque seu conhecimento está baseado em observações de numerosos casos e na sua precisão no ato de dar o diagnóstico, o qual deve ser confiável.

A 3ª etapa é o processamento de dados para distinguir entre padrão normal e anormal, baseado nas características obtidas no 2° passo e na "regra", a qual pode ser estabelecida em função da posição da lesão e em outros padrões normais. Outras técnicas usadas incluem análise discriminante, redes neurais artificiais (ANN), e o método decisão por árvore *(decision-tree)*.

Fica claro que o conceito básico de um sistema CAD é amplo e geral, sendo aplicável a todas as modalidades de imagens, incluindo projeção convencional radiográfica, Tomografia Computadorizada (CT), Ressonância Magnética (MRI), imagem de Ultra-som e imagem de Medicina Nuclear. Esquemas computadorizados CAD podem ser desenvolvidos para todos os tipos de exames em várias partes do corpo. Entretanto as pesquisas para aplicação do CAD são limitadas, sendo que o maior foco recentemente foi no campo da mamografia para detecção precoce do câncer de mama.

2.2.3 RELEVÂNCIA DOS SISTEMAS CAD

Segundo Schmidt (1998), vários casos de câncer não são detectados pelos mamogramas convencionais. Os médicos que fazem as leituras destas imagens enfrentam problemas como fadigas, distrações e variações nos seus níveis de treinamento. Há pouco tempo que os radiologistas estão sendo treinados a analisar os mamogramas. Além do que, no programa de residência somente 2 meses dos 4-5 anos do currículo dos radiologistas, são dedicados à mamografia.

Basset et al (1987)* apud Giger (1998) afirmam que 10% a 30% das mulheres que têm câncer de mama e se submeteram a um rastreamento mamográfico, têm interpretações mamográficas negativas. A detecção errada pode ser devido à natureza sutil dos índices encontrados na mamografia, à qualidade pobre da imagem, à fadiga no olho humano, ou a equívoco do radiologista. Foi sugerido que sejam feitas duas leituras (dois radiologistas), isso provocou uma diminuição no número de lesões encontradas de 15%, (MURPHY, 1990) e (BRENNER, 1991). Assim segundo Giger e Vyborny (1994) é importante desenvolver métodos computacionais para auxiliar o radiologista, alertando-o para uma estrutura que necessita de uma melhor observação deixando a decisão final e o tratamento do paciente para o médico.

Os pesquisadores Jiang et al. (1999) observaram, a performance de dez radiologistas na tarefa de decidir qual mulher necessitava de exame de biopsia. O estudo foi realizado em duas diferentes condições: condições normais sem o computador e com o auxílio do computador. Os resultados mostraram que houve uma melhora significativa na performance quando os radiologistas usaram as indicações do computador. A sensibilidade aumentou de 73% para 87% e a especificidade de 32% para 42%. Com esses resultados, cada radiologista, em média, recomendou 6,4 biopsias a mais em casos de lesões malignas e evitou 6,0 exames de biopsias no caso de lesões benignas. Esses resultados indicaram que os computadores podem reduzir os números de biopsias desnecessárias.

Um outro estudo envolvendo a eficiência na classificação de lesões como massas foi desenvolvido por (Huo, 1995 e 1998). Os autores extraíram as lesões dos mamogramas usando técnicas automáticas. A resposta do computador forneceu uma estimativa da probabilidade de malignidade aproximadamente igual a um radiologista experiente em mamografia e muito melhor quando se fez a comparação com um radiologista com menos experiência. Usando uma base de dados o esquema alcançou um valor preditivo (número de decisões corretas dividido pelo número total de casos) de 83%, o qual foi 12% maior que o esperado pelo melhor radiologista e 21% maior do que a média esperada do radiologista inexperiente.

---

* **BASSET, LW.; GOLD, RH.** (1987) Breast cancer Detection: Mammography and other Methods in breast imaging. Grune and Stratton, New York. Vozes apud **GIGER, M.L.** (1998). Overview of Computer-Aided Diagnosis in Breast Imaging. Proceedings of the First International Workshop on Computer-Aided Diagnosis Chicago, U.S.A. 20-23 september, p. 167-176.

2.3 PADRÕES MAMOGRÁFICOS

2.3.1 PADRÃO DE WOLFE

WOLFE (1967a, 1967b) apud Heine e Malhotra (2002) observou, através de radiografia, uma relação entre o padrão ductal proeminente e o câncer de mama. O autor acreditava que a gordura, o tecido conectivo e os elementos epiteliais deviam ser vistos como uma displasia e que fibroses periductais eram representadas pelo predomínio do padrão ductal. No passado o termo displasia era usado como referência para doenças fibrocísticas da mama ou para áreas de aumento da densidade mamográfica. Wolfe observou que em mamas supostamente saudáveis a aparência ductal era mais aparente e mais freqüente quando o tecido lobular diminuía com a idade. Também notou similar tendência para pacientes no começo da doença. O padrão ductal foi mais proeminente em mulheres jovens que apresentavam doenças malignas e tecido lobular denso, sendo que alguns casos de cânceres podiam ser diagnosticados pela aparência de um grupo de dutos dilatados. Mas, em um estudo subseqüente, Wolfe (1969) notou que a presença de dutos proeminentes é uma anormalidade causada pelo depósito de colágeno subepitelial e que o diâmetro do duto é diretamente proporcional à quantidade de colágeno. Esses dutos aparecem como cordões de densidades aumentadas que se espalham a partir da região subareolar e tiveram um padrão de espalhamento descontínuo em uma mama severamente envolvida. O autor concluiu que a presença do câncer altera a distribuição do padrão ductal resultando em dutos aumentados e sinuosos.

WOLFE JN., (1967a), A study of breast parenchyma by mammography in the normal woman and those with benign and malignant disease. Radiology, v. 89, p. 201–205.

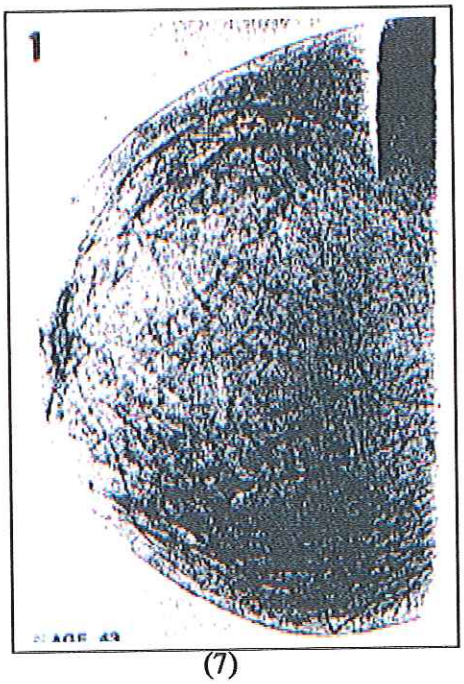
WOLFE JN.(1967b), Mammography: ducts as a sole indicator of breast carcinoma. Radiology, v. 89, p.206–210.

Wolfe (1976a) apresentou um esquema de classificação compreendendo 4 padrões relacionados com o risco de câncer de mama. A classificação foi feita de acordo com a aparência da mama (como realizado usualmente pelo radiologista). Em relação ao risco relativo (RR) as designações são: N1, P1, P2 e DY e não levam em conta a anormalidade (nódulos, microcalcificações), a vascularidade ou alguma outra característica ou fator de risco.

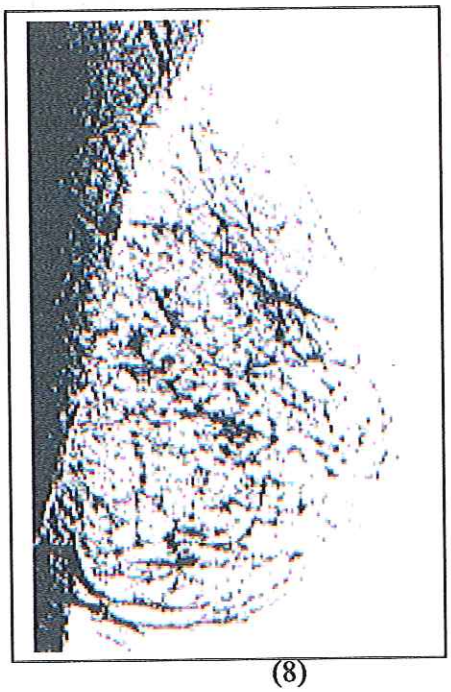
A composição mamária foi classificada por Wolfe (1967a) como:

- **N1** – considera o padrão normal ou negativo – baixíssimo risco. A imagem mostra tecidos adiposos sem dutos visíveis (Figura 7).
- **P1** - É uma classificação de baixo risco. A imagem apresenta tecidos adiposos com dutos visíveis, ocupando um dos quadrantes do volume da região subareolar. Esses dutos aparecem como cordilheiras ou estruturas lineares, em alguns casos estendem-se por todo o quadrante (Figura 8).
- **P2** - alto risco. A imagem mostra um padrão ductal proeminente que ocupa mais de ¼ do volume da mama e pode envolver todo o parênquima (Figura 9).
- **DY** – altíssimo risco. A imagem apresenta um padrão de displasia, com densidade maior que a da gordura com ausência de um padrão ductal proeminente (Figura 10).

A quinta categoria **QDY** (Figura 11) foi acrescentada para mulheres entre 40 – 45 anos, que tiveram suas imagens mamográficas com densidades aumentadas radiograficamente, mas não com muita severidade, neste padrão a maioria provavelmente irá retornar ao padrão de pouco risco com a idade.

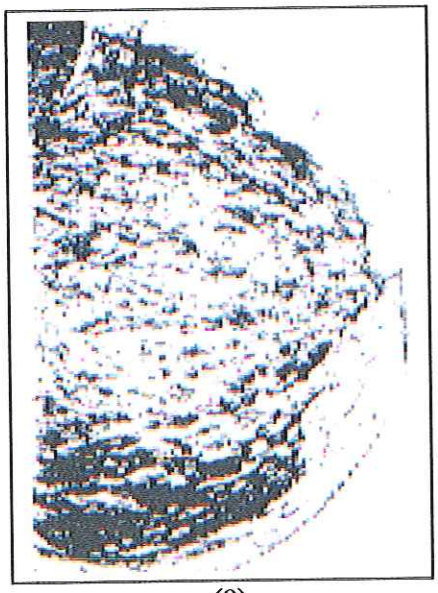


(7)

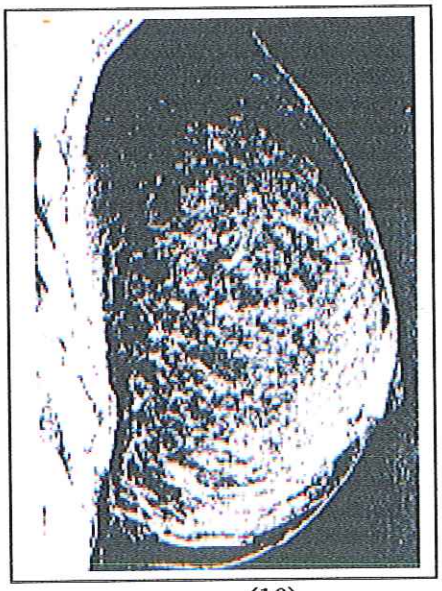


(8)

Figura 7- Classificação N1,
Figura 8- Classificação P1.
Fonte: WOLFE (1976b).



(9)



(10)

Figura 9- Classificação P2.
Figura 10-Classificação DY.
Fonte: WOLFE (1976b).

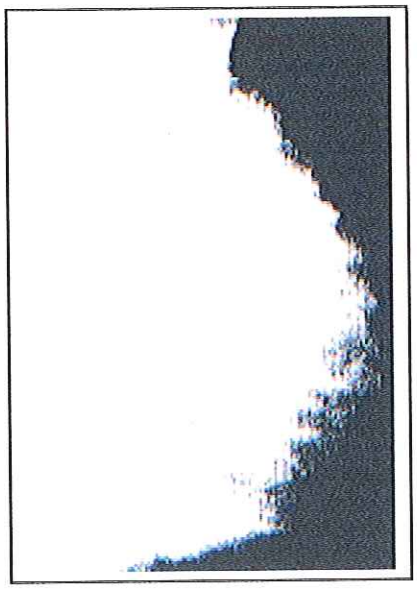


Figura 11- Classificação QDY.
Fonte: WOLFE (1976b).

O padrão DY acontece em muitas mulheres com câncer, mas também em mulheres na faixa etária dos 50 anos. Depois desta idade, a proporção de mama com o padrão DY é significativamente reduzida, com o padrão P2 acontecendo em muitos casos.

Neste estudo foram mostradas evidencias que a incidência do risco de câncer progride com o padrão em uma fatia apreciável da população estudada. Foi também notado que nem todos os carcinomas são acompanhados de padrão ductal proeminente, e o autor sugeriu que haviam dois tipos diferentes de patogenias relativas a idade, que indicavam risco de câncer.

Em estudo subseqüente Wolfe (1976b) descobriu que mulheres, que tinham padrão de baixo ou alto risco no processo inicial, não tinham demonstrado muitas mudanças com o decorrer do tempo e que, com a idade o padrão, QDY algumas vezes regressava para um padrão de baixo risco.

Em 1977, Wolfe observou que o risco de incidência de câncer elevado nos padrões P2 e DY, não era usual para mamas classificadas como N1. Ele também mostrou que há um intervalo, em média 48 meses, entre o mamograma negativo e o tempo histológico do carcinoma. Isto reforça a idéia de que casos de cânceres supostamente ignorados por descuido no exame de rastreamento inicial, na realidade

não podiam ser detectados. Ele também mostrou que quando o padrão DY muda ele reverte para P2, assim o risco não diminui porque o padrão P2 é estável no tempo.

Em resumo:

- N1 muda para P1 em 14% dos casos,
- P1 muda para N1 em 7% dos casos,
- DY regride para: - P2 em 33% dos casos
  - P1em 11% dos casos
  - N1 em 8% dos casos
- P2 raramente reverte para outro padrão.

Welling e Wolfe (1978) correlacionaram dados histológicos com composição da mama de 143 biópsias específicas. O mais alto grau de anormalidade pré-cancerosa epitelial ocorreu freqüentemente no grupo combinado P2 DY, pouco no grupo P1 e muito raramente no grupo N1. Isto deve ser visto com cuidado porque o material usado era de biópsia, o qual é um tipo de tecido a priori anormal.

Os trabalhos de Wolfe foram muito importantes para a radiologia devido a sugestões, tais como:

a) Pode-se obter uma medida do fator de risco mensurável a partir de uma imagem.
b) O intervalo entre rastreamentos pode ser ajustado em função da composição da mama.
c) A prática do rastreamento, em geral, pode ser baseada no padrão parenquimal.

Isto gerou muitas discussões, alguns autores comprovaram o padrão de Wolfe [Egan et al (1977), Wellings et al (1978), Wilkinson et al (1977), Krook (1978a), Hainline et al (1978)], enquanto que outros afirmaram que não é possível demonstrar a associação entre os tecidos da mama e o aumento no risco de câncer, Mendell et al (1977), Egan et al (1979), Moskowitz et al (1980), Buchanan et al (1981), Verbeek et al (1984).

As pesquisas que concordam com o padrão de Wolfe foram baseadas no trabalho original e apresentaram resultados semelhantes. Em geral eram três classes de estudos:

a) Caso controle - comparar um grupo de doentes (casos) e um grupo de pessoas não doentes (controles). Nos casos-controle o objetivo é verificar se

os casos doentes diferem significantemente em relação à presença de algum fator de risco. Este problema se resume em comparar a proporção de fator de risco entre os doentes e entre os não doentes.

b) "Cohort" - Estudos nos quais são identificados subconjuntos de uma população definida. Estes grupos podem ou não ser expostos aos fatores que influenciam a ocorrência de uma doença particular. São definidas populações de "cohort", que como um todo, são seguidos em uma tentativa de determinar características distintas de subgrupo.

c) Prevalência do estudo – Número de casos da doença mais os casos novos.

Tabar e Dean (1982) observaram a taxa da prevalência do câncer para mulheres com mais de 50 anos no grupo N1P1, assemelhando-as com as mais jovens do grupo P2DY. Entretanto outros autores Bland et al (1982) e Janzon et al (1982) mostraram um risco menor com o aumento da idade para mulheres com padrão DY. Wilkinson et al (1977) encontrou um aumento no risco de câncer para mulheres de todas as idades com padrão DY. Gravelle et al (1986) forneceram fortes evidências para uma associação do padrão de risco para mulheres com padrão P2 e DY. Krook et al (1978b) demonstraram que há um risco significativamente elevado para mulheres com 45 anos e que pertençam ao grupo P2 DY.

2.3.2 PADRÃO DE DENSIDADE MAMOGRÁFICA

Alguns pesquisadores relacionam a densidade mamográfica ao risco. Roebuck (1982) acredita que a medida de densidade pode ser um indicador mais fiel do risco do que o padrão de WOLFE. Mas Brisson et al (1982a e 1982b) apresentaram a terminologia mais aceita descrevendo densidade em conjunto com o padrão de WOLFE. Eles classificaram densidades como nodulares e homogêneas. Eles mostraram também que mulheres com padrão P2 tinham extensiva densidade nodular, quando mulheres com padrão DY tinham extensiva densidade homogênea, tão bem como densidades nodulares. A densidade nodular foi menor e mais concentrada em mulheres com padrão DY. Os autores levantaram também alguns pontos interessantes:

- Para mulheres com 60 anos, um aumento da porcentagem da mama mostrando densidades nodulares esta correlacionado com o aumento no risco de câncer, tendendo para densidades homogêneas.
- Um aumento no tamanho médio ou grau de concentração de densidade nodular está associado com o aumento do risco de câncer de mama. A relação ente risco e porcentagem de densidade nodular é fortalecida quando o peso e a altura estão sendo levados em conta.

Segundo Heine e Malhotra (2002) algumas pesquisas envolvendo densidade e risco apresentaram imagens digitalizadas com uma área chamada de densa e outra não densa. Se se aceita a hipótese que o tecido denso é o tecido de risco e que quanto mais denso maior o risco, duas importantes hipóteses devem ser observadas:

1. A binarização da imagem pode perder ou distorcer a informação relacionada ao risco devido aos ruídos inerentes a imagem digitalizada;
2. Há evidencias indicando que as densidades homogênea e nodular se comportem diferentemente com respeito ao risco, o que indica que uma medida de textura pode ajudar o diagnostico.

2.3.3 PADRÃO SCC – (SEIS CATEGORIAS DE COMPOSIÇÃO)

Boyd et al (1992, 1995a) classificaram as Composições mamárias em seis Categorias (SCC), dividindo subjetivamente a porcentagem da imagem da mama ocupada pela densidade mamográfica. As seis categorias são:

Categoria A: Nenhuma Densidade;
Categoria B: Densidade < que 10% da área total da mama;
Categoria C: Densidade de 10-25% da área total da mama;
Categoria D: Densidade de 25-50% da área total da mama;
Categoria E: Densidade de 50-75% da área total da mama;
Categoria F: Densidade > que 75% da área total da mama.

A Figura 12 ilustra a classificação SCC descrita por BOYD (1992, 1995a).

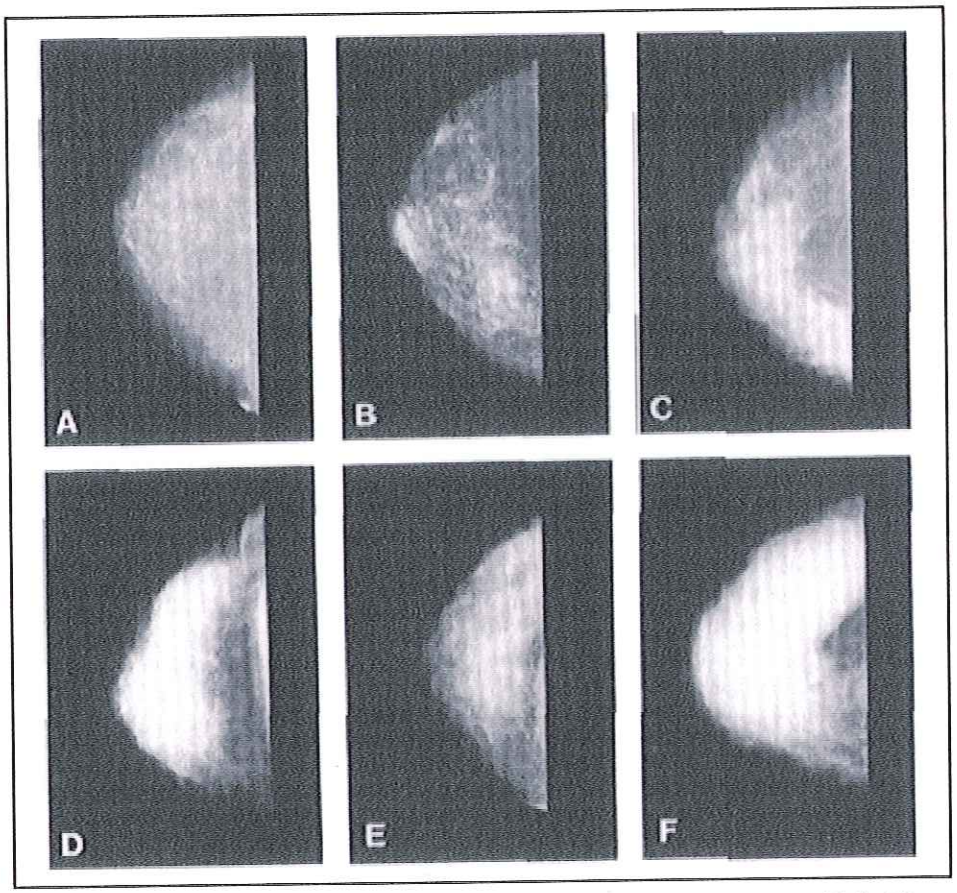


Figura 12 - Seis tipos de classificação da densidade mamária (SCC). A) Nenhuma Densidade; B) Densidade < que 10% da área total da mama; C) Densidade de 10-25% da área total da mama; D) Densidade de 25-50% da área total da mama; E) Densidade de 50-75% da área total da mama; F) Densidade > que 75% da área total da mama.

Fonte: BYNG (1994).

Em outro estudo Byrne ainda encontrou que o risco aumenta quando a área absoluta da mama com tecido denso aumenta. Segundo Byrne (2002) embora o padrão parenquimal inclua ambas aparência e densidade em suas definições, a extensão da medida de densidade como uma porcentagem da área da mama, discrimina melhor o risco de câncer. O autor também lembra que a densidade mamográfica dificulta a detecção do câncer e varia com base em características das mulheres conforme; idade, índice de massa corporal, reprodução, menopausa e uso de hormônios, ou também com o grau de compressão da mama, posicionamento na hora do exame e contraste da imagem.

A densidade da mama deveria ser incorporada no modelo do diagnóstico, assim, quando uma mulher com alta densidade mamária faz um mamograma, qualquer dúvida deveria ser vista como uma suspeita. O mesmo não precisa acontecer em casos onde a mulher possui baixo risco de desenvolver câncer.

Sabe-se que mamas densas são bastante comuns. Entretanto, embora a proporção de mulheres com câncer de mama que tem densidade aumentada é alta, o contrário não é verdadeiro, a proporção de mulheres com mamas densas que tem câncer de mama ainda é baixa, BYRNE (2002).

Segundo Heine e Malhotra (2002), há outros métodos para caracterizar densidade mamográfica em conjunto ou independentemente do padrão de WOLFE. O método mais usado envolve a medida da densidade radiográfica da imagem, isto é, análise dos tecidos que estão associados com a gordura (áreas radiolucentes).

2.3.4 DENSIDADE E PADRÃO DE WOLFE JUNTOS COMO FATOR DE RISCO

Alguns pesquisadores tem considerado risco, classificação de Wolfe e densidade simultaneamente. Brisson et al (1982a) relacionaram as porcentagens de densidades nodulares e homogêneas com o padrão de WOLFE:

- N1 = 0 % nodular e 0,4 % homogêneo;
- P1 = 13 % nodular e 1,9 % homogêneo;
- P2 = 45,9 % nodular e 3,4% homogêneo;
- DY = 24 % nodular e 36,7 % homogêneo.

Wolfe et al. (1987) encontraram que mama com característica:

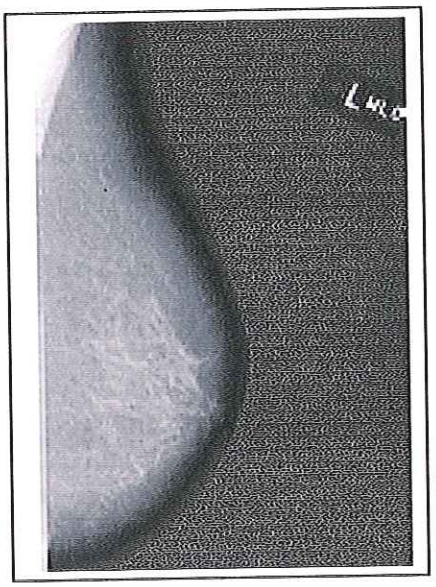
- N1 contém menos que 0,01 % de densidade,
- 80% de mamas P1 continham menos de 25% de densidade,
- 98% de mamas P2 continham no mínimo 25% de densidade
- Todas as DY continham mais de 25% de densidade.

2.3.5 PADRONIZAÇÃO DOS LAUDOS MAMOGRÁFICOS – BI-RADS

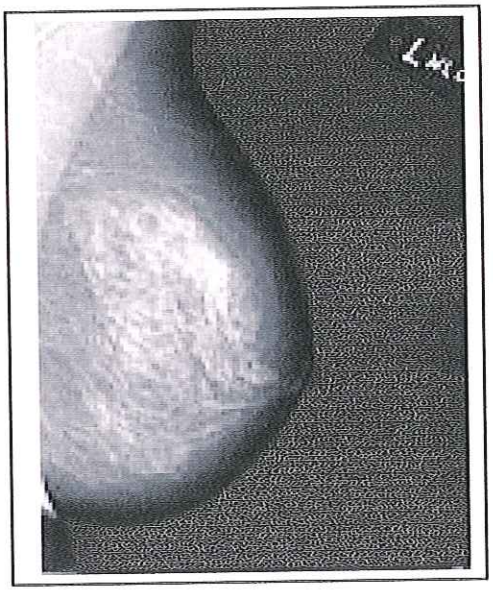
Em 1992, um estudo colaborativo entre membros de vários comitês do Colégio Americano de Radiologia juntamente com o Colégio Americano de Cirurgiões, com o objetivo de tornar mais clara a comunicação entre o radiologista e o médico solicitante, criou o modelo de padronização dos laudos de mamografia conhecido como BI-RADS (Breast Imaging Reprorting and Data System). Trata-se do mais completo sistema já elaborado para garantir a padronização do laudo mamográfico. Em 1998 a Terceira edição do BI-RADS foi lançada com algumas modificações. Desde 1999 esse padrão tem sido utilizado no Brasil. Nesta ultima edição encontra-se a descrição das características da densidade mamária baseada na atenuação média dos raios X pela lesão, em relação ao tecido fibroglandular. Ela permite indicar uma relativa possibilidade de haver uma lesão oculta no tecido normal. Geralmente se refere à consistência adiposa, mista ou densa.

Com a finalidade de uniformizar a classificação de densidade das mamas de todas as pacientes, devem ser consideradas as seguintes composições:

- **Composição I** – A mama é quase inteiramente adiposa (Figura 13);
- **Composição II** – Há áreas esparsas de densidades fibroglandulares (Figura 14);
- **Composição III** – O tecido mamário é heterogeneamente denso. (Figura 15);
- **Composição IV** – O tecido mamário é extremamente denso. O que pode mascarar lesões na mamografia (Figura 16).



(13)

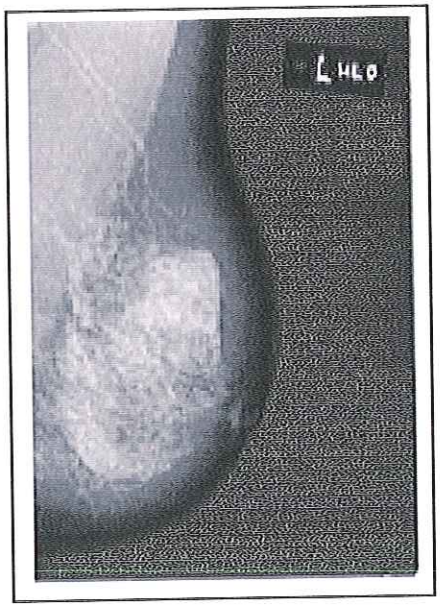


(14)

Figura 13 – Composição segundo BI-RADS – Composição I.

Figura 14 – Composição segundo BI-RADS - Composição II.

Fonte: (BOVIS, 2002)



(15)

(16)

Figura 15 – Composição segundo BI-RADS - Composição III,

Figura 16 – Composição segundo BI-RADS - Composição IV

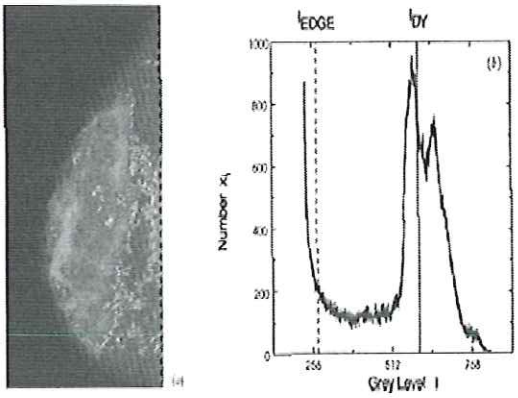
Fonte: (BOVIS, 2002)

CAPÍTULO 3

ESTADO DA ARTE - MÉTODOS COMPUTADORIZADOS PARA CLASSIFICAÇÃO DA DENSIDADE DAS IMAGENS MAMÁRIAS

3.1 CLASSIFICAÇÃO SEMI-AUTOMÁTICA DA DENSIDADE ATRAVÉS DE TÉCNICA DE LIMIARIZAÇÃO

Byng et al (1994) quantificaram e analisaram a densidade mamográfica utilizando técnica de Limiarização como ilustrada na Figura 17. Na digitalização da imagem, cada "pixel" tem um valor de nível de cinza i, que varia entre zero e i_{max}, onde i_{max} deve valer 1024. Para separar a imagem da mama do fundo, o valor do limiar é escolhido manualmente quando o gráfico da limiarização é mostrado na tela do computador.



(a) (b)

Figura 17 - (a) A linha pontilhada sinaliza a borda da mama, enquanto a linha sólida identifica o tecido fibroglandular. (b) O histograma mostra o limiar selecionando a i_{borda} e a densidade i_{DY}. A área no histograma i_{borda} é a área A, descrita na eq. (1).

Fonte: Byng et al (1994).

Obs: Áreas escuras na imagem têm um baixo valor de nível de cinza e contribuem para o lado esquerdo do gráfico. Áreas claras estão associadas com valores altos de níveis de cinza e contribuem para o lado direito do gráfico.

O segundo limiar i_{DY} é então selecionado para identificar a borda da região significativa de tecido denso na imagem, e é sinalizado na imagem (17b) pela linha sólida. A proporção da densidade radiográfica é calculada como a porcentagem PD, da área projetada dentro da mama, A, consistindo de "pixels" "densos" eq. (2).

$$A = \sum_{i=iborda}^{i\max} x_i \qquad (1)$$

$$PD = \left(\sum_{i=iborda}^{i\max} x_i \right) / A \times 100\% \qquad (2)$$

3.2 CLASSIFICAÇÃO AUTOMÁTICA DA DENSIDADE ATRAVÉS DE MEDIDA LOCAL DE ASSIMETRIA

Taylor et al (1984) propõem que o sistema computacional detecta anormalidades em todas as imagens, mas que desse maior atenção às imagens que representam mamas mais adiposas, que são as mais fáceis para leituras, deixando de fora as imagens que contem estruturas suspeitas, pois essas precisam de maior atenção dos radiologistas. Os autores para classificar os mamogramas em duas categorias: mamas adiposas e densas, utilizaram uma análise de textura local dividindo a imagem em quadrados, obtendo um valor estatístico para cada um e uma média geral para todos, fornecendo assim um meio de comparação. Os autores utilizaram uma janela de 128x128 "pixels", para ter um exemplo representativo do tecido em cada par da imagem, evitando o músculo peitoral, pele e mamilo. Para cada quadrado foram calculados, os graus de assimetria sobre um valor central da distribuição, o desvio padrão, a média, medidas de textura e fractais.

A medida local da assimetria do nível de cinza é o elemento básico do procedimento adotado. Taylor et al (1984) observaram que tecidos adiposos aparecem como uma estrutura pouco brilhante sob um fundo preto, dando uma distribuição de assimetria do nível de cinza em direção a parte escura do gráfico, enquanto que tecidos

densos são predominantemente brilhantes. Os autores concluíram que as duas texturas têm distribuição com assimetrias opostas, embora possam ter variações idênticas cujas medidas transmitam pouca informação. A seleção da região da mama foi desenvolvida usando seleção manual e em seguida um procedimento automático foi aplicado em toda a imagem. Pelo motivo do procedimento não ser totalmente automático foi necessário um conjunto de limiar para a segmentação da área da mama.

A curva ROC que relaciona as porcentagens de denso verdadeiro (proporção da imagem classificada pelo limiar como densa) com os falsos densos (proporção da imagem classificada pelos radiologistas como adiposas e como densa pelo limiar), foi usada para avaliar o método. Cada medida fornece um número único para cada quadrado da imagem. Os resultados de Taylor et al (1984) indicam que uma medida automática baseada na assimetria local poderia ser usada para separar 2/3 das imagens adiposas diminuindo a sobrecarga dos radiologistas.

3.3 CLASSIFICAÇÃO SEMI-AUTOMÁTICA DA DENSIDADE (WOLFE) ATRAVÉS DA ANÁLISE DE FOURIER

Tahoces et al (1995) desenvolveram um método computadorizado para separar 117 mamogramas nos quatro padrões descritos por WOLFE (1976a); N1, P1, P2 e DY.

Os autores dividiram os mamogramas em três regiões (Figura 18), e aplicaram um algoritmo que calcula o gradiente dos níveis de cinza para detectar a borda da região II. As bordas das regiões I e III foram marcadas manualmente.

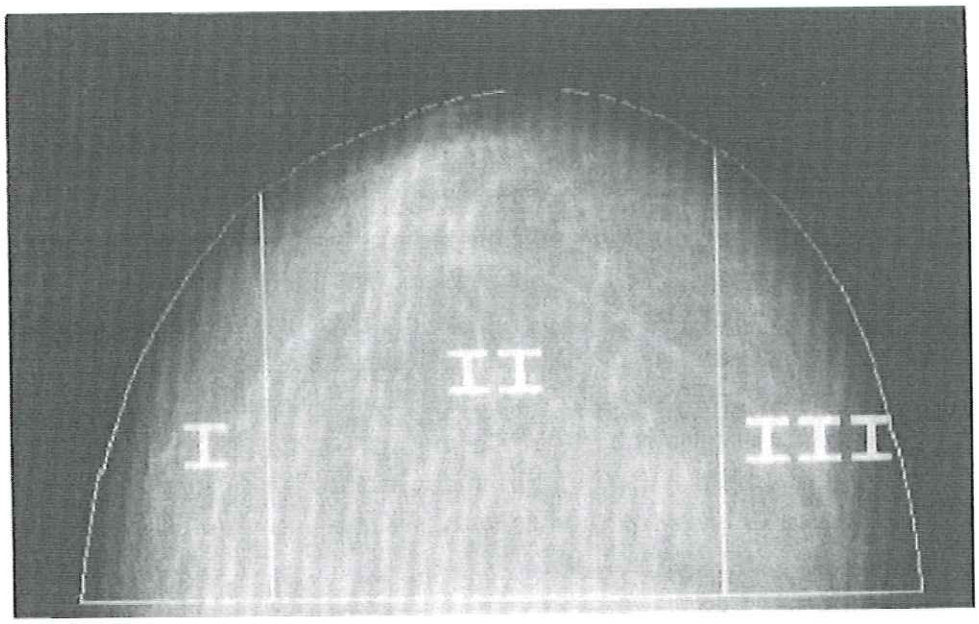


Figura 18 – Divisão da mama em três Regiões.
Fonte: TAHOCES (1995).

Após o cálculo da borda da região II, dois tipos de regiões diferentes foram selecionados pelo computador automaticamente referentes à borda da mama, como mostra as Figuras 19a e b, com matrizes 64 x 64, localizada na parte medial (região III, Fig. 18) e central da mama (região II, Fig. 18). Para se fazer à escolha das outras regiões que deveriam ser observadas foi feito um estudo em 500 mamogramas (CC) para localizar a presença ou ausência do tecido parenquimal, Figuras 19c a f.

Os autores utilizaram três algoritmos para separar as imagens de acordo com as regiões.

1. Algoritmo baseado na Transformada de Fourier;
2. Relação espacial através dos níveis de cinza;
3. Algoritmo baseado em valores absolutos dos níveis de cinza.

Os dois primeiros foram aplicados em regiões de interesse, selecionando algumas características do tecido glandular como, por exemplo, a arquitetura do tecido. Foram obtidos valores altos para o primeiro momento da Transformada de Fourier no padrão DY e diferentes contrastes entre P1 e P2 quando computados em uma região medial da mama (região III). O terceiro tipo de algoritmo foi aplicado em regiões irregulares de interesse, tentando separar mamas densas (P2 e DY) de adiposas (N1 e P1).

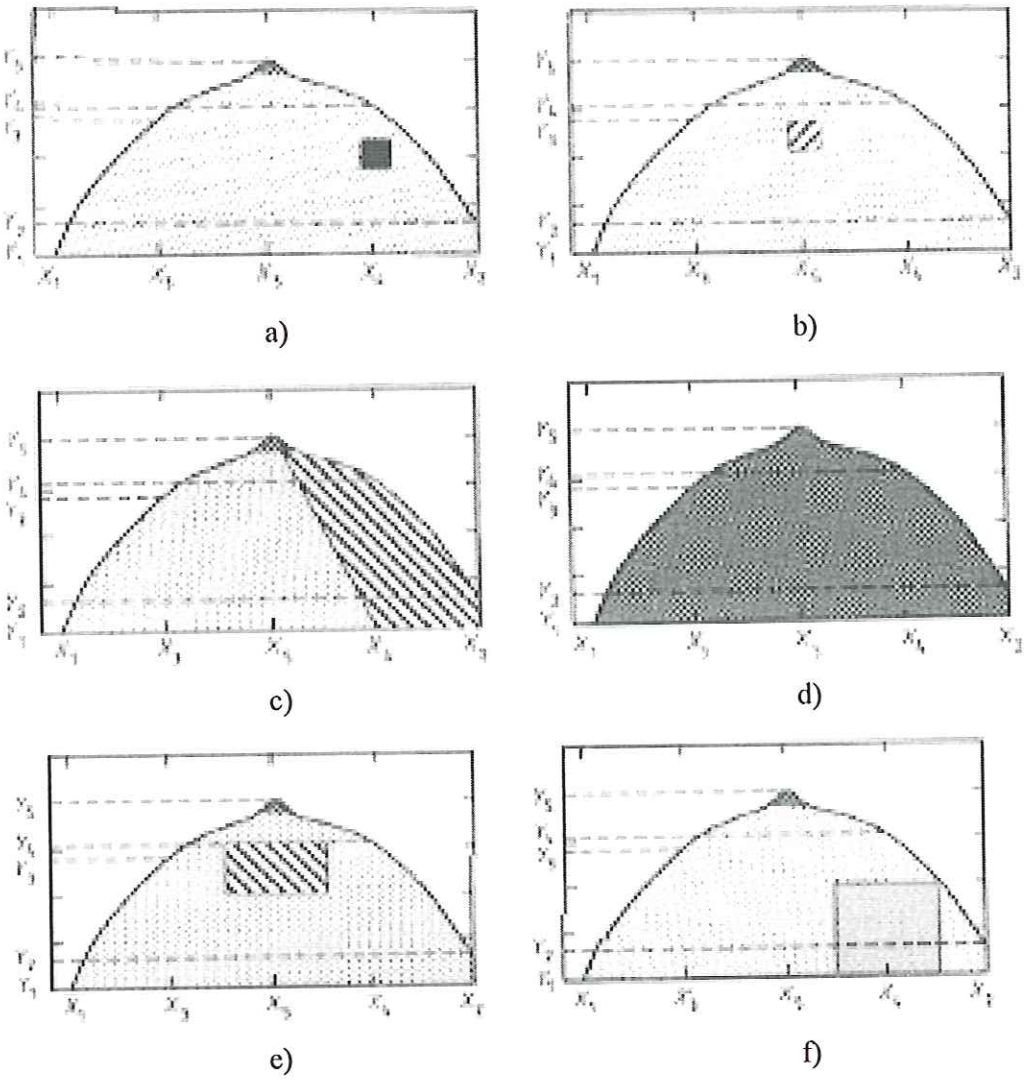


Figura 19 – Regiões da mama selecionadas pelo computador para serem analisadas.
Fonte: TAHOCES (1995).

Os resultados obtidos automaticamente foram comparados com os obtidos por cinco radiologistas que também haviam classificado as mesmas imagens de acordo com os padrões de WOLFE. A concordância foi moderada para os padrões N1 e P2, razoável para P1 e boa para DY.

3.4 CLASSIFICAÇÃO AUTOMÁTICA DA DENSIDADE ATRAVÉS DE LIMIARIZAÇÃO E TÉCNICAS DE FRACTAIS

Byng et al (1996) realizaram uma classificação da densidade de maneira automática, utilizando limiarização e técnicas de fractais. Para os autores a análise da densidade mamográfica em imagens digitais está relacionada basicamente com o brilho e a variação da textura. A dimensão fractal foi calculada para caracterizar a textura inerente na região da imagem. A Figura 20 mostra a imagem de uma mama com um pequeno grau de densidade podendo ter uma aparência de textura "aspera" para bom contraste entre o tecido conectivo e o fibro-glandular. Similarmente, quando há alto grau de densidade mamográfica a imagem pode aparecer "reprimida" como deve ser refletida por uma dimensão de fractal baixa.

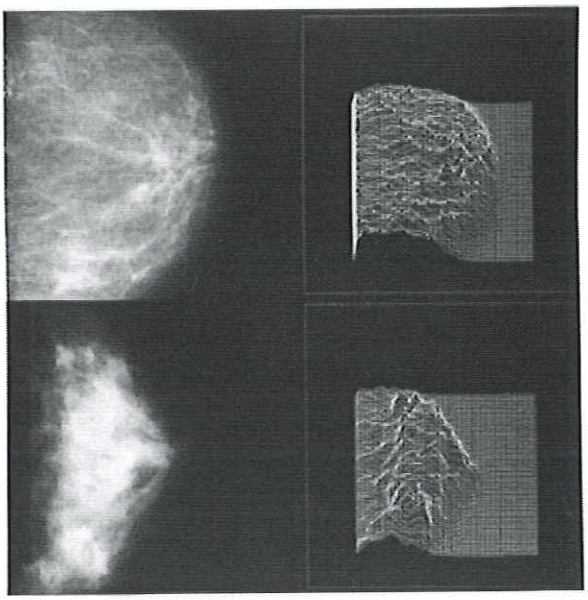


Figura 20 – Imagem de uma superfície que pode ser associada com uma imagem de mama. Na representação digital, o nível de cinza de cada "pixel" determina a altura da superfície. As imagens de cima representam densidades menores, enquanto as de baixo representam maior densidade.
Fonte: (BYNG 1996).

Os autores realizaram uma parametrização do histograma, o qual está representando a variação e distribuição do brilho da imagem cuja limiarização pode ser calculada de maneira completamente automática. A forma do histograma pode ser representada pelo momento de distribuição sobre a média dos níveis de cinza. Byng et al (1996) utilizaram o terceiro momento para caracterizar a assimetria do histograma.

Para calcular a dimensão fractal, o brilho i de cada "pixel" (x, y) em uma imagem digitalizada pôde ser tratado como uma dimensão vertical acompanhada de duas dimensões fixadas do "pixel" no processo de digitalização.

No procedimento de limiarização, a identificação da borda está baseada na seleção do limiar de um nível de cinza $i_{(borda)}$ separando a mama do fundo da imagem. O histograma é calculado para "pixels" tendo $i > i_{(borda)}$. A medida de assimetria considera que a região do tecido fibroglandular como aparece mais brilhante que a do tecido adiposo, deve tender em direção ao mais alto nível de cinza, como ilustra a Figura 21.

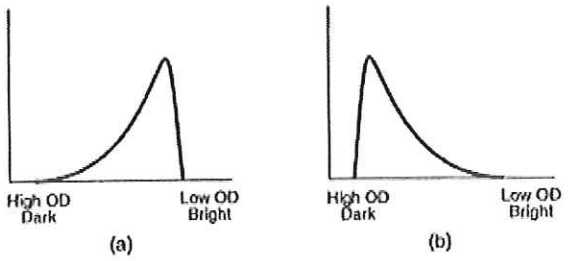


Figura 21 – Distribuição da assimetria, a) negativo e b) positivo.

Fonte: BYNG (1996).

Para regiões de alta densidade a distribuição é de uma assimetria negativa, Figura 21a, e para regiões de pouca densidade a distribuição é de uma assimetria positiva, Figura 21b. Medidas de assimetria aplicadas em pequenas regiões da mama são mais sensíveis para composições de tecidos do que para uma região global. Para isso a região da mama foi dividida em pequenas regiões de 24 x 24 "pixels".

Os autores calcularam para cada região um valor de assimetria, e com esses valores obtiveram o valor médio para cada imagem de mama. As regiões não densas devem ter valores mais positivos de assimetria, ou seja, quanto mais positivo for o valor de assimetria mais tecido adiposo tem a mama e quanto mais negativo for o valor de assimetria mais tecido fibroglandular tem a mama. Os autores utilizaram um classificador Bayesiano para separar as imagens no padrão de Boyd. Um teste de concordância entre as classificações feitas pelos radiologistas e pelo método automático apresentou um resultado de aproximadamente 95%.

3.5 CLASSIFICAÇÃO AUTOMÁTICA DA DENSIDADE ATRAVÉS DA TÉCNICA AUTOMÁTICA DE LIMIARIZAÇÃO GLOBAL

Karssemeijer (1998) desenvolveu um método para classificar 615 mamogramas na posição ML (médio lateral) em quatro categorias de densidades, correspondentes a porcentagem de densidade, <5, 5-25, 25-75 e 75 a 100. Para a segmentação do fundo da imagem, o autor aplicou uma técnica automática de limiarização global, ressaltando, entretanto um problema sério com a característica do histograma que é sua dependência da aquisição da imagem, tanto da curva característica do filme quanto da compressão da mama.

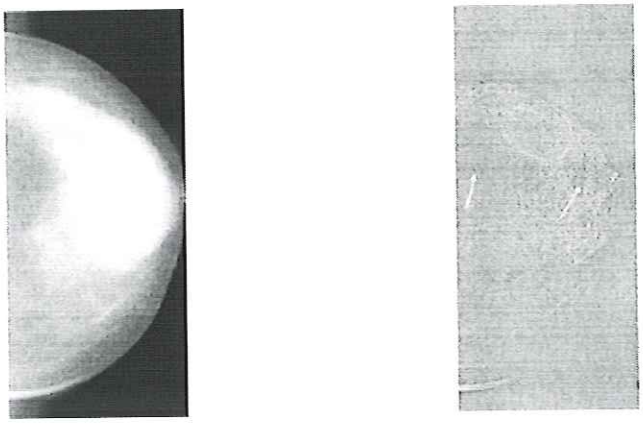
Karssemeijer (1998) calculou o desvio padrão e a assimetria do histograma, nas distâncias 20, 40, 70 e 90, entre o músculo peitoral e a pele, usando o músculo peitoral como referência para a interpretação de regiões de densidades, sendo que regiões menos brilhantes que o músculo peitoral foram consideradas como tecidos adiposos e regiões de brilhos similares foram entendidas como tecidos fibroglandulares. O classificador KNN foi usado para classificar o padrão parenquimal com as características mencionadas acima. Foram treinados dois classificadores um com e outro sem as características do músculo peitoral. Como resultado em 67% dos casos houve uma concordância entre o classificador e a resposta dos radiologistas. Para mamogramas mais recentes, feitos após 1991, foi obtida uma concordância de 80%.

3.6 CLASSIFICAÇÃO AUTOMÁTICA DA DENSIDADE ATRAVÉS DE UMA METODOLOGIA ESTATÍSTICA BASEADA EM UMA ANÁLISE DE PROBABILIDADE UTILIZANDO O χ^2

Neste trabalho Heine e Velthuizen (2000) utilizaram uma metodologia estatística baseada em uma análise de probabilidade $\chi 2$ (qui-quadrado) que permite a separação automatizada do tecido adiposo e do tecido fibroglandular em mamogramas digitalizados seguindo a classificação do parênquima mamário descrita por (WOLFE, 1976a).

Os autores desenvolveram um modelo onde à imagem é derivada de um processo de filtragem linear. O modelo é facilmente expresso no domínio de Fourier. A

Figuras 22a ilustra uma imagem de mama normal com densidade visível e a 22b mostra a imagem com uma transformada logarítmica aplicada.

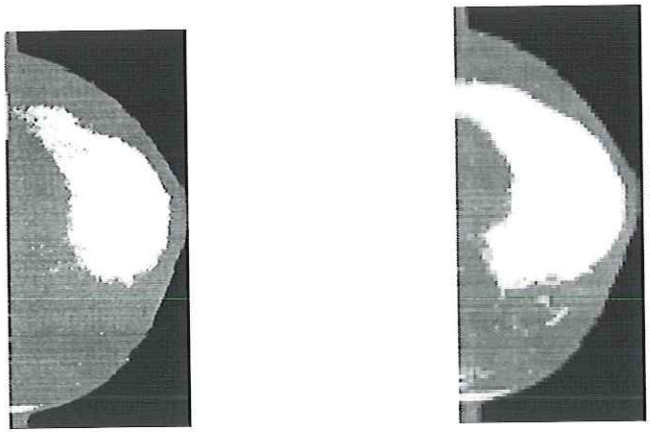


a) b)

Figuras 22 – a) imagem normal de mama. b) imagem com transformada logarítmica aplicada, note que a região apontada com a seta da esquerda representa área de tecido adiposo, enquanto a seta da direita tecido denso.

Fonte: HEINE e VELTHUIZEN (2000).

Para separar tecido adiposo de tecido fibroso os autores estimaram a variância global por comparações estatísticas baseadas em análise de probabilidade $\chi 2$. Heine e Velthuizen (2000) construíram uma matriz de 8 x 8mm que percorria toda a imagem para fazer o calculo da variância local. O limiar é derivado da distribuição do $\chi 2$ com (n+1) grau de liberdade e nível de significância de 10%. O resultado da Figura 22a para um nível de significância SL = 0,10 é mostrado na Figura 23, onde a densidade mamográfica está em branco.



a) b)

Figuras 23 – a) Resultado da primeira interação com o SL = 0,10, b) Resultado da segunda interação com o SL = 0,001.

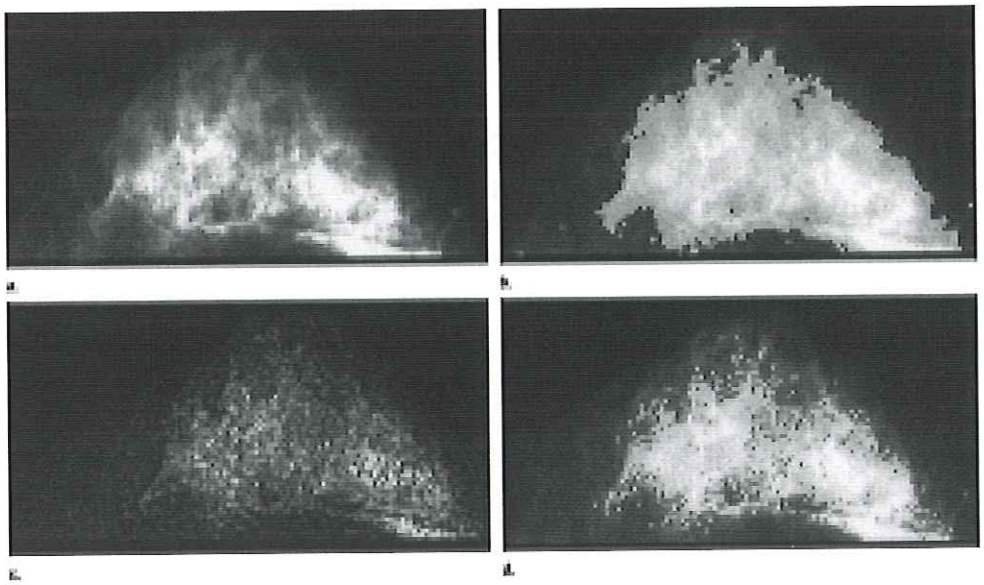
Fonte: HEINE e VELTHUIZEN (2000).

3.7 CLASSIFICAÇÃO AUTOMÁTICA DA DENSIDADE ATRAVÉS DO LIMIAR ÓTIMO DE KITTLER E ILLINGWORT'S

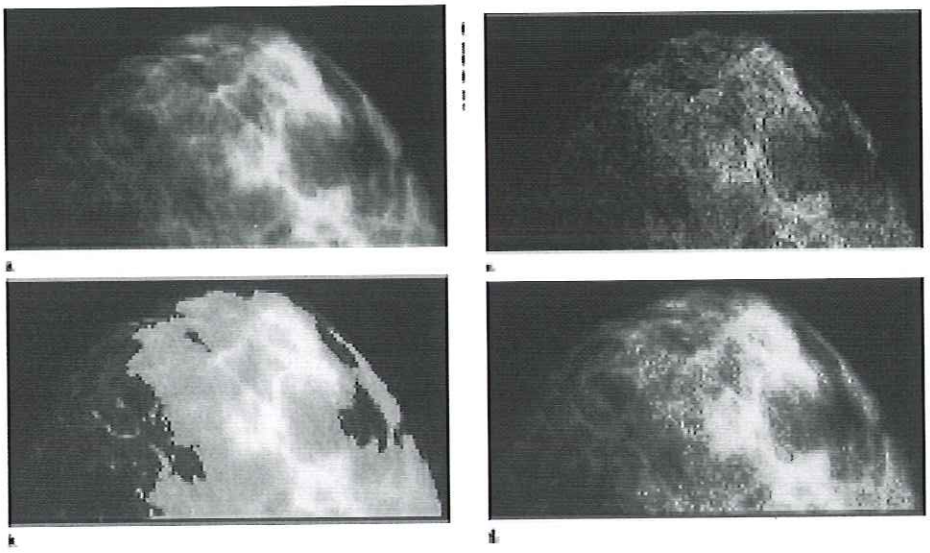
A proposta de Sivaramakrishna et al (2001) é de avaliar um método completamente automático para estimar a densidade mamária, baseada no limiar ótimo de Kittler e Illingwort's (1986) que é uma técnica de segmentação baseada em histograma envolvendo a hipótese que o objeto (mama) e o fundo da imagem podem ser distinguidos pelos seus valores de níveis de cinza. O algoritmo calcula automaticamente um valor de limiar ótimo pela maximização da variância dos níveis de cinza entre a mama e o fundo. Os autores realizaram uma conversão para a variação da imagem, a qual é calculada como segue: Em todo "pixel" da imagem *M*, a variação de *V* normalizada foi calculada sobre uma janela de tamanho (2N+1), onde *V* é dado pela eq. (3):

$$V(I,J)=\frac{1}{S}\sqrt{\sum_{i=-N}^{N}\sum_{j=-N}^{N}\left(\frac{M(i,j)-M(I,J)^2}{M(I,J)}\right)} \qquad (3)$$

e *S* é o numero total dos "pixels" da vizinhança que fazem parte da região da mama. Todo "pixel" correspondente à região da mama foi dividido pelo valor correspondente de *V* e o limiar ótimo de Kittler e Illingwort's foi calculado sobre essa variação da imagem. Essa conversão mostra que quando existem altos valores de "pixels" com baixos valores de *V* a variação da imagem é convertida para altos valores de níveis de cinza. Sivaramakrishna et al (2001) calcularam a porcentagem de densidade como a razão do numero de "pixels" na imagem segmentada pelo numero de "pixels" da área da mama multiplicada por 100. Esse algoritmo foi utilizado em 32 mamogramas e em todos os casos a segmentação foi considerada visualmente muito boa. A performance desse algoritmo foi comparada com a técnica da limiarização global. As Figuras 24a e 25a são imagens originais; 24b e 25b são imagens originais com a segmentação produzida por um limiar regular; 24c e 25c são imagens com variação e 24d e 25d imagens com segmentação pelo algoritmo desenvolvido. Essas figuras mostram claramente que o limiar regular global inclui tecido adiposo, em compensação o algoritmo desenvolvido segmenta somente regiões densas da mama.



Figuras 24 – a) Mamograma de uma mulher de 71 anos, b)segmentação da densidade por limiar regular global, c) Variação da imagem, d) Segmentação da mesma imagem pelo método desenvolvido.
Fonte: SIVARAMAKRISHNA et al (2001).



Figuras 25 – a) Mamograma de uma mulher de 43 anos, b)segmentação da densidade por limiar regular global, c) Variação da imagem, d) Segmentação da mesma imagem pelo método desenvolvido.
Fonte: SIVARAMAKRISHNA et al (2001).

As mamografias segmentadas pelo método proposto foram fotografadas e mostradas para radiologistas experientes. Uma correlação aproximadamente de 92% foi encontrada entre o método e os radiologistas experientes.

3.8 CLASSIFICAÇÃO AUTOMÁTICA DA DENSIDADE ATRAVÉS DE ESTATÍSTICAS DA IMAGEM, MÉTODO DA VARIAÇÃO, MORFOLOGIA MATEMÁTICA, TEXTURA DE BORDA, E SUBTRAÇÃO GAUSSIANA.

Segundo Chang et al (2002) a tarefa de classificação do tecido mamário é bastante útil, porém ela é muito subjetiva e relativamente grosseira em conseqüência a grande variação entre e inter observadores. Para superar algumas dessas limitações Chang et al (2002) propuseram uma aproximação computacional que fornece uma quantificação objetiva e reproduzível da composição do tecido mamário.

Três radiologistas classificaram 200 mamogramas digitais em 4 composições (BI-RADS) com escalas de 0 a 4 e incremento de 0,1. Sendo que para pertencer a Composição I (tecido quase inteiramente gorduroso), as imagens deveriam ter sido classificadas de 0 a 1,0; Composição II (tecido fibroglandular disperso) de 1,1 a 2,0; Composição III (tecido heterogeneamente denso de 2,1 a 3,0 e Composição IV (tecido extremamente denso) de 3,1 a 4,0. Os três conjuntos de resultados foram utilizados para estabelecer um consenso para a avaliação e desenvolvimento do método computacional. O maior problema não foi estabelecer um valor para as Composições 1 e 4, mas sim para a Composição 2 e 3.

Antes da análise das características, os autores utilizaram um método de segmentação semi-automático para remover áreas de fundo e selecionar a região de interesse R para cada imagem.

Para avaliar quantitativamente a composição do tecido da mama, os autores extraíram várias características usando: estatísticas da imagem, método da variação, morfologia matemática, textura de borda, e subtração gaussiana.

- Estatísticas da imagem - desvio padrão, assimetria e curtose, foram calculadas para a região de interesse R.
- Método da variação - Foi calculada a máxima variação do "pixel" dentro de uma área *B* (área circular) em relação aos seus vizinhos.
- Morfologia Matemática – foi calculado dilatação e erosão da imagem.
- Textura de borda – O valor de gradiente de cada "pixel" é o *nonoverlapping neighborhoods* e foi usado para estimar o valor de gradiente total na região R. Cada "pixel" é então associado a um valor de gradiente.

- Subtração gaussiana – Cada imagem foi filtrada usando um conjunto de 5 funções Gaussianas. Essas imagens foram subtraídas "pixel" a "pixel" das imagens originais para gerar 5 imagens filtradas por filtros passa baixa.

Em resumo os autores investigaram 4 características estatísticas dos valores do "pixel" e 4 atributos para caracterizar a composição do tecido.

A região R foi dividida em duas regiões, uma contendo a região mais densa (R1) e outra a menos densa (R2). Valores de limiares foram usados para determinar a área que converge para mais densa (R1) em cada mamograma. Para cada mamograma, a média de todos os valores dos "pixels" dentro da área *R* foi selecionada como limiar, então 50% dos "pixels" foram designados para a região R1 e 50% para a região R2.

Usando um conjunto de características Fk, $k = 0,1,2,...,M$ os autores investigaram um índice de inferência (I) para cada imagem definida como segue a eq. (4):

$$CI(i) = \sum_{K=0}^{M} Fk \times Wk \qquad (4)$$

onde Wk tem peso diferente para cada característica, $F_0 = 1$ e M é o número de característica.

Todas as características foram analisadas individualmente para determinar se elas mostravam correlação (coeficiente Pearson) com a classificação feita pelos radiologistas.

Os 200 mamogramas foram divididos aleatoriamente em dois grupos de 100, uma para treinamento e outro para teste. Do grupo para treinamento foi dividido em 4 subgrupos de imagens 20, 40, 60, 80 e 100.

Para se obter um consenso entre os radiologistas, foi feito uma correlação entre eles. Essa correlação por ter sido estatisticamente significante foi usada como um consenso na classificação dos mamogramas. O resultado desse consenso foi comparado com o método computadorizado e obteve-se um coeficiente de correlação de $r = 0,88$ no conjunto de treinamento e $r = 0,87$ para o conjunto de teste. Em 14 pacientes (7%) o programa não trabalhou satisfatoriamente e foi preciso traçar a região R manualmente.

CAPÍTULO 4

METODOLOGIA

Neste capítulo apresentamos o método desenvolvido para a classificação das mamografias, em função da composição da densidade do padrão BI-RADS, a descrição dos bancos de imagens e dos algoritmos desenvolvidos para a classificação e exibição dos resultados.

4.1 BANCO DE IMAGENS

4.1.1 BANCO DE IMAGENS DA FACULDADE DE MEDICINA DE RIBEIRÃO PRETO (FMRP)

A base de dados do Hospital Universitário de Ribeirão Preto pertencente à Faculdade de Medicina de Ribeirão Preto (FMRP – USP) que utilizamos, possui 320 casos de pacientes que fizeram exames mamográficos, ou seja, 1280 imagens, sendo duas imagens de cada mama - uma na posição CC e outra na ML. Laudos mamográficos acompanham as imagens fornecendo informações como código de cada paciente, data do exame, classificação na categoria e na composição da densidade mamária de acordo com (BI-RADS), conclusão e descrição feita pelo médico.

A digitalização das imagens foi realizada com um "scanner" Vidar modelo DiagnosticPro com 300dpi e 12 bits de quantização do Laboratório do Prof. Dr. Paulo M. de A. Marques (FMRP-USP).

Para o estudo selecionamos as imagens na posição CC de apenas uma mama, pois Byng et al (1996) mostraram que um valor representativo da densidade mamária

pode ser obtido com somente uma imagem da paciente. Utilizamos para desenvolver o algoritmo um banco de treinamento composto de 152 imagens classificadas pela equipe médica de acordo com o padrão da Composição de densidade descrito no BI-RADS da seguinte forma:

- Composição I – 67 imagens
- Composição II – 39 imagens
- Composição III – 29 imagens
- Composição IV – 17 imagens.

No banco da FMRP 120 imagens não tinham sido classificadas. Para testar a validade do algoritmo utilizamos o método desenvolvido para separar essas imagens em função das 4 Composições de Densidades descritas no BI-RADS. Chamamos essa segunda parte do banco da FMRP de banco de Teste. Os resultados obtidos com o banco de testes foram avaliados por um médico da FMRP.

4.1.2 BANCO DE IMAGENS DA UNIVERSIDADE DE GIFU (JAPÃO) - FUJILAB

Para validar o método utilizamos também o Banco do Laboratório do Prof. Dr. Fujita da Universidade de Gifu (Japão), que consiste de 132 imagens mamográficas obtidas de 33 pacientes, em 4 projeções ML (direita e esquerda) e CC (direita e esquerda). Utilizamos uma projeção Crânio Caudal (CC) de cada paciente sendo no total 33 imagens. A classificação da Composição da densidade mamária (BI-RADS) dessas imagens foi feita por um médico radiologista do Hospital das Clínicas de Ribeirão Preto da seguinte forma:

- Composição I – 03 imagens
- Composição II – 11 imagens
- Composição III – 09 imagens
- Composição IV – 10 imagens

4.2 MÉTODOS DE PRÉ PROCESSAMENTOS UTILIZADOS

4.2.1 DIVISÃO DAS IMAGENS EM REGIÕES

Inicialmente desenvolvemos um algoritmo que divide a imagem mamográfica em janelas para obter as características de pequenas regiões. As janelas permitem não ter de processar, todos os "pixels" da imagem (valor global), o que tem um custo computacional elevado. Utilizamos o MATLAB, pois possui um ambiente genérico de computação cientifica, linguagem de programação simples, compacta e muito próxima da linguagem matemática, assim como ferramentas importantes na área de processamento de imagens. A divisão em "janelas" (pixels) de 5x5, 10x10, 20x20, 25x25, 30x30, 50x50 e 100x100 foi feita de maneira automática de acordo com o tamanho de cada imagem. Só informamos para o algoritmo o tamanho da janela, em seguida o programa divide a largura e a altura da imagem pelo número de "pixels" da janela desejada. Por exemplo, uma janela de 10x10 "pixels" em uma imagem de 1024x780 "pixels" fornece 7956 regiões, 102 regiões na largura e 78 na altura.

4.2.2 DETECÇÃO DA BORDA DA MAMA

Iniciamos o processo com a detecção da região da mama realizando a binarização global da imagem a partir de informações locais das regiões. Neste procedimento utilizamos o método automático de detecção de limiar de Wong & Sahoo (1989), que determina o valor ótimo baseado na maximização da entropia de Shannon (1948) para calcular os limiares de cada janela. De posse desses valores, que podem ser observados como uma distribuição de limiares, calculamos o desvio padrão que utilizamos como valor fronteira fundo-mama para binarização da imagem.

A entropia total é um indicativo importante da quantidade de informação presente na região, ou seja, quanto maior o seu valor, maior também a presença de formas não pertencentes ao fundo, indicando assim região não uniforme. Segundo Wiener apud Edwards (1971), a entropia de um sistema é a medida de seu grau de desorganização. Surgindo então a oportunidade de medir a quantidade de

desorganização, ou quantidade de incerteza, e utilizá-la como quantidade de informação presente em uma imagem, para que a mesma possa servir como um critério de decisão.

O método que desenvolvemos está baseado na entropia *H da* eq. (5).

$$H = -\sum_{i=1}^{n} p_i \log p_i + (1 - p_i)\log(1 - p_i) \qquad (5)$$

a qual apresenta duas possibilidades de probabilidades que são *p* e *(1-p)*, utilizadas para expressar a probabilidade do "pixel" pertencer ao fundo ou não. Esse método encontra o valor de limiar para cada região da imagem da mama, quando a entropia da mesma é máxima.

A REGIÃO DE FRONTEIRA, é a região que contém ao mesmo tempo a imagem da mama e o fundo da imagem.

Essas regiões são determinadas pela seguinte regra:

[(Entropia de região > 1) e (Limiar de região = 1)]

A REGIÃO DE FUNDO DE IMAGEM obedece à seguinte regra:

[(Entropia de região <= 1) e (Limiar de região = 1)]

Otimizamos essas informações para binarizar a imagem sendo que todo "pixel" pertencente ao fundo possui valor igual a *"0"* e todo "pixel" pertencente à mama recebeu valor igual a "1". Realizamos uma operação *AND* entre duas imagens a binarizada e a original. A imagem resultante desse procedimento simples é uma imagem filtrada, ou seja, o fundo é separado da região de interesse (mama) e passa a ter nível de cinza zero (preto total).

Após esse processo realizamos uma varredura na imagem para encontrar a borda da mama. O algoritmo inicia sua leitura no canto esquerdo superior da imagem, realizando uma busca por "pixels" que possuam níveis de cinza diferentes de zero, ou seja, diferentes do valor de fundo. Uma vez encontrados, o algoritmo coloca um ponto branco nesses "pixels", determinando assim a borda da mama em toda a imagem. Esses "pixels" brancos não irão interferir nos resultados de processamento, pois são colocados apenas para visualização em uma imagem espelho.

Na imagem da mama introduzimos os valores das estruturas da mamografia original.

4.3 MÉTODOS DE CLASSIFICAÇÃO DE DENSIDADES UTILIZADOS

Para a classificação da densidade, utilizamos primeiro parâmetros estatísticos, como os resultados alcançados não foram maiores que 70% desenvolvemos um algoritmo que segue o padrão do radiologista quando examina um mamograma.

4.3.1 EXTRAÇÃO DOS PARÂMETROS ESTATÍSTICOS

Para a extração de informações de cada região utilizamos primeiramente duas abordagens estatísticas de primeira ordem relacionadas com a distribuição dos níveis de cinza da imagem. As características de primeira ordem são: desvio padrão, média, coeficiente de assimetria e curtose. Calculamos as 52 características estatísticas relacionadas na TABELA 1.

TABELA 1- Características estatísticas da imagem

Siglas	Significados
1-A- Nome	Nome da Imagem
2-B- xfront(1)	Região de Fronteira da mama
3-C- xent_0(1)	Região onde a entropia é zero
4-D- xa(1)	Região onde há densidade na mama
5-E- xb(1)	Região de tecido adiposo
6-F- razao_ab	Razão entre a região de densidade e a de tecido adiposo
7-G- porc	Porcentagem da mama ocupada por densidade
8-H- somareg(1)	Soma de todas as "janelas"
9-I- med_reg_densa	Média do nível de cinza da região densa
10-J- dsv_reg_densa	Desvio médio do nível de cinza da região densa
11-K- med_reg_adiposo	Média do nível de cinza da região de tecido adiposo
12-L- dsv_reg_adiposo	Desvio médio do nível de cinza da região de tecido adiposo
13-M- acc_med_total	Média dos níveis de cinza total da mama
14-N- media_thr_mama	Media do "Threshold" total da mama
15-O- v_med_enta	Media da entropia da região densa da mama
16-P- v_desv_enta	Desvio da entropia da região densa da mama

Siglas	Significados
17-Q- v_med_entb	Media da entropia da região adiposa da mama
18-R- v_desv_entb	Desvio da entropia da região adiposa da mama
19-S- v_med_thra	Media do "Threshold" da região densa da mama
20-T- v_desv_thra	Desvio do "Threshold" da região densa da mama
21-U- v_med_thrb	Media do "Threshold" da região adiposa da mama
22-V- v_desv_thrb	Desvio do "Threshold" da região adiposa da mama
23-W –ent_thr_a	Entropia dos"Threshold" da região densa da mama
24-X- ent_thr_b	Entropia dos "Threshold" da região adiposa da mama
25-Y- cov_thr_alta	Covariância do "Threshold" da região densa da mama
26-Z- cov_thr_baixa	Covariância do "Threshold" da região adiposa da mama
27-AA- var_thr_alta_thr	Variância do "Threshold" (peso=thr) da região densa da mama
28-AB- var_thr_alta_ent	Variância do "Threshold" (peso=ent) da região densa da mama
29-AC- var_thr_baixa_thr	Variância do "Threshold" (peso=thr) da região adiposa da mama
30-AD- var_thr_baixa_ent	Variância do "Threshold" (peso=ent) da região adiposa da mama
31-AE- cov_ent_alta	Covariância da entropia da região densa
32-AF- cov_ent_baixa	Covariância da entropia da região adiposa
33-AG- var_ent_alta_thr	Variância da entropia (peso=thr) da região densa da mama
34-AH- var_ent_alta_ent	Variância da entropia (peso=ent) da região densa da mama
35-AI- var_ent_baixa_thr	Variância da entropia (peso=thr) da região adiposa da mama
36-AJ- var_ent_baixa_ent	Variância da entropia (peso=ent) da região adiposa da mama
37-AK- ent_med	Media da Entropia
38-AL- ent_median	Mediana da Entropia
39-AM- desvent	Desvio da Entropia
40-AN- cvp_ent	Coeficiente de Variação de Pearson da entropia
41-AO- As_ent	Assimetria da entropia
42-AP- thr_med	Media dos "Thresholds"
43-AQ- thr_median	Mediana do "Threshold"
44-AR- desvthr	Desvio do "Threshold"
45-AS- cvp_thr	Coeficiente de Variação de Pearson do "Threshold"
46-AT- As_thr	Assimetria do "Threshold"
47-AU- tote1	Entropia da distribuição das entropias da mama toda
48-AV- toth1	Entropia da distribuição dos "Threshold" da mama toda
49-AW- var_thr1	Variância do "Threshold" (peso=ent)
50-AX- var_thr1_thr1	Variância da entropia (peso=thr) da região densa da mama
51-AY- var_tot1	Variância da entropia (peso=ent) da região adiposa da mama
52-AZ- var_tot1_tot1	Variância da entropia (peso=ent) da região adiposa da mama

4.3.2 CLASSIFICAÇÃO COM FASTMAPDB

O *FastMapDB* desenvolvido por Traina Jr et al. (1999) e Traina et al, (2001) permite visualizar os resultados da combinação dos parâmetros descritos no item anterior para classificar as imagens nas Composições I, II, III, e IV de densidade mamária BI-RADS. O objetivo da ferramenta é gerar mapeamentos de dados armazenados em bases de dados relacionais permitindo a visualização tri-dimensional.

Essa ferramenta foi desenvolvida com o intuito de possibilitar ao usuário "ver" a distribuição dos dados sem basear-se em qualquer propriedade espacial intrínseca possivelmente presente neles. Ela permite, por exemplo, verificar a existência de "outliers", verificar a formação de agrupamentos ("clusters") e auxiliar o usuário a escolher conjuntos reduzidos de atributos para minerar. A ferramenta possui recursos para que o usuário crie interativamente uma função de distância vetorial, a partir de qualquer quantidade dos atributos de uma ou mais relações de uma base de dados conectada via BDE (*Borland Database Engine*) ou ODBC (*Open Database Connectivity*). Os atributos utilizados pela função de distância podem ser quaisquer dados, sejam eles numéricos, textuais ou datas, e podem ser ponderados, normalizados, e/ou utilizados em escala linear ou logarítmica.

4.3.3 MÉTODO DE CLASSIFICAÇÃO SEMELHANTE AO USADO PELO MÉDICO

4.3.3.1 O MÉTODO

As características que o médico analisa para que a mamografia seja classificada nas Composições de densidades descritas no BI-RADS, podem ser atendidas pelo algoritmo como descrito a seguir:

- **Composição I**

A mama para ser classificada na Composição I é considerada pelo observador como "quase inteiramente adiposa". Na mamografia, a mama pode ser considerada

adiposa se a distribuição dos níveis de cinza da sua imagem é uniforme. Sendo que os níveis de cinza devem ser próximos aos níveis do fundo da imagem (embora esses valores variem com a exposição e a espessura da mama). A imagem só pode apresentar áreas esparsas de tecido glandular. Essas áreas são representadas na mamografia por manchas bem mais claras que o resto do disco glandular. Com base no preenchimento de regiões, chegamos a eq. (6) que separa corretamente essas manchas. A adequação da formula foi verificada por observação visual dos resultados.

$$\text{Nível para densidade 1} = \text{THRMÉDIO} + \text{DESVTHR} \qquad (6)$$

onde, THRMÉDIO é o valor médio de limiar de toda a imagem e DESVTHR é o desvio padrão dos valores de limiar.

Os médicos consideram como Composição I as imagens cujo percentual de densidade 1 ocupada na mama é pequeno. Como esse critério é subjetivo, analisamos no banco laudado, qual é a porcentagem considerada pequena pela equipe médica e introduzimos esse limiar no algoritmo.

- **<u>Composição II</u>**

Nessa classe a mama contém mais tecido glandular, portanto a diferença entre o fundo da imagem do disco glandular e as manchas claras é menor (contraste menor) o que foi considerado no algoritmo pela eq. (7):

$$\text{Nível para densidade 2} = \text{THRMÉDIO} + \text{DESVTHR}/2 \qquad (7)$$

Entretanto as áreas ocupadas pelas manchas claras (densidade 2) são maiores que na Composição I, mas menores do que na Composição III. Para entender o que a equipe médica entende como "maior" e "menor" o algoritmo analisa os limiares de porcentagens ocupadas na mama pelas densidades II e III no banco de treinamento já laudado.

- **Composição III**

O padrão BI-RADS classifica como Composição III a mama onde "o tecido mamário é considerado heterogeneamente denso". Para atender esse critério adotamos o Segundo Momento Angular (SMA) que fornece a medida quantitativa de regularidade da distribuição dos níveis de cinza na imagem. Essa medida serviu para separar as Composições III e IV.

O valor de corte para o SMA foi definido analisando nas imagens do banco de treinamento já classificadas pelo laudo o que a equipe medica considera como heterogeneamente denso e, portanto como pertencentes à Composição III.

- **Composição IV**

O padrão de heterogeneidade adotado para a Composição III foi considerado também para a Composição IV. O padrão BI-RADS classifica na Composição IV a mama onde "o tecido mamário é extremamente denso", ou seja, com áreas de densidades que possuem altos valores do Segundo Momento Angular (SMA) na janela local.

Neste caso também o critério "extremamente" utilizado pela equipe médica responsável pela classificação, foi extraído da análise do banco de treinamento já laudado.

4.3.3.2 CÁLCULO DO SEGUNDO MOMENTO ANGULAR (SMA)

As três abordagens principais usadas em processamento de imagens para a descrição de textura são a estatística, a estrutural e a espectral. As abordagens estatísticas levam a caracterizações de textura como suave, áspera, granular e assim por diante. As técnicas estruturais tratam de arranjos de primitivas de imagem, como a descrição de textura baseada em linhas paralelas regularmente espaçadas. As espectrais baseiam-se em propriedades do espectro de Fourier, sendo usadas basicamente na detecção de periodicidade global em uma imagem através da identificação de picos de alta-energia no espectro (GONZALEZ 2000).

Ainda segundo Gonzalez (2000), embora não exista nenhuma definição formal de textura, o segundo momento angular intuitivamente fornece medidas de propriedades como suavidade, rugosidade e regularidade. O valor do SMA é calculado através de Matrizes de Co-ocorrência (*GLCM – gray level coocurrence matrix*). Essas matrizes são muito úteis para discriminar diferentes análises de texturas, mas não são muito eficientes em termos de memória computacional. Por isso o método adotado deve considerar apenas pequenas regiões da imagem, diminuindo assim o tempo de processamento.

O posicionamento relativo do "pixel" na imagem pode ser resumido em matriz de co-ocorrência bidimensional, calculada para diferentes distancias e orientações. A matriz de co-ocorrência de textura considera a relação entre dois "pixels" por vez, um chamado de "pixel" referência e o outro de "pixel" vizinho. O "pixel" vizinho escolhido pode, por exemplo, estar à leste (direita) de cada "pixel" referência. Isto pode ser expresso como uma relação *(1,0)*: 1 "pixel" na direção *x*, 0 "pixel" na direção y.

Cada "pixel" dentro da imagem torna-se o "pixel" referência, iniciando no canto superior esquerdo e procedendo até o inferior direito. Os "pixels" situados na margem direita não têm vizinhos da direita, então eles não são utilizados para esta contagem. Então os elementos de textura são "pixels" e as propriedades são os níveis de cinza. Por exemplo, para um relacionamento com um ângulo $\theta = 0^o$, $P\ (i, j;\ d,\ \theta)$ é a média aritmética da probabilidade de uma transição, da esquerda para a direita, do nível de cinza *i* para o nível de cinza *j*, com uma distância *d*. Essas matrizes podem ser normalizadas através da divisão de cada entrada da matriz pelo número de "pixels" vizinhos usados para calcular aquela matriz, (AKSOY, 1998).

O Segundo Momento Angular pode representar muito bem a homogeneidade local da imagem. Esse momento é o oposto à medida da Entropia. Quando a imagem é heterogênea o valor do SMA é baixo e para imagem homogênea o valor de SMA é alto. A eq. (8) descreve esse momento:

$$SMA = \sum_{i=0}^{n-1} \sum_{j=0}^{n-1} [P(i,j,d,\theta)]^2 \qquad (8)$$

onde *P* é a matriz normalizada, *i* e *j* são valores de níveis de cinza da matriz de coocorrência, *d* e θ são distância e direção que um "pixel" de referência possui em

relação ao "pixel" vizinho. Calculamos essas medidas nas quatro direções (0°, 45°, 90° e 135°) e as médias dos quatro valores.

4.3.3.3 FLUXOGRAMA UTILIZADO

Para classificar as imagens nas composições de densidades mamárias descritas pelo método BI-RADS, utilizamos o seguinte procedimento (Figura 26):

1. Seleção das imagens.
2. Divisão da imagem em regiões.
3. Detecção da borda da mama.
4. Extração de informações de cada região.
5. Cálculo das densidades 1 e 2.
6. Cálculo da porcentagem de área da mama ocupada pelas densidades 1 e 2
7. Levantamento de histogramas para estabelecer o valor de corte utilizado no banco de treinamento entre a Composição I e as demais imagens.
8. Classificação das imagens da Composição I.
9. Exclusão do banco das imagens classificadas como Composição I.
10. Levantamento de histogramas para estabelecer o valor de corte no banco de treinamento entre a Composição II e as demais imagens.
11. Classificação das imagens da Composição II.
12. Exclusão do banco das imagens classificadas como Composição II.
13. Cálculo dos valores do Segundo Momento Angular (SMA) das imagens restantes.
14. Levantamento de histogramas para estabelecer os valores de corte entre as Composições III e IV.
15. Classificação das imagens da Composição III.
16. Classificação das imagens da Composição IV.

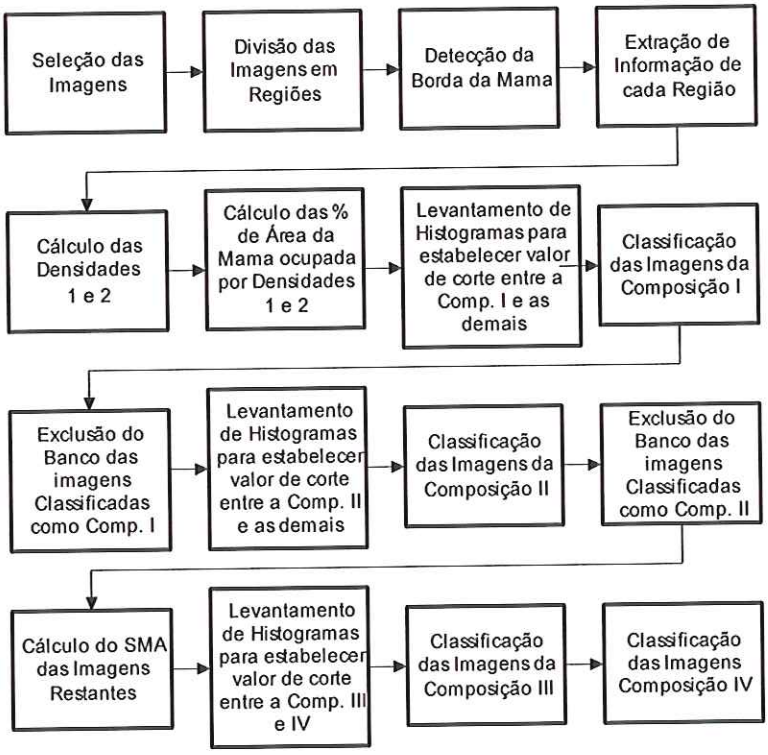


Figura 26 – Fluxograma para classificar as imagens nas composições de densidades mamárias descritas pelo método BI-RADS.

CAPÍTULO 5

RESULTADOS

5.1 DIVISÃO DA IMAGEM E DETECCÃO DA BORDA

A separação nas imagens da região ocupada somente pela mama, daquelas que apresentam outras informações que interfeririam nos resultados e no tempo de processamento é exemplificada a seguir. A FIGURA 27 mostra um dos mamogramas da base de dados, antes do processamento.

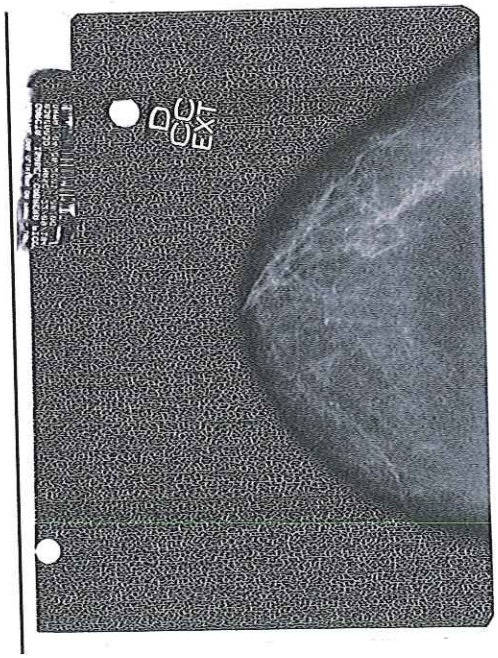


Figura 27 – Mamograma digitalizado do FMRP antes do processamento.

Para tanto utilizamos o IMAGE J, pois as imagens estavam no padrão *tif* e não são todos os "softwares" que abrem esse formato de imagem. A Figura 28 mostra a imagem após esse procedimento.

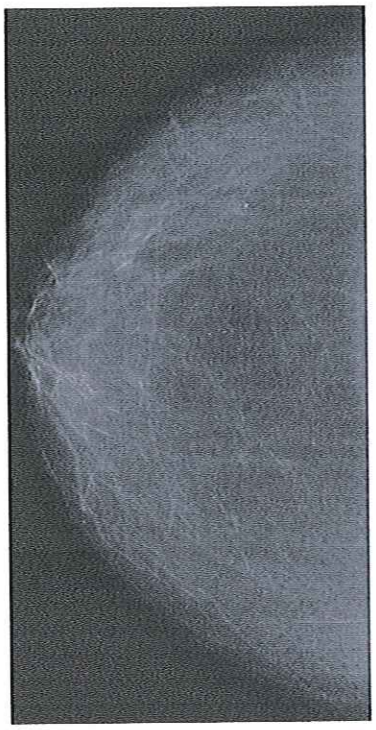


Figura 28 – Região de interesse do mamograma original selecionada utilizando o IMAGE J.

O próximo passo foi dividir as imagens em regiões para obter de cada área as informações necessárias para a detecção da borda, extração da região de densidade e classificação de acordo com o BI-RADS. A Figura 29 mostra uma imagem de 414x1332 "pixels" dividida em 29 regiões sendo janelas de 82x266 "pixels". Não mostramos a imagem de 10x10 "pixels" porque não é possível visualizar essas divisões sem o recurso da tecla "zoom".

Figura 29 – Imagem da mama dividida em 25 regiões.

Com a imagem dividida o próximo passo foi à detecção da borda. A Figura 30 mostra o resultado do processo de binarização da imagem onde o fundo se torna preto.

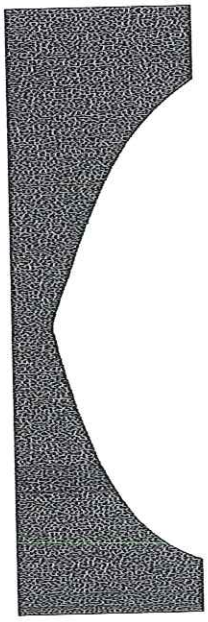


Figura 30 – Imagem binarizada da mama.

Sobre essa imagem binarizada foi feito a operação *AND* com a imagem original tendo como resultado a FIGURA 31 onde as estruturas originais da mamografia são preservadas.

Figura 31 – Mamografia com o fundo preto (zero) e a região da mama com as características originais.

O passo seguinte é a detecção da borda da região da mama. A Figura 32 mostra uma imagem com a borda da mama detectada.

Figura 32 – Imagem com a detecção da borda da mama.

A seguir aplicamos os algoritmos para extrair as características mencionadas no item 4.3.1. O processamento foi feito para todas as imagens sendo as respostas armazenadas em arquivos texto.

5.2 CLASSIFICAÇÃO DAS IMAGENS DO BANCO DE TREINAMENTO DA FACULDADE DE MEDICINA DE RIBEIRAO PRETO NO PADRÃO DE DENSIDADE MAMÁRIA BI-RADS

5.2. 1 PRIMEIRA CLASSIFICAÇÃO

Processamos as 152 imagens do Banco de treinamento da FMRP e obtivemos 52 parâmetros estatísticos de primeira e segunda ordem. Analisamos quais desses parâmetros teriam alguma relevância para classificar a densidade segundo o padrão BI-RADS e selecionamos os 20 parâmetros listados a seguir:

- Média global dos níveis de cinza da região de interesse (acc_med_total).
- Quantidade de "pixels" brancos (xa) e pretos (xb).
- Razão de "pixels" brancos e pretos (xa/xb).
- Média dos níveis de cinza das regiões claras e escuras (méd_reg_densa) (méd_reg_adiposa).
- Desvio padrão das regiões claras e escuras (desv_reg_densa) (desv_reg_adiposa).
- Região de fronteira da mama (xfront).
- Região onde a entropia é zero (xent_0).
- Total de regiões (somareg).
- Porcentagem de região de alta densidade (porc).
- Média, mediana e desvio das entropias (méd_ent), (median_ent) e (desv_ent).
- Média, mediana e desvio dos "Threshold" (méd_thr), (median_thr) e (desv_thr).
- Entropia das distribuições das entropias das regiões (tote1).
- Entropia das distribuições dos "Threshold" das regiões (toth1)

Para exemplificar apresentamos os valores obtidos com a imagem fmrp 002:

Nome = fmrp002

----------valores por regiao da mama----------

- xfront=481 || xent_0=8401 || xa=1963 || xb=18798 || razao(xa/xb)=0.1044
- Porcent_alta_dens = 9.4552 || Total de regioes=29643
- Media n.cinza regiao clara = 77.2384 || Desvio n.cinza regiao clara = 3.0715
- Media n.cinza regiao escura = 57.8398 || Desvio n.cinza regiao escura= 2.5380
- Media n.cinza da mama total = 59.6740
- Media thresh da mama total = 59.5613
- Media entr. regiao clara = 5.3924 || Desvio entr. regiao clara = 1.5907
- Media entr. regiao escura = 4.3940 || Desvio entr. regiao escura = 1.4945
- Media thr. regiao clara = 76.7117 || Desvio thr. regiao clara = 3.6692
- Media thr. regiao escura = 57.2979 || Desvio thr. regiao escura = 10.4815
- Entr. dos thr regiao clara = 6.1496 || Entr. dos thr regiao escura = 17.8728
- Cov ou Var thr regiao clara= 13.4632 || Covariância thr regiao escur= 109.8610
- Variância de thr (peso=thr) regiao clara = 14.2518
- Variância de thr (peso=ent) regiao clara = 14.8101
- Variância de thr (peso=thr) regiao escura= 79.5963
- Variância de thr (peso=ent) regiao escura= 123.5298
- Covariância ent regiao clara= 2.5304 || Covariância ent regiao escur= 2.2335
- Variância de ent (peso=thr) regiao clara = 2.5719
- Variância de ent (peso=ent) regiao clara = 3.6396
- Variância de ent (peso=thr) regiao escura= 1.8146
- Variância de ent (peso=ent) regiao escura= 4.4601

----------valores globais da mama----------

- Media das entr=4.6128 || Mediana das entr=4.6128 || Desvio das entr=1.8997
- Coef. Variação Pearson entr= 0.4118 || Coef. Ass. Pearson entr= 0.0000
- Media dos thr = 57.8171 || Mediana dos thr= 57.8171 || Desvio dos thr = 14.3107
- Coef. Variação Pearson thr= 0.2475 || Coef. Ass. Pearson thr= 0.0000
- Entropia da distr. de entropias da mama toda = 4.1073
- Entropia da distr. de threshold da mama toda = 44.6315
- Variância de thr (peso=ent) = 302.6522 || Variância de thr (peso=thr) = 130.4345
- Variância de ent (peso=thr) = 2.1623 || Variância de ent (peso=ent) = 7.8310
- Covariância da distr. threshold da mama toda = 802.3787
- Covariância da distr. entropia da mama toda = 6.9075
- Valor diagonal da Covariância = 49.9675

Analisamos esses parâmetros para as 152 imagens do banco e combinamos a média dos níveis de cinza, a porcentagem da mama ocupada por região de densidade e a diferença entre os níveis de cinza claro e escuro (contraste) de cada mamografia.

Entretanto os resultados obtidos não foram satisfatórios porque não conseguimos acertos maiores que 60% a saber:

- Composição I (BI-RADS) = 58% de concordância.
- Composição II (BI-RADS) = 54% de concordância.
- Composição III (BI-RADS) = 51% de concordância.
- Composição IV (BI-RADS) = 60% de concordância.

5.2.2 CLASSIFICAÇÃO COM O FASTMAP

Com o auxílio do *FastMapDB* desenvolvido por Traina Jr et al. (1999) e Traina et al, (2001) combinamos esses 52 parâmetros de várias maneiras para classificar as imagens nas Composições I, II, III, e IV de densidade mamária descritas no BI-RADS.

A Figura 33 mostra a distribuição das Composições em representações tridimensionais utilizando o FASTMAP.

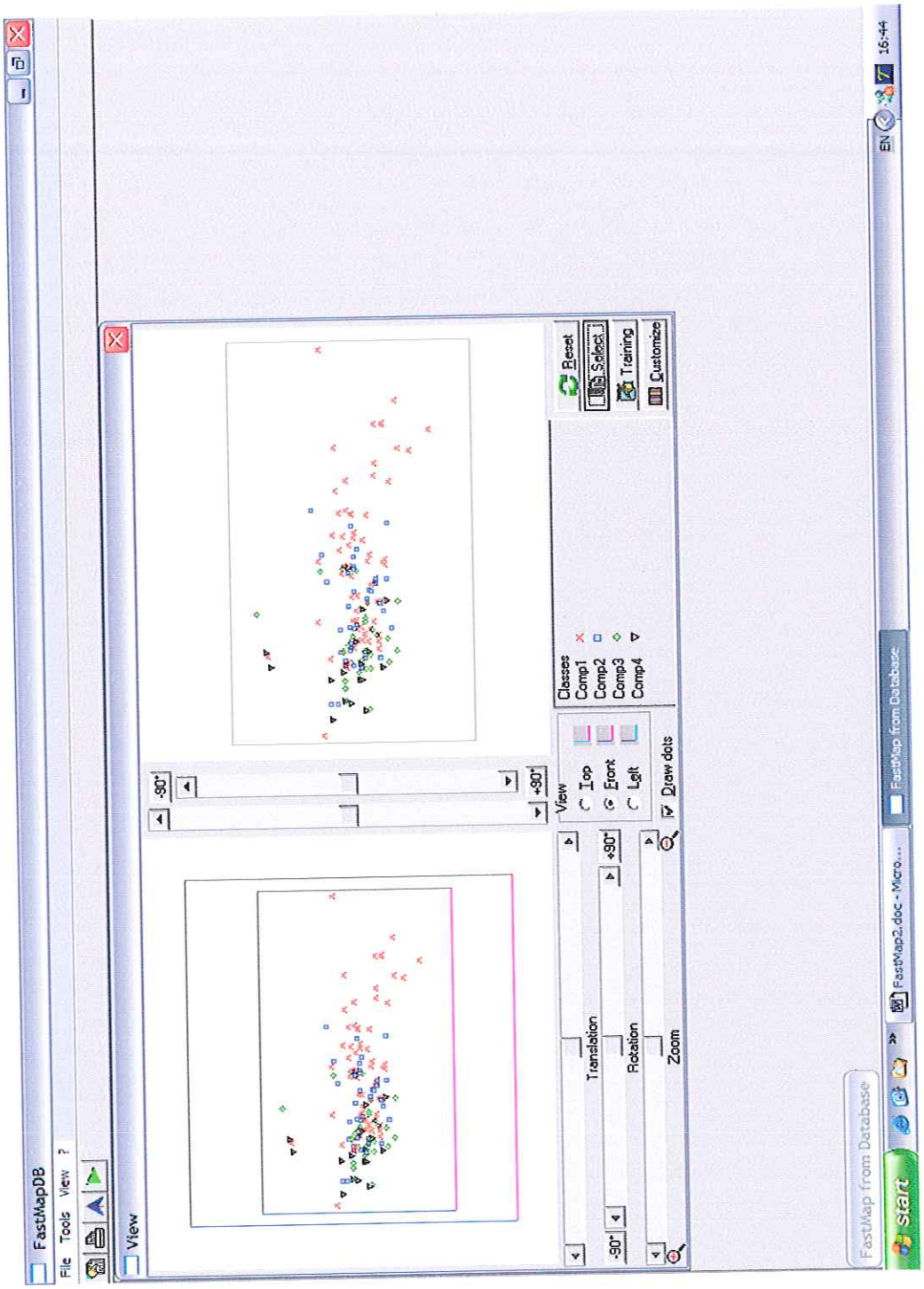


Figura 33 – Distribuição das Composições I, II, III e IV, utilizando o FastMap.

A Figura 34 apresenta os resultados das distribuições, mas considerando as composições II, III e IV juntas formando uma só classe, neste caso chamada de II.

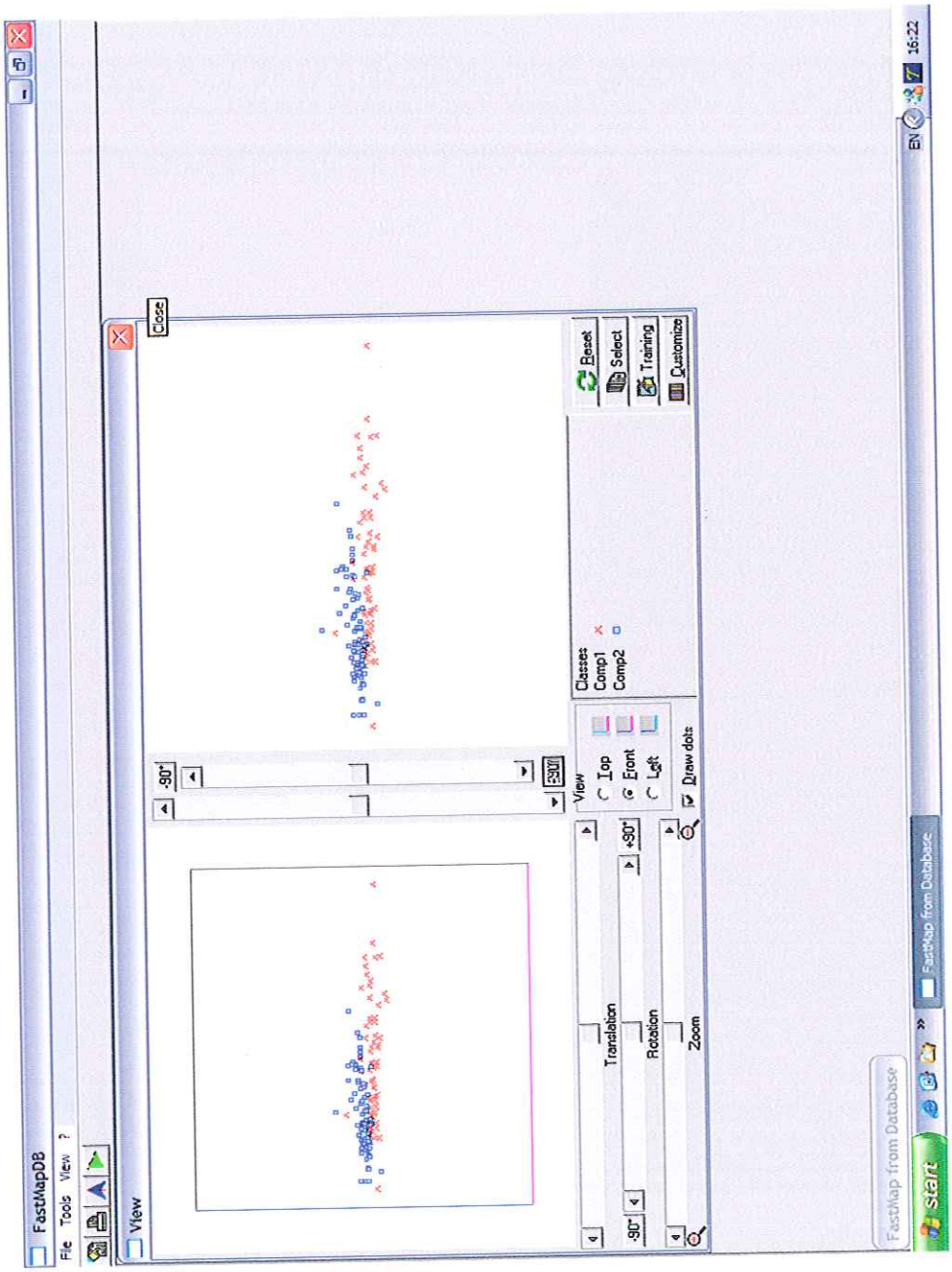


Figura 34 – Distribuição da Composição I e a classe II, com visualização no FastMap.

O programa sugeriu que deveríamos considerar as características xá, xb e a razão a/b. A Figura 35 representa as composições utilizando somente as características sugeridas pelo FastMap.

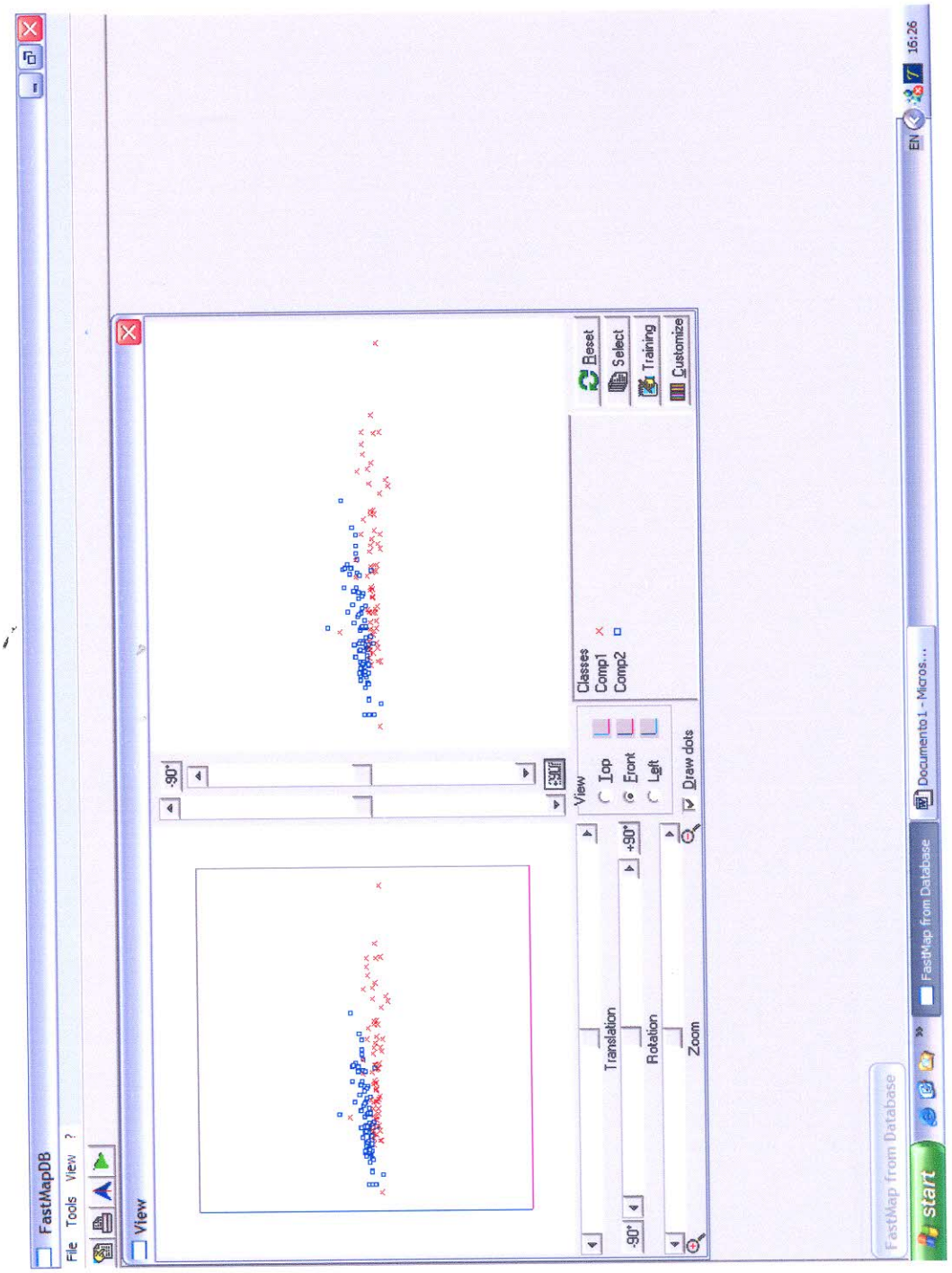


Figura 35– Distribuição da Composição I e classe II utilizando somente os valores de xa, xb e razão a/b (FastMap).

Observamos que as Figuras 34 e 35 são as mesmas. Não sendo necessário utilizar todas as características.

Fizemos o mesmo procedimento, mas juntando as Composições I, II e III em uma só classe I, deixando a Composição IV separada. A Figura 36 mostra esse agrupamento.

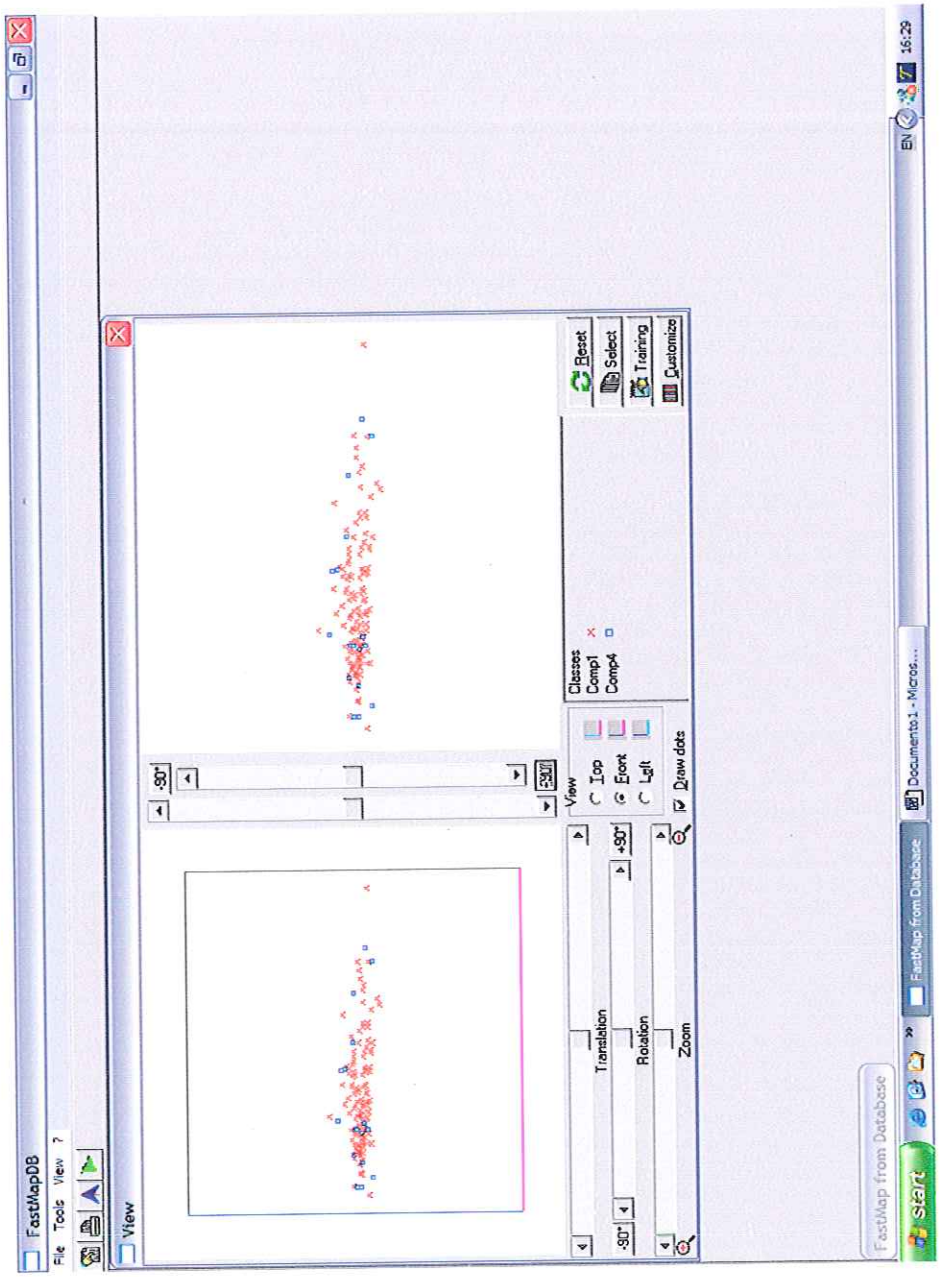


Figura 36 – Distribuição da Composição IV, com as composições I, II e III agrupadas na classe I (FastMap).

O programa sugeriu que deveríamos considerar as características xa, xb e a razão a/b. A Figura 37 representa as composições utilizando somente as características sugeridas pelo FastMap.

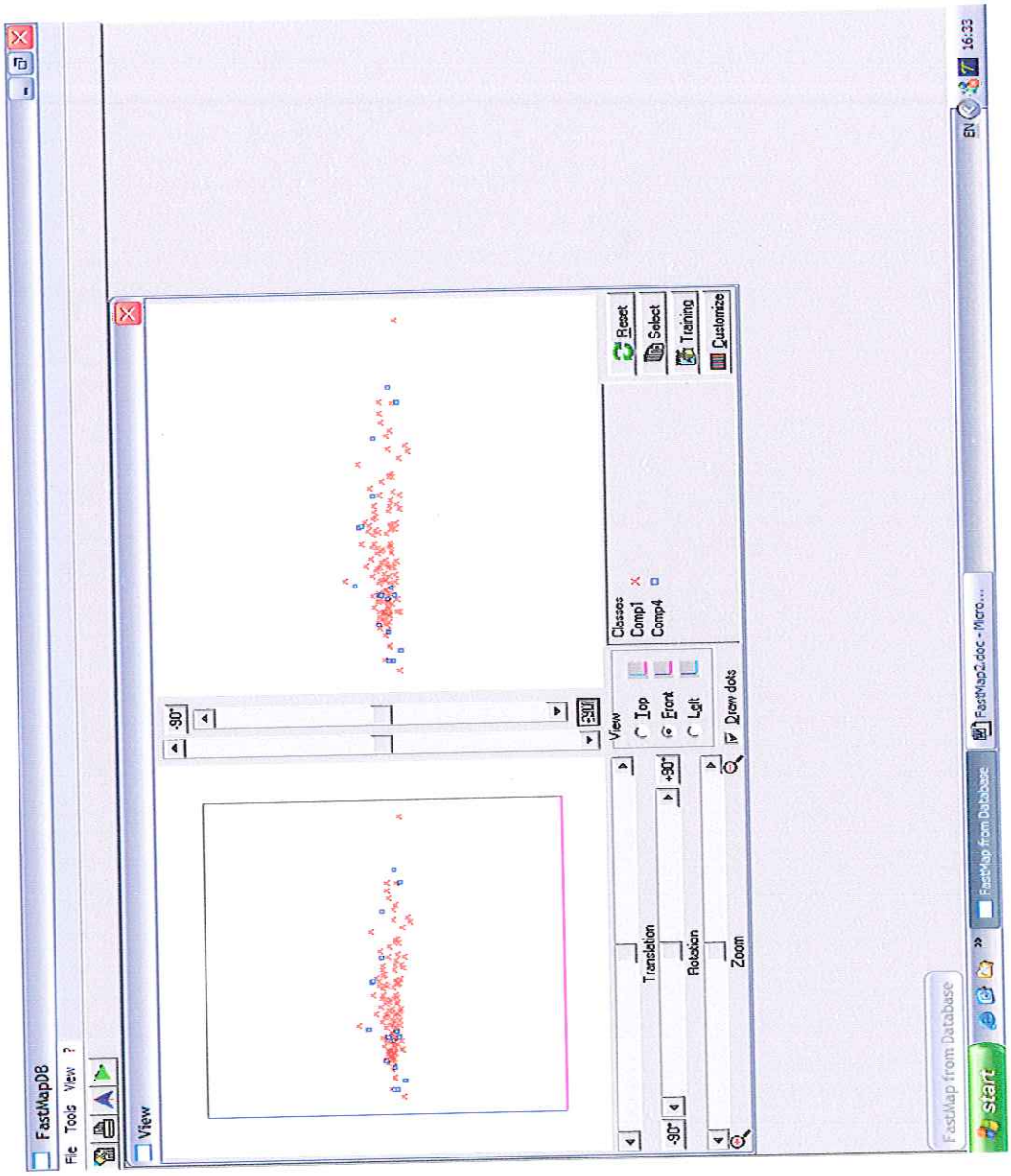


Figura 37 – Distribuição das Composições I, II, III e IV, utilizando somente os valores de xa, xb e razão a/b sendo que as I, II e III foram chamadas de I, (FastMap).

Observamos que as Figuras 36 e 37 são as mesmas. Não sendo necessário usar todas as características.

Um dos auxílios ao usuário disponível no FastMap, informou que os melhores atributos a serem utilizados na classificação seriam: razão ab, porcentagem da mama ocupada por densidade, média dos níveis de cinza das regiões densas, desvio das regiões densas e media das regiões adiposas. Com estes dados obtivemos uma concordância media de 67% entre o laudo e o resultado do classificador.

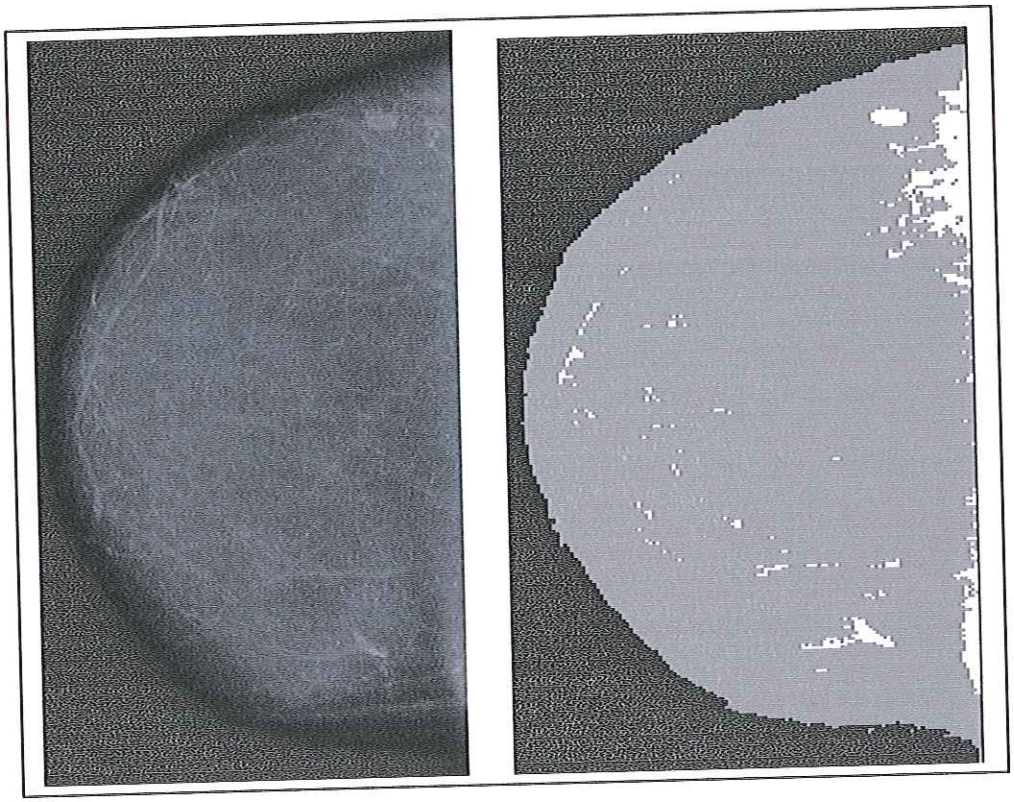
Esses resultados não são inteiramente satisfatórios, para melhorá-los deveríamos escolher outros limiares para esses parâmetros o que é um procedimento trabalhoso, portanto, desenvolvemos um algoritmo que analisa as imagens de maneira semelhante ao realizado pelo médico.

5.2.3 CLASSIFICAÇÃO DAS IMAGENS UTILIZANDO DENSIDADE E HOMOGENEIDADE

Calculamos a porcentagem da mama ocupada pelas densidades, 1 e 2 assim como o Segundo Momento Angular (SMA) que avalia a homogeneidade da imagem. Com essas duas características classificamos as imagens do primeiro banco de treinamento da FMRP em função das Composições de Densidades Mamográficas apresentadas no BI-RADS.

Inicialmente para definir automaticamente o que o radiologista considerou como densidade, independentemente das técnicas de exposição, variamos o limiar de contraste até que as áreas demarcadas pelo programa nas imagens do banco 1 (FMRP), coincidam com a avaliação do médico. Chamamos as áreas marcadas com este limiar de densidade 1.

A Figura 38a mostra uma mamografia sem processamento e a Figura 38b a mesma imagem com a área de densidade 1 marcada.



(38a) (38b)

Figura 38 a) imagem original e b) mamografia com densidade 1 marcada pelo limiar 1.

Utilizamos o *Software Origin 5.0* para mostrar a distribuição das imagens em função da porcentagem ocupada na mamografia pela densidade 1. A Figura 39 apresenta o gráfico obtido com todas as imagens do banco. A Figura 40 mostra a porcentagem de densidade 1 somente para as imagens da Composição I, a Figura 41 a porcentagem para a Composição II, a Figura 42 a porcentagem para a Composição III e a Figura 43 a porcentagem para a Composição IV.

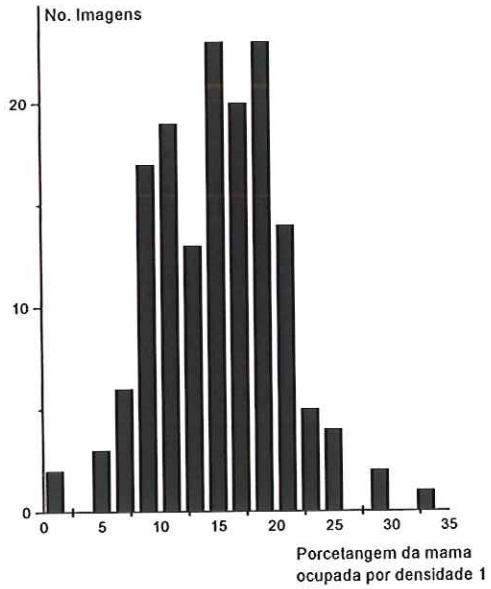


Figura 39 – Distribuição dos valores de porcentagens da mamografia ocupada pela densidade 1 para todas as imagens do banco da FMRP-I.

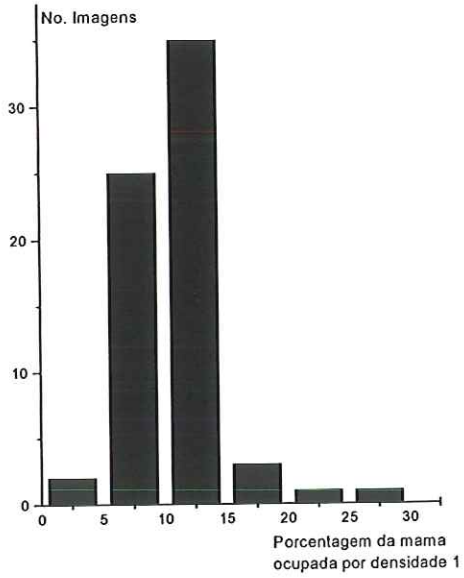


Figura 40 –Distribuição dos valores de porcentagens de densidade 1 para as imagens do banco da FMRP-I com Composição I.

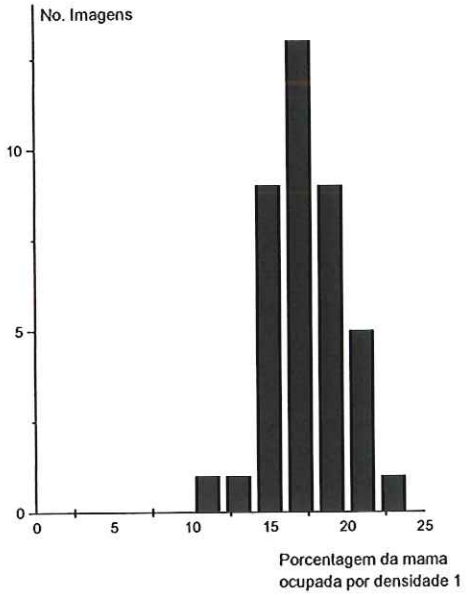


Figura 41 - Distribuição dos valores de porcentagens de densidade 1 para as imagens do banco da FMRP-I com Composição II.

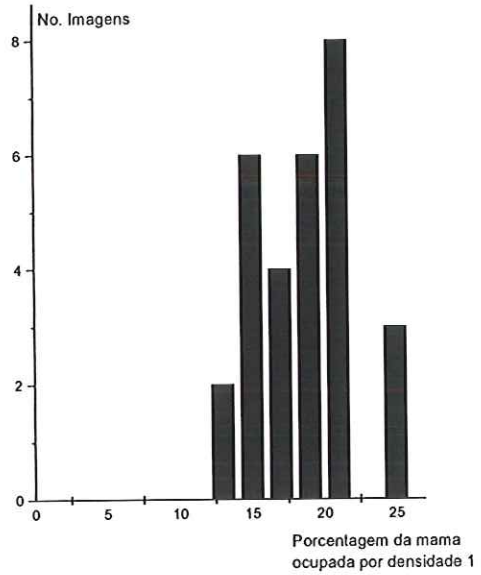


Figura 42 –Distribuição dos valores de porcentagens de densidade 1 para as imagens do banco da FMRP-I com Composição III.

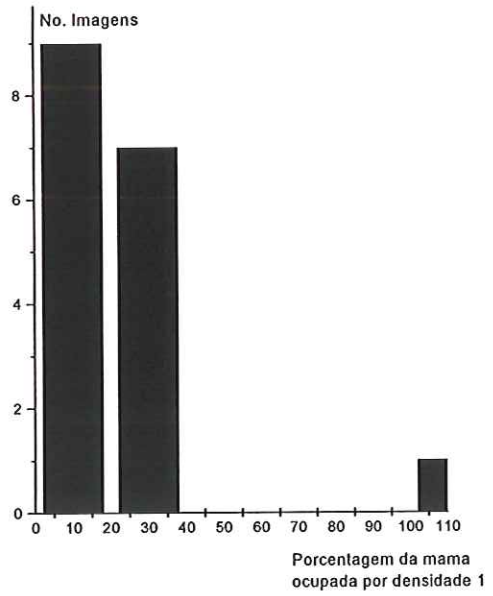
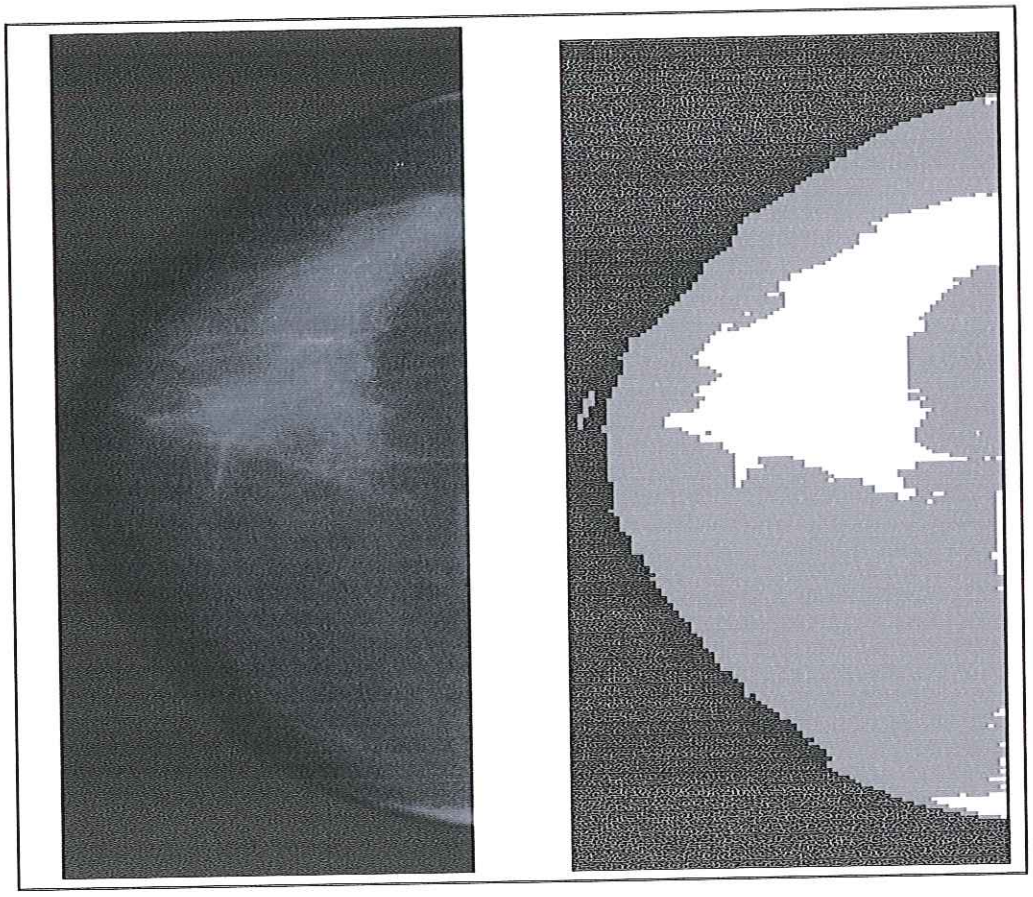


Figura 43 –Distribuição dos valores de porcentagens de densidade 1 para as imagens do banco da FMRP-I com Composição IV.

Estes histogramas mostram que todas as imagens classificadas pelos médicos como pertencentes à Composição I apresentam até 15% da mama ocupada por densidade 1. Adotamos portanto esse limiar para o banco da FMRP e retiramos as imagens consideradas pelo algoritmo como Composição I da base de dados.

A seguir processamos o banco FMRP-I para marcar as imagens com densidade 2. A Figura 44a mostra uma mamografia original e a Figura 44b a mesma imagem com a área de densidade 2.



(44a) (44b)

Figura 44 - a) imagem original e b) mamografia marcada com densidade 2.

A Figura 45 mostra a porcentagem para todas as imagens do banco FMRP-I. A Figura 46 mostra a porcentagem de densidade 2 somente para as imagens na Composição II, a Figura 47 a porcentagem para a Composição III, a Figura 48 a porcentagem para a Composição IV.

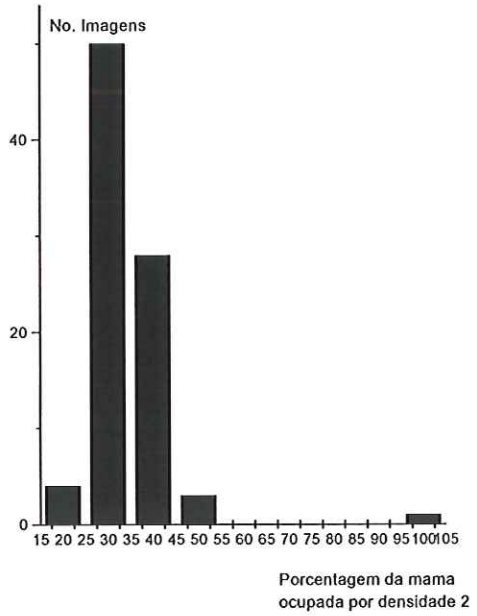


Figura 45 – Distribuição dos valores de porcentagens de densidade 2 para as imagens do banco da FMRP-I com Composição II, III e IV.

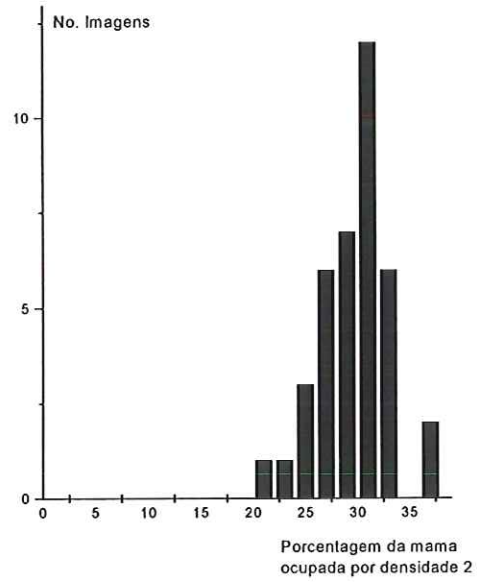


Figura 46 – Distribuição dos valores de porcentagens de densidade 2 para as imagens do banco da FMRP-I com Composição II.

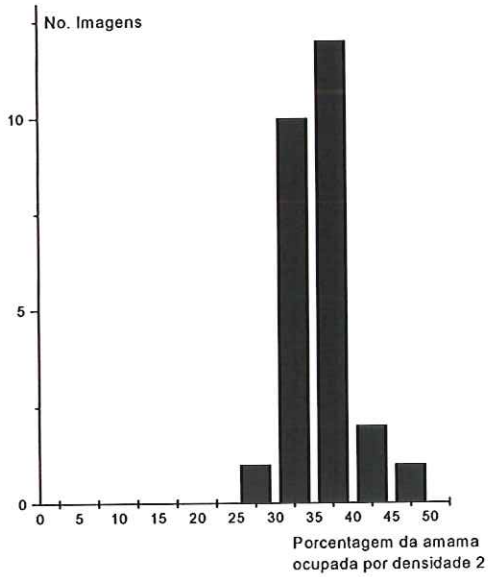


Figura 47 – Distribuição dos valores de porcentagens de densidade 2 para as imagens do banco da FMRP-I com Composição III.

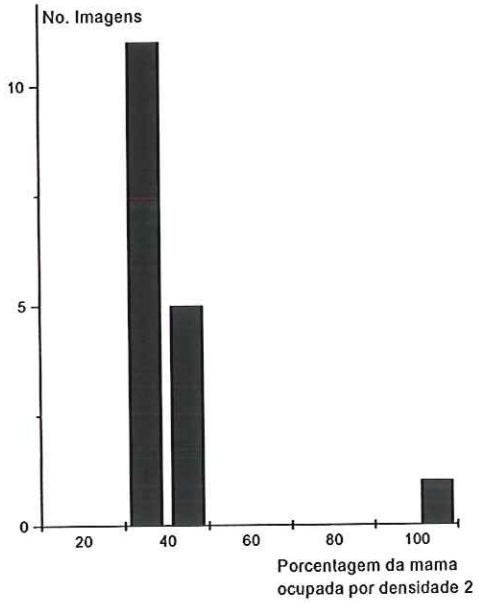


Figura 48 – Distribuição dos valores de porcentagens de densidade 2 para as imagens do banco da FMRP-I com Composição IV.

O histograma da Figura 46 mostra que a maioria das imagens laudadas como Composição II tem até 33% de sua área total ocupados por valores representativos da densidade 2 (limiar 2). Entretanto esse parâmetro não distingue as composições III e IV. Portanto, o algoritmo classifica como Composição 2 as imagens que apresentam até 33% de densidade 2 e como Composição III ou IV as imagens com valores acima.

Retiramos as imagens classificadas na Composição II processamos as demais para obter os valores do Segundo Momento Angular. Esses valores foram colocados em histogramas para determinar o valor de corte entre as Composições III e IV. As Figuras 49, 50 e 51 mostram como os valores do SMA se distribuem nos histogramas.

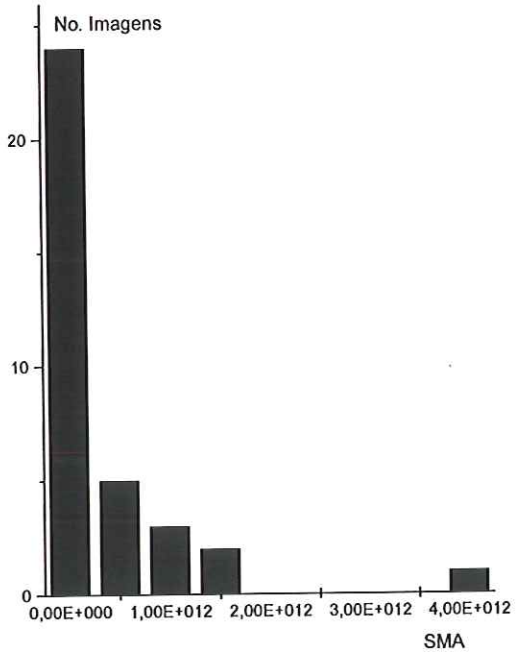


Figura 49 – Valores do Segundo Momento Angular obtidos processando as imagens não classificadas como composição I e II do banco da FMRP-I.

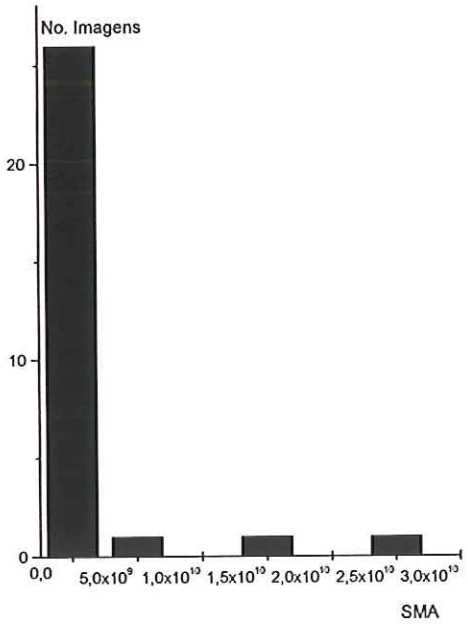


Figura 50 – Valores do Segundo Momento Angular obtidos processando as imagens da Composição III do banco da FMRP-I.

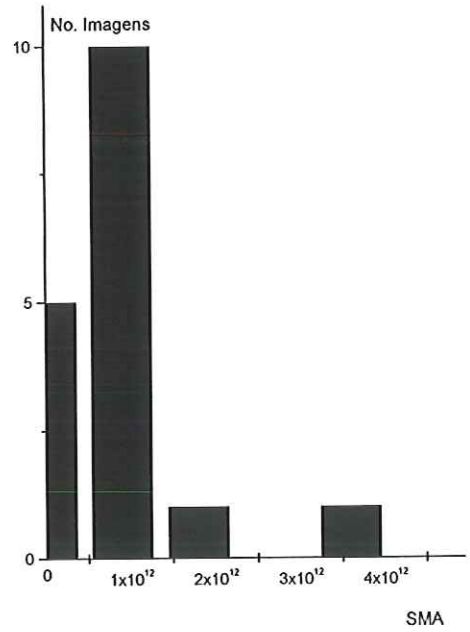


Figura 51 – Valores do Segundo Momento Angular obtidos processando as imagens da Composição IV do banco da FMRP-I.

Consideramos como Composição III imagens que apresentam valores até $5{,}0x10^{11}$ e acima deste valor como Composição IV.

A Tabela 2 apresenta os valores das porcentagens das densidades 1 e 2, os valores dos SMAs, os resultados do classificador e do laudo e a concordância entre o resultado e o laudo para o banco de imagens da FMRP-I (banco de treinamento).

Tabela 2 – Relação das imagens com as porcentagens de densidade 1 e 2, os SMAs e a concordância entre o laudo e as regras para o banco FMRP-I.

	Nome da imagem	% da mama ocupada por densidade 1	% da mama ocupada por densidade2	SMA	Laudo Médico	Resultado com Classifica-dor	Concordancia
1	fmrp002	9,4	-		COMP1	COMP1	VP
2	fmrp006	10,2	-		COMP1	COMP1	VP
3	fmrp010	15,3	37,7	129.222	COMP1	COMP3	FALSO
4	fmrp011	10,8	-		COMP1	COMP1	VP
5	fmrp016	11,8	-		COMP1	COMP1	VP
6	fmrp023	13,2	-		COMP1	COMP1	VP
7	fmrp026	11,2	-		COMP1	COMP1	VP
8	fmrp031	14,3	-		COMP1	COMP1	VP
9	fmrp032	4,4	-		COMP1	COMP1	VP
10	fmrp033	8,9	-		COMP1	COMP1	VP
11	fmrp036	6,7	-		COMP1	COMP1	VP
12	fmrp037	10,5	-		COMP1	COMP1	VP
13	fmrp040	11	-		COMP1	COMP1	VP
14	fmrp042	9,1	-		COMP1	COMP1	VP
15	fmrp047	10,8	-		COMP1	COMP1	VP
16	fmrp048	1,1			COMP1	COMP1	VP
17	fmrp051	10,4	-		COMP1	COMP1	VP
18	fmrp056	12,4	-		COMP1	COMP1	VP
19	fmrp068	8,1	-		COMP1	COMP1	VP
20	fmrp074	8,6	-		COMP1	COMP1	VP
21	fmrp084	13,1	-		COMP1	COMP1	VP
22	fmrp091	14,8	-		COMP1	COMP1	VP
23	fmrp109	10,3	-		COMP1	COMP1	VP
24	fmrp110	12,6	-		COMP1	COMP1	VP
25	fmrp123	12,9	-		COMP1	COMP1	VP
26	fmrp125	12,1	-		COMP1	COMP1	VP
27	fmrp140	7,9	-		COMP1	COMP1	VP
28	fmrp147	9,2	-		COMP1	COMP1	VP
29	fmrp153	10,7	-		COMP1	COMP1	VP
30	fmrp158	14,3	-		COMP1	COMP1	VP
31	fmrp159	10,5	-		COMP1	COMP1	VP
32	fmrp170	10,6	-		COMP1	COMP1	VP
33	fmrp171	14,8	-		COMP1	COMP1	VP

	Nome da imagem	% da mama ocupada por densidade 1	% da mama ocupada por densidade2	SMA	Laudo Médico	Resultado com Classifica-dor	Concordancia
34	fmrp174	5,3	-		COMP1	COMP1	VP
35	fmrp178	14	-		COMP1	COMP1	VP
36	fmrp180	9,9	-		COMP1	COMP1	VP
37	fmrp183	9,6	-		COMP1	COMP1	VP
38	fmrp187	12,2	-		COMP1	COMP1	VP
39	fmrp186	8,3	-		COMP1	COMP1	VP
40	fmrp198	8,9	-		COMP1	COMP1	VP
41	fmrp200	10,7	-		COMP1	COMP1	VP
42	fmrp204	5,8	-		COMP1	COMP1	VP
43	fmrp210	11,4	-		COMP1	COMP1	VP
44	fmrp211	7,2	-		COMP1	COMP1	VP
45	fmrp212	9,6	-		COMP1	COMP1	VP
46	fmrp214	6,7	-		COMP1	COMP1	VP
47	fmrp218	12,1	-		COMP1	COMP1	VP
48	fmrp221	11,1	-		COMP1	COMP1	VP
49	fmrp226	10,2	-		COMP1	COMP1	VP
50	fmrp232	13,9	-		COMP1	COMP1	VP
51	fmrp235	20,8	28,9		COMP1	COMP2	FALSO
52	fmrp236	8,6	-		COMP1	COMP1	VP
53	fmrp237	9,5	-		COMP1	COMP1	VP
54	fmrp238	7,3	-		COMP1	COMP1	VP
55	fmrp242	17	30,7		COMP1	COMP2	FALSO
56	fmrp243	14,8	-		COMP1	COMP1	VP
57	fmrp261	9,2	-		COMP1	COMP1	VP
58	fmrp267	12,2	-		COMP1	COMP1	VP
59	fmrp269	9,3	-		COMP1	COMP1	VP
60	fmrp289	8,2	-		COMP1	COMP1	VP
61	fmrp296	11,2	-		COMP1	COMP1	VP
62	fmrp304	17,7	29,5		COMP1	COMP2	FALSO
63	fmrp306	9,8	-		COMP1	COMP1	VP
64	fmrp310	14	-		COMP1	COMP1	VP
65	fmrp317	29,5	52,2	135.168.869	COMP1	COMP3	FALSO
66	fmrp330	10,5	-		COMP1	COMP1	VP
67	fmrp331	7,9	-		COMP1	COMP1	VP

	Nome da imagem	% da mama ocupada por densidade 1	% da mama ocupada por densidade2	SMA	Laudo Médico	Resultado com Classifica-dor	Concordancia
68	fmrp001	18,9	27,8		COMP2	COMP2	VP
69	fmrp004	18,9	27,2		COMP2	COMP2	VP
70	fmrp007	17,7	31,7		COMP2	COMP2	VP
71	fmrp008	19,0	32,9		COMP2	COMP2	VP
72	fmrp009	16,1	23,6		COMP2	COMP2	VP
73	fmrp014	19,1	30,1		COMP2	COMP2	VP
74	fmrp021	17,7	25,3		COMP2	COMP2	VP
75	fmrp025	15	27,9		COMP2	COMP2	VP
76	fmrp027	15,1	24,3		COMP2	COMP2	VP
77	fmrp029	16,2	30,8		COMP2	COMP2	VP
78	fmrp030	16,2	30,8		COMP2	COMP2	VP
79	fmrp034	20,2	36,4	27.737.208	COMP2	COMP3	FALSO
80	fmrp035	20,6	31,2		COMP2	COMP2	VP
81	fmrp038	19,9	31,7		COMP2	COMP2	VP
82	fmrp039	15,8	29,4		COMP2	COMP2	VP
83	fmrp045	15,8	29,8		COMP2	COMP2	VP
84	fmrp050	16,7	28,9		COMP2	COMP2	VP
85	fmrp052	16,8	30,9		COMP2	COMP2	VP
86	fmrp062	19,4	29,7		COMP2	COMP2	VP
87	fmrp065	16,2	28,7		COMP2	COMP2	VP
88	fmrp070	16,5	29,2		COMP2	COMP2	VP
89	fmrp076	15,6	37,1	583.727	COMP2	COMP3	FALSO
90	fmrp085	13,9	28,9		COMP2	COMP1	FALSO
91	fmrp092	18	32,8		COMP2	COMP2	VP
92	fmrp102	15,4	32,7		COMP2	COMP2	VP
93	fmrp108	16,7	26,1		COMP2	COMP2	VP
94	fmrp128	15,5	32,9		COMP2	COMP2	VP
95	fmrp142	20,6	30,2		COMP2	COMP2	VP
96	fmrp143	17,4	26,1		COMP2	COMP2	VP
97	fmrp176	21,2	32,4		COMP2	COMP2	VP
98	fmrp195	15,7	21,1		COMP2	COMP2	VP
99	fmrp201	16,3	27,8		COMP2	COMP2	VP
100	fmrp215	17,2	24,1		COMP2	COMP2	VP
101	fmrp233	21,6	31,6		COMP2	COMP2	VP

	Nome da imagem	% da mama ocupada por densidade 1	% da mama ocupada por densidade2	SMA	Laudo Médico	Resultado com Classifica-dor	Concordancia
102	fmrp245	22,6	30,9		COMP2	COMP2	VP
103	fmrp247	19,7	32,6		COMP2	COMP2	VP
104	fmrp249	19,1	30,9		COMP2	COMP2	VP
105	fmrp256	10,5	-		COMP2	COMP1	FALSO
106	fmrp307	15,6	31,4		COMP2	COMP2	VP
107	fmrp012	21,6	33,3	63.628	COMP3	COMP3	VP
108	fmrp013	17,3	33,1	3.480.475.192	COMP3	COMP3	VP
109	fmrp015	21,1	33,2	304.852	COMP3	COMP3	VP
110	fmrp024	19,7	33,7	188.022	COMP3	COMP3	VP
111	fmrp028	16,3	39,1	375.867	COMP3	COMP3	VP
112	fmrp053	15,8	35,2	142.250	COMP3	COMP3	VP
113	fmrp060	15,2	36,4	312.856.917	COMP3	COMP3	VP
114	fmrp069	19,9	37,6	45.398	COMP3	COMP3	VP
115	fmrp077	13,9	-		COMP3	COMP1	FALSO
116	Fmrp080	12,4			COMP3	COMP1	FALSO
117	fmrp087	24,2	35,6	91.858	COMP3	COMP3	VP
118	fmrp121	19,3	39,6	8.578.787.690	COMP3	COMP3	VP
119	fmrp124	15,7	34,2	103.810	COMP3	COMP3	VP
120	fmrp127	19,8	39,5	86.294	COMP3	COMP3	VP
121	fmrp148	16,2	43,1	2.088.463.836	COMP3	COMP3	VP
122	fmrp150	15,9	37,7	25.818.739.807	COMP3	COMP3	VP
123	fmrp154	18,8	34,3	202.618	COMP3	COMP3	VP
124	fmrp155	21,1	35,9	309.991.872	COMP3	COMP3	VP
125	fmrp162	20,2	35,5	87.234	COMP3	COMP3	VP
126	fmrp164	20,1	38,8	376.889	COMP3	COMP3	VP
127	fmrp181	15,8	45,4	440.260.679	COMP3	COMP3	VP
128	fmrp188	20,1	34,7	2.602.907.499	COMP3	COMP3	VP
129	fmrp216	24	36,5	153.936	COMP3	COMP3	VP
130	fmrp231	14,7	-		COMP3	COMP1	FALSO
131	fmrp234	20,8	33,4		COMP3	COMP3	VP
132	fmrp255	19,4	34,5	672.465.201	COMP3	COMP3	VP
133	fmrp258	24,4	33,8	74.130	COMP3	COMP3	VP
134	fmrp321	17,9	26,5		COMP3	COMP2	FALSO
135	fmrp329	20,9	40,3	20.890.505.990	COMP3	COMP3	VP

	Nome da imagem	% da mama ocupada por densidade 1	% da mama ocupada por densidade2	SMA	Laudo Médico	Resultado com Classifica-dor	Concordancia
136	fmrp003	18,8	34,9	1.679.679.509.752	COMP4	COMP4	VP
137	fmrp020	18,0	40,9	751.153.113.568	COMP4	COMP4	VP
138	fmrp022	19,8	39,2	462.923.709.281	COMP4	COMP4	VP
139	fmrp055	29,5	44,1	231.207.905.473	COMP4	COMP4	VP
140	fmrp120	19	39,5	670.626.010.874	COMP4	COMP4	VP
141	fmrp129	23,3	40,1	1.328.252.473.309	COMP4	COMP4	VP
142	fmrp141	23,5	36,3	138.433.313.051	COMP4	COMP4	VP
143	fmrp152	100*	100	259.038.850.272	COMP4	COMP4	VP
144	fmrp161	18,9	31,4		COMP4	COMP2	FALSO
145	fmrp224	22,1	34,7	753.142.337.109	COMP4	COMP4	VP
146	fmrp246	18,2	38,2	640.609.742.443	COMP4	COMP4	VP
147	fmrp250	24,1	36	1.144.164.527.753	COMP4	COMP4	VP
148	fmrp262	16,9	42	548.724.848.984	COMP4	COMP4	VP
149	fmrp268	19,6	38,7	3.781.794.209.400	COMP4	COMP4	VP
150	fmrp293	23,4	31,1		COMP4	COMP2	FALSO
151	fmrp311	33,4	49	114.348.091.979	COMP4	COMP4	VP
152	fmrp316	18,4	39,2	718.059.120.654	COMP4	COMP4	VP

A Tabela 3 apresenta os resultados de VP, FN, FP, FN e Concordância encontrados na classificação.

Tabela 3 – Resultados de VP, FN e FP.

Composições (BI-RADS)	Total imagens	VP	FP	FN	Concordância
COMP. I	67	62	5	4	92,5%
COMP. II	39	35	6	4	90%
COMP. III	29	25	4	4	86,2%
COMP. IV	17	15	0	2	88,2%
Total	152	137	15	14	

Levantamos a taxa de acerto e de erros. Conseguimos um acerto de *92,5%* na Composição I e *90%* na Composição II, *86,2%* na Composição III e *88,2%* na Composição IV. Com o índice de Falsos Positivos *7,46%* para a Composição I, *15,38%*

para a Composição II, *13,79%* para a Composição III e *0%* para a Composição IV. Os Falsos Negativos foram *5,97%* para a Composição I, *10,2%* para a Composição II, *13,79%* para a Composição III e *11,7%* para a Composição IV.

Num total de 152 imagens apenas 15 não foram classificadas pelo algoritmo na mesma composição determinada pelo médico. Conseguimos, portanto uma concordância de 89,2%.

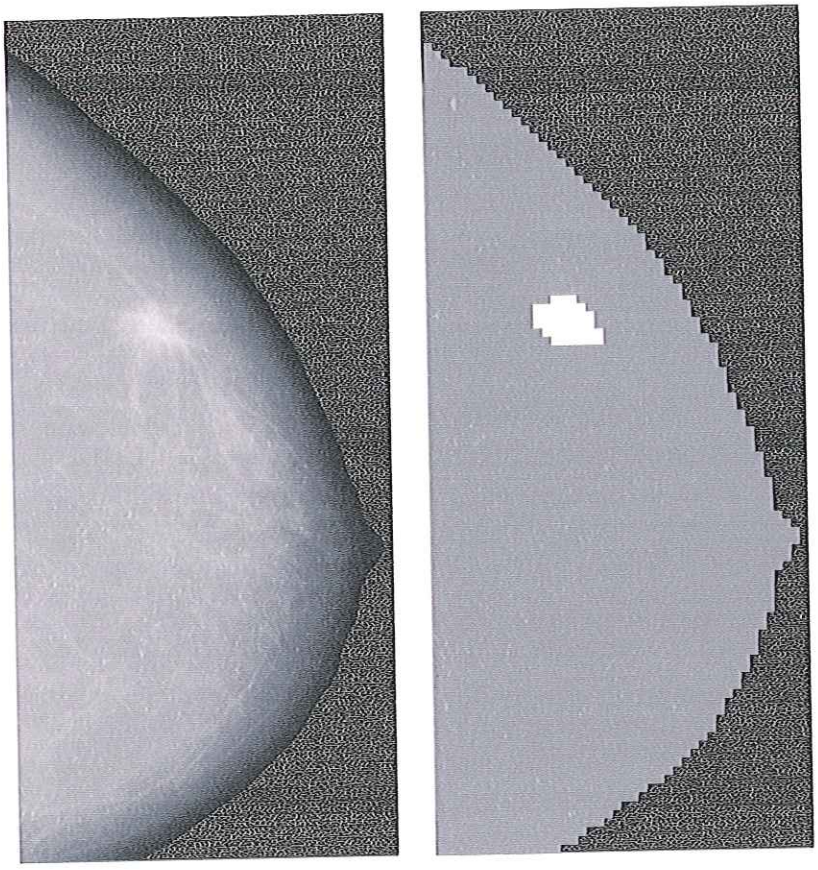
Analisamos a tabela e constatamos que as imagens fmrp 010, fmrp 233, fmrp 161, fmrp 231, fmrp 077, fmrp 085, fmrp 034 e fmrp 231 eram próximas do valor de corte da composição (mais ou menos 10%) são, portanto imagens que poderiam pertencer a uma ou outra composição dependendo do observador. Para as demais classificadas erroneamente, o algoritmo deve pedir ao médico a confirmação de seu laudo.

5.3 VALIDAÇÃO DO MÉTODO DESENVOLVIDO

5.3.1 APLICACÃO NO BANCO DO FUJILAB

Aplicamos este mesmo método em outro banco de imagens. O Banco do Laboratório do Prof. Dr. Fujita da Universidade de Gifu - Japão, FujiLab. A classificação da Composição da densidade mamária (BI-RADS) dessas imagens foi feita por um médico radiologista do Hospital das Clínicas de Ribeirão Preto.

As Figuras 52 a e b mostram que o limiar 1 determinado para a marcação das regiões de densidade 1 do banco da FMRP também esta correto para o banco do FujiLab.



(52a) (52b)

Figura 52 - a) imagem original do FujiLab, b) mamografia com densidade 1 marcada pelo limiar 1.

A Figura 53 mostra a distribuição das porcentagens de densidade 1 para todas as imagens do banco FujiLab. A Figura 54 mostra a porcentagem de densidade 1 somente para as imagens laudadas como Composição I, a Figura 55 a porcentagem para a Composição II, a Figura 56 a porcentagem para a Composição III e a Figura 57 a porcentagem para a Composição IV.

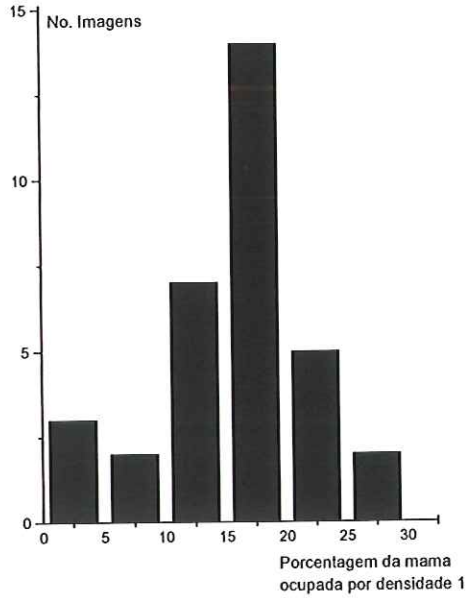


Figura 53 – Distribuição de porcentagens de densidades 1 para todas as imagens do banco FujiLab.

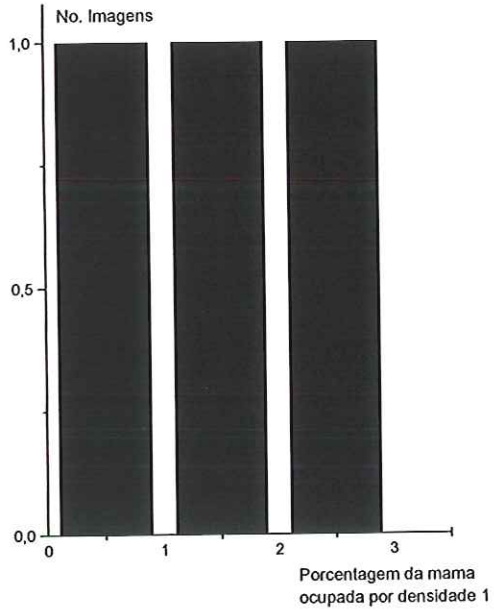


Figura 54 – Distribuição de porcentagens de densidades 1 para as imagens do banco FujiLab laudadas como Composição I.

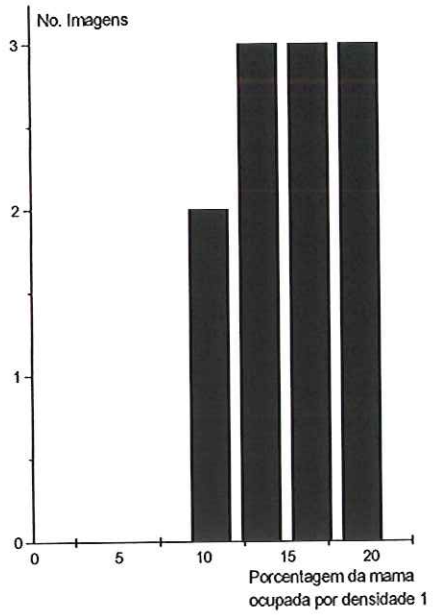


Figura 55 – Distribuição de porcentagens de densidades 1 para as imagens do banco FujiLab laudadas como Composição II.

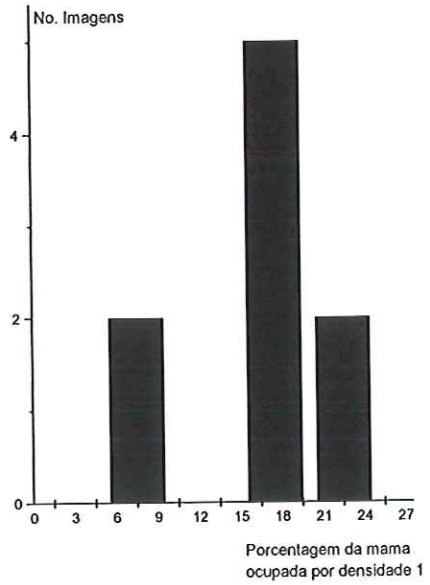


Figura 56 – Distribuição de porcentagens de densidades 1 para as imagens do banco FujiLab laudadas como Composição III.

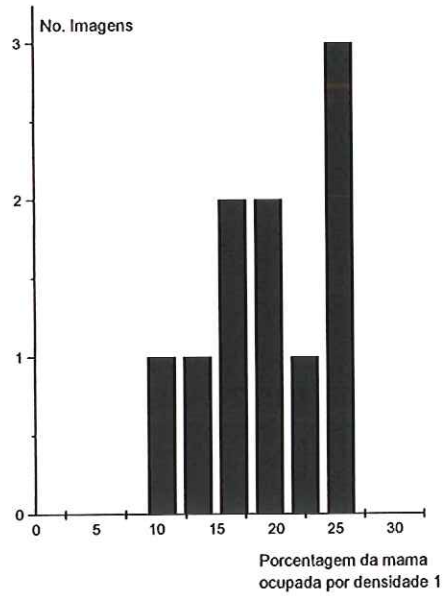
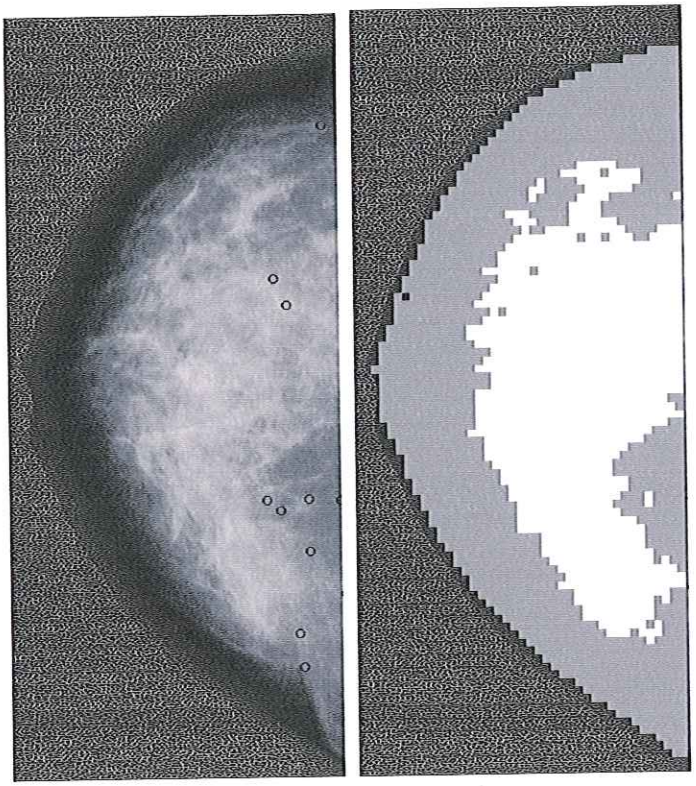


Figura 57 – Distribuição de porcentagens de densidades 1 para as imagens do banco FujiLab laudadas como Composição IV.

O histograma da Figura 54 mostra que o conceito de *"quase inteiramente adiposa"* utilizado no padrão BI-RADS para classificação na Composição I foi interpretado pelo radiologista como apresentando menos que 5% da mamografia ocupada por densidade. Utilizamos, portanto esse limiar para o classificador.

As imagens consideradas como Composição I foram retiradas da base de dados e processamos as imagens marcando as áreas densidade 2 com o mesmo contraste definido para o banco de imagens de Ribeirão Preto. A Figura 58a mostra uma mamografia original e a Figura 58b a mesma imagem com a área de densidade 2 marcada pelo limiar 2 para o FujiLab.



(58a) (58b)

Figura 58 - a) imagem original do FujiLab, b) mamografia com densidade 2 marcada pelo limiar 2.

A Figura 59 mostra a distribuição das porcentagens de densidade 2 para as imagens do banco FujiLab das Composições II, III e IV. A Figura 60 mostra a porcentagem de densidade 2 somente para as imagens laudadas como Composição II, a Figura 61 a porcentagem para a Composição III, a Figura 62 a porcentagem para a Composição IV.

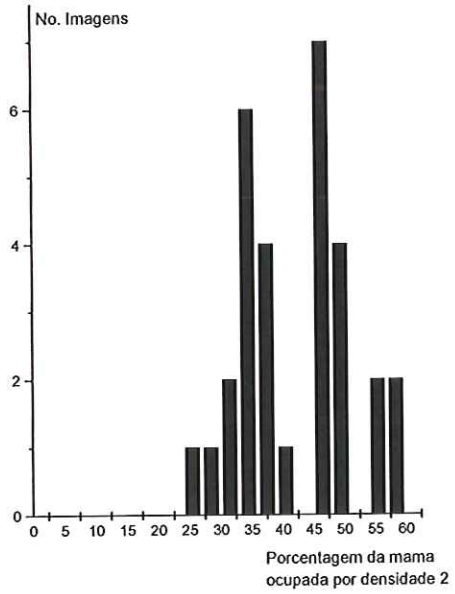


Figura 59 –Distribuição de porcentagens de densidade 2 para as todas as imagens do banco Fujilab.

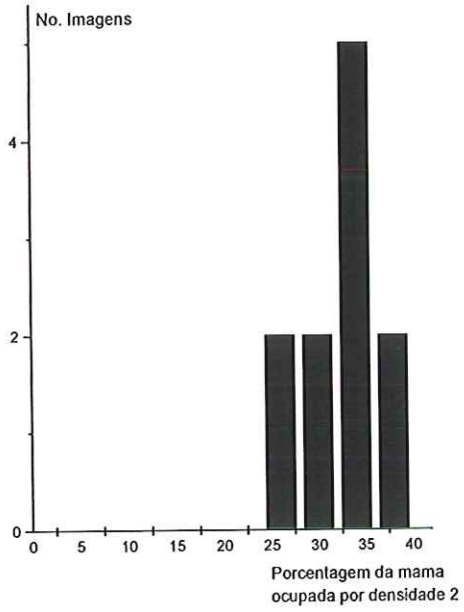


Figura 60 –Distribuição de porcentagens de densidade 2 para as imagens do banco Fujilab laudadas como Composição II.

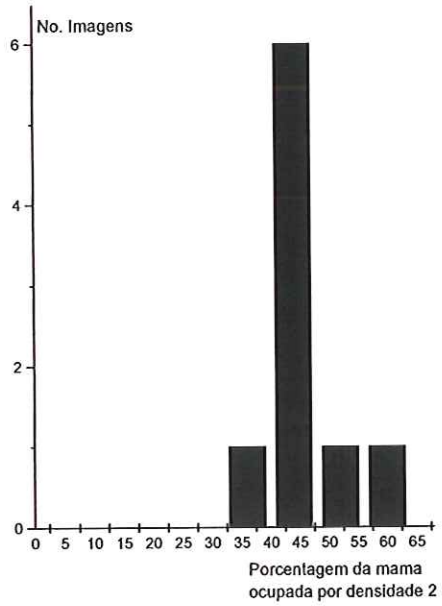


Figura 61 –Distribuição de porcentagens de densidades 2 para as imagens do banco Fujilab laudadas como Composição III.

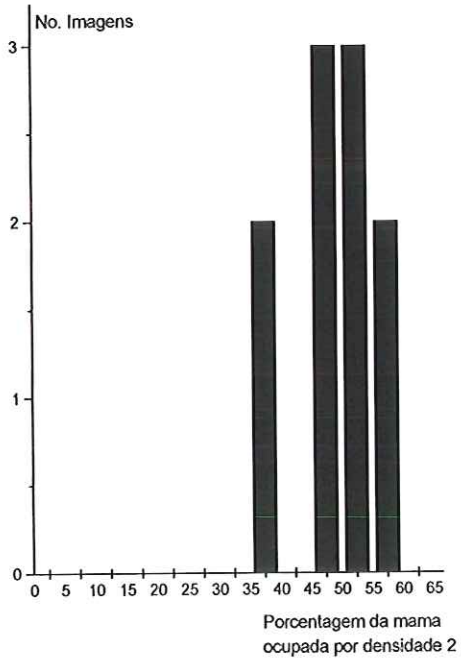


Figura 62 –Distribuição de porcentagens de densidade 2 para as imagens do banco Fujilab laudadas como Composição IV.

Deduzimos que imagens com porcentagem da mama ocupada por até 37% de densidade 2 foram consideradas pelo médico como pertencentes a Composição II.

Os valores do SMA foram obtidos processando as imagens não classificadas como Composição I e II. O algoritmo considerou Composição III imagens com valores de SMA menores que $30x10^9$ e para valores maiores Composição IV, Figura 63. As Figuras 64 e 65 mostram como os valores do SMA se distribuem nos histogramas para as imagens laudadas como Composições III e IV.

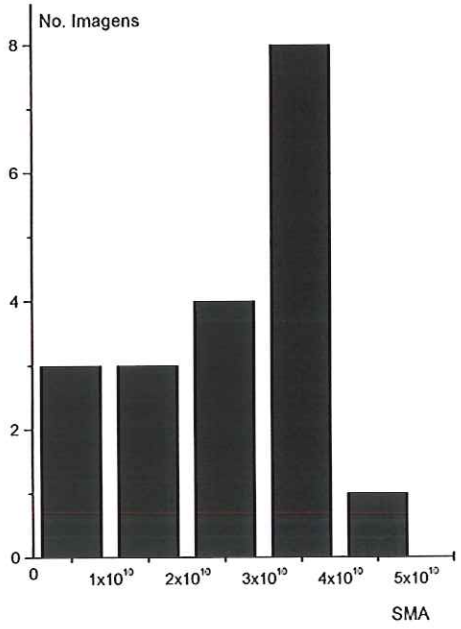


Figura 63 - Histograma dos valores de SMA para as Composições laudadas como III e IV do FujiLab.

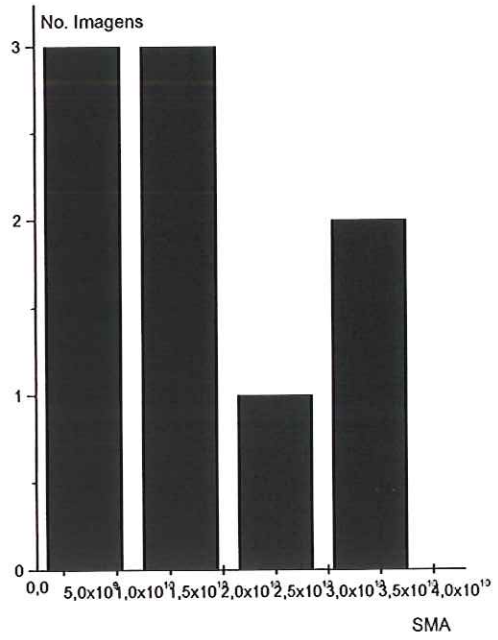


Figura 64 - Histograma dos valores de SMA para a Composição laudadas como III do FujiLab.

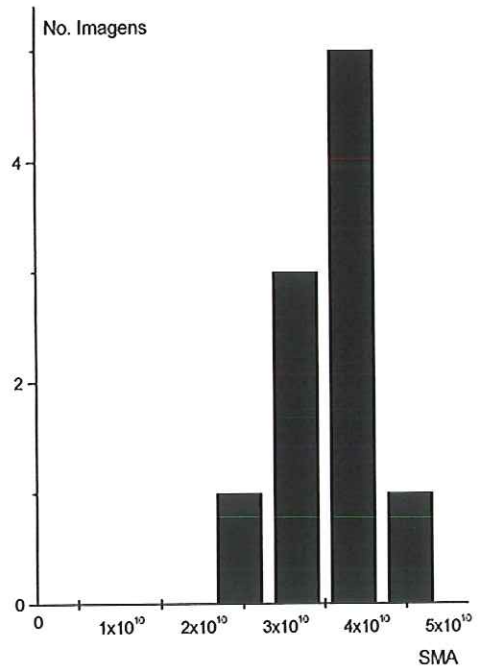


Figura 65 - Histograma dos valores de SMA para a Composição laudadas como IV do FujiLab.

A Tabela 4 apresenta os resultados obtidos com o classificador e a concordância entre o método e o laudo médico.

Tabela 4 – Resultado da classificação dada pelo médico e a concordância com o laudo, banco do FujiLab.

Número da imagem	**% da mama ocupada por densidade 1**	**% da mama ocupada por densidade 2**	**SMA**	**Laudo Médico**	**Resultado com Classificador**	**Concordância**
Fujilab_001	0,19	-	-	Comp 1	Comp 1	VP
Fujilab_002	2,17	-	-	Comp 1	Comp 1	VP
Fujilab_003	1,33	-	-	Comp 1	Comp 1	VP
Fujilab_004	13,2	31,08	-	Comp 2	Comp 2	VP
Fujilab_005	15,14	35,04	-	Comp 2	Comp 2	VP
Fujilab_006	14,76	34,32	-	Comp 2	Comp 2	VP
Fujilab_007	15,6	35,19	-	Comp 2	Comp 2	VP
Fujilab_008	14,03	25,92	-	Comp 2	Comp 2	VP
Fujilab_009	10,98	27,10	-	Comp 2	Comp 2	VP
Fujilab_010	18,71	33,54	-	Comp 2	Comp 2	VP
Fujilab_011	16,90	36,57	-	Comp 2	Comp 2	VP
Fujilab_012	10,72	35,75	-	Comp 2	Comp 2	VP
Fujilab_013	20,55	31,52	-	Comp 2	Comp 2	VP
Fujilab_014	18,21	36,01	-	Comp 2	Comp 2	VP
Fujilab_015	7,00	47,18	6.259.681.567	Comp 3	Comp 3	VP
Fujilab_016	7,37	58,87	3.744.293.087	Comp 3	Comp 3	VP
Fujilab_017	18,01	45,8	23.035.603.393	Comp 3	Comp 3	VP
Fujilab_018	17,80	40,3	14.182.501.275	Comp 3	Comp 3	VP
Fujilab_019	15,68	49,11	27.836.301.374	Comp 3	Comp 3	VP
Fujilab_020	15,53	45,79	14.414.476.537	Comp 3	Comp 3	VP
Fujilab_021	21,38	35,87	1.935.285.945	Comp 3	Comp 2	FALSO
Fujilab_022	18,43	47,63	14.313.873.933	Comp 3	Comp 3	VP
Fujilab_023	21,24	47,16	27.064.950.538	Comp 3	Comp 3	VP
Fujilab_024	12,95	54,24	43.139.223.436	Comp 4	Comp 4	VP
Fujilab_025	17.25	46,36	33.751.516.551	Comp 4	Comp 4	VP
Fujilab_026	19,91	38,05	21.545.711.504	Comp 4	Comp 3	FALSO
Fujilab_027	24,29	50,16	32.304.378.456	Comp 4	Comp 4	VP
Fujilab_028	25,25	58,94	39.830.797.808	Comp 4	Comp 4	VP

Número da imagem	% da mama ocupada por densidade 1	% da mama ocupada por densidade 2	SMA	Laudo Médico	Resultado com Classificador	Concordância
Fujilab_029	25,34	47,08	39.750.978.994	Comp 4	Comp 4	VP
Fujilab_030	19,82	38,25	37.967.355.279	Comp 4	Comp 4	VP
Fujilab_031	22,99	50,11	34.451.409.652	Comp 4	Comp 4	VP
Fujilab_032	15,22	49,25	37.615.251.149	Comp 4	Comp 4	VP
Fujilab_033	10,96	55,14	36.779.322.772	Comp 4	Comp 4	VP

A Tabela 5 apresenta os resultados de VP, FN, FP e Concordância encontrados na classificação para o FujiLab.

Tabela 5– Resultados de VP, FN e FP para o FujiLab.

Composições (BI-RADS)	Total imagens	VP	FN	FP	Concordância
COMP. I	3	3	0	0	100%
COMP. II	11	11	0	0	100%
COMP. III	9	8	1	0	88,8%
COMP. IV	10	9	1	0	90%
Total	33	31	2	0	

Utilizando os limiares (1 e 2) para o FujiLab conseguimos um acerto de *100%* na Composição I, *100%* na Composição II, *88,8%* na Composição III e *90%* na Composição IV. Com o índice de Falsos Negativos de *11%* para a Composição III e *10%* para a Composição IV. Não obtivemos Falsos Positivos para as Composições I, II, III e IV e Falsos Negativos para as Composições I e II.

Num total de 33 imagens apenas duas não foram classificadas pelo algoritmo na mesma composição determinada pelo médico. Conseguimos, portanto uma concordância de 93,9%.

As Figuras 66 e 67 apresentam as imagens que não foram classificadas pelo algoritmo de acordo com o médico. Essas duas imagens apresentam valores próximos do limiar escolhido para o corte. As imagens estão visualmente no limite entre as composições podendo pertencer a uma composição ou outra.

Uma divisão do BI-RADS que contemplaria mais composições talvez conseguiria uma melhor classificação.

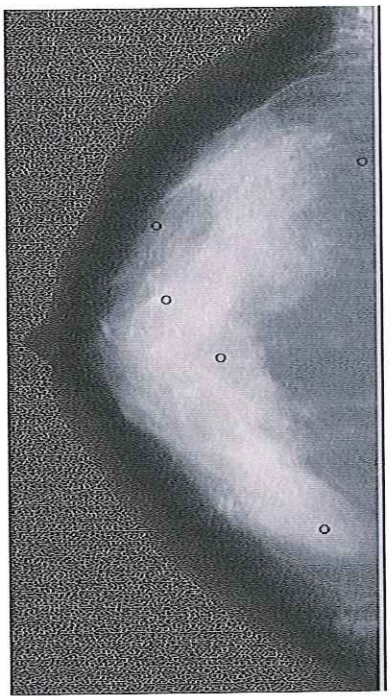


Figura 66 – Imagem do FujiLab da Composição III, classificada pelo algoritmo como Composição II.

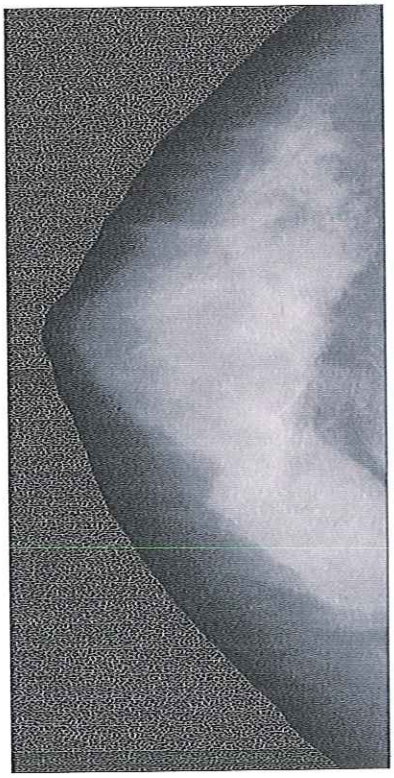


Figura 67 – Imagem do FujiLab da Composição IV, classificada pelo algoritmo como Composição III.

O método poderia ser relevante nesse caso, chamando a atenção do médico sobre o acerto de sua classificação.

5.3.2 APLICAÇÃO NO BANCO DE TREINAMENTO (FMRP-II)

Aplicamos o método nas 120 imagens do banco da Faculdade de Medicina de Ribeirão Preto (FMRP-II) que não continham no laudo a classificação no BI-RADS, utilizando os mesmos critérios adotados nas outras 154 imagens.

Solicitamos que um médico radiologista integrante da equipe que classificou as imagens que serviram de base para o método, analisasse os resultados.

A Tabela 6 apresenta os resultados obtidos com o classificador e a concordância entre o método e o laudo médico.

Tabela 6 – Relação das imagens com as porcentagens de densidade 1 e 2, os SMAs e a concordância entre o laudo e as regras.

Nome da imagem	% da mama ocupada por densidade1	% da mama ocupada por densidade2	SMA	Laudo	Resultado do Algoritmo	Concordância
fmrp049	8,8	32,6	1.392.370.331.744	COMP1	COMP1	VP
fmrp064	14,5	22,2	3.953.532.575.904	COMP1	COMP1	VP
fmrp067	11,1	33,5	1.529.909.252.824	COMP1	COMP1	VP
fmrp071	11,5	29,1	1.057.803.551.827	COMP1	COMP1	VP
fmrp083	11,4	31	1.503.204.054.072	COMP1	COMP1	VP
fmrp088	14,7	23,2	1.967.108.251.554	COMP2	COMP1	FALSO
fmrp089	11,9	22,6	1.296.276.350.411	COMP1	COMP1	VP
fmrp090	13,3	31,1	1.907.584.129.123	COMP1	COMP1	VP
fmrp107	13,4	28,7	2.208.454.312.122	COMP1	COMP1	VP
fmrp114	14,0	28,8	2.456.187.529.566	COMP1	COMP1	VP
fmrp115	13,9	25	1.719.886.678.282	COMP3	COMP1	FALSO
fmrp116	7,8	27,2	2.745.168.753.343	COMP1	COMP1	VP
fmrp119	9,7	31	1.525.276.571.157	COMP1	COMP1	VP
fmrp126	11,7	35,7	804.134.134.476	COMP1	COMP1	VP
fmrp135	14,2	26,4	2.201.575.155.293	COMP1	COMP1	VP
fmrp137	12,2	37,8	404.982.798.277	COMP4	COMP1	FALSO

Nome da imagem	% da mama ocupada por densidade1	% da mama ocupada por densidade2	SMA	Laudo	Resultado do Algoritmo	Concordância
fmrp139	13,6	25,4	1.729.984.302.704	COMP1	COMP1	VP
fmrp157	12,8	36,1	1.310.866.874.601	COMP1	COMP1	VP
fmrp166	13,5	19,8	1.792.085.080.967	COMP1	COMP1	VP
fmrp167	14,2	30,4	3.078.990.164.370	COMP1	COMP1	VP
fmrp177	8,3	29	31.614.095.162.600	COMP1	COMP1	VP
fmrp182	13,3	32,5	1.136.183.442.100	COMP1	COMP1	VP
fmrp185	14,8	36,1	281.394.829.838	COMP1	COMP1	VP
fmrp193	12,2	32,8	848.292.978.839	COMP1	COMP1	VP
fmrp194	8,9	37,3	914.010.642.728	COMP1	COMP1	VP
fmrp197	14,2	26,3	3.041.679.750.077	COMP1	COMP1	VP
fmrp202	14,9	37	669.688.853.318	COMP2	COMP1	FALSO
fmrp203	11,4	32,7	3.916.663.080.602	COMP1	COMP1	VP
fmrp206	13	21,9	1.312.147.217.891	COMP1	COMP1	VP
fmrp207	14,6	33,6	1.091.987.328.324	COMP2	COMP1	FALSO
fmrp213	11,8	29,4	1.690.091.218.156	COMP1	COMP1	VP
fmrp220	14,8	27,6	1.201.232.391.322	COMP1	COMP1	VP
fmrp222	14,1	32,2	2.561.247.492.476	COMP1	COMP1	VP
fmrp225	12,7	21,4	3.108.512.266.157	COMP2	COMP1	FALSO
fmrp229	12,5	37,9	3.089.962.401.286	COMP1	COMP1	VP
fmrp239	14,5	27,7	2.034.640.089.309	COMP1	COMP1	VP
fmrp240	9,7	21,9	1.813.994.835.620	COMP1	COMP1	VP
fmrp248	11,2	22	1.696.608.869.995	COMP1	COMP1	VP
fmrp253	12,7	26	54.693.904.817	COMP1	COMP1	VP
fmrp257	11,1	27,2	3.854.157.821.097	COMP1	COMP1	VP
fmrp263	8,8	18,9	3.122.995.725.290	COMP1	COMP1	VP
fmrp272	8,9	29,9	2.558.340.094.775	COMP1	COMP1	VP
fmrp277	12,3	21,1	442.995.163.993	COMP1	COMP1	VP
fmrp278	7,9	32,1	2.666.146.231.070	COMP1	COMP1	VP
fmrp279	12,8	34,8	2.872.275.082.155	COMP1	COMP1	VP
fmrp283	13,2	30	4.149.574.137.091	COMP1	COMP1	VP
fmrp284	13,7	24	868.867.496.855	COMP1	COMP1	VP
fmrp285	14,7	31,5	1.382.892.130.210	COMP1	COMP1	VP
fmrp286	10,1	32,4	3.296.614.707.111	COMP1	COMP1	VP
fmrp288	13,7	24,1	2.508.219.512.046	COMP1	COMP1	VP

Nome da imagem	% da mama ocupada por densidade1	% da mama ocupada por densidade2	SMA	Laudo	Resultado do Algoritmo	Concordância
fmrp291	13,8	31,9	3.092.500.073.232	COMP1	COMP1	VP
fmrp292	8,5	39,4	264.016.551.590	COMP4	COMP1	FALSO
fmrp297	7,7	19,4	2.479.769.303.470	COMP1	COMP1	VP
fmrp298	14,1	28,7	763.622.739.206	COMP1	COMP1	VP
fmrp299	15,0	26,3	1.015.157.570.990	COMP1	COMP1	VP
fmrp301	14,7	33,6	4.430.543.023.344	COMP1	COMP1	VP
fmrp303	14,1	31,2	1.989.058.639.429	COMP1	COMP1	VP
fmrp313	12,3	28,6	4.430.543.023.344	COMP1	COMP1	VP
fmrp315	9,5	26,7	1.365.735.800.838	COMP1	COMP1	VP
fmrp046	26,6	27,3	78.695.031.531	COMP4	COMP2	FALSO
fmrp054	21,6	30	793.624.650.935	COMP2	COMP2	VP
fmrp073	19,8	30,4	720.025.196.896	COMP2	COMP2	VP
fmrp078	17	28,7	864.954.809.932	COMP2	COMP2	VP
fmrp081	15,2	31,6	519.235.299.603	COMP2	COMP2	VP
fmrp086	24,9	30,8	554.449.156.992	COMP2	COMP2	VP
fmrp105	17,2	31,7	1.666.274.573.694	COMP2	COMP2	VP
fmrp111	17	25,4	860.218.874.172	COMP2	COMP2	VP
fmrp113	18,1	26,5	2.680.025.093.269	COMP2	COMP2	VP
fmrp117	22	32,6	2.161.529.013.004	COMP2	COMP2	VP
fmrp118	16,8	21,9	499.657.742.899	COMP2	COMP2	VP
fmrp132	21,5	29,4	656.486.050.998	COMP2	COMP2	VP
fmrp134	17,3	29,7	846.943.894.705	COMP2	COMP2	VP
fmrp136	17,8	28,7	1.499.467.749.846	COMP2	COMP2	VP
fmrp138	15,5	28,9	3.207.400.022.527	COMP2	COMP2	VP
fmrp149	16	28,6	890.994.217.289	COMP2	COMP2	VP
fmrp172	18	28,3	1.097.626.209.753	COMP2	COMP2	VP
fmrp173	22	27,1	2.646.183.353.566	COMP3	COMP2	FALSO
fmrp175	20,5	32,4	535.610.193.726	COMP2	COMP2	VP
fmrp189	18,6	28,5	911.841.369.892	COMP2	COMP2	VP
fmrp190	15,4	23,3	1.232.478.986.731	COMP2	COMP2	VP
fmrp191	19,3	30	1.254.383.586.774	COMP2	COMP2	VP
fmrp192	21,4	30	5.823.146.304.185	COMP2	COMP2	VP
fmrp205	15,6	25,2	526.314.948.190	COMP2	COMP2	VP
fmrp208	18,5	30,2	1.586.097.204.656	COMP2	COMP2	VP

Nome da imagem	% da mama ocupada por densidade1	% da mama ocupada por densidade2	SMA	Laudo	Resultado do Algoritmo	Concordância
fmrp209	15,6	26	1.350.090.366.600	COMP2	COMP2	VP
fmrp227	17,4	28,3	1.639.126.977.344	COMP2	COMP2	VP
fmrp228	16,8	27,3	5.438.797.482.756	COMP2	COMP2	VP
fmrp230	16,1	25,7	1.221.437.708.653	COMP2	COMP2	VP
fmrp244	17	31,2	3.517.285.416.900	COMP2	COMP2	VP
fmrp251	16,6	24,6	12.068.760.888.650	COMP2	COMP2	VP
fmrp259	18,9	32,8	2.757.900.310.456	COMP2	COMP2	VP
fmrp260	20,4	30,5	1.796.144.359.747	COMP2	COMP2	VP
fmrp264	17,4	27,4	1.774.514.469.559	COMP2	COMP2	VP
fmrp265	20,3	31,8	451.736.612.414	COMP3	COMP2	FALSO
fmrp270	20,1	31,2	540.218.796.480	COMP2	COMP2	VP
fmrp274	16	31,5	4.847.372.253.050	COMP2	COMP2	VP
fmrp276	15,7	26,7	2.063.798.902.442	COMP2	COMP2	VP
fmrp281	21,1	29,9	2.069.941.932.696	COMP3	COMP2	FALSO
fmrp290	16,1	30,2	3.621.595.712.312	COMP2	COMP2	VP
fmrp302	17,8	31,1	4.430.543.023.344	COMP2	COMP2	VP
fmrp055	28,2	44,3	155.246.742.724	COMP3	COMP3	VP
fmrp079	24,5	37,6	90.310.293.716	COMP3	COMP3	VP
fmrp103	22,2	43,6	160.229.726.105	COMP4	COMP3	FALSO
fmrp131	20,5	37,2	415.101.459.517	COMP3	COMP3	VP
fmrp179	18,9	39,9	189.917.915.631	COMP3	COMP3	VP
fmrp254	21,1	40,4	10.728.600.590	COMP3	COMP3	VP
fmrp043	17,7	38,5	639.484.307.800	COMP4	COMP4	VP
fmrp057	16,7	40,2	643.649.394.896	COMP4	COMP4	VP
fmrp061	15,5	35,7	1.050.994.528.900	COMP2	COMP4	FALSO
fmrp093	19,5	34	583.833.981.198	COMP4	COMP4	VP
fmrp104	27,1	40,2	767.649.939.974	COMP4	COMP4	VP
fmrp122	17,1	35,3	1.660.487.339.105	COMP4	COMP4	VP
fmrp151	18,2	34,9	885.090.087.810	COMP4	COMP4	VP
fmrp168	18,5	36,9	1.786.186.112.057	COMP4	COMP4	VP
fmrp199	20,5	37,5	1.959.069.365.056	COMP4	COMP4	VP
fmrp252	18,2	40,7	1.223.299.024.235	COMP4	COMP4	VP
fmrp280	20,1	33,5	7.398.183.178.494	COMP4	COMP4	VP
fmrp282	24,1	38	7.398.183.178.494	COMP4	COMP4	VP

Nome da imagem	% da mama ocupada por densidade1	% da mama ocupada por densidade2	SMA	Laudo	Resultado do Algoritmo	Concordância
fmrp294	19,8	33,9	542.753.696.911	COMP4	COMP4	VP
fmrp305	23,1	36,2	4.221.637.526.471	COMP4	COMP4	VP

A Tabela 7 apresenta os resultados de VP, FN, FP e Concordância encontrados na classificação para o FMRP-II.

Tabela 7– Resultados de VP, FN e FP para o FMRP-II.

Composições (BI-RADS)	Total imagens	VP	FN	FP	Concordância
COMP. I	59	52	7	0	88%
COMP. II	41	37	4	5	90,3%
COMP. III	6	5	1	4	83,3%
COMP. IV	14	13	1	3	93%
Total	120	107	13	12	

5.4 ESTABELECIMENTO DE UM BANCO DE IMAGENS PADRÃO

O método permite também selecionar as imagens mais representativas do banco utilizado pelo serviço do hospital. Para o banco da FMRP o histograma da Figura 40 mostra que as imagens mais representativas da Composição I são aquelas que apresentam porcentagem de densidade 1 em torno de 13%. A Figura 46 mostra que as imagens mais representativas da Composição II são aquelas que apresentam porcentagem de densidade 2 em torno de 33%. A Figura 50 mostra que as imagens mais representativas da Composição III são aquelas que apresentam SMA em torno de $2x10^{11}$. A Figura 51 mostra que as imagens mais representativas da Composição IV são aquelas que apresentam SMA em torno de $1x10^{12}$.

As imagens do banco padrão da FMRP são mostradas nas Figuras 68 a, b, c e d.

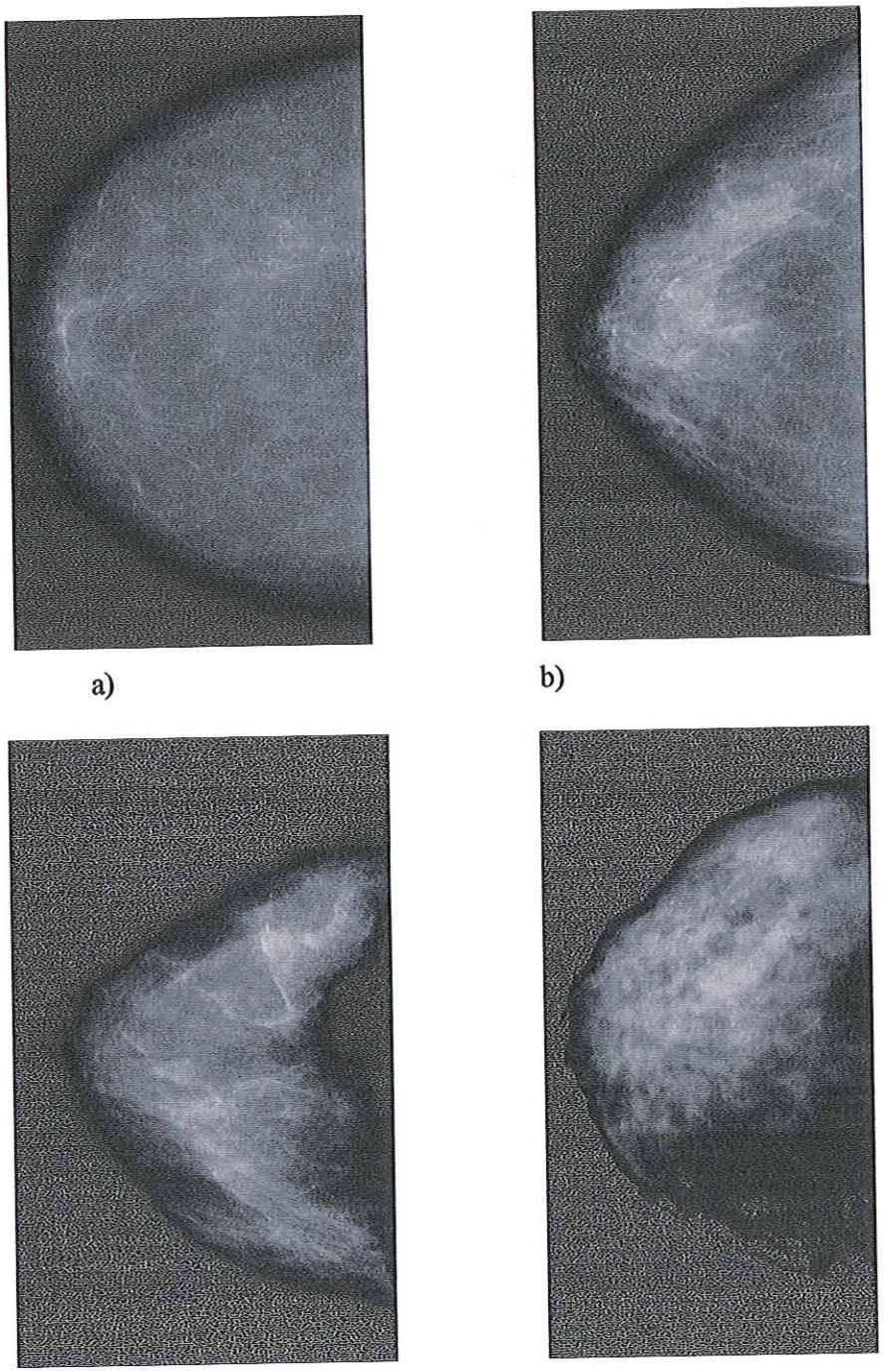


Figura 68 – **a)** imagem mais representativa da Composição I, **b)** imagem mais representativa da Composição II, **c)** imagem mais representativa da Composição III e **d)** imagem mais representativa da Composição IV considerando os critérios da equipe médica da FMRP.

Para estabelecer o banco padrão do FujiLab o histograma da Figura 54 mostra que as imagens mais representativas da Composição I são aquelas que apresentam porcentagem de densidade 1 em torno de 5%. A Figura 60 mostra que as imagens mais representativas da Composição II são aquelas que apresentam porcentagem de densidade 2 em torno de 37%. A Figura 64 mostra que as imagens mais representativas da Composição III são aquelas que apresentam SMA em torno de $5x10^9$. A Figura 65 mostra que as imagens mais representativas da Composição IV são aquelas que apresentam SMA em torno de $4x10^{10}$.

Para o FujiLab as imagens do banco padrão são mostradas nas Figuras 69 a, b, c e d.

EESC-USP Biblioteca Campus de S. Carlos

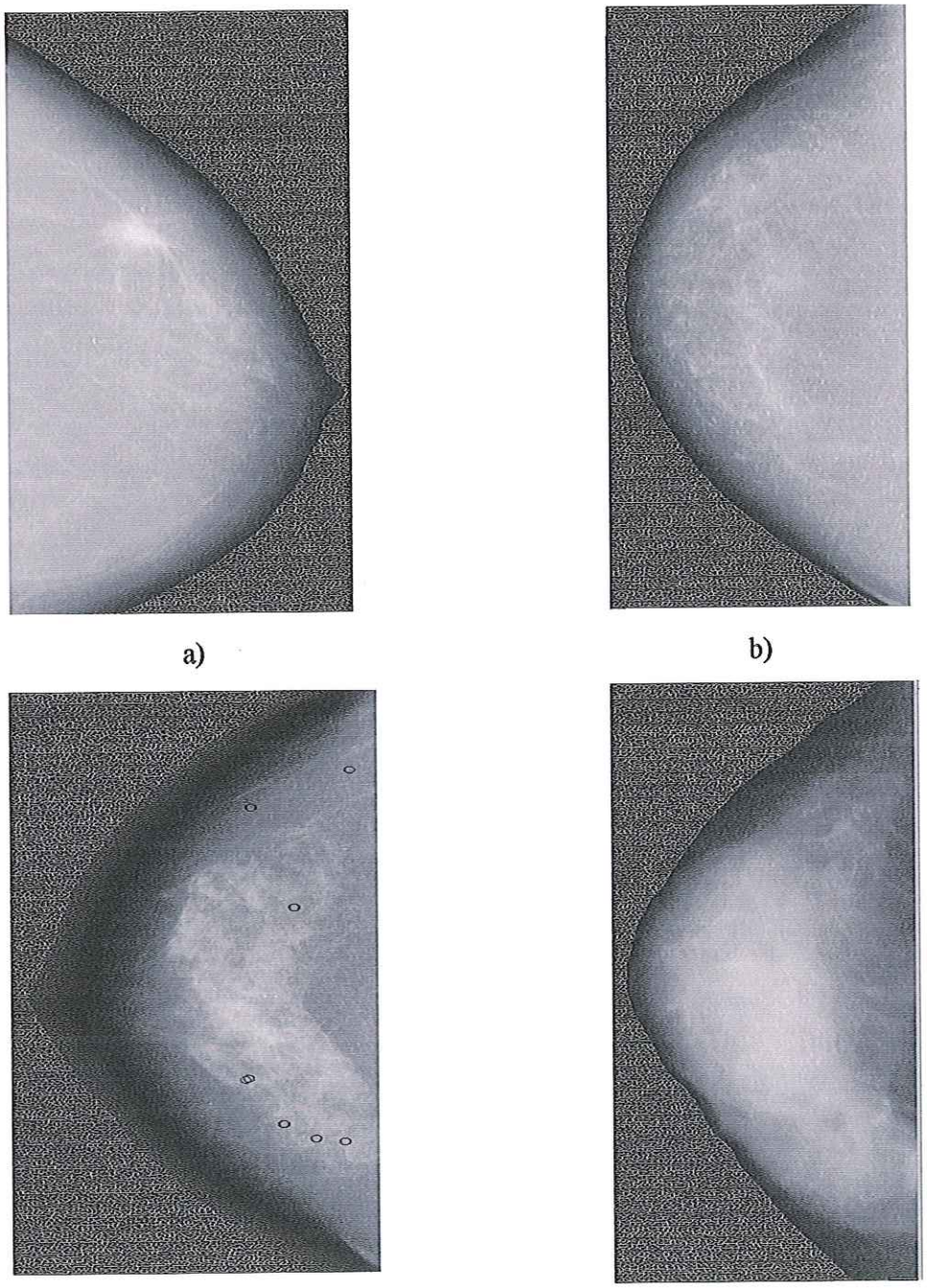


Figura 69 – **a)** imagem mais representativa da Composição I, **b)** imagem mais representativa da Composição II, **c)** imagem mais representativa da Composição III e **d)** imagem mais representativa da Composição IV do banco FujiLab.

5.5 DISCUSSÃO

Durante a execução do trabalho, pode-se perceber a grande variação de interpretação da densidade mamográfica entre os radiologistas. Entretanto, algumas imagens representam muito bem cada composição, ou seja, mesmo sendo analisadas por vários radiologistas, elas não causam duvidas e recebem sempre a mesma classificação. Então, acreditamos que a taxa de erro cometido pelo método desenvolvido pode ser menor dependendo do radiologista que classificou a imagem. Isto pode ser observado quando analisamos as imagens classificadas erroneamente. Essas imagens estavam bem próximas aos limiares entre as composições. Chang et al (2002) também observaram a dificuldade em separar as imagens pertencentes às composições III e IV, essa dificuldade também foi encontrada nesse trabalho.

Um outro fator que também interfere nos resultados é o pré-processamento, isso ficou nítido quando usamos as imagens do FujiLab, as quais eram pré-processadas enquanto que as da Faculdade de Medicina de Ribeirão Preto não.

Um outro problema foi referente à pequena quantidade de imagens pertencentes à Composição IV, por se tratar de uma pequena amostra os resultados não são muitos significativos.

CAPÍTULO 6

CONCLUSÕES

6.1 DETECÇÃO DA BORDA DA MAMA

A primeira contribuição trazida por este trabalho é um método de detecção automática da borda da mama, simples eficiente para os propósitos deste trabalho e de baixo custo operacional.

Para a detecção da borda, o método não necessita de marcação manual como feito por Tahoces et al (1995) e não perde algumas regiões como acontece com o método apresentado por Sivaramakrishina (2001), que além do mais tinha um alto custo computacional.

6.2 CLASSIFICAÇÃO NO PADRÃO DE DENSIDADE BI-RADS

Outra contribuição é o método para realizar a classificação computacional automática da composição mamária de acordo com o BI-RADS. Um método para esse fim já tinha sido desenvolvido por Chang et al (2002), mas os autores utilizaram um método de segmentação semi-automático, argumentando que as imagens não tem um padrão de comportamento previsível, ou seja, uma imagem é diferente da outra. No método que desenvolvemos o procedimento automático utiliza o contraste da imagem

do disco glandular sendo essa uma característica relativa que depende pouco da técnica utilizada.

Chang et al (2002) propuseram uma aproximação computacional que fornece uma quantificação da composição do tecido mamário usando estatística da imagem (média, desvio padrão, assimetria e Curtose), variação do método (máxima local em valor de "pixel" em uma determinada área), morfologia matemática (erosão e dilatação), textura (valor máximo do gradiente na região de interesse) e subtração gaussiana (cada imagem foi filtrada utilizando 5 filtros gaussianos). Os resultados apresentaram concordância de 87% com os especialistas. O método que desenvolvemos segue o raciocínio do medico quando ele faz a classificação, provavelmente é por isso que alcançamos 92% de concordância. Alem disso o método é de simples aplicação e custo computacional baixo.

6.3 CLASSIFICAÇÃO EM OUTROS PADRÕES

O método que desenvolvemos permite também classificar as mamografias em outros padrões apresentados na literatura.

Tahoces et al (1995) desenvolveram um método de classificação que utiliza o padrão de Wolfe e obtiveram uma concordância moderada para os padrões N1 e P2, razoável para P1 e boa para DY, mas não classificaram as imagens em função do padrão de dutos. Inicialmente pretendíamos também adotar esse padrão, mas constatamos que poucas imagens de um banco evidenciam os dutos, portanto acreditamos que um método que identifica automaticamente essa característica seria de pouca utilidade. O método que desenvolvemos também pode classificar as imagens nos padrões N1, P1, P2 e DY (sem dutos).

A classificação em imagem Homogênea ou não Boyd et al (1992, 1995), é interessante porque constatamos que essa característica é relevante no desempenho do CAD. A classificação das imagens no padrão Boyd et al (1992, 1995) poderá ser feita facilmente com os algoritmos que desenvolvemos.

Byng et al (1996) realizaram uma classificação de densidade mamográfica de maneira automática utilizando limiarização e técnicas de fractais para relacionar o brilho da imagem com a variação da textura da mama. Os autores utilizaram um classificador Bayesiano para separar as imagens no padrão de Boyd. Um teste de

concordância entre as classificações feitas pelos radiologistas e pelo método automático apresentou um resultado aproximadamente de 95%.

Karssemeijer (1998) desenvolveu um método para classificar 615 mamogramas em quatro categorias de densidades, <5, 5-25, 25-75 e 75 a 100. O autor calculou o desvio padrão e a assimetria do histograma, da parte interna da mama até a pele, usando o músculo peitoral como referência para a interpretação de regiões de densidades. O classificador KNN foi usado para classificar o padrão parenquimal. Como resultado em 67% dos casos houve uma concordância entre o classificador e a resposta dos radiologistas. Para mamogramas mais recentes, feitos após 1991, foi obtida uma concordância de 80%.

A proposta de Sivaramakrishna et al (2001) é um método completamente automático para estimar a densidade mamária, baseado em histograma envolvendo a hipótese que o objeto (mama) e o fundo da imagem podem ser distinguidos pelos seus valores de níveis de cinza. Os autores estimam a porcentagem da mama ocupada por densidade, mas perdem grande parte da área no processamento, eles não utilizam o método BI-RADS e comparam seus resultados com a técnica de técnica do "threshold" global, que não é eficiente.

Heine e Velthuizen (2000) apresentaram um método que permite a separação automatizada do tecido adiposo e do tecido fibroglandular. Para a detecção da densidade mamária os autores estimaram a variância global por comparações estatísticas, aplicando um procedimento de detecção baseado em análise de probabilidade $\chi 2$. Uma das etapas da classificação que utilizamos permite também separar o tecido fibro-glandular dos demais.

Vyborny et al (2000), analisaram o desempenho de determinado CAD em função da composição de densidade descrito no BI-RADS. Mostrando 78% de casos corretamente marcados pelo CAD para a Composição I, 77% para a Composição II, 74% para a Composição III e 71% para a Composição IV. Portanto, esse trabalho poderá ser utilizado no pré-processamento das imagens no CAD melhorando a taxa de acertos dos sistemas.

6.4 CONCLUSÕES FINAIS

O método desenvolvido permite classificar as imagens no padrão BI-RADS independentemente da subjetividade próprio à observação visual de quem organizou o banco ou da técnica utilizada. Os resultados foram superiores a 92% mesmo para bancos de imagens totalmente diferentes. Além do mais ressaltamos que esses resultados foram obtidos respeitando as interpretações de diferentes equipes médicas, (RODRIGUES, 2004a).

O método desenvolvido também poderá ser bastante útil para o ensino e treinamento médico na área de radiologia da mama. A equipe poderá estabelecer o padrão desejado e os estagiários serem treinados para classificar as densidades no padrão BI-RADS respeitando as particularidades locais de cada banco de imagem, (RODRIGUES, 2004b).

O método permite também uniformizar o banco de imagens dando sugestão aos médicos para uma classificação mais próxima do padrão utilizado no serviço. Por exemplo, se o médico considera como Composição 2 uma imagem que o classificador consideraria como Composição 1, o sistema poderá solicitar a confirmação do laudo.

O método não utiliza um processamento complicado, pois toda a programação foi feita no MATLAB, que possui um ambiente genérico de computação científica, linguagem de programação simples, compacta e muito próxima da linguagem matemática, assim como ferramentas bem conhecidas na área de processamento de imagens, portanto ele deve ser facilmente implementado e utilizado nos serviços radiológicos interessados na classificação automática de suas imagens.

6.5 PROPOSTA PARA TRABALHOS FUTUROS

- Averiguar a sensibilidade do classificador desenvolvido, pois ele reproduz a classificação feita pelo médico em quatro grupos e não foi testado quanto à sensibilidade em relação a pequenas variações dos tecidos glandulares e adiposos.

- Utilizar a metodologia desenvolvida no doutorado para detectar essa assimetria, pois o método independente da base de dados utilizada. A análise de assimetria mamária é um procedimento feito pelo radiologista quando analisa um mamograma. Assimetrias estruturais podem indicar uma possível presença de tumor (maligno ou benigno).
- Com essa metodologia desenvolvida, estudar a existência da relação entre a assimetria da mama com o diagnóstico *precoce* do câncer mamário.
- Também estudar com esta metodologia a relação que existe entre a assimetria mamária e o câncer de mama.
- Implementar essa metodologia desenvolvida no setor de radiologia hospitalar, ajudando na rapidez da análise de mamogramas.

REFERÊNCIAS

AMERICAN COLLEGE OF RADIOLOGY (1995). *Breast imaging reporting and data system (BI-RADS)*. 2nded. Reston: American College of Radiology.

AMOROSA, J.K. (2002). *Essentials of radiology*. Disponível em:<www.vasilev.com/medinter>. Acesso em: 04/setembro.

AKSOY, S.; HARALICK, R.M. (1999). Using texture in image similarity and retrieval. In: INTERNATIONAL WORKSHOP ON TEXTURE ANALYSIS IN MACHINE VISION, (2000), Finland. *Proceedings World Scientific,* M. Pietikainen, Ed. n. 20, p.111-117.

BLAND, K.I. et al. (1982). A clinic pathologic correlation of mammographic parenchymal patterns and associated risk factors for human mammary carcinoma. *Annals of Surgery*, Philadelphia, v.195, p.582–594.

SCHILLING, K. (2002). *Women's imaging update*. PalmBeach: Bocaradiology. Disponível em:<www.bocaradiology.com/womens_center/WIUpdate.html>. Acesso em: 01 Sept.

BONTRAGER, K. (1999). *Tratado de técnica radiológica e base anatômica*. 4.ed. Rio de Janeiro: Guanabara Koogan.

BOVIS, K.; SINGH, S. (2002). Classification of mammographic breast density using a combined classifier paradigm. In: 4th INTERNATIONAL WORKSHOP ON DIGITAL MAMMOGRAPHY, p. 177-180.

BOYD, N.F. et al. (1992). Relationship between mammographic and histological risk factors for breast cancer. *Journal of the National Cancer Institute*, Washington, v.84, n.15, p.1170–1179, August.

_____. (1995). Quantitative classification of mammographic densities and breast cancer risk: results from the Canadian National Breast Screening Study. *Journal of the National Cancer Institute*, Washington, v.87, n. 9, p.670–675.

BRISSON, J. et al. (1982a). Mammographic features of the breast and breast cancer risk. *American Journal of Epidemiology*, Baltimore, v.115, n. 3, p.428–437, Mar.

_____. (1982b). The relation of mammographic features of the breast to breast cancer risk factors. *American Journal of Epidemiology*, Baltimore, v.115, n.3, p.438–443, Mar.

BUCHANAN, J.B. et al. (1981). Selected prognostic variables for mammographic parenchymal variables. *Cancer*, Bruxelles, v.47, n.9, p.2135–2137, May.

BYNG, J.W. et al. (1994). The quantitative analysis of mammographic densities. *Physics Medicicine and Biology*, Bristol, v.39, n.10, p.1629-1638, Oct.

_____. (1996). Automated analysis of mammographic densities. *Physics Medicine and Biology*, Bristol, v.41, n.5, p.909-923, May.

BYRNE, C. (2002). Mammographic density: a breast cancer risk factor or diagnostic indicator. *Academic Radiology*, Oak Brook, v.9, n.3, p.252-253.

CHANG, Y.H. et al. (2002). Computerized assessment of tissue composition on digitized mammograms. *Academic Radiology*, Oak Radiology, v.9, n.8, p.899-905, August.

DANCE, D.R. (1996). Physical principles of breast imaging. In: INTERNATIONAL WORKSHOP ON DIGITAL MAMMOGRAPHY, 3, 1996, Chicago. *Proceedings 3th International Workshop on Digital Mammography*, p.427-430.

DOI, K. et al. (1991). Computer-aided diagnosis: present and future. In: A NEW horizon in medical physics and biomedical engineering. Amsterdam: Elsevier Science. p. 59-66.

_____. (1999). Computer-aided diagnosis in radiology: potential and pitfalls. *European Journal of Radiology*, Amsterdam, v.31, n.2, p.97-109, Aug.

EDWARDS, E. (1971). *Introdução à teoria de comunicação*. São Paulo: Cultrix.

EGAN, R.L.; MCSWEENEY, M.B. (1979). Mammographic parenchymal patterns and risk of breast cancer. *Radiology*, Syracuse, v.133, n.1, p.65–70, Oct.

EGAN, R.L.; MOSTELLER, R.C. (1977). Breast cancer mammography patterns. *Cancer*, Bruxelles, v.40, n.5, p.2087–2090, Nov.

ELMORE, J.G. et al. (1994). Variability in radiologists interpretation of mammograms. *The New England Journal of Medicine*, Walthom, v.331, n.2, p.1493-1499.

GIGER, M.L. (1998). Overview of computer-aided diagnosis in breast imaging. In: INTERNATIONAL WORKSHOP ON COMPUTER-AIDED DIAGNOSIS, 1998, Chicago, Amsterdam: Elsevier. p.167-176.

GONZALEZ, R.C.; WOODS, R.E. (2000). *Processamento de imagens digitais*. São Paulo: Edgard Blücher.

GRAVELLE, I.H. et al. (1986). A prospective study of mammographic parenchymal patterns and risk of breast cancer. *British Journal of Radiology*, London, v.59, n.701, p.487–491, May.

HAINLINE, S. et al. (1978). Mammographic patterns and risk of breast cancer. *American Journal of Roentgenology*, Springfield, v.130, n. 6, p.1157–1158, Jun.

HEINE, J.J.; MALHOTRA, P. (2002). Mammographic tissue, breast câncer risk serial image analysis and digital mammography, part 1- tissue and related risk factors. *Academic Radiology*, Oak Brook, v.9, n. 3, p.298-316, Mar.

HEINE, J.J.; VELTHUIZEN, R.P. (2000). The statistical methodology for mammographic density detection. *Medical Physics*, New York, v.27, n.12, p. 2644–2651, Dec.

HOMER, M. J. (1997). *Mammographic interpretation*: a pratical approach. Boston: McGraw-Hill.

HUO, Z. et al. (1995). Analisys of spiculation in the computerized classification of mammography masses. *Medical Physics*, New York, v.22, n.10, p.1569-1579, Oct.

_____. (1998). Automated computerized classification of malignant and benign mass lesion on digitized mammograms. *Academic Radiology*, Oak Brook, v.5, n. 3, p.155-168, Mar.

INSTITUTO NACIONAL DE CÂNCER. (2004). *Câncer de mama.* Disponível em:<www.inca.gov.br/conteudo_view.asp?id=336>. Acesso em: 12 abr.

INFOCANCER (2004). *Câncer de mama.* Disponível em:<www.infocancer.hpg.ig.com.br/mama.html>. Acesso em: 12 abr.

JANZON, L.; ANDERSSON, I.; PETERSSON, H. (1982). Mammographic patterns as indicators of risk of breast cancer: a cross-sectional population study. *Radiology*, v.143, n. 2, p.417–419, May.

JIANG,Y. et al. (1999). Improvement breast cancer diagnosis with computer-aided diagnosis. *Academic Radiology*, Oak Brook, v.6, n. 1, p.22-33, Jan.

KARSSEMEIJER, N. (1998). Automated classification of parenchymal patterns in mammograms. *Physics in Medicine Biology*, Bristol, v.43, n.2, p.365–378, Feb.

KITTLER, J.; ILLINGWORTH, J. (1986). Minimum error thresholding. *Pattern Recognition*, Oxford, v.19, n.1, p.41-47.

KROOK, P.M. (1978). Mammographic parenchymal patterns as risk indicators for incident cancer in a screening program: an extended analysis. *American Journal of Roentgenology*, Springfield, v.131, n. 6, p.1031–1035, Dec.

KROOK, P.M. et al. (1978). Mammographic parenchymal patterns as a risk indicator for prevalent and incident cancer. *Cancer*, Bruxelles, v.41, n.3, p.1093–1097, Mar.

MENDELL, L.; ROSENBLOOM, M.; NAIMARK, A. (1977). Are breast patterns a risk index for breast cancer?: a reappraisal. *American Journal of Roentgenology*, Springfield, v.128, n.4, p.547, April.

MOSKOWITZ, M.; GARTSIDE, P.; MCLAUGHLIN, C. (1980). Mammographic patterns as markers for high-risk benign breast disease and incident cancers. *Radiology*, Syracuse, v.134, n.2, p.293–295, Feb.

MURPHY JR., W.A.; DESTOUET, J.M.; MONSEES, B.S. (1990). Professional quality assurance for mammography screening programs. *Radiology*, Syracuse, v.175, n.2, p.319-320, May.

PINOTTI, J.A. (1991). *Compêndio de mastologia*. São Paulo: Manole.

RODRIGUES, S. C. M.; FRÈRE, A. F.; RODRIGUES, E. L. L (2004a). "Organização Automática de Bancos de Mamografias no Padrão de Densidade BI-RADS". III Congresso Latino-americano de Engenharia Biomédica, João Pessoa, setembro de 2004.

RODRIGUES, S. C. M.; FRÈRE, A. F.; RODRIGUES, E. L. L (2004b). "Análise dos Critérios Utilizados na Organização de Bancos de Mamografias no Padrão de Densidade BI-RADS". IX Congresso Brasileiro de Informática Médica, Ribeirão Preto-SP, novembro de 2004 (aceito).

ROEBUCK, E.J. (1982). The importance of mammographic parenchymal patterns. *British Journal of Radiology*, London, v.55, n.654, p.387–398, June.

SCHMIDT, R.A. (1998). The role of CAD in mammography and missed lesions. In: INTERNATIONAL WORKSHOP ON COMPUTER-AIDED DIAGNOSIS, 1998, Chicago. *Proceedings...* Amsterdam: Elsevier. p.177-184.

SHANNON, C.E. (1948). A mathematical theory of communication. *The Bell System Technical Journal*, New York, v.27, n.3, p.379-423.

SIVARAMAKRISHNA, R. et al. (2001). Automatic segmentation of mammographic density. *Academic Radiology*, Oak Brook, v.8, n.3, p.250–256, Mar.

SOUEN, J. (1998). A mamografia: acertos e desacertos. *Femina*, Rio de Janeiro, v.26, n.6, p.537-538, jul.

TABAR, L.; DEAN, P.B. (1982). Mammographic parenchymal patterns: risk indicator for breast cancer?. *JAMA*: Journal of the American Medical Association, Chicago, v.247, n. 2, p.185–189, January.

TAHOCES, P.G. et al. (1995). Computer assisted diagnosis: the classification of mammographic breast parenchymal patterns. *Physics in Medicine and Biology*, Bristol, v.40, n.1, p.103–117, Jan.

TAYLOR, P. et al. (1994). Measuring image texture to separate "difficult" from "easy" mammograms. *The British Journal of Radiology*, London, v.67, n.797, p.456-463, May.

TRAINA, A.J.M. et al. (2001). Visualização de dados em sistemas de bases de dados relacionais. In: SIMPÓSIO BRASILEIRO DE BANCO DE DADOS, 16., 2001, Rio de Janeiro. *Anais...* Rio de Janeiro: COPPE/UFRJ. p. 95-109.

TRAINA JR., C.; TRAINA, A.J.M.; FALOUTSOS, C. (1999). *FastMapDB user's manual*. Local: Carnegie Mellon University, School of Computer Science.

VERBEEK, A.L.M. et al. (1984). Mammographic breast pattern and the risk of breast cancer. *Lancet*, Barcelona, v. 323, n. 8377, p.591–593, Mar.

VYBORNY, C.J.; GIGER, M.L. (1994). Computer vision and artificial intelligence in mammography. *American Journal of Roentgenology*, Springfield, v. 162, n. 3, p. 699-708, Mar.

WELLINGS, S.R.; WOLFE, J.N. (1978). Correlative studies of the histological and radiographic appearance of the breast parenchyma. *Radiology*, Syracuse, v. 129, n. 2, p. 299–306, Nov.

WILKINSON, E. et al. (1977). Mammographic parenchymal pattern and risk of breast cancer. *Journal of the National Cancer Institute*, Washington, v.59, n.5, p.1397–1400, Nov.

WOLFE, J.N. (1969). The prominent duct pattern as an indicator of cancer risk. *Oncology*, Basel, v. 23, n. 2, p.149–158.

_____. (1976a). Risk for breast cancer development determined by mammographic parenchymal pattern. *Cancer*, Bruxelles, v.37, n.5, p.2486–2492, May.

_____. (1976b). Breast parenchymal patterns and their changes with age. *Radiology*, Syracuse, v. 121, n. 3, p. 545–552, Dec.

_____. (1977). Risk of developing breast cancer determined by mammography: process in clinical and biological research. Prog Clin Biol Res. v.12, p.223–238.

WOLFE, J.N.; SAFTLAS, A.F.; SALANE, M. (1987). Mammographic parenchymal patterns and quantitative evaluation of mammographic densities: a casecontrol study. *American Journal of Roentgenology*, Springfield, v.148, n. 6, p.1087–1092, Jun.

WONG, A.K.C.; SAHOO, P.K. (1989). A grey-level threshold selection method based on maximum entropy principle. IEEE Transactions on Systems, Man and Cybernetics, New York, v.SMC-19, n.4, p.866-871.